Advanced Textbooks in Economics
Editors C. J. BLISS and M. D. INTRILIGATOR

The Strategy of Social Choice

H. Moulin

North-Holland

THE STRATEGY OF SOCIAL CHOICE

H. MOULIN

Université Paris IX – Dauphine and
Laboratoire d'Econométrie de l'Ecole Polytechnique

1983

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

ISBN: 0444 86371 0

Publishers

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Sole distributors for the U.S.A. and Canada

ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 VANDERBILT AVENUE
NEW YORK, N.Y. 10017

PRINTED IN THE NETHERLANDS

INTRODUCTION TO THE SERIES

The aim of the series is to cover topics in economics, mathematical economics and econometrics, at a level suitable for graduate students or final year undergraduates specializing in economics. There is at any time much material that has become well established in journal papers and discussion series which still awaits a clear, self-contained treatment that can easily be mastered by students without considerable preparation or extra reading. Leading specialists will be invited to contribute volumes to fill such gaps. Primary emphasis will be placed on clarity, comprehensive coverage of sensibly defined areas, and insight into fundamentals, but original ideas will not be excluded. Certain volumes will therefore add to existing knowledge, while others will serve as a means of communicating both known and new ideas in a way that will inspire and attract students not already familiar with the subject matter concerned.

The Editors

PREFACE

The theory of voting attempts to classify actual voting methods with respect to their ethical and strategical features. It provides theoretical foundations to the neoclassical approach to democracy, a rapidly developing field which we refer to as "public choice" and to welfare economics in general. It is intimately related to the incentive problem, a main issue in current economic theory.

While textbooks on social choice (Sen, Fishburn) on the one hand, or game theory (Owen, Case) on the other hand are available, only the pioneering book of Farquharson addresses the conceptual problem of strategic voting by itself. Most of the fundamental mathematical results (e.g. the Gibbard–Satterthwaite theorem) emerged only in the seventies, some of them in the late seventies (e.g. Maskin's implementation concept). At that time a technical unification took place: through the concepts of strong monotonicity and effectivity functions most of the statements that the theory had produced were rearranged consistently from Arrow's impossibility to implementation by strong equilibrium.

This volume is entirely self-contained and uses only elementary mathematical techniques: there are finitely many candidates and voters, and only ordinal preferences enter the picture. However, the conceptual difficulty of some of the definitions, and the mathematical complexity of some of the proofs, is fairly high. The economically oriented reader will find a long introduction relating the proposed results to the literature, especially to welfare economics and the economics of incentives. The mathematically minded reader will find nearly 30 problems, some of them difficult, to check his or her understanding of the theorems.

The book seems most appropriate for use in a course on social choice theory and/or public choice for first- or second-year graduate students. It could also be used in any course on game theory.

CONTENTS

CHAPTER 1

INTRODUCTION

1. Prescriptive judgements or descriptive analysis

Social choice theory provides theoretical foundations to the neoclassical approach to democracy, a rapidly developing field which we refer to as "public choice" (see Mueller, 1979), and to welfare economics in general. Its major concern, pioneered by the eighteenth-century political philosophers, is normative decision-taking: several agents have to decide on some issue of collective interest whereas their opinions (preferences) about the issue might differ. Despite their conflicting interests, they must agree on one particular "final" decision. A "social" choice is any proposed solution of this problem, where society is taken to be the group of agents concerned by the issue, no matter how small.

To solve the possibly very sharp opposition of contradictory opinions, social choice theory takes an axiomatic route: it explores systematically all conceivable rules by which all potential conflicts are each given one particular resolution. Such rules are called social choice functions (or correspondences). They are the basic ingredient of the theory and their mathematical study its main activity.

The relevance of this approach to collective decision crucially depends on the agents' willingness and ability to bargain. Let us take an archetypical example. Given a committee, i.e. a fixed set of agents, who must together elect one among a well-defined set of candidates (these candidates being physical persons or any political or economic issue), we may think of two opposite patterns of collective behaviour. At one extreme the committee will go to an omniscient godfather, someone that every agent sees as wise and just, tell him in full detail how the various interests conflict and let him decide sovereignly. This is the ideal context where pure social choice theory operates: accordingly, it has much to say on virtually every prescriptive judgement arising in a political (e.g. what apportionment method is more fair, more faithful to the "one man one vote" doctrine?) or economic context (what is the just consumption level of a given public good? what compensatory side-payments are just among heirs sharing indivisible goods?).

At the other extreme no consensus exists on a supra authority that could settle the matter and the agents will bargain until some issue is agreed upon. Whereas in the social choice approach the decision power was indivisible into the hands of some legal or moral emanation from the collectivity (a central planner, a judge, or a referee), here the power flows pervasively and changingly among agents and coalitions of agents. Under this jungle-style decision process, the only rule is no rule and very little "social" behaviour is likely to emerge. Game theory at large addresses itself to the formalization of patterns of behaviour arising in such environments. There it provides powerful descriptive tools to analyse the subtle mixture of struggle and cooperation inherent in any social interaction.

These two approaches possess similar defects: a host of different possibly inconsistent, prescriptive judgements make it difficult for a well-meaning referee to arbitrate among arbitration methods, just as the great many game-theoretic equilibrium concepts kill any hope of making predictive use of the theory.

2. Normative versus positive approach to voting

Modern modelization of collective decision-making always involves both normative features (prescriptive value judgements as represented by social choice mappings) and positive features (strategic behaviour as represented by game-theoretic equilibrium concepts). This point is perhaps best illustrated by the literature on voting methods, since the corresponding formalization entails a minimum of a priori assumptions (as opposed to, say, the prerequisites of microeconomics) and the conceptual added value is therefore more transparent.

A voting method is a constitution attached to a given set of agents and a given set of candidates (or outcomes, or issues). Viewed as a social choice mapping this constitution scrutinizes every conceivable opinion by every individual agent over the set of candidates and selects for each particular profile (i.e. every particular choice of an opinion for each agent) a subset of "good" candidates (ideally a singleton). Viewed as a "game form" it allocates the decision power among individual agents by assigning to each agent a message space (a strategy space) from which he or she sovereignly selects one. Then, for every particular choice of one message by each agent, the game form determines the "winning" candidate.

The key observation is that every democratic voting method must simultaneously be thought of as a social choice mapping and as a game form.

Democracy, in its neoclassical acceptation, means that the goals of collective action must rely on the opinion of individuals and depend on these opinions only. Accordingly, the only relevant information for collective decision is the profile itself: as soon as an agent realizes that through his or her message he or she influences the final decision, he or she will use it strategically, i.e. manipulate the decision-making mechanism for the sake of his or her own interest. At this point freedom of speech means a right to lie, and taking into account the quoted manipulations amounts to recognizing that the agent will send a sincere message only if it is selfishly rational to do so.

3. The implementation problem

To say that the agents involved in any voting method realize that they actually face a game, the strategies of which are their own messages, induces a genuine complexity of the analysis. This is the same as saying that when firms realize that they do influence the price we switch from a perfect competition framework to oligopoly analysis. In both cases the conceptual difficulty results from the potential variety of strategic behaviour by the agents: if anything, game theory tells us that no straightforward equilibrium concept exists to describe this behaviour.

The concept of implementation proves to be central for describing the relationship between the normative (in terms of equity or efficiency axioms) and the positive (in terms of how utility-maximizing agents can be expected to distort their messages) properties of voting methods. The feasibility problem is to ask whether or not a particular ethic (represented by a social choice mapping) can be materialized in a world of selfish individual agents (i.e. the decision power can be decentralized via a particular game form) according to some pattern of behaviour (i.e. some equilibrium concept). At this level of generality the implementation problem bears upon the essence of social decision-making: given that society views as desirable certain ethics of collective decision, is it possible, and if so how, to decentralize the decision power among individual agents in such a way that by freely exercising this decision power the agents eventually select the very outcome(s) recommended as a priori desirable? Only a limited number of economic and political issues have been explored so far along this line (see section 7 below). However, the relevance of this approach to economic and political decision-making is self-evident.

Each game-theoretical equilibrium concept gives rise to a different implementation notion. Specifically, we explore the implementation problem for two non-cooperative concepts: i.e. dominant strategy equilibrium and sophisticated perfect equilibrium, and two cooperative concepts: i.e. strong equilibrium and Nash equilibrium (which in our context involve a fair amount of coalitional agreement). Non-cooperative behaviour allows us to implement social choice functions (i.e. a deterministic single candidate is chosen for all profiles) whereas cooperative behaviour implements only social choice correspondences (i.e. the choice set contains several outcomes for one profile).

4. Classifying voting methods

As the main output of the implementation technology, we will prove many theorems, some of them difficult. Altogether we classify, with respect to their main strategic features, familiar (Condorcet, Borda) and less familiar (voting by veto) voting methods. Although one can conceive infinitely many, three main families clearly emerge.

First, the scoring methods translate the utilitarian view into a voting method. Assume that every agent has a fixed scale utility, and the various utility levels (which might differ from one agent to the other) are common knowledge. Then an agent's message is simply an ordering of the various candidates and those candidates which score best on total (collective) utility are elected. Two typical examples are the Borda voting method, where the utility levels are common to each agent and in arithmetical progression, and plurality voting, where each agent casts a vote for his or her (supposedly) top candidate and the candidate(s) with the maximal number of votes are elected. These methods are characterized by a very appealing ethical property: Young's consistency property (see section 5 below). On the other hand, they give rise to wild strategic manipulations, both by non-cooperative and cooperative agents.

A second family of voting methods follows from the Condorcet majority principle: if a candidate happens to beat every other candidate in pairwise contests, in which case we call it a Condorcet winner, then it should be unconditionally elected. If a Condorcet winner exists, then the social choice that it defines is robust against any (individual and/or coalitional) manipulations. If no Condorcet winner exists (a situation often quoted as the Condorcet paradox), then cooperative instability occurs. The family of

voting by binary choices does implement the Condorcet majority principle: the final outcome is reached by a predetermined sequence of binary votes, each one taken by majority vote. These methods implement a well-defined social choice function when players behave non-cooperatively (and this social choice function picks the Condorcet winner if there is one) but are highly vulnerable to coalition formation when the Condorcet paradox occurs.

The third family is the voting by veto method: the agents are endowed with a certain number of veto tokens (say one token per agent when the number of candidates exceeds the number of voters by exactly one) and they successively exercise their veto rights, ultimately leaving just one candidate on the floor, which is then elected.

These voting rules illustrate the minority principle, which claims that any minority should be given the right to veto as many candidates as is compatible with the feasibility of the decision process. This principle prevents any minority from being crushed by the opposing majority, as should indeed be the case according to the majority principle. The voting by veto methods are the natural constitutions that overcome the cooperative instability following from the Condorcet paradox. They yield a non-empty core for every possible profile and the outcome of non-cooperative voting behaviour always belongs to this core. The consistency of the cooperative and non-cooperative voting behaviour is a unique feature of voting by veto methods. Therefore the social choice correspondences that these methods implement can be justified both on normative and positive grounds as a reasonable answer to the social choice problem, especially when the electorate is small and cooperation among the voters plausible.

The most remarkable feature of the game-oriented attack of the social choice problem is the positive character of the analysis: our important results are not impossibility theorems (stating that such a list of strategic and ethical properties are inconsistent – see Arrow's famous impossibility result as well as the Gibbard–Satterthwaite theorem). On the contrary, we associate to each behavioural scenario (formalized as a specific equilibrium concept) several families of voting methods where this scenario is likely to yield a socially satisfactory outcome. These voting methods allow an egalitarian distribution of decision power among the individual agents, thus ruling out dictatorial trivialities.

The next three sections provide the historical and bibliographical information necessary to relate this course with the social choice and economics literature.

5. Voting and the non-strategic theory of social choice

During the nineteenth century various authors, including C.L. Dodgson, alias Lewis Carroll, proposed and discussed the ethical properties of several voting methods. There strategical considerations are virtually absent, except as a moral pollution of the decision process: "principles of voting make an election more of a game of skill than a real test of the wishes of the electors. My own opinion is that it is better for elections to be decided according to the wish of the majority than of those who happen to have most skill at the game" (C.L. Dodgson quoted by Farquharson, 1969).

Contemporary voting theory brought to an end the controversy first opened during the late eighteenth century by Condorcet (1785) and Borda (1781). For that purpose an axiomatic characterization of both the Condorcet majority principle and the (utilitarianist) scoring functions was built up.

May (1952) proposed a set of conditions that characterize the binary majority relation, namely aRb, iff a is preferred to b by at least as many agents as b is preferred to a. Since the binary majority relation is not in general transitive (owing to the Condorcet paradox), two routes are open to convert it to a decision rule. The first route is to pick a social choice function or correspondence that always selects a Condorcet winner if there is one, and some natural ersatz if there isn't one. Several methods were suggested, which we review in Chapter 2, section 5. Another route is to approximate the binary majority relation by a "close" ordering, thus achieving a social welfare function that aggregates a profile into a collective preference ordering. This route was taken by Kemeny (1959), Kemeny and Snell (1960), and later systematically explored by Young and Levenglick (1978). For a comprehensive survey of this approach see Barthelemy and Monjardet (1980).

A fairly simple characterization of the scoring functions was obtained by Young (1974, 1975) by considering a variable sized electorate body and using a generalized unanimity property: "if two committees meeting separately arrive at the same consensus ordering, then meeting together this should still be their consensus". Then a social choice correspondence is a scoring function if and only if it is anonymous, neutral and consistent.

The growing concern of economists for the social choice problem was initiated by Arrow's seminal book (Arrow, 1963) who formalized first the Benthamite approach to collective welfare as an aggregation of preference operators. Arrow's famous impossibility theorem points out the logical inconsistency of the Benthamite approach with the ordinalist requirement

that only preference ordering should matter. In our framework we shall derive Arrow's theorem from the (technically more profound) strong monotonicity of social choice correspondences: no (deterministic) social choice function exists which is strongly monotonic and non-dictatorial (theorem 1, Chapter 3). Since the inclusion of minimal strongly monotonic s.c.c. (i.e. the more deterministic strongly monotonic s.c.c.) play a central role in cooperative voting (see Chapter 7) we regard implementation theory as the synthetic viewpoint that embodies a large number of possibility results and incorporates the traditional impossibility results as interesting particular cases, not as disturbing mathematical dead-ends. This is how Arrow himself invites us to deal with his result: "The philosophical and distributive implications of the paradox of social choice are still not clear. Certainly there is no simple way out. I hope that others will take this paradox as a challenge rather than a discouraging barrier" (Arrow, 1972).

Most of the developments of non-strategic social choice theory are to be found in the books by Sen (1970), Fishburn (1973), Pattanaik (1971) and Kelly (1979).

6. Development of the strategic theory of voting

It is self-evident that voting methods are relevant to virtually every issue within the field of public choice: taxes, public services, as well as policy orientations can be decided by vote. The economic viewpoint at these non-marketed decision problems is to regard them as games where individuals pursue their own interest within the mutual dependency pattern imposed by the decision rule.

Immediately following the massively influential book by Von Neumann and Morgenstern (1944), the pioneers of social choice theory were quite aware of the conflicting implications of any collective decision rule. Most clearly, Black (1958) and Guilbaud (1952) took into account the strategic behaviour of the electoral body. Their formal analysis of this phenomenon remains elementary, however, just as game theory at the time did not provide the necessary equilibrium concepts to analyse the subtle mixture of conflictual and cooperative behaviour generated by most decision mechanisms. Not surprisingly, then, Arrow's seminal axiomatization of the social choice problem does not include any explicit game-theoretic considerations. Next, from the early fifties to the late sixties the whole development and diversification of social choice theory ignores the manipulation problem.

Farquharson's book (1969) can be approximately taken as the starting point of the game-oriented social choice literature (see also Dummett and Farquharson, 1961) and the Gibbard–Satterthwaite theorem (see Gibbard, 1973; Satterthwaite, 1973, 1975) as its seminal result. The theorem states that when at least three candidates are to be compared, the only strategy-proof social choice functions are the dictatorial ones. In other words, no non-dictatorial voting method exists where the secret ballot paradigm applies: a voter's sincere message cannot be his or her optimal strategy no matter how the other agents vote. If the vote is actually taken by secret ballot, an incentive does exist for some agents in some profiles to acquire the information about other agents' opinions: polls are a strategical necessity.

To escape the Gibbard–Satterthwaite impossibility result, two lines of research have been investigated: one restricts the domain of feasible preferences, the other changes the equilibrium concept.

6.1. Restricted domains of preferences

Black (1958) had already noticed that when the candidates are linearly ordered and the agents' preferences are single-peaked with respect to that ordering, then a Condorcet winner exists and yields a strategy-proof voting rule.

Next Inada (1969) and Sen and Pattanaik (1969) characterized the restricted domains guaranteeing a transitive majority relation for any number of agents. By seeking restricted domains that guarantee only the existence of a Condorcet winner (the underlying majority relation being not necessarily transitive over the whole set of outcomes), many more satisfactory domains obtain. Typically the outcomes (candidates) form a tree configuration and the preferences are single-peaked with respect to that tree (Demange, 1980). This context is applicable to the location of public service problems (Hansen and Thisse, 1980).

Another generalization of Black's result consists of characterizing all non-manipulable voting rules when the preferences are single-peaked (see Moulin, 1980a, for the real line case, and Chichilnisky and Heal, 1981, for the multi-dimensional case). Chapter 4, section 5, is devoted to these various results.

Another typical restriction of the domain of preferences considers voting rules that allow some randomness in the final decision. Thus, the game form aggregates individual messages into a lottery over the candidates. Next the agents are endowed with a cardinal utility over (deterministic) candidates

and compare lotteries by means of their expected utility (which amounts to restricting the domain of feasible preferences). The social choice problem was first attacked within that framework by Intriligator (1973). In that context a random dictator mechanism, where an agent is chosen to be the dictator according to a fixed probability distribution, is a fairly trivial strategy-proof decision rule. Gibbard (1977, 1978), and then Hylland (1980) proved conversely that every strategy-proof voting method is essentially a random dictator mechanism under a mild attainability assumption (when all agents agree on the top candidate a, then a should be elected with probability one). This negative result severely narrows the interest of random voting mechanisms. Another approach to this result is proposed in Barbera and Sonnenschein (1977) and McLennan (1980). Without the attainability condition, Barbera (1977, 1979) derives nice strategy-proof probabilistic voting rules.

Assuming that the preferences of individual agents are dichotomous, i.e. each agent can partition the candidates in one subset of equally good candidates and one subset of equally bad candidates, Brams and Fishburn (1978) and Weber (1978) observe that approval voting is a strategy-proof voting method. An agent casts a subset of candidates – presumably the set of his or her "good" candidates – and those candidates with a maximal number of votes are elected. The strategy-proofness feature of approval voting – although only true on a highly restricted domain of preferences – makes it an appealing alternative to plurality voting.

A general, although hardly applicable, characterization of the restricted domains on which strategy-proof social choice functions as well as Arrow's social welfare function can be constructed, is proposed in Kalai and Muller (1977). See also Kalai and Ritz (1980).

Other restrictions of preferences allow us to overcome the Gibbard–Satterthwaite impossibility result: they derive from the microeconomic theory of preferences. Typically the set A of outcomes must be endowed with some additional structure – usually A is a subset of some euclidean space – and assumptions like separability, linearity and concavity can be made on preferences. We review the main results in that direction in section 7 below.

6.2. *Changing the equilibrium concept*

Since the strategy-proofness requirement is too demanding, the obvious way out of the impossibility is to weaken the equilibrium notion. This is what Farquharson (1969) actually did when he originally introduced the idea of