Laboratoire d'Analyse et Modélisation de Systèmes pour l'Aide à la Décision
UMR 7243

CAHIER DU LAMSADE

377

Février 2017

Majority Judgment vs.Majority Rule

Michel Balinski, Rida Laraki

Majority Judgment
vs.
Majority Rule*†

Michel Balinski
CNRS, École Polytechnique, Paris, France

and

Rida Laraki
CNRS, Université Paris-Dauphine,
PSL Research University, LAMSADE, Paris, France, and
Department of Economics, École Polytechnique, Paris, France

February 6, 2017

Abstract

The validity of majority rule in an election with but two candidates—and so also of Condorcet consistency—is challenged. Axioms based on evaluating candidates—paralleling those of K. O. May characterizing majority rule for two candidates based on comparing candidates—lead to another method, majority judgment, that is unique in agreeing with the majority rule on pairs of "polarized" candidates. It is a practical method that accommodates any number of candidates, avoids both the Condorcet and Arrow paradoxes, and best resists strategic manipulation. It may also be viewed as a "solution" to Dahl's (reformulated) intensity problem in that an intense minority sometimes defeats an apathetic majority.

Key words: majority, measuring, ranking, electing, Condorcet consistency, domination paradox, tyranny of majority, intensity problem, majority-gauge, strategy-proofness, polarization.

Subject classification: Games/group decisions: voting. Government: politics.

*R. Laraki's work was supported by grants administered by the French National Research Agency as part of the Investissements d'Avenir program (Idex [Grant Agreement No. ANR-11-IDEX-0003-02/Labex ECODEC No. ANR-11- LABEX-0047]) and ANR-14-CE24-0007-01 CoCoRIco-

†The authors are deeply indebted to Jack Nagel for very helpful comments and suggestions, particularly with regard to Dahl's intensity problem.

1 Introduction

Elections *measure*.

Voters express themselves, a rule amalgamates them, the candidates' scores—measures of support—determine their order of finish, and the winner.

Traditional methods ask voters to express themselves by *comparing* them:

- ticking one candidate at most (majority rule, first-past-the-post) or several candidates (approval voting), candidates' total ticks ranking them;
- rank-ordering the candidates (Borda rule, alternative vote or preferential voting, Llull's rule, Dasgupta-Maskin method, ...), differently derived numerical scores ranking them.

However, no traditional method contains a hint about how to measure the support of *one* candidate.[1] Instead, the theory of voting (or of social choice) elevates to a basic distinguishing axiom the faith that in an election between *two* candidates majority rule is the only proper rule. Every traditional method of voting tries to generalize majority rule to more candidates and reduces to it when there are only two candidates. Yet majority rule decides unambiguously only when there are two candidates: it says nothing when there is only one and its generalizations to three or more are incoherent.

Using the majority rule to choose one of two candidates is widely accepted as infallible: since infancy who in this world has not participated in raising their hand to reach a collective decision on two alternatives? A. de Tocqueville believed "It is the very essence of democratic governments that the dominance of the majority be absolute; for other than the majority, in democracies, there is nothing that resists" ([72], p. 379)[2]. Judging from W. Sadurski's assertion—"The legitimating force of the majority rule is so pervasive that we often do not notice it and rarely do we question it: We usually take it for granted" ([65], p. 39)—that conviction seems ever firmer today. Students of social choice well nigh universally accept majority rule for choosing between two alternatives, and much of the literature takes *Condorcet consistency*—that a candidate who defeats each of the others separately in majority votes must be the winner (the *Condorcet-winner*)—to be either axiomatic or at least a most desirable property.

Why this universal acceptance of majority rule for *two* candidates? First, the habit of centuries; second, K.O. May's [51] axiomatic characterization; third, the fact that it is strategy-proof or incentive compatible, i.e., that a voter's optimal strategy is to vote truthfully; fourth, the Condorcet jury theorem [25].

Regrettably, as will be shown, the majority rule can easily go wrong when voting on but *two* candidates. Moreover, asking voters to compare candidates when there are three or more inevitably invites the Condorcet and Arrow paradoxes. The first shows that it is possible for a method to yield a non-transitive

[1] A significant number of elections, including U.S. congressional elections, have but one candidate.

[2] Our translation of: "Il est de l'essence même des gouvernements démocratiques que l'empire de la majorité y soit absolu; car en dehors de la majorité, dans les démocraties, il n'y a rein qui résiste."

order of the candidates, so no winner [25]. It occurs, e.g., in elections [46], in figure skating ([5] pp. 139-146, [9]), in wine-tasting [7]. The second shows that the presence or absence of a (often minor) candidate can change the final outcome among the others. It occurs frequently, sometimes with dramatic global consequences, e.g., the election of George W. Bush in 2000 because of the candidacy of Ralph Nader in Florida; the election of Nicolas Sarkozy in 2007 although all the evidence shows François Bayrou, eliminated in the first-round, was the Condorcet-winner.

How are these paradoxes to be avoided? Some implicitly accept one or the other (e.g., Borda's method, Condorcet's method). Others believe voters' preferences are governed by some inherent restrictive property. For example, Dasgupta and Maskin [28, 27] appeal to the idea that voters' preferences exhibit regularities—they are "single-peaked," meaning candidates may be ordered on (say) a left/right political spectrum so that any voter's preference peaks on some candidate and declines in both directions from her favorite; or they satisfy "limited agreement," meaning that for every three candidates there is one that no voter ranks in the middle. Another possibility is the "single crossing" restriction (e.g., [13, 62]): it posits that both candidates and voters may be aligned on a (say) left/right spectrum and the more a voter is to the right the more she will prefer a candidate to the right. Such restrictions could in theory reflect sincere patterns of preference in some situations; nevertheless in *all* such cases voters could well cast strategic ballots of a very different stripe.

All the sets of ballots that we have studied show that *actual* ballots are in no sense restricted. There are many examples. (1) An approval voting experiment was conducted in parallel with the first-round of the 2002 French presidential election that had 16 candidates. 2,587 voters participated and cast 813 different ballots: had the preferences been single-peaked there could have been at most 137 sincere ballots ([5] p. 117). (2) A Social Choice and Welfare Society presidential election included an experiment that asked voters to give their preferences among the three candidates: they failed to satisfy any of the above restrictions [21, 64]. (3) A majority judgment voting experiment was conducted in Orsay in parallel with the first-round of the 2007 French presidential election that had 16 candidates ([5] pp. 112-115, [6]). 1,733 valid ballots expressed the opinions of voters according to a scale, so their preferences could be deduced. 1,705 ballots were different and many seemed devoid of any discernible political explanation. For example, there were three major candidates, a rightist (Sarkozy), a centrist (Bayrou) and a leftist (Royal). 14.3% evaluated Sarkozy and Royal the same, 4.1% gave both their highest evaluation; 17.9% evaluated Sarkozy and Bayrou the same, 10.6% gave both their highest evaluation; 23.3% evaluated Bayrou and Royal the same, 11.7% gave both their highest evaluation; and 4.8% evaluated all three the same, 4.1% gave all three their highest evaluation. (4) A recent study of voting in Switzerland uses principal component analysis to show that at least three dimensions are necessary to map voters' preferences, so the one-dimensional single-peaked condition cannot be met [34]. Real ballots—the voters' true opinions or strategic choices—eschew ideological divides.

One of the reasons that things go so wrong is that the majority rule on two candidates and first-past-the-post on many candidates measure badly. Voters are charged with nothing other than to tick the name of at most one candidate. Thus voters are not given the means to express their opinions. A striking example of poor measurement is the 2002 French presidential election. There were 16 candidates, among them J. Chirac (the outgoing rightist President), L. Jospin (the outgoing socialist Prime Minister), and J.-M. Le Pen (the extreme right leader). France expected a run-off between Chirac and Jospin, and most polls predicted a Jospin victory. Chirac had 19.88% of the votes, Le Pen 16.86%, so Jospin's 16.18% eliminated him (another instance of Arrow's paradox). Chirac's 82.2% in the run-off with Le Pen in no way measured his support in the nation. Another example is the 2007 French presidential election (already mentioned) that saw the Condorcet-winner Bayrou eliminated.

All of this, we conclude, shows that the domain of voters' preferences is in real life unrestricted. An entirely different approach is needed that gives voters the means to better express their opinions.

This has motivated the development of majority judgment [5, 9] based on a different paradigm: instead of comparing candidates, voters are explicitly charged with a solemn task of expressing their opinions precisely by evaluating the merit of every candidate in an ordinal scale of measurement or language of grades. Thus, for example, the ballot in a presidential election could ask:

> "Having taken into account all relevant considerations, I judge, in conscience, that as President of the United States of America each of the following candidates would be:"

The language of grades constituting the possible answers may contain any number of grades though in elections with many voters five to seven have proven to be good choices. Two scales that have been used are:

Outstanding, Excellent, Very Good, Good, Fair, Poor, To Reject;

and

a *Great*, a *Good*, an *Average*, a *Poor*, a *Terrible* President.

The method then specifies that majorities determine the electorate's evaluation of each candidate and the ranking between every pair of candidates—necessarily transitive—with the first-placed among them the winner.

This article presents a completely new argument for majority judgment that responds to the objection that it does not agree with the majority rule on two candidates, i.e., that it is not Condorcet-consistent. It first gives a new description of majority judgment that shows it naturally emerges from the majority principle when candidates are assigned grades. It shows that majority rule on two candidates is not incontestable, it is open to serious error because it admits the "domination paradox": it may elect a candidate whose grades are dominated by another candidate's grades. Then, starting from the basic principles that justify majority rule (May's axioms), majority judgment is characterized

as the unique method based on evaluations that satisfies those same principles, avoids the Condorcet and Arrow paradoxes, and coincides with majority rule on two candidates when the electorate is "polarized," that is, when the higher (the lower) a voter evaluates one candidate the lower (the higher) she evaluates the other, so there can be no consensus. Majority rule does not violate domination on polarized candidates, and since it is incentive compatible on that domain, so is majority judgment, and that is when voters are most tempted to manipulate. It is shown that the condition is necessary to combat manipulation.

The infallibility of majority rule on two candidates has been challenged before. R. A. Dahl charged: "By making 'most preferred' equivalent to 'preferred by most' we deliberately bypassed a crucial problem: What if the minority prefers its alternative much more passionately than the majority prefers a contrary alternative? Does the majority principle still make sense? This is the *problem of intensity*. ... [W]ould it be possible to construct rules so that an apathetic majority only slightly preferring its alternative could not override a minority strongly preferring its alternative?" ([26], pp. 90, 92, our emphasis). He proposed no such rules: majority judgment meets his objectives to the extent that they may be met, so provides a solution to his problem.

Majority judgment has been criticized by the proponents of the traditional paradigm in articles, reviews, and talks. Most are answered in this paper.[3] Majority judgment's characterization in terms of polarized candidates addresses the principal charge, namely, that it is not Condorcet-consistent: it *is* Condorcet-consistent when majority rule makes most sense, namely, when the domination paradox cannot occur and voters are most tempted to manipulate. The last sections address the various other objections, summarize its good properties, and illustrate them with new real uses of majority judgment.

Our aim has been to develop a practical, easily usable method to rank candidates and competitors. Majority judgment avoids the major drawbacks of the traditional theory and combats manipulation, yet agrees with majority rule when it makes sense. We believe its proven properties together with the evidence of its use in practice and in experiments shows—to borrow an expression used in [28, 27]—that it "works well."

2 Evaluating to rank

Walter Lippmann wrote in 1925,

> "But what in fact is an election? We call it an expression of the popular will. But is it? We go into a polling booth and mark a cross on a piece of paper for one of two, or perhaps three or four names. Have we expressed our thoughts ...? Presumably we have a number of thoughts on this and that with many buts and ifs and ors. Surely the cross on a piece of paper does not express them.... [C]alling a

[3]The no-show paradox charge is answered in a forthcoming paper.

vote the expression of our mind is an empty fiction." ([49], pp. 106-107)

Except for elections the practice in virtually every instance that ranks entities is to evaluate each of them (see [5], chapters 7 and 8). The *Guide Michelin* uses stars to rate restaurants and hotels. Competitive diving, figure skating, and gymnastics use carefully defined number scales. Wine competitions use words: *Excellent, Very Good, Good, Passable, Inadequate, Mediocre, Bad.* Students are graded by letters, numbers, or phrases. Pain uses sentences to describe each element of a scale that is numbered from 0 ("Pain free") to 10 ("Unconscious. Pain makes you pass out."), a 7 defined by "Makes it difficult to concentrate, interferes with sleep. You can still function with effort. Strong painkillers are only partially effective."

In the political sphere polls, seeking more probing information about voter opinion, also ask more. Thus a Harris poll: "... [H]ow would you rate the overall job that President Barack Obama is doing on the economy?" Among the answers spanning 2009 to 2014 were those given in Table 1.

	Excellent	*Pretty good*	*Only fair*	*Poor*
March 2009	13%	34%	30%	23%
March 2011	5%	28%	29%	38%
March 2013	6%	27%	26%	41%

Table 1. Measures evaluating the performance of Obama on the economy [44].

It is not only natural to use measures to evaluate one alternative—a performance, a restaurant, a wine, or a politician—but necessary.

To be able to measure the support a candidate enjoys, a voter must be given the means to express her opinions or "feelings." To assure that voters are treated equally, voters must be confined to a set of expressions that is shared by all. To allow for meaningful gradations—different shades ranging from very positive, through mediocre, to very negative—the gradations must faithfully represent the possible likes and dislikes. Such finite, ordered sets of evaluations are common and accepted in every day life. Call it a *scale* or *common language of grades* Λ linearly ordered by $\succ$.

An electorate's *opinion profile* on a candidate is the set of her, his, or its grades $\alpha = (\alpha_1, \dots \alpha_n)$, where $\alpha_j \in \Lambda$ is voter j's evaluation of the candidate.

Since voters must have equal voices, only the grades can count: which voter gave what grade should have no impact on the electorate's global measure of a candidate. The number of times each grade occurs or their percentages (as in Table 1) is called the candidates's *merit profile* (to distinguish it from an *opinion profile* that specifies the grade given the candidate by each of the judges). A candidate's merit profile will always be written from the highest grades on the left down to the lowest on the right.

2.1 A new description of majority judgment

Majority judgment naturally emerges from the majority principle.

What is the electorate's *majority* opinion of a candidate with grades $\alpha = (\alpha_1, \ldots, \alpha_n)$? An example best conveys the basic idea. In the spring of 2015 majority judgment was used by a jury of six (J_1 to J_6) at LAMSADE, Université Paris-Dauphine to rank six students (A to F) seeking fellowships to prepare Ph.D dissertations. The jury agreed their solemn task was to evaluate the students and chose the language of grades

Excellent, Very Good, Good, Passable, Insufficient.

The opinion profile of candidate C was

	J_1	J_2	J_3	J_4	J_5	J_6
C:	*Passable*	*Excellent*	*Good*	*V. Good*	*V. Good*	*Excellent*

Since voters or judges must have equal voices, only the grades count, not which voter or judge gave what grade. Accordingly the number of times each grade occurs or their percentages is called the candidate's *merit profile*, always written from the highest grades on the left down to the lowest on the right. C's merit profile was

C:	*Excellent*	*Excellent*	*V. Good*	↕	*V. Good*	*Good*	*Passable*

The middle of C's grades in the merit profile is indicated by a two-sided arrow.

There is a majority of $\frac{6}{6}$—unanimity—for C's grade to be at most *Excellent* and at least *Passable*, or *for* [*Excellent*, *Passable*]; a majority of $\frac{5}{6}$ for C's grade to be at most *Excellent* and at least *Good*, or *for* [*Excellent*, *Good*]; and a majority of $\frac{4}{6}$ for C's grade to be at most *Very Good* and at least *Very Good*, or *for* [*Very Good*, *Very Good*]. The closer the two—equally distant from the middle—are to the middle, the closer are their values, so the more accurate is the majority decision. When the two are equal it is the "majority-grade" and it suffices to specify one grade. If n is odd there is certain to be an absolute majority for a single grade; if n is even and the middlemost grades are different (very rare in a large electorate) there is a majority consensus for two grades.

In general, letting $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ be a candidate A's set of n grades written from highest to lowest, $\alpha_i \succeq \alpha_{i+1}$ for all i, there is a majority of (at least) $\frac{n-k+1}{n}$ for A's grade to be at most α_k and at least α_{n-k+1}, for all $1 \leq k \leq (n+1)/2$. Call this the $(\frac{n-k+1}{n})$-*majority for* $[\alpha_k, \alpha_{n-k+1}]$. When $k > h$ the two grades of the $(\frac{n-k+1}{n})$-majority are closer together than (or the same as) those of the $(\frac{n-h+1}{n})$-majority: they are *more accurate*.

A measure of A's global merit is the most accurate possible majority decision on A's grades. C's majority-grade is *Very Good*, Obama's majority-grade in each of the evaluations of his performance on the economy (Table 1) is *Only fair*.

How does a majority of an electorate rank candidates having sets of grades? For example, how does it rank two distributions of grades of Obama's performance at different dates (Table 1)? Had the March 2011 or 2013 distributions been identical to that on March 2009 the electorate would have judged the performances to be the same. In fact, Obama's March 2009 evaluations "dominate"

those of the same month in 2011 and 2013 and so the electorate clearly ranks it highest; but it is not clear how to compare the evaluations in 2011 and 2013.

In general, a candidate A's merit profile $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ *dominates* B's merit profile $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ (both written from highest to lowest) when $\alpha_i \succeq \beta_i$ for all i and $\alpha_k \succ \beta_k$ for at least one k (equivalently, when A has at least as many of the highest grade as B, at least as many of the two highest grades,..., at least as many of the k highest grades for all k, and at least one "at least" is "more").[4] Any reasonable method of ranking should *respect domination*: namely, evaluate one candidate above another when that candidate's grades dominate the other's. Surprisingly, some methods do not (as will be seen).

With m candidates the basic input is an electorate's *opinion profile*: it gives the grades assigned to every candidate by each voter and may be represented as a matrix $\alpha = (\alpha_{ij})$ of m rows (one for each candidate) and n columns (one for each voter), α_{ij} the grade assigned to candidate i by voter j. Table 2a gives the LAMSADE Jury's opinion profile. The *preference profile* of the traditional theory—voters' rank-orderings of the candidates—may be deduced from the opinion profile whenever the language of grades is sufficient for a voter to distinguish between any two candidates when he evaluates their merit differently. Thus, for example, J_1's preferences are $A \approx B \succ D \approx F \succ E \succ C$. Note that no judge used all five grades even though there were six candidates.

	J_1	J_2	J_3	J_4	J_5	J_6
A:	*Excellent*	*Excellent*	*V. Good*	*Excellent*	*Excellent*	*Excellent*
B:	*Excellent*	*V. Good*	*V. Good*	*V. Good*	*Good*	*V. Good*
C:	*Passable*	*Excellent*	*Good*	*V. Good*	*V. Good*	*Excellent*
D:	*V. Good*	*Good*	*Passable*	*Good*	*Good*	*Good*
E:	*Good*	*Passable*	*V. Good*	*Good*	*Good*	*Good*
F:	*V. Good*	*Passable*	*Insufficient*	*Passable*	*Passable*	*Good*

Table 2a. Opinion profile, LAMSADE Jury.[5]

To see how majority judgment (MJ) ranks the candidates of the LAMSADE Jury consider the corresponding merit profile given in two equivalent forms: extensively (Table 2b) and by counts of grades (Table 2c).

A:	*Excellent*	*Excellent*	***Excellent***	***Excellent***	*Excellent*	*V. Good*
B:	*Excellent*	*V. Good*	***V. Good***	***V. Good***	*V. Good*	*Good*
C:	*Excellent*	*Excellent*	***V. Good***	***V. Good***	*Good*	*Passable*
D:	*V. Good*	*Good*	***Good***	***Good***	*Good*	*Passable*
E:	*V. Good*	*Good*	***Good***	***Good***	*Good*	*Passable*
F:	*V. Good*	*Good*	***Passable***	***Passable***	*Passable*	*Insufficent*

Table 2b. Merit profile (extensive), LAMSADE Jury.

[4] In other contexts domination is called first-order stochastic dominance.

[5] The order of the students is chosen to coincide with the MJ-ranking for clarity.

The $\frac{4}{6}$-majorities are indicated in bold in Tables 2b,c. A's $\frac{4}{6}$-majority dominates all the others, so A is the MJ-winner; B's and C's dominate the remaining candidates, so are next in the MJ-ranking; and D and E—tied in the MJ-ranking since their sets of grades are identical—follow in the MJ-ranking since their $\frac{4}{6}$-majorities dominate F's. How are B and C to be compared (Table 2d)?

	Excellent	*Very Good*	*Good*	*Passable*	*Insufficient*
A:	**5**	1			
B:	1	**4**	1		
C:	2	**2**	1	1	
D:		1	**4**	1	
E:		1	**4**	1	
F:		1	1	**3**	1

Table 2c. Merit profile (counts), LAMSADE Jury.

Their $\frac{4}{6}$-majorities are identical but their $\frac{5}{6}$-majorities (indicated by square brackets in Table 2d) differ: B's is for [*Very Good, Very Good*] and C's for [*Excellent, Good*]. Since neither pair of grades dominates the other and there is more consensus[6] for B's grades than for C's, MJ ranks B above C. Thus the MJ ranking is $A \succ_{MJ} B \succ_{MJ} C \succ_{MJ} D \approx_{MJ} E \succ_{MJ} F$.

B:	*Excellent*	[***V. Good***	***V. Good***	↕	***V. Good***	***V. Good***]	*Good*
C:	*Excellent*	[***Excellent***	***V. Good***	↕	***V. Good***	***Good***]	*Passable*

Table 2d. Merit profile, B and C, LAMSADE Jury.

In general, if A's grades are $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and B's $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$, both written from highest to lowest, suppose the most accurate majority where the candidates differ is the $(\frac{n-k+1}{n})$-majority for $[\overline{\alpha}, \underline{\alpha}] \neq [\overline{\beta}, \underline{\beta}]$. Thus for A, $\overline{\alpha} = \alpha_k$ and $\underline{\alpha} = \alpha_{n-k+1}$, similarly for B, and $\alpha_j = \beta_j$ for $k < j < n-k+1$. Comparing them, A's *middlemost block* of grades is $[\alpha_k, \alpha_{k+1}, \ldots, \alpha_{n-k}, \alpha_{n-k+1}]$ and B's is $[\beta_k, \beta_{k+1}, \ldots, \beta_{n-k}, \beta_{n-k+1}]$.

The *majority-ranking* $\succeq_{MJ}$ ranks A above B when (a) A's middlemost block dominates B's or (b) A's middlemost block is more consensual than B's:

$$A \succ_{MJ} B \text{ when } \begin{cases} \text{(a) } \overline{\alpha} \succeq \overline{\beta} \text{ and } \underline{\alpha} \succeq \underline{\beta}, \text{ with at least one } \succeq \text{ strict, or} \\ \text{(b) } \overline{\beta} \succ \overline{\alpha} \succeq \underline{\alpha} \succ \underline{\beta}; \end{cases} \tag{1}$$

otherwise, their sets of grades are identical and $A \approx_{MJ} B$. The most accurate majority is either for a single grade called the *majority-grade*, or, when n is even, it may be for two grades (very rare in a large electorate).

It is immediately evident that MJ respects domination; moreover, one candidate is necessarily ranked above another unless their grade distributions are

[6] In other contexts this is called second-order stochastic dominance (favoring less "risk.") It is related to Hammond's equity principle [43] and implied by the Rawlsian criterion, B's minimum being above C's.

exactly the same. It is simple to show that MJ gives a transitive ordering when there are more than two candidates.

It is of interest to contrast the MJ ranking with the rankings obtained by traditional methods based on comparisons, so on the preference profile. *Condorcet's method* is the majority rule whenever the result is transitive. *Borda's method* ranks the candidates according to the average of each candidate's votes against all others, their *Borda scores*. Assuming a higher grade for one candidate against another means a preference for him and a tied grade indifference (so the scale is sufficient to express preferences and indifferences), and counting an indifference a $\frac{1}{2}$-win, Table 2e gives the pertinent data (e.g., A's row shows A defeats C with 5 votes—4 preferences, 2 indifferences).

	A	C	B	D	E	F	Borda score
A	–	5	5	6	5.5	6	5.5
C	1	–	3.5	5	4	5	3.7
B	1	2.5	–	5.5	5	6	4.0
D	0	1	0.5	–	3.5	5	2.0
E	0.5	2	1	2.5	–	4	2.0
F	0	1	0	1	2	–	0.8

Table 2e. Face-to-face majority rule votes, LAMSADE Jury.

The Condorcet ranking differs from the MJ ranking, $A \succ_{Condo} C \succ_{Condo} B \succ_{Condo} D \succ_{Condo} E \succ_{Condo} F$. On the other hand the Borda ranking is identical to the MJ ranking, $A \succ_{Borda} B \succ_{Borda} C \succ_{Borda} D \approx_{Borda} E \succ_{Borda} F$. Here the majority rule between D and E makes D the winner (2 preferences, 3 indifferences) though they have identical sets of grades; it might with equal chance have gone the other way with the same set of grades.

2.2 Majority judgment with many voters

When there are many voters simpler arithmetic is almost always sufficient to determine the MJ ranking. This is due to two facts: the most accurate majority decision concerning a candidate is generically a single grade—the *majority-grade*—and (1b) almost never occurs, so it suffices to detect when a difference in grades first occurs according to (1a). An example shows why this simplification works.

Terra Nova, a Parisian think-tank, sponsored a national presidential poll carried out by OpinionWay April 12-16, 2012 (just before the first-round of the election on April 22) to compare MJ with other methods. 993 participants voted with MJ and also according to usual practice—first-past-the-post (FPP) among all ten candidates, followed by a MR run-off between every pair of the expected five leaders in the first-round. Since the results of FPP varied slightly from the actual national percentages on election day (up to 5%) a set of 737 ballots was found for which those tallies are closely matched and are presented here.

An important practical aspect of MJ—that has theoretical implications discussed in Section 9.3—needs to be repeated. An MJ ballot should pose a precise

question in asking for a voter's response that depends on the particular application. For this poll the question was: "As President of France, in view of all relevant considereations, I judge, in conscience, that each of these candidates would be:" to which the voter is asked to answer with one of the available evaluations. In the case of the LAMSADE Jury the question was clearly posed when the judges chose the scale of grades.

To begin consider one candidate's merit profile ("H" for Hollande). 50% of the grades are to the left of the middle, 50% are to the right:

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	↕	*Good*	*Fair*	*Poor*	*To Reject*
H	12.48%	16.15%	16.42%	**4.95%**	↕	**6.72%**	14.79%	14.25%	14.24%

Hollande's majority-grade is *Good* because there is a $(50+\epsilon)$%-majority[7] for [*Good, Good*] for every ϵ, $0 < \epsilon \leq 4.95$. Similarly, there is a $(54.95+\epsilon)$%-majority for [*Very Good, Good*] for every ϵ, $0 < \epsilon \leq 1.77$, and a $(56.72+\epsilon)$%-majority for [*Very Good, Fair*] for every ϵ, $0 < \epsilon \leq 8.07$. The x%-majority decision—in general, a pair of grades—may be found for any $x\% > 50\%$.

The MJ-ranking with many voters is determined in exactly the same manner as when there are few: the most accurate majority where two candidates differ decides. Compare, for example, Hollande (H) and Bayrou (B):

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	↕	*Good*	*Fair*	*Poor*	*To Reject*
H	12.48%	16.15%	16.42%	**4.95%**	↕	**6.72%**	14.79%	14.25%	14.24%
B	2.58%	9.77%	21.71%	**15.94%**	↕	**9.30%**	20.08%	11.94%	8.69%

Both have $(50+\epsilon)$%-majorities for [*Good, Good*] for every ϵ, $0 < \epsilon \leq 4.95$, so both have the majority-grade *Good*. But for $0 < \epsilon \leq 1.77$ Hollande has a $(54.95+\epsilon)$%-majority for [*Very Good, Good*] whereas Bayrou has a $(54.95+\epsilon)$%-majority for [*Good, Good*]. Since Hollande's middlemost block dominates Bayrou's, MJ ranks Hollande above Bayrou. This happens because $4.95 < \min\{6.72, 15.94, 9.30\}$. Had the smallest of these four numbers 4.95 been Hollande's but to the right of the middle, his $(54.95+\epsilon)$%-majority would have been for [*Good, Fair*] whereas Bayrou's $(54.95+\epsilon)$%-majority would have remained [*Good, Good*], putting Bayrou ahead of Hollande. Finding the smallest of these four numbers is the same as finding the highest percentage of each candidate's grades strictly above and strictly below their majority-grades: if that highest is the percentage above the majority-grade it puts its candidate ahead, if that highest is the percentage below it puts its candidate behind. The rule has another natural interpretation: of the four sets of voters who disagree with the majority-grades the largest tips the scales.

The general rule for ranking two candidates with many voters makes the generic assumption that there is a $(50+\epsilon)$%-majority, $\epsilon > 0$, for each candidate's majority-grade. Let p_A be the percentage of A's grades strictly above her majority-grade α_A and q_A the percentage of A's grades strictly below α_A.

[7] ϵ may be thought of as one grade.

A's *majority-gauge* (MG) is (p_A, α_A, q_A). The *majority-gauge rule* $\succ_{MG}$ ranks A above B when

$$A \succ_{MG} B \text{ when } \begin{cases} \alpha_A \succ \alpha_B \text{ or,} \\ \alpha_A = \alpha_B \text{ and } p_A > \max\{q_A, p_B, q_B\} \text{ or,} \\ \alpha_A = \alpha_B \text{ and } q_B > \max\{p_A, q_A, p_B\}. \end{cases} \tag{2}$$

The MG-rule is generically decisive, i.e., there are (unique) majority-grades α_A and α_B, and a unique maximum among the four numbers p_A, q_A, p_B and q_B. This is as certain as for the majority rule to be decisive. The majority-grade may be endowed with a sign, α_A+ when $p_A > q_A$, otherwise α_A-. When the MG-rule is decisive its ranking is identical to that of the MJ-rule (specified in (1)) by construction.

The full merit profile of the French 2012 presidential poll is given in Table 3a. The MJ(=MG)-ranking is given in Table 3b together with the first-past-the-post (FPP) ranking to show the marked differences between them.

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	*Fair*	*Poor*	*To Reject*
F. Hollande	12.48%	16.15%	16.42%	11.67%	14.79%	14.25%	14.24%
F. Bayrou	2.58%	9.77%	21.71%	25.24%	20.08%	11.94%	08.69%
N. Sarkozy	9.63%	12.35%	16.28%	10.99%	11.13%	7.87%	31.75%
J.-L. Mélenchon	5.43%	9.50%	12.89%	14.65%	17.10%	15.06%	25.37%
N. Dupont-Aignan	0.54%	2.58%	5.97%	11.26%	20.22%	25.51%	33.92%
E. Joly	0.81%	2.99%	6.51%	11.80%	14.65%	24.69%	38.53%
P. Poutou	0.14%	1.36%	4.48%	7.73%	12.48%	28.09%	45.73%
M. Le Pen	5.97%	7.33%	9.50%	9.36%	13.98%	6.24%	47.63%
N. Arthaud	0.00%	1.36%	3.80%	6.51%	13.16%	25.24%	49.93%
J. Cheminade	0.41%	0.81%	2.44%	5.83%	11.67%	26.87%	51.97%

Table 3a. Merit profile,[8] 2012 French presidential poll (737 ballots) [8].

	Above Maj-grade p	Maj-grade $\alpha\pm$	Below Maj-grade q	FPP
F. Hollande	**45.05%**	*Good*+	43.28%	28.7%
F. Bayrou	34.06%	*Good*−	**40.71%**	9.1%
N. Sarkozy	**49.25%**	*Fair*+	39.62%	27.3%
J.-L. Mélenchon	**42.47%**	*Fair*+	40.43%	11.0%
N. Dupont-Aignan	**40.57%**	*Poor*+	33.92%	1.5%
E. Joly	36.77%	*Poor*−	**38.53%**	2.3%
P. Poutou	26.19%	*Poor*−	**45.73%**	1.2%
M. Le Pen	46.13%	*Poor*−	**47.63%**	17.9%
N. Arthaud	24.83%	*Poor*−	**49.93%**	0.7%
J. Cheminade	48.03%	*To Reject*	–	0.4%

Table 3b. MJ and first-past-the-post rankings, 2012 French presidential poll (737 ballots) [8].

[8]The row sums may differ from 100% due to round-off errors.

2.3 Point-summing methods

A *point-summing method*[9] chooses (ideally) an ordinal scale—words or descriptive phrases (but often undefined numbers)—and assigns to each a numerical grade, the better the evaluation the higher the number. There are, of course, infinitely many ways to assign such numbers to ordinal grades. Every voter evaluates each candidate in that scale and the candidates are ranked according to the sums or averages of their grades. A point-summing method clearly respects domination, but it harbors two major drawbacks.

The Danish educational system uses six grades with numbers attached to each: *Outstanding* 12, *Excellent* 10, *Good*, 7, *Fair* 4, *Adequate* 2 and *Inadequate* 0 [75]. Its numbers address a key issue of measurement theory [45]: the scale of grades must constitute an *interval scale* for sums or averages to be *meaningful*. An interval scale is one in which equal intervals have the same meaning; equivalently, for which an additional point anywhere in the scale—going from 3 to 4 or from 10 to 11—has the same significance. In practice—in grading divers, students, figure skaters, wines or pianists—when (say) the scale is multiples of $\frac{1}{2}$ from a high of 10 to a low of 0, it is much more difficult and much rarer to go from 9 to $9\frac{1}{2}$ than from $4\frac{1}{2}$ to 5, so adding or averaging such scores is—in the language of measurement theory—meaningless. The Danes specified a scale that they believed constitutes an interval scale (for an extended discussion of these points see [5], pp. 171-174, or [9]). "Range voting" is a point-summing method advocated on the web where voters assign a number between 0 and a 100 to each candidate, but the numbers are given no meaning other than that they contribute to a candidate's total number of points and they do not constitute an interval scale.

A second major drawback of point-summing methods is their manipulability. Any voter who has not given the highest (respectively, the lowest) grade to a candidate can increase (can decrease) the candidate's average grade, so it pays voters or judges to exaggerate up and down. A detailed analysis of an actual figure skating competition [9] shows that with point-summing every one of the nine judges could alone manipulate to achieve precisely the order-of-finish he prefers by changing his scores. A companion analysis of the same competition shows that with majority judgment the possibilities for manipulation are drastically curtailed.

Long time users of point-summing methods such as the Fédération International de Natation (FINA) have recently tried to combat strategic manipulation. Divers must specify the dives they will perform, each of which has a known degree of difficulty expressed as a number. Judges assign a number grade to each dive from 10 to 0 in multiples of $\frac{1}{2}$: *Excellent* 10, *Very Good* 8.5-9.5, *Good* 7.0-8.0, *Satisfactory* 5.0-6.5, *Deficient* 2.5-4.5, *Unsatisfactory* 0.5-2.0, *Completely failed* 0 (the meanings of each are further elaborated). There are five or seven judges. The highest and lowest scores are eliminated when there are five judges and the two highest and two lowest are eliminated when there are seven judges. The sum of the remaining three scores is multiplied by the degree of difficulty

[9] Point-summing methods are characterized in [5], chapter 17.

to obtain the score of the dive. Competitions in skating and gymnastics have chosen similar methods. Had any of them gone a little further—eliminating only what must be to distinguish a difference—they would have used MJ: increasingly practical people choose methods approaching MJ.

2.4 Approval voting

Analyses, experiments, and uses of approval voting have deliberately eschewed ascribing any meaning to *Approve* and *Disapprove*—except that *Approve* means giving one vote to a candidate and *Disapprove* means giving none—leaving it entirely to voters to decide how to try to express their opinions [74, 20]. Thus, for example, the Social Choice and Welfare Society's ballot for electing its president had small boxes next to candidates' names with the instructions: "You can vote for any number of candidates by ticking the appropriate boxes," the number of ticks determining the candidates' order of finish. In this description AV may be seen as a point-summing method where voters assign a 0 or 1 to each candidate and the electorate's rank-order is determined by the candidates' total sums of points.

Recently, however, that view has changed: "the idea of judging each and every candidate as acceptable or not is fundamentally different" from either voting for one candidate or ranking them ([18], pp.vii-viii). This implies a belief that voters are able to judge candidates in an ordinal scale of merit with two grades. With this paradigm approval voting becomes MJ with a language of two grades: "approval judgment." For example, if *Approve* meant *Good* or better the AV results of the LAMSADE Jury would be those given in Table 2f.

Student:	A	B	C	D	E	F
AV-score:	6	6	5	5	5	2
AV-ranking:	1^{st}	1^{st}	2^{nd}	2^{nd}	2^{nd}	3^{rd}

Table 2f. AV-scores and -ranking, *Approve* means *Good* or better, LAMSADE Jury.

When there are few voters AV's two grades are not sufficient to distinguish the competitors. Further evidence[10] shows two grades are too few even when there are many voters [9, 10].

3 Majority rule characterized for two candidates

Majority rule (MR) in a field of two elects that candidate preferred to the other by a majority of the electorate. May proved [51] that the majority rule is the one rule that satisfies the following six simple properties in an election with two candidates. This theorem is considered to be a major argument in its favor [33].

[10] AV in the 2012 French presidential election is discussed in section 9.4.

Axiom 1 (Based on comparing) *A voter's opinion is a preference for one candidate or indifference between them.*[11]

Thus the input is a *preference profile* that specifies the preference or indifference of each voter.

Axiom 2 (Unrestricted domain) *Voters' opinions are unrestricted.*

Axiom 3 (Anonymity) *Interchanging the names of voters does not change the outcome.*

Axiom 4 (Neutrality) *Interchanging the names of candidates does not change the outcome.*

Anonymity stipulates equity among voters, neutrality the equitable treatment of candidates.

Axiom 5 (Monotonicity) *If candidate A wins or is tied with his opponent and one or more voters change their preferences in favor of A then A wins.*

A voter's change in favor of A means changing from a preference for B to either indifference or a preference for A, or from indifference to a preference for A.

Axiom 6 (Completeness) *The rule guarantees an outcome: one of the two candidates wins or they are tied.*

Theorem 1 (May [51]) *For $n = 2$ candidates majority rule $\succeq_{MR}$ is the unique method that satisfies Axioms 1 through 6.*

Proof. The argument is simple. That MR satisfies the axioms is obvious.

So suppose the method $\succeq_M$ satisfies the axioms. Anonymity implies that only the numbers count: the number of voters n_A who prefer A to B, the number n_B that prefer B to A, and the number n_{AB} that are indifferent between A and B. Completeness guarantees there must be an outcome. (1) Suppose $n_A = n_B$ and $A \succ_M B$. By neutrality switching the names results in $B \succ_M A$: but the new profile is identical to the original, a contradiction that shows $A \approx_M B$ when $n_A = n_B$. (2) If $n_A > n_B$ change the preferences of $n_A - n_B$ voters who prefer A to B to indifferences to obtain a valid profile (by Axiom 2). With this profile $A \approx_M B$. Changing them back one at a time to the original profile proves $A \succ_M B$ by monotonicity. ■

There are three further arguments in favor of MR for *two* candidates. First, its simplicity and familiarity. Second, its incentive compatibility: the optimal strategy of a voter who prefers one of the two candidates is to vote for that candidate [12]. Third, the Condorcet jury theorem.

In its simplest form, the jury theorem supposes that one of the two outcomes is correct and that each voter has an independent probability $p > 50\%$ of voting

[11] In a footnote May had the wisdom to admit, "The realism of this condition may be questioned."

for it, concludes that the greater the number of voters the more likely the majority rule makes the correct choice, and that furthermore, in the limit, it is certain to do so. In most votes between two alternatives, however, there is no "correct" choice or "correct" candidate: all opinions are valid judgments, disagreement is inherent to any democracy, and must be accepted. A mechanism that choses the consensus is what is needed.

4 The domination paradox

The dangers of a "tyranny of the majority" have been discussed for centuries: the phrase was used by John Adams in 1788, Alexis de Tocqueville in 1835 and John Stuart Mill in 1859. Robert A. Dahl's now classic treatise [26] considers it at length in the context of what he calls "Madisonian Democracy" and poses the problem of intensity: "[J]ust as Madison believed that government should be constructed so as to prevent majorities from invading the natural rights of minorities, so a modern Madison might argue that government should be designed to inhibit a relatively apathetic majority from cramming its policy down the throats of a relatively intense minority" ([26], p. 90).

Given voters' evaluations of candidates, MR for two candidates harbors a major drawback heretofore unrecognized. It admits the *domination paradox*: a candidate A's evaluations dominate B's evaluations but MR elects B not A.

Contrast the merit profiles of Hollande and Sarkozy in the national poll of the 2012 French presidential election (Table 4a). Hollande's grades very generously dominate Sarkozy's. But this merit profile could come from the opinion profile of Table 4b where Sarkozy is the MR-winner with 59.57% of the votes to Hollande's 26.19%, 14.24% rejecting both.

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	*Fair*	*Poor*	*To Reject*
Hollande	12.48%	16.15%	16.42%	11.67%	14.79%	14.25%	14.24%
Sarkozy	9.63%	12.35%	16.28%	10.99%	11.13%	7.87%	31.75%

Table 4a. Merit profile, Hollande-Sarkozy, 2012 French presidential election poll.

	9.63%	12.35%	11.67%	4.61%	10.18%	11.13%	14.24%
Hollande:	*Exc.*	*V.Good*	*Good*	*Fair*	*Fair*	*Poor*	*Rej.*
Sarkozy:	*Outs.*	*Exc.*	*V.Good*	*V.Good*	*Good*	*Fair*	*Rej.*

	0.81%	7.87%	3.80%	6.52%	4.07%	3.12%
Hollande:	*Outs.*	*Outs.*	*Outs.*	*Exc.*	*V.Good*	*Poor*
Sarkozy:	*Good*	*Poor*	*Rej.*	*Rej.*	*Rej.*	*Rej.*

Table 4b. Possible opinion profile, Hollande-Sarkozy (giving the merit profile of table 4a), national poll, 2012 French presidential election.

Those voters who rate Sarkozy above Hollande do so mildly (with small differences in grades, top of profile), whereas Holland is rated above Sarkozy

intensely (with large differences in grades, bottom of the profile): this is a situation Dahl elicited in questioning the validity of MR. In the actual national vote—and in the poll—Hollande won by a bare majority of 51.6% to 48.4%, suggesting that the possibility for MR to err on two candidates is important and real. To elect a candidate whose evaluations are dominated by another is clearly unacceptable.

5 May's axioms for more than two candidates

Given a fixed scale of linearly ordered grades Λ, a *ranking problem* is defined by an electorate's opinion profile Φ, an m by n matrix of grades when there are m candidates and n voters. A *method of ranking* $\succeq_M$ is an asymmetric binary relation between all pairs of candidates.

With the grading model voters' inputs are grades. With the traditional model voters' inputs are comparisons. Individual rationality implies that a voter's preference is a rank order over all the candidates. It may be deduced from a voter's grades when the scale is *sufficient* to distinguish between any two candidates whenever the voter believes their merit to be different.

The following are May's axioms extended to any number of candidates.

Axiom 1 (Based on comparing) *A voter's input is a rank-order of the candidates.*

Axiom 2 (Unrestricted domain) *Voters' opinions are unrestricted.*

Axiom 3 (Anonymity) *Interchanging the names of voters does not change the outcome.*

Axiom 4 (Neutrality) *Interchanging the names of candidates does not change the outcome.*

In the traditional model a voter's input *becomes better* for a candidate A if A rises in his rank-order. In the grading model a voter's input *becomes better* for A if A is given a higher grade.

Axiom 5 (Monotonicity) *If $A \succeq_M B$ and one or more voters' inputs become better for A then $A \succ_M B$.*

Axiom 6 (Completeness) *For any two candidates either $A \succeq_M B$ or $A \preceq_M B$ (or both, implying $A \approx_M B$).*

With more than two candidates the Condorcet and Arrow paradoxes must be excluded.

Axiom 7 (Transitivity) *If $A \succeq_M B$ and $B \succeq_M C$ then $A \succeq_M C$.*

Axiom 8 (Independence of irrelevant alternatives (IIA)) *If $A \succeq_M B$ then whatever other candidates are either dropped or adjoined $A \succeq_M B$.*

This is the Nash-Chernoff formulation of IIA defined for a variable number of candidates [59, 24], not Arrow's definition for a fixed number of candidates, and it implies Arrow's. It is this conception that is often violated in practice (e.g., elections, figure skating, wines).

Theorem 2 (Arrow's impossibility [1]) *For $n \geq 3$ candidates there is no method of ranking $\succeq_M$ that satisfies Axioms 1 through 8.*

This is a much watered-down form of Arrow's theorem, based on more axioms—though all necessary in a democracy—and is very easily proven. It is stated to contrast the two models: comparing candidates versus evaluating them.

Proof. Take any two candidates A and B. By IIA it suffices to deal with them alone to decide who leads. Axioms 1 through 6 imply that the method $\succeq_M$ must be MR. Since the domain is unrestricted Condorcet's paradox now shows MR violates transitivity. So there can be no method satisfying all the axioms. ■

Now replace Axiom 1 above by:

Axiom 1* (Based on measuring) *A voter's input is the grades given the candidates.*

Theorem 3 *For $n \geq 1$ candidates there are an infinite number of methods of ranking $\succeq_M$ that satisfy Axioms 1* and 2 through 8. Their rankings depend only on candidates' merit profiles and they respect domination.*

Proof. Majority judgment and any point-summing method clearly satisfy all axioms, so there are as many methods as one wishes that satisfy the axioms.

All methods satisfying the axioms depend only on their merit profiles. To see this compare two competitors. By Axiom 8 (IIA) it suffices to compare them alone. If two candidates A and B have the same set of grades, so that B's list of n grades is a permutation σ of A's list, it is shown that they must be tied. Consider, first, the opinion profile ϕ^1 of A and a candidate A' who may be added to the set of candidates by IIA,

$$\begin{array}{rccccc} \phi^1: & & v_1 & \cdots & v_{\sigma 1} & \cdots & v_n \\ \hline & A: & \alpha_1 & \cdots & \alpha_{\sigma 1} & \cdots & \alpha_n \\ & A': & \alpha_{\sigma 1} & \cdots & \alpha_1 & \cdots & \alpha_n \\ \hline \end{array}$$

where A''s list is the same as A's except that the grades given by voters v_1 and $v_{\sigma 1}$ have been interchanged. ϕ^1 is possible since the domain is unrestricted. Suppose $A \succeq_M A'$. Interchanging the votes of the voters v_1 and $v_{\sigma 1}$ yields the profile ϕ^2

$$\begin{array}{rccccc} \phi^2: & & v_{\sigma 1} & \cdots & v_1 & \cdots & v_n \\ \hline & A: & \alpha_{\sigma 1} & \cdots & \alpha_1 & \cdots & \alpha_n \\ & A': & \alpha_1 & \cdots & \alpha_{\sigma 1} & \cdots & \alpha_n \\ \hline \end{array}$$

Nothing has changed by Axiom 3 (anonymity), so the first row of ϕ^2 ranks at least as high as the second. But by Axiom 4 (neutrality) $A' \succeq_M A$, implying $A \approx_M A'$. Thus $A \approx_M A'$ where A''s first grade agrees with B's first grade.

Compare, now, A' with another added candidate A'', with profile ϕ^3

$$\phi^3: \quad \begin{array}{c|cccccc} & v_{\sigma 1} & v_2 & \cdots & v_{\sigma 2} & \cdots & v_n \\ \hline A': & \alpha_{\sigma 1} & \alpha_2 & \cdots & \alpha_{\sigma 2} & \cdots & \alpha_n \\ A'': & \alpha_{\sigma 1} & \alpha_{\sigma 2} & \cdots & \alpha_2 & \cdots & \alpha_n \end{array}$$

where A'''s list is the same as A''s except that the grades of voters v_2 and v_{σ_2} have been interchanged. Suppose $A' \succeq_M A''$. Interchanging the votes of the voters v_2 and $v_{\sigma 2}$ yields the profile ϕ^4

$$\phi^4: \quad \begin{array}{c|cccccc} & v_{\sigma 1} & v_{\sigma 2} & \cdots & v_2 & \cdots & v_n \\ \hline A': & \alpha_{\sigma 1} & \alpha_{\sigma 2} & \cdots & \alpha_2 & \cdots & \alpha_n \\ A'': & \alpha_{\sigma 1} & \alpha_2 & \cdots & \alpha_{\sigma 2} & \cdots & \alpha_n \end{array}$$

so as before conclude that $A' \approx_M A''$. Axiom 7 (transitivity) now implies $A \approx_M A''$ where A'''s first two grades agree with B's first two grades.

Repeating this reasoning shows $A \approx B$, so which voter gave which grade has no significance. Therefore a candidate's distribution of grades—his merit profile—is what determines his place in the ranking with any method that satisfies the Axioms. It has a unique representation when the grades are listed from the highest to the lowest.

Suppose A's grades α dominates B's grades β, both given in order of decreasing grades. Domination means $\alpha_j \succeq \beta_j$ for all j, with at least one strictly above the other. If $\alpha_k \succ \beta_k$ replace β_k in β by α_k to obtain $\beta^1 \succ_M \beta$ by monotonicity (Axiom 5). Either $\beta^1 \approx_M \alpha$ proving that $\alpha \succ_M \beta$, or else $\alpha \succ_M \beta^1$. In the second case, do as before to obtain $\beta^2 \succ_M \beta^1$, and either $\beta^2 \approx_M \alpha$, or else $\alpha \succ_M \beta^2$. If $\beta^2 \approx_M \alpha$ then $\beta \prec_M \beta^1 \prec_M \beta^2 \approx_M \alpha$ and transitivity implies $\beta \prec_M \alpha$. Otherwise, repeating the same argument shows that $\alpha \succ_M \beta$.

Monotonicity implies domination is respected. ■

A reasonable method should certainly avoid the domination paradox. This theorem shows that any method based on evaluations that satisfies May's axioms and avoids the Condorcet and Arrow paradoxes does so. When majority rule fails to respect domination it differs from all such methods: why then the persistent insistence on agreeing with the majority rule?

6 Polarization

Are there reasons to choose one method among all that meet the demands of Theorem 3?

All ranking methods that satisfy IIA (Axiom 8) are determined by how they rank pairs of candidates. So what makes sense when two candidates are to be ranked? In particular, are there circumstances when majority rule for two candidates *is* acceptable?

One instance leaps to mind: jury decisions. The goal is to arrive at the truth, the correct decision, either the defendant is guilty or is not guilty. A juror may be more or less confident in his judgment: the higher his belief that one decision is correct the lower his belief that the opposite decision is correct. In this context Condorcet's jury theorem strongly supports MR. But this context is very different from that of most elections between two candidates where gradations of opinion are inherent and an excellent opinion of one does not necessarily imply a low opinion of the other.

"Political polarization" is a phenomenon that is more and more observed and discussed in the United States and other countries (see e.g., [11, 22]). It refers to a partisan cleavage in political attitudes in support of ideological extremes—pro-abortion vs. anti-abortion, pro-evolution vs. anti-evolution, left vs. right—and applies to voters, elites, candidates, and parties. The concept necessarily concerns an opposition between two. The word is used when (say) large majorities of Democratic and Republican voters are vehemently on opposite sides in their evaluations of issues or candidates. The notion evokes the idea that most voters are at once intensely for one side and intensely against the other, so the situation approaches that of a jury decision where there is no question of (in Dahl's words) pitting a passionate minority against an apathetic majority.

Consider the two major opponents of the 2012 French presidential election poll (Table 3a), Hollande (moderate left) and Sarkozy (traditional right). Table 5a gives the electorate's opinion profile concerning them (where, e.g., 1.63% in the first column give Sarkozy *Fair* and Hollande *Outstanding*). Tables 5b and 5c contain respectively the distributions and cumulative distributions of the grades given Hollande for each of the grades given Sarkozy (Table 5b is obtained from Table 5a by normalizing each line to sum to 100%).

		Hollande							
		Outs.	*Exc.*	*V.G.*	*Good*	*Fair*	*Poor*	*Rej.*	Total
S	*Outs.*	0.14%	0.00%	0.41%	1.09%	2.04%	2.99%	2.99%	09.63%
a	*Exc.*	0.27%	1.09%	0.95%	2.17%	2.71%	2.71%	2.44%	12.35%
r	*V.G.*	0.27%	1.22%	2.04%	3.12%	2.99%	3.93%	2.71%	16.28%
k	*Good*	1.22%	1.09%	1.76%	1.76%	2.85%	1.63%	0.68%	10.99%
o	*Fair*	1.63%	2.44%	2.58%	1.09%	2.31%	0.68%	0.41%	11.13%
z	*Poor*	1.75%	2.58%	1.09%	0.27%	0.54%	0.81%	0.81%	07.87%
y	*Rej.*	7.19%	7.73%	7.60%	2.17%	1.36%	1.49%	4.21%	31.75%
	Total	12.48%	16.15%	16.42%	11.67%	14.79%	14.25%	14.25%	

Table 5a. Opinion profile, Hollande-Sarkozy, 2012 French presidential poll.

		Hollande						
		Outs.	*Exc.*	*V.Good*	*Good*	*Fair*	*Poor*	*Rej.*
S	*Outs.*	01.41%	00.00%	04.23%	11.27%	21.13%	30.99%	30.99%
a	*Exc.*	02.20%	08.79%	07.69%	17.58%	21.98%	21.98%	19.78%
r	*V.Good*	01.67%	07.50%	12.50%	19.17%	18.33%	24.17%	16.67%
k	*Good*	11.11%	09.88%	16.05%	16.05%	25.93%	14.81%	06.17%
o	*Fair*	14.63%	21.95%	23.17%	09.76%	20.73%	06.10%	03.66%
z	*Poor*	22.41%	32.76%	13.79%	03.45%	06.90%	10.34%	10.34%
y	*Rej.*	22.65%	24.36%	23.93%	06.84%	04.27%	04.70%	13.25%

Table 5b. Distibutions of Hollande's grades for each of Sarkozy's grades, 2012 French presidential poll.

		Hollande						
		Outs.	$\succeq$ *Exc.*	$\succeq$ *V.Good*	$\succeq$ *Good*	$\succeq$ *Fair*	$\succeq$ *Poor*	$\succeq$ *Rej.*
S	*Outs.*	01.41%	01.41%	05.64%	16.91%	38.04%	69.03%	100%
a	*Exc.*	02.20%	10.99%	18.68%	36.26%	58.24%	80.23%	100%
r	*V.Good*	01.67%	09.17%	21.67%	40.84%	59.17%	83.34%	100%
k	*Good*	11.11%	20.99%	37.04%	53.09%	79.02%	93.83%	100%
o	*Fair*	14.63%	36.58%	59.75%	69.51%	90.24%	96.34%	100%
z	*Poor*	22.41%	55.17%	68.96%	72.41%	79.31%	89.65%	100%
y	*Rej.*	22.65%	47.01%	70.94%	77.78%	82.05%	86.75%	100%

Table 5c. Cumulative distibutions of Hollande's grades for each of Sarkozy's grades, 2012 French presidential poll.

Comparing any two lines of Table 5c, the lower (almost) always dominates the upper. Thus the lower the grade given Sarkozy the higher the distribution of grades given Hollande: statistically there is diametrical opposition in their grades. The same occurs with the distributions of Sarkozy's grades for each of Hollande's grades. Moreover, as may be seen in Table 5d, 47.36% of the voters are polarized on the two candidates.

	7.19%	7.73%	7.60%	2.17%	1.36%	0.54%	2.31%
Hollande:	*Outs.*	*Exc.*	*V.Good*	*Good*	*Fair*	*Fair*	*Fair*
Sarkozy:	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Poor*	*Fair*

	2.85%	2.99%	3.93%	2.71%	2.44%	2.99%
Hollande:	*Fair*	*Fair*	*Poor*	*Rej.*	*Rej.*	*Rej.*
Sarkozy:	*Good*	*V.Good*	*V.Good*	*V.Good*	*Exc.*	*Outs.*

Table 5d. Polarized part (47.36%) of true opinion profile, Hollande-Sarkozy, 2012 French presidential election poll (for merit profile of Table 4a).

On the other hand, analyzing in the same manner the grades given to two extreme left candidates, Mélenchon and Poutou, the higher the grade given one the higher the distribution of grades given the other: statistically there is a concordance of grades. All of this confirms common sense.

Given an electorate's merit profile take any two candidates. Different opinion profiles on the two candidates have this same merit profile. An opinion profile on a pair of candidates will be said to be "polarized" when the higher a voter evaluates one candidate the lower the voter evaluates the other. Specifically:

> Two candidates A and B are *polarized* if, for any two voters, v_i evaluates A higher (respectively, lower) than v_j then v_i evaluates B no higher (respectively, no lower) than v_j.

This definition includes and extends the usual (loosely defined) notion of a polarized electorate and formulates a "pure" form of a type of distribution of grades found in practice. It brings to mind the single crossing opinion profile of the traditional theory [13, 62].

The merit profile of Hollande and Sarkozy (Table 4a or 5a) showed that Hollande's grades largely dominate Sarkozy's, and yet there are opinion profiles

(e.g., Table 4b) that make Sarkozy the overwhelming MR-winner. The polarized opinion profile that has the same merit profile is given in Table 5e.

	12.48%	16.15%	3.12%	7.87%	5.43%	5.70%	5.97%
Hollande:	*Outs.*	*Exc.*	*V.Good*	*V.Good*	*V.Good*	*Good*	*Good*
Sarkozy:	*Rej.*	*Rej.*	*Rej.*	*Poor*	*Fair*	*Fair*	*Good*

	5.02%	9.77%	6.51%	7.74%	4.61%	9.63%
Hollande:	*Fair*	*Fair*	*Poor.*	*Poor*	*Rej.*	*Rej.*
Sarkozy:	*Good*	*V.Good*	*V.Good*	*Exc.*	*Exc.*	*Outs.*

Table 5e. Polarized opinion profile, Hollande-Sarkozy, 2012 French presidential election poll (for merit profile of Table 4a).

The polarized opinion profile makes Hollande the MR-winner with a score of 50.75% to Sarkozy's 43.28%, 5.97% giving both the grade *Good*, so agrees with Hollande's domination in grades. With this profile MR makes the right decision.

It would seem that it is precisely when an electorate is polarized—or when a jury seeks the correct answer between two opposites—and there can be no consensus that the "strongly for or strongly against" characteristic of MR should render the acceptable result. The electorate is not polarized on Hollande vs. Sarkozy (Table 5a): it is only statistically polarized. Completely polarized pairs are rare; however, significant parts may be polarized (Tables 5d,h). Consider, for example, the breakdown of the grades given to Poutou (extreme left) and Le Pen (extreme right) in the 2012 French presidential poll (see Table 5f).

		Poutou							
		Outs.	*Exc.*	*V.G.*	*Good*	*Fair*	*Poor*	*Rej.*	Total
	Outs.	0.00%	0.14%	0.14%	0.00%	0.27%	0.95%	4.48%	05.97%
L	*Exc.*	0.00%	0.14%	0.41%	0.41%	1.36%	1.90%	3.12%	07.33%
e	*V.G.*	0.00%	0.00%	0.81%	0.14%	1.09%	3.39%	4.07%	09.50%
	Good	0.00%	0.00%	0.41%	1.09%	1.22%	1.36%	5.29%	09.36%
P	*Fair*	0.00%	0.14%	0.41%	0.68%	2.58%	4.75%	5.43%	13.98%
e	*Poor*	0.00%	0.27%	0.14%	0.27%	1.09%	1.90%	2.58%	06.24%
n	*Rej.*	0.14%	0.68%	2.17%	5.16%	4.88%	13.84%	20.76%	47.63%
	Total	0.14%	1.36%	4.48%	7.73%	12.48%	28.09%	45.73%	

Table 5f. Opinion profile, Le Pen-Poutou, 2012 French presidential poll.

The polarized opinion profile that agrees with the merit profile is given in Table 5g (MR places Poutou with 47.63% ahead of Le Pen with 46.14%). It does not agree with the true opinion profile.

	0.14%	1.36%	4.48%	7.73%	12.48%	21.44%	6.24%
Poutou:	*Outs.*	*Exc.*	*V.Good*	*Good*	*Fair*	*Poor*	*Poor*
Le Pen:	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Poor*

	0.41%	13.57%	9.36%	9.50%	7.33%	5.97%
Poutou:	*Poor*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*
Le Pen:	*Fair*	*Fair*	*Good*	*V.Good*	*Exc.*	*Outs.*

Table 5g. Polarized opinion profile, Le Pen-Poutou, 2012 French presidential election poll (for merit profile of Table 4a).

	0.14%	0.68%	2.17%	5.16%	4.88%	13.84%	20.76%
Poutou:	*Outs.*	*Exc.*	*V.Good*	*Good*	*Fair*	*Poor*	*Rej.*
Le Pen:	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*

	2.58%	5.43%	5.29%	4.07%	3.12%	4.48%
Poutou:	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*	*Rej.*
Le Pen:	*Poor*	*Fair*	*Good*	*V.Good*	*Exc.*	*Outs.*

Table 5h. Polarized part (72.60%) of true opinion profile, Le Pen-Poutou, 2012 French presidential election poll (for merit profile of Table 4a).

However, a large part of the true opinion profile—almost three-quarters—is polarized on Le Pen and Poutou as may be seen in Table 5h, and every opinion profile will have parts that are polarized.

Polarization, as defined here, is a significant phenomenon in practice. It is also of significance in theory. It is proven in the next section—and was observed in the example of Table 5e—that when two candidates are polarized and one's grades dominate the other's MR (when decisive) makes the correct decision: it *is* acceptable.

7 A new chacterization of majority judgment for large electorates

A method of ranking $\succeq_M$ is *consistent with the majority rule on polarized pairs of candidates* if both give the identical ranking between every pair of polarized candidates whenever both are decisive (meaning no ties). Generically, ties have a zero probability of occurring when the number of voters becomes large.

MR on two candidates has a particularly important property: it resists manipulation or is incentive compatible. Any method that agrees with it inherits the property. MR may err, but it does not when a pair of candidates is polarized, and that is precisely the situation when voters are most tempted to manipulate. Thus for a method to agree with MR on polarized pairs is a very attractive property.

Not every method satisfying Axioms 1* and 2 – 8 is consistent with MR on polarized electorates. Point-summing methods, for example, are not. Take the polarized opinion profile of Poutou and Le Pen (Table 5g). MR makes Poutou the winner with 47.63% of the votes to Le Pen's 46.14%, 6.24% evaluating both to be *Poor*. But if *Outstanding* is worth 6 points, *Excellent* 5, ..., down to *To Reject* 0, Le Pen's 172.75 points easily defeats Poutou's 101.80. Furthermore, it will be proven that consistency with MR on polarized pairs is necessary to combat strategic manipulation (Theorem 6).

Theorem 4 *A method of ranking $\succeq_M$ that satisfies Axioms 1* and 2-8 and is consistent with the majority rule on polarized pairs of candidates must coincide with the majority-gauge rule $\succ_{MG}$ when the scale of grades Λ is sufficient.*

The assumption of a sufficient scale of grades is necessary only to compare the rankings of a method based on evaluations with the rankings based on preferences; otherwise, several preference profiles may be associated with the same opinion profile (depending on how two candidates with the same grade are compared: indifferent or one of the two preferred).

Proof. Given an opinion profile with any number of candidates, IIA (Axiom 8) implies that the order between any two must be determined between the two alone. By Theorem 3, the method $\succeq_M$ depends only on the merit profile—the distributions of the candidates' grade—not on which voters gave which grades. Thus, the order between them determined by the method is the same as that determined by the polarized opinion profile.

	$x_1\%$		$x_j\%$			$x_{k-1}\%$		$x_k\%$		$x_{k+1}\%$				$x_s\%$
A	$\lambda_A^1 \succeq$	$\cdots$	λ_A^j	$\cdots$	$\succeq$	λ_A^{k-1}	$\succeq$	λ_A^k	$\succeq$	λ_A^{k+1}	$\succeq$	$\cdots$	$\succeq$	λ_A^s
	$\succ$	$\cdots$	$\succ$	$\cdots$		$\succ$		$=$		$\prec$		$\cdots$		$\prec$
B	$\lambda_B^1 \preceq$	$\cdots$	λ_B^j	$\cdots$	$\preceq$	λ_B^{k-1}	$\preceq$	λ_B^k	$\preceq$	λ_B^{k+1}	$\preceq$	$\cdots$	$\preceq$	λ_B^s

Table 6. Polarized opinion profile.

So consider the polarized opinion profile of two candidates, A's grades going from highest on the left to lowest on the right and B's from lowest to highest, as displayed in Table 6. A's grades are non-increasing and B's non-decreasing and for any j either $\lambda_A^j \succ \lambda_A^{j+1}$ or $\lambda_B^j \prec \lambda_B^{j+1}$, so the corresponding grades can be equal at most once (as indicated in the middle line of the profile).[12]

Suppose A is the MR-winner. Then $x_A = \sum_1^{k-1} x_i > \sum_{k+1}^{s} x_i = x_B$. By assumption $x_A = 50\%$ is excluded (since $\succ_{MG}$ must be decisive).

If $x_A > 50\%$ then A's majority-grade is at least λ_A^{k-1} and B's at most λ_B^{k-1}, so A is ahead of B by the MG rule.

Otherwise $x_B < x_A < 50\%$ implying $x_A + x_k > 50\%$ and $x_B + x_k > 50\%$, so the candidates' majority-grades are the same, $\lambda_A^k = \lambda_B^k$. Notice that $p_A \leq x_A$ and $q_B \leq x_A$ but one of the two must be an equality; similarly $q_A \leq x_B$ and $p_B \leq x_B$ but one of the two must be an equality as well. Thus $x_A = \max\{p_A, q_B\}$ and $x_B = \max\{p_B, q_A\} < x_A$. If $p_A = x_A$ then p_A is the largest of the p's and q's, and MG puts A above B; if $q_B = x_A$ then q_B is the largest of the p's and q's, and MG puts B below A, as was to be shown.

Now assume the MG rule places A above B, $(p_A, \lambda_A, q_A) \succ_{MG} (p_B, \lambda_B, q_B)$. Either A's majority-grade is above B's, $\lambda_A \succ \lambda_B$; or they are equal and either $p_A > \max\{q_A, p_B, q_B\}$ or $q_B > \max\{p_A, q_A, p_B\}$.

In the first case, at least $(50 + \epsilon)\%$ (for some $\epsilon > 0$) of the voters gave the grade λ_A or better to A and λ_B or worse to B. They constitute a majority of at least $(50 + \epsilon)\%$ that makes A the MR-winner.

In the second case, they are equal and are the kth column of the polarized opinion profile: $\lambda_A = \lambda_B = \lambda_A^k = \lambda_B^k$. As before, $x_A = \sum_1^{k-1} x_i = \max\{p_A, q_B\}$ and $x_B = \sum_{k+1}^{s} x_i = \max\{p_B, q_A\}$. Thus, $p_A > \max\{q_A, p_B, q_B\}$

[12]This brings to mind the single crossing property mentioned earlier.

implies $x_A = p_A > \max\{p_B, q_A\} = x_B$ and $q_B > \max\{p_A, q_A, p_B\}$ implies $x_A = q_B > \max\{p_B, q_A\} = x_B$, so in both instances A is the MR-winner. ■

Corollary *Majority rule on a pair of polarized candidates elects (when decisive) a candidate whose grades dominate the other's.*

Proof. The majority-gauge rule elects the candidate whose grades dominate the other's. But as just shown the majority rule coincides with it when the candidates are polarized, proving the corollary. ■

With majority judgment any number of voters may rank any number of candidates and two candidates are tied only when their sets of grades are identical. However, whenever the majority-gauge is decisive it ranks the candidates exactly as does majority judgment (as seen above, also proven in [5] pp. 236-239). This will almost certainly be the case in an election with many voters, e.g., with over a hundred voters and a scale of six or seven grades (in actual experience to date the majority-gauge has sufficed to determine the rank-order with 19 voters and fewer [9]). Thus for all intents and purposes the majority judgment ranking *is* the majority-gauge ranking when there are many voters.

If the scale of grades is not sufficient—i.e., a voter is forced to give a same grade to two candidates whereas she has a preference between them—the theorem fails since it is impossible to deduce the ranking from the grades (a same grade for two candidates could bear three meanings, preference for one of the two or indifference). For example, with two grades MJ becomes approval voting, a voter may *Approve* both candidates or *Disapprove* both without actually being indifferent between them, so that agreement with majority rule calculated on the basis of preferences is impossible even when the electorate is polarized. If *Approve* means *Good* or better in the polarized profile of Poutou versus Le Pen (Table 5f), approval voting elects Le Pen with 32.16% not Poutou with 13.71%, reversing the outcome with the full complement of grades.

To guarantee agreement with the majority rule on polarized pairs the scale of grades must faithfully represent the possible diversity of opinion. The optimal number of grades to use depends on the application and is discussed in section 10.2.

8 Dahl's intensity problem

Dahl directly challenged the validity of MR on two candidates, suggesting that when a minority ardently supports one candidate whereas the majority only mildly prefers the other (such as may be seen in Table 4b), the minority candidate should win: "If there is any case that might be considered the modern analogue to Madison's implicit concept of tyranny, I suppose it is this one" ([26], p. 99).

To attack this question Dahl proposed using an ordinal "intensity scale" obtained "simply by reference to some observable response, such as a statement of one's feelings ..." ([26], p. 101), and argued that it is meaningful to do so: "I think that the core of meaning is to be found in the assumption that the

uniformities we observe in human beings must carry over, in part, to the unobservables like feeling and sensation" ([26], p. 100). This is precisely the role of Axiom 1* and relates to a key problem raised by measurement theorists, the *faithful representation problem*: when measuring some attribute of a class of objects or events how to associate a scale "in such a way that the properties of the attribute are faithfully represented ..."([45], p.1). Practice—in figure skating, wine tasting, diving, gymnastics, assessing pain, etc.—has spontaneously and naturally resolved it. However, some social choice theorists continue to express the opinion that a scale of intensities or grades is inappropriate in elections.

Why should intensities be valid—indeed, be necessary—and practical in judging competitions but not in elections? The validity of using intensities as inputs in voting as versus using rankings as inputs is not a matter of opinion or of mathematics: it is a practical, experimental issue. Repeated experiments and real uses of MJ ([5] pp. 9-16 and chapter 15; [42]; [9] pp. 496-497 and 504-509; and others referred to in this article) show that nuances in evaluations are as valid for candidates in elections as they are for figure skaters, divers or wines in competitions, though the criteria and scales of measures must be crafted for each individually. Participants have uniformly found it easy to assign grades and have done so quickly. Some practical-minded people are also persuaded that MJ should be used in elections: Terra Nova—"an independent progressive think tank whose goal is to produce and diffuse innovative political solutions in France and Europe"—has included majority judgment in its recommendations for reforming the presidential election system of France [71].

Dahl's challenge was: "If a collective decision is involved, one that requires voting, would it be possible to construct rules so that an apathetic majority only slightly preferring its alternative could not override a minority strongly preferring its alternative?" He went on to say he would not do so, but believed it would "answer a problem that ... has frequently disturbed democratic theorists."

However, in discussing intensities he erred in the formulation of the problem in three particulars. (1) He concentrated *only* on the choice between two alternatives A and B with intensities that are not attached to the alternatives themselves: instead the intensities express the degree to which a voter prefers A as *compared* with B (or vice versa), in keeping with the traditional paradigm. Thus it is not possible to apply or build on his ideas when there are more than two alternatives. (2) He failed say what he means by "a majority slightly prefers one alternative and a minority strongly prefers a mutually exclusive alternative " ([26] p. 102). The objective is clear—to permit exceptions to majority rule in limited circumstances—but surely the relative sizes of the majority and minority matter, and he never mentions them. (3) He did not realize that with intensities MR can go so far wrong as not to respect domination.

Given two alternatives A and B and the voters' grades or intensities attached to each, when *should* A be the winner? There is one indisputable case: when A's merit profile dominates B's. As was seen, every method based on evaluations and satisfying Axioms 2 to 8 makes A the winner in such cases. However, MR may fail to do so.

Given the opinion profiles of each candidate (e.g., Table 7a) the majority principle has two possible interpretations: the traditional meaning based on comparisons (MR, or the "vertical" view of the opinion profile) and the meaning based on each candidate's grades (the MJ-ranking, or the "horizontal" view of the opinion profile). Are there circumstances when these two points of view may be reconciled? As was proven, there is: when the candidates are polarized and no consensus is possible; in extreme cases when A's supporters give A very high grades and B very low ones and B's supporters do the opposite. This is precisely the situation where MR makes the most sense. It is not Dahl's apathetic majority opposed to an intense minority. However, sometimes polarized candidates pit apathetic majorities against intense minorities with MJ giving the result that Dahl seeks, as for example, the seven voters with the opinion profile given in Table 7a. The first four voters form the apathetic majority in favor of A; the last three form the intense minority in favor of B: MR elects A, MJ elects B.

	v_1	v_2	v_3	v_4	v_5	v_6	v_7
A	*Good*	*Very Good*	*Fair*	*Poor*	*Fair*	*Poor*	*To Reject*
B	*Fair*	*Good*	*Poor*	*To Reject*	*Excellent*	*Very Good*	*Very Good*

Table 7a. Opinion profile, A the MR-winner, B the MJ-winner.

A more extreme example is the possible opinion profile of Hollande and Sarkozy in the national poll of the 2012 French presidential election given in Table 4b: the majority of 59.67% that rates Sarkozy above Hollande does so mildly (top of table), the minority of 26.19% that rates Hollande above Sarkozy does so emphatically (bottom of table): MR places Sarkozy above Hollande, MJ places Hollande above Sarkozy. On the other hand, the opinion profile of Table 7b seems to be a typical example of what Dahl had in mind but MR and MJ concord in making B the winner with an apathetic majority: here the candidates are polarized so the two methods must agree.[13]

	v_1	v_2	v_3	v_4	v_5	v_6	v_7
A	*Excellent*	*Excellent*	*Excellent*	*Fair*	*Fair*	*Fair*	*Fair*
B	*Poor*	*Poor*	*Poor*	*Good*	*Good*	*Good*	*Good*

Table 7b. Opinion profile, B the MR- and MJ-winner.

To illustrate the importance of the relative sizes of the majorities and minorities take a scale of three evaluations (say 2, 1, and 0). Suppose the opinion profile on two candidates is:

$$\begin{array}{cccc} & \overbrace{}^{n_1\%} & \overbrace{}^{n_2\%} & \overbrace{}^{n_3\%} \\ A\text{:} & 2\cdots2 & 1\cdots1 & 0\cdots0 \\ B\text{:} & 1\cdots1 & 0\cdots0 & 2\cdots2 \end{array}$$

[13]With the point-summing method obtained by giving 6 points to *Outstanding*, 5 to *Excellent*, ..., 0 to *To Reject*, A is the winner. Some point-summing methods elect A, others elect B.

where $(n_1 + n_2)\%$ of the voters slightly prefer A to B and $n_3\%$ strongly prefer B to A. When n_3 is sufficiently large Dahl would presumably wish B to be elected.

(1) When $n_3 > \max\{n_1, n_2\}$ this is exactly what MJ does because B's merit profile dominates A's (since $n_3 > n_1$ and $n_3 + n_1 > n_2 + n_1$). (2) On the other hand when $n_3 < \min\{n_1, n_2\}$ Dahl would presumably wish A to be elected because the minority that strongly prefers B to A is too small. Again, this is exactly what MJ does because A's merit profile dominates B's (since $n_1 > n_3$ and $n_1 + n_2 > n_3 + n_1$). Note that if Dahl placed no limit on the size of the minority, B would be the winner violating domination. (3) When neither of the conditions hold there is no domination, at times A is the MJ-winner, at other times B is the MJ-winner.

Thus MJ may be viewed as a possible solution to Dahl's reformulated intensity problem in that it is based on a model that permits intensities to be expressed by voters when there are any number of candidates, respects domination, sometimes declares a candidate supported by an enthusiastic minority to be the winner in a two-person election, and agrees with MR on two candidates when they are polarized so that no consensus is possible.[14]

9 Majority judgment: good properties

9.1 Permits voters to express themselves

MJ enables voters to express their opinions much more precisely than any other input including rank-orders, as does any method based on measuring when the scale is sufficient. One telling instance is the 2007 experiment mentioned in the introduction: over a third of the voters gave their highest grade to at least two candidates. Moreover, voters assign grades to candidates quickly, so doing so must be cognitively simple in comparison with rank-ordering.[15] Practice suggests rank-ordering is difficult. In figure skating the inputs to the method of ranking were for years rank-orders but judges were asked for grades from which their rank-orders were deduced. In Australia, a voter must rank-order all candidates for her ballot to be valid. Some ten to twelve candidates run for a seat in the House of Representatives: "how to vote cards" are provided by each of the parties at polling places to show voters how to place their 1's, 2's, ..., and 10's that specify their rankings. In senatorial elections there may be as many as sixty candidates: few voters practice "below-the-line voting," meaning specifying their personal rank-order; most (some 90%) practice "above-the-line voting," meaning they simply tick one box that indicates a party's list [3, 4].

[14] Storable votes [23]—reinterpreted as a method for ranking many candidates—may also be considered an approach to the intensity problem. It gives to each voter a total of (say) 100 points to distribute among the candidates: their total points rank them. It is, however, subject to the Arrow paradox: when a candidate drops out the distribution of his points to others may change the outcome.

[15] A back of the envelope calculation suggests that rank-ordering n candidates takes $n(n + 1)/2$ time as versus $2n$ time for grading them.

9.2 Permits one candidate to be evaluated

MJ enables an expression of a jury's opinion on *one* competitor (wines, restaurants, skaters) and of an electorate's opinion on *one* candidate (in a significant number of elections there is only one candidate, e.g., to the U.S. House of Representatives), as does any method based on measuring. The majority-grade has meaning for one competitor and is a significant signal to a candidate, rivals, and the electorate; moreover, many voters have expressed delight that a candidate be given a "final grade." Absolute evaluations are of obvious importance for wines (medals), restaurants and hotels (stars), as well as athletic performances that require judges.

A method is *strategy-proof* when it is an optimal strategy for every voter to express his opinion honestly. The majority-grade of one candidage is strategy-proof. This follows from Moulin's theorem [54] because a voter's preference over the grades for a candidate is single peaked: the closer the majority-grade is to what he wishes the happier he is. The proof is simple. Suppose a voter wishes the candidate A with majority-grade α_A to have the majority-grade α_*. If $\alpha_* \succ \alpha_A$ giving a higher grade to A changes nothing, giving a lower grade either changes nothing or decreases the majority-grade (not in his intent); and if $\alpha_* \prec \alpha_A$ giving a lower grade to A changes nothing, giving a higher grade either changes nothing or increases the majority-grade (not in his intent). A similar argument shows it is *group strategy-proof in grading*: a group whose inputs are higher (or lower) than α_A can only lower (raise) the majority-grade.

Moreover, a candidate's distribution of grades is revealing. Contrast, for example, the merit profiles of Hollande (left), Sarkozy (right) and Bayrou (center) of Table 7 (taken from Table 3a). Many high and many low grades indicate a polarizing candidate. Sarkozy is polarizing, Hollande is too but less so, whereas Bayrou is more consensual. No such information is available when candidates are ranked instead of evaluated, or when only two grades are used.

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	*Fair*	*Poor*	*To Reject*
F. Hollande	12.48%	16.15%	16.42%	11.67%	14.79%	14.25%	14.24%
F. Bayrou	2.58%	9.77%	21.71%	25.24%	20.08%	11.94%	8.69%
N. Sarkozy	9.63%	12.35%	16.28%	10.99%	11.13%	7.87%	31.75%

Table 7. Merit profile (partial), 2012 French presidential poll (737 ballots).

9.3 Combats manipulation

The Gibbard-Satterwaite impossibility theorem [41, 66] proves that no method of ranking satisfying Axioms 1-7—so based on comparisons—is strategy-proof when there are at least three candidates. Yet, "It is important to ask under what circumstances it would be possible to design non-trivial strategy-proof decision rules, because strategy-proofness, when attainable, is an extremely robust and attractive property" ([12] Sect. 5).

Suppose a method M satisfying Axioms 1* and 2-8 ranks A above B, $A \succ_M B$ (IIA implies strategic manipulation may be studied for any two candidates

alone). A voter who prefers B to A can try to manipulate by lowering A's grade and raising B's. A method is *strategy-proof in ranking* if a voter who grades A above B can neither raise B's global measure[16] nor lower A's. As will be shown, there is no such method. So formulate a less demanding property: a method is *partially strategy-proof in ranking* if a voter who grades A above B can raise B's global measure implies he cannot lower A's, and if he can lower A's implies he cannot raise B's. A voter who gives the same grade to two candidates has no preference between them (when the scale is sufficient) so has no incentive to see one ranked above the other; indeed, his primary objective may be for them to receive the majority-grade he has given. Point-summing methods are clearly not partially strategy-proof in ranking (as was noted earlier, section 2.2).

Theorem 5 *(1) No method of ranking satisfying Axioms 1* and 2-8 is strategy-proof in ranking on the entire domain of possible opinions. (2) The majority-gauge rule $\succ_{MG}$ is partially strategy-proof in ranking on the entire domain when the language of grades is sufficient.*

The first part (1) is a watered-down form of the Gibbard-Satterwaite theorem, paralleling the version of Arrow's theorem given above. Their proofs both consist of showing that some of the conditions imply MR must be used for any two candidates, contradicting transitivity (because of the Condorcet paradox).

Proof. (1) Suppose a method $\succeq_M$ satisfying the axioms is strategy-proof in ranking on the entire domain. Axiom 8—IIA—implies it must be strategy-proof on any two candidates A and B with profile Φ. If a voter j gave A the higher grade change both of j's grades, giving A the highest and B the lowest; if voter j gave B the higher grade change both, giving B the highest and A the lowest; and if voter j gave both the same grade change nothing (she is indifferent). Strategy-proofness or monotonicity implies $\succeq_M$ must give the same outcome for this new profile on A and B. Doing the same for every voter in turn results in an opinion profile Φ' for which the method $\succeq_M$ necessarily gives the same outcome between A and B as does Φ.

In Φ' let m_A be the number of voters who give A the highest and B the lowest grade, m_B the number who give B the highest and A the lowest, and m_{AB} the number who gave both the same grade (m_{AB} is also the number who gave the same grade in Φ). If $m_A > m_B$ then A's grades dominate B's and monotonicity implies $A \succ_M B$; and if $m_A = m_B$ their sets of grades are identical so by Theorem 3 $A \approx_M B$. But this implies $\succeq_M$ coincides with the majority rule $\succeq_{MR}$. Since the domain is unrestricted, Condorcet's paradox shows transitivity is violated, a contradiction.

(2) Suppose $A \succ_{MG} B$, A's majority-grade is α^A and B's α^B, and voter j ranks B above A, $\alpha_j^B \succ \alpha_j^A$. If j can raise (p_B, α^B, q_B) then $\alpha_j^B \preceq \alpha^B$, implying $\alpha_j^A \prec \alpha_j^B \preceq \alpha^B \preceq \alpha^A$ so j cannot lower (p_A, α^A, q_A); whereas if j can lower (p_A, α^A, q_A) then $\alpha_j^A \succeq \alpha^A$, implying $\alpha_j^B \succ \alpha_j^A \succeq \alpha^A \succeq \alpha^B$ so j cannot raise (p_B, α_B, q_B). ■

[16] Examples of global measures are a candidate's mean of his grades for point-summing, his number of ticks for AV, his majority-gauge for MJ.

Lemma *A method of ranking $\succeq_M$ that satisfies Axioms 1* and 2-8 and is strategy-proof on the limited domain of polarized pairs of candidates must be consistent with the majority rule $\succeq_{MR}$ on polarized pairs, when the language of grades is sufficient.*

Proof. The proof is very similar to that of Theorem 5. Suppose $\succeq_M$ satisfies Axioms 1* and 2-8 and is strategy-proof on pairs of polarized candidates. Take any such pair with profile Φ. If voter j gave A the higher grade change both grades, giving A the highest and B the lowest; if voter j gave B the higher grade change both, giving B the highest and A the lowest; and if voter j gave both the same grade change nothing (he is indifferent). Since the new opinion profile on the two candidates is again polarized, strategy-proofness or monotonicity implies that $\succeq_M$ gives the same outcome. Doing the same for all voters gives an opinion profile Φ^* for which the outcome is the same as for Φ. The last part of the proof of Theorem 5 (1) now shows that $\succeq_M$ between any two polarized opinion profiles is $\succeq_{MR}$. ■

Theorem 6 *A method of ranking $\succeq_M$ that satisfies Axioms 1* and 2-8 and is strategy-proof on the limited domain of polarized pairs of candidates must coincide with the majority-gauge rule $\succeq_{MG}$ when the language of grades is sufficient.*

Proof. The Lemma together with Theorem 4 proves it. ■

Theorem 6 provides new evidence that MJ best resists strategic manipulation. It says that when two candidates are polarized—and sizable parts of an electorate often are—the majority-gauge rule is strategy-proof on those candidates. This is important for it is in these cases that voters are the most tempted to vote strategically but cannot; moreover, it is the unique method satisfying Axioms 1* and 2-8 that does so.

By bootstrapping ballots from several electoral experiments and polls (including the poll described above, Table 3a) the vulnerability of various methods to manipulation have been compared; the experimental results confirm that the majority-gauge better resists than first-past-the-post, approval voting, point-summing and Borda's method ([5] pp. 343-349, [8], [9], [42]). The unique method that competes with majority judgment in resisting manipulability is MR which is hardly surprising, but not the main point.

In comparing the strategic behavior of voters in elections using methods based on comparisons as versus MJ, Nagel [58] suggests that MJ may make it more likely for voters to indulge in exaggerating their grades up and down, not for rational reasons based on outside information (e.g., polls) but for psychological reasons made possible by the choice of one among several grades. "None of the binary or preferential systems gives an incentive to vote strategically in the absence of information, but majority judgment does." As he states, there is no way of knowing whether this will happen: the only test is experience. Experience to date shows it does not happen (see Section 9.2 below).

Take another perspective to explore what might happen. Suppose there are but two candidates, A and B, and that with the voters' honest evaluations A's grades dominate B's (this is not infrequent, see below). The MG-rule (and all

methods that satisfy the basic axioms) necessarily ranks A ahead of B, as it should. Moreover, if the pair is polarized the result must agree with MR, and the strategy-proofness of the latter implies voters cannot change the outcome by manipulating. When the pair is not polarized voters might manipulate by giving their preferred candidate the highest possible grade, the other the lowest possible. There are, then, two possible outcomes. (1) Their manipulation does not succeed (perhaps because the MG-rule is partially strategy-proof on the entire domain). Thus the MG-rule has resisted strategic manipulation, as it should, and implemented the correct outcome. (2) Their manipulation does succeed; but then, since the result changed, it must agree with MR. Yet A's grades dominated B's, so the majority rule with the voters' honest evaluations errs.

The MG-rule corrects the MR-rule in some cases: it is a step in the right direction. When many voters manipulate, at worst the MG-rule elects the MR-winner; when only few voters manipulate, majority judgment places A (whose grades dominate those of B) above B.

Strategy-proofness is a very important property of a method *if* it implements correct decisions. Dictatorship, for example, is strategy-proof, yet that attribute is not a sufficient reason to adopt it. Strategy-proofness is no virtue in a method that implements wrong decisions (as MR sometimes does). This is an important point that does not seem to have been appreciated in the social choice literature. First and foremost a method should guarantee a correct decision when voters or judges express themselves honestly; subject to that caveat, the method should resist manipulation as best as possible.

9.4 Is meaningful

The traditional theory eschews any scale as inputs: only comparisons count. Point-summing and other methods based on cardinal measures implicitly assume that there is an interval scale, which is rare; so in the language of measurement theory such methods are "meaningless." Between these extremes—no scale and a cardinal scale—lies an ordinal scale. Used by MJ—relying on the fact that (in Dahl's terms) there are uniformities in the expression of human feelings—it is "meaningful" according to measurement theory.

To be meaningful the particular representation used for a set of grades should make no difference in the ultimate outcome (e.g., comparisons of distances should not change when the scale is in meters rather than yards). Specifically, a method of ranking is *order consistent* if the order between any two candidates for some opinion profile Φ is the same for an opinion profile Φ' obtained from Φ by any increasing, continuous transformation ϕ of the grades.

Given the merit profile of a candidate $(\alpha_1, \ldots, \alpha_n)$, $\alpha_i \in \Lambda$ and $\alpha_i \succeq \alpha_{i+1}$ for all i, the kth-*order function* is α_k, the kth in the list of grades. The order functions are clearly order consistent. MJ is order consistent because it may be described as a *lexi-order* method: a first order function is applied and if it does not distinguish the order between some two candidates, a second order function is applied; if it does not distinguish between them, a third order function is

applied; and so on until the order between the two is distinguished. In fact, as was first shown in the context of the welfarist approach, the unique monotone and order consistent methods of ranking are the lexi-order functions [43, 29]. Welfarists have concentrated on the *leximin* social welfare ordering: the lowest order function α_n is applied first, if ties occur the next α_{n-1} is applied, and so on. MJ is quite different and may be described as follows: the "middlemost" order function is applied first (if there are two middlemosts because there are an even number of grades, the "middlemost" is the lower of the two); next the middlemost of the remaining grades is applied, and so on.

9.5 A reconsideration of the welfarist approach

Welfarism is an approach "in which social welfare is taken to be function of—and only of—individual utilities" (in the words of its originator A. Sen [69]). Motivated by Arrow's framework but including more information to avoid impossibilities, it is a very general, all encompassing theory that seeks to evaluate the relative goodness to society of alternative courses of action—and so to make a best choice—as a function of the utility or well-being each choice affords to its individual members (e.g., [67, 68, 56, 15, 16, 38]). It avoids the Condorcet paradox and also the Arrow paradox, the latter for a fixed number of alternatives. It has been said that "The grading model [of this paper] is formally identical to the model of social choice with interpersonally comparable utilities [that is, the welfarist model]" [37].

We disagree that the models are identical, but agree that they have important elements in common. Both approaches depend on measuring, welfarism uses utilities, MJ uses merit. Both are axiomatic, both pay attention to the meaningfulness problem, both appeal to IIA (though to its different formulations), both seek to avoid the traditional paradoxes. Welfarism goes beyond GNP, concerned as it is with the satisfaction of each individual in society. MJ seeks a practical methodology to rank skaters, wines and candidates that in addition combats strategic manipulation.

However, grades are not judges' or voters' utilities. There is no practical way by which voters in elections or judges of juries can express their utilities as inputs to a collective decision because utility is a function of the output not the input. To illustrate the point imagine n candidates are to be ranked and every place is important (as in skating). There would have to be $n!$ numbers, one for each possible ranking. And even if only one candidate is to be elected with first-past-the-post the satisfaction of a voter almost certainly depends on the scores of other candidates and not solely on who is the winner (which may explain why voters vote for candidates having no chance of winning). Reducing the utility of a voter to a single number for each possible candidate is unrealistic.

Utilities are measures of the satisfaction obtained with—or the relative well-being found in—alternate outcomes. Thus for example, Arrow [2] repeatedly uses the term "satisfaction" when talking about utilities, and von Neumann and Morgenstern assume that "the consumer desires to obtain a maximum of utility or satisfaction" ([73] p. 8)). Grades are absolute measures of the merit

of candidates used as inputs, not satisfactions. Utilities are relative measures used by voters to evaluate their satisfaction with the ouputs. A good example of the difference is the 2002 French presidential election. The utility to leftists for a Chirac (rightist) victory over Jospin (socialist) was the lowest possible; but their utility for a Chirac victory over Le Pen (extreme rightist) was the highest possible, so they were delighted to see Chirac defeat Le Pen with 82% of the votes in the second-round. However, these same voters would most likely have given to Chirac (who had a mere 19.9% of the votes in the first round even though the incumbent President) a grade of *Fair* or *Poor* (and to Le Pen the grade *To Reject*) were he standing against Jospin, Le Pen or anyone else. MJ makes no overall assumptions concerning voters' or judges' (unkown) utilities on the outputs. Note, however, that if the inputs or grades depend on the set of candidates—i.e., if voters change their grades when candidates are added or dropped—then the Arrow paradox cannot be avoided. To avoid this the scale must be fixed and absolute so that more or fewer candidacies does not change the inputs or messages of other candidates.

There are technical differences between welfarism and the grading model as well. The first axiom (called anonymity) of an axiomatic account of welfarism ([56] p. 33) states that a social welfare ordering function is indifferent between any two identical sets of utilities, independent of which agents expressed which utilities. In the grading model this is a result (Theorem 3, that shows only the merit profile counts, not the opinion profile). This difference stems from the fact that welfarism uses the Arrow version of IIA (for a fixed number of alternatives) whereas the grading model uses the Nash-Chernoff formulation (for a variable number of alternatives).

Utilitarianism is an instance of the welfarist approach where each alternative is evaluated in terms of the total sum of utilities it affords its members; its realization in voting is a point-summing method. Relative utilitarianism is utilitarianism where each individual's utility is scaled so that her preferred alternative is 1, her least preferred is 0 [30]. This erases all absolute evaluations and gives a 1 to a candidate when highly regarded by a voter or when lowly regarded by another (so long as it is the highest: but many voters have a low regard for all candidates). It does not satisfy Nash-Chernoff-IIA since adjoining an alternative can rescale the inputs so change the outputs.

Sen [69] goes on to plead for positive results: "One object of noting an impossibility is to question the initial choice of axioms and to suggest what variations in the axiomatic structure should be reconsidered." This, of course, has been the primary thrust of the grading model in the context of voting. Recently a change in the welfarist program has been proposed using ordinal scales to measure well-being [53]. The authors point out that "the social welfare functional approach to social choice theory fails to distinguish between a genuine change in individual well-beings from a merely representational change due to the use of different measurement scales" ([53], p. 127). This makes it possible for a change in well-beings to be compensated by a change in scale, hiding the real change, and so resulting in the same ordering of the alternatives. To distinguish states of well-being they take a utility scale to be values that are

linearly ordered from best to worst and that are endowed with interpretations of what these values mean and thereby (as they state) formulate a model that is in the spirit of the grading model for voting.

In any case, a model's meaning, formally or informally—how it is interpreted, how it fits the purpose for which it is formulated—is what counts.

9.6 Close to practice

Increasingly practitioners use measures as inputs instead of rank-orders, and increasingly they head in the direction of MJ in that they eliminate equal numbers of highest and lowest grades to combat strategic manipulation. One example is diving (described above). A more insightful example is figure skating.

Since the end of the 19th century judges' inputs in figure skating were and continue to be grades. Before 2004 their only role was to determine each judge's rank-ordering of the competitors, the rank-orders constituting the inputs to determine the jury's rank-order. Then came the 1997 men's European championships. A French skater, Candeloro, who with the other first five finishers had already performed, was in 3rd place. After the performance of the (eventual) 6th place finisher, Vlascenko, Candeloro leaped to 2nd place: Arrow's paradox—a "flip-flop"—once again, this time the epsilon that tipped the scales. There were nine judges, four placed Vlascenko 6th among the top six contestants, three placed him 5th, one 4th, and one 1st. The use of comparisons as inputs to the method for determining the outcome of competitions together with the (suspected strategic) vote of one judge led to the adoption of the OBO ("one-to-one") rule in 1998. Invented internally the OBO rule was independently proposed by Dasgupta and Maskin in 2004 [28, 30]. Its inputs are rank-orders. It orders the contestants by their number of MR wins in all head-to-head encounters, with ties broken by Borda's rule. The 2001 Four Continents Figure Skating Championships shows that in practice it is subject to both the Condorcet and Arrow paradoxes [9]. A year later, at the 2002 Olympic Games held in Salt Lake City, OBO's vulnerability to strategic manipulation in the pairs-skating caused a major scandal and mayhem in the skating world. The story, in short: a Russian pair placed 1st, a Canadian pair 2nd; the public and some experts expressed outrage; a French judge confessed (then later denied) having favored the Russians under pressure; the decision was changed to a tie for 1st; deep divisions ensued in the skating world over what method to use; finally, in 2004, OBO was dropped and a new system adopted.

The new system uses grades but not to deduce judges' rank-orders. There are 12 judges. To combat strategic manipulation the system first eliminates the scores of three randomly selected judges, then for each performance eliminates the highest and lowest grades of the nine remaining judges, using the average of the seven scores that remain. It is doubtful that the random elimination of three judges does anything more than discard information: the three could be honest so the impact of any remaining manipulators would be greater. Gymnasts are ranked with a roughly similar system (as is diving, as seen above).

Practice, accumulated experience—"market forces"—has increasingly pushed

practical people to (1) replace comparisons as inputs by measures and (2) ignore equal numbers of highest and lowest grades to combat manipulation. Majority judgment simply goes all the way: it eliminates as many equal numbers of highest and lowest grades as required to distinguish any two competitors.

9.7 Eliminates the Condorcet and Arrow paradoxes

A principal aim in developing MJ was to avoid the Condorcet and Arrow paradoxes. It avoids Arrow's paradox because using a scale of absolute grades implies that a removal from or addition to the list of candidates or competitors does not change the voters' or judges' evaluations of the others of the list.

As was emphasized earlier, these paradoxes are important: they occur. The Condorcet paradox is not often observed in voting because few methods ask voters for rank-orders as inputs. But even when they do, as for example the Australian system, the raw data is unavailable so it is impossible to check: instead tables give the numbers of times each candidate is in first-place at each stage of the alternative vote procedure. Estimates of the probability for the Condorcet paradox to occur have been made by many. It grows to certainty as the number of candidates and of voters increases. Under the impartial culture assumption Fishburn [36] estimates the probability of the Condorcet paradox to be .369 for seven candidates as the number of voters goes to infinity. However, as early as 1876 C. L. Dodgson (aka Lewis Carroll) noted that strategic behavior could—and in his experience and subsequent analysis would—provoke the Condorcet paradox, what he called a "cyclic majority" [31]. In any event it is a clear imperative that elections and competitions must have transitive outcomes.

With traditional methods the Arrow paradox occurs frequently. When the performances of competitors follow one another it occurs openly (as in skating). In elections its occurrence can only be inferred, but its potential impact is enormous. Recall the election of George W. Bush instead of Al Gore in 2000 due to the candidacy of Ralph Nader in Florida; the election of Jacques Chirac in 2002 instead of Lionel Jospin due to the presence of two other socialist candidates; the election of Nicolas Sarkozy instead of François Bayrou (eliminated in the first round) in 2007; the election of Bill Clinton in 1992 due to the candidacy of H. R. Perot; the election of Woodrow Wilson in 1912 instead of Teddy Roosevelt or perhaps W. H. Taft (both Republicans though Roosevelt ran as the candidate of the Progressive or Bull Moose Party), both candidates.

9.8 Works well

MJ is used. It ranked the five finalists for the 2009 Louis Lyons Award for Conscience and Integrity in Journalism given by the Nieman Foundation of Harvard University [9]. It has been tested in French presidential elections: in an experiment carried out in parallel with the first round of the 2007 election in Orsay ([5] pp. 10-16); in two national polls preceding the 2012 presidential election conducted by OpinionWay (one of which is described above [8], the other in [9]); in several experiments carried out in parallel with the first round

of the 2012 French Socialist Party presidential primaries [42, 9]; and two French web sites invited viewers to use MJ online before the 2012 Socialist primary and then before the 2012 national election (designed independently by them [63, 70]). The fusion of two research groups into one computer sciences laboratory at Paris-Diderot University in 2015 required the choice of a name. Ten names were considered, 95 persons expressed their opinions using MJ with a scale of five grades: *Good, Rather Good, Not Bad Not Good, Rather Bad, Bad*. MJ has been used for several years to rank-order applicants for faculty positions by the mathematics department of the University of Montpellier; it was also used for the same purpose in 2015 by the economics department at the École Polytechnique, the computer sciences department at Paris-Dauphine University, and the industrial engineering department at the University of Chile.

In these instances voters or judges displayed no difficulty in assigning grades to candidates. Evaluating even many candidates or choices in a scale of grades that carry understandable meanings seems to be cognitively simple. In contrast, voters faced with many candidates using first-past-the-post have a considerably more difficult task even when they wish to express themselves honestly but realistically since their "favorite" may have no chance of being elected.

Recent Pew Research Center national political surveys ([60], pp. 11-12) provide a striking example of MJ in use and show that polling professionals believe asking registered voters to evaluate candidates is a natural way of eliciting opinions. Pew conducted four such surveys in mid-January, late March, mid-August, and late October 2016 (just before the election on November 8). One of the questions asked in each survey was:

> "Regardless of who you currently support, I'd like to know what kind of president you think each of the following would be if elected in November 2016. . . . [D]o you think (he/she) would be a great, good, average, poor or terrible president?"

These words provide a reasonable scale of evaluations.

Pew Research did not know that this information may be used to deduce a rank-order of the candidates. In late March there were three remaining Republican candidates—Cruz, Kasich, and Trump—and two remaining Democratic candidates—Clinton and Sanders. The results are given in Table 8a.

	Great	*Good*	*Average*	*Poor*	*Terrible*	*Never heard of*
John Kasich	5%	28%	39%	13%	7%	9%
Bernie Sanders	10%	26%	26%	15%	21%	3%
Ted Cruz	7%	22%	21%	17%	19%	4%
Hillary Clinton	11%	22%	20%	16%	30%	1%
Donald Trump	10%	16%	12%	15%	44%	3%

Table 8a. Pew Research Center poll results, March 17-27, 2016 [60].

The March MJ-ranking is given in Table 8b. *Never heard of*—as damning in this context as *Terrible*—is taken as its equivalent.

	Above Maj-grade p	Maj-grade $\alpha \pm \max\{p, q\}$	Below Maj-grade q
John Kasich	**33%**	*Average+*	29%
Bernie Sanders	36%	*Average−*	**39%**
Ted Cruz	29%	*Average−*	**40%**
Hillary Clinton	33%	*Average−*	**47%**
Donald Trump	38%	*Poor−*	**47%**

Table 8b. MJ-ranking, Pew Research Center poll, March 17-27, 2016 [60].

MR, used in the primaries of both parties, led to the nominations of Clinton evaluated *Poor* or worse by 47% and Trump evaluated *Poor* or worse by a majority of 62%: the least respected major candidates in more than a half century according to many independent assessments.

The distributions of Clinton's and Trump's grades in answer to the four 2016 Pew polls are given in Table 8c, their respective majority-gauges in Table 8d. In contrast with many MR polling results, the evaluations of Clinton and Trump and, in particular, their majority-gauges remain remarkably stable throughout the entire year despite the ups and downs—the many bombastic revelations and accusations—of the campaign.

		Great	*Good*	*Average*	*Poor*	*Terrible*	*Never heard of*
Clinton:	January	11%	24%	18%	16%	28%	2%
	March	11%	22%	20%	16%	30%	1%
	August	11%	20%	22%	12%	33%	2%
	October	8%	27%	20%	11%	34%	1%
Trump:	January	11%	20%	12%	14%	38%	5%
	March	10%	16%	12%	15%	44%	3%
	August	9%	18%	15%	12%	43%	3%
	October	9%	17%	16%	11%	44%	2%

Table 8c. Pew Research Center poll results, 2016 [61].

	Clinton			Trump		
	p	$\alpha\pm$	q	p	$\alpha\pm$	q
January	35%	*Average−*	**46%**	43%	*Poor−*	**43%**
March	33%	*Average−*	**47%**	38%	*Poor−*	**47%**
August	31%	*Average−*	**47%**	42%	*Poor−*	**46%**
October	35%	*Average−*	**46%**	42%	*Poor−*	**46%**

Table 8d. Majority-gauges, Pew Research Center polls, 2016.

Another real use of MJ was initiated by several concerned French citizens in 2016—an engineer, a lawyer and a mathematician. Called LaPrimaire.org the idea was for citizens independent of any political parties to freely nominate a candidate for the 2017 French presidential election. The entire effort was carried out on the web, including campaigning and voting. The organizers considered

the various known voting methods and chose to use MJ. An initial slate of some 200 candidates was whittled down to twelve. Then in a first vote each voter was asked to evaluate five prescribed candidates of the twelve on the scale *Excellent, Very Good, Good, Passable, Insufficient.* The assignment of five (of the twelve) candidates to voters was done randomly but in such manner that each candidate was evaluated by approximately the same number of voters. The twelve were then ranked by MJ and the five leaders designated to participate in the final campaign and vote [47]. The reason for not directly voting on all twelve was to incite voters to invest the time to make a careful comparative study of the proposals and approaches of each candidate: evaluating twelve candidates was seen as asking for too great an investment of time and effort.

The final election among the five was conducted using MJ with the same scale; 32,685 persons voted; each candidate's grades dominated those of the candidates lower in the MJ-ranking. The winner's majority-grade was *Excellent*, the runner up's was *Very Good*, all of the last three's were *Good* (see [48] for a complete description) The results were well accepted and MJ was included in the electoral reforms proposed by the winner and others. The winner will try to qualify as a candidate in the first-round of the official presidential election. This means obtaining 500 signatures from elected officials throughout France.

10 Majority judgment: objections

10.1 Not Condorcet-consistent, admits "no-show paradox"

Critics of MJ have primarily focussed on these two points [17, 19, 32, 35, 37].

The first has been clarified: while Condorcet-consistency is not a desirable property in all circumstances, Condorcet-consistency on the restricted domain of polarized pairs of candidates characterizes the majority-gauge. It is reasonable to recall, however, that most of the widely advocated methods are not Condorcet-consistent when voters express themselves honestly, including approval voting, point-summing, Borda's, first-past-the-post, and the alternative vote (or instant run-off voting). However, MJ is solidly based on the fundamental idea of majority—not on the majority's preference on pairs of candidates—but on the majority's evaluation of each candidate's merit.

The no-show paradox occurs when it is better for a voter not to vote than to express his opinion sincerely because his vote can tip the scales against his favorite candidate. However, every Condorcet-consistent method admits the no-show paradox [55]. Moreover, the only methods based on measuring that exclude the no-show paradox are point-summing methods ([5] chpt. 17), but they should be rejected for reasons given earlier. In any case the no-show paradox is less likely to occur with MJ than for MR to result in a tie and is of little importance in practice (as argued in [9] and a forthcoming paper [10]). Its importance is insignificant when compared with the serious problems of methods of election, the Arrow and Condorcet paradoxes and strategic manipulation.

10.2 Grades: for experts not voters

One critic declared [57]: "[As you have shown] methods based on grades are necessary when the judgments of experts are aggregated since nuances in their opinions have objective value (recall the French judge of figure skating who was punished precisely because it could be said that her grade was wrong). In political elections [however] there is no question of obtaining such objectivity and grades make no sense. Thus the two approaches have distinct domains of application..." "I remain convinced that grades will never work in certain kinds of collective decision making including political elections and small committees when stakes are high." Another [32] stated: "More importantly, there is a fundamental difference between experts grading a wine competition and voters grading candidates. One can easily imagine two voters agreeing completely on all objective aspects of a candidate, e.g., policy preferences, honesty, future states of the world should he be elected, etc., but radically disagree on the grade. If one of the voters is pro-choice and the other is anti-abortion it is certainly plausible that one voter will grade the politician *Excellent* and the other as *Reject*. But if that is the case, in what sense is grading objective since they agree on all objective measures of the candidate?" A third [19] wrote "[I]n many political elections, I'm afraid, voters would [...give] their favorites the maximum grade and their most serious competitors the minimum grade."

A thermometer is objective. If judging figure skaters or wines was "objective" there would be an instrument to measure their merits; or, one single judge would suffice. The same is true of judging candidates for office. The fundamental difficulty is that judging skaters, wines, and candidates is subjective precisely because judges and voters have different opinions concerning them in a host of different dimensions: there is no thermometer that "objectively" measures any opinion in any dimension.

Juries and elections are attempts to make what is subjective "objective," or as close to objective as possible. It is incontestable that voters have nuances of opinion on candidates, just as judges have nuances of opinion on the varying performances of skaters or the varying properties of wines: every skating and wine competition proves it, as does every use and experiment with majority judgment. For example, in the famous 1976 "Judgment of Paris" [7] that pitted Bordeaux against California wines, the jury consisted of eleven well-known connoisseurs. They used the standard French grading system of 0 (low) to 20 (high). Not one of the ten wines received a higher grade than 17, not one a lower grade than 2. Despite the judges' expertise and objectivity and the very high quality of all the wines, one wine received[17] 17, 15, 13, 12, 11, 10, 10, 9, 8, 7, 2. A voter who bases her vote primarily or even completely on the intensity of a candidate's stance on the pro-choice to anti-abortion spectrum displays no greater lack of "objectivity"; nor, indeed, does a voter that bases his evaluations on the glamor of candidates. In a democracy voters must have the right to express themselves as they wish according to their conceptions of "objectivity,"

[17]There is a slight discrepancy in how these grades are reported. Elsewhere we have erred, reporting 11.5 for the fifth. It should be 11.

not academics' views of what is "objective."

The stakes of an election may be high and critics may be afraid that voters will be driven to give their highest grades to their favorites, the lowest to their most serious competitors, but experience does not confirm these opinions. Simple inspection shows it is not true of the LAMSADE Jury (Table 2a). It is not true of the 2007 French presidential voting experiment (where twelve candidates ran, MJ used a scale of six grades, and 1,733 voters participated [5] pp. 112-115, [6]), nor of the 2012 French presidential poll described above (where ten candidates ran, MJ used a scale of seven grades, and 737 voters participated) as may be seen in Tables 9a,b.

	Outs.	*Exc.*	*V.Good*	*Good*	*Fair*	*Poor*	*Rej.*	Total
2007:	–	0.69	1.25	1.50	1.74	2.27	4.55	12
2012:	0.38	0.64	1.00	1.15	1.49	1.86	3.48	10

Table 9a. Average number of grades per ballot, 2007 French presidential election experiment and 2012 French presidential poll.

	Outs.	*Exc.*	*V.Good*	*Good*	*Fair*	*Poor*	*Rej.*	Total
2007:	–	52%	37%	9%	2%	0%	1%	100%
2012:	34%	35%	23%	6%	1%	0%	0%	100%

Table 9b. Distributions of highest grades per ballot, 2007 French presidential election experiment and 2012 French presidential poll.

In every experiment or poll using MJ in political elections, the lower the grade, the more it is used. This simply reflects the growing frustration with politicians. In the 2007 experiment half of the voters did not use the maximum grade; in the 2012 poll two-thirds of the voters did not use the maximum grade and a third used neither of the two highest. The significance of these distributions have been challenged by some who say that "[voters gave intermediary grades] because they knew this was only an experiment and thus felt that they could afford to be honest" [19].[18]

The Doctoral Council of the Université Paris-Dauphine had one extra fellowship—possibly two—to award in the early summer of 2015. There were three candidates, one in computer sciences S_c, one in economics S_e, and one in mathematics S_m. Sixteen members of the Council $J_1, \ldots, J_{16}$, representing the university's five disciplines (law and social sciences in addition to the three mentioned), voted. In keeping with an often used scheme for classifying students, the scale of grades used was A (highest), B, and C (lowest). This was not an experiment. The stakes were high as all who are in academia know. The opinion profile is in Table 10a, the merit profile and face-to-face votes together with the Borda counts in Table 10b.

The merit profile shows that S_e dominates S_m which in turn dominates S_c, so any method satisfying the axioms of Theorem 3 ranks the students in that order (including majority judgment, approval voting, and point-summing).

[18]This point of view denies the significance of polls.

Borda's method puts S_e in first place but reverses the order of the other two; Condorcet's also places S_e first but declares a tie between S_m and S_c. Voters did not confine themselves to the highest (A) and lowest (C) grades: 12 B's were given (together with 20 A's and 16 C's). Only five of the 16 voters confined themselves to A's and C's. Moreover, although there were three grades and three candidates only six voters rank-ordered them by using all three grades.

	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8
S_e	B	A	A	A	A	A	A	B
S_m	B	C	B	B	A	A	C	A
S_c	A	C	C	A	C	B	B	C

	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}	J_{15}	J_{16}
S_e	A	B	A	B	A	B	C	C
S_m	C	A	C	C	B	C	A	A
S_c	C	C	B	A	A	A	C	C

Table 10a. Opinion profile, extra fellowship, Doctoral Council, Université Paris-Dauphine, summer 2015.

Merit profile:

	A	B	C
S_e	9	5	2
S_m	6	4	6
S_c	5	3	8

Face-to-face majority votes:

	S_e	S_m	S_c	Borda Score
S_e	–	10.5	10.0	10.25
S_m	5.5	–	8.0	6.50
S_c	6.0	8.0	–	7.00

Table 10b. Merit profile and face-to-face majority votes, extra fellowship, Doctoral Council, Université Paris-Dauphine, summer 2015.

Recall that the LAMSADE Jury's behavior was similar (Table 2a): every judge used but four grades though five were available and there were six candidates; one judge did not used the highest grade; and only one used the lowest.

With rare exceptions voters and judges have not to date confined themselves to maximum and minimum grades in experiments, polls, and *real uses* of MJ.

10.3 Not faithful: dissimilar understandings of grades

Some have said that voters do not have a common understanding of the meanings of grades, the scales are not faithful. But what about all other methods? Do voters mean the same thing when casting first-past-the-post or approval votes? Clearly not. Do voters mean the same thing when their inputs are rank-orders? Again, clearly not, for in rank-ordering candidates two voters who put a candidate in first (or any other) place may have vastly different opinions about them ranging from excellent down to mediocre. MJ's grades can only improve the commonality of meaning: two *Excellents* have much more in common than two first-ranked candidates, two *To Rejects* have much more in common than two last-ranked candidates or two *Disapproves*.

A crucial feature of a MJ ballot aims at promoting interpersonal comparisons of grades: judges and voters are charged with answering precise questions (see Section 2.2). This is not true of other methods. First-past-the-post and AV, for example, ask nothing other than ticking boxes: voters give one or several ticks in response to their own questions. The Australian system asks voters for their rank-orders on candidates but as just mentioned two identical rank-orders may carry vastly different meanings. So in all these cases there is no valid interpersonal comparisons of inputs.

The grades used in the following instances are clear for their purposes:

- *Good, Rather Good, Not Bad Not Good, Rather Bad, Bad*, devised and used at Paris-Diderot University to choose the name of a new laboratory;
- 0 to 10 in jumps of 1/2 (together with detailed explanations of their meanings) as used and carefully explained in diving;
- any of the many traditional academic scales such as the American A down to E or F, the French 0 up to 20, or the German *sehrgut* (very good), *gut* (good), *befriedigend* (satisfactory), *ausreichend* (sufficient), *mangelhaft* (deficient);
- *Excellent, Very Good, Good, Passable, Poor, To Reject*, as has been offered as possible answers to a precise question in voting.

The words have understandable meanings and the numbers are defined. Both acquire—and have acquired—meaning in use. Wittgenstein's famous phrase—"the meaning of a word is its use in the language"—is operative. There may, of course, be severe judges who tend to give low grades and lenient ones who tend to give high grades (routinely accepted when evaluations pretend to be objective as in grading students). In voting—where there is no objective measure of the candidates and no reason for voters to agree—differences must all the more be respected as expressions of voters' opinions, their democratic right.

The six grades given above were used in the 2007 French presidential experiment at Orsay by the 1,733 persons who participated [5, 9]. They constituted a wide spectrum from different walks of life. Mild shades of dissimilarity may exist in the use of two neighboring grades such as *Good* and *Very Good*, but to a much lesser degree than the inputs of other methods of voting. Extensive analysis of the 2007 Orsay experiment showed that the frequencies of use of each of the grades were substantially the same throughout the electorate (of 1,733 voters), supporting the idea that voters had similar understandings of the grades ([5] chpt. 15). Table 11 is but one piece of evidence for this.

Precinct	*Excellent*	*Very Good*	*Good*	*Passable*	*Poor*	*To Reject*
1st	0.7	1.2	1.5	1.7	2.3	4.7
6th	0.7	1.2	1.4	1.7	2.3	4.6
12th	0.7	1.4	1.6	1.8	2.2	4.3

Table 11. Average number of grades per ballot (total of 12), 2012 French presidential experiment (Orsay).

Finally, voting experiments have conformed with intuition: candidates may often be identified simply by the distributions of their grades (seeing how consensual, how polarized, how held in esteem or in contempt they may be). Moreover, when the entire electorate expresses itself (in polls or experiments) in a presidential vote as many as half the voters never use *Excellent* and *Very Good* whereas in a party's presidential primaries when only the party faithful express themselves the grades are much higher with *Excellent* and *Very Good* used frequently. Uses and experiments with MJ lead us to believe that for voters grades have clear meanings.

These (and many other) examples are instances of faithful scales; in any case, they are much more faithful than inputs of first-past-the-post, AV, or rank-orders.

10.4 Order changes when neighboring grades merged

A method based on a scale of grades Λ is *scale-stable* if when it ranks one candidate above another it does the same when two neighboring grades are merged into one grade. Scale-stability seems a desirable property.

Theorem 7 *No method satisfying Axioms 1* and 2-8 is scale-stable on the entire domain of opinions. Every method satisfying the Axioms is scale-stable on the limited domain of pairs of candidates where one's grades dominate the other's.*

Proof. Suppose a method M satisfying the axioms and using the grades $\Lambda : \lambda_1 \succ \lambda_2 \succ \cdots \succ \lambda_k$ is scale-stable and that $A \succ_M B$ when A's grades do not dominate B's. No domination implies there is at least one grade λ_i $(i < k)$ for which B has more grades equal to λ_i and higher than does A. Merge λ_i with the next higher grade, then the two with the next higher, and continue until λ_i and all higher grades are merged into one grade $\lambda_\uparrow$; and similarly merge all grades lower than λ_i into one grade $\lambda_\downarrow$. Since M is scale-stable $A \succ_M B$ using the two grades $\lambda_\uparrow \succ \lambda_\downarrow$. But on the two grades B's dominate A's, a contradiction that proves the first assertion. Theorem 3 proves the second assertion. ■

In practice dominations are frequent, so scale-stability is satisfied for those pairs. The LAMSADE Jury (Table 2b) had 6 competitors, 15 comparisons, 13 were dominations. The 2012 Socialist presidential primaries had 6 competitors, 15 comparisons: in two different precincts 13 were dominations (the rankings were different) [42]. The 2007 French presidential election had 12 candidates, 66 comparisons: in the Orsay experiment that included 1,733 participants, 50 were dominations [5, 6]. The French 2012 national presidential poll reported on above (Table 3a) had 10 candidates, 45 comparisons, 34 were dominations. There were 10 possible names for the newly formed computer sciences laboratory at Paris-Diderot University, 45 comparisons, 40 were dominations. LaPrimaire had 5 candidates, 10 comparisons, all were dominations.

The results of this paper all suggest that there should be a scale with a sufficient number of grades to avoid any voter from giving the same grade to

two candidates when she sees a difference in their merit. For if there are too few grades in the scale a voter who sees a difference but believes two candidates merit the same of the available grades could be motivated to manipulate the grades to express his difference of opinion, thereby violating the conclusion of Theorem 6. And if A is ranked above B with a sufficient set of grades, Theorem 7 shows that a restricting the set may induce error by changing the outcome.

However, practicality, the need for a common usage of grades, and experimental evidence pushes for fewer grades. In the most cited paper of the first hundred years of the *Psychological Review* G. A. Miller [52] concluded, "There is a clear and definite limit to the accuracy with which we can identify absolutely the magnitude of a unidimensional stimulus variable. I would propose to call this limit the *span of absolute judgment*, and I maintain that for unidimensional judgments this span is usually somewhere in the neighbourhood of seven." Voters and judges are asked for unidimensional evaluations in that they must themselves integrate their appreciation for all aspects of a performance or a candidate's competence. Experts in judging figure skaters, divers, gymnasts, or wines may, however, have the finesse to discern refinements that demand finer scales; in figure skating, for example, judges use a scale of 13 grades (for each part of a performance) and in diving a scale of 21 (for each dive).

No. of grades:	1	2	3	4	5	6	7	Total
2007 (12 candidates):	1%	2%	10%	31%	42%	14%	–	100%
2012 (10 candidates):	1%	6%	13%	31%	36%	13%	1%	100%

Table 12. Percentages of ballots with k grades, 2007 French presidential election experiment and 2012 French presidential poll.

The evidence with MJ in voting concords with Miller's finding (Table 12). In the 2007 experiment where voters were offered a scale of six grades only 14% used all six although there were twelve candidates. In the 2012 poll voters had a scale of seven grades, yet the number of grades they used is remarkably similar to that of the participants in the 2007 experiment, only 1% using all seven grades to evaluate ten candidates. This suggests that at least five grades are necessary, six is a good number, seven is unnecessary.

	Hollande	Bayrou	Sarkozy	Mélenchon	Le Pen	Borda Score	FPP
Hollande	–	51.6%	53.9%	68.5%	64.1%	59.5%	1st
Bayrou	48.4%	–	56.5%	59.4%	70.5%	58.7%	5th
Sarkozy	46.1%	43.5%	–	50.5%	65.7%	51.4%	2nd
Mélenchon	31.5%	40.6%	49.5%	–	59.7%	45.3%	4th
Le Pen	35.9%	29.5%	34.3%	40.3%	–	35.0%	3rd

Table 13. Face-to-face majority votes, 2012 French presidential poll.

Experimental evidence shows too few grades can induce error. The 2012 presidential poll asked participants to vote in the ten face-to-face races between each pair of the five principal candidates using the usual MR (asking for all 45 races was asking them too much). This suffices to find the rank-order among the

five according to the Condorcet and Borda methods (Table 13). The methods of Condorcet (or MR), Borda and MJ—all of which use richer inputs—concord to give the same ranking. The first-past-the-post (FPP) result is very different.

In an experiment done in parallel with the 2012 French presidential election using approval voting [14] the data was adjusted to conform with the actual first-round national results as was done for the results given in Tables 3b and 13, so the two results are comparable. Approval voting orders the five candidates differently than the common order determined by the methods of Condorcet, Borda and majority judgment as may be seen in Table 14. Whereas the latter three methods place Bayrou comfortably ahead of Sarkozy, approval voting places him behind; and Le Pen drops to 5th not 8th as she does with the richer language of majority judgment. This suggest that two grades are insufficient.

Hollande	Sarkozy	Bayrou	Mélenchon	Le Pen
49.44%	40.47%	39.20%	39.07%	27.43%

Joly	Poutou	Dupont-Aignan	Arthaud	Cheminade
26.69%	13.28%	10.69%	8.35%	3.23%

Table 14. Approval voting results, 2012 French presidential experiment (Strasbourg, Louvigny and Saint-Etienne) [14].

It is well known that different questions or words elicit different answers. Designing a scale is a practical question that goes well beyond mathematical modelling. Users of MJ to date—e.g., the Louis Lyons jury, university juries charged with rank-ordering applicants for faculty positions, staff determining a name for their research group, the Pew Research Center's surveys seeking to understand the electorate's opinions, LaPrimaire's vote on the web—have had no difficulty in choosing a scale and using it. Nor have those who have chosen and carefully defined numerical scales for evaluating divers, figure skaters or gymnasts. In elections a scale once chosen should be treated as a constitutional clause: fixed and unchanging (but with difficulty subject to amendment). The law cannot mandate restrictions on a voter's domain of opinions but it can mandate a definitive choice of a common language of grades.

10.5 Nothing but the median with tie-breaking rules

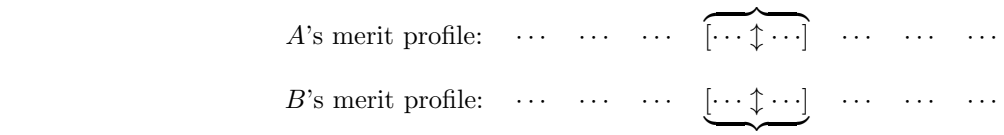
MJ has been described as a rule of ranking based on the median (as opposed, for example, to the average) together with "an elaborate set of rules for breaking ties. These are plausible, but there are other tie-breaking rules that would probably work just as well" [19]. This statement misleads: MJ invokes no tie-breaking rules, it is based on one simple majoritarian idea.

A succinct complete description of MJ is this:

> For each pair of competitors ignore as many equal numbers of highest and lowest grades of their merit profiles as possible so that domination or consensus in the middlemost blocks decides when there are

few judges, or domination in the middlemost blocks decides when there are many voters.

A simple diagram pictures the method:



The middlemost blocks (bracketed around the center, ↕) for which A's grades dominate B's or is more consensual than B's places A above B. There are the same number or percentage of grades to the left and to the right of the center; the same number to the left and to the right of the center within the brackets; and within the brackets the competitors' grades are identical except for the left- and right-most. Sir Francis Galton had the seeds for the right idea in 1907 [39, 40] when he asked "How can the right conclusion be reached?" and answered "I wish to point out that the [evaluation] to which least objection can be raised is the *middlemost* [evaluation]" (he used the word "estimate" instead of "evaluation").

B:	*Excellent*	*V. Good*	*V. Good*	*V. Good*	*V. Good*	*Good*
C:	*Excellent*	*Excellent*	*V. Good*	*V. Good*	*Good*	*Passable*

B:	*V. Good*	*V. Good*	*V. Good*	*V. Good*
C:	*Excellent*	*V. Good*	*V. Good*	*Good*

Table 15a. Merit profile and truncated merit profile to rank students B and C, LAMSADE Jury.

For an example with few judges take students B and C of the LAMSADE Jury (Table 15a). Eliminating the highest and lowest grades of the merit profiles gives the smallest middlemost truncated merit profiles that differ: in the latter the inside grades are identical, only the extremities differ. B's truncated profile is more consensual than C's so MJ ranks B above C.

For an example with many voters take Hollande (H) and Bayrou (B) of the 2012 national presidential poll (Table 15b). Eliminating[19] $(50 - 4.95 - \epsilon)\%$ of the highest and the lowest grades of the merit profiles yields the smallest middlemost truncated merit profiles that differ: in the latter the inside grades are all *Good* ($2 \times 4.95 = 9.90\%$ of them), only the extremities differ. The lowest grades of each are the same, *Good*; Hollande's highest is *Very Good*, Bayrou's *Good*; Hollande's truncated profile dominates Bayrou's, so MJ ranks Hollande above Bayrou. That is the majority decision.

[19] ϵ may be thought of as one single grade.

	Outst-anding	*Excel-lent*	*Very Good*	*Good*	↕	*Good*	*Fair*	*Poor*	*To Reject*
H	12.48%	16.15%	16.42%	4.95%	↕	6.72%	14.79%	14.25%	14.24%
B	2.58%	9.77%	21.71%	15.94%	↕	9.30%	20.08%	11.94%	8.69%

	Very Good	*Good*	*Good*	↕	*Good*
H	ϵ%	0%	4.95%	↕	4.95+ϵ%
B		ϵ%	4.95%	↕	4.95+ϵ%

Table 15b. Merit profiles and truncated merit profiles to rank Hollande and Bayrou, 2012 French presidential poll.

This description highlights the consensual, majoritarian nature of MJ. It shows that MJ is based on a genuine generalization of the notion of median to the middlemost grades that invokes no "tie-breaking" rules.[20]

11 Conclusion

The intent of this article is to convince readers of a number of main points.

Several concern the drawbacks of methods based on the traditional paradigm where voters compare candidates and the "ideal" winner is a candidate who is preferred by a majority to another candidate:

- The domination paradox shows that ideal has a major flaw: majority rule for electing one of two candidates is not, with the exception of polarized electorates, a good method.
- Condorcet consistency, in consequence, is not a desirable property, contrary to the widely held view, and should certainly not be considered axiomatic.
- Comparisons as inputs to methods of voting are insufficient expressions of opinions and should be replaced by ordinal measures.

Several others concern methods based on voters' evaluations of candidates:

- The essential property or "ideal" is that a candidate whose grades dominate another's should lead the other (no domination paradox).
- An infinity of different methods that obey the essential properties (Axioms 1* and 2 through 8) meet that ideal.
- Among them MJ is the one that best combats strategic manipulation.

[20]Given the majority-gauge (p, α, q) of a candidate various rules for determining the order come naturally to mind when two candidates' majority-grades α are the same. They have the allure of tie-breaking rules: e.g., rank them according to $p - q$, or the highest p, or the lowest q. However, as is proven, only the majority-gauge rule coincides with the majority rule and combats strategic manipulation on polarized pairs.

- In particular, MJ is the one method that agrees with the traditional ideal when the electorate is polarized (i.e., when the domination paradox is excluded and the traditional ideal is acceptable).

The last concern MJ practice:

- The scale of grades should be tailored to the particular application: sufficient to allow distinctions between competitors whenever judges or voters see a distinction; yet limited to ease a common understanding. In voting, five to seven have proven to be good choices.
- Theoretical and practical evidence combine to show majority judgment is a serious contender for juries and electorates to use in designating winners and rankings among competing alternatives and candidates.

References

[1] Arrow, K. J. 1951. *Social Choice and Individual Values*. Yale University Press, New Haven, CT. 2nd edition 1963, Wiley, New York, NY.

[2] Arrow, K. J. 1958. "Utilities, attitudes, choices: a review note." *Econometrica* 26: 1-23.

[3] Australian Electoral Commission. Accessed Dec. 11, 2015. http://www.aec.gov.au/Elections/index.htm

[4] AustralianPolitics.com. Accessed Dec. 11, 2015. http://australianpolitics.com/voting/electoral-system/proportional-representation-voting-in-australia

[5] Balinski, M., and R. Laraki. 2011. *Majority Judgment: Measuring, Ranking, and Electing*. MIT Press, Cambridge, MA.

[6] Balinski, M., and R. Laraki. 2011. "Election by majority judgment: experimental evidence." In B. Dolez, B. Grofman, and A. Laurent (eds.), *In Situ and Laboratory Experiments on Electoral Law Reform: French Presidential Elections*: 13-54. Springer, Heidelberg.

[7] Balinski, M., and R. Laraki. 2013. "How best to rank wines: majority judgment." In E. Giraud-Héraoud and M.-C. Pichery (eds.), *Wine Economics: Quantitative Studies and Empirical Applications*: 149-172. Palgrave Macmillan, Basingstoke, UK.

[8] Balinski, M., and R. Laraki. 2014. "What *should* 'majority decision' mean ?" In [33]: 103-131.

[9] Balinski, M., and R. Laraki. 2014. "Judge: *Don't vote* !" *Operations Research* 62: 483-511.

[10] Balinski, M., and R. Laraki. "Majority judgment vs. approval voting." Working paper, Laboratoire d'économétrie, École Polytechnique, Palaiseau, France.

[11] Barber, M., and N. McCarty. 2013. "Causes and consequences of polarization." In Jane Mansbridge and Cathie Jo Martin (eds.), *Negotiating Agreements in Politics*, Report of the task force, American Political Science Association, Washington, DC.

[12] Barberà, S. 2010. "Strategy-proof social choice." In K. J. Arrow, A. K. Sen, and K. Suzumura (eds.), *Handbook of Social Choice and Welfare. Vol. 2*: chapter 25, 731-831. North-Holland, Netherlands.

[13] Barberà, S. and B. Moreno. 2011. "Top monotonicity: A common root for single peakedness, single crossing and the median voter result." *Games and Economic Behavior* 73: 345-359.

[14] Baujard, A., F. Gavrel, H. Igersheim, J.-F. Laslier and I. Lebon. 2013. "Vote par approbation, vote par note. Une exprimentation lors de l'élection présidentielle du 22 avri 2012." *Revue Économique* 64: 345-356.

[15] Blackorby, C., W. Bossert, and D. Donaldson. 1980. *Population Issues in Social Choice Theory, Welfare Economics, and Ethics.* Cambridge University Press, Cambridge, UK.

[16] Blackorby, C., D. Donaldson, and J.A. Weymark. 1984. "Social choice with interpersonal utility comparisons: a diagrammatic introduction." *International Economic Review* 25: 327-356.

[17] Bogomolny, A. 2011. Review of [5]. Mathematical Association of America. Accessed February 1, 2016. http://www.maa.org/press/maa-reviews/majority-judgement-measuring-ranking-and-electing

[18] Brams, S. J. 2010. Preface of J.-F. Laslier and M. Remzi Sanver (eds), *Handbook on Approval Voting*: vii-ix. Springer-Verlag, Berlin and Heidelberg.

[19] Brams, S. J. 2011. "Grading candidates." (Book review of [5].) *American Scientist* 99: 426-427.

[20] Brams, S. J., and P. C. Fishburn. 1983. *Approval Voting.* Birkhauser, Boston, MA.

[21] Brams, S. J., and P. C. Fishburn. 2001. "A nail-biting election." *Social Choice and Welfare* 18: 409-414.

[22] Brennan Center. 2010. "Ungovernable America? The causes and consequences of polarized democracy." Jorde Symposium. Accessed February 14, 2015, https://www.brennancenter.org/blog/ungovernable-america-jorde-symposium

[23] Casella, A. 2011. *Storable Votes: Protecting the Minority Voice.* Oxford University Press, Oxford, UK.

[24] Chernoff, H. 1954. "Rational selection of decision functions." *Econometrica* 22: 422-443.

[25] Condorcet, le Marquis de. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* L'Imprimerie Royale, Paris.

[26] Dahl, R. A. 1956. *A Preface to Democratic Theory.* The University of Chicago Press, Chicago, IL.

[27] Dasgupta, P., and E. Maskin. 2008. "On the robustness of majority rule." *Journal of the European Economics Association* 6: 949-973.

[28] Dasgupta, P., and E. Maskin. 2004. "The fairest vote of all." *Scientific American* 290 (March): 92-97.

[29] d'Aspremont, C., and L. Gevers. 1977. "Equity and the informational basis of collective choice." *Review of Economic Studies* 44: 199-209.

[30] Dhillon, A., and J.-F. Mertens. 1999. "Relative utilitarianism." *Econometrica* 67: 471-498.

[31] Dodgson, C. L. 1876. "A method of taking votes on more than two issues." In, D. Black, *The Theory of Committees and Elections.* Cambridge University Press, Cambridge UK, 224-234.

[32] Edelman, P. H. 2012. "Michel Balinski and Rida Laraki: Majority judgment: measuring, ranking, and electing." *Public Choice* 151: 807-810.

[33] Elster, J. and S. Novak (eds.). 2014. *Majority Decisions.* Cambridge University Press, Cambridge, UK.

[34] Etter, E., J. Herzen, M. Grossglauser, and P. Thiran. 2014. "Mining democracy." In *Proceedings of the Second ACM Conference on Online Social Networks.* ACM, New York, NY, 1-12. Accessed Sept. 15, 2016, http://cosn.acm.org/2014/files/cosn060f-etterA.pdf

[35] Felsenthal, D. S., and M. Machover. 2008. "The majority judgment voting procedure: a critical evaluation." *Homo œconomicus* 25: 319-334.

[36] Fishburn, P. C. 1973. *The Theory of Social Choice.* Princeton University Press, Princeton, NJ.

[37] Fleurbaey, M. 2014. Book review of [5]. *Social Choice and Welfare* 42: 751-755.

[38] Fleurbaey, M., and F. Maniquet. 2011. *A Theory of Fairness and Social Welfare.* Cambridge University Press, Cambridge.

[39] Galton, F. 1907."One vote, one value." *Nature* 75: 414 (Feb. 28).

[40] Galton, F. 1907. "Vox populi." *Nature* 75: 450-451 (Mar. 7).

[41] Gibbard, A. 1973. "Manipulation of voting schemes: a general result." *Econometrica* 41: 587-601.

[42] Gonzalez-Suitt, J., A. Guyon, T. Hennion, R. Laraki, X. Starkloff, S. Thibault, and B. Favreau. 2014. "Vers un système de vote plus juste?" Département d'Économie, École Polytechnique, Cahier No. 2014-20.

[43] Hammond, P. J. 1976. "Equity, Arrow's conditions, and Rawls' difference principle." *Econometrica* 44: 793-804.

[44] Harris – A Nielson Company. October 2014. "Ratings rebound some for President Obama's job performance." Accessed January 16, 2015, http://www.harrisinteractive.com/NewsRoom/HarrisPolls.aspx

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

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

[47] LaPrimaire.org. 2016. Results. Accessed February 2, 2017. https://articles.laprimaire.org/rsultats-du-1er-tour-de-laprimaire-org-c8fe612b64cb https://articles.laprimaire.org/rsultats-du-2nd-tour-de-laprimaire-org-2d61b2ad1394

[48] LaPrimaire.org. 2016. Tutorial. Accessed February 2,217. https://laprimaire.org/citoyen/vote/tutorial

[49] Lippmann, W. 1925. "The task of the public." In *The Phantom Public*, MacMillan Co., New York. Also in *The Essential Lippmann – A Political Philosphy for Liberal Democracy*, C. Rossiter and J. Lare (eds.), Random House, New York, 1963.

[50] Madison, J. 1787. In A. Hamilton, J. Madison, and J. Jay, *The Federalist Papers*, No. 10.

[51] May, K. O. 1952. "A set of necessary and sufficient conditions for simple majority decision." *Econometrica* 20: 680-684.

[52] Miller, G. A. 1956. "The magical number seven, plus or minus 2: Some limits on our capacity for processing information." *Psychological Review* 63: 81-7.

[53] Morreau, M., and J. A. Weymark. 2016. "Measurement scales and welfarist social choice." *Journal of Mathematical Psychology* 75: 127-136.

[54] Moulin, H. 1980. "On strategy-proofness and single peakedness." *Public Choice* 35: 437-455.

[55] Moulin, H. 1988. "Condorcet's principle implies the no show paradox." *Journal of Economic Theory* 45: 53-64.

[56] Moulin, H. 1988. *Axioms of Cooperative Decision Making*. Cambridge University Press, Cambridge.

[57] Moulin, H. 2015. Personal communications.

[58] Nagel, J. H. 2012. "Strategic evaluation in majority judgment." Paper presented July 18, Celebration of Michel Balinski's 78 years, University of Stony Brook. Accessed March 10, 2016 at: http://www.gtcenter.org/Archive/2012/Conf/Nagel1552.pdf

[59] Nash, J. F. 1950. "The bargaining problem." *Econometrica* 18: 43-58.

[60] Pew Research Center. 2016."March 2016 political survey." Accessed May 5, 2016, http://www.people-press.org/files/2016/03/03-31-2016-Political-topline-for-release.pdf

[61] Pew Research Center. 2016. "As election nears, voters divided over democracy and 'respect."' Accessed January 10, 2017, http://assets.pewresearch.org/wp-content/uploads/sites/5/2016/11/03170033/10-27-16-October-political-release.pdf

[62] Puppe, C., and A. Slinko. 2015. "Condorcet domains, median graphs and the single crossing property." Working paper, July.

[63] *Rue89*. 2011. "Testez une autre façon de voter à la primaire." Accessed March 25, 2014. http://www.rue89.com/making-of/2011/10/05/testez-autre-facon-de-voter-a-la-primaire-225043

[64] Saari, D. G. 2001. "Analyzing a nail-biting election." *Social Choice and Welfare* 18: 415-430.

[65] Sadurski, W. 2008. "Legitimacy, political equality, and majority rule." *Ratio Juris* 21: 39-65.

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

[67] Sen, A. 1970. *Collective Choice and Social Welfare.* Holden-Day, Inc. San Francisco CA.

[68] Sen, A. 1977. "On weights and measures: informational constraints in social welfare analysis." *Econometrica* 45: 1539-1572.

[69] Sen, A. 1988. “Forward.” In [56], xi-xiii.

[70] Slate.fr. 2011. Accessed March 25, 2014. “Jugement majoritaire: Arnaud Montebourg au second tour.” http://www.slate.fr/story/43791/primaire-ps-jugement-majoritaire

[71] Terra Nova. 2011. “Rendre les lections aux lecteurs : le jugement majoritaire.” Accessed February 17, 2015, http://www.tnova.fr/note/rendre-les-lections-aux-lecteurs-le-jugement-majoritaire

[72] Tocqueville, A. de. 1831. *De la démocratie en Amérique*, Vol. 1. Edition M.-Th. Gérin, Librairie Médicis, Paris, 1951.

[73] von Neumann, J., and O. Morgenstern. 1944. *Theory of Games and Economic Behavior*. Third edition 1964, Science Editions, New York.

[74] Weber, R. J. 1977. “Comparison of public choice systems.” Cowles Foundation discussion paper 498. Yale University, New Haven, CT.

[75] Wikipedia. 2014. “Academic grading in Denmark.” Accessed May 7, 2014.