Social Choice and Individual Values

THIRD EDITION

Kenneth J. Arrow

Winner of the Nobel Memorial Prize in Economics

Foreword by Eric S. Maskin

"This book . . . demands the attention of everyone concerned with economic policy."—*Journal of Political Economy*

THIRD EDITION

Social Choice and Individual Values

Kenneth J. Arrow

Foreword by Eric S. Maskin

Yale

UNIVERSITY PRESS

New Haven and London

Printed in the United States of America.

Library of Congress Control Number: 2011945943
ISBN 978-0-300-17931-6 (pbk.)

A catalogue record for this book is available from the British Library.

10 9 8 7 6 5 4 3 2 1

COWLES FOUNDATION

for Research in Economics at Yale University

Monograph 12

COWLES FOUNDATION

For Research in Economics at Yale University

The Cowles Foundation for Research in Economics at Yale University, established as an activity of the Department of Economics in 1955, has as its purpose the conduct and encouragement of research in economics and related social sciences with particular emphasis on the development and application of logical, mathematical, and statistical methods of analysis. The professional research staff have, as a rule, a departmental appointment and some teaching responsibility.

The Cowles Foundation continues the work of the Cowles Commission for Research in Economics founded in 1932 by Alfred Cowles at Colorado Springs, Colorado. The Commission moved to Chicago in 1939 and was affiliated with the University of Chicago until 1955. In 1955 the professional research staff of the Commission accepted appointments at Yale and, along with other members of the Yale Department of Economics, formed the research staff of the newly established Cowles Foundation.

A list of Cowles Foundation Monographs appears at the end of this volume.

FOREWORD TO THE THIRD EDITION

Eric S. Maskin

Kenneth Arrow is a towering figure in economics and the social sciences more generally. And so I felt extraordinarily honored when Yale University Press asked me to write a foreword to a new edition of his landmark monograph *Social Choice and Individual Values*. Ken also happens to be my former teacher and a mentor and role model—facts that made the invitation a great personal pleasure as well, and an excuse for some nostalgia.

I first met Ken when I was an undergraduate at Harvard in the early 1970s. I was a math major but somehow wandered into his graduate course on information economics. It was a hodgepodge of some of the cutting-edge topics he was thinking about then—e.g., mechanism design, adverse selection, and communication in organizations—and proved to be utterly fascinating. I can't say that Ken—despite his interest in the theory of organizations—was the most organized of teachers. His lectures sounded improvised, largely because they were: he apparently decided what to talk about on his way over to the classroom (and sometimes not even then). On one unusual occasion, he prepared a lecture in advance—on a highly technical result called the Gibbard-Satterthwaite theorem (not yet published at the time)—but then forgot to bring his notes. He worked out a new and detailed proof of the theorem for us on the spot.

One feature that made the course so gripping was Ken's ability to pack so much into so little time. This was partly because he talks unusually fast. But even Ken's rapid-fire speaking style couldn't keep pace with his lightning-quick mind. So his presentations had an elliptical quality—he would leave out the ends of sentences so that he could race ahead to the next thought. His listeners would have to stay alert to fill in the missing words.

Another thing that struck us students was the breadth of Ken's knowledge, which extended far beyond economics. He would be lecturing on some technical point when an apt quotation from Maimonides or an analogy from thermodynamics would occur to him. Ken evidently knew more on virtually any subject than any of

the rest of us. There was a story at Harvard that a group of junior faculty once concocted a plan by which they could finally appear to outshine their erudite senior colleague. They read up on the most arcane topic they could think of: the breeding habits of gray whales. On the appointed day they gathered in the coffee room and waited for Ken to come in. Then they started talking about the elaborate theory of a marine biologist named Turner on how gray whales find their way back to the same breeding spot year after year. Ken was silent . . . they had him at last! With a delicious sense of triumph, they continued to discuss Turner, while Ken looked increasingly perplexed. Finally, he couldn't hold back: "But I thought Turner's theory was discredited by Spenser, who showed that the supposed homing mechanism couldn't possibly work."

With its publication in 1951, *Social Choice and Individual Values* initiated the modern theory of social choice, the study of how a society should choose among its various options based on the preferences of the individual members of society. There had been sporadic literature on the subject before Arrow, going back (at least) to Jean-Charles Borda and the Marquis de Condorcet in the late eighteenth century. But the earlier essays lacked the generality and power of Arrow's approach, and the subject did not really take off until *Social Choice*. But take off it did: by the time the second edition was published, in 1963, there were already several hundred works building on the book. A recent count on Google Scholar turned up over ten thousand citations.

Let me try to explain why the monograph has turned out to be so immensely influential—of interest to political scientists, sociologists, lawyers, and philosophers as well as economists. First, Arrow's abstract formulation of the social choice problem makes it very widely applicable. He begins with a society and a set of social alternatives (the different possible options from which society must choose), which, depending on the context, could be almost anything. For example, in a setting where a town is considering whether or not to build a bridge across the local river, "society" comprises the citizens of the town, and the social alternative set consists of just two options: "build the bridge" or "don't build it." In a context of pure distribution where there is a jug of milk and a plate of cookies to be divided among a group of children, the children are the society, and the social alternative set consists of the different ways the milk and cookies could be allocated to them. In a setting where a committee is interested in electing a chairman, society is the committee, and the social alternatives are the various candidates for the chairmanship.

Arrow's definition of a *social welfare function* (SWF) is also very general. An SWF is any rule for determining *society's* preferences

over the social alternative set on the basis of the preferences of the *individual members*. More precisely, because individuals' preferences might not be known in advance, the SWF is a function: it must determine social preferences for every different configuration of preferences that individuals could have, i.e., for every *profile* of possible preferences.

The most famous example of an SWF is probably *majority rule*, which Condorcet himself particularly advocated for elections. Under majority rule, alternative a is socially preferred to b for a given profile if a majority of individuals prefer a to b for that profile.

A second reason for the great impact of Arrow's monograph is the powerful and unexpected Impossibility Theorem that constitutes its central finding. Arrow discovered that there is *no* SWF that satisfies all of a few natural and seemingly undemanding conditions. These conditions are:

Unrestricted Domain (U): the SWF must determine social preferences for *all logically possible* profiles; in other words, there are no limitations on the preferences that individuals might have.

Pareto Property (P): if all individuals prefer alternative a to b, then a must be socially preferred to b.

Independence of Irrelevant Alternatives (IIA): if there are two profiles and each individual ranks alternatives a and b the same way in both of them, then the social preference between a and b must also be the same for both. In other words, the social preferences between a and b depend only on individuals' preferences between *these two* alternatives—and not on preferences involving some third alternative.

Nondictatorship (ND): there does not exist a member of society who always gets his way, in the sense that, for any profile, the social preferences coincide with his preferences.

Transitivity (T): for any profile, if a is socially preferred to b and b is socially preferred to c, then a must be socially preferred to c.

The Impossibility Theorem establishes that if there are at least three alternatives in the set of social alternatives, then there is no SWF that satisfies all of U, P, IIA, ND, and T.

It is worth noting why majority rule is not a counterexample to the theorem. As Condorcet himself pointed out, majority rule violates T. Suppose there are three alternatives a, b, and c, and consider a profile in which 35 percent of individuals prefer a to b and b to c, 33 percent

prefer b to c and c to a, and the remaining 32 percent prefer c to a and a to b. Then, 67 percent of individuals prefer a to b, and so a is socially preferred to b. Similarly, b is socially preferred to c (because 68 percent of individuals prefer b). But c is socially preferred to a (65 percent prefer c), and so T is violated.

Much of the literature stemming from *Social Choice* amounts to trying to avoid the Impossibility Theorem by relaxing one or more of the Arrow conditions. Condition ND is already so mild that relaxing it further seems pointless. The same might be said for condition P (but there is interesting work in this direction by Robert Wilson). By contrast, condition T is judged by some (see, in particular, James Buchanan) to be too strong a requirement. Nevertheless, relaxations of T turn out not to take us very far away from impossibility.

But relaxing condition U—which amounts to restricting the domain of preferences that an individual might have—has proved very fruitful. Even before Arrow, Duncan Black showed that in some applications, a natural restriction on preferences ensures that majority rule will satisfy T (in other words, it rules out the sort of T-violating example I gave above). In subsequent work, Amartya Sen characterized all possible restrictions under which majority rule is transitive, while Ehud Kalai and Eitan Muller and, independently, I characterized all possible restrictions under which there exists *some* SWF satisfying Arrow's remaining conditions. In recent work, Partha Dasgupta and I have shown that there is a sense in which majority rule satisfies Arrow's conditions (somewhat strengthened) for more domains of preferences than any other SWF.

Weakenings of IIA—the most controversial of Arrow's conditions—have spawned the biggest literature of all. As formulated, IIA rules out interpersonal comparisons: if in a two-member society, individual 1 prefers a to b and individual 2 prefers b to a in each of two profiles, then IIA requires that the social preferences between a and b be the *same* for both profiles—despite the possibility that in one profile, individual 1 has a strong preference for a over b and 2 only slightly prefers b to a, and the opposite is true for the other profile. Hence, in the extensive literature on interpersonal comparisons, IIA is weakened so that differences in preference intensity across individuals can be reflected in the SWF (although there are serious difficulties with trying to take account of such intensity differences when conducting elections). Recently, there has also been work developing other important relaxations of IIA, notably that of Marc Fleurbaey and François Maniquet, and also of Michel Balinski and Rida Laraki.

IIA also lies behind my final explanation for *Social Choice*'s great sway: the condition is intimately connected with the vast literature on mechanism design theory. Mechanism design was developed

to overcome an important obstacle to social choice, namely, that individuals' preferences—the inputs to the SWF—may not be publicly known. An SWF is said to be *implementable* if it is possible to design a *mechanism*—a procedure or game for individuals to follow—that leads to the choice of the top-ranked alternative according to the SWF (despite this lack of public knowledge). Arrow's monograph was a crucial foundation for mechanism design, because it turned out that conditions closely related to IIA are the key to an SWF's being implementable.

A book's importance can be crudely gauged by how many other works cite it. But perhaps a better measure is its longevity: how long it continues to inspire new work. By that criterion, *Social Choice and Individual Values* is an amazing success: having passed its sixtieth birthday, it continues to generate a steady stream of original research. I suspect that the same will be true when it reaches one hundred.

To Selma, my wife

PREFACE TO SECOND EDITION

The literature on the theory of social choice has grown considerably beyond the few items in existence at the time the first edition of this book appeared in 1951. Some of the new literature has dealt with the technical, mathematical aspects, more with the interpretive. My own thinking has also evolved somewhat, although I remain far from satisfied with present formulations. The exhaustion of the first edition provides a convenient time for a selective and personal stocktaking in the form of an appended commentary entitled, "Notes on the Theory of Social Choice, 1963," containing reflections on the text and its omissions and on some of the more recent literature. This form has seemed more appropriate than a revision of the original text, which has to some extent acquired a life of its own.

K. J. A.

Tokyo, Japan
August 1963

ACKNOWLEDGMENTS

First Edition

This study was initiated in the summer of 1948 while I was on leave from the Cowles Commission as a consultant to The RAND Corporation, which is engaged in research under contract with the United States Air Force. It was further developed and assumed its present form at the Cowles Commission during the period of October, 1948, to June, 1949, as part of the general research program of the Commission which is conducted under a grant from the Rockefeller Foundation. During part of this period support was also received from The RAND Corporation under a contract between RAND and the Cowles Commission for the study of resource allocation. To these organizations I wish to express my appreciation for the interest shown in this study and for the facilities accorded to me.

I wish to express my indebtedness to the following individuals at RAND: A. Kaplan, University of California at Los Angeles, and J. W. T. Youngs, University of Indiana, for guidance in formulating the problem, and D. Blackwell, Howard University, and O. Helmer for other helpful discussions. The manuscript has been read by A. Bergson and A. G. Hart, Columbia University, and by T. C. Koopmans, Cowles Commission for Research in Economics and the University of Chicago, and I owe much, both in improvement of presentation and in clarification of meaning, to their comments. Development of the economic implications of the mathematical results has been aided by the comments of F. Modigliani, Cowles Commission and the University of Illinois, T. W. Schultz, University of Chicago, and H. Simon, Cowles Commission and Carnegie Institute of Technology. I have had the benefit of comments by J. Marschak, Cowles Commission and the University of Chicago, on the question of the measurability of utility, touched on in Chapter II. The section in Chapter VII on the decision process as a value in itself has benefited from suggestions by P. J. Bjerve, Central Statistical Bureau, Oslo, Norway, then a guest of the Cowles Commission, and M. Friedman, University of Chicago. For guidance in the unfamiliar realms of political philosophy, I must thank D. Easton, University of Chicago. The mathematical exposition has been considerably improved as a result of comments by T. W. Anderson, Cowles Commission and Columbia University, E. Nagel, also of Columbia University, J. C. C. McKinsey, The RAND Corporation, and J. W. T.

Youngs. I must also mention the stimulation afforded by several staff meetings of the Cowles Commission in which my thesis was submitted to the research group for discussion and criticism. Needless to say, any error or opacity remaining is my responsibility.

I cannot fully acknowledge here the great debt I owe to my many teachers, but I cannot refrain from singling out H. Hotelling, now of the University of North Carolina, to whom I owe my interest in economics and particularly my interest in the problems of social welfare.

Acknowledgment is also due Mrs. Jane Novick, Editorial Secretary of the Cowles Commission, for preparing the manuscript for publication and seeing it through the press and to Miss Jean Curtis, Editorial Assistant, for valuable aid in proofreading and preparation of the index.

SECOND EDITION

I wish to thank Miss Laura Staggers, of the Stanford Institute for Mathematical Studies in the Social Sciences, for her patient typing and retyping of the manuscript for Chapter VIII.

K. J. A.

CONTENTS

Chapter I

INTRODUCTION

1. The Types of Social Choice

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 emerging democracies with mixed economic systems, Great Britain, France, and Scandinavia, the same two modes of making social choices prevail, though more scope is given to the method of voting and decisions based directly or indirectly on it and less to the rule of the price mechanism. Elsewhere in the world, and even in smaller social units within the democracies, social decisions are sometimes made by single individuals or small groups and sometimes (more and more rarely in this modern world) by a widely encompassing set of traditional rules for making the social choice in any given situation, e.g., a religious code.[1]

[1] The last two methods of making social choices are in a sense extreme opposites, developments of conflicting tendencies in a democracy. The rule of the single individual is the extreme of administrative discretion, the rule of a sacred code the extreme of rule by law. But in dynamic situations the rule of a sacred code leads by insensible steps to dictatorship. The code needs interpretation, for conditions change, and, no matter how explicit the code may have been in the first place in determining how society shall act in different circumstances, its meaning becomes ambiguous with the passage of time. It might conceivably happen that the job of interpretation passes to society as a whole, acting through some democratic process—"vox populi, vox dei." Or it can happen that interpretation passes to the hands of the people individually and not collectively; in this case, as soon as differences of opinion arise, the religious code loses all its force as a guide to social action. See, for example, the ultimate consequences in the field of economic ethics of the Protestant insistence on the right of each individual to interpret the Bible himself (R. H. Tawney, *Religion and the Rise of Capitalism*, London: J. Murray, 1926, pp. 97–100). But more likely, in view of the authoritarian character of the sacred code, the interpretation will pass into the hands of a single individual or a small group alone deemed qualified.

The classification of methods of social choice given here corresponds to Professor Knight's distinction among custom, authority, and consensus, except that I have subdivided consensus into the two categories of voting and the market (F. H. Knight, "Human Nature and World Democracy," in *Freedom and Reform*, New York: Harper and Bros., 1947, pp. 308–310).

The last two methods of social choice, dictatorship and convention, have in their formal structure a certain definiteness absent from voting or the market mechanism. In ideal dictatorship there is but one will involved in choice, in an ideal society ruled by convention there is but the divine will or perhaps, by assumption, a common will of all individuals concerning social decisions, so in either case no conflict of individual wills is involved.[2] The methods of voting and the market, on the other hand, are methods of amalgamating the tastes of many individuals in the making of social choices. The methods of dictatorship and convention are, or can be, rational in the sense that 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?

It should be emphasized here that the present study is concerned only with the formal aspects of the above question. That is, we ask if it is formally possible to construct a procedure for passing from a set of known individual tastes to a pattern of social decision-making, the procedure in question being required to satisfy certain natural conditions. An illustration of the problem is the following well-known "paradox of voting." Suppose there is a community consisting of three voters, and this community must choose among three alternative modes of social action (e.g., disarmament, cold war, or hot war). It is expected that choices of this type have to be made repeatedly, but sometimes not all of the three alternatives will be available. In analogy with the usual utility analysis of the individual consumer under conditions of constant wants and variable price-income situations, rational behavior on the part of the community would mean that the community orders the three alternatives according to its collective preferences once for all, and then chooses in any given case that alternative among those actually available which stands highest on this list. A natural way of arriving at the collective preference scale would be to say that one alternative is preferred to another if a majority of the community prefer the first

[2] It is assumed, of course, that the dictator, like the usual economic man, can always make a decision when confronted with a range of alternatives and that he will make the same decision each time he is faced with the same range of alternatives. The ability to make consistent decisions is one of the symptoms of an integrated personality. When we pass to social decision methods involving many individuals (voting or the market), the problem of arriving at consistent decisions might analogously be referred to as that of the existence of an integrated society. Whether or not this psychiatric analogy is useful remains to be seen. The formal existence of methods of aggregating individual choices, the problem posed in this study, is certainly a necessary condition for an integrated society in the above sense; but whether the existence of such methods is sufficient or even forms an important part of the sufficient condition for integration is dubious.