

Preface

There was a risk in theorizing. I had witnessed, close up, the fatal, comic effect upon professors and students of hypotheses which had become unconscious convictions. And thus warned, I had thrown overboard, as a reporter facing facts, many of my college-bred notions . . . It was hard to do; ideas harden like arteries; indeed, one theory of mine is that convictions are identical with hardened arteries. But the facts . . . forced me to drop my academic theories one by one; and my reward was the discovery that it was as pleasant to change one's mind as it was to change one's clothes. The practice led one to other, more fascinating—theories.

—Lincoln Steffens

Kenneth Arrow begins his classic, *Social Choice and Individual Values*, with the sentence, “In a capitalist democracy there are essentially two methods by which social choices can be made: voting, typically used to make ‘political’ decisions, and the market mechanism, typically used to make ‘economic’ decisions.” In the second paragraph he poses the problem of social choice. “The methods of voting and the market . . . are methods of amalgamating the tastes of many individuals in the making of social choices . . . [Any] individual can be rational in his choices. Can such consistency be attributed to collective modes of choice, where the wills of many people are involved?” (1951, 1–2). His celebrated impossibility theorem answers *no!* for voting.

Arrow's conclusion begs refutation.

But Arrow carefully explained the limitations of his analysis. First, he explicitly assumed that “the behavior of an individual in making choices is describable by means of a preference scale [an ordered list]” (11). Second, he deliberately ignored the strategic aspects of voting, in his words, the “game aspects of social choice” (7). Third, he clearly stated his acceptance of the standard though unreasonable view that “individual values . . . are not capable of being altered by the nature of the process itself” (8). This assumes that a voter's or a judge's expression—his or her vote, his or her evaluation—is not altered by the actions of other voters or judges, nor by the mechanism by which all the

expressions are amalgamated. In so doing, Arrow anticipated future developments, showing where to seek an escape from the logical conclusions of his analysis: a brand new model of social choice not bound by these restrictive and unrealistic assumptions. Practice, it turns out, suggests how the model should be formulated.

A mechanism is “an instrument or process, physical or mental, by which something is done or comes into being.”¹ Society routinely uses mechanisms left unmentioned by Arrow to collectively designate who or what is best, second-best, and so on, down to worst: it *measures* in one way or another, then ranks in accordance with the measures and declares the winner to be the one with the highest score. Students are graded, then ranked according to their grades; the attributes of wines (e.g., tannicity, finish, bouquet, body) are assigned numerical values, then ranked according to the values; figure skaters, gymnasts, and divers are given marks for a particular exercise, then ranked according to their marks. In each of these instances there may be several or many judges who assign scores that may be quite different, yet each uses a language that is common to all and is understood by all. “Kenneth is an A+ student” is a meaningful statement (or was before the age of grade inflation). The statement that “Sonja’s free skating performance is worth 5.9” when the traditional 0–6 scale was used, or that “her skating skills component is 7.75” with the newly adopted scale, or that “Xu Sang’s inward flying $1\frac{1}{2}$ somersault was a 9.0” means something specific to figure skating or diving enthusiasts, though all may not agree with the score that was assigned.

The market mechanism—perhaps better thought of as an invisible hand, since its process is but loosely understood—itself provides the world with a measure: *price* expressed in terms of money. Money is, of course, a complex concept that plays many roles in the dynamic workings of an economy, but its units are mere convention, a numéraire expressed in euros, shekels, dinars, or dollars, invented to simplify or solve problems, just as are letter grades or numerical levels of performance. In past times the units differed: grain in ancient Babylon, rice in Japan, cigarettes in concentration camps, though gold and silver—more durable commodities—have enjoyed a longer-lasting use. Price measures, and as a consequence ranks. An instance is the famous classification of 1855, said to be of the Bordeaux wines, though in fact limited to the Médocs and Sauternes and exactly one Graves, Haut-Brion. The Exposition universelle de Paris of that year prompted an official request for a “complete and satisfactory description of the wines of the department” (Debuigne 1970). This ranking

1. *American Heritage Dictionary*. 2d college ed. Boston: Houghton Mifflin, 1982.

of the “grands crus” has steadfastly maintained its importance to this day; it was determined by the prices of the wines prevalent in those years (Markham 1997). Auctions—the traditional English ascending-price auction, the sealed second-price auction of Vickrey, the Dutch descending-price auction—are other well-defined mechanisms that also use price as a measure to determine winners.

Measurement as a means to choose and to rank has accumulated a rich experience and taught lessons well worth learning. Ranking figure skating competitions is a good example: it is an ongoing, dynamic process that reveals many of the defects known in methods of elections. The traditional system often saw the lead of one skater over another reverse as a result of a *third*'s performance—to the obvious consternation of the spectators. This is, of course, nothing but a violation of Arrow's “independence of irrelevant alternatives”: no system of ranking should allow the relative positions of two competitors to be influenced by the (irrelevant) performance of a third. Moreover, cheating—when a judge exaggerates the rank of competitors up or down—has provoked major scandals at the Olympics. This is nothing but strategic manipulation, a problem well known in voting theory and central to the theory of games. What did the International Skating Union do? It invented a new method that appears to satisfy Arrow's condition and pretends to combat the possibility of cheating. It is a method that measures. It is a mechanism that has flaws—as do the market mechanism when markets are “rigged” and auction mechanisms that are often plagued by the “winner's curse”²—but these flaws may be overcome.

Up to now, the problem of ranking competitions among athletes, goods, or musicians has remained completely separated from the problem of elections. No organized body of knowledge has accompanied ranking competitors. The methods are, by and large, “home-grown,” invented by skating enthusiasts for skaters, by oenologists for classifying wines, by piano maestros for discerning prizes to pianists. Nevertheless, the methods show good sense: increasingly, they use measures.

In contrast, the specialists in how to elect and rank have almost exclusively devoted their attentions to elections, either in small committees such as juries or in large electorates such as nations, where a voter submits either a vote for one or several candidates or an ordered list of his preferences among all candidates. The focus of the work has remained resolutely the same for almost a millennium. It is the model analyzed by Arrow: how to transform the so-called preference lists of individuals or judges into a preference list of society

2. The “winner's curse” occurs when the actual value of the good obtained is well below what was paid for it. This happens typically because the winning bidder made the highest estimate of a good of unknown value, as for example when bidding for the right to exploit an off-shore oil field.

or the jury. Despite Arrow's devastating result showing the traditional theory of social choice has no truly acceptable solution to the problem of how to elect and how to rank, that model—and the manner in which voters may express themselves—has remained unquestioned.

We show that this traditional model is fundamentally flawed, for reasons that go well beyond the classical paradoxes. First, it assumes that the preferences of voters and judges are expressed as rank-orderings. This is clearly not true: a voter confronted with (say) twelve candidates knows from personal experience that he does *not* formulate a list of all candidates from first to last. We present evidence that proves that this conception of what judges or voters have in their minds is simply false. Second, the traditional model does not make a clear distinction between the judges' or voters' preferences and their votes, which are the *messages* they are allowed to send by the method of voting that is used. A voter who announces a rank-ordering—or who votes for one candidate among many—is not and cannot be expressing all of his preferences; he is only sending a very limited and strategically chosen message. Third, there is a profound difference between the problem of electing one candidate and the problem of ranking several or many candidates, though this has not been widely appreciated. H. Peyton Young (1986) is the first to have made a clear distinction between them, showing how a same line of reasoning can yield a ranking of the candidates and a winner who is *not* the first-place candidate of the ranking. This, it turns out, is an irreconcilable difference of the traditional model; a new impossibility theorem shows how and why the two problems are incompatible.

In summary, insofar as it concerns ranking and electing, social choice theory hypothesizes a faulty model of reality to produce an inconsistent theory. So why on earth use it?

The analogy with wine, sports, and music opens the door to another view. Lord Kelvin's celebrated warning may be seen on the façade of the Social Sciences Research Building at the University of Chicago: "If you cannot measure, your knowledge is meager and unsatisfactory" (see Kuhn 1961, 178). The economist Frank Knight is reported to have quipped that for social scientists this means, "If you cannot measure, measure anyhow" (Kuhn 1961, 183). His remark makes more sense than most people at first may suppose.

We have studied what is done in practice. We have learned from the insights of oenologists, sportsmen, pianists, and others that the fundamental question should be posed differently. Instead of trying to translate many individual rankings of competitors into a single collective ranking, or many individual lists of preferences into a single collective list of preferences, a common language to measure should be defined, individuals should measure and assign grades to

each competitor in that language, and the many individual grades should then determine the single collective grade of each competitor. In short, the central problem becomes *how to transform many individual grades of a common language into a single collective grade*, when the many individuals have unknown preferences that are too complex to be formulated. Sharing a common language of grades makes no assumptions about a voter's or a judge's utilities or preferences. Utilities measure the satisfaction of a voter or a judge, grades measure the merits of competitors. The basic atoms of Arrow's model are the comparisons between pairs of alternatives, competitors, or candidates. The basic atoms of our model are the grades of a common language assigned to alternatives, competitors, or candidates. Grades yield rankings, but rankings assuredly do not yield grades.

The celebrated market mechanism works because it uses a common measure that facilitates comparisons of goods, services, assets, debts: monetary units. The evaluations in terms of a common language of grades are no more the utilities of judges or voters making a collective decision than the evaluations of items bought and sold in terms of money are the utilities of agents in a market. *The common language of the market is money. The "money" of collective decision is—in our model—the common language of grades.*

The change in point of view—in the premises of the underlying model of social choice—changes everything. The method and theory that emerges is simple. It must be, to be practical. We have called it *majority judgment*. It may be used to elect officials, to classify wines, and to rank contestants for international piano prizes and Olympic competitors in skating, diving, and gymnastics. It has been tested in classifying wines and electing a candidate to political office. A simple theory characterizes the methods that satisfy all the "good" properties that were stated in Arrow's axioms. Beyond overcoming Arrow's impossibility, the model makes it possible to address another important question: What mechanisms are the most robust against cheating and strategic manipulation of the judges or the voters? Or, what mechanisms make the judges' and voters' optimal strategies be to give the grades they believe the competitors and candidates merit? Understanding and discovering the psychology or the possible secret cabals of judges is one perfectly reasonable approach to combating strategic manipulation. Another is to design methods that make it impossible for cheating to take place or, if that ideal is unattainable, that minimize the possibility of manipulation. When grades replace orders, "possibility theorems" and "strategy-proof" or "least manipulative" methods replace impossibility theorems. The majority judgment method uniquely best combats strategic manipulation while satisfying the desirable properties of classical social choice theory. In fact, the approach suggests a new mechanism in the

context of the traditional model (the “Borda-majority judgment method”). The aim of this book is to state, to explain, to prove and to apply these theorems.

Brand new ideas are rare indeed. Anyone, at any time, cannot but build upon the accumulation of all of past knowledge, consciously or unconsciously. Confronted by a real problem—how should wines be judged and ranked?—we stumbled on a simple but fundamental idea, realized with some amazement how useful it could be, and set out to explore and develop it. Only when we began to write did we discover that others, most notably Sir Francis Galton and the Marquis de Laplace, had seen one or another aspect of these ideas before. The mass of knowledge of the theory of social choice and welfare—the concepts, theorems, paradoxes, and mechanisms—provided the road signs for the development of a unified theory of measuring, ranking, and electing. The crux of the matter is a new model in which the traditional paradigm—*to compare*—is replaced by a new one—*to evaluate*. In the words of Thomas Kuhn, “Since new paradigms are born from old ones, they ordinarily incorporate much of the vocabulary and apparatus, both conceptual and manipulative, that the traditional paradigm had previously employed. But they seldom employ these borrowed elements in quite the traditional way. Within the new paradigm, old terms, concepts, and experiments fall into new relationships one with the other. The inevitable result is what we must call, though the term is not quite right, a misunderstanding between the two competing schools” (1970, 149).

George Dantzig, the acknowledged father of linear programming and its many generalizations, begins the preface of his magnum opus (1963) with an often forgotten yet naked truth: “The final test of a theory is its capacity to solve the problems which originated it.” Our ambition is to establish a theory that meets this test. And so this book is addressed—in addition to the economists, mathematicians, operations research analysts, and political scientists who are the specialists of the theory of social choice—to anyone who is confronted with problems of electing and ranking, including the just plain interested voter (theorems and proofs may be skipped by those who are only interested in *how* to elect and rank).

Acknowledgments

We are indebted to many, friends and colleagues and institutions.

Friends and Colleagues Francis Bloch, B. Curtis Eaves, David Gale, Edi Karni, Yukio Koriyama, Friedrich Pukelsheim, Jérôme Renault, Ludovic Renou, Martin Shubik, Tristan Tomala, and H. Peyton Young supported and challenged us in discussions and communications, and made important suggestions and

critiques. We are deeply indebted to Jack Nagel, Hannu Nurmi, Maurice Salles, and John Weymark, each of whom analyzed the entire manuscript with great care and made very substantial recommendations. It is a particular pleasure to acknowledge the contributions of two exceptionally gifted graduate students in mathematics, Andrew Jennings and Cheng Wan. Andy Jennings read the entire manuscript, checking the proofs, the statements, the reasoning, and the prose, thereby eliminating errors, ambiguities, and infelicities. Cheng Wan carried out all the extensive computer analyses of the 2007 Orsay experiment. Claude Henry and Vincent Renard gave us their unstinting support and encouragement throughout. Anna Kehres-Diaz accepted and espoused the importance of the ideas for real, practical use, and backed the application for a patent and the development of professional software that permitted several experiments to be carried out on the Web. Last, yet also first, we thank Jacques Blouin. Dissatisfied with the existing methods for evaluating wines, he asked us whether we could find a better way; that is how this work began.

Institutions The C.N.R.S. (Centre National de Recherche Scientifique), where both of us have held permanent positions, is wonderful for those who wish to pursue long-term projects: they are free to go ahead and do it! The École Polytechnique, more particularly, the Laboratoire d'Économétrie, provided the ideal interdisciplinary environment in which to pursue this project that is at once mathematics, economics, political science, operations research, and statistics, with occasional dashes of linguistics, psychology, sociology, and philosophy. The D.R.I.P. (Direction des relations industrielles et des partenariats) of the École Polytechnique has steadfastly supported every endeavor to realize practical applications of our ideas.

The 2007 Orsay experience could not have been realized without the generous support of Orsay's Mayor, Madame Marie-Hélène Aubry, the staff of the Mayor's office, and our friends and colleagues who sacrificed their Sunday (a beautiful spring day) to urging voters to participate and explaining the idea: Pierre Brochot, Stéphanie Brochot Laraki, David Chavalarias, Sophie Chemarin, Clémence Christin, Maximilien Laye, Jean-Philippe Nicolai, Matias Nuñez, Vianney Perchet, Jérôme Renault, Claudia Saavedra, Gilles Stoltz, Tristan Tomala, Marie-Anne Valfort, and Guillaume Vigerat. Thanks to them, the experiment was successful and its expense limited to the costs of ballots, envelopes, and posters.