9 New Model

Revolution is not the uprising against pre-existing order, but the setting-up of a new order contradictory to the traditional one.
—José Ortega y Gasset

When we mean to build,
We first survey the plot, then draw the model;
And when we see the figure of the house,
Then must we rate the cost of the erection;
Which if we find outweighs ability,
What do we then but draw anew the model
In fewer offices, or at last desist
To build at all?
—William Shakespeare

Over seven hundred years of effort and a host of impossibility theorems show that the "Arrovian model", where many individual rankings are to be resolved into a single collective ranking, cannot be made to work: there is no satisfactory mechanism for doing what is wanted. Experience shows, on the other hand, that it is a relatively simple matter to invent grades, scores, levels, or measures to evaluate the performances of students, figure skaters, divers, and musicians, the qualities of wines and cheeses, and the intensities of seismic events, and so by inference to determine the relative merits of competitors in any situation. With use, the grades take on meaning, so they come to constitute a common language of evaluation. Experience also shows that what to do with judges' scores—how to resolve them into a single score—is far from evident. Practical people have devised many different mechanisms.

The first step is to formulate the problem precisely. That is the aim of this chapter. It presents the basic model, which consists of a common language, a set of judges, and a set of competitors.

9.1 Inputs

A *language* Λ is a set of *grades* (words, levels, or categories) denoted by lowercase letters of the Greek alphabet, α, β, ... It is strictly ordered; specifically, supposing $\alpha, \beta, \gamma \in \Lambda$, (1) any two levels may be compared, $\alpha \neq \beta$ implies either $\alpha \prec \beta$ (β is the higher grade) or $\alpha \succ \beta$ (α is the higher grade); and (2) transitivity holds, $\alpha \succ \beta$ and $\beta \succ \gamma$ imply $\alpha \succ \gamma$. Otherwise, there is no restriction: a language Λ may be either finite or a subset of points of an interval of the real line. $\alpha \succeq \beta$ means that either α is a higher grade than β or $\alpha = \beta$.

There is a finite set of m *competitors* (alternatives, candidates, performances, competing goods) $\mathcal{C} = \{A, \ldots, I, \ldots, Z\}$. Individual competitors are denoted by uppercase Latin letters.

There is also a finite set of n *judges* $\mathcal{J} = \{1, \ldots, j, \ldots, n\}$. Individual judges are denoted by lowercase Latin letters, typically i, j, k.

A problem is completely specified by its *inputs*, or a *profile* $\Phi = \Phi(\mathcal{C}, \mathcal{J})$: it is 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 $I \in \mathcal{C}$. Thus, if $\mathcal{C}$ is a collection of wines, $\mathcal{J}$ is a jury of five œnologists, and Λ is a language of six grades or levels—say, *excellent, very good, good, mediocre, poor, bad*. Each judge gives to each wine one of the six grades, and the profile Φ is a matrix of grades with five columns and as many rows as there are wines in the collection to be tasted.

9.2 Social Grading Functions

A *method of grading* is a function F that assigns to any profile Φ—any set of grades in the language Λ assigned by judges to competitors—one single grade in the same language for every competitor:

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

Thus $F(\Phi)$ is a vector whose Ith component is the collective or *final grade* assigned to competitor $I \in \mathcal{C}$ by the mechanism F. As an example, suppose three wines A, B, and C were to be evaluated by five judges in the language postulated earlier. A possible profile Φ is the matrix that is the argument of the function F, and a possible set of grades for the three wines A, B, and C (in that order) is on the right:

$$F\begin{pmatrix} \textit{very good} & \textit{good} & \textit{good} & \textit{mediocre} & \textit{good} \\ \textit{excellent} & \textit{good} & \textit{good} & \textit{poor} & \textit{very good} \\ \textit{mediocre} & \textit{excellent} & \textit{poor} & \textit{good} & \textit{bad} \end{pmatrix} = \begin{pmatrix} \textit{good} \\ \textit{very good} \\ \textit{mediocre} \end{pmatrix}.$$

In Arrow's model the inputs are the judges' rankings of the candidates; there is no language or measure. A ranking function—or what he calls a social welfare function—assigns to any preference-profile, one single ranking of society. In terms of wines this would mean that every judge rank-orders all of them, and the ranking function deduces one collective rank-order among them. But the primary aim of the grading model is to *classify* competitors or alternatives, to give them final grades as students are given final grades. The final grades may be used to rank competitors, but only up to a point, because several competitors appreciated differently by the judges may have a same final grade.

Many different grading methods F may be imagined. When the language is numerical, say grades range from 0 to 100, the most often encountered example is an F that assigns to each competitor the average of the grades given her by the judges. Other possibilities would be to assign each competitor the lowest of all her grades or the highest of all of them. But F should obey some minimal requirements to be deemed acceptable. What should they be?

They are directly inspired by the requirements imposed on the traditional model of social choice theory. There should be no inherent advantage or disadvantage given to any one or more competitors: all should be treated equally. So if, for example, the three wines A, B, and C were listed in a different order—say, B followed by A, then by C—then F should yield the same answer: B very good, A good, C mediocre. When the rows (or competitors) of a profile Φ are permuted, F should give the identical answer permuted in the same way. Axiom 9.1 states this formally.

Axiom 9.1 *F is neutral, $F(\rho\Phi) = \rho F(\Phi)$, for any permutation ρ of the competitors (or rows).*

Similarly, all judges are to have the same influence on the grades, meaning that when the columns (or judges) of a profile are permuted, F should give the identical answer. In some situations—to protect minority rights, for example—it may be desirable to give certain judges more weight than others. It suffices to count their inputs several times.

Axiom 9.2 *F is anonymous, $F(\Phi\tau) = F(\Phi)$, for any permutation τ of the judges (or columns).*

A method of grading is *impartial* when it is both neutral and anonymous.

Three other properties naturally impose themselves.

First, if every judge is in agreement on the grade to be given to a competitor, then he must be assigned that final grade.

Axiom 9.3 *F is unanimous: If a competitor is given an identical grade α by every judge, then F assigns him the grade α.*

That is, $F(\Phi)(I) = \alpha$ when $\Phi(I, j) = \alpha$ for every $j \in \mathcal{J}$.

Next, if in comparing two profiles, a competitor I's grades in the second are all the same or lower than in the first, then F cannot assign the competitor a higher grade in the second case than in the first. Moreover, if all the competitor's grades are strictly lower in the second profile, then F must assign him a strictly lower grade in the second case.

Axiom 9.4 *F is monotonic: If two inputs Φ and Φ' are the same 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 strictly higher grades to competitor I in Φ than in Φ', then $F(\Phi)(I) \succ F(\Phi')(I)$.*

In other words, if $\Phi(I, j) \succeq \Phi'(I, j)$ for every $j \in \mathcal{J}$, then $F(\Phi)(I) \succeq F(\Phi')(I)$; and if $\Phi(I, j) \succ \Phi'(I, j)$ for every $j \in \mathcal{J}$, then $F(\Phi)(I) \succ F(\Phi')(I)$. When F satisfies only the first of the two properties, it will be said to be *weakly monotonic*; when only the second, it will be said to be *strictly monotonic* (sometimes referred to as Pareto efficiency or unanimity).

Finally, Arrow's famous independence of irrelevant alternatives is a very natural condition when translated into the context of this model. The collective grade of a competitor should depend on her grades alone: it should certainly not depend on any other competitor's grades.

Axiom 9.5 *F is independent of irrelevant alternatives in grading (IIAG): If the lists of grades assigned by the judges to a competitor $I \in C$ in two profiles Φ and Φ' are the same, then $F(\Phi)(I) = F(\Phi')(I)$.*

That is to say, if $\Phi(I, j) = \Phi'(I, j)$ for every $j \in \mathcal{J}$, then $F(\Phi)(I) = F(\Phi')(I)$.

These axioms are already sufficient to reduce the choice of a method of grading to a manageable, well-defined class. A function

$$f : \Lambda^n \to \Lambda$$

that transforms grades given to one competitor into a final 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*: $f(\alpha, \alpha, \ldots, \alpha) = \alpha$;
- *monotonicity*:

$$\alpha_j \preceq \beta_j \quad \text{for all } j \Rightarrow f(\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n) \preceq f(\alpha_1, \ldots, \beta_j, \ldots, \alpha_n)$$

and

$$\alpha_j \prec \beta_j \quad \text{for all } j \Rightarrow f(\alpha_1, \ldots, \alpha_n) \prec f(\beta_1, \ldots, \beta_n).$$

When f only satisfies the first of the two monotonicity properties, it will be said to be *weakly monotonic*.

A language Λ is often parameterized as a bounded interval of the non-negative rational or real numbers. In either case an obvious example of an aggregation function assigns the mean value of its arguments. Other examples are those that assign the geometric, the harmonic, or any other of the well-known means; those that assign the minimum or the maximum value of its arguments; or more generally, those that assign the kth largest of the arguments for $1 \leq k \leq n$ (called order statistics by probabilists).

Theorem 9.1 *A method of grading F is impartial, unanimous, monotonic, and independent of irrelevant alternatives in grading if and only if $F(\Phi)(I) = f(\Phi(I))$ for every $I \in \mathcal{C}$, for some one aggregation function f.*

Proof If there is an aggregation function f that defines F as in the statement of the theorem, then the axioms are obviously met by F. On the other hand, suppose F satisfies the axioms. IIAG and neutrality imply that F determines the grade of a competitor $I \in \mathcal{C}$ solely on the basis of the grades assigned to I; so call the function that does this f. The other three axioms—anonymity, unanimity, and monotonicity—immediately establish the corresponding properties of f, so it must be an aggregation function. ■

The theorem is obvious. Yet it already eliminates Arrow-type impossibilities.

In practice, grades are almost always numbers, final grades almost always averages or truncated averages. The language of a judge's grades may be discrete—in Australian wine tastings, for example, the grades of individual judges can range from 0 to 20, and they are assigned in multiples of $\frac{1}{2}$—but the language of final grades is richer: with five judges a wine's final grade can be 17.10. Or to take a more probing example, consider the seven grades used by the Union Internationale des Œnologues (U.I.Œ.) (see chapter 7) and assign them numbers: excellent 6, very good 5, good 4, passable 3, inadequate 2, mediocre 1, bad 0. The average of a five-person jury giving the grades—very good, good, good, good, bad—would be 3.40; that might reasonably be described as a passable+. Though that word is not in the judges' original language, it almost surely *becomes* a word in the language. So why not let it as well as its sisters and its cousins and its aunts enter all at once? The point is that in practice final grades are often more detailed than the grades a judge is allowed to give, yet they cannot help but take on meanings of their own. Were they then to be used in the judges' language, a further enrichment of the final grades would ensue. Why not simply take all the possible numbers as grades to begin with?

Accordingly, in conformity with most practical applications, the common language is parameterized as a subset of real numbers and whatever aggregation is used, small changes in the parameterization or the input grades should naturally imply small changes in the outputs or the final grades. The analysis in the rest of this book could equally well have chosen any open or half-open interval, including $[0, \infty)$. Even if the original language is finite, the possibility of taking an arbitrary aggregation function implies that all possible parameterizations must be considered. Accordingly, the common language will be taken to be $[0, R]$, as did Laplace. Whereas Americans may like $R = 100$, the French $R = 20$, and the Danes $R = 13$, most mathematicians probably prefer $R = 1$. The choice is unimportant. Grades are almost always bounded. The grades used by the International Organization of Vine and Wine (OIV) are an exception: 0 is the best, ∞ the worst.

Suppose that Λ and Λ' are the number grades corresponding to two languages or parameterizations that are *ϵ-close*: $r \in \Lambda$ implies there exists an $r' \in \Lambda'$ with $|r - r'| < \epsilon$, and symmetrically, $r' \in \Lambda'$ implies there exists an $r \in \Lambda$ with $|r - r'| < \epsilon$. It is then clear that a method of grading F should be defined by an aggregation function f that satisfies $|f(r_1, \ldots, r_n) - f(r'_1, \ldots, r'_n)| < \eta(\epsilon)$ when $\max_{j \in \mathcal{J}} |r_j - r'_j| < \epsilon$ for some positive function $\eta(\epsilon)$ that converges to 0 when ϵ approaches 0. That is, f should be uniformly continuous, so, since $[0, R]$ is compact, f should be continuous.

Axiom 9.6 *F (and its aggregation function f) is continuous.*

Two lists of grades that are very similar should clearly be assigned final grades that differ by very little. 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 discrete languages as well. *The characterizations in the next chapters sometimes require the language to be sufficiently rich and the functions to be continuous; but the properties of the functions that are characterized hold for arbitrary finite languages.* Many theorems do not require all axioms.

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.

The number of candidates or of judges may be varied to study certain properties and phenomena.

9.3 Social Ranking Functions

With Arrow's model, "one of the consequences of the assumption of rationality is that the choice to be made from any set of alternatives can be determined by the choices made between pairs of alternatives" (Arrow 1951, 20). But, as Arrow's theorem shows, this ideal cannot be realized. One of the consequences of rationality in the new model is that every single alternative has a final grade that is independent of all other alternatives: there is no ambiguity in grading. But what happens when the aim is to rank alternatives?

Given a finite language Λ, judges assign grades to any number of competitors that are the inputs or profile

$$\Phi = \begin{pmatrix} \vdots & \vdots & \cdots & \vdots & \vdots \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{n-1} & \alpha_n \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \beta_1 & \beta_2 & \cdots & \beta_{n-1} & \beta_n \\ \vdots & \vdots & \cdots & \vdots & \vdots \end{pmatrix}.$$

Imagine that a competitor A is assigned the list of grades $\alpha = (\alpha_1, \ldots, \alpha_n)$ and a competitor B the list $\beta = (\beta_1, \ldots, \beta_n)$. A *method of ranking* is a nonsymmetric binary relation $\succeq_S$ that compares any two competitors, A and B, whose grades belong to some profile Φ. By definition, $A \approx_S B$ if $A \succeq_S B$ and $B \succeq_S A$; and $A \succ_S B$ if $A \succeq_S B$, but it is not true that $A \approx_S B$. Thus $\succeq_S$ is a complete binary relation.

Any reasonable method of ranking should possess certain minimal properties.

Axiom 9.7 *The method of ranking $\succeq_S$ is neutral: $A \succeq_S B$ for the profile Φ implies $A \succeq_S B$ for the profile $\sigma\Phi$, for σ any permutation of the competitors (or rows).*

Axiom 9.8 *The method of ranking $\succeq_S$ is anonymous: $A \succeq_S B$ for the profile Φ implies $A \succeq_S B$ for the profile $\Phi\sigma$, for σ any permutation of the judges (or columns).*[1]

Axiom 9.9 *The method of ranking $\succeq_S$ is transitive: $A \succeq_S B$ and $B \succeq_S C$ implies $A \succeq_S C$.*

1. So, for example, A's grades $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ are permuted to $(\alpha_{\sigma 1}, \alpha_{\sigma 2}, \ldots, \alpha_{\sigma n})$.

Axiom 9.10 *The method of ranking $\succeq_S$ is independent of irrelevant alternatives in ranking* (IIAR): *If $A \succeq_S B$ for the profile Φ, then $A \succeq_S B$ for any profile Φ' obtained from Φ by eliminating or adjoining some other competitor (or row).*

Axiom 9.9 demands that the Condorcet paradox be avoided. Axiom 9.10 is strong independence of irrelevant alternatives, as defined in chapter 3. It demands that Arrow's paradox be avoided. These are the two important paradoxes that have been observed to occur in real competitive situations.

A method of ranking *respects grades* if the rank-order between two candidates A and B depends only on their sets of grades.

Thus, the preference lists induced by the grades must be forgotten: it matters not which judge gave which grade. In other words, if two judges or voters switch the grades they give to a candidate, then nothing changes in the jury's or the electorate's ranking of all candidates.

A method of ranking *respects ties* if when any two competitors A and B have an identical set of grades, they are tied, $A \approx_S B$.

Respecting grades together with impartiality implies respecting ties.

Theorem 9.2 *A method of ranking is neutral, anonymous, transitive, and independent of irrelevant alternatives in ranking if and only if it is transitive, and respects ties and grades.*

Proof To compare the sets of grades of two competitors A and B it suffices to compare them alone (by IIAR).

Suppose A's list of grades is $(\alpha_1, \alpha_2, \dots, \alpha_n)$ and B's list of grades is a permutation σ of them, $(\alpha_{\sigma 1}, \alpha_{\sigma 2}, \dots, \alpha_{\sigma n})$.

To begin, consider the profile

$$\Phi^1 = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{\sigma 1} & \cdots & \alpha_n \\ \alpha_{\sigma 1} & \alpha_2 & \cdots & \alpha_1 & \cdots & \alpha_n \end{pmatrix},$$

where the grades of A are in the first row and those in the second row are called those of A$'$. Suppose $A \succeq_S A'$. Permuting the grades of the two judges 1 and $\sigma 1$ changes nothing by anonymity,

$$\Phi^{1*} = \begin{pmatrix} \alpha_{\sigma 1} & \alpha_2 & \cdots & \alpha_1 & \cdots & \alpha_n \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{\sigma 1} & \cdots & \alpha_n \end{pmatrix},$$

so the first row of Φ^{1*} ranks at least as high as the second; but by neutrality $A' \succeq_S A$, so that $A \approx_S A'$. Thus $(\alpha_1, \alpha_2, \dots, \alpha_n) \approx_S (\alpha_{\sigma 1}, \alpha_2, \dots, \alpha_n)$ and the second list agrees with B's in the first place.

Now consider the profile

$$\Phi^2 = \begin{pmatrix} \alpha_{\sigma 1} & \alpha_2 & \cdots & \alpha_{\sigma 2} & \cdots & \alpha_n \\ \alpha_{\sigma 1} & \alpha_{\sigma 2} & \cdots & \alpha_2 & \cdots & \alpha_n \end{pmatrix},$$

and permute judges 2 and $\sigma 2$ to conclude, as in the first step, together with transitivity, that $(\alpha_1, \alpha_2, \ldots, \alpha_n) \approx_S (\alpha_{\sigma 1}, \alpha_{\sigma 2}, \ldots, \alpha_n)$, the second list agreeing with B's in the first two places. Continuing, in at most n steps, $(\alpha_1, \alpha_2, \ldots, \alpha_n) \approx_S (\alpha_{\sigma 1}, \alpha_{\sigma 2}, \ldots, \alpha_{\sigma n})$, so ties are respected.

The order between any two lists α and β respects grades because each of them is equivalent to a unique representation in which the list is written from the highest to the lowest grade and it suffices to compare them, so grades are respected. The converse is immediate. ■

Notice that no axiom asks that the language of grades be understood in the same way by all voters or judges. The only implicit assumption is that the scale of grades be absolute for each individual judge or voter. This implies that if some competitor is added or dropped, a judge's true grade remains the same. On the other hand, as will be seen anon, for the final decisions to be meaningful, the scale of grades must be common to all judges or all voters.

A *social ranking function* (SRF) is a method of ranking that satisfies the four ranking axioms.

This simple theorem is essential; it says that if Arrow's and Condorcet's paradoxes are to be avoided, then the preferences induced by the grades must be forgotten. *Who gave what grade cannot be taken into account*: only the sets of grades themselves may be taken into account. It also says that in the new model Arrow's ideal can be realized: "The choice to be made from any set of alternatives can be determined by the choices made between pairs of alternatives."

9.4 The Role of Judges' Utilities

Nothing has been said yet concerning the behavior of the judges or the voters, their complex and often secret aims, their likes and dislikes. The tradition in the theory of social choice is to assume that judges and voters have preferences, invariably expressed as rank-orders. But the word "preference" misleads. A judge in a court of justice evaluates in conformity with the law, which has nothing to do with his preferences; a judge may dislike a wine presented in a competition yet give it a high grade because of its merits, or he may like one and without qualm give it a low grade because of its demerits; an elector may cast a vote not in accord with his personal opinions of the candidates but rather in the hope of making the correct decision by electing the best candidate for the job (see, e.g., Goodin and Roberts 1975; but Llull, Cusanus, Condorcet, and all the early thinkers formulated the problem in these terms, as do also some philosophers today, e.g., Estlund 2008).

Thus whereas the traditional model pretends to aggregate the preferences of judges and voters, in fact it does nothing of the kind. It amalgamates individuals' rankings of the candidates—the input—into society's ranking of the candidates—the output. The possible outputs are *rankings*, yet the inputs say nothing about how a judge or a voter compares the rankings.

In the real world the deep preferences or utilities of a judge or a voter are a very complicated function that depends on a host of factors, including the decision or output, the messages of the other judges (a judge 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 a "democratic" function to one rendered by an "oligarchic" function, or the contrary), and the message she thinks is the right one (a judge may prefer honest behavior, or not). We contend that the deep preferences of judges or voters *cannot* be the inputs of a practical model of voting. *A judge's input is simply a message, no more no less*. But her input, chosen strategically, depends, of course, on her deep preferences or utilities.

Amartya Sen's model (1970) and the subsequent work on "welfarism" (Blackorby, Donaldson, and Weymark 1984; Bossert and Weymark 2004)—often referred to as social welfare functionals—postulates real number utilities on candidates as the input, a rank-ordering of candidates as the output. The motivation is the study of social welfare judgments in the context of Arrow's framework but with more information in order to avoid the impossibilities. Sen makes no claim that this approach is valid in the context of voting. As with any mathematical model, the mathematical symbols may be given very different interpretations. At a formal level, reinterpreting the symbols of the inputs of Sen's model as the grades of a language yields a social ranking function. But this ignores the essential concept of a common language. By contrast, utilities measured in an absolute scale play no significant role in the social welfare functional literature, which focuses instead on weaker information invariance assumptions (although they are often assumed for simplicity, e.g. Blackorby, Bossert, and Donaldson 2005).

Social welfare functionals are not intended to enable a comparison of rankings. For, how are two outputs—two rankings of the candidates—to be compared by a voter or judge on the basis of his utilities for individual candidates? If the answer is by looking at the first-place candidate of the rankings, then all $(m-1)!$ rankings that have the same first-place candidate must be taken as giving him equal satisfaction. This is too restrictive for a theory (or practice) that designates winners and orders of finish.

The language of grades has nothing to do with utilities (viewed as measures of individual satisfaction). Grades are absolute measures of merit. In the context of voting and judging, utilities are relative measures of satisfaction. The 2002 French presidential elections offer a perfect example of the difference. The voters of the left would have hated to see Jacques Chirac defeat Lionel Jospin: their utilities for a Chirac victory would have been the lowest possible. The same voters were delighted to see Chirac roundly defeat Le Pen in the second round: their utilities for a Chirac victory were the highest possible. On the other hand, these same voters would probably have given Chirac a grade of *Acceptable* or *Poor* (on a scale of *Excellent, Very Good, Good, Acceptable, Poor, To Reject*) were he standing against Jospin, Le Pen, or anyone else.

Formally, a distinction must be made between two different types of scales of measurement. An *absolute* scale measures each entity individually (height, area, merit). A *relative* scale measures each entity with respect to a collective of like entities (velocity, satisfaction). Were voters to be asked their satisfaction as inputs, adjoining or eliminating candidates would alter their answers, provoking the possibility of Arrow's paradox. A common language must be an absolute scale of measurement.

Utility plays another, important role in voting and judging. A decision maker is routinely assumed to behave in such manner as to try to maximize his utility. But what is it? In theory the utility of a judge or voter j may be imagined to be a function $u_j(\Phi^*, \Phi, f, C, \Lambda)$, where $\Phi^* = (\alpha^*_{ij})$, with α^*_{ij} the grade judge j believes candidate i merits, $\Phi = (\alpha_{ij})$, with α_{ij} the grade judge j actually gives candidate i, f is the aggregation function, C the set of competitors, and Λ the common language that is used. The utility of judge j could include a term $-|\alpha^*_{ij} - \alpha_{ij}|$ if she wished to grade candidate i honestly; it could include a term $-\sum_{k \neq j} |\alpha^*_{ik} - \alpha_{ik}|$ if she wished that the other judges graded i honestly; it could include a term $-|\Lambda - \Lambda^*_j|$ if she wished that the common language was Λ^*_j rather than Λ; and the reader can no doubt invent many other utility functions that a judge might have, or plausible components of them. One hypothesis is to imagine that a judge's utility is single-peaked in the grade of each candidate i, $u_j = \sum_i -|\alpha^*_{ij} - f(\alpha_{i1}, \ldots, \alpha_{in})|$: the further the final grade $f(\alpha_{i1}, \ldots, \alpha_{in})$ is from what judge j believes it should be, the less her satisfaction. Another is that a judge's utilities depends solely on the winner, which is usually assumed in the analysis of voting games. In fact, of course, judges' utilities, judges' beliefs, their beliefs about the others' beliefs, their likes and dislikes for the decision mechanism or the language are all completely unknown and often hidden, and they change from one competition to another (indeed, perhaps a voter's utility on a sunny election day differs from that on a rainy election day).

In the terms of the current technical jargon, we are faced with a problem of *mechanism design*: "[Individuals'] actual preferences are not publicly observable. As a result ... individuals must be relied upon to reveal this information . . . [The problem is] how this information can be elicited, and the extent to which the information revelation problem constrains the ways in which social decisions can respond to individual preferences" (Mas-Colell, Whinston, and Green 1995, 857). This is often seen as a problem of the theory of games where the information is incomplete. The standard approach postulates that every individual is of a certain *type* and associates to each type a utility function that depends only on the outcome. Typically, the individuals' types are drawn from a set of types by some known prior probability distribution, and the utility functions are all of some common analytical form (whose parameters vary with the different types). The methods are then, of course, dependent on the utilities that are postulated.

In contrast, the methods we develop make no overall assumptions concerning utilities. They are similar, in this regard, to Vickrey's "second price" mechanism, which allocates the good up for auction to the highest bidder at a price equal to the second highest bid (Vickrey 1961): it depends only on the bidders' bids—their "private values"—not their utilities. Our mechanisms depend *only* on what in practice can be known. Knowing the judges' or voters' true utilities is unnecessary to much of the analysis. The mechanisms that emerge as the only ones that separately satisfy each of several desirable properties are strategy-proof for large classes of reasonable utility functions, though not all. When they are not strategy-proof, they are unique in being the least manipulable methods in several well-defined senses.