A Theory of Measuring, Electing and Ranking

Michel Balinski, Rida Laraki

ECOLE POLYTECHNIQUE
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

A Theory of Measuring, Electing and Ranking

Michel Balinski
Rida Laraki

28 novembre 2006

Cahier n° 2006-11

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

A Theory of Measuring, Electing and Ranking[1]

Michel Balinski[2]
Rida Laraki[3]

28 novembre 2006

Cahier n° 2006-11

Résumé: Le modèle traditionnel du choix social admet une multitude d'impossibilités et n'a abouti à aucune méthode incontestable pour élire un candidat ou ranger des compétiteurs. Un nouveau modèle permet de contourner les impossibilités avec une méthode éminemment pratique, "la valeur-majoritaire".

Abstract: The traditional model of the theory of social choice admits a host of impossibility theorems and has led to no satisfactory methods for electing candidates or ranking competitors. A new model allows the impossibilities to be avoided with an eminently practical method: "the majority-value."

Mots clés : Choix social, théorème d'impossibilité d'Arrow, non-manipulabilité, fonctions de notation-sociale, fonctions de rangement-social, note-majoritaire, rangement majoritaire, valeur-majoritaire.

Key Words : Social choice, Arrow's impossibility theorem, strategy-proofness, social-grading functions, social-ranking functions, majority-grade, majority-ranking, majority-value.

JEL classification: D7, D6, C72.
AMS classification: 91B12, 91B14, 91F10, 91C05.

---

[1] This theory was first presented at the Social Choice and Welfare Society meetings, Istanbul, on July 14, 2006, and subsequently at: U. of Virginia, Johns Hopkins U., NYU and Columbia U., CIRANO (U. de Montréal), Séminaire Parisien de Théorie des Jeux, GREQAM (Aix-en-Provence), and Tilburg University. We are indebted to the organizers who invited us and the participants for their remarks and questions, in particular: S. Anderson, H. Attouch, J. Blouin, S. Brams, J. Harrington, D. Jestaz, H. Moulin, J. Rieskamp, R. Sanver, A. Soubeyran, H. de Swart, M.-A. Valfort, J. Weibull, J. Weymark, and W. Zwicker. We are especially indebted to Claude Henry, Edi Karni, Friedrich Pukelsheim, Vincent Renard, Jérôme Renault, Bernard Sinclair-Desgagné, Sylvain Sorin, Tristan Tomala and Peyton Young for their criticisms and suggestions on the first version of our forthcoming book that describes the theory in detail.
[2] michel.balinski@shs.polytechnique.fr, C.N.R.S., Laboratoire d'économétrie, Ecole polytechnique, Paris
[3] rida.laraki@shs.polytechnique.fr, C.N.R.S., Laboratoire d'économétrie, Ecole polytechnique, Paris

A Theory of Measuring, Electing and Ranking

Michel Balinski and Rida Laraki

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

Introduction

The theory of social choice concerns methods for amalgamating the appreciations or evaluations of many individuals into one collective appreciation or evaluation. It has two principal applications. (1) Voting: electors in a democracy choose one among several candidates, or committee members decide on one among several courses of action. (2) Jury decisions: judges evaluate competitors—figure skaters, gymnasts, pianists, wines, ...—and rank them or classify them by level of excellence.[1]

The fundamental problem is to find a *social decision function (SDF)* whose *inputs* are *messages* of judges or voters and whose *outputs* are the jury or electoral *decisions*, usually rank-orderings of competitors and winners. Much of the theory of social choice has blurred the distinction between a judge's complex aims, ends, purposes and wishes—in short, his or her *preferences* or *utilities*—and the messages he or she is allowed to send.[2]

In the traditional model, consecrated by some seven centuries of use,[3] each individual judge's or voter's rank-ordering of the competitors is at once his or her message *and* his or her preferences. Does it mean that the judge prefers this

[1]Our thesis is capsuled in this article. A complete account, including proofs, many other results and references, is given in the forthcoming book, Michel Balinski and Rida Laraki, *One-Value, One-Vote: Measuring, Electing and Ranking.*

[2]The word "preferences" contains a misleading connotation: voters do not merely express what they prefer, they may well express what they believe is right (*e.g.*, R. Goodin and K. Roberts, "The ethical voter," *American Political Science Review* 69 (1975) 926-928); a judge in a court of justice is supposed to evaluate conformity with the law, not merely express his preferences. The real, deep preferences of a judge or voter cannot come close to being expressed in a rank-ordering of competitors.

[3]Ramon Lull proposed a refinement of Condorcet's method in 1297: it is known today as Copeland's method. Nicolaus Cusanus put forth what is today known as Borda's method in 1433. See, I. McLean, "The Borda and Condorcet principles: three medieval applications," *Social Choice and Welfare* 7 (1990) 99-108; G. Hägele and F. Pukelsheim, "Llull's writings on electoral systems," *Studia Lulliana* 41 (2001) 3-38; Gunter Hägele and Friedrich Pukelsheim, "The electoral systems of Nicolaus Cusanus," to appear in Hg. G. Christianson, T.M. Izbicki, and C.M. Bellitto, eds., *The Church, the Councils and Reform: Lessons from the Fifteenth Century*, Catholic University of America Press, Washington, DC.

rank-ordering above all others; or, that the judge wishes the first competitor on his list to be the winner, the second to be the winner if the first is not, the third to be the winner if the first two are not, and so on down the list; or, is the rank-ordering required and chosen strategically by the judge given his or her "true" rank-ordering.

In the real world, a judge's message is simply a message, nothing more. It depends on the judge's preferences, but it is not and cannot be his or her preferences. And in the real world, a judge's or a voter's preferences or utilities depends on a host of factors that include the decision (or output), the messages of the other judges (a judge or voter may wish to differ from the others, or on the contrary resemble the others), the social decision function that is used (a judge may prefer a decision given by "democratic" function to one rendered by an "oligarchique" function, or the contrary) and the message he or she thinks is the right one (a judge may prefer honest behavior, or not).

Kenneth Arrow, in the first deep theoretical analysis of the theory of social choice, uses the traditional model: each judge's input message is a rank-ordering, routinely interpreted to be a complete expression of his "preferences" (strategic considerations are absent); the output is a rank-ordering and a winner (the first-ranked competitor of the order). His celebrated "impossibility" theorem shows that there exists no *social welfare function SWF* satisfying three reasonable properties for obtaining a decision given any inputs (unless there are only two competitors).[4] Donald Saari's algebraic analysis explains why and how the traditional model cannot but lead to paradoxical outcomes.[5] Amartya Sen models each judge's inputs as a numerical "utility" over the competitors, *i.e.*, the judge assigns a real number to every competitor; the output is a rank-ordering whose utility to a judge is not specified. The model has theoretical interest but no practical significance because a voter's individual utility is a much more complex concept. In any case, Arrow's theorem emerges again unless the utilities are assumed to be comparable (that bugabear of economists!).[6] The model used to prove the well-known Gibbard-Satterthwaite impossibility theorem assumes the output is a winner (indeed, how could preferences be modelled if the output were a rank-ordering?); each judge has "true" preferences expressed as a rank-ordering; but a judge's input is a strategically chosen rank-ordering. The theorem states that there exists no SWF that makes it a dominant strategy for every judge to report his true preferences.[7]

Refined, extended, and reformulated in many variants, the traditional approach has continued to produce a host of related impossibility theorems. We add to this list a negative theorem of a new kind: a fundamental incompatibility between winners and rank-orderings as outputs of the traditional model. It

[4]Kenneth J. Arrow, *Social Choice and Individual Values*, John Wiley & Sons, Inc., 1951.

[5]Donald Saari, "Mathematical structure of voting paradoxes. I. Pairwise votes, and II. Positional voting," *Economic Theory* 15 (2000) 1-53 and 55-102.

[6]Amartya K. Sen, *Collective Choice and Social Welfare*, Holden-Day, 1970.

[7]Allan Gibbard, "Manipulation of voting schemes: a general result,"*Econometrica* 41 (1973) 587-601; and Mark A. Satterthwaite, "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions," *Journal of Economic Theory* 10 (1973) 187-217.

devolves from a simple observation: if the output is to be a rank-ordering and inputs are interpreted as preferences, should not an individual's input message be his preferences over rank-orderings rather than a single rank-ordering?

Given all of these negative results, it is not surprising that the debate over what method of voting *should* be used in practice goes on unabated. By and large, it may be said to pit the supporters of Lull (alias Condorcet) against those of Cusanus (alias Borda), though some argue for a new method, "approval voting,"[8] and diverse hybrids are regularly proposed.

We contend: (1) Arrow's and all the other impossibility and incompatibility results show that the fundamental problem has no acceptable solution in the context of the traditional model. (2) The traditional approach does not adequately model the messages or the purposes of the judges and voters. (3) A new model is necessary.

Practice, curiously enough, suggests a different formulation of the inputs. Olympic competitions in figure skating and gymnastics, wine competitions, competitions among pianists, flautists or orchestras, ..., all use measures or grades. As Lord Kelvin proclaimed, "If you cannot measure, your knowledge is meager and unsatisfactory." Indeed, Arrow[9] himself states "there are essentially two methods by which social choices can be made, voting,... and the market mechanism": the second uses a measure, *price* expressed in terms of money.

A *measure* or *grade* is a message that has strictly nothing to do with a utility. A judge may dislike a wine yet give it a high grade because of its merits; he or she may also like a wine and yet with great satisfaction give it a low grade because of its demerits. A measure provides a *common language*—be it numerical, ordinal or verbal—to grade and to classify. In this perspective, Arrow's theorem means that without a common language there can be no consistent collective decision.

When the messages are grades expressed in a common language then one method of classifying competitors, candidates or alternatives—the *majority-grade*—and one method for ranking them—the *majority-ranking*—emerge as the only ones that satisfy each of various desirable properties. They are compatible. Moreover, they best resist strategic manipulations of judges and voters under varying assumptions concerning the judges' and voters' preferences or utilities.

1 The traditional model

There is a finite set of *competitors* (alternatives, candidates, performances, or competing goods) $\mathcal{C} = \{A, \dots, I, \dots, Z\}$, and a finite set of n *judges* (or voters) $\mathcal{J} = \{1, \dots, j, \dots, n\}$. Each judge's input *message* is a rank-ordering of the competitors. Together all of the input messages constitute a *preference profile* (in keeping with traditional terminology we use the word "preference(s)" as a synonym for rank-orderings in this section). A *social welfare function (SWF)* renders an output—a rank-ordering—for any inputs or preference profile.

[8] See S. J. Brams and P. C. Fishburn, *Approval Voting*, Birkhäuser, Boston, 1983.
[9] Arrow, *op. cit.*, p. 1.

Take $A \succ B$ to mean that a judge ranks A ahead of B, and $A \succ_S B$ to mean that the SWF (or "Society") ranks A ahead of B; in examples an integer followed by a rank-ordering is the number of judges sending that input message.

Condorcet is the first to have realized the essential difficulty of the problem. Consider one of his examples with the 60-judge preference profile:

$$23 : A \succ B \succ C \qquad 2 : B \succ A \succ C \qquad 17 : B \succ C \succ A$$

$$10 : C \succ A \succ B \qquad 8 : C \succ B \succ A$$

If majority rule decides the order between each pair of competitors separately, the result is as follows (with numbers in parentheses giving the respective votes):

$$A(33) \succ_S B(27) \qquad B(42) \succ_S C(18) \qquad C(35) \succ_S A(25)$$

Thus majority decision says $A \succ_S B \succ_S C \succ_S A$: this is the *Condorcet-paradox*, no competitor is favored to all others.

Arrow showed that this is an inescapable conundrum. He imposed three conditions that any SWF should satisfy. *Unanimity*: If every judge prefers competitor A to competitor B, *i.e.*, $A \succ B$, then the SWF ranks A ahead of B, *i.e.*, $A \succ_S B$. *Non-dictatorship*: The input of no one judge can determine the output of the SWF whatever the inputs of all the other judges. *Independence of irrelevant alternatives (IIA)*: whether the SWF yields $A \succ_S B$ or the contrary $A \prec_S B$, depends only on the judges' preferences between A and B. His theorem shows that when there are at least three competitors, there is no SWF that satisfies the three conditions for all possible inputs of the judges.

Nevertheless, people vote and judges rank: how? A judge accords k *Borda-points* to a competitor if k opponents are ranked below him.[10] A competitor's *Borda-score* is the sum of his Borda-points over all judges; equivalently, it is the sum of the votes he receives in all pair by pair votes.[11] The *Borda-ranking* ranks the competitors by their Borda-scores, from highest to lowest; the highest designates the *Borda-winner*. The Borda-ranking for Condorcet's 60-judge example is (each competitor's Borda-score is in parentheses): $B(69) \succ_S A(58) \succ_S C(53)$.

Condorcet attacked Borda's method. His argument was that when there exists a *Condorcet-winner*—a competitor who has a majority against every other competitor—then he must be the winner, a property that Borda's method violates, as the following 81-judge example of Condorcet shows:

$$30 : A \succ B \succ C \qquad 1 : A \succ C \succ B \qquad 29 : B \succ A \succ C$$

[10] Laplace justified the Borda-points by imagining that each judge wishes to assign a positive real score in some interval $[0, R]$ to each competitor but is asked instead to rank them. Laplace computed the average of the lowest points, of the next to lowest, on up to the highest, and found them to be proportional to the Borda-points. See, Pierre-Simon, Marquis de Laplace, *Théorie analytique des probabilités*, 3rd ed., in *Œuvres Complètes de Laplace*, t. 7: pp. v and clii-cliii.

[11] Le Chevalier Jean-Charles de Borda, "Mémoire sur les lections au scrutin," *Histoire de l'Académie royale des sciences* (1784) 657-665. A footnote states that the ideas were presented before the Academy on June 16, 1770. Cusanus proposed exactly the same method in 1433.

$$10 : B \succ C \succ A \qquad 10 : C \succ A \succ B \qquad 1 : C \succ B \succ A.$$

Here the Borda-ranking is $B(109) \succ_S A(101) \succ_S C(33)$, yet A is the Condorcet-winner. This argument was widely accepted before the work of Donald Saari.

Consider Condorcet's 81-judge example. 30 of the 81 judges have preferences that constitute a *Condorcet-component*, a perfect symmetry among the rank-orderings of equal numbers of judges:

$$10 : A \succ B \succ C \qquad 10 : B \succ C \succ A \qquad 10 : C \succ A \succ B.$$

The 30 judges together say that the competitors are tied. There is another Condorcet-component as well:

$$1 : A \succ C \succ B \qquad 1 : C \succ B \succ A \qquad 1 : B \succ A \succ C.$$

The 33 judges simply cancel each other out, their messages may be dropped, and the decision should depend on the remaining judges:

$$20 : A \succ B \succ C \qquad 28 : B \succ A \succ C.$$

B is the clear winner. Moral: the Condorcet-winner—when he exists—is certainly *not* the candidate who should win in every case! Borda's method (and any "positional" method) avoids this difficulty because the total number of points awarded to every competitor in a Condorcet-component is the same.

Saari's theory generalizes this fundamental cancellation idea. He asserts that "all election difficulties" come from Condorcet-components and, to a lesser extent, from more intricate symmetries in the preference profiles; and concludes, "the [Borda-count] applied to all n-candidates is the unique ranking which avoids all of the indicated problems."[12] But is Borda's method good for ranking, or for designating winners, or both?

Condorcet proposed a method explicitly for ranking.[13] A voter contributes k *Condorcet-points* to an arbitrary rank-ordering $A \succ_S B \succ_S C \succ_S \ldots \succ_S Z$ if his input agrees in k pair-by-pair comparisons. The *Condorcet-count* of the rank-ordering is the sum of the Condorcet-points over all voters. The *Condorcet-ranking* is the ranking that maximizes the Condorcet-count. It ranks the Condorcet-winner first, and the *Condorcet-loser*—a competitor who loses against every other competitor—last, when either exist.[14] Is it good for ranking, or for designating winners and losers, or both?

The two outputs—a rank-ordering and a winner—have usually been treated as two sides of one coin: given a rank-ordering, the winner is the first-placed

[12] Saari, *op. cit.*, p. 57. However, he claims that to counter manipulation, "Instant-Borda-Runoff" should be used to elect a winner: namely, obtain the Borda-ranking, drop the bottom candidate, and repeat until one candidate remains. This method *always* elects the Condorcet-winner when he exists. See, Donald Saari, *Chaotic Elections!: A Mathematician Looks at Voting*, American Mathematical Society, 2001, p. 103.

[13] Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris, 1785, l'Imprimerie royale.

[14] The Condorcet-ranking of his 60-judge example is $B \succ_S C \succ_S A$; that of his 81-judge example, $A \succ_S B \succ_S C$. Neither agrees with the Borda-ranking.

competitor; given a mechanism for determining a winner, place him first then apply the mechanism to the remaining competitors and place that winner second, and continue.[15] These are questionable practices in the context of the traditional model.

To see why, consider the preference profile

$$333: A \succ B \succ C \qquad 333: B \succ C \succ A \qquad 333: C \succ A \succ B.$$

The 999 judges constitute a Condorcet-component, so cancel each other out. Borda and Condorcet agree on the winners: A, B and C are tied. Condorcet (reasonably) says the three stated rank-orderings are tied for first; Borda (ridiculously) that all six possible rank-orderings are tied for first. Now consider the situation with one additional judge

$$333: A \succ B \succ C \quad 333: B \succ C \succ A \quad 333: C \succ A \succ B \quad 1: A \succ C \succ B.$$

Borda (reasonably) declares A the winner and B the loser, but (ridiculously) $A \succ_S C \succ_S B$, since only 1 judge agrees with it, 666 partially agree, 333 totally disagree. Condorcet (reasonably) declares $A \succ_S B \succ_S C$ and $C \succ_S A \succ_S B$ are tied: 333 agree, 667 partially agree. But A and C should certainly not be tied as winners. Borda's method is appropriate for designating winners and losers, Condorcet's for designating rank-orderings, a fact already appreciated by Peyton Young.[16] In fact, the situation is much worse: the two outputs cannot be reconciled.

A SWF is *unanimous* if whenever the inputs of all judges are the same rank-ordering, its output is that one. It is *choice-consistent* if the first-placed and last-placed competitors of the output remain the same when a Condorcet-component of the profile is eliminated. It is *rank-consistent*[17] if removing a first-placed or last-placed competitor of the output does not change the order among the remaining competitors. Borda's method is unanimous and choice-consistent but not rank-consistent; Condorcet's is unanimous and rank-consistent but not choice-consistent.

Theorem 1 (Incompatibility) *There exists no unanimous, choice- and rank-consistent method.*

2 Grading: the basic model

A thorough investigation of practice shows that scores, measures or grades have been invented to classify and to rank in an incredibly wide variety of circumstances. Practical people needing practical solutions have increasingly devised

[15] Alternatively, given a mechanism for determining a loser, place him last then apply the mechanism to the remaining competitors and place that loser next to last, and continue.

[16] H. P. Young, "Optimal ranking and choice from pairwise comparisons," in B. Grofman and G. Owen, eds., *Information Pooling and Group Decision Making*, JAI Press, Greenwich CT, 1986. We give new characterizations of Borda's and Condorcet's methods that make this clear, see Balinski and Laraki, *op. cit.*

[17] This property is equivalent to "local stability" in H. P. Young, "Condorcet's theory of voting,"" *American Political Science Review* 82 (1988) 1231-1244.

mechanisms to transform judges' grades (instead of rank-orderings) into a jury's grades to determine final rank-orderings. A set of grades—numbers from 0 to 20, to 25 or to 100; medal nominations from none to bronze, silver or gold; letters from F to A; or words or phrases from bad to excellent—becomes, in effect, a *common language* used to assess performances, just as grades determine the standing of students in schools and universities.[18]

Formally, a *common language* Λ is a set of *grades* $\alpha, \beta, \ldots$ that are strictly ordered. It may be finite or an interval of the real numbers. $\alpha \succeq \beta$ means that either α is a higher grade than β, in symbols, $\alpha \succ \beta$, or $\alpha = \beta$.

There is a finite set of m *competitors* $\mathcal{C} = \{A, \ldots, I, \ldots, Z\}$, and a finite set of n *judges* (or voters) $\mathcal{J} = \{1, \ldots, j, \ldots, n\}$. A problem is completely specified by an *input* or *profile* $\Phi = \Phi(\mathcal{C}, \mathcal{J})$: an m by n matrix of the grades $\Phi(I, j) \in \Lambda$ assigned by each of the judges $j \in \mathcal{J}$ to each of the competitors $I \in \mathcal{C}$.

A *method of grading* is a function F that assigns to any input or profile Φ one *output or final grade* in the same language for every competitor:

$$F : \Lambda^{m \times n} \to \Lambda^m.$$

Designed to assign grades, it must satisfy certain basic properties.

Axiom 1 *F is* neutral*: $F(\rho\Phi) = \rho F(\Phi)$, for any permutation ρ of the competitors (or rows).*

Axiom 2 *F is* anonymous*: $F(\Phi\tau) = F(\Phi)$, for any permutation τ of the judges (or columns).*

Axiom 3 *F is* unanimous*: If a competitor is given an identical grade α by every judge, then F assigns him the grade α.*

Axiom 4 *F is* monotonic*: If $\Phi = \Phi'$ except that one or more judges give higher grades to competitor I in Φ than in Φ', then $F(\Phi)(I) \succeq F(\Phi')(I)$. Moreover, if all the judges give higher grades to competitor I in Φ than in Φ', then $F(\Phi)(I) \succ F(\Phi')(I)$.*

Axiom 5 *F is* independent of irrelevant alternatives (IIA)*: if the grades assigned by the judges to a competitor $I \in \mathcal{C}$ in two profiles Φ and Φ' are the same then $F(\Phi)(I) = F(\Phi')(I)$.*

A function $f : \Lambda^m \to \Lambda$ that transforms a set of judge's grades into a single grade will be called an *aggregation function* if it satisfies the following three properties:

- *anonymity*: $f(\ldots, \alpha, \ldots, \beta, \ldots) = f(\ldots, \beta, \ldots, \alpha, \ldots)$;
- *unanimity*: $f(\alpha, \alpha, \ldots, \alpha) = \alpha$; and

[18] The history of the use of grades to evaluate students and employees; competitive pianists and flutists; Olympic skaters, divers and gymnasts; cities; and wines proves these claims. See, Balinski and Laraki, *op. cit.*

- *monotonicity*:

$$\alpha_j \preceq \beta_j \Rightarrow f(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n) \preceq f(\alpha_1, \ldots, \beta_j, \ldots, \alpha_n)$$

  and

$$\alpha_1 \prec \beta_1, \ldots, \alpha_n \prec \beta_n \Rightarrow f(\alpha_1, \ldots, \alpha_n) \prec f(\beta_1, \ldots, \beta_n).$$

Theorem 2 (Possibility) *A method of grading F satisfies the five axioms if and only if $F(\Phi)(I) = f\big(\Phi(I)\big)$ for every $I \in \mathcal{C}$, for some one aggregation function f.*

The average or mean value function is the universally used aggregation function in practice, though sometimes highest and lowest grades are dropped. This means that almost always the output language is richer than the language of the input grades (inputs are usually restricted to discrete levels).

In conformity with most practical applications, the common language is parametrized as a subset of real numbers and whatever aggregation is used, small changes in the parametrization or the input grades should imply small changes in the output or the final grades. Hence, even if the initial language is finite, all possible parametrizations must be considered. It is thus natural to take the common language to be $[0, R]$ for some positive real R (as did Laplace), and impose

Axiom 6 *F (or its aggregation function f) is* continuous.

> A *social grading function* (SGF) F is a method of grading that satisfies the six axioms of the basic model.

Thus F defines—and is defined by—a unique continuous aggregation function f. *In the sequel—since a SGF and its aggregation function go hand in hand—properties are defined in terms of aggregation functions, theorems stated in terms of SGFs. Also, $\mathbf{r} = (r_1, \ldots, r_n)$ represents a competitor's grades, superscripts designate competitors.*

Enriching a language by embedding it into a real interval opens the door to vastly more possible methods of grading, but it will turn out that the aggregation functions that emerge as those that *must* be used are directly applicable in the seemingly more restrictive finite languages as well.

3 Order functions

A judge of the jury knows the SGF (equivalently, the aggregation function f) that determines the final grades: what strategies will he use in the "game" of assigning his grades? A judge undoubtedly wishes to give the grade he believes is the "right one;" he may, however, assign it so that the final grade is as close as possible to the "right one;" or, he may try to manipulate the outcome for extraneous reasons (as did a judge of the pairs figure skating in the 2002 Olympic games). This is why in practice, highest (1 or 2) and lowest (1 or 2) grades are often eliminated.

The "utility" of a judge j is a complicated function $u_j(\mathbf{r}^*, \mathbf{r}, f, \Lambda)$, where $\mathbf{r}^* = (r_1^*, \dots, r_n^*)$ are the grades the judges believe are the right ones and $\mathbf{r} = (r_1, \dots, r_n)$ the grades they give. The utility of judge j might include a term $-|r_j^* - r_j|$ if he wished to grade honestly; it might contain a term $-\sum_{i \neq j} |r_i^* - r_i|$ if he wished that the other judges graded honestly; it might include a term $-|\Lambda - \Lambda_j^*|$ if he wished a language Λ_j^* were used; and it is often assumed to be "single-peaked," $u_j(\mathbf{r}^*, \mathbf{r}, f, \Lambda) = -|r_j^* - f(r_1, \dots, r_n)|$. In fact, judges' utilities, judges' beliefs, their beliefs about the others' beliefs, ..., are all completely unknown and change from one competition to another. The theory we develop depends *only* on what in practice can be known.[19] It does not formulate a formal game theoretic model (as does classical "mechanism design"). The mechanisms that are singled out are "strategy-proof" for large classes of reasonable "utilities;" when they are not, they best combat manipulability.

Suppose that r is a competitor's final grade. An aggregation function is *strategy-proof-in-grading* if, when a judge's input grade is $r^+ > r$, any change in his input can only lead to a lower grade; and if, when a judge's input grade is $r^- < r$, any change in his input can only lead to a higher grade.

Strategy-proof-in-grading implies that it is a *dominant strategy* for a judge to honestly assign the grade he believes is the correct one, whenever the more a grade deviates from the correct one the less he likes it ("single-peaked preference," a reasonable assumption for most judges who grade). There is a class of SGFs that is easily seen to be strategy-proof: the *order functions*.

The kth highest grade is called the *kth-order function* f^k.

Theorem 3 *The unique strategy-proof-in-grading SGFs are the order functions.*

They are "group strategy-proof-in-grading" as well.[20] The result holds without the continuity assumption and also when the language is finite.

How can the effects of strategic manipulation be countered when judges' appreciations or utilities are more complex? To manipulate a judge must be able to raise or to lower the final grade by raising or lowering the grade he assigns. In some situations a judge can only change the final grade by increasing his grade, in others only by decreasing his grade. A judge who can both lower and raise the final grade has greater opportunity to manipulate.

Theorem 4 *There exists no SGF that, for every profile of grades, prevents every judge from both increasing and decreasing the final grade. The unique SGFs for which at most one judge may both increase and decrease a final grade are the order functions.*

Given an aggregation function f and input grades $\mathbf{r} = (r_1, \dots, r_n)$, let $\mu^-(f, \mathbf{r})$ be number of judges who can decrease the final grade, $\mu^+(f, \mathbf{r})$ be

[19] As does Vickrey's second price auction mechanism.

[20] Hervé Moulin proves a related technical result but in an entirely different context. See, H. Moulin, "On strategy-proofness and single peakedness," *Public Choice* 35 (1980) 437-455.

the number of judges who can increase the final grade, and $\mu(f,\mathbf{r}) = \mu^-(f,\mathbf{r}) + \mu^+(f,\mathbf{r})$. Define the *manipulability of* f, to be

$$\mu(f) = \max_{\mathbf{r}=(r_1,\ldots,r_n)} \mu(f,\mathbf{r}).$$

At worst, a judge can both increase and decrease the final grade, so $\mu(f) \leq 2n$. In particular, when f is taken to be the arithmetic mean of the grades (as does Borda's method) the manipulability is maximized, $\mu(f) = 2n$. On the other hand, when f is the kth-order function, $\mu(f) = n + 1$.

Theorem 5 *The unique SGFs that minimize manipulability are the order functions.*

Suppose that after the members of a jury have assigned their grades, some judge wishes to revise his grade by assigning a grade closer to the final grade of the jury. An aggregation function f is *reinforcing* when $f(r_1,\ldots,r_k,\ldots,r_n) = r$ and $r_k > \hat{r}_k \geq r$ or $r \geq \hat{r}_k > r_k$ implies $f(r_1,\ldots,\hat{r}_k,\ldots,r_n) = r$.

Theorem 6 *The unique reinforcing SGFs are the order functions.*

If every judge assigns a grade in a subset of the grades, then the final grade should belong to that subset. This may be seen as restricting outputs to the language of inputs, or generalizing unanimity. An aggregation function f *conforms* with the assigned grades if $\{r_1,\ldots,r_n\} \subset S$ implies $f(r_1,\ldots,r_n) \in S$.

Theorem 7 *The unique SGFs that conform with the assigned grades are the order functions.*

The particular language used in grading should make no difference in the ultimate outcomes. An aggregation function should give equivalent grades when one language is faithfully translated into another. This is the "meaningfulness" problem of measurement theory in the context of a jury decision.[21] An aggregation function f is *language-consistent* if $f\big(\phi(r_1),\ldots,\phi(r_n)\big) = \phi(f(r_1,\ldots,r_n)\big)$ for all increasing, continuous functions $\phi : [0,R] \to [\underline{R},\overline{R}],\ \phi(0) = \underline{R},\ \phi(R) = \overline{R}$.

Theorem 8 *The unique language-consistent SGFs are the order functions.*

This result depends crucially on the judges using a common language. When there is no common language, a judge's only meaningful input is the order of his grades. An aggregation function f is *preference-consistent* if $f(r_1,\ldots,r_n) \geq f(s_1,\ldots,s_n)$ implies $f\big(\phi_1(r_1),\ldots,\phi_n(r_n)\big) \geq f\big(\phi_1(s_1),\ldots,\phi_n(s_n)\big)$, for all increasing, continuous functions $\phi_j : [0,R] \to [\underline{R},\overline{R}],\ \phi_j(0) = \underline{R},\ \phi_j(R) = \overline{R}$.

Theorem 9 (Arrow's impossibility) *There exists no preference-consistent SGF.*

[21]See, D. H. Krantz, R. D. Luce, P. Suppes and A. Tversky, *Foundations of Measurement, Vol. I: Additive and Polynomial Representation,* Academic Press, New York, 1971; J. Pfanzagl, *Theory of Measurement,* Physica-Verlag, Wien, 1971; and F. S. Roberts, *Measurement Theory, with Applications to Decision Making, Utility and the Social Sciences*, Addison-Wesley, Reading, MA, 1979.

This shows that in order to arrive at meaningful final grades it is *essential* for judges to share a common language: otherwise, the road is barred by Arrow's fundamental result. But that only stands to reason: imagine the leaders of the world's powers negotiating an agreement with no common language (and no translators)![22]

4 The majority-grade

The evidence supports the use of order functions when juries grade. There are many. Different arguments single out one method. Sir Francis Galton had the key idea[23] just one century ago, namely: "[The] middlemost estimate, the number of votes that it is too high being exactly balanced by the number of votes that it is too low. Every other estimate is condemned by a majority of voters as being either too high or too low ... The number of voters may be odd or even. If odd, there is one middlemost value ... If the number of voters be even, there are two middlemost values, the mean of which must be taken ..."[24] He erred in the even case.

A *middlemost* aggregation function f, for $r_1 \geq \ldots \geq r_n$, is

$$f(r_1, \ldots, r_n) = r_{(n+1)/2} \text{ when } n \text{ is odd, and}$$

$$r_{n/2} \geq f(r_1, \ldots, r_n) \geq r_{(n+2)/2} \text{ when } n \text{ is even.}$$

When n is odd, it is the order function $f^{(n+1)/2}$. When n is even, there are infinitely many; in particular, $f^{n/2}$ and $f^{(n+2)/2}$ are the *upper-middlemost* and *lower-middlemost* order functions. The *middlemost interval* is the point $r_{(n+1)/2}$ when n is odd, and the interval $[r_{(n+2)/2}, r_{n/2}]$ when n is even.

Whatever the parity of n, every grade other than a grade in the middlemost interval is condemned by an absolute majority of the judges as being either too high or too low. Three other properties justify the middlemost aggregation functions; it is then shown why there is a single best choice.

Theorem 10 *The unique aggregation functions that assign a final grade of r when a majority of judges assign r are the middlemost.*

[22] Properties close to language- and preference-consistency are known under different names in the literature on measurement theory; in particular, Theorems 8 and 9 are known in one guise or another. See: A. Orlov, "The connection between mean quantities and admissible transformations," *Mathematical Notes* 30 (1981) 774-778; S.-R. Kim, "On the possible scientific laws," *Mathematical Social Sciences* 20 (1990) 19-36. Work on welfarism initiated by Sen, *op. cit.*, has considered similar invariance properties; see also W. Bossert and J. A. Weymark, "Utility in social choice," in S. Barberà, P. J. Hammond and C. Seidl, eds., *Handbook of Utility Theory, Vol. 2: Extensions*, Kluwer Academic Publishers, Boston, 2004, pp. 1099-1177.

[23] Discarded as "a small contribution" by Duncan Black, T*he Theory of Committees and Elections*, Cambridge University Press, 1958, p. 188.

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

Practical mechanisms of grading often eliminate extremes to counter cheaters, to guard against cranks, and to emphasize the significance of place in order rather than magnitude. A SGF *counters crankiness*[25] if for $r_1 \geq \ldots \geq r_n$, $n \geq 3$, its aggregation function f satisfies $f(r_1, r_2, \ldots, r_{n-1}, r_n) = f(r_2, \ldots, r_{n-1})$, where in going from left to right the highest and lowest grades have been dropped.[26] Iterating, $f(r_1, r_2, \ldots, r_{n-1}, r_n) = f(r_+, r_-)$ where $[r_-, r_+]$ is the middlemost interval (a point $r_- = r_+$ when n is odd).

When a judge dislikes a final grade the further it departs from his ideal grade, it is a dominant strategy for him to assign his ideal grade. But judges may have different incentives. A judge may wish to either increase or decrease the final grade. The kth-order function allows $n-k+1$ judges to increase the final grade and k to decrease it. It is desirable to thwart potential manipulation as much as possible. Letting λ be the probability a judge wishes to increase the grade and $1-\lambda$ that he wishes to decrease it, the *probability of effective-manipulability* of the aggregation function f is

$$EM(f) = \max_{\mathbf{r}=(r_1,\ldots,r_n)} \max_{0 \leq \lambda \leq 1} \frac{\lambda \mu^+(f, \mathbf{r}) + (1-\lambda)\mu^-(f, \mathbf{r})}{n}.$$

Theorem 11 *The unique aggregation functions that minimize the probability of effective-manipulability or that counter crankiness are the middlemost that depend only on the middlemost interval.*

Many physical measures have the property that equal intervals have the same significance: they are "interval measures" in the jargon of measurement theory. The grades invented to assign to competing skaters, pianists or politicians could be interval measures, but more likely are not. As a grade approaches "perfection," each additional point often represents much more than an additional point added to a middling grade; and at the other end of the scale, the same phenomenon exists. It is reasonable to suppose that an interval measure exists in theory. *In fact*, points are routinely added and averaged, so treated as if they were interval measures! It suffices to postulate the existence of much less to imply the existence of an interval measure.

For suppose there exists a distance function d that measures the judge's discontent: when the judge assigns the grade r and the final grade is s, his disutility is $d(r, s) \geq 0$. Thus d satisfies: $d(r, r) = 0, d(r, s) = d(s, r)$, and $r < s < t$ implies $d(r, s) + d(s, t) = d(r, t)$. The last equation says that the improvement in a competitor's performance in going from a grade of r to s plus the improvement in going from s to t equals that of going from r to t; or, that the disutility of a judge who believes the grade should be r when the final grade is t equals his disutility when the final grade is s plus his disutility when he believes it should be s when the final grade is t. This accommodates the possibility that, for example, on a scale of 0 to 100, $d(98, 99) = 5d(75, 76)$. Several points of view suggest that it is reasonable to assume that all the judges

[25]The word honors Galton, who wished to avoid giving "power to 'cranks' in proportion to their crankiness," *ibid.*

[26]The two fs are in fact different, but expressing the idea in this manner simplifies notation.

share this same language: one judge teaches all others; or, the rules impose by fiat that the meaning of the grades meets exactly this test; or, equity among the judges imposes that their disutilities must be modelled identically.

A SGF with aggregation function f *maximizes the social grade* when the final grade $f(r_1, \ldots, r_n) = r$ minimizes the total disutility of all of the judges, $\Delta(r) = \sum_{j \in \mathcal{J}} d(r, r_j)$.

Theorem 12 *The unique aggregation functions that maximize the social grade are the middlemost.*

Thus imposing an equity condition—namely, that judges compare performances with the same measure—together with the assumption that the measure is a distance function, implies that the optimal mechanism must be a majority decision. A distance function is equivalent to the existence of an interval measure (not necessarily compact).[27]

Characterization A SGF *rewards consensus* when all of A's grades belong to the middlemost interval of B's grades implies that A's final grade is at least as high as B's final grade (when the number of judges is odd the condition means A's grades are all equal to B's unique middlemost grade).

> The *majority-grade* f^{maj} is the SGF defined by the order function $f^{(n+1)/2}$ when n is odd, and by the lower-middlemost order function $f^{(n+2)/2}$ when n is even.

Theorem 13 *The unique middlemost aggregation function that rewards consensus is the majority-grade* f^{maj}.

5 The majority-ranking

A competitor bestowed a higher grade than another is naturally ranked higher in the order of the competitors than the other: grades imply orders. The essential incompatibility between the designation of winners (or losers) and rank-orderings inherent to the traditional model (Theorem 1) simply does not arise in the context of grades. On the other hand, although in some applications a complete ordering is not sought—*e.g.*, wine competitions—there are other applications—notably sports and elections—where an ordered list from first to last and a clear winner is absolutely necessary.

When rank-orderings are the principal goal instead of grades, the strategic behavior of the judges may change. A SGF is *strategy-proof-in-ranking* if for any judge j final grades $r^A < r^B$ opposed to the judge's grades $r_j^A > r_j^B$ implies he can neither decrease B's final grade nor increase A's final grade; and it is *partially strategy-proof-in-ranking* if any judge in the same situation can decrease B's final grade implies he cannot increase A's and if he can increase A's final grade implies he cannot decrease B's.

[27] Defining $\phi(s) = d(\frac{R}{2}, s)$ if $s \geq \frac{R}{2}$ and $= -d(\frac{R}{2}, s)$ if $s \leq \frac{R}{2}$ implies $d(r, s) = |\phi(r) - \phi(s)|$.

Theorem 14 *There exists no SGF that is strategy-proof-in-ranking. The unique SGFs that are partially strategy-proof-in-ranking are the order functions.*

Given a profile of grades Φ, let $\rho = (\rho_I)$ be independent permutations of each of the rows I. The anonymity of f implies

$$F(\rho\Phi)(I) = f\big(\Phi(I)\big).$$

This says that since only a candidate's set of grades count, there are $(n!)^m$ other profiles that give the identical final grades. Among them is the *ordered profile*

$$\Phi^* = \begin{pmatrix} r_1^A \geq & \dots & \geq r_j^A \geq & \dots & \geq r_n^A \\ \vdots & & \vdots & & \vdots \\ r_1^I \geq & \dots & \geq r_j^I \geq & \dots & \geq r_n^I \\ \vdots & & \vdots & & \vdots \\ r_1^Z \geq & \dots & \geq r_j^Z \geq & \dots & \geq r_n^Z \end{pmatrix}.$$

The majority-grades $f^{maj}(I)$ of the competitors $I \in \mathcal{C}$ are the entries of the $(n+1)/2$-th column when n is odd, and those of the $(n+2)/2$-th column when n is even. When the majority-grades of two competitors A and B differ, the one with the higher majority-grade ranks ahead of the other. When the majority-grades of two competitors are equal, no more useful information concerning these two competitors can be drawn from this grade.

> The *majority-ranking* ($\succ_{maj}$) between two competitors is determined by:
> (1) If $f^{maj}(A) > f^{maj}(B)$ then $A \succ_{maj} B$.
> (2) If $f^{maj}(A) = f^{maj}(B)$, the majority-grade is dropped from the grades of each of the competitors, and the procedure is repeated.

Theorem 15 *The majority-ranking always ranks one competitor ahead of another unless the two are assigned an identical set of grades by the judges.*

The *1st-majority-grade* of a competitor is the majority-grade of the entire jury; the *2nd-majority-grade* is the majority-grade of the grades that remain after the 1st-majority-grade has been dropped; the *ith-majority-grade* is the majority-grade of the grades that remain after the first $i-1$ majority-grades have been dropped. A competitor's *majority-value* is a vector of n components that assigns, in order, his 1st-, 2nd-, 3rd,..., nth-majority-grades.

Theorem 16 *$A \succ_{maj} B$ if and only if A's majority-value is lexicographically higher than B's.*

The majority-value assigns a specific value to each competitor expressed in terms of the common language. It may be transformed into a rational number when the language is finite. E.g., if the language of grades is 0, 1, ..., 13 (as are school grades in Denmark) and a competitor receives the grades $(7, 7, 9, 10, 11)$

then his majority-value is $9,07100711$. Dividing by 1.01010101 rescales the final values so that the minimum is 0, the maximum 13, and a candidate assigned the same grade α by all judges has a rescaled majority-value of α.

Juries of different sizes In practice, contests with many competitors have many juries which sometimes may be of different sizes. How are two competitors to be compared *if they have the same majority-grade* bestowed by juries of different sizes? Two procedures naturally suggest themselves.

> *Annex the majority-grade* of each competitor considered by the smaller jury to its grades as many times as necessary to equal the number of grades of the larger jury.

The rationale for this is that the most reliable collective information concerning the final grade of a competitor given by any jury is its majority-grade, so it is added. It has been characterized as the only method that "compensates fairly" when a grade is adjoined. The second procedure takes the dual point of view.

> *Remove the majority-grade* of each competitor evaluated by the larger jury from its grades. If either their new (or 2nd-majority) grades distinguish them from the competitors considered by the smaller jury or their number of grades is the same as those of the smaller jury, they can be ranked. Otherwise, repeat.

The two procedures result in identical rankings. The *general majority-ranking* ($\succ_{gmaj}$) between two competitors A and B when B's jury is no larger than A's jury is defined by the majority-ranking ($\succ_{maj}$) applied to two sets of grades of equal size: A's grades, and B's grades supplemented, to the extent necessary, by its majority-grade.

Theorem 17 *The general majority-ranking $\succ_{gmaj}$ is a complete, transitive order of all competitors.*

Hence a judge need not assign a grade to every competitor.

Characterization Given input grades $\mathbf{r}^A = (r_1^A, \ldots, r_n^A), \mathbf{r}^B = (r_1^B, \ldots, r_n^B)$ of two competitors, how should they be ranked? Write $A \succ_S B$ to mean A is ranked ahead of B, and $A \succeq_S B$ to mean either A is ahead of B or they are tied. If $n \geq 3$, the highest and the lowest grade are $\mathbf{r}$'s *residual grades*; its *set of center grades* is obtained by dropping the residual grades.

A *social ranking function (SRF)* should satisfy several properties. It should be (1) *monotone*: if $A \succeq_S B$ and one judge raises the grade he gives to A then $A \succ_S B$. It should be (2) *decisive for the center grades*: the ranking between A and B is the ranking determined by the center grades unless that ranking is a tie; in that case, the ranking is determined by the residual grades. Finally, it should (3) *reward consensus.*

Theorem 18 *The majority-ranking is the unique monotone social ranking function (SRF) that is decisive for the center and rewards consensus.*[28]

[28] The majority-value with Borda-points or Condorcet-points as inputs provide SWFs for the

The majority-ranking is IIA in the sense of Arrow: the order between two competitors depends only on their respective grades. It is also a *grading compatible SRF*: it depends only on the ordered profile Φ^* of grades.

Remark This theory is not "cardinal": adding grades is meaningless. Nor is it "ordinal": a voter's message depends upon the particular words of the language that is used, so different common languages with the same number of words may lead to different rankings. And yet, a grade in a common language has an absolute meaning.

6 Practice

Jury decisions The majority-grade and ranking are described in the new edition of the French "Bible" on wines.[29] The majority-grade was tested in a wine competition, *Les citadelles du vin*, held June 15-17, 2006 in the Bordeaux region of France. 1,247 different wines competed, 60 judges were organized in juries of five members (sometimes fewer). As usual, judges completed the "Sensorial analysis tasting sheet for wine judging competitions" of the *Organisation Internationale de la Vigne et du Vin* for each wine: fourteen attributes of the wine were assigned points (from 0 "bad" to 6 or 8 "excellent"); their sum determined whether the wine was bestowed a gold, silver, bronze or no medal at all. But the sum misses the point because it "has difficulty in detecting exceptional wines by overly favoring those that are 'taste-wise correct.'"[30] Moreover, there is strong evidence showing that judges work "backwards"—they first decide the grade they wish to bestow, then assign points to attributes whose sum yield the grade. The judges preferred to answer "*For you, this wine is:*" with one of five descriptions that constituted the common language in this experiment: *Excellent, Very good, Good, Average*, or *Mediocre*. A preliminary evaluation of the experiment concludes: "The 'majority-grade' correctly distinguished . . . the wines, in accordance with the traditional objectives of wine competitions. This system seems better adapted than the [old system] . . . [However], the scale of five levels—97% of the grades were confined to three levels—should be extended."[31]

Voting *Approval voting (AV)* uses a common language of two words, 1 "approve" and 0 "disapprove." The majority-ranking with a 0,1-language is the approval voting ranking, so AV is a special case of the majority-value. AV has

traditional model that combat strategic manipulation. The 1rst-majority grade with Borda-points was used to rank figure skaters prior to 2004, with *ad hoc* rules to resolve ties. See Balinski and Laraki, *op. cit.*

[29] Émile Peynaud and Jacques Blouin, *Le goût du vin*, Dunod, Paris, 2006. We are deeply indebted to Jacques Blouin for having contacted us in 2002, suggesting that our work on voting should be useful in the classification of wines. His suggestion led to the development of our approach.

[30] É. Peynaud and J. Blouin, *Découvrir le goût du vin*, Dunod, Paris, 1999, p. 109.

[31] J. Blouin, "Essai d'application de la 'note majoritaire' à un concours de vins," unpublished note, July 2006.

been tested in a variety of settings, notably professional scientific societies. It was also tested in parallel with the first round of the French presidential election of 2002, when sixteen candidates presented themselves: voters were clearly happy to be able to express themselves better with AV than casting at most one solitary vote.[32] The arguments for and against AV have been cast in the context of the traditional model and have not addressed the real problem. The only solid results assume "dichotomous preferences,"[33] but "like" and "dislike" (or the very different "for" and "against") is much too limited a language.

Why do electors vote at all, since they hardly expect to determine the outcome?[34] They feel the moral imperative to express themselves: why else do so many cast blank votes? A richer language should encourage greater public participation. Even in an election having one candidate, the majority-grade permits electors to express themselves whatever language is used! Exactly what common language should be used in, say, a presidential election, is not clear. Perhaps the common language should be the grading system used in the nation's educational system: from a low of F to a high of A in the USA, from 0 to 20 in France, or 0 to 13 in Denmark. Alternatively, an election of an official might ask each voter: "For you, this candidate is *Exceptional, Accomplished, Capable, Average, Limited*, or *Incompetent* to undertake the high responsibilities of [the office]."

Common language How to define a common language in general remains an open question, though in many applications—skating, diving, gymnastics, piano competitions—languages already exist. Different applications naturally call for different common languages. Experimentation will be necessary to define a language; and, as is true of any language, it will alter over time. The *Les citadelles du vin* experiment suggests that judges (and voters?) shun the highest and lowest grades. It may be best to define a language with an even number of words in order to prevent voting in the middle, or not. The nature of the words or numbers used will illicit different voting and judging behavior: the words themselves matter! The environment in which judging and voting take place may also. Just imagine, how would *responsible voters* behave were they to read Ramon Lull's solemn proclamation of 1297 before casting their ballots:

> ...[It] is necessary to ascertain that in the election three things should be considered, of which the first is honesty and holiness of life, the second is knowledge and wisdom, and the third is a suitable disposition of the heart. Each person having a vote in the chapter should take an oath by the holy gospels of God to consider these three things and to always elect the person in whom they are best [embodied].[35]

[32] M. Balinski, R. Laraki, J.-F. Laslier and K. van der Straeten, "Résultats du vote par assentiment," *Pour la science*, June, 2002, p. 13.

[33] S. Barberà, H. Sonnenschein, and L. Zhou, "Voting by committees," *Econometrica* 59 (1991) 595-609; and A. Bogomolnaia, H. Moulin, and R. Strong, "Collective choice under dichotomous preferences," *Journal of Economic Theory* 122 (2005) 165-184.

[34] Goodin and Roberts, *op. cit.*

[35] Quoted in G. Hägele and F. Pukelsheim, "Llull's writings on electoral systems," *op. cit.*