Social Choice and the Mathematics of Manipulation

Honesty in voting, it turns out, is not always the best policy. Indeed, in the early 1970s, Allan Gibbard and Mark Satterthwaite, building on the seminal work of Nobel Laureate Kenneth Arrow, proved that with three or more alternatives there is no reasonable voting system that is non-manipulable; voters will always have an opportunity to benefit by submitting a disingenuous ballot. The ensuing decades produced a number of theorems of striking mathematical naturality that dealt with the manipulability of voting systems. This book presents many of these results from the last quarter of the twentieth century – especially the contributions of economists and philosophers – from a mathematical point of view, with many new proofs. The presentation is almost completely self-contained and requires no prerequisites except a willingness to follow rigorous mathematical arguments.

ALAN D. TAYLOR is the Marie Louise Bailey Professor of Mathematics at Union College, where he has been since receiving his Ph.D. from Dartmouth College in 1975. His research interests have included logic and set theory, finite and infinitary combinatorics, simple games, and social choice theory. He is the author of *Mathematics and Politics: Strategy, Voting, Power, and Proof* and coauthor of *Fair Division: From Cake-Cutting to Dispute Resolution* and *The Win-Win Solution: Guaranteeing Fair Shares to Everybody* (both with Steven J. Brams) and *Simple Games: Desirability Relations, Trading, and Pseudoweightings* (with William S. Zwicker).

OUTLOOKS

PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS AND THE MATHEMATICAL ASSOCIATION OF AMERICA

Mathematical content is not confined to mathematics. Eugene Wigner noted the unreasonable effectiveness of mathematics in the physical sciences. Deep mathematical structures also exist in areas as diverse as genetics and art, finance and music. The discovery of these mathematical structures has in turn inspired new questions within pure mathematics.

In the *Outlooks* series, the interplay between mathematics and other disciplines is explored. Authors reveal mathematical content, limitations, and new questions arising from this interplay, providing a provocative and novel view for mathematicians, and for others an advertisement for the mathematical outlook.

Social Choice and the Mathematics of Manipulation

ALAN D. TAYLOR
Union College

Mathematical Association of America

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
40 West 20th Street, New York, NY 10011-4211, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521810524

MATHEMATICAL ASSOCIATION OF AMERICA
1529 Eighteenth St. NW, Washington, DC 20036, USA

First published 2005

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Taylor, Alan D., 1947–

Social choice and the mathematics of manipulation / Alan D. Taylor.
p. cm. – (Outlooks)

Includes bibliographical references and index.

ISBN 0-521-81052-3 (hardback : alk. paper) – ISBN 0-521-00883-2 (pbk. : alk. paper)

1. Voting – Mathematical models. 2. Social choice – Mathematical models.
3. Political Sciences – Mathematical models. 4. Game theory.
I. Title. II. Series.

JF1001. T39 2005
324.6′01′5193–dc22 2004054521

ISBN-13 978-0-521-81052-4 hardback
ISBN-10 0-521-81052-3 hardback

ISBN-13 978-0-521-00883-9 paperback
ISBN-10 0-521-00883-2 paperback

Contents

Preface

Were honesty always the best policy, this indeed might be a better world. But there seems to be a place for the little white lie, and there is certainly reason for many supporters of Ralph Nader in the state of Florida – as they watched Albert Gore concede the U.S. presidential election to George W. Bush on the evening of December 12, 2000 – to regret having cast sincere ballots, the result of which was a victory for their third choice (Bush) instead of their second choice (Gore).

We have nothing to say here about the little white lie. In this book, however, we collect much of what is known regarding a single fundamental question of obvious political importance and surprising mathematical naturality: In what election-theoretic contexts is honesty in voting the best policy?

For example, consider an election in which there are three or more candidates from which a unique winner must be chosen, and in which each voter casts a ballot that gives his or her ranking of the candidates from best to worst with no ties. Can one, in this situation, devise a voting procedure such that each candidate wins at least one hypothetical election and with which no voter can ever gain by unilaterally changing his or her ballot?

As stated, this turns out to be a trivial question. If there are n voters, then a moment's reflection reveals n such voting procedures, each obtained by fixing one of the voters and taking the winner to be his or her top-ranked candidate. Dismissing these – they are, after all, dictatorships – leaves the better question: Are there any others?

The answer, quite remarkably, is no. This is precisely what the Gibbard–Satterthwaite theorem of the early 1970s asserts, and it is this result that gives rise to most of what follows in this book. That theorem is related to – indeed, some would say equivalent to – the celebrated 1950 result known as Arrow's impossibility theorem.

If there is a weakness to the Gibbard–Satterthwaite theorem, it is the assumption that winners are unique. But if we drop the uniqueness of winners as an assumption, then there are voting systems that intuitively seem to be non-manipulable. For example, the voting procedure that declares everyone to be tied for the win regardless of the ballots (a very uninteresting example) or the one that takes as winners all candidates with at least one first-place vote (certainly a more interesting example).

Why do we speak of these two procedures as being only "intuitively" non-manipulable? The problem is that if a voter's preferences are given by a list, then it is not at all clear what it means to say that he or she prefers one *set* of candidates to another *set* of candidates. For example, if a voter ranks alternative a over alternative b over alternative c over alternative d, does he or she then prefer the set $\{a, d\}$ to $\{b, c\}$ or vice versa? It's certainly not obvious.

Thus, one of our objectives is to collect many of the definitions, theorems, and questions that arise when one asks about single-voter manipulability in election-theoretic contexts in which winners are not necessarily unique. Most of the results we present – whether in the concrete setting of the twenty voting rules that we introduce in Chapter 1 or the more abstract context of theorems like that of Gibbard and Satterthwaite – are organized around four kinds of manipulability that we call, from strongest to weakest, single-winner manipulability, weak-dominance manipulability, manipulability by optimists and pessimists, and expected-utility manipulation.

Our undertaking is interdisciplinary in the sense that it is, in large part, a mathematician's presentation of some major contributions that economists and philosophers have made to the field of political science. Thus, few of the results in this book originated with the author, but many of the proofs did. For example, we give unified proofs of three important manipulability results in three of the major voting-theoretic contexts: the Gibbard–Satterthwaite theorem in the context wherein the outcome of an election is a single winner (Chapter 3), the Duggan–Schwartz theorem in the context wherein the outcome of an election is a set of winners (Chapter 4), and the Barberá–Kelly theorem in the context wherein the outcome of an election is a choice function (Chapter 5).

There are virtually no prerequisites for reading this book, except that a certain degree of what is usually called mathematical maturity is required beginning with Chapter 3. The nine chapters in the book are organized into three parts, each consisting of three chapters. We comment on each part in turn.

Part I of the book is presented at a suitably accessible level for use in a number of undergraduate or graduate courses in mathematics, economics, and political science. In particular, Chapter 1 is an introduction to social choice theory that provides (i) an explicit discussion of the different contexts in which one works; (ii) something of the history of the field; (iii) accurate statements of Arrow's impossibility theorem for voting rules, social choice functions, and social welfare functions; and (iv) a wide range of examples of voting rules. Chapter 2 is an introduction to the notion of manipulability, largely in the context of twenty specific voting procedures, and Chapter 3 presents a careful proof of the Gibbard–Satterthwaite theorem and deals with the question of the extent to which it is equivalent to Arrow's theorem.

There are more than seventy-five exercises in Part I, and these range from routine verifications to additional development of the material in the chapter, with hints (or outlines) provided where needed. Each exercise is labeled with a "C" for computational, an "S" for short answer, or a "T" for theory. Hopefully these labels will be of some use, but all three terms are being used somewhat metaphorically.

Part II of the book begins with a treatment of manipulation in the contexts wherein the outcome of an election is a set of winners (Chapter 4) and a social choice function (Chapter 5). Each chapter contains a direct proof of the main result that mimics what was done with the Gibbard–Satterthwaite theorem in Chapter 3 by combining a new idea – down-monotonicity for singleton winners – with a number of classical ideas, most of which arose in later treatments of Arrow's impossibility theorem.

In Chapter 6, we move to the case of infinitely many voters, and we present the known ultrafilter versions of Arrow's theorem and the Gibbard–Satterthwaite theorem, the latter of which requires finitely many alternatives and coalitional non-manipulability. We also provide an extension to the case where there are infinitely many alternatives as well as infinitely many voters. The proofs in Chapter 6 are completely self-contained, though less pedestrian than the presentations in Part I.

Part III of the book contains a number of additional results in cases of both single winners (Chapter 7) and multiple winners (Chapter 8). Finally, in Chapter 9, we conclude the book with a brief treatment of some voting-theoretic situations in which ballots and election outcomes are different from those of Chapters 1–8.