The welfare consequences of strategic voting

De gevolgen van strategisch stemmen voor welvaart

Thesis

to obtain the degree of Doctor from the Erasmus University Rotterdam
by command of the Rector Magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board

The public defence shall be held on
10 October 2007 at 11.45 hrs

by

Aki Petteri Lehtinen

born in Raisio, Finland

Doctoral Committee

Promotors:

Prof.dr. J.J. Vromen and Prof.dr. I.U. Mäki

Other members:

Prof.dr. M.V.B.P.M. van Hees
Prof.dr. M.C.W. Janssen
Prof.dr. H.J. Nurmi
Prof.dr. K.M. Dowding
Dr. B. Visser

Contents

Preface

I never planned to study strategic voting. I just dislike not being able to finish a project I have once started. This thesis has come about because I have thought for a long time that there is something seriously wrong with the methodology of social choice and the theory of strategic voting. In particular, I have always been puzzled by the reluctance of theorists working in these fields to provide models in which voters' intensities of preference have a role, even though they clearly seem to affect voters' behaviour. However, I realised early on that scholars in these fields are used to discussing philosophical and methodological issues by way of constructing formal models. Merely pointing out that this particular behavioural assumption concerning preference intensity is unrealistic would not have been very convincing without an account of how changing it affects our view of strategic voting. I therefore began this project of 'philosophy of science in practice' about a decade ago. Despite the fact that three of the essays are written with an audience of economists in mind, I consider this work mainly a methodological one.

I have studied strategic voting because I believe I have something important to say about it, and because I find it theoretically intriguing. These motivations, viz. theoretical curiosity and stubbornness I have inherited from my father Pekka T. Lehtinen. I believe he has had the most important influence on my work, despite the fact that as a biologist he is unlikely to understand my topic very well. I thank him for having set the example for critical thinking and unflinching search for knowledge.

There are many people who have been helpful along the way. The philosophy departments at the University of Helsinki and Erasmus Institute for Philosophy and Economics (EIPE) have provided stimulating environments for intellectual development. I want to mention Tomi Kokkonen, Jani Raerinne, Tarja Knuuttila, Erika Mattila, Matti Sintonen, Janne Hiipakka, Hanne Ahonen and Pekka Mäkelä from the former and Emrah Aydinonat, Jorma Sappinen, Caterina Marchionni, Gűlbahar Tezel Pot, and Peter Marks from the latter even though all of them have not been directly involved with the development of this thesis. I had the opportunity to use supercomputers at the Finnish Centre for Scientific Computing. This research was made possible by financial aid from the Yrjö Jahnsson Foundation, EIPE, and the Academy of Finland.

The body of the thesis consists of a compilation of scientific publications. They are listed here:

- Chapter 1: Lehtinen, Aki (2006): "Signal extraction for simulated games with a large number of players", *Computational Statistics and Data Analysis*, vol. 50, pp. 2495-2507.
- Chapter 2: Lehtinen, Aki (2007): "The welfare consequences of strategic voting in two commonly used parliamentary agendas", *Theory and Decision*, vol. 63, no. 1, pp. 1-40.
- Chapter 3: Lehtinen, Aki (forthcoming): "The Borda rule is also intended for dishonest men", *Public Choice*.
- Chapter 4: Lehtinen, Aki & Kuorikoski, Jaakko (2007): "Unrealistic assumptions in rational choice theory", *Philosophy of the Social Sciences*, vol. 37, no. 2, pp. 115-138.

I continue to enjoy working with Jaakko as much as I did when we wrote the paper included in this collection. I would like to thank Markus Haavio and Ruurik Holm for helpful comments on the first and the second and Petri Ylikoski for the fourth essay. Olli Serimaa and Antti Nevanlinna helped me with computer programming for the second and the third essay. Hannu Nurmi and Martin van Hees have commented extensively on all the chapters, including the concluding fifth essay. Cecilia Therman and Joan Nordlund have checked the language of almost all essays. (You will easily find out which essays have not passed through their finishing touch.) My supervisors Jack Vromen and Uskali Mäki have really done their best given the difficulty of some of the topics in the thesis. They have both been intellectually extremely important to me, but paradoxically enough, their influence and helpfulness may be even more evident in my future intellectual endeavours.

PhDs are not produced without love. Thank you äiti, and thank you Cecilia.

Chapter 1

Introduction

An incomplete information framework for studying the welfare consequences of strategic voting

Arrow's theorem (1963) and the Gibbard-Satterthwaite theorem (Gibbard 1973, Satterthwaite 1975) have long been regarded as cornerstones of social choice theory. The former is often interpreted as demonstrating that there is no perfect voting rule. The latter states, roughly, that all voting rules are vulnerable to strategic misrepresentation of preferences: at least one individual voter always has an incentive to vote strategically, and no voting rule is *strategy-proof*. Under most voting rules strategic voting means giving one's vote to an alternative that is not considered the best. It has been taken for granted in the vast literature on strategy-proofness that strategic voting is to be avoided as far as possible. However, there are very few contributions that endeavour to evaluate its welfare consequences in an explicit and welfarist way. This is precisely the gap that this PhD thesis endeavours to fill.

Assume that there are three parties: Right, Center, and Left. Voter X prefers Right to Center and Center to Left, but he believes that Right has the least chance of winning. If he greatly prefers Right to Center and is almost indifferent between Center and Left, he is less likely to switch his vote from Right to Center than if he slightly prefers Right to Center but abhors Left. (Downs 1957, p. 49)

This intuitive reasoning behind a strategic vote is so familiar that any layman can recognise it. Downs expresses the idea that preference intensities for the choice alternatives, and beliefs concerning their chances of winning, are crucial. Given how obvious the logic of beliefs and intensities is, it may be surprising that models of strategic voting that do not take preference intensities and incomplete information into account continue to be accepted for publication (e.g., Felsenthal 1996, Saari 2003*b*).

Using mere preference orderings in a voting model implies making an assumption that is known not to correspond to the facts. To be sure, the reason for neglecting intensities in the current mainstream literature on social choice does not derive from the belief that they are behaviourally or normatively irrelevant. It is rather that they are considered relevant in both senses, but they have not been used because they have been considered epistemologically problematic, or because it has been too difficult to construct mathematical models in which their behavioural consequences can be studied.

Models of incomplete information that formulate conditions for strategic voting for individual voters have been around for decades. McKelvey and Ordeshook's (1972) model for the plurality rule was the first, Merrill (1981*a*, 1981*b*) formulated similar conditions for the Borda rule and approval voting, while Enelow (1981) and Ordeshook & Palfrey (1988) devised such models for amendment agendas.

This literature explicitly models preference intensities, but it has not been integrated with exercises in normative social choice in which voting rules are characterised and critically compared in terms of their properties. Most theories of normative social choice have thus been behaviourally simple: voters are typically assumed to record their preference ordering sincerely. The reason for this, I presume, is that formulating strategic-voting conditions for individual voters is not sufficient for characterising the consequences of strategic behaviour at the aggregate level. The models referred to above take preferences and beliefs as exogenously given. An account of deriving or determining voters' beliefs in a reasonable way is thus necessary. Black (1978) and Hoffman (1982) show how to compute so-called pivot probabilities, but these contributions do not show how to aggregate voters' decisions. What is needed is either equilibrium analysis or computer simulation.

I have chosen to work with computer simulations for two reasons. First, since the problem of multiple equilibria is ubiquitous in voting contexts, it is usually impossible to derive the beliefs from of a game-theoretical solution concept in a reasonable way. Secondly, since the methods for computing pivot probabilities are already very complex, building tractable analytical models of strategic voting has thus far been beyond the capacities of even the best game theorists.[1]

The main methodological contribution of this thesis is to provide an incomplete-information framework in which the welfare consequences of strategic voting can be systematically evaluated.

Desiderata for models that aim to normatively evaluate the consequences of strategic voting

Since preference intensities inevitably affect voting outcomes, they should *also* be taken into account in evaluating the welfare consequences of strategic voting. If and when voters *themselves* consider preference intensities important, they need to be taken into account both in modelling voting games and in evaluating the outcomes. Essays two and three are based on the idea of bringing preference intensities into a framework that unifies the positive assumptions con-

[1] See, however, Myerson & Weber (1993) and Myatt (2007).

cerning voter behaviour, and the normative evaluation of the outcomes through the notion of a utilitarian winner. In short, the idea is to simply see how taking intensities into account affects the normative evaluation of strategic voting and voting rules. Taking preference intensities into account in an evaluation of the welfare consequences of strategic voting implies using utilitarian criteria. Given that the intuition behind strategic voting so self-evidently relies on preference intensities, it is unlikely that nobody has happened to think of studying its intensity-related welfare consequences: it is rather that there has been an articulated methodological unwillingness to do so. The epistemological and philosophical problems concerning the sum-of-utilities criterion were the very starting point of social choice theory.

How should the desirability of strategic voting be evaluated and investigated? I will briefly describe the existing approaches, and then present my own proposal. The approach that most social choice theorists have used is the intuitive one of assuming that since strategic voting means 'misrepresenting' or 'lying' about one's true preference ordering, it should be ruled out as far as possible.

Kelly (1988, p. 103) provides a list of explicit arguments upon which intuitive judgments rely, and which is applicable to the strategy-proofness condition.[2]

1. Manipulation introduces an element of randomness into collective decisions.

2. Unequal manipulative skills may lead to the destruction of our efforts to design rules with an equal treatment of individuals.

3. Voters are led to waste resources in manipulation calculations.

4. We are led to try to reduce manipulation by others of us by concealing our preferences, thus reducing the flow of information that might help in collective decision-making.

5. Manipulation by representatives blurs their voting record and makes it difficult for us to determine if they are really representing our interests.

None of these arguments refers to how well individual preferences are satisfied when people vote strategically rather than sincerely. The arguments are not welfarist. It is clear that points 2, 3, 4 and 5 are more relevant to the voting behaviour of parliamentary representatives than to citizens' voting behaviour in mass elections. Therefore, in order to evaluate whether strategic voting is beneficial or harmful in parliaments or courts, we need some further knowledge about the relevance of the five arguments compared to a utilitarian evaluation of its consequences. In the case of mass elections, none of the arguments seems to be compelling. An increasing number of scholars have stated that they do not consider strategic voting morally questionable.[3] This may signal the fact that Satterthwaite's arguments are no longer fully accepted.

The voting models discussed in this thesis provide a framework for turning arguments 1, 2, and 4 into open research questions. This is a first step in for-

[2] These arguments were originally presented in Mark Satterthwaite's PhD dissertation: The Existence of a Strategy Proof Voting Procedure (University of Wisconsin, 1973). See Van Hees & Dowding (forthcoming) for a comprehensive discussion.

[3] See Kolm (1996), Buchanan & Yoon (2006), and for a more detailed argument Van Hees & Dowding (forthcoming).

mally studying the importance of these arguments, and the possible differences between voting rules. Although this thesis does not offer tools for studying arguments 3 and 5, there seems to be no reason to suppose that they, in turn, could not be turned into empirical questions.

The literature on the uncovered set, the Banks set, and related works (Miller 1980, Banks 1985) have given a somewhat more positive assessment of strategic voting because it has been shown that it restricts the set of outcomes a monopoly agenda setter may achieve to a small subset of the outcome space. A third approach proceeds by constructing models of strategic voting, and then studying whether some axiomatic conditions are satisfied under various voting rules (e.g., Felsenthal 1990). Usually the results of these exercises show that very few conditions are satisfied if the voters engage in strategic behaviour. Since voting rules satisfy very few axioms if the voters may vote strategically, all of these approaches suffer from the fact that comparisons between different voting rules are difficult to evaluate.[4]

It is impossible to *obtain information* on von Neumann-Morgenstern utilities in such a way that attitudes towards risk are not involved. One reason for the unwillingness to use preference intensities in normative voting models is based on the idea that since von Neumann-Morgenstern utilities incorporate attitudes towards risk, they are not suitable for normative assessments. However, it is not necessary to assume that attitudes towards risk are an intrinsic property of preferences. I do not need to join the general philosophical discussion concerning whether attitudes towards risk should be seen as belonging to a person's preferences here.[5] It suffices for my purposes to say that since those attitudes are irrelevant in terms of normatively evaluating the outcomes in voting games, voters' utilities should not incorporate attitudes towards risk. These attitudes are relevant only to the choices they make, and should be modelled through beliefs rather than through preferences in voting models. The first essay shows how to do this.[6]

Table 1.1 summarises the desiderata for voting models.

The second column 'Is X relevant for evaluating the outcomes normatively'

[4] We must bear in mind that the use of strategy-proofness is not restricted to voting theory. It has become an integral part of the implementation-theoretical toolbox through the *revelation principle*, according to which truth-telling, direct revelation mechanisms can generally be designed to achieve the Nash equilibrium outcome of other mechanisms. This can be proven in a large category of mechanism-design cases. It is used most often to prove something about the whole class of mechanism equilibria, by selecting the simple direct revelation mechanism, proving a result about that, and applying the revelation principle to assert that the result is true for all mechanisms in that context. Insofar as the principle is used merely as a mathematical tool in deriving theorems, my results have no relevance for the revelation principle. However, since the sum of utility provides a social welfare function, the finding that strategic voting increases utilitarian efficiency provides a case in which the intuition behind the revelation principle is falsified. These results do not provide a counter-example for this principle, however, because the models in which utilitarian efficiency is increased are not based on game-theoretical equilibria.

[5] Hansson (1988) and Rabin (2000) have argued that risk-aversion ought not to be conceptualised in terms of utility functions.

[6] It is clear that this model cannot measure attitudes towards risk in terms of the traditional Arrow-Pratt coefficient because it is defined in terms of the utility function.

X	Is X relevant for evaluating the outcomes normatively?	Is it desirable to include X in a realistic model of voter decisions?
intensity	yes	yes
risk attitudes	no	yes
beliefs	no	yes

Table 1.1: Desiderata for models of strategic voting

has a 'yes' entry if an item x (intensity, risk attitude, or beliefs) ought to be taken into account in evaluating voting outcomes, and the third column has a 'yes' entry if a voting model represents real-world voting realistically only by including X. Let me emphasise that when I talk about beliefs here, I only mean beliefs concerning the winning prospects of the various alternatives,and they should not be understood here as concerning the *content* of the policy alternatives or the characteristics of a candidate. The same comment applies to attitudes towards risk. It is clear that they should be included in, say, evaluating whether a new nuclear power plant should be built. They are thus irrelevant for normative evaluations only insofar as they affect voters' beliefs concerning other voters' voting behaviour, but since they inevitably affect the outcomes, the third column has a 'yes' entry. Risk attitudes, beliefs and intensities should be included in a realistic model of the behaviour of voters, but only preference intensities should be taken into account in evaluating voting outcomes.

In the framework that I am proposing the welfare consequences of strategic voting are evaluated by comparing voting outcomes under *sincere* and *strategic* voting *behaviour*. Given a welfarist standpoint, this would seem to be the obvious way of evaluating strategic voting. The structure of a model aimed at evaluating the welfare consequences of strategic voting on the basis of individual preferences has to be the following. It has to specify two different *behavioural assumptions*, expected utility- maximising (strategic) behaviour and sincere voting behaviour, and to compare the utilitarian efficiency under these two behavioural assumptions with a *fixed* set of preferences. If utilitarian efficiency is higher under strategic behaviour than under sincere behaviour, strategic voting may be said to be *welfare-increasing* compared to sincere voting. If the converse holds, strategic voting is *welfare-decreasing*.

Computer simulations are particularly well-suited for such analysis because they allow the generation of a fixed set of preferences and comparison of the outcomes of two different behavioural assumptions. Since voters' preferences have to be fixed for this kind of comparison, there has to be a *behaviour-independent* way of evaluating the alternatives. It must be possible to compare the alternatives in terms of voters' utilities. However, in that case the alternatives cannot be compared without making interpersonal comparisons of voters' utilities concerning them. The welfare consequences of strategic voting thus cannot be evaluated in a welfarist manner without making interpersonal comparisons. A model in which the welfare consequences of voting rules and strategic voting are evaluated has to separate the analysis of *voting behaviour* and of *normative judgements* concerning the outcomes.

The signal extraction model of incomplete information

The first essay, '*Signal extraction for simulated games with a large number of players*', provides one possible way of modelling incomplete information in mass elections; it describes a signal extraction model embedded in a computer-simulation framework. The main result of this paper is an expression for an agent's belief, given that he or she has obtained a perturbed signal concerning the preference profile. This result is directly applied in the second essay, and it was instrumental in developing the signal extraction model for the third essay. In computer simulations voters are assumed to derive their beliefs using statistical signal extraction. Nothing in this model restricts its use to voting contexts, but having been formulated for the *impartial anonymous culture* (IAC) assumption it is likely to be particularly suitable for studying voting. IAC states that each possible event, such as a voter's type, is equally likely to occur.

In incomplete information models of voting, voters need to formulate beliefs concerning how other voters are likely to vote. There is thus a need for assigning probabilities to the agents in one way or another. The IAC assumption, combined with the usual common prior assumption, does not lead to very interesting results in voting theory because it means that all voters have the same priors for the relevant quantities. A Bayesian model does not provide any advantage over a complete information model in such circumstances because uniform priors imply that all voters always vote sincerely unless, of course, they learn from previous voting choices and thereby formulate posterior probabilities.

My model solves this problem by assuming that voters obtain a *perturbed signal* of the preference or utility profile. In other words, a voter profile is first generated by computer, and each voter is given a perturbed signal of the relevant aspects of this realised profile. Typically, the relevant aspect is whether an alternative has a chance of winning.

Voters formulate their probabilities by considering the perturbed signal as the *mean* of the relevant random variable. They also need to assume something about the *variance* of the distribution of errors with which their signals are perturbed. This variance of the error distribution is taken as an exogenous variable in the simulations. Thus, using the incomplete information model in voting simulations amounts to testing how the *reliability* of the perturbed signals affects the results. The parameter value that describes reliability could be assumed to be known or not known by the voters. If they do not know it, the expected value could also be taken as an exogenous parameter that is tested in the simulations. The expected value of the reliability is called the *degree of confidence* in the signals because this term best conveys the intuitive interpretation of this parameter.

This method is particularly suitable for modelling *a large number of heterogeneous agents.* The crux of the model is that the voters' beliefs can be derived by characterising the *reliability* of their signals and their *confidence* in them. As the signals are randomly perturbed, each agent obtains a somewhat different signal, and thus derives a somewhat different belief. This facilitates the avoidance of the unhappy consequences of the standard Bayesian assumption of common priors. The model does not violate this assumption, however, because the priors could be considered as common before the agents obtain their signal.

Voting theorists have often expressed concern over the gap between complete information models and the paucity of information with which real voters are assumed to operate. The signal extraction model goes a long way towards solving this problem because it can be used to model informational conditions ranging from very close to complete information to wildly inaccurate information. Furthermore, since voters' degree of confidence in their signals may be separately modelled, it also allows for studying setups in which voters are hesitant to engage in strategic voting because they believe that their information is not fully reliable.

The welfare consequences of strategic voting under the Borda rule and in parliamentary agendas

Essays two and three show how intensities affect voting outcomes by formulating expected-utility conditions for strategic voting under some commonly used voting rules, and evaluate its welfare consequences. The second essay '*The welfare consequences of strategic voting in two commonly used parliamentary agendas*' shows that if voters engage in strategic behaviour under amendment and elimination agendas, their average utility is higher than if they always vote sincerely. The third essay '*The Borda rule is also intended for dishonest men*' provides a similar study of the Borda count. Here the outcomes are evaluated in terms of utilitarian efficiency, i.e., the frequency with which the utilitarian winner (the alternative with the highest sum of utility) is selected. The main finding in these essays is that strategic voting increases utilitarian efficiency and average utility. Given a utilitarian evaluation of voting outcomes, strategic voting turns out to be beneficial rather than harmful. This result prompts a new interpretation of the aforementioned theorems: if strategic voting is beneficial, it is no longer reasonable to interpret them as implying that there is something wrong with voting rules.

The main reason for this result is what I call the *counterbalancing* of strategic votes. In an electorate with a large number of voters who make their decisions based on incomplete information some individuals typically have an incentive to vote strategically for an alternative, say x, but at the same time, others have an incentive to strategically desert this very same alternative x by voting for some other alternative y. The conditions for voting strategically in terms of the structure of voters' preferences and their information are different in different voting rules, but they always indicate that, ceteris paribus, an alternative that has a large sum of utility is likely to obtain many and lose few strategic votes. Thus, many strategic votes for the utilitarian winner are likely to be counterbalanced by few strategic votes against.

The remaining three essays comprising the thesis provide the necessary philosophical and methodological support for the voting models. The rest of this introduction describes how essays one, four, and five relate to the voting models.

Unrealistic assumptions in rational choice theory

There are two epistemological difficulties concerning preference intensities. The first is that it is difficult to observe intensities, even *intra*personal ones. One particular formulation of this argument is that it is impossible to distinguish between intensities and attitudes towards risk in the reference-lottery technique (see e.g., Hirshleifer & Riley 1992) that lies behind the von Neumann-Morgenstern utility construction. The second is that since it is impossible to observe interpersonal differences in preference intensities, interpersonal comparisons are meaningless with von Neumann-Morgenstern utilities, and are generally considered to be methodologically questionable because it is not possible to derive such comparisons from individual choices.

The fourth essay '*Unrealistic assumptions in rational choice theory*' is related to models of strategic voting in two ways. First, the concept of utilitarian efficiency presupposes interpersonal comparability of preference intensities, and the results of the voting models thus critically hinge on the possibility of responding to the criticism in a satisfactory way. The finding that strategic voting increases utilitarian efficiency was tested for robustness with respect to different interpersonal comparisons of intensities.[7] The results of such robustness analysis reported in essay two show that although the exact numerical results are affected by different interpersonal comparisons, strategic voting remains welfare-increasing if the interpersonal variation is not allowed to exceed the one-man-one-vote principle too frequently. This essay provides a theoretical justification for robustness by explicating its epistemic credentials.

Secondly, the essay provides support for a particular assumption upon which voting simulation models are based; the IAC. As a description of real electorates this assumption is wildly inaccurate. However, its use may be perfectly legitimate if the purpose of the theory employing it is to provide an account of how two or more alternative institutions are able to cope with a particular problem. The purpose of some assumptions is to exaggerate some characteristic in order to bring it into focus. In the case of strategic-voting models, the IAC assumption is used because it maximises the number of simulated games in which strategic

[7] By different interpersonal comparisons I mean different numerical values for the parameters that define a voter's preference scale. The different comparisons are all defined within full cardinal comparability. It is thus not a matter of different comparability assumptions in the sense of allowable transformations for utility functions as in standard accounts of interpersonal comparisons, but rather a question of different 'actual' comparisons given full comparability. See e.g., Blackorby & Bossert (2006) for a recent survey on interpersonal comparisons.

voting may affect the results. It is thus used because it distorts reality in the right way: it allows for studying the welfare consequences of strategic voting. It must thus be borne in mind that the voting models presented in this thesis are not meant to be realistic with respect to how commonly strategic voting occurs.

The concluding essay: a methodological critique of social choice theory

The concluding fifth essay is entitled '*A farewell to IIA*'. Conclusions from the voting models are drawn, and supporting arguments provided. The discussion focuses on two perennial questions in social choice theory: intensities of preference and the normative validity of the Independence of Irrelevant Alternatives (IIA) condition (Arrow 1963). I would never have ventured to resurrect a topic that had been extensively discussed decades previously, if I did not have something new and important to say about it.

A further aim in this essay is to contribute to the current discussion on the IIA condition. Donald Saari (1998, 2001) argued that it is not acceptable because a voting rule that satisfies it cannot differentiate between rational and irrational voters. There are, however, those who continue to think that it is acceptable, mainly on the grounds that it is closely related to precluding strategic voting (e.g., Risse 2004), and strategic voting is considered undesirable.

Saari uses his arguments against IIA in defence of the Borda rule and as an attack on the majority rule with various agendas, i.e., the rules that are discussed in this thesis. It is well-known that the Borda rule violates IIA, but since it asks voters to provide a full ordering of the alternatives, it collects information on the transitivity of the preferences.

It is shown in this essay that the independence condition is violated under the majority rule with an amendment agenda if voters engage in strategic behaviour. Given that IIA is also violated under a voting rule in which it is commonly thought to be satisfied, it is evident, and hardly requires proof, that it is violated under all commonly used voting rules. This has implications for the recent debate between Saari (2006, 2003*b*) and Risse (2001, 2004, 2005): IIA is violated under *all* voting rules if voters are rational. Furthermore, if voters engage in strategic behaviour, even the majority rule discloses connecting information between pairs of alternatives.

Therefore, Saari's argument against IIA is not valid *qua* the argument for the Borda rule and against other voting rules. On the other hand, the Borda rule is commonly taken to be highly vulnerable to strategic voting, and this has been considered the main argument against it.[8] The third essay shows that strategic voting is also beneficial under the Borda rule. The whole discussion of IIA and the comparison of voting rules must thus be rethought. It is worthwhile emphasising that my argument is directed against the common interpretation of impossibility theorems rather than for or against any particular voting rule.

[8] Saari (2006) does not accept this judgment, however.

Indeed, my point is that it is not possible to use IIA for arguing for or against any such rule.

What has traditionally been considered the *main justification for IIA* is invalid. The strongest argument for it has been that it is closely related to precluding the strategic misrepresentation of preferences, and this misrepresentation has been considered undesirable. My results show that strategic voting results in better outcomes in terms of utilitarian efficiency than sincere voting, and therefore this argument fails. My argument boils down to the old criticism that IIA does not take preference intensities into account. Mackie (2003) provides a recent version of this criticism. What I have done is to show that this argument need not rest on the mere intuition that intensities are important. If the welfare consequences of strategic voting are explicitly studied, it turns out that violating IIA through strategic voting has beneficial intensity-related consequences at the aggregate level.

I thus argue that IIA ought to be violated under many commonly used voting rules. Providing further support for this claim entails going through the various arguments that have been presented for the IIA condition. The essay is organised around three of these: the observability, the epistemological-moral and the strategic-voting arguments. I have discussed the last of these above. The moral part of the epistemological-moral argument is the idea that cardinal von Neumann-Morgenstern utilities should not be used in social-welfare judgments because they incorporate attitudes towards risk, and these attitudes are morally irrelevant in such judgments. The epistemological part is that it is not possible to distinguish between attitudes towards risk and intensities in individual behaviour.

This argument is perfectly acceptable. However, the methodological injunction to use only preference orderings in evaluating the alternatives follows from it only if it is possible to obtain reliable information on preference orderings. The observability argument must thus be discussed. The postulation here is that it is not possible to obtain reliable information on preference orderings *either* in a voting context. The main reason for this lies in strategic voting itself: even though voters are able to explicitly express only their preference orderings under most voting rules, their choices are also influenced by their preference intensities. If we observe that a voter gives his or her support to x, this does not necessarily mean that this alternative stands highest in his or her preference ordering. It is, of course, easier to obtain information on individual preference orderings than on preference intensities and interpersonal comparisons. However, given that it is impossible to obtain fully reliable information on preference orderings, and given that collecting information on real preference profiles is not part of the practice of theorists endeavouring to provide normative comparisons concerning voting rules, the injunction to restrict the preference information to orderings is without proper warrant.

The thesis amounts to a fundamental methodological criticism of social choice theory. Since strategy-proofness and IIA are not normatively acceptable, social choice theorists have been trying to answer the wrong question. The bulk of the current literature on strategic voting represents attempts to find strategy-proof mechanisms, or to determine the degree to which different voting rules can be manipulated. The simulation results presented in these essays on strategic voting show that these questions are misguided and provide misleading results on voting rules. The important task here is to determine the circumstances in which strategic voting is welfare-increasing or welfare-diminishing. The circumstances include different voting rules, and assumptions concerning voters' information and the utility profile.

Arrow's theorem and the Gibbard-Satterthwaite theorem are logically unassailable in the sense of being deductive. Their interpretation, however, is a different matter. They are not false, but their current interpretation is misleading. Arrow's theorem cannot show that all voting rules are problematic if one of its conditions is not normatively compelling. Whether the fact that all voting rules are vulnerable to strategic voting is a good or a bad thing is something that only research can establish. Under all voting rules I have studied thus far, strategic voting is beneficial.

Chapter 2

Signal extraction for simulated games with a large number of players

Abstract

A signal extraction problem in simulated games is studied. A modelling technique is proposed for deriving beliefs for players in simulated games. Since standard Bayesian games provide conditions for beliefs on the basis of the common prior assumption, they do not allow for non-uniform beliefs unless the game has some dynamic structure that allows for learning. The framework presented allows for deriving beliefs by characterizing the reliability of the signals, and the players' degree of confidence in these signals. This makes it particularly suitable for games with a large number of heterogenous players.

Keywords: Signal extraction; Simulated games; Beliefs; Heterogeneous players

2.1 Introduction

In standard Bayesian incomplete information models, the players' actions are independent of the realisations of random variables, because they are assumed to know the probability distributions for the relevant random variables, but not the realisations of these variables (e.g., Harsanyi 1967-8, 1995). These models assume that the players start with common priors and update them with Bayes' rule as the play unfolds. Bayesian models have proven to be very useful in game theory but they are not applicable in all circumstances. For example, in games where the players do not have the opportunity of updating their beliefs by observing other players' choices, they may update their beliefs only if they obtain a signal that contains some valuable information for them.

A decade ago Carlsson & van Damme (1993) proposed an alternative to the Bayesian approach of analysing incomplete information in games; the *global games*.[1]

In this paper, we will propose a modelling technique that is similar to global games and to *statistical signal extraction* in that the players are assumed to observe a *perturbed signal* of the underlying true game. However, instead of presenting an analytical model, our model is best applied in simulated games with a large number of heterogenous players.

A model where the players receive perturbed signals concerning the true game[2] is particularly appropriate when the preference profile (the set of preferences for all players) is drawn from some symmetric distribution, and the players need to know something beyond their priors about the characteristics of a large population of heterogenous players. We will study a setting where such characteristics include the *realised* distribution of players, i.e. the *number* of players with some particular type of preferences. More particularly, we will show how to derive beliefs in simulated games where the profile of player types is generated with a uniform distribution such that each player type is equally likely.

It is difficult to model the beliefs of a large number of heterogenous players because it is practically impossible to collect information on such beliefs. Our approach provides one possible way of dealing with such situations, because we characterise the players' information by the *reliability of signals* they receive, and by the *degree of confidence* that they have for these signals. The framework allows us to derive beliefs for a large number of players with heterogenous preferences who receive different signals.

Since the terms 'reliability' and 'degree of confidence' have various meanings in different frameworks, let us emphasize at the outset that the reliability of the players' information is a property of the *signals* rather than an intentional state of the players in our model. It is formalised as the standard deviation of the perturbations. The degree of confidence also concerns the signals rather than the beliefs derived from them. The degree of confidence *affects* the players' beliefs, but it is conceptually and formally different from those beliefs. The degree of confidence can, however, be considered as a subjective property of the players.

An obvious area of application for such models is voting theory where simulations with the uniform distribution on player types is known as the *impartial culture* assumption.[3] In principle, the technique is general enough to be ap-

[1] See Morris and Shin (2003) for a review of the literature on global games. Frankel et al. (2003) generalise the results of Hansson and van Damme to arbitrary numbers of players and actions.

[2] Such models are not to be confused with *signaling models*, where the players themselves send signals.

[3] See Tsetlin, Regenwetter & Grofman (2003) and Gehrlein (2002) for recent discussions of impartial culture.

plicable in any situation with a large number of players, but the fact that we derive beliefs for players whose types are drawn from a uniform distribution of course limits the applicability of the model.

The structure of the paper is the following. Section 2.2 delineates the similarities and differences of our approach to global games. Section 2.3 describes the signals. The beliefs are derived from these signals in appendix A. In section 2.4 we discuss how the reliability of the signals and the players' confidence in these signals affect the beliefs. Section 2.5 compares our concept of the degree of confidence to some previous conceptualizations.

2.2 Global games

A *global game* (Carlsson and van Damme 1993) is an incomplete information game where the actual payoff structure is determined by a random draw from a given class of games and where each player makes a noisy signal of the selected game.

Consider a situation in which the players know that some game in the class of games G will be played, but they do not know which one. A class of games is a set of games with a set of players I $(i = 1, 2, ..., N)$ and a set of possible payoff profiles Π.

Initially, the players have common prior beliefs represented by a probability distribution with support on G. Before choosing an action, each player gets additional private information in the form of a perturbed signal of the actual game g to be played. The resulting incomplete information game is thus called a global game. It can be described by the following steps:

1. Nature selects a game g from G.
2. Each player observes g with some noise.
3. Players choose actions simultaneously.
4. Payoffs are determined by g and by the player's choices.

Player i:s signal is described by a random variable S_i^ε which is defined by:

$$S_i^\varepsilon = g + \varepsilon r_i,$$

where r_i is a realisation of a random variable R, and ε is a scale parameter. The players are thus assumed to observe the realisation of a random variable g, plus an error term εr_i. The players' signals are correlated, because they are noisy signals of the *true game*.

Our approach differs from global games as follows. First, most contributions in global games derive limit uniqueness results assuming that ε approaches zero. In this sense the signals in global games are 'close' to being correct, whereas in our approach parameter ε may in principle be of any size whatsoever. Our players are thus allowed to be 'less' informed than the players in global games.

Second, it is usually assumed in global games that as $\varepsilon \to 0$, each player becomes certain that she and her opponent have observed the true game. In contrast, we may consider parameter ε as unobservable and not necessarily known. Therefore, even if the perturbance term were zero, we need not assume the players are certain that they have observed the true game.

Third, our model is better suited for situations with parametric rationality than strategic rationality (Elster 1983). We derive *beliefs* for the players on the basis of what they know about the payoff profile, but we do not derive equilibrium strategies. Our model of signals and beliefs can be used *together* with different models that formulate the players' expected utilities. Since we do not present any particular application in this paper, we will not present an account for how the players can take other players' optimal strategies into account.[4]

2.3 The signals

The players' preferences are defined on a set X of items. X contains some items that the players rank or compare with each other. The most typical application of such a model is one where the players are interested in whether a majority of players prefer j to k or vice versa. The model is obviously applicable to any number of pairwise comparisons of items.

Let N denote the number of players and $n^g(j \succ k)$ the number of players who prefer item j to item k $(j, k \in X)$ in a simulated game g. Since all the symbols will be defined for a given simulated game g, we will not subscript our variables by the 'g' in the sequel.

Let $n_i(j \succ k) = 1$, if player i prefers j to k, and $n_i(j \succ k) = 0$, if player i prefers k to j. In this simple setting, a player's *type* refers merely to whether she prefers j to k. Since we assume that the profile of player types for a simulated game g is generated with *impartial culture*, each type is equally likely. $n(j \succ k)$ can thus be viewed as a sum of N Bernoulli trials, $n(j \succ k) = \sum_{i=1}^{N} n_i(j \succ k)$, and the probability p that such a Bernoulli trial results in the outcome $n_i(j \succ k) = 1$ is $\frac{1}{2}$.

The players are assumed to obtain a perturbed signal of the number of players who prefer j to k. One way of writing the signal is as follows:

$$s_i^{\varepsilon} = \frac{n(j \succ k)}{N} + \varepsilon r_i, \tag{2.1}$$

where ε is a scaling factor and r_i is a realisation of a random variable P. ε reflects the *reliability of the signal*. In this paper we will assume that the variable P is standard normal; $P \sim N(0, 1)$.

What we want to do is to derive the probability that the number of players who prefer j to k is larger than $\frac{N}{2}$, given a signal S_i^{ε}. Let $p_i(j, k)$ denote such

[4] For an example of how this can be done, see Lehtinen (2006*b*).

a probability for player i:

$$\begin{aligned} p_i(j,k) &= prob(n(j \succ k) > n(k \succ j)) && (2.2)\\ &= prob(n(j \succ k) > \frac{N}{2}) && (2.3)\\ &= prob(\frac{2n(j \succ k)}{n(j \succ k) + n(k \succ j)} > 1). && (2.4) \end{aligned}$$

The derivation of such probabilities requires knowledge of the variance of the variable $n(j \succ k)$. In simulated games generated with the impartial culture assumption, this variance is Np^2.

Since $n(j \succ k)$ is the sum of N Bernoulli trials, the Central Limit Theorem implies that the random variable $n(j \succ k)/N$ can be approximated with a normally distributed random variable $\mathcal{N}(j \succ k)$. Naturally, invoking the central limit theorem restricts the applicability of this model to games with a relatively large number of players.

Let us write $R_i = \varepsilon r_i$ The signal can now be written as a sum of two normally distributed random variables:

$$s_i^\varepsilon = \mathcal{N}(j \succ k) + R_i. \qquad (2.5)$$

Before deriving beliefs from such signals, let us point out that it will usually be more convenient to use a *standardized sum* of Bernoulli trials, $Q(j \succ k)$, instead of the variable $n(j \succ k)$ itself. The standardised sum is given by:

$$Q(j \succ k) = \frac{n(j \succ k) - Np}{\sqrt{Np^2}}. \qquad (2.6)$$

In models with impartial culture $p = \frac{1}{2}$, so that this is

$$Q(j \succ k) = \frac{2n(j \succ k) - N}{\sqrt{N}}. \qquad (2.7)$$

A *standardised signal* of player i is then given by

$$\begin{aligned} s_i^\varepsilon &= \frac{2n(j \succ k) - N}{\sqrt{N}} + \varepsilon \cdot r_i && (2.8)\\ &= Q(j \succ k) + R_i. && (2.9) \end{aligned}$$

Deriving the beliefs from such signals involves standard statistical inference. The derivation is relegated to an Appendix because it is somewhat tedious. Equation A.13 in Appendix A shows that the players' beliefs are given by:

$$p_i(j,k) = 1 - \Phi\left(-\frac{1}{\varepsilon\sqrt{1+\varepsilon^2}} s_i^\varepsilon\right). \qquad (2.10)$$

2.4 Reliability of signals and confidence

We will now consider how the degree of confidence and the reliability of the signals are interpreted in our model. Let us define the random variable X as follows: $X = (Q(j \succ k)|S = s_i^\varepsilon)$. X is the conditional value of the standardised variable $Q(j \succ k)$, given the signal s_i^ε. Inserting the standard deviations $\sigma_R = \varepsilon$, and $\sigma_Q = 1$ into equation (A.8) in the appendix gives the density of variable X:

$$f_X(x) = \frac{\sqrt{1+\varepsilon^2}}{\sqrt{2\pi}\varepsilon} \exp\left(-\frac{1}{2}\frac{(1+\varepsilon^2)}{\varepsilon^2}\left(x - \frac{s_i^\varepsilon}{(1+\varepsilon^2)}\right)^2\right). \tag{2.11}$$

The expected value of X is thus

$$E[X] = E[Q(j \succ k)|S = s_i^\varepsilon] = \frac{s_i^\varepsilon}{(1+\varepsilon^2)}. \tag{2.12}$$

Equation (2.12) has a natural interpretation. The smaller the variance $\left(\varepsilon^2\right)$ of the error term R, the more exact information the signal provides of the variable $Q(j \succ k)$, and the more it will be rational to update the beliefs.

Note that

$$\lim_{\varepsilon\to 0} E[Q(j \succ k)|S = s_i^\varepsilon] = s_i^\varepsilon. \tag{2.13}$$

Hence, as the error term R $(= \varepsilon r_i)$ approaches zero, the signal provides more and more exact information on the ratio $\frac{2n(j\succ k)}{n(j\succ k)+n(k\succ j)}$ or the corresponding standardised ratio $\frac{2n(j\succ k)-N}{\sqrt{N}}$. Furthermore,

$$\lim_{\varepsilon\to \infty} E[Q(j \succ k)|S = s_i^\varepsilon] = 0. \tag{2.14}$$

Hence, as the variance of the error term approaches infinity, the expected value of the conditional value of the standardised variable Q approaches zero. This means that the signals become more and more uninformative as the variance of the perturbations increases.

It may not be realistic to assume that the players know the reliability ε of their signals. In such cases the players may be assumed to formulate expectations $E_i(\varepsilon)$ concerning the reliability of their signals ε. The player's beliefs can then be derived using a modified version of equation 2.10:

$$p_i(j,k) = 1 - \Phi\left(-\frac{1}{E(\varepsilon)\sqrt{1+E(\varepsilon)^2}} s_i^\varepsilon\right). \tag{2.15}$$

Let us say that ε denotes the *reliability of signals*, and $E_i(\varepsilon)$ the *degree of confidence* in these signals. Considering equation 2.8, we may now define the following concepts:

Definition 1 *The* ***reliability of signals****,* ε*, is the standard deviation of the perturbation of the signals* s_i^ε $(i = 1, ..., N)$.

Definition 2 *The **degree of confidence** in the signals is the expectation of the reliability of the signals $E(\varepsilon)$.*

Here is how the proposed model can be used in computer simulations. We can test how the outcomes differ when we keep the preference profile fixed, but vary the reliability of the signals and the degree of confidence. It is usually convenient to assume that all players and player types have the same reliability of signals and the same degree of confidence, but this is by no means necessary. It is also possible to study cases where the players are systematically over-confident ($E(\varepsilon) < \varepsilon$) or under-confident ($E(\varepsilon) > \varepsilon$).

Definition 3 *The players have a **correct degree of confidence** if their degree of confidence in their signals equals the degree of reliability of these signals; $E_i(\varepsilon) = \varepsilon_i$ for all $i \in I$.*

The correct degree of confidence means that the players' beliefs about the quality of their signals reflects the real quality of those signals. The standard way of distinguishing between objective and subjective interpretations of probabilities is to say that probabilities can be interpreted as objective if they are based on a known probabilistic process. Probabilities are subjective if they are not based on such processes. It may thus be said that models assuming correct degree of confidence incorporate objectively interpreted probabilities.

The smaller ε is, the more reliable a player's signals are, and the smaller $E_i(\varepsilon)$ is, the greater the player's degree of confidence in her signals. If $\varepsilon = 0$ for all $i \in I$, we say that the players have *perfectly reliable information.* However, if the players are not assumed to know the value of ε, even though the players have the same signals as they would have in a corresponding complete information game, this does not yet imply that they act in the same way as players with complete information.

If $\varepsilon = E_i(\varepsilon) = 0$ for all players, i.e. if the players have both perfectly reliable information and a correct degree of confidence in their signals, the players' signals correspond to the knowledge of players in a corresponding complete information game. In this sense, complete information games can be viewed as a special case of our information model. To see this, note first that

$$\begin{cases} p_i(j,k) > \frac{1}{2} \Leftrightarrow s_i^\varepsilon > 0 \\ p_i(j,k) = \frac{1}{2} \Leftrightarrow s_i^\varepsilon = 0 \\ p_i(j,k) < \frac{1}{2} \Leftrightarrow s_i^\varepsilon < 0. \end{cases}$$

With $\varepsilon = E(\varepsilon)$, and inserting 2.8 into 2.15, we have

$$\begin{aligned} \lim_{E(\varepsilon)\to 0} p_i(j,k) &= \lim_{E(\varepsilon)\to 0} \left\{ 1 - \Phi\left(-\frac{1}{E(\varepsilon)\sqrt{1+E(\varepsilon)^2}} \left[\frac{2n(j \succ k) - N}{\sqrt{N}} + \varepsilon \cdot r_i \right] \right) \right\} \\ &= 1 \Leftrightarrow \frac{2n(j \succ k) - N}{\sqrt{N}} > 0 \\ &= 0 \Leftrightarrow \frac{2n(j \succ k) - N}{\sqrt{N}} < 0. \end{aligned}$$

Even if $\varepsilon \neq 0$, it is possible that a player's perturbed signal corresponds exactly to the true value of the variable Q, if r_i happens to be exactly zero. This is very unlikely, of course, because the perturbations are normally distributed. If $r_i = 0$, and $E_i(\varepsilon) \neq 0$, player i's signal is essentially the same as in a corresponding complete information game, but she will not act as if she had complete information, because player i's beliefs are not degenerate (0 or 1) when she does not have full confidence in her signals. In other words, if a player happens to guess the ratio $\frac{2n(j \succ k)-N}{\sqrt{N}}$ correctly, she is not willing to act on the basis of this guess if she believes it is based on highly dubious evidence.

2.5 Relation to some previous literature

Many previous accounts have considered the degree of confidence *in one's beliefs*. Here, however, we model the degree of confidence *in one's signals*. This is why we need not invoke second-order beliefs (e.g., Marschak 1975, Borch 1975), or intervals of beliefs (e.g., Good 1962, Gärdenfors and Sahlin 1982) to take into account the players' confidence. These approaches suffer from well-known weaknesses. The degree of confidence should already be taken into account in the first-order probabilities and thus the second-order probabilities are superfluous (see Savage 1954, p. 58, de Finetti 1977). If the upper and lower probabilities in the interval do not yield the same recommendations for action, there is no evident way to choose between the different actions (e.g., Skyrms 1990, p. 113).

Second-order beliefs and intervals of beliefs have been proposed as a solution to Ellsberg's (1961) paradox. The literature that has tried to respond to Ellsberg's experiments[5] has been concerned with two related concepts; the *degree of confidence* in one's probability judgments and the *ambiguity* of the players' information. Our approach is not designed nor suitable for modelling ambiguity because the players are always assumed to know the form of the distribution that is of interest to them. At the same time, the degree of confidence has a natural interpretation in our model.

Savage (1954, p. 68) denies that the degree of confidence in one's information can have an effect on a person's judgment of probabilities: "...the particular personalistic view sponsored here does not leave room for optimism or pessimism, however these traits be interpreted, to play any role in the person's judgment of probabilities". But since we model the degree of confidence *in the signals* rather than the degree of confidence *in the probability judgments*, we arrive at unique probabilities that may be used in standard expected utility calculations. This is why we can sidestep Skyrms' criticism even though we explicitly model the players' degree of confidence.

[5] See also the papers on the third kind of solution; non-additive probabilities (Gilboa 1987), (Schmeidler 1989).

2.6 Conclusions

Our account of perturbed signals is particularly well suited for modelling situations where a large number of players have heterogenous preferences and beliefs. It is designed to be used as a part of a larger expected utility model, where the minor differences in the players' beliefs play an important role.

Chapter 3

The welfare consequences of strategic voting in two commonly used parliamentary agendas

Abstract

This paper studies the welfare consequences of strategic voting in two commonly used parliamentary agendas by comparing the average utilities obtained in simulated voting under two behavioural assumptions: expected utility maximising behaviour and sincere behaviour. The average utility obtained in simulations is higher with expected utility maximising behaviour than with sincere voting behaviour under a broad range of assumptions. Strategic voting increases welfare particularly if the distribution of preference intensities correlates with voter types. (JEL classification numbers: D71, D81)

Keywords: strategic voting; agendas; welfare; simulation; counterbalancing

3.1 Introduction

This paper investigates whether strategic voting is beneficial or harmful in two commonly used parliamentary voting rules; amendment and elimination agendas. It is widely acknowledged that strategic voting may be beneficial because it may contain the power of an agenda-setter[1] but usually the possibility of strategic voting is considered an undesirable characteristic of a social decision

[1] See Miller (1980), Shepsle & Weingast (1984), and Banks (1985).

mechanism.[2] Thus far, however, the welfare consequences of strategic voting have not been studied by explicitly comparing strategic voting behaviour with sincere voting behaviour (but see Chen & Yang 2002).[3]

The welfare consequences of strategic voting are evaluated by comparing voters' *average utility* obtained with *Expected Utility maximising voting behaviour* (EU behaviour) and with *Sincere Voting behaviour* (SV behaviour). In SV behaviour *all* voters always vote sincerely. In EU behaviour, voters may vote strategically or sincerely in any given stage of voting depending on the expected utility of the choice options. If the average utility obtained with EU behaviour is higher than with SV behaviour, strategic voting is said to be *welfare-increasing*. Otherwise it is *welfare-decreasing*.

The idea that strategic voting may result in better outcomes than sincere voting on the aggregate level may be surprising because strategic voting means voting for an alternative that is not highest in one's preference order. The mechanism of *counterbalancing of strategic votes* explains why, when, and how strategic voting may lead to desirable outcomes on the aggregate level. In a large group of voters, there are usually incentives to vote strategically both for and against a given alternative. Strategic votes for an alternative are counterbalanced by strategic votes against this same alternative. An *intensively* supported alternative gets *more* strategic votes than a less intensively supported alternative.[4] Strategic voting thus increases the chance that an intensively supported alternative beats an alternative which has less intense support but a broader base of supporters. If an intensively supported alternative would lose against an alternative with a larger number of supporters in a sincere pair-wise first-round vote between the two, strategic voting may increase welfare by increasing the chance that an intensively supported alternative is selected in an early stage of voting.

Some scholars have lamented that the widespread use of majority rule has not been properly explained, particularly in view of the negative impossibility and instability (McKelvey 1976, Schofield 1978) results in social choice theory. It has been widely acknowledged that preference intensities are relevant for social welfare judgements[5], but there are very few models that explicitly try to study how these intensities affect voting outcomes (but see Blais & Nadeau 1996). The traditional criticism of majority rule is that it does not take into account preference intensities. The results presented here provide a more positive perspective on majority rule than many previous results in voting theory because it will be shown that strategic voting not only may, but is likely to be beneficial in the sense that the outcomes reflect preference intensities *if* and only if *voters vote strategically*.

[2] However, Miller (1977) shows, by way of an example, that strategic voting may select the Condorcet winner when sincere voting does not. The Condorcet winner is an alternative that the majority of voters prefer to all other alternatives.

[3] Vote-trading is also a way to vote strategically, and its welfare consequences have been investigated. See Shepsle & Weingast (1994) for a fairly recent review.

[4] The literature on vote-trading has also acknowledged that strategic voting allows for expressing preference intensities (see e.g., Stratmann 1997).

[5] See e.g. Hildreth (1953), Coleman (1966) and Mackay (1980, p. 42).

Preference intensities in agenda voting can be explicitly modelled only in a model with incomplete information. The model of incomplete information is based on statistical *signal extraction* since voters obtain *noisy signals* of the true structure of the game, and formulate beliefs on the basis of these signals[6]. The signal extraction model is explained in more detail in Lehtinen (2006*a*). The model has been applied to Borda rule (Lehtinen forthcoming) and to plurality and runoff rules (Lehtinen 2006*b*).

Instead of presenting an analytical model, computer simulations are used for modelling voters' belief formation and behaviour. Simulations are used for the following reasons. First, welfare-increasing strategic voting is what the literature on computer simulations calls an "emergent property", it emerges only when the individual votes are combined. The mechanism of counterbalancing strategic votes explains why an "invisible hand" result is obtained. Although it may be possible to derive such a result analytically, it is very difficult to analyse the interaction of hundreds of heterogeneous voters with an analytical model. Second, the purpose of the simulations is to examine *how much* voters' preference intensities must correlate with voter types, and *how reliable* must voters' signals be, in order for strategic voting to be welfare-increasing. This is why the degree of reliability and the degree of correlation are taken as exogenous parameters.

The existence of a *Condorcet winner (CW)* is usually considered sufficient for satisfactory performance in majority rule. If there is a Condorcet winner among the alternatives, this alternative will be the outcome under amendment agendas if all voters vote sincerely (Black 1958), or if they maximise utility with complete information.[7] However, various results have established that the existence of a Condorcet winner is highly unlikely, especially if the number of alternatives and/or voters is large.[8]

Simulation approaches to voting have evaluated and compared voting rules by investigating how frequently a Condorcet winner is chosen in a voting rule (assuming that it exists), or by investigating how frequently a *utilitarian winner* (the alternative with the largest sum of utility) is chosen (see e.g. Merrill 1988).

All well-known incomplete information models of strategic voting in majority rule (Enelow 1981, Jung 1987, Ordeshook & Palfrey 1988) assume that voters condition their choices on the possibility that they are pivotal in the sense that they make their choices by comparing the expected utility of voting for each of the alternatives. Enelow's model differs from the other models, however, in that it does not assume that the voters *formulate beliefs* by conditioning on the assumption of being pivotal in the first round of voting. If a voter who conditions her choices on being pivotal has poor knowledge of the type

[6] This model of incomplete information is also similar to global games (Carlsson & van Damme 1993). See Morris & Shin (2003) for a review. See also Frankel, Morris & Pauzner (2003).

[7] See McKelvey & Niemi (1978), Moulin (1979), and Sloth (1993).

[8] See McKelvey (1990) and Austen-Smith & Banks (1999) for surveys on the analytical literature on the existence of a Condorcet winner. Mueller (1989) and Gehrlein (2002) provide overviews of the simulation approaches.

distribution of voters, she may well obtain a worse outcome for herself by voting strategically than she would have obtained by voting sincerely. Therefore, while conditioning one's *choice* on being pivotal is rational, conditioning one's *beliefs* on being pivotal is irrational.

The paper is organised as follows. EU behaviour under amendment agendas is based on Enelow's (1981) expected utility model, which is introduced in section 3.2.1. Section 3.2.2 presents a similar expected utility model for elimination agendas. These basic building blocks are sufficient for understanding the logic of the welfare consequences of strategic voting. Simple examples in section 3.3 show that the utilitarian winner rather than the Condorcet winner may be selected if voters engage in strategic voting under incomplete information.

In the rest of the paper, the circumstances in which strategic voting increases or diminishes welfare are investigated using computer simulation. A model of incomplete information is introduced in section 3.4 by describing the assumptions related to voters' signals and beliefs.

Section 3.5 explains in detail how the counterbalancing of strategic votes affects the welfare consequences of strategic voting. Section 3.6 describes the structure of the simulation framework. The behavioural assumptions of EU behaviour and SV behaviour are analysed with *setups*. A setup is a collection of assumptions on voters' preferences, beliefs, behaviour, and the institutional structure. This section also establishes the criteria for evaluating voting outcomes.

Section 3.7 presents simulation results. Since the results depend on a utilitarian welfare function, it will be necessary to discuss interpersonal comparisons of utilities. Section 3.7.3 presents simulation results with various different interpersonal comparisons. The results from these various setups indicate that strategic voting increases welfare irrespective of what kinds of interpersonal comparisons are made. The purpose of these simulations is thus to show that the results are robust with respect to interpersonal comparisons. Section 3.8 presents the conclusions.

3.2 Expected utility models for agendas

3.2.1 Amendment agendas: Enelow's model

Let $X = \{x, y, z\}$ denote a set of available alternatives[9], $I = \{1, 2, ..., i, ..., N\}$ a set of voters, and U_i voters i's utility. Let U_1, U_2, and U_3 denote the utilities for the best, second-best, and the worst alternatives, respectively. (The subscript i denoting the individual is dropped here in order to avoid clutter.) The possible voter types are shown in Table 3.1.

Let us say that U_2 denotes a voter's *intensity* of preference. There are six different *types of voters*, $t^1, t^2, ..., t^6$. A voter's type refers only to his or her

[9] Only the case with three alternatives is studied in this paper. Extending the model to any number of alternatives is possible but so complicated that it requires another paper (Lehtinen 2002).

type of voter						
t^1	t^2	t^3	t^4	t^5	t^6	utility
x	y	z	x	y	z	U_1
y	z	x	z	x	y	U_2
z	x	y	y	z	x	U_3

Table 3.1: Voter types and utilities with three alternatives

order of preferences here, it does not include a specification of his or her beliefs. All preferences are assumed to be strict.

Alternatives are put to a sequence of pair-wise majority comparisons in an *amendment agenda* or in an *elimination agenda*.[10] An amendment agenda is constructed as follows: two alternatives (say x and y) are put to a majority vote against each other in the first round of voting. The winner of this first contest is then put to vote against the third alternative (z) in a second round of voting. Figure 3.1 presents this amendment agenda. Since path-dependence is not studied in this paper, other possible voting orders in amendment agendas are not shown here.

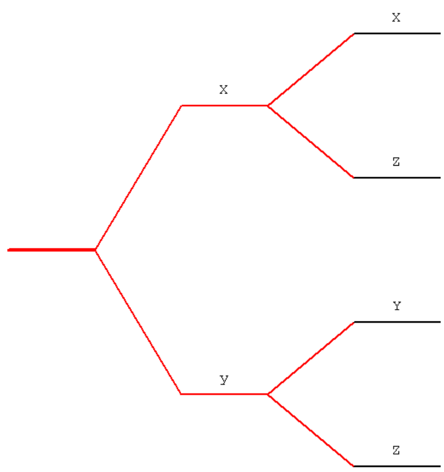

Figure 3.1: An amendment agenda with three alternatives

Voter i's subjective probability that a given alternative j beats k ($j, k \in X$) in a pair-wise second-round contest is denoted $p_i\,(jBk)$. In the first round of voting, voters' choice options are lotteries on the second-round outcomes.

In the first round of voting, voters choose by evaluating lotteries $(x, z;\ p_i\,(xBz)\,, 1 - p_i\,(xBz))$ and $(y, z;\ p_i\,(yBz)\,, 1 - p_i\,(yBz))$. Maximizing expected

[10] See Ordeshook (1986), Ordeshook & Schwartz (1987), and Miller (1995) for discussions on different agendas.

utility implies giving one's vote for the branch of the voting tree with the greatest expected utility. A voter will vote for the upper branch (i.e. for x) if

$$\begin{aligned} & p_i\,(xBz)\,U_i(x) + (1 - p_i\,(xBz))U_i(z) \\ \geq\ & p_i\,(yBz)\,U_i(y) + (1 - p_i\,(yBz))U_i(z). \end{aligned} \tag{3.1}$$

If the expected utility is the same for the two branches, the voter is assumed to vote sincerely. Voters of types 2 and 4 have a dominant strategy to vote sincerely (Farquharson 1969). Enelow uses a zero-one normalization for utilities for formulating the model. Although this normalization is not used in the simulations, the examples presented in later sections are formulated using this normalization in order to simplify the presentation.

3.2.2 Elimination agendas

Although most of this paper is concerned with amendment agendas, elimination agendas are also briefly considered. Under an elimination agenda, alternative x first put to vote against the other alternatives. If x wins it is elected, if not, the winner is decided by a pairwise vote between y and z. This agenda is denoted $([x]\,yz)$, and is shown in Figure 3.2.

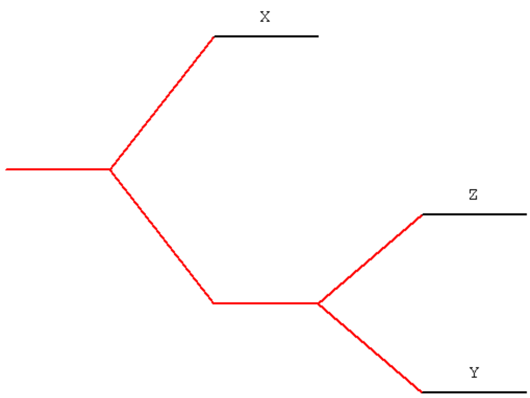

Figure 3.2: An elimination agenda with three alternatives

The expected utility of voting for the upper branch is $U_i\,(x)$, and the expected utility of voting for the lower branch is $p_i\,(yBz)\,U_i\,(y)+[1 - p_i\,(yBz)]\,U_i\,(z)$. A voter thus votes for the upper branch if

$$U_i\,(x) \geq p_i\,(yBz)\,U_i\,(y) + [1 - p_i\,(yBz)]\,U_i\,(z)\,. \tag{3.2}$$

Voter types 1, 2, 4, and 6 have dominant strategies to vote sincerely (see e.g., Miller 1995, pp. 48-52).

3.3 The logic of the welfare consequences of strategic voting: three examples

In the examples that follow, we will say that a voter's beliefs are *reasonable* if they could have been derived from relatively reliable signals. The purpose of using the term "reasonable" here is that it is merely a shorthand for "could have been derived from relatively reliable signals". The examples below are meant to provide an intuitive understanding of how the quality of the beliefs affects the voting outcomes. The term "reasonable beliefs" does not have any role in the theory and nothing depends on it. It is thus introduced merely for the heuristic purpose of making the logic of the model more salient. The following examples involve only three voters, but the model of incomplete information is based on applying the Central Limit Theorem, and it is thus not directly applicable for a society of three voters.[11] This is why the signal extraction model is not used in discussing the examples here. It is hoped, however, that these examples provide the reader with an easier access to the intuition of the model than one with a large amount of voters. The signal extraction model is introduced in section 3.4. Section 3.5.1 will then present another example with 29 voters in which this signal extraction model is used for determining voters' beliefs.

If the Condorcet winner (CW) is not the same alternative as the utilitarian winner (UW), the latter ought to be selected according to the utilitarian welfare criterion. Strategic voting may lead to the choice of the UW even if some other alternative is a CW, but this usually requires that most voters' beliefs are reasonable.

Example 1 illustrates such a situation. Assume that the preferences of three voters A, B, and C, can be described with Table 3.2.

A	B	C
y (1)	y (1)	x (1)
x (0.9)	x (0.9)	z (0.9)
z (0)	z (0)	y (0)

Table 3.2: Example 1

x is the utilitarian winner here. The numbers in parentheses denote voters' utilities for the alternatives. If they vote sincerely, y will beat x in the first round and z in the second round, and the Condorcet winner, y, is chosen.

Assume now that all three voters have identical beliefs such that $p(xBz) = 0.9$, and $p(yBz) = 0.7$. Voters thus consider it likely that y beats z, but even more likely that x beats z in the second round of voting.

Let $U^t(j)$ denote a type t voter's utility for alternative j. Voters A and B are of type 5. They will vote strategically for x in the first round if $p(xBz)\,U^5(x) +$

[11] This approximation restricts the applicability of the model to situations with a fairly large number of voters. Thirty observations is sometimes given as a very rough guess on the validity of the Central Limit Theorem.

$0 > p\,(yBz) \cdot 1 + 0 \quad \leftrightarrow U^5(x) > \frac{p(yBz)}{p(xBz)}$, i.e. if $0.9 > \frac{0.7}{0.9} = 0.7778$. Since this is true, A and B will vote strategically for x in the first round of voting. Voter C has a weakly dominant strategy to vote for x in the first round of voting. x is the outcome if all voters maximise expected utility because it beats y in the first round and z in the second round. The utilitarian winner x is thus chosen if voters maximise expected utility, but the Condorcet winner y is chosen if all voters vote sincerely. Example 1 also shows that a Condorcet winner is not necessarily chosen in majority rule, and that this may happen under fairly reasonable assumptions on voters' beliefs. This result has already been proven by Ordeshook & Palfrey (1988), but their model is based on implausible assumptions. In particular, given that they assume incomplete information, it is implausible to assume that the players condition their beliefs on the assumption that exactly three of the six possible types of players may be playing the voting game.

Consider now an example in which voters' beliefs are not reasonable. Let the preferences of three voters D, E and F be as follows:

D	E	F
x (1)	y (1)	x (1)
y (0.9)	z (0.9)	z (0.9)
z (0)	x (0)	y (0)

Table 3.3: Example 2

Here $x = UW = CW$. Let us now assume that voters have identical beliefs such that $p\,(xBz) = 0.3$ and $p\,(yBz) = 0.7$. They now believe that z will beat x in the second round even though $x = CW = UW$, and z is the worst alternative in utilitarian terms. $p\,(yBz)$ is reasonable, because y beats z in the second round if it survives the first. Voter D will vote strategically for y in the first round, because $U^1(y) = 0.9$ is larger than $\frac{p(xBz)}{p(yBz)} = 0.428$. Voter E has a weakly dominant strategy to vote for y in the first round. Voter F has a weakly dominant strategy to vote sincerely for x in the first round. Thus, if all voters maximise expected utility, y beats x in the first and z in the second round, and emerges as the outcome. Here strategic voting leads to an outcome (y) which is worse in utilitarian terms than the outcome if all voters vote sincerely (x).

Examples 1 and 2 illustrate that the welfare consequences of strategic voting depend on how accurate voters' beliefs are. If they are clearly inaccurate, as in example 2, strategic voting can diminish welfare, but if they are relatively accurate, as in example 1, strategic voting may increase welfare. If voters have complete information, the Condorcet winner wins in both cases. Hence, strategic voting with incomplete information may increase welfare when compared to strategic voting with complete information (example 1). However, strategic voting with complete information never has the catastrophic consequences that strategic voting with incomplete and poor information may have (example 2).

These examples also show that if a voter thinks that her information is highly unreliable, she *should not* take the risk of voting strategically because she might well obtain a worse outcome for herself. In example 2, voter D obtained a worse outcome (y) by voting strategically than she would have obtained if she had voted sincerely (x). Furthermore, the *benefit* from this "foolish" strategic voting *accrued to voter E (who voted sincerely)*, who obtained a better outcome than she would have obtained if D had voted sincerely.

In example 1, the strategic voting of A and B resulted in an outcome that has a lower utility for them than the alternative that would have been chosen if they had voted sincerely. Nevertheless, their actions increased the average utility of *all* voters because voter C's utility increases more than their own utility decreases. Hence, EU behaviour may be welfare-increasing on the aggregate level even though those who vote strategically may diminish their own utility.

Strategic voting may also be welfare-increasing and increase the utility of those who engage in it. The famous Condorcet paradox in example 3 illustrates such a case.

G	H	I
x (1)	y (1)	z (1)
y (0.8)	z (0.1)	x (0.9)
z (0)	x (0)	y (0)

Table 3.4: Example 3

If voters engage in SV behaviour, x beats y in the first round, and z beats x in the second. If they maximise expected utility with $p\,(xBz) = 0.1$ and $p\,(yBz) = 0.9$, G votes strategically for y in the first round ($0.8 > \frac{0.1}{0.9}$), and the others continue to vote sincerely. y beats x in the first round and z in the second. Voter G obtains a better result for herself than she would have obtained by voting sincerely. y is also better than z in terms of the sum of utility. Notice, however, that y is not a utilitarian winner. Strategic voting resulted in a clearly better outcome than sincere voting, but the utilitarian winner was not selected.

3.4 A model of incomplete information in simulated voting games

The previous section showed that in some situations strategic voting is welfare-increasing and in some others it is not. These examples may provide some insight into the logic of strategic voting, but it will be important to know whether strategic voting is *typically* beneficial or not. The examples were also silent on how voters are assumed to formulate their beliefs. Let us now give an account of the voters' beliefs in a framework of simulated voting games.

A standard Bayesian model of incomplete information would assume that the players start with common priors and update them with Bayes' rule. Voters may be able to update their beliefs after the first round of voting, but they are not able to *benefit* from these updated beliefs when there are only three alternatives because all voters vote sincerely in the second round of voting. The model can be extended to four or more alternatives, but introducing an updating model is beyond the scope of this paper because four or more alternatives also bring other complications that should be dealt with.[12]

A model that starts with common priors does not provide interesting results under amendment agendas with three alternatives because all priors before the first round of voting would be equal to $\frac{1}{2}$ if voters knew that all voter types are equally likely. It can be checked that all voters will vote sincerely if this value $\frac{1}{2}$ is inserted into the condition that determines strategic voting presented in equation 3.1. For these reasons, voters need to have some information on the preferences of the other voters *before* the *first* round of voting.

Voters are thus assumed to obtain *perturbed signals* of the other voters' preferences before the first round of voting. They formulate beliefs on the basis of these noisy signals. This information model is embedded in simulated games for which the voter types are generated with the *impartial anonymous culture* (IAC) assumption. This assumption means that each voter type is equally likely. If the preferences for a pair of alternatives j and k is considered, it means that each voter is equally likely to prefer j to k as the reverse. The IAC assumption over-emphasises the prevalence of strategic voting when compared to real-world situations. The use of this assumption is legitimate in this model, however, because the purpose is not to evaluate how common strategic voting is, but rather what its consequences are when it occurs and is significant. The IAC assumption is the best possible assumption for this purpose because it generates the maximum amount of very tight elections and thereby a maximum amount of cases in which strategic voting matters.[13]

A *simulated game* g consists of a set of utilities created by a random number generator, beliefs based on these utilities, voters' perturbed signals, and voting outcomes under the different behavioural assumptions. Let $\succ_i^g$ denote voter i's preference relation in a simulated game g. Let $n^g(j \succ k)$ denote the number of voters who prefer alternative j to alternative k in simulated game g, and $n^g(k \succ j)$ the amount of voters with opposite preferences. If alternatives j and k are put to vote against each other in the last round, j beats k if $n^g(j \succ k) > n^g(k \succ j)$.

Since all the symbols to be defined in what follows concern a single simulated game g, the superscript will be omitted in the sequel. Let $n_i(j \succ k) = 1$, if voter i prefers j to k, and $n_i(j \succ k) = 0$, if voter i prefers k to j. Then $n(j \succ k)$ can be viewed as a sum of N Bernoulli trials. The total number of supporters for j against k is thus given by $n(j \succ k) = \sum_{i=1}^{N} n_i(j \succ k)$. Let p denote the probability that such a Bernoulli trial results in the outcome that $n_i(j \succ k) = 1$. The impartial culture implies that $p = \frac{1}{2}$. $n(j \succ k)$ can thus be viewed as a random variable with a binary distribution $n(j \succ k) \sim B(N, \frac{1}{2})$.

[12] It is argued in (Lehtinen 2002) that updating is difficult even if there are more than three alternatives. See also Enelow & Hinich (1983).

[13] See Krehbiel & Rivers (1990), Eckel & Holt (1989), Calvert & Fenno (1994), Volden (1998), Wilkerson (1999), and Gilmour (2001) for discussions on the prevalence of strategic voting. See Tsetlin, Regenwetter & Grofman (2003) and Gehrlein (2002) for recent discussions of impartial culture.

3.4.1 Signals

The voters are assumed to obtain a perturbed signal of the number of voters who prefer j to k. It will be more convenient to use a standardized sum of Bernoulli trials $Q(j \succ k)$ instead of the variable $n(j \succ k)$ itself:

$$Q(j \succ k) = \frac{n(j \succ k) - Np}{\sqrt{Np^2}}. \tag{3.3}$$

Since $p = \frac{1}{2}$, this is $Q(j \succ k) = \frac{2n(j \succ k) - N}{\sqrt{N}}$. A *signal* of voter i concerning the preferences of all voters for alternatives j and k , $S_i(j,k)$, is given by

$$S_i(j,k) = \frac{2n(j \succ k) - N}{\sqrt{N}} + \varepsilon \cdot r_i\,(j,k)\,, \tag{3.4}$$

where $r_i\,(j,k)$ is a realization of an i.i.d. standard normal random variable, and ε is a scaling factor that reflects the *reliability* of the signals. Let $R_i\,(j,k) = \varepsilon \cdot r_i\,(j,k)$. The signal can then be written as follows:

$$S_i(j,k) = Q(j \succ k) + R_i\,(j,k)\,. \tag{3.5}$$

The brief term "signal" is used here, even though the longer expression "a voter's conception of an aspect of the game to be played" would be more accurate. A voter's conception of the game may be the result of several observations.

A signal is the only constraint imposed on a voter's beliefs. In particular, beliefs that constitute a cycle are allowed; $p_i\,(xBz) > \frac{1}{2}$, $p_i\,(zBy) > \frac{1}{2}$, and $p_i\,(yBx) > \frac{1}{2}$. The reason for this is that if the underlying preferences are cyclical, the beliefs for them may well be cyclical as well.

Deriving beliefs from these signals involves applying the Central Limit Theorem and standard statistical inference. Voters are thus modelled as amateur econometricians involved in a *signal extraction* problem. Lehtinen (2006*a*) shows that voters' beliefs are given by equations (3.6) and (3.7).

$$p_i(xBz) = 1 - \Phi\left(\frac{-s_i(x,z)}{\varepsilon\sqrt{1+\varepsilon^2}}\right), \tag{3.6}$$

and

$$p_i(yBz) = 1 - \Phi\left(\frac{-s_i(y,z)}{\varepsilon\sqrt{1+\varepsilon^2}}\right). \tag{3.7}$$

Voters are assumed to know that the voter types are drawn from a uniform distibution. Hence, they cannot use their own type for deriving a belief about others because their own type does not provide them with new information.

Let us say that ε is the *reliability of the signals*. The smaller ε is, the more reliable a voter's signals are. In this paper, voters are assumed to know the reliability of their signals. This assumption can be relaxed as explained in Lehtinen (2006*a*).

3.5 Counterbalancing of strategic votes

The mechanism of *counterbalancing* strategic votes explains when and why strategic voting is welfare-increasing. Four different types of voters may vote strategically under amendment agendas. Voters of types 5 and 6 may vote strategically for x, while voters of types 1 and 3 may vote strategically for y. Let us now reformulate equation 3.1 as follows:

$$\begin{aligned} L_i = {} & p_i\,(xBz)\,U_i(x) + (1 - p_i\,(xBz))U_i(z) \\ & - p_i\,(yBz)\,U_i(y) - (1 - p_i\,(yBz))U_i(z). \end{aligned} \quad (3.8)$$

This equation says that if $L_i \geq 0$, the voter votes for the upper branch (x). It is easy to see that $\frac{\partial L_i}{\partial U_i(y)} < 0$, and that $\frac{\partial L_i}{\partial U_i(x)} > 0$. The signs of these derivatives mean that the higher is the utility of y for voters of type 1 and 3, the more likely they are to vote strategically for y. Similarly, the higher the utility of x for voters of type 5 and 6, the more likely they are to vote strategically for x.

Hence, if the utility for x is *almost as high* as the utility of y for many voters of types 5 and 6, and if the utility of y is *significantly lower* than the utility of x for many voters of types 1 and 3, a larger number of voters of types 5 and 6 than of types 1 and 3 vote strategically. This means that x gets *more* strategic votes than y. Furthermore, strategic votes for x are at the same time strategic votes *against* y.

Ceteris paribus, if many $U^5(x)$ and $U^6(x)$ are almost as high as $U^5(y)$ and $U^6(y)$, respectively, and if many $U^1(x)$ and $U^3(x)$ are significantly higher than $U^1(y)$ and $U^3(y)$, respectively, the sum of utility for alternative x is relatively large, and the sum of utility for alternative y is relatively small. Hence, under these assumptions on individual utilities, x is likely to have a larger sum of utility than y. Counterbalancing means that both x and y will obtain strategic votes, but x is likely to obtain more strategic votes than y if it has a larger sum of utilities than y.

3.5.1 An example of counterbalancing

Consider now an example that purports to show how counterbalancing affects the voting results. There are 29 voters whose utilities are the result of a simulation. Their signals were formulated with $\varepsilon = 1$. Table 3.5 on page 23 displays voters' types (t), decisions (D), preference intensities $U_{2,i}$, beliefs (p_i(xBz) and p_i(yBz)), perturbation terms (R_i(x,z) and and R_i(y,z)), and expected utilities for the two branches of a voting tree (EU_i(U) for Upper (a vote for x) and EU_i(L) for Lower (a vote for y)). When a voter votes sincerely $D = S$, and when a voter votes strategically $D = T$.

The sums of utilities are $U\,(x)$=15.43, $U\,(y)$=13.88, and $U\,(z)$=12.85. y is the Condorcet winner because

$n_1 + n_3 + n_4 = n(x \succ y) = 6 + 2 + 5 = 13,$
$n_2 + n_5 + n_6 = n(y \succ x) = 7 + 4 + 5 = 16,$
$n_1 + n_4 + n_5 = n(x \succ z) = 6 + 5 + 4 = 15,$

no.	t	D	$U_{2,i}$	p_i(xBz)	p_i(yBz)	$R_i(x,z)$	$R_i(y,z)$	EU_i(U)	EU_i(L)
1	1	S	0.22	0.41	0.86	-0.53	0.59	0.41	0.19
2	1	T	0.6	0.37	0.62	-0.66	-0.49	0.37	0.37
3	1	S	0.42	0.68	0.52	0.47	-0.87	0.68	0.22
4	1	S	0.23	0.96	0.79	2.3	0.22	0.96	0.18
5	1	S	0.25	0.7	0.7	0.57	-0.18	0.7	0.17
6	1	S	0.58	0.76	0.83	0.83	0.44	0.76	0.49
7	2	S	0.67	0.49	0.96	-0.22	1.55	0.34	0.99
8	2	S	0.56	0.86	0.4	1.35	-1.3	0.08	0.73
9	2	S	0.66	0.66	0.96	0.41	1.63	0.22	0.99
10	2	S	0.81	0.53	0.86	-0.08	0.6	0.38	0.97
11	2	S	0.57	0.44	0.88	-0.41	0.71	0.32	0.95
12	2	S	0.39	0.41	0.83	-0.52	0.41	0.23	0.90
13	2	S	0.14	0.68	0.82	0.49	0.36	0.04	0.84
14	3	S	0.74	0.46	0.32	-0.33	-1.58	0.88	0.68
15	3	S	0.61	0.95	0.96	2.14	1.54	0.63	0.04
16	4	S	0.36	0.9	0.55	1.6	-0.76	0.93	0.16
17	4	S	0.05	0.44	0.73	-0.41	-0.04	0.47	0.01
18	4	S	0.5	0.58	0.87	0.09	0.66	0.79	0.07
19	4	S	0.8	0.86	0.45	1.37	-1.09	0.97	0.44
20	4	S	0.34	0.78	0.76	0.9	0.07	0.85	0.08
21	5	T	0.72	0.83	0.5	1.18	-0.93	0.6	0.50
22	5	T	0.88	0.66	0.52	0.4	-0.85	0.58	0.52
23	5	S	0.85	0.18	0.86	-1.48	0.6	0.15	0.86
24	5	S	0.63	0.31	0.46	-0.88	-1.09	0.2	0.46
25	6	S	0.53	0.68	0.74	0.47	-0.01	0.32	0.65
26	6	T	0.42	0.38	0.86	-0.63	0.63	0.62	0.50
27	6	T	0.03	0.13	0.4	-1.77	-1.28	0.87	0.61
28	6	S	0.08	0.36	0.35	-0.71	-1.47	0.64	0.68
29	6	S	0.52	0.74	0.33	0.73	-1.54	0.26	0.84

Table 3.5: Example 4

$n_2 + n_3 + n_6 = n(z \succ x) = 7 + 2 + 5 = 14$,
$n_1 + n_2 + n_5 = n(y \succ z) = 6 + 7 + 4 = 17$, and
$n_3 + n_4 + n_6 = n(z \succ y) = 2 + 5 + 5 = 12$.

The standardized numbers of voters are $Q(x \succ z){=}\frac{2n^g(x \succ z)-N}{\sqrt{N}}{=}\frac{2 \cdot 15-29}{\sqrt{29}}{=}$ $0.185\,70{\approx}0.19$, and $Q(y \succ z) = \frac{2n^g(y \succ z)-N}{\sqrt{N}} = 0.928\,48 \approx 0.93$. If a voter would have obtained a *perfectly reliable signal* ($\mathrm{R}_i{=}0$), he or she would have formulated the following probabilities $p\,(xBz){=}1 - \Phi\left(\frac{-0.19}{1\sqrt{1+1^2}}\right){=}0.55$, and $p\,(yBz){=}1 - \Phi\left(\frac{-0.93}{1\sqrt{1+1^2}}\right){=}0.74$. Probabilities that are *close* to these values could be considered "reasonable". It should now be easier to understand why the inexact notion of reasonable beliefs was used and what it could mean. One might argue that reasonable beliefs are those that correspond to reality, and that this would mean that reasonable beliefs must be degenerate zeros or ones. But if you know that the signals on which your probabilities are based are not fully reliable, it is not rational to assign probabilities one and zero to anything of concern to you. Furthermore, we have seen that if a voter engages in strategic voting with poor information, she may lose rather than gain in utility by doing so. It is natural to take perfectly reliable signals as a measuring rod for what counts as a reasonable belief. A voter's belief is the more reasonable, the closer her signals are to being perfectly reliable. It seems plausible to say that there is a continuum of beliefs from reasonable to (highly) unreasonable between the extremes of, say, $(\varepsilon = 0.01, R = 0)$ and $(\varepsilon = 100,\, R = 100)$ or $(\varepsilon = 100,\, R = -100)$, even though there are no non-arbitrary values of R_i and ε that make a belief based on these parameters reasonable.

To see how the actual beliefs are derived in this example, consider voter 2 as an example. Applying equation 3.4, it is seen that $s_i(x, z) = \frac{2n(x \succ z)-N}{\sqrt{N}} + \varepsilon \cdot r_i\,(x, z) = 0.19{+}1{\cdot}(-0.66) = -0.47$. Applying equation 3.6 it is seen that $\mathrm{p}_2(\mathrm{xBz}){=}1 - \Phi\left(\frac{-s_i(x,z)}{\varepsilon\sqrt{1+\varepsilon^2}}\right) = 1 - \Phi\left(\frac{0.47}{\sqrt{2}}\right) = .3698 \approx .37$. A similar calculation applies to $\mathrm{p}_2(\mathrm{yBz})$.

This example is analogous to example 1 in that the Condorcet winner y is chosen with SV behaviour, but the utilitarian winner x is chosen with EU behaviour. Voter 2 gives a strategic vote for y, but this is counterbalanced by four strategic votes for x by voters 21, 22, 26, and 27. The fact that x receives more strategic votes is not a coincidence. The average preference intensity for x (0.7383) is clearly higher than that for y (0.3527). In contrast, the perturbations are distributed relatively equally for all voter types. What matters for the voter's choice is not only the size of the perturbations, but also whether the perturbation for $Q(x \succ z)$ mutually reinforces the perturbation for $Q(y \succ z)$, i.e. whether the sign of the two perturbations is the same or not. If $|R\,(x, z) - R\,(x, z)\,| > 1$ is taken as a criterion, voters 1, 3, 4, 7, 8, 9, 11, 14, 16, 19, 21, 22, 23, 26, and 29 have mutually reinforcing perturbations.

Of these, voters 1, 14, 21, 22, and 26 have perturbations that increase the probability of voting strategically as compared with zero perturbations. Considering only the beliefs, voters 1 and 14 *could have voted strategically* for y, but they voted sincerely. Notice that voter 1's intensity for y (0.22) is relatively low, and voter 14's intensity for x (0.74) is relatively high. In contrast, voters 21, 22, and 26 do vote strategically because the intensities for x ($U_{21}(x) = 0.72$, and $U_{22}(x) = 0.83$) are relatively high, and the intensity for y ($U_{26}(y) = 0.42$) relatively low. Counterbalancing thus implies that alternatives with high average utility will get more and lose less strategic votes than other alternatives.

3.6 Simulation and setups

A *simulated EU-game* g consists of a profile of utilities, $\Pi^g(\Psi) = \{U_1^g, U_2^g,...,U_N^g\}$, as determined by a rule Ψ, and a profile of beliefs computed on the basis of ε and $\Pi^g(\Psi)$. All simulations had $N = 201$ voters.

An expected utility *setup* (EU-setup) is a collection of assumptions $S = \{\mathbb{I}, X, \Pi(\Psi), A, \varepsilon, C, IPC\}$. There are $G = 10000$ simulated games in a setup. $\mathbb{I} = \{I^1,I^2,...,I^g,...,I^G\}$ is a collection of G sets of voters, and $\Pi(\Psi) = \{\Pi^1,\Pi^2,...,\Pi^G\}$ is a collection of utility profiles, one set for each simulated game. A is an agenda. C and IPC denote parameters that will be explained shortly.

In what will be called *uniform setups*, the rule Ψ that determines voters' types and preference intensities is a combination of the impartial culture assumption and the assumption that the utilities are derived from a uniform distribution on [0,1]. Since the logic of counterbalancing suggests that strategic voting should be more welfare-increasing if there are systematic differences between voters' relative utilities that are not reflected in the preference orderings, setups in which the preference intensities for the second-best alternatives are systematically different for different voter types will be studied. In order to generate such *setups with correlation between preference intensities and voter types* without affecting the interpersonal comparisons or the preference orderings, the individual utilities were derived in the following way.

U_1, U_2, and U_3 were first generated from the uniform distribution on [0,1] for each voter, but U_2 was not used for any purpose. Instead, a standardized utility $\widetilde{U}_2$ for the second-best alternative was generated from the uniform distribution on [0,1]. This standardized utility expresses what a voter's utility for the second-best alternative would be if his or her scale of utility was [0,1]. These standardized second-best utilities will be referred to as *intra*personal *intensities.* In setups with intensity correlation, these intensities were multiplied with a parameter C, $0.5<C\leq1$ for those who put y second (voter types 1 and 6) so that the new correlated intensities $\widetilde{U}_2^{C,1}$ and $\widetilde{U}_2^{C,6}$ were given by

$$\widetilde{U}_2^C = C\widetilde{U}_2.$$

To compensate the decreases in utility for voter types 1 and 6, the intensities for voters of type 3 and 5 (i.e. for x) were given by

$$\widetilde{U}_2^C = 1 - C\widetilde{U}_2.$$

These adjustments make the average utilities for x higher and the average utilities for y lower than in the uniform setups while keeping the overall average utility fixed. In uniform setups, $C = 1$. C thus denotes the *degree of correlation* between preference intensities and voter types.

These standardized intensities were then scaled back into the original $[U_3, U_1]$ utility scale. Let U_2^* denote a voter's intensity expressed in terms of the original $[U_3, U_1]$ scale. Since the relationship between the standardized intrapersonal utility for the second-best alternative and the original scale of utility is given by

$$\widetilde{U}_2^C = 1 - \frac{U_1 - U_2^*}{U_1 - U_3}, \tag{3.9}$$

U_2^* is given by:

$$U_2^* = U_3 + \widetilde{U}_2^C (U_1 - U_3). \tag{3.10}$$

3.6.1 Criteria for evaluating the welfare consequences of strategic voting

The shorter expression $\mathcal{S}(\varepsilon,\ C)$ will be used to refer to an EU-setup, because an EU-setup is essentially a set of simulated games in which the reliability of signals ε and the degree of correlation C are the same for all voters. The winner of voting is denoted W_{SV}^g in a simulated SV-game, and $W_{EU}^g(\varepsilon, C)$ in a simulated EU-game. Let $U^i(W_{EU}^g, \varepsilon, C)$ and $U^i(W_{SV}^g)$ denote voter i's utility in simulated game g in an EU-setup and a SV-setup, respectively. The *Average Utility in an EU-setup* $\mathcal{S}(\varepsilon, C)$, $AU_{EU}(\varepsilon)$, is:

$$AU_{EU}(\varepsilon, C) = \frac{\sum_{g=1}^{G} \sum_{i=1}^{N} U^i(W_{EU}^g, \varepsilon, C)}{G * N}. \tag{3.11}$$

The *Average Utility in the SV-setup*, AU_{SV}, is:

$$AU_{SV} = \frac{\sum_{g=1}^{G} \sum_{i=1}^{N} U^i(W_{SV}^g)}{G * N}. \tag{3.12}$$

EU behaviour is **welfare-increasing** *in setup $S(\varepsilon, C)$ if the average utility of all voters is larger in this EU-setup than in the SV-setup:*

$$AU_{EU}(\varepsilon, C) > AU_{SV} \tag{3.13}$$

If the converse holds, EU behaviour is **welfare-decreasing**. Let us also say that *strategic voting is welfare-increasing* in a setup if EU behaviour is welfare-increasing in that setup.

3.7 Simulation results

3.7.1 Amendment agendas

Figure 3.3 displays average utilities from setups with ε = [0,0.4,..., 1.6] and C=[1,..,0.5].[14]

[14] The results are presented only as graphs here. All numerical results in tabular form, as well as the FORTRAN codes to generate them are available from the author on request.

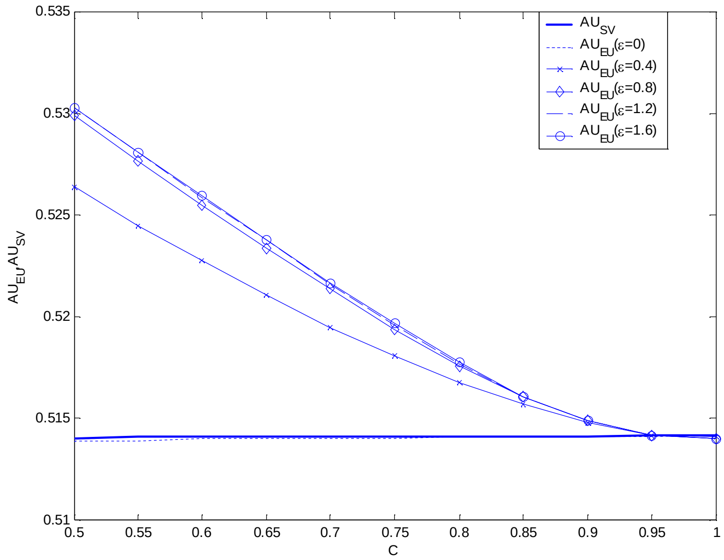

Figure 3.3: Average utilities in amendment agendas

Since the variance of $Q(j \succ k)$ is 1, the reliability of the voters' signals with $\varepsilon = 1.6$ is very low; chance is more important in determining the signal than the real preference profile in setups with $\varepsilon > 1$. Such a large range of parameter values were studied in order to ensure that the relevant parameter range, and more, is covered.

The following observations can be made from these simulation results. EU behaviour increases welfare in almost all setups. In uniform setups the average utilities are virtually the same under the two behavioural assumptions. As expected, welfare-increasing strategic voting becomes more and more important as the correlation between voter types and preference intensities increases. EU behaviour with complete information ($\varepsilon = 0$) yields lower average utilities than EU behaviour with incomplete information. As long as information is not complete, the quality of voters' information does not seem to be particularly important for the results. In fact, the average utilities are highest when the perturbations are large and when the correlation is strong.

What happens if the intensity for z rather than y is decreased (or increased) in setups with intensity correlation? The results from such a setup are presented in Figure 3.4.

The difference in average utilities between SV behaviour and EU behaviour is now considerably lower. Furthermore, the average utilities are lower under both behavioural assumptions. These results can be explained as follows. Since

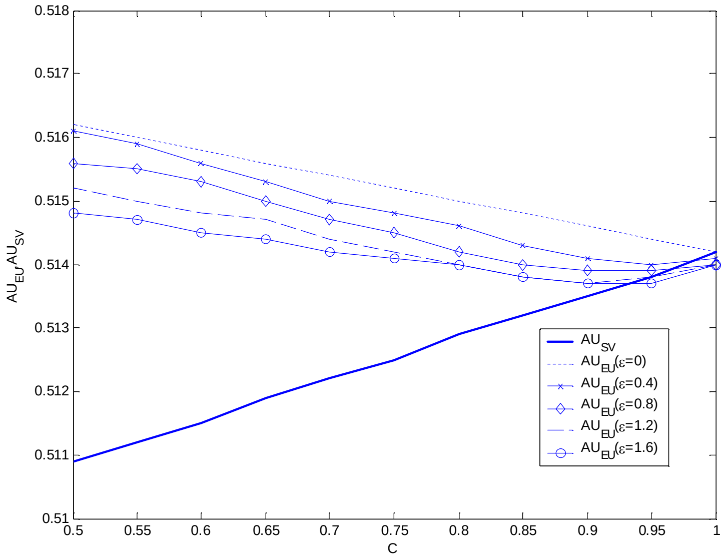

Figure 3.4: Average utilities in setups in which the intensities of x are high and the intensities of z low.

the utility of z is low, it is natural that the average utilities are lower; z always participates in the second-round contest, and wins one-half of them. The difference in average utility between the two behavioural assumptions is now lower because in the setups with low utilities for y, strategic voting is effective in eliminating y in the first voting round, but in the latter setups this matters less because the low-utility z is always waiting in the second round of voting. If the roles of z and x are reversed by decreasing the intensities for x and increasing the intensities for z, the results are again similar to the ones presented in Figure 3.3. They are presented in Figure 3.5.

As expected, the average utilities under EU behaviour remain similar to what they were in previous setups, but now the average utility under SV behaviour increases slightly with an increase in the degree of correlation.

3.7.2 Counterbalancing once again

Uniformly distributed preference intensities generate very small differences in intensities between the different voter types. This is why uniform setups provide the *least favourable* comparison between EU behaviour and SV behaviour.

It can be seen from Figure 3.3 that even a very weak correlation between intensities and voter types makes strategic voting welfare-increasing in all setups.

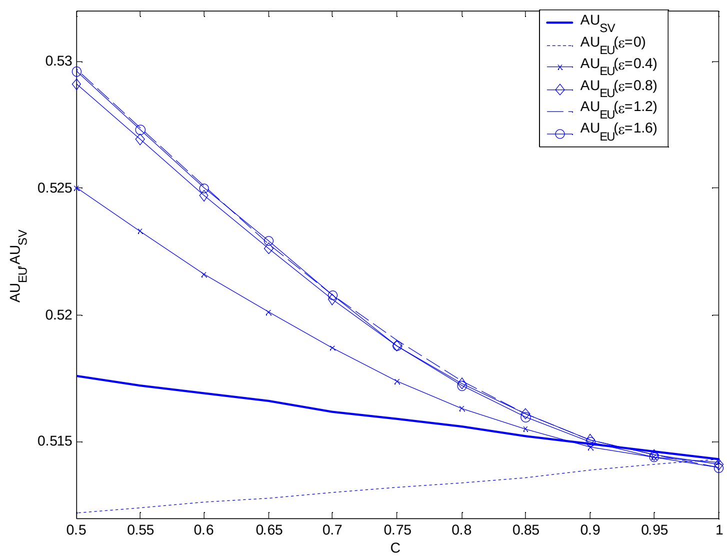

Figure 3.5: Average utilities in amendment agendas when the intensities of z are high and the intensities of x are low

It is reasonable to assume that typically some amendments are widely endorsed as second-best alternatives. The setups with somewhat high correlation may well represent the reality better than the setups with weak correlation.

Let us now look at the logic of counterbalancing by considering some comparisons between uniform and correlated setups. Let $V(\sim x)^m$ denote the number of voters who prefer x to y, but who vote for y in simulated game m. The average percentage of votes against candidate x is the relative frequency of voters who prefer x to y but who vote strategically for y. The average percentage of votes *against* alternative x, AX, is thus given by

$$AX = \sum_{m=1}^{G} \frac{V(\sim x)^m}{G * N} * 100. \tag{3.14}$$

Let $V(\sim y)^m$ denote the number of voters who prefer y to x, but who vote for x in simulated game m. The average percentage of votes *for* alternative x, FX, is given by

$$FX = \sum_{m=1}^{G} \frac{V(\sim y)^m}{G * N} * 100. \tag{3.15}$$

Since a strategic vote for x is simultaneously a strategic vote against y, AX and FX also provide the percentages of strategic votes for and against y. Figure 3.6

displays the average percentages of strategic votes for and against x in various setups.

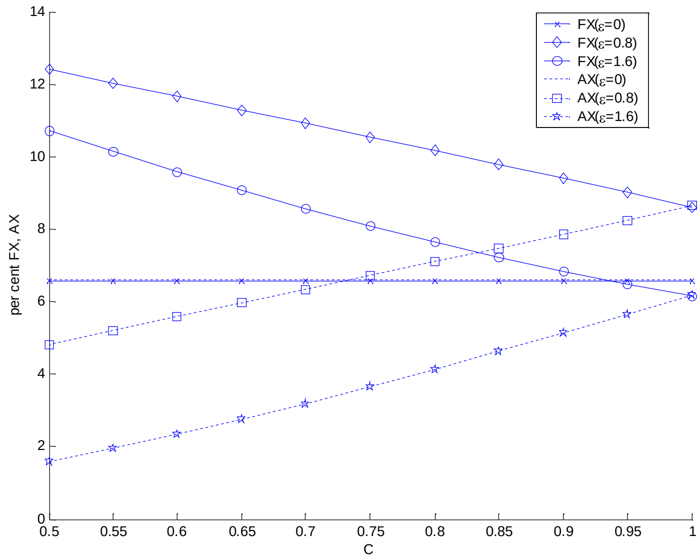

Figure 3.6: Strategic votes for and against x and y in various setups

It is to be expected that in uniform setups where all preference intensities are taken from the uniform distribution, all candidates should obtain and lose about the same amount of strategic votes. Figure 3.6 shows that this is indeed the case. Furthermore, the more there is correlation between voter types and preference intensities, the more y loses and the more x gains strategic votes.

3.7.3 Robustness with respect to interpersonal comparisons

Since the results are based on average utilities, it is necessary to make interpersonal comparisons of utilities. Furthermore, *interpersonal comparisons of preference intensities* are also needed because it is necessary to assume that one person's utility may be added to another person's utility. In the simulations conducted thus far, random interpersonal comparisons of preference intensities have been used because the utilities have been derived from the uniform distribution on the [0,1] interval. This particular assumption creates some variation in the minimum and maximum values of utilities for different voters. If it is considered likely and important that different individuals in fact attach different importance to the different issues, this way of modeling is justifiable. Another possibility is to derive the utilities in such a way that the maximum and minimum utilities are given the values 1 and 0, respectively, and the utility for the second-best alternative is something in between these extremes. This way of modeling may be justified on the normative grounds that each voter should have the same weight in determining the best outcome. It could be seen as an expression of the one-man one-vote principle that takes preference intensities into account.

Irrespective of the way of modeling chosen, it may be argued that our choice of interpersonal comparisons is arbitrary. This arbitrariness ultimately derives from the fact that it is impossible to obtain exact information on individual differences in utilities. Epistemological considerations thus indicate that we will never know which interpersonal comparison is correct. Unfortunately, the results depend crucially on interpersonal comparisons of preference intensities. These are generally considered as the most suspect kinds of comparisons.

Strategic voting is beneficial *only* because it allows voters to express intensities indirectly even in voting rules in which such information is not explicitly collected. Therefore, if only ordinal welfare measures are used, it is to be expected that strategic voting is welfare-decreasing. Consider, however, what using only ordinal welfare measures implies if the expected utility model of voter behaviour is accepted. It implies that using intensity-based welfare measures are not accepted even though one acknowledges the relevance of intensities for individual voters' behaviour. But if intensities are important for the individuals, they should be normatively important for the whole electorate.

Fortunately, it is possible to accommodate the criticism that our choice of interpersonal comparisons is arbitrary. If the result that strategic voting increases average utility obtains with *all* different and at least mildly reasonable interpersonal comparisons, then this result does not depend on any particular interpersonal comparison. If the result is *robust* to interpersonal comparisons in such a way, we can be assured that we know something more about the consequences of strategic voting even though we do not know which interpersonal comparison is correct.

Several different variations on interpersonal comparisons were thus tried in order to see whether the results are robust or not.[15] In order to retain comparability to previous results, all these variations need to change interpersonal comparisons without changing the preference orderings, the *intra*individual preference intensities, or the average utility of all alternatives. It is thus necessary to hold the parameters that determine individual *behaviour* fixed in evaluating robustness to interpersonal comparisons.

One interpersonal comparability variation is to preserve the original preference orderings and relative intraindividual intensities but redraw the minimum and maximum values for the utility scales (i.e. U_3 and U_1) randomly from the same uniform distribution as before. The results from this variation are almost identical to those presented before, and will therefore not be presented. This

[15] The idea of simulating different interpersonal comparisons was suggested to me by Emrah Aydinonat.

variation is admittedly quite slight because it merely changes the realizations of the random variables in one particular random assignment of utilities.

In order to make more dramatic changes, the utility scales must be changed in such a way that they are systematically different between different voter types. The utilities of voters of types 1, 3, 5, and 6 were again changed. The average utility for each voter type was retained, but the utility scale, i.e. the difference between the maximum and minimum utilities was made smaller (larger) for voters of types 1 and 6, and the utility scale for voters of type 3 and 5 was made larger (smaller). The utility scales of those who put alternative y second were thus shrunk and the utility scales of those who put alternative x second were stretched. Bearing in mind that in setups with correlated intensities the intensities for x are higher than for y on average, this variation effectively diminishes the importance of those who put y second and increases the importance of those who put x second. This variation on interpersonal comparisons will be referred to as the "mutually reinforcing correlation setup" because the *intra*personal intensities are high on average for the *same* voter types whose *inter*personal intensities weigh most in the sum of utilities. A second variation reverses the interpersonal correlation but retains the intrapersonal correlation by stretching the scales for voters of types 1 and 6, and shrinking the scales for voters of types 3 and 5. The second variation will be referred to as the "negative correlation setup".

Let IPC denote a parameter that reflects how much voters' scales are shrunk or stretched. The original utilities are U_1, U_2^*, and U_3. Let $\underline{U}_1$ and $\overline{U}_3$ denote the maximum and minimum utilities for voters of types one and six after their scales have been shrunk ($\underline{U}_1 < U_1$ and $\overline{U}_3 > U_3$). Since the idea is to subtract as much from U_1 as is added to U_3, $\overline{U}_3 - U_3 = U_1 - \underline{U}_1$. $\underline{U}_1$ (and $\overline{U}_3$) is obtained by adding to (subtracting from) the midpoint of the utility scale $\frac{U_1+U_3}{2}$ a part of the individual's scale $\frac{IPC\cdot(U_1-U_3)}{2}$ so that

$$\underline{U}_1 = \frac{[U_1 + U_3 + IPC \cdot (U_1 - U_3)]}{2}, \tag{3.16}$$

and

$$\overline{U}_3 = \frac{[U_1 + U_3 - IPC \cdot (U_1 - U_3)]}{2} = U_1 + U_3 - \underline{U}_1. \tag{3.17}$$

Similarly, let $\overline{U}_1$ and $\underline{U}_3$ denote the maximum and minimum utilities for voters of types three and five after their utility scales have been stretched. The idea now is to add as much, on average, to U_1 as was subtracted from voters of types one and six. Thus, the difference between $\frac{U_1-U_3}{2}$ and $\frac{IPC\cdot(U_1-U_3)}{2}$ is added to the original U_1 so that

$$\overline{U}_1 = U_1 + \frac{(1 - IPC)(U_1 - U_3)}{2} = 2 \cdot U_1 - \underline{U}_1. \tag{3.18}$$

Again it is required that $U_3 - \underline{U}_3 = \overline{U}_1 - U_1$ so that

$$\underline{U}_3 = U_3 - \overline{U}_1 + U_1 = U_3 - U_1 + \underline{U}_1. \tag{3.19}$$

What remains is to rescale the interpersonal intensities in such a way that their *intrapersonal* relative values remain unchanged. Let U_2^{shrink} and $U_2^{stretch}$ denote these two intensities. Then

$$U_2^{shrink} = \overline{U}_3 + U_2^*(\underline{U}_1 - \overline{U}_3), \tag{3.20}$$

and

$$U_2^{stretch} = \underline{U}_3 + U_2^*(\overline{U}_1 - \underline{U}_3). \tag{3.21}$$

In the mutually reinforcing correlation setup voter types 1 and 6 have utilities $(\underline{U}_1, U_2^{shrink}, \overline{U}_3)$, voter types 3 and 5 have utilities $(\overline{U}_1, U_2^{stretch}, \underline{U}_3)$, and voter types 2 and 4 have utilities (U_1, U_2, U_3).

The results from these two setups with $\varepsilon = 0.4$ are displayed in Figures 3.7 and 3.8. The results are similar with other values of ε.

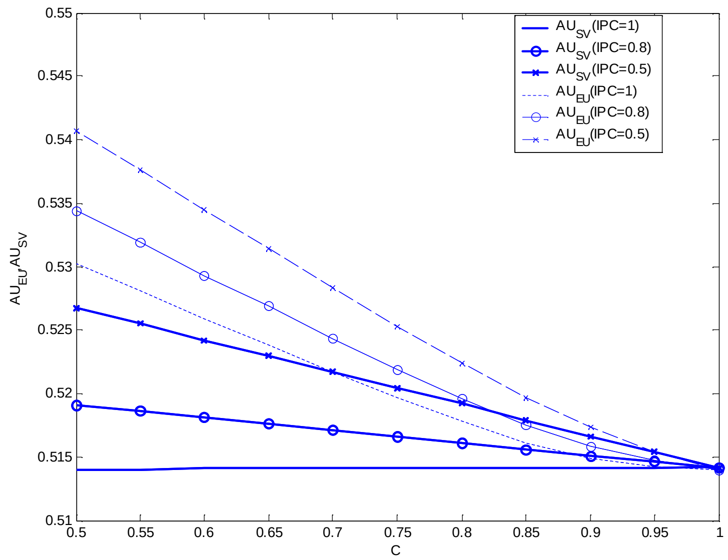

Figure 3.7: Average utilities in mutually reinforcing correlation setups

As expected, the more the interpersonal intensities correlate with the intrapersonal intensities (i.e. the smaller IPC and C are), the higher are the average utilities in the mutually reinforcing correlation setup. This result can be explained as follows. The lower IPC is, the more the utilities of voter types 3 and 5 weigh in the sum of utility. Since these voters put x second, and since the sum of utility of *all* these voters is higher for x than for y or z, the average utility is higher for x than it was without the reinforcing correlation. Since these voters also vote strategically for x, their actions make the average utility relatively high. As Figure 3.8 shows, reversing the interpersonal correlation

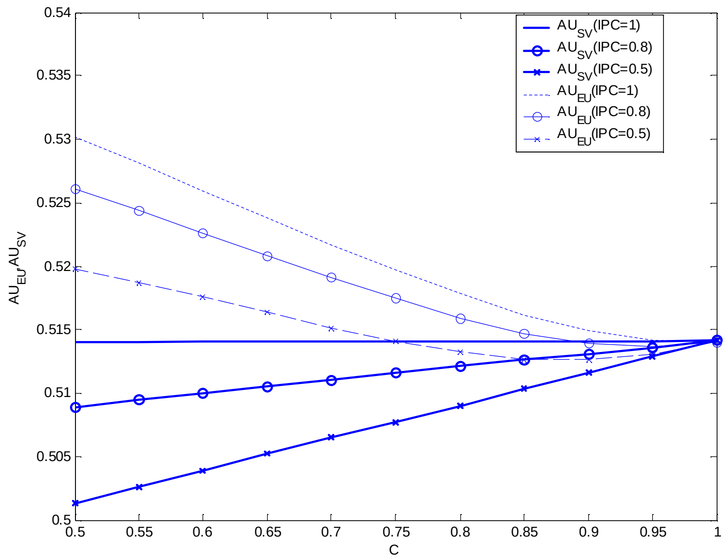

Figure 3.8: Average utilities in setups negative correlation setups

while keeping the intrapersonal correlation makes average utilities lower. Notice, however, that EU behaviour remains welfare-increasing even in negative correlation setups.

Yet another interpersonal comparison consists in making all three utilities higher for some voter types than for some others. In "shift x upwards setups" the utility of voter types 3 and 5 was diminished by substracting the parameter IPC from their utilities and the utility of voter types 1 and 6 was increased by adding IPC to their utilities. In "shift y upwards setups" the roles of the voter types were again reversed. The results from these two setups are displayed in Figures 3.9 and 3.10.

It is easy to see that the intrapersonal differences in these setups are much more important than the interpersonal ones. There is a slight difference however. Shifting the utilities of voter types 3 and 5 upwards, and those of types 1 and 6 downwards increases average utilities slightly. Reversing the voter types has the opposite effect. These results can be explained as follows. Voters of types 3 and 5 put alternative x second. Increasing their utilities increases their weight in the sum of utility. Such a shift slightly increases average utilities under EU behaviour because voter types 3 and 5 also vote strategically for x.

The simulation results from the setups studying different interpersonal comparisons can be summarized as follows. Making different interpersonal comparisons does change the results, but EU behaviour *remains welfare-increasing*

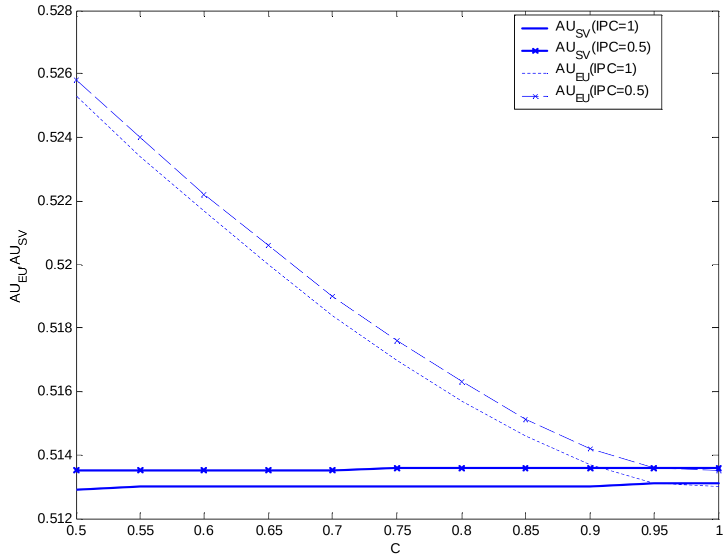

Figure 3.9: Average utilities for setups with an upward shift in utility for voter types 1 and 6

with each different interpersonal comparison. The results are thus robust with respect to interpersonal comparisons.

3.7.4 Elimination agendas

The important difference between elimination agendas and amendment agendas is that in the former fewer voter types may have an incentive to vote strategically. Figure 3.11 displays the simulation results with an elimination agenda $([x]\, yz)$.It is easy to see from this Figure that strategic voting is welfare-increasing in all setups under elimination agendas if the preference intensities for alternative x are systematically higher than for y and z. It is relatively easy to explain why strategic voting is welfare-increasing under elimination agendas when the intensities for x are high. The average utility under SV behaviour is relatively low because alternative x is seldom selected, but, at the same time, there is approximately an equal number of supporters for each of the three alternatives. Hence, under elimination agenda $([x]\, yz)$, strategic voting may cause x to be selected, and this is what increases average utility in the EU-setups when compared to the SV-setup. x is not the only alternative that obtains strategic votes, but under agenda $([x]\, yz)$, it is likely to be the only alternative for which the strategic votes matter; if y or z is intensively preferred, it will be selected also

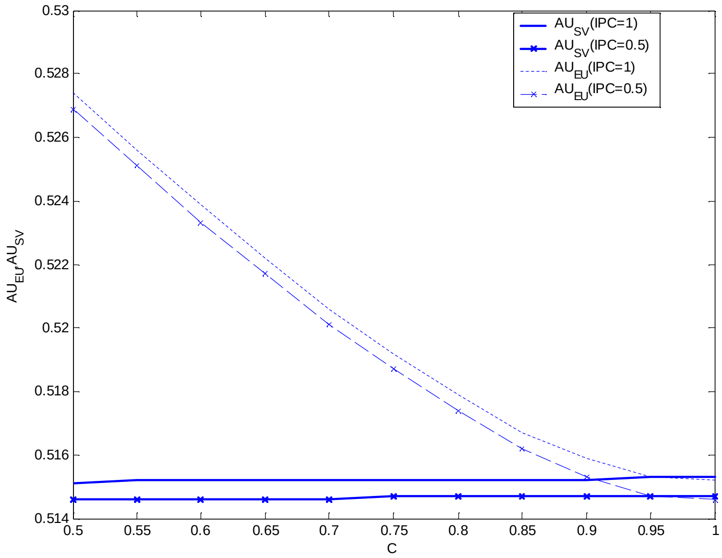

Figure 3.10: Average utilities in setups with an upward shift in the utilities of voter types 3 and 5

under sincere behaviour.

However, if the average utility for x is decreased, and that of y (or z) increased, the results are quite different. Figure 3.12 shows average utilities under elimination agendas when the intensities correlate with the voter types such that the intensities of x are decreased and the intensities of y increased.

EU behaviour is now welfare-increasing only if the degree of correlation is not very high. Notice, however, that the average utilities under EU behaviour in two two different cases are very similar. The main difference lies in the average utility under SV behaviour. The average utilities under SV behaviour are low under elimination agendas when the intensities of x are high because these high intensities are not reflected in voters' sincere behaviour in any way. Strategic voting thus at least gives a chance to an alternative that is introduced early under an elimination agenda. When the intensities of x are lower and the intensities of y higher, x is selected just as seldom as in all SV-setups. However, since the high-utility y always participates in the second-round contest, the average utility becomes relatively high under SV behaviour. In setups with strong correlation, and in which the intensities of x are low, strategic voting decreases average utility because the strategic votes for the alternative that has a low sum of utility on the average are more likely to matter than the strategic votes for the other alternatives.

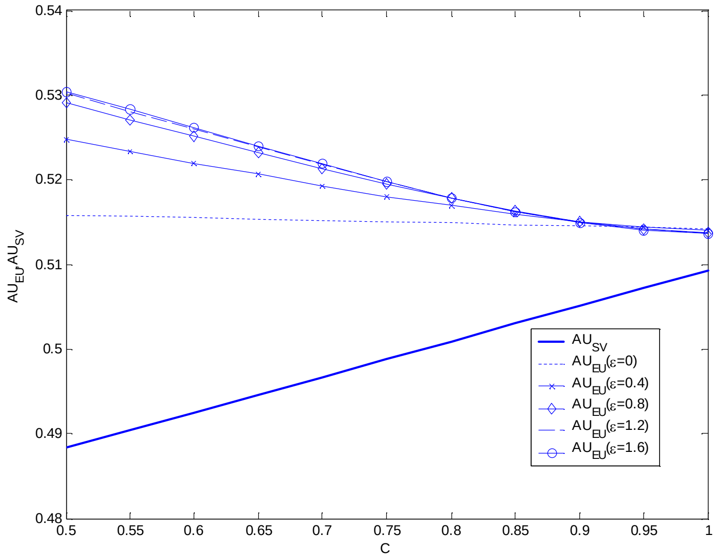

Figure 3.11: Average utilities in an elimination agenda when the intensities of x are high

There is one important qualification to the results from both voting rules. If some voter types engage in EU behaviour and some in SV behaviour, the systematic absence of balancing strategic votes suggests that the welfare consequences of strategic voting are less beneficial or welfare-decreasing.[16] This is an important consideration, because it may well be reasonable to assume that some voter types are more prone to strategic voting than some others. The complexity of this matter, however, prevents us from presenting a discussion of it here.

3.8 Conclusions

The main conclusion that may be drawn from the simulation results is that welfare-increasing strategic voting is not a mere theoretical possibility in parliamentary voting. Indeed, it may well be the typical case.

The most important and widely discussed condition in Arrow's (1963) impossibility theorem is the Independence of Irrelevant Alternatives (IIA). IIA is

[16] If some voters engage in EU-behaviour and some others in SV-behaviour, but engaging in EU-behaviour does not correlate with being of a certain voter type, the simulation results are similar (but weaker) to the results obtained here.

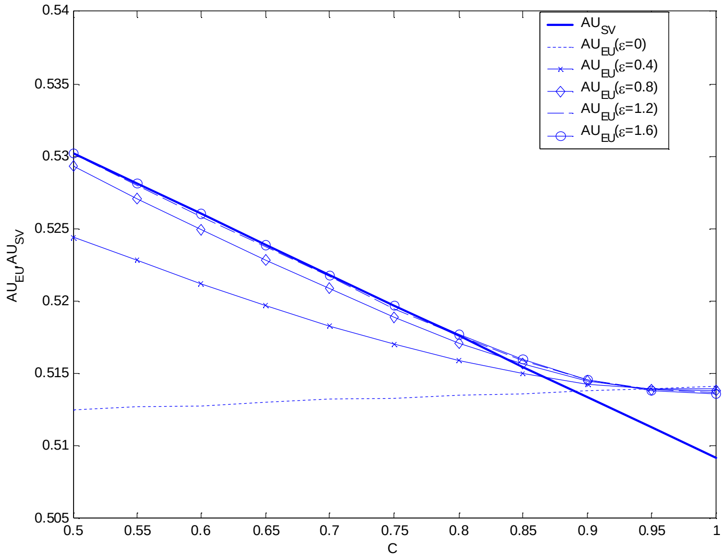

Figure 3.12: Average utilities in elimination agendas when the intensities for x are low

closely connected to strategy-proofness.[17] The idea that strategic voting should be precluded in a voting rule is the only justification for strategy-proofness, and a crucial argument for IIA (e.g. Blin 1976). The results imply that the precise interpretation of these conditions may need to be re-evaluated in voting theory because the results presented here indicate that strategic voting may well be beneficial. Imagine that there was a strategy-proof voting rule. By definition, this would mean that voters would not have an incentive for changing their behaviour by voting strategically. Strategic voting could not be welfare-increasing or welfare-decreasing because the individuals would not have an incentive to engage in it. The point is that it is not possible to determine whether strategic voting and thereby strategy-proofness are desirable or not, a priori, without explicitly investigating the welfare consequences of strategic voting in each voting rule. The possibility that strategic voting is beneficial implies that the rationale for the so called *manipulability measures* (e.g. Saari 1990*b*, Smith 1999) is put into question. The important question to study is not which voting rules are best in selecting outcomes that are "close" to those that would have ensued from sincere voting, but rather which rules result in best outcomes when individuals vote strategically.

In some contexts (other than voting) strategy-proofness may be *intrinsically*

[17] See Gibbard (1973), Satterthwaite (1975) and Blin & Satterthwaite (1978).

important because it may be important to know the preferences of every agent. It should be borne in mind that the results here concern only two specific, although commonly used voting rules, whereas the scope of the impossibility theorems is considerably broader.

Strategic voting increases average utility compared to sincere voting because the former allows preference intensities to influence voting outcomes but the latter does not. Uniform setups yield the worst possible welfare consequences of strategic voting because the intensity differences between the alternatives are as small as they can possibly be. If the correlation between voter types and intensities is strong, strategic voting is very clearly welfare-increasing. This result also shows that if voters vote strategically, the criticism that majority rule does not take preference intensities into account is false. Furthermore, the larger the differences in the intensities are, the more welfare-increasing strategic voting is. If the correlation of intensities is strong, welfare-increasing strategic voting does not even require reliable signals.

A particular configuration of utilities under elimination agendas provides an exception. If the intensities for the alternative that may be eliminated on the first round are low on the average, strategic voting may be welfare-decreasing. We have seen, however, that even in this case the average utilities under EU behaviour are relatively high. It is just that the average utilities under SV behaviour are even higher because the unpopular alternative will often be eliminated with a sincere vote.

These findings suggest that strategic voting is a virtue rather than a vice in commonly used parliamentary agendas if all voters engage in expected utility maximising behaviour.

Chapter 4

The Borda rule is also intended for dishonest men

Abstract

his paper examines the welfare consequences of strategic voting under the Borda rule in a comparison of utilitarian efficiencies in simulated voting games under two behavioural assumptions: expected utility- maximising behaviour and sincere behaviour. Utilitarian efficiency is higher in the former than in the latter. Strategic voting increases utilitarian efficiency particularly if the distribution of preference intensities correlates with voter types. The Borda rule is shown to have two advantages: strategic voting is beneficial even if some but not all voter types engage in strategic behaviour, and even if the voters' information is based on unreliable signals. (JEL classification numbers: D71, D81)

Keywords: Strategic voting; Borda rule; Welfare; Simulation

4.1 Introduction

One of the main criticisms of the Borda rule is that it is highly susceptible to strategic voting.[1] Voting strategically *for* (*against*) a candidate means giving a higher (lower) Borda score than the voter's preference ordering would imply.[2] Borda is famous for having exclaimed, "My scheme is intended only for honest men" (quoted in Black 1958, p. 182), when the susceptibility of his rule to strategic manipulation was pointed out.

[1] See e.g., Saari 1990*b*, Smith 1999, Favardin et al. 2002, and Taylor (2005).

[2] Strategic manipulation of the Borda rule by introducing a new alternative, and manipulation by coalitions are not considered in this paper. See Dummett (1998) and Saari (1990*a*) on the former and Lepelley & Mbih (1994) and Lepelley & Valognes (2003) on the latter.

This paper examines the welfare consequences of strategic voting under the Borda rule by means of computer simulations. As in (2006*b*, 2007), the welfare consequences of strategic voting are evaluated by comparing the *utilitarian efficiency* obtained with *Expected Utility- maximising voting behaviour* (EU behaviour) and with *Sincere Voting behaviour* (SV behaviour). In the former *all* voters always vote sincerely, while in the latter the voters may vote strategically or sincerely depending on their preferences and beliefs. *Utilitarian efficiency* is defined as the percentage of simulated voting games in which the candidate with the highest utility sum (the utilitarian winner) is selected.

The main finding is that strategic voting increases utilitarian efficiency compared with sincere voting behaviour when the voters engage in expected utility-maximising behaviour under conditions of incomplete information. Under the utilitarian evaluation of voting outcomes, what has been thought of as a major disadvantage of the Borda rule turns out to be an argument *for* it. However, since strategic voting increases utilitarian efficiency in most of the commonly used voting rules (Lehtinen 2006*b*, 2007), the results reported here do not provide an unambiguous argument for using the Borda rule instead of some other voting rule. On the other hand, and in contradistinction to the majority rule in amendment agendas, it will be shown that the Borda rule yields high utilitarian efficiencies even when voters' information on other voters' preferences is fairly unreliable, and even if some but not all *voter types* engage in strategic behaviour.

Proponents of the Borda rule have traditionally argued that it selects fair compromises as outcomes. Indeed, Borda himself seems to have defended it by referring to cardinal utilities. He argued that, given three candidates and in the absence of further knowledge, preference for the second-best candidate could be assumed to be midway between the best and the worst (de Borda 1995[1784], p. 85). I will show that strategic voting is less welfare-increasing precisely when the utilities for the voters' second-best candidates are, on average, midway between the worst and the best. It is also less welfare-increasing when the utilities for the middle candidates are uniformly distributed in the interval $[0, 1]$ between the worst and the best candidates, but most welfare-increasing when the second-placed utilities for *some* candidates are typically higher than the average of the uniform distribution (i.e. higher than one half), and the second-placed utilities for some *other* candidates are typically lower than this average. Computer simulations that feature such assumptions are described as *setups with correlation between preference orderings (voter types) and intensities*, because the preference intensities (utilities for the second-best candidates) correlate with the voter types in the sense that voters with given preference orderings have typically a high or low utility for their middle candidate.

It will be shown that strategic voting increases utilitarian efficiency mainly because it allows the voters to express intensities of preference, thereby providing fuller information on such intensities than sincere voting.[3] This suggests that

[3] Donald Saari (e.g., 2001; 2003*b*; 2003*a*; 2006) has consistently argued in favour of using the Borda rule instead of majority rule. One of his arguments is that the latter throws away

the Borda rule may not need to be made fuzzy (Marchant 2000, García-Lapresta & Martínez-Panero 2002) or probabilistic (Heckelman 2003) in order to yield information on intensities.

Under many voting rules, strategic voting may be considerably less welfare-increasing or welfare-diminishing if some but not all voter types engage in strategic behaviour (see Lehtinen 2006*b*, 2007). I will argue that the beneficial welfare consequences of strategic voting under the Borda rule do not depend crucially on the assumption that *all* voter types engage in strategic behaviour: unlike other voting rules, it is fairly robust to this kind of heterogeneity in behavioural disposition.

Since the results of this study concern utilitarian efficiencies, they rely on full interpersonal comparability of preference intensities. Given that assumptions of such interpersonal comparability are generally considered to be suspect, it is necessary to justify their use. The main justification is that the results are highly robust with respect to different interpersonal comparisons.

The model of incomplete information is based on *statistical signal extraction* in the sense that voters are assumed to obtain noisy signals concerning the preference profile before they vote, and they derive their beliefs concerning whether one candidate has a higher Borda score than another from these signals. These beliefs are then used in an expected utility model of voting. The model is not game-theoretical in the sense that the voters are not assumed to be able to take other voters' strategic choices into account when they formulate their beliefs concerning the expected Borda scores. Since the determination of these beliefs is independent of the determination of the actions, I will explain, for expository reasons, 'where the beliefs come from' only after giving an account of how voters act with their given beliefs.

The structure of the paper is the following. Section 4.2 formulates an expected utility model of strategic voting under the Borda rule. Sections 4.3 and 4.6.3 explain the logic of the model in terms of why utilitarian winners are likely to obtain many and lose few strategic votes by explaining the 'counterbalancing' of strategic votes. In section 4.4 I describe the incomplete information model by showing how to derive beliefs from perturbed signals concerning the preference profile. Section 4.5 describes the simulation framework and setups, and section 4.6 presents the simulation results. Section 4.7 provides a discussion on interpersonal comparisons, and section 4.8 concludes the paper.

information but the former does not.

4.2 A model of strategic voting under the Borda rule

Let X={x,y,z} denote the set of candidates with generic members j and k.[4] Let 1, 2, and 3 denote an individual voter's best, second-best, and worst candidate. Let U_1^i, U_2^i, and U_3^i denote voter i's utility for his or her best, second-best and worst candidate, respectively. The six possible types of voters and their preference orderings are presented in Table 4.1 below. I will refer to a voter's

type of voter						
t_1	t_2	t_3	t_4	t_5	t_6	U^i
x	y	z	x	y	z	U_1
y	z	x	z	x	y	U_2
z	x	y	y	z	x	U_3

Table 4.1: Voter types and utilities

utility for her second-best candidate U_2^i as *intensity of preference*.

The Borda rule is defined as follows.[5] Let n denote the number of candidates. Voters are asked to provide a full ranking list of all candidates, assigning n-1 marks for the top candidate, n-2 for the second ,.., 0 for the worst candidate. The Borda winner is defined as the candidate who obtains the largest sum of marks, i.e. the largest *Borda score*.

Voters are assumed to have beliefs concerning whether any given candidate j will obtain a higher Borda score than another candidate k. How these beliefs are derived is explained in the next section. Let $p^i(12)$ denote voter i's (degree of) belief that the candidate he or she considers the best will obtain a higher Borda score than the second-best candidate: $p^i(13)$ and $p^i(23)$ are similarly defined. Reporting the ordering 123 then means voting sincerely, and reporting any other ordering means voting strategically.

There are two possible motivations for voting strategically in Borda rule.[6]

Situation 1 A voter do not like his or her second-best candidate very much. In order to increase the victory chances of the candidate he or she condiders best, he or she gives the lowest score to the second-best candidate. He or she must simultaneously believe that his or her strategic vote is not likely to make the worst candidate win. The voter must thus believe that his or her best and second-best candidates are the most likely winners, and

[4] The present model is restricted to three candidates. The framework of this paper (the signal extraction information model) could easily be extended to incorporate more than three candidates, all that is needed is an account of expected utility maximization with more than three candidates.

[5] See Pattanaik (2002) for a review of the axiomatic literature on the Borda rule and other positional methods.

[6] To the best of my knowledge, there are no incomplete-information models of strategic voting in the Borda rule. Black (1976) and Ludwin (1978) provide an account that resembles the first situation, and Felsenthal (1996) considers a case that resembles the second.

that the race between them is tight. The voter thus weighs the chance of the most preferred candidate winning the whole contest if he or she votes strategically, against the chance that putting the worst candidate second and the second-best candidate third will bring victory to the worst candidate.

- This situation is characterised by the following kinds of beliefs and preferences: $p^i(13)$ high, $p^i(23)$ high, $p^i(12)$ close to $\frac{1}{2}$, and U_2^i low.
- When a voter votes according this motivation, he or she reports 132 instead of 123.

Situation 2 A voter believes that his or her best candidate does not have a chance of winning, but that his or her second-best candidate will have a close race with the worst candidate, and he or she has fairly strong positive feelings about the second-best candidate. In order to increase the chance that this second-best candidate will win, he or she puts it first, the best candidate second, and the worst candidate last. The trade-off is now between the chance that the second-best candidate will be selected and the possibility of an error of judgment in that the best candidate would have won after all, had he or she not been strategically deserted by the voter.

- This situation is characterised by the following kinds of beliefs and preferences: $p^i(12)$ low, $p^i(13)$ low, $p^i(23)$ close to $\frac{1}{2}$, and U_2^i high
- When a voter votes with this motivation, she reports 213 instead of 123.

Let us now consider some examples in order to check that these behaviour rules are rational. Suppose that $p^i(12)=1$, $p^i(13)=0.8$ and $p^i(23)=0.7$.[7] Since the voter believes that his or her most preferred candidate is sure to obtain a higher Borda score than the second-best candidate, there is no need for strategic voting. Note, however, that a rational voter could not have this combination of beliefs because they are not transitive. Since the final Borda scores always constitute a transitive ordering, voters' beliefs about these scores must be transitive. Here the voter expects the aggregate Borda ordering to be 123 such that $p^i(12)=1$ and $p^i(13)=0.8$, but these beliefs are not transitive because $p^i(12)>p^i(13)$. Let us update the beliefs so as to end up with $p^i(12)=1$, $p^i(13)=1$ and $p^i(23)=0.7$. In this case the beliefs are consistent with the transitivity restriction, but it does not seem to make much sense to vote strategically because the voter's best candidate is selected in any case. What if $p^i(12)=0$, $p^i(13)=0.8$ and $p^i(23)=0.7$? It now seems that strategic voting makes no sense because the voter's best candidate does not have a chance of beating his or her second-best candidate, but a strategic vote for the worst candidate might bring

[7] An anonymous referee proposed this example as a criticism of an earlier and admittedly incorrect model of strategic voting under the Borda rule. Since the example led me to construct an entirely new model, my gratitude to the referee is hereby acknowledged.

victory to this candidate. Thus, the race between a voter's most preferred and second-best candidate must be tight.

Note that in situation 1 the voter believes that the worst candidate does not have a chance of winning, and in situation 2 that the best candidate does not have a chance of winning. If the second-best candidate does not have a chance of winning, there is no point in strategic voting because it may only increase the chance that his or her worst candidate wins (situation 1) or increase the chance of victory for the worst candidate (situation 2).

Voters are assumed to make their choice between sincere and strategic voting on the basis of whether the expected utility gain from voting strategically is higher than the expected utility loss. A standard starting point in voting models is that voters should condition their strategic vote on its being pivotal. As Myatt and Fisher (2002) pointed out in the context of the plurality rule, what is important is the *relative* rather than the *absolute* probability of being pivotal. In the model under discussion, and in situation 1, the voters condition their choice on the probability that they are pivotal between the best and second-best candidates (i.e. an individual voter's best and second-best candidates), and between the second-best and worst candidates. In situation 2 the relative probability concerns being pivotal between the second-best and the worst and between the best and the second-best.

Let us now formulate a decision rule that adequately reflects the trade-offs. Voters assess the *possible utility gain* (PUG) and the *possible utility loss* (PUL) from voting strategically against the probability of realisation. PUG is the potential gain in utility from voting strategically, and PUL is the potential loss in utility incurred by voting strategically if the probability estimates turn out to be incorrect.

In situation 1, a voter's PUG is the difference in utility between the best and second-best candidate: $U_1^i - U_2^i$. This gain is most relevant when the race between the two is tight. (In what follows, the superscript denoting the individual voter is dropped from all expressions in order to avoid clutter.) What is thus needed is a function P that correctly weighs the utility gain depending on how likely it is to materialise. The following functional form gives weight 1 to the utility gain when p(12)=$\frac{1}{2}$, and weight 0 when p(12)=0 or p(12)=1:

$$P = 1 - 2(|p(12) - \frac{1}{2}|),$$

where the vertical bars denote absolute values. The expected utility gain from reporting 132 $EU(G)$ is thus

$$EU(G) = [1 - 2(|p(12) - \frac{1}{2}|)](U_1 - U_2). \tag{4.1}$$

The possible utility loss from voting strategically depends on which candidate is expected to win. If the voter expects the aggregate Borda ordering to be 123, it is $U_1 - U_3$, and if she expects it to be 213, it is $U_2 - U_3$. Given that the voters do not know whether the best or the second candidate will win, but they have beliefs about it, they need to weigh the losses against the probability that the best candidate will beat the second-best candidate. The expected utility loss is thus

$$\begin{aligned} EU\,(L) &= p\,(12)\,[1-p(13)][1-p(23)](U_1-U_3)+ \\ &\quad [1-p\,(12)][1-p(13)][1-p(23)](U_2-U_3). \end{aligned} \tag{4.2}$$

A voter thus votes strategically by reporting the ordering 132 if

$$EU(G) - EU(L) > \tau_1. \tag{4.3}$$

τ_1 is a parameter that reflects the voters' propensity to engage in strategic voting. This can be expressed as follows:

$$\begin{aligned} &[1-2(|p(12)-\frac{1}{2}|)](U_1-U_2) \\ &-p\,(12)\,[1-p(13)][1-p(23)](U_1-U_3) \\ &-[1-p\,(12)]\,[1-p(13)][1-p(23)](U_2-U_3) > \tau_1. \end{aligned} \tag{4.4}$$

Let us now consider situation 2. The PUG is $U_2 - U_3$, and the PUL is $U_1 - U_2$. The expected utility gain from strategic voting is

$$EU\,(G) = [1-2(|p(23)-\frac{1}{2}|)](U_2-U_3),$$

and the expected utility loss is

$$EU\,(L) = p(12)p(13)\,(U_1-U_2)\,.$$

A voter votes strategically by reporting 213 if

$$[1-2(|p(23)-\frac{1}{2}|)](U_2-U_3) - p(12)p(13)\,(U_1-U_2) > \tau_2.\text{[8]} \tag{4.5}$$

4.3 The logic of the model: counterbalancing

f a strategic vote is based on poor information, it may be counter-productive. Assume, for example, that a voter expects the Borda ordering to be 123 with, say, p(12)=0.7, p(13)=0.8 and p(23)=0.7. Let $U_1 = 0.9$, $U_2 = 0.5$, $U_3 = 0.1$, and $\tau_1 = 0$. He or she will then vote strategically by reporting 132 because applying equation 4.4 yields $[1 - 2 * (0.7 - \frac{1}{2})] * (0.9 - 0.5) - 0.7 * [1 - 0.8] * [1 - 0.7] * [0.9 - 0.1] - [1 - 0.7] * [1 - 0.8] * [0.5 - 0.1] = 0.1824 > 0$. However, if he or she was wrong about the likely outcomes, the strategic vote will bring about the worst outcome! It is the very nature of uncertainty that something that is considered unlikely but possible may happen. However, it is not likely

[8] It is possible to give parameters τ_1 and τ_2 different values, but in this paper they were assumed to be the same in all except the simulations in which one of them was so large (i.e. at least 1) that there was no strategic voting in that situation.

that a voter will obtain a worse outcome by voting strategically than by voting sincerely. In most cases a strategic vote benefits both the voter and the whole electorate. The explanation lies in the *counter-balancing* of strategic votes.

Under incomplete information, some voters may have the strategic incentive to decrease the Borda score for candidate x and increase that of y while at the same time some others may have the incentive to increase the score of candidate y and decrease that of candidate x. It follows from the set out above that voting strategically *for* a candidate is more likely when the preference intensity for that candidate is high than when it is low. Table 4.2 summarises the effects of strategic voting by all voter types by showing the candidate whose Borda score is increased (↑) or decreased (↓) when a voter of a given type gives a strategic vote.

voter type	1		2		3		4		5		6	
situation	1	2	1	2	1	2	1	2	1	2	1	2
↑	z	y	x	z	y	x	y	z	z	x	x	y
↓	y	x	z	y	x	z	z	x	x	y	y	z

Table 4.2: Directions of change due to strategic voting

For example, if type-one voters vote strategically, they do so by increasing the Borda score for z at the expense of y in situation 1, and by increasing the score of y at the expense of x in situation 2.

Consider situation 1. Let us assume that the intensities for candidate x are higher on average than those for candidate y. The simulations with intensity correlation formalise these assumptions, the implication being that x is likely to be the utilitarian winner, and y the worst candidate in utilitarian terms. If type-one or type-six voters vote strategically in situation 1, their vote *decreases* the Borda score of y, and increases that of z and x, respectively, while strategic voting by type-three or type-five voters *decreases* the Borda score of x, and increases that of y and z, respectively.

Equation 4.4 implies that, under the above assumptions, type-one and type-six voters are likely to vote strategically *more often* than those of types three and five. For example, a type-three voter will vote strategically by reporting zyx rather than zxy if $T_3 = [1-2(|p(xz)\text{-}\frac{1}{2}|)]\,[U(z)\text{-}U(x)]\text{-}[1\text{-}p(xz)]\,p(yz)p(xy)[\mathrm{U}(z)\text{-}\mathrm{U}(y)] - p(xz)\,p(yz)p(xy)\,[U(x)\text{-}U(y)] > \tau_1$. Since $[1-2(|p(xz)-\frac{1}{2}|)] \geq 0$ and $p(xz)\,p(yz)p(xy) \geq 0$ for all possible values of the probabilities, $\frac{\partial T_3}{\partial U(x)} < 0$. Thus, the higher the intensity for x, $\mathrm{U}_i(\mathrm{x})$, the less likely this person is to vote strategically against x by reporting zyx rather than the sincere zxy. Similarly, since $T_5 = [1-2(|p(xy)-\frac{1}{2}|)]\,[U(y)-U(x)] - [1-p(xy)]\,[1-p(yz)][1-p(xz)][U(y)-U(z)] - p(xy)\,[1-p(yz)][1-p(xz)][U(x)-U(z)]$, $\frac{\partial T_5}{\partial U(x)} < 0$, there will be few voters of type five who report yzx rather than yxz if their intensity for x is high on average. On the other hand, for voters of type one $T_1 = [1-2(|p(xy)-\frac{1}{2}|)]\,[U(x)-U(y)] - p(xy)\,[1-p(xz)][1-p(yz)](U(x)-U(z)) - [1-p(xy)]\,[1-p(xz)][1-p(yz)](U(y)-U(z))$ so that $\frac{\partial T_1}{\partial U(y)} < 0$. Thus, the smaller U(y) is, the more likely it is that these voters will vote strategically for z and against y by reporting xzy. A similar argument shows that $\frac{\partial T_6}{\partial U(y)} < 0$.

Let us now turn to situation 2. A high average intensity for x implies that *many* type-three and -five voters will vote strategically by raising the Borda score of x at the expense of y and z. A low average intensity for y implies that *few* voters of types one and six will vote strategically. They are thus likely to refrain from making y's Borda score higher. Consider again a type-three voter as an example. Since $\frac{\partial[1-2(|p(xy)-\frac{1}{2}|)][U(x)\text{-}U(y)]-[1\text{-}p(xz)][1\text{-}p(yz)][U(z)\text{-}U(x)]}{\partial U(x)}$=[1-2($|p(xy)$-$\frac{1}{2}|$)]+[1-$p(xz)$] [1-$p(yz)$] > 0 for all values of the probabilities, the higher the intensity for x, the more likely this voter is to vote strategically for x by reporting xzy in situation 2. Thus, a high average value of U_i(x) in this case implies that many of these voters will vote strategically for x by reporting xzy. Similar arguments show that many type-five voters will vote strategically by reporting xyz, and that few voters of types one and six will vote strategically by reporting yxz or yzx, respectively.

In conclusion, relatively *many* strategic votes *for* the utilitarian winner x are likely to be counter-balanced by relatively *few* strategic votes *against* it. I will return to the matter of counter-balancing in section 4.6.3.

4.4 The voters' signals and beliefs

The basic idea behind this information model is that voters formulate probabilities based on noisy signals concerning other voters' preferences, i.e. a preference profile. In real life voters obtain this kind of information from opinion polls, television broadcasts and conversations with friends, for example. All these possible sources of information are assumed to be modeled by the noisy signals. This signal-extraction framework allows the derivation of a heterogeneous set of probabilities for a large population of voters by characterizing the *reliability* of the signals.[9] Each voter obtains a slightly different signal, but since the signals are based on the realised preference profile, his or her beliefs are constrained by the realities of the situation.

Voters are assumed to take this perturbed information about the realised profile as a relevant proxy for the expected aggregate Borda scores. This assumption is not reasonable if the relationship between the preference profile and the realised Borda scores is systematically distorted, and if the voters could be assumed to know how it is distorted. If the results reported here are correct, i.e. if strategic voting increases utilitarian efficiency, they imply that voters would get better information on the expected Borda scores if they were also able to obtain information on preference intensities and on the behavioural dispositions of other voters. It is possible to take intensities into account in the signals by assuming, for example, that they are based on the sums of utilities. However, the signals referred to in this paper are based only on the ordinal preference profile. They are thus 'systematically distorted' in the sense that voters are as-

[9] Lehtinen (2006*a*) discusses a similar signal-extraction model in more detail.

sumed not to be able to take intensity information concerning other voters into account. This assumption is made because it considerably simplifies the application of the signal extraction model to the various setups. It is also perfectly possible to study how other kinds of systematic distortions affect the results, but since this is already a long paper, this is left for a future study.

This signal extraction model is embedded in simulated games that are based on the *impartial culture* (Tsetlin, Regenwetter & Grofman 2003, Gehrlein 2002) assumption. A *simulated game g* consists of a set of randomly generated payoffs, beliefs based on these payoffs, and other informational assumptions, as well as voting outcomes under the different behavioural assumptions. The uniform distribution on [1,2,...,6] was used to generate a profile of N=201 voters in each simulated game g.[10] The voters then obtained a perturbed signal on each candidate's Borda score which was based on the realised number of voters who preferred one candidate to another.

Let $N^j_{1,i}$ be a random variable that obtains the value 1 when voter i ranks candidate j highest, and zero otherwise. The number of voters who rank candidate j first, $\mathcal{N}^j_1$, could then be viewed as the sum of N random variables $N^j_{1,i}$, one for each voter i: $\mathcal{N}^j_1 = \sum_{i=1}^N N^j_{1,i}$. Similarly, the number of voters who rank candidate j second, $\mathcal{N}^j_2$, could be viewed as the sum of random variables $N^j_{2,i}$: $\mathcal{N}^j_2 = \sum_{i=1}^N N^j_{2,i}$. The impartial culture assumption implies that the probability that such a Bernoulli trial (for example, $N^j_{1,i}$ and $N^j_{2,i}$) will result in outcome 1 is $\frac{1}{3}$. N^j_1, N^j_2, N^k_1, and N^k_2 could thus be viewed as random variables with a binary distribution $N^j_1 \sim B(N, \frac{1}{3})$.

If all voters voted sincerely, the Borda scores for candidates j and k would be given by

$$B^j = 2\mathcal{N}^j_1 + \mathcal{N}^j_2,$$

and

$$B^k = 2\mathcal{N}^k_1 + \mathcal{N}^k_2.$$

Voter i's expected Borda scores for candidates j and k are given by the following *signals:*

$$S^j_i = 2\mathcal{N}^j_1 + \mathcal{N}^j_2 + \varepsilon R^j, \tag{4.6}$$

and

$$S^k_i = 2\mathcal{N}^k_1 + \mathcal{N}^k_2 + \varepsilon R^k, \tag{4.7}$$

where R^j and R^k are standard normally distributed random variables, and ε is a scaling factor that reflects the *reliability* of the signals. The probability that candidate j will obtain a higher Borda score than candidate k, given the signals S^j and S^k can be derived by formulating another random variable for

[10] This particular number was chosen mainly in order to obtain comparability with some earlier simulation studies of voting rules (Merrill 1984, 1988). It is easy to study the effect of the number of voters on the results, but this was not done for the present paper.

the *difference*[11] between the two signals, S^{jk}:

$$S^{jk} = S^j - S^k. \tag{4.8}$$

Let us also define $B^{jk} = 2\mathcal{N}_1^j + \mathcal{N}_2^j - 2\mathcal{N}_1^k - \mathcal{N}_2^k$, and $R^{jk} = \varepsilon R^j - \varepsilon R^k$. Then the signal S^{jk} could be written as

$$S^{jk} = B^{jk} + R^{jk}, \tag{4.9}$$

i.e. as the sum of two independent random variables. Let n_t denote the realised number of voters of type t. For a comparison between x and y, $S^{xy} = 2N_1^j + N_2^j + \varepsilon R^j - (2N_1^k + N_2^k + \varepsilon R^k)$ is

$$\begin{aligned} S^{xy} &= 2\,(n_1 + n_4) + n_3 + n_5 - 2(n_2 + n_5) - n_1 - n_6 + R^{xy} \\ &= n_1 - 2n_2 + n_3 + 2n_4 - n_5 - n_6 + R^{xy}. \end{aligned}$$

For a comparison between x and z it is

$$S^{xz} = 2n_1 - n_2 - n_3 + n_4 + n_5 - 2n_6 + R^{xz},$$

and for a comparison between y and z it is

$$S^{yz} = n_1 + n_2 - 2n_3 - n_4 + 2n_5 - n_6 + R^{yz}.$$

It is shown in Appendix B that the variance of B^{jk} is $2N$, and that of R^{jk} is $2\varepsilon^2$. According to the central limit theorem, S^{jk} can be approximated with a normally distributed random variable with a mean of zero, and a variance of $2N + 2\varepsilon^2$, $S^{jk} \sim N(0, 2N + 2\varepsilon^2)$. This approximation restricts the applicability of the model to situations with a fairly large number of voters. Since the mean of S_{jk}, $\mu_{S^{jk}}$ is obviously zero, normalising S^{jk} yields

$$Q^{jk} = \frac{S^{jk} - \mu_{S^{jk}}}{\sqrt{\sigma_S^2}} = \frac{S^{jk}}{\sqrt{2N + 2\varepsilon^2}} = \frac{2\mathcal{N}_1^j + \mathcal{N}_2^j + \varepsilon R^j - (2\mathcal{N}_1^k + \mathcal{N}_2^k + \varepsilon R^k)}{\sqrt{2N + 2\varepsilon^2}}. \tag{4.10}$$

The probability that candidate j will obtain a higher Borda score than candidate k, $p^i(jk)$ is thus given by the standard normal cumulative distribution function Φ:

$$p^i(jk) = 1 - \Phi(\frac{S_i^{jk}}{\sqrt{2N + 2\varepsilon^2}}). \tag{4.11}$$

[11] The signals could be formulated in such a way that the difference in the Borda scores is taken first, and the random variable R^{jk} is added to this expression; $S^{jk} = 2\mathcal{N}_1^j + \mathcal{N}_2^j - 2\mathcal{N}_1^k - \mathcal{N}_2^k + \varepsilon R^{jk}$. I chose to add the random component to each Borda score because doing so automatically precludes cyclic beliefs: the beliefs are derived *after* each Borda score has been perturbed. The ordering of the *realized* signals for Borda scores is automatically transitive. This way of formulating the signals thus obviates the need to specify how the voters should update their beliefs once the signal information has provided intransitive beliefs. The downside is that the range of reasonable values for the parameter ε now inevitably depends on the number of voters.

Applying this equation gives more familiar-looking expressions for beliefs. For example, a type-one voter ranks the candidates in the order xyz, and his or her beliefs are given by $p^i(12)=p^i(xy)=1-\Phi(\frac{S_i^{xy}}{\sqrt{2N+2\varepsilon^2}})=$ $1-\Phi(\frac{2\mathcal{N}_1^x+\mathcal{N}_2^x+\varepsilon R^x-(2\mathcal{N}_1^y+\mathcal{N}_2^y+\varepsilon R^y)}{\sqrt{2N+2\varepsilon^2}})$, $p^i(13)=p^i(xz)=1-\Phi(\frac{S_i^{xz}}{\sqrt{2N+2\varepsilon^2}})$, for example. Note that the variables concerning the preference profile ($\mathcal{N}_1^x$, $\mathcal{N}_2^x$, $\mathcal{N}_1^y$, and $\mathcal{N}_2^y$) are the same for all voters, but the random perturbances (R^x and R^y) are different for each one, making each voter's belief slightly different.

4.5 Setups and simulation

A setup is a combination of assumptions used in a set of $G = 2000$ simulated games. Expected utility setups differ with respect to the degree of reliability of the voters' information (ε), their propensity to vote strategically (τ), and the degree of correlation between voter types and preference intensities (C) (see next paragraph). In *uniform setups* each voter's utilities are drawn from a uniform distribution on [0,1], while in *setups with intensity correlation* voter types three and five have systematically higher preference intensities for their second-best candidate (x), and voter types one and six have systematically lower preference intensities for their second-best candidate (y). These setups are identical to the corresponding uniform setups with respect to all parameters except voters' preference intensities. In order to generate setups with correlation between this parameter and voter types without affecting the interpersonal comparisons or the preference orderings, the individual utilities were derived as follows.

U_1, U_2, and U_3 were first generated from the uniform distribution on [0,1] for each voter. U_1 and U_3 were then used for defining the voter's utility scale as the $[U_3,U_1]$ interval, but U_2 was not used for any purpose. (The reason for this is explained in section 4.7.) Instead, a standardised intensity $\widetilde{U}_2$ for each voter's second-best candidate was generated from the uniform distribution on [0,1]. This standardised intensity expresses what a voter's utility for his or her second-best candidate would be if the scale was the [0,1] interval. Let us refer to these standardised second-best utilities as *intra*personal *intensities.* In setups with intensity correlation, these intensities were multiplied by a parameter C, $0.5<C\leq 1$ for those who put y second (voter types one and six) so that the new correlated intensities $\widetilde{U}_2^{C,1}$ and $\widetilde{U}_2^{C,6}$ were given by

$$\widetilde{U}_2^C = C\widetilde{U}_2. \tag{4.12}$$

In order to compensate the decreases in utility for voter types one and six, the intensities for voters of types three and five (i.e. for x) were given by

$$\widetilde{U}_2^C = 1 - C\widetilde{U}_2. \tag{4.13}$$

These adjustments make the average utilities for x higher and the average utilities for y lower than in the uniform setups, while keeping the overall average utility fixed. In uniform setups, $C = 1$. C thus denotes the *degree of correlation* between preference intensities and voter types.

These standardised intensities were then scaled back into the original $[U_3, U_1]$ utility scale. Let U_2^* denote voter's intensity expressed in terms of the original $[U_3, U_1]$ scale. Since the relationship between the standardised intrapersonal utility for the second-best candidate and the original scale is given by

$$\widetilde{U}_2^C = 1 - \frac{U_1 - U_2^*}{U_1 - U_3}, \tag{4.14}$$

U_2^* is given by:

$$U_2^* = U_3 + \widetilde{U}_2^C (U_1 - U_3). \tag{4.15}$$

Let UW^g denote the utilitarian winner in a simulated game g, and $W_{EU}^g(\varepsilon, \tau, C)$ the winner in an EU-behaviour setup. Let UWW^g denote an indicator function that obtains the value 1 when $UW^g = W_{EU}^g(\varepsilon, \tau, C)$. *Utilitarian Efficiency in an EU setup* is thus given by:

$$UE_{EU}(\varepsilon, \tau, C) = \frac{1}{G} \sum\nolimits_{g=1}^{G} UWW^g * 100. \tag{4.16}$$

Utilitarian Efficiency in the SV setup, UE_{SV}, is calculated similarly. Let us say that EU behaviour is *welfare-increasing* in a setup if the utilitarian efficiency is higher than in the SV setup:

$$UE_{EU}(\varepsilon, \tau, C) > UE_{SV}. \tag{4.17}$$

If the converse holds, EU behaviour is *welfare-diminishing*. Let us also say that *strategic voting is welfare-increasing* in a setup if EU behaviour is welfare-increasing.

4.6 Simulation results

4.6.1 Preliminaries

What is of interest is how the degree of correlation (C), the reliability of the voters' information (ε), and the voters' propensity to engage in strategic voting (τ) affect utilitarian efficiency. All these parameters affect the results, but at least two must be fixed each time the results are reported.[12] The simulations were conducted with $C = 0, 0.05, ..., 0.5$, $\varepsilon = 0, 4, ..., 16$, and $\tau = 0, 0.25, ..., 1$. The choice of parameter values is arbitrary to some extent, but there are some alternatives. For example, since the signal extraction model cannot be used with

[12] The FORTRAN codes for generating the results, and the result tables with all combinations of parameter values are obtainable from the author on request. In order to check the computer code with the number of runs used (2000) for each setup, the IMSL library of FORTRAN codes and access to a supercomputer are required. The simulations were conducted with a Sun Fire 25K server (UltraSPARC IV processor) at the Center for Scientific Computing, Otaniemi, Espoo, Finland.

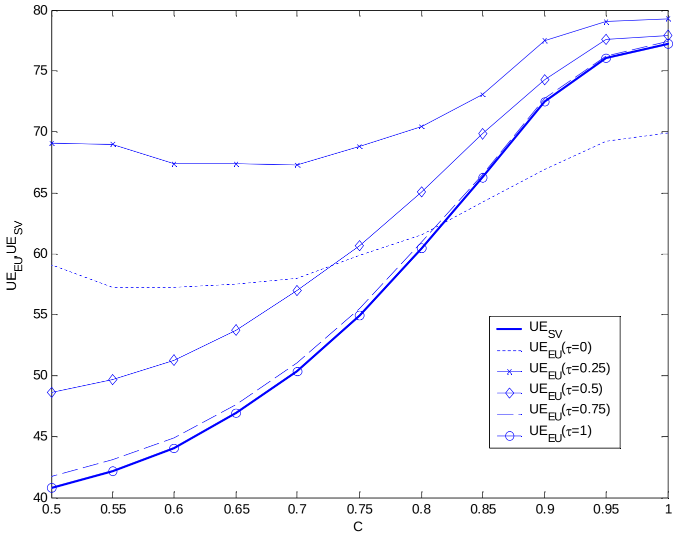

Figure 4.1: Utilitarian efficiencies in setups with different degrees of correlation. The degree of reliability is not excessively low ($\varepsilon = 4$).

$\varepsilon = 0$ because it would involve dividing by zero, the choice was not to present the results with this value of ε. The results for $\tau = 1$ were not always presented either because the conditions for strategic voting imply that nobody votes that way strategically[13]. Randomness affects the voters' beliefs as much as the real preference profile when the variance of B^{jk} equals the variance of R^{jk}, i.e. when $2N=2\varepsilon^2$. A somewhat natural maximum value for ε is thus $\varepsilon = \sqrt{N} \approx 14.177$.

4.6.2 The degree of intensity correlation and the reliability of the signals

Figure 4.1 shows the utilitarian efficiencies in setups with different degrees of intensity correlation. It is easy to see from this figure that strategic voting is welfare-increasing in all setups except those in which $\tau = 0$, and C is higher than about 0.81. Choosing a range of reasonable values for parameter τ can be done by evaluating the percentage of voters that vote strategically. When $\tau = 0$, about 60 to 62 per cent of the voters actually gave a strategic vote. When $\tau = 0.25$ and 0.5 the percentage figures were about 10-12, and 1.5-1.8 respectively. Estimates for how common strategic voting is (e.g., Alvarez, Boehmke & Nagler

[13] Simulations were conducted with $\tau = 1$, however, in order to check the consistency of the computer model: the results for SV behaviour and EU behaviour with $\tau = 1$ have to be the same.

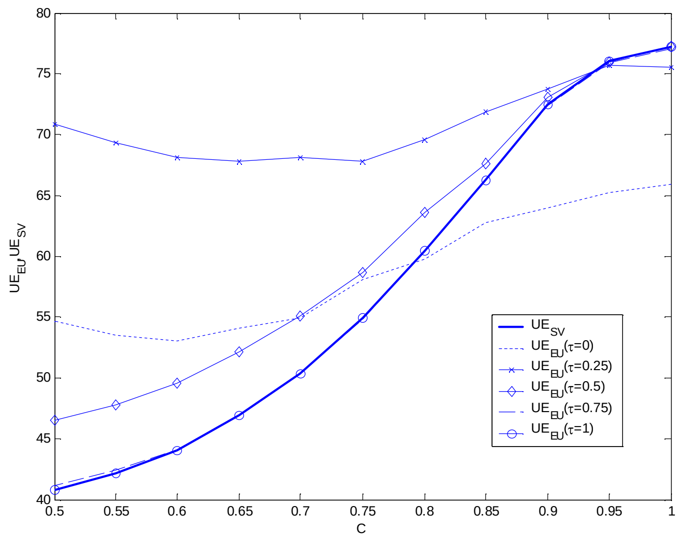

Figure 4.2: Utilitarian efficiencies in various setups with $\varepsilon = 16$

2006, Cox 1997) suggest that $\tau = 0.25$ is the most plausible value. Results that can be displayed with only one value of τ at a time will thus be reported as $\tau = 0.25$.

Figure 4.2 shows utilitarian efficiencies in setups with $\varepsilon = 16$. A comparison of Figures 4.1 and 4.2 shows that strategic voting is slightly more welfare-increasing when the reliability of the voters' information is high than when it is low. The reliability of the voters' signals thus turned out to be less important than expected, and less important than other parameters. As before, strategic voting was more welfare-increasing in setups with intensity correlation than in uniform setups.

4.6.3 What happens if some voter types do not engage in strategic behaviour?

The logic of counter-balancing implies in many voting rules that if some voter *types* never vote strategically, strategic voting may be welfare-diminishing. If the strategic votes for a candidate are not counter-balanced with strategic desertions of the same candidate, the voting results no longer adequately reflect the differences in preference intensities between the candidates. It is unlikely, however, that the welfare consequences of strategic voting are affected at all if the selection of those who engage in it is random, i.e. if there are no systematic differences in the behavioural dispositions of different types of voters. If it is

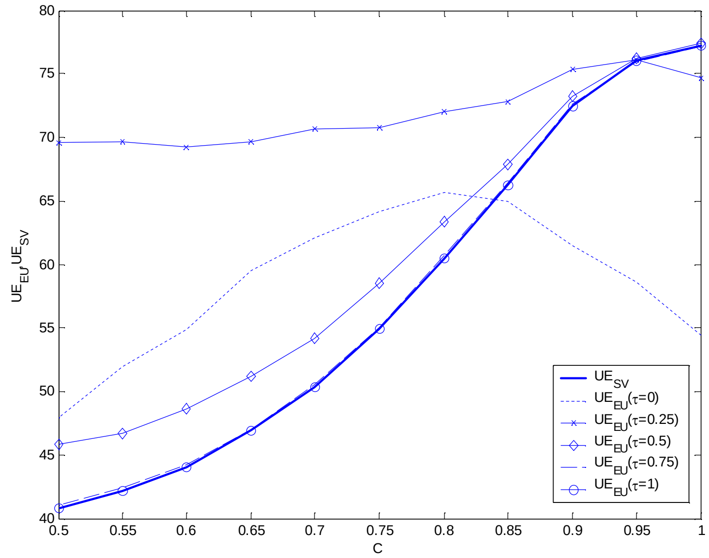

Figure 4.3: Utilitarian efficiencies with $\varepsilon = 4$ in setups with different degrees of correlation when type-four voters engage in SV-behaviour

assumed, for example, that a randomly selected half of the population of voters engages in SV behaviour, there will simply be less strategic voting, and the model will yield the same welfare effects as that in which all voters engage in EU behaviour, although these effects will be weaker.[14]

The Borda rule differs from other voting rules in that the beneficial welfare consequences of strategic voting do not depend heavily on whether all voter *types* engage in strategic behaviour or not. The reason for this is that a single voter type may confront two different strategic situations, and the incentive structures of these two situations provide *partial counterbalancing within a single voter type*. Although a voter *may* have a strategic incentive to report a higher Borda score than his or her preference ordering implies for a given candidate in situation 1, another voter of the same type *may* have a strategic incentive to report a lower Borda score than his or her preference ordering implies for this same candidate in situation 2. The conditions for strategic voting, together with the transitivity restrictions, however, imply that a single voter cannot have an incentive to vote strategically in both situations at the same time.

It also matters which voter type(s) do not engage in EU behaviour. A further look at Table 4.2 on p. 48 shows that the Borda score of candidate y may be both increased and decreased by the strategic actions of type-one voters. Coun-

[14] This was also verified in the simulations.

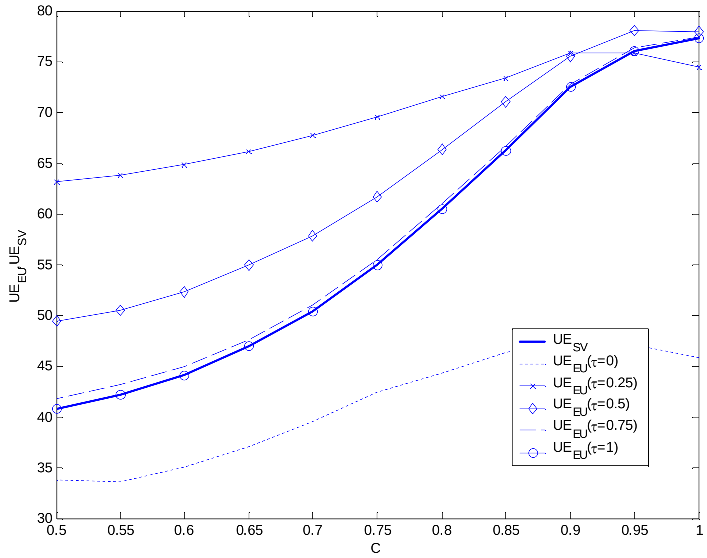

Figure 4.4: Utilitarian efficiencies when type-two voters engage in SV-behaviour

terbalancing thus works partially in the sense that the strategic votes for y are counterbalanced by the strategic votes against y cast by voters of the same type, but the strategic votes for z and those against x are not counterbalanced by the strategic actions of type-one voters. Note, however, that such counterbalancing within a voter type requires heterogeneity of preference intensities *and* beliefs because the two strategic situations depend on systematically different beliefs. It is thus unlikely that this kind of counterbalancing will occur if the voters have exact information on other voters' preferences.

Given the preference structures in setups with correlation between intensities and voter types and the directions of change as presented in table 4.2, utilitarian efficiencies should be highest in setups with correlation in which only type-four voters engage in SV behaviour, and lowest in setups in which only type-two voters engage in SV behaviour. The difference should be rather small, however, if only one voter type refrains from strategic voting, because in that case there remain many voters who may have the incentives to vote strategically for all three candidates.

Figure 4.3 shows utilitarian efficiencies in setups in which type-four voters engaged in SV behaviour and other voter types engage in EU behaviour, and figure 4.4 shows similar results when type-two voters engage in SV-behaviour.

Comparison with figure 4.1 shows that, although the utilitarian efficiencies are somewhat lower when some voter types do not engage in strategic behaviour,

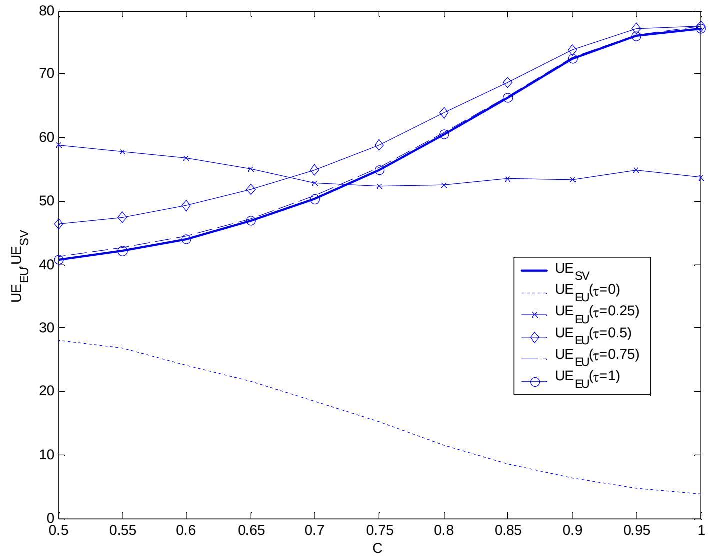

Figure 4.5: Utilitarian efficiencies in setups in which voters engage in strategic voting only in situation 1

the effect is not particularly strong. Furthermore, the difference between the setups in which different voter types engaged in sincere behaviour is relatively small. Setups in which type-one and type-three voters engaged in SV behaviour were also investigated. As expected, the utilitarian efficiencies were broadly speaking between those derived from the extreme cases in which type-two or -four voters engaged in SV behaviour.[15]

4.6.4 Sincere and non-sincere manipulation

Van Hees and Dowding (forthcoming) have recently argued that there are two kinds of manipulation, and that although one may be normatively suspect the other is less so. A voter engaged in 'sincere manipulation' gives a vote to a candidate j in order to increase the chance that *this* candidate will win, whereas one engaged in 'non-sincere manipulation' gives a vote to candidate j in order to increase the chance that *another* candidate k will win. Since van Hees and Dowding consider the Borda rule an example of a voting rule in which non-sincere manipulation may occur, it may well be justified to consider situation 1 as representing non-sincere manipulation and situation 2 as representing sincere manipulation.

[15] These results are available from the author on request.

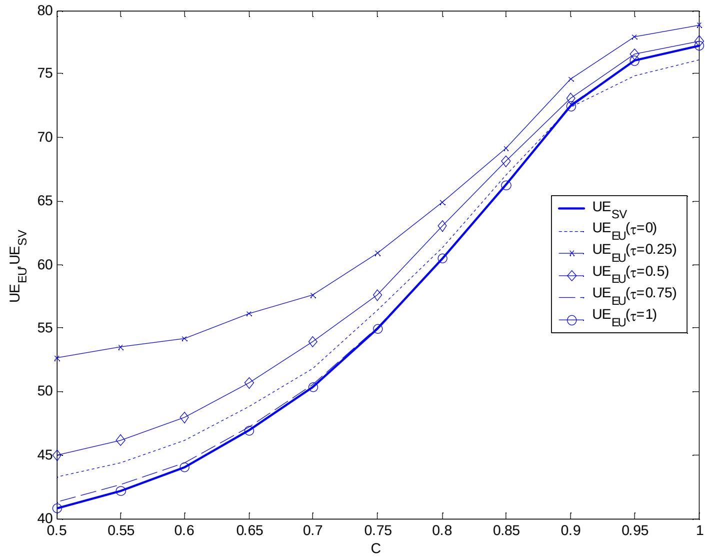

Figure 4.6: Utilitarian efficiencies in a setup where the voters engage in strategic behaviour only in situation 2

It is natural to ask how the welfare consequences differ between sincere and non-sincere manipulation given that only one situation affects voters' decisions in a setup. Figure 4.5 shows the utilitarian efficiencies from setups in which the voters engage in strategic voting in situation 1 but not in situation 2 (only non-sincere manipulation), and Figure 4.6 displays similar results for setups in which the voters engage in strategic behaviour only in situation 2 (only sincere manipulation). There are clear differences in welfare implications between the two situations, but it is rather difficult to say whether the results are in favour of sincere or non-sincere manipulation. With only non-sincere manipulation it would be better if the voters engaged in EU behaviour only if the correlation between intensities and voter types were strong, whereas sincere manipulation seems to be welfare-increasing irrespective of the degree of correlation.

4.7 Interpersonal comparisons

Since these results are based on utilitarian efficiencies, it is necessary to make *interpersonal comparisons of preference intensities* because it must be assumed that it is possible to add one person's utility to another person's utility. Since such comparisons are generally considered the most suspect for epistemic (choices do not provide easily interpretable information: (Myerson 1985)) and conceptual (they are meaningless with vNM utility functions) reasons, the welfare criterion has to be justified.[16]

Condorcet efficiency is the percentage of voting games in which the Condorcet winner is selected, given that it exists. Figure 4.7 displays Condorcet efficiencies. Many voting theorists would no doubt consider the result that strategic

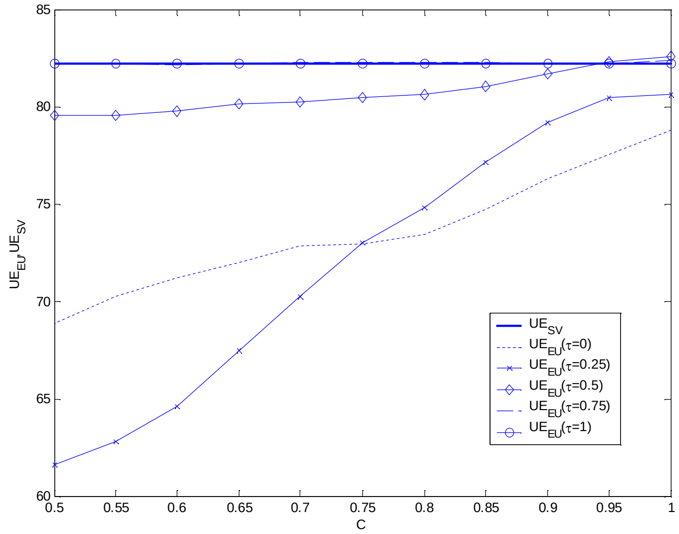

Figure 4.7: Condorcet efficiencies with $\varepsilon = 0.4$

voting is welfare-increasing acceptable only if Condorcet efficiency was used as a normative criterion, on the grounds that it does not require the comparability of different voters' utility scales. Given that Figure 4.7 shows unambiguously that strategic voting decreases Condorcet efficiency, taking this position would mean that the results show precisely the reverse of what was claimed: strategic voting will always be welfare-diminishing. However, such an interpretation of the results is not correct for the following reasons.

Refusing to use an intensity-based welfare measure in a model in which intensities affect the voters' behaviour implies a methodological bias. If intensities are important for individuals, they should be normatively important for the whole electorate. Strategic voting is beneficial *only* because it allows voters to express *intensities* indirectly, even under a voting rule in which such information is not explicitly collected.

If the result that strategic voting increases utilitarian efficiency obtains with *all* different and at least mildly reasonable preference scales, then it does not

[16]See Hammond (1991*b*, 2004) for surveys on interpersonal comparisons.

depend on any particular interpersonal comparison. If it is thus *robust* to interpersonal comparisons, we can be assured that we know something more about the consequences of strategic voting even though we do not know which interpersonal comparison is correct. The model was tested with various different interpersonal comparisons. Since the ways in which the results were tested for robustness, as well as the qualitative conclusions from the robustness analysis were exactly the same as those presented in Lehtinen (2007) for parliamentary agenda voting, the analysis will not be reproduced here. The robustness analysis showed that EU behaviour *remains welfare-increasing* irrespective of the interpersonal comparison used.

The simulations conducted thus far have featured random interpersonal comparisons of preference intensities because the utilities are derived from the uniform distribution on the [0,1] interval. It could always be argued that the choice of individual utility scales is arbitrary. This arbitrariness ultimately derives from the fact that it is impossible to obtain exact information on individual differences in utilities. Epistemological considerations thus indicate that we will never know which interpersonal comparison is correct. Robbins (1938) noted that even though a Brahmin's claim that he is ten times more capable of happiness than an untouchable may be repugnant, he cannot demonstrate his own more egalitarian view by scientific means. This would seem to imply that fixing a preference scale is entirely arbitrary because *any* scale will be equally good from a scientific point of view. This argument is valid, but it does not necessarily follow that we should not impose any bounds on individual utility scales because we are involved in the normative evaluation of a voting scheme, and in such an enterprise ethical judgments are also important.

Consider the following example with three voters A, B, and C.

A	B	C
$x(2)$	$y(\frac{1}{2})$	$y(\frac{1}{2})$
$y(\frac{1}{2})$	$z(\frac{1}{4})$	$z(\frac{1}{4})$
$z(0)$	$x(0)$	$x(0)$

Table 4.3: An example with three voters

The numbers in parentheses indicate cardinal interpersonally comparable utilities. Here the utility sums are 2, $1\frac{1}{2}$, and 0 for x, y and z, respectively. No doubt, many of us might think that y rather than x should be selected *even if* the utilities were interpersonally comparable and even if A's high utility for x outweighed B's and C's utilities for y. We would be willing to argue that voter A's great satisfaction from x does not compensate for the fact that *two* voters would be obtaining their worst outcome. We would thus be willing to say that y, the Condorcet winner, should be selected. If we are using this argument, however, we are looking at interpersonal comparability, although this is different from the utilitarian argument. The comparison consists in the idea that alternatives x and y are compared in terms of the *number* of individuals who would gain utility in passing from x to y as opposed to the number who would lose (cf. Hildreth 1953).

If we are willing to grant the normative relevance of preference intensities in the first place, the proper conclusion to be drawn from this example is not that Condorcet winners should be used, but rather that the utility scales cannot vary boundlessly. That the individual scales of utility are somewhat similar is based on the normative judgment that each individual's utility should weigh somewhat equally in the social- evaluation function.[17] Making the utility scales different for different individuals will thus accommodate the fact that different individuals are likely to care about the results of the vote to different degrees. However, making it *unlikely* that one voter's utility scale will be ten times wider than another voter's scale will prevent too much divergence from the normative one-man-one-vote principle. The variability in individual utility scales is not limited due to the belief that real people's scales do not vary all that much, but rather because this methodological choice provides a way of taking into account important normative considerations.

4.8 Conclusions

Strategic voting is welfare-increasing under the Borda rule in various configurations of assumptions. All the results reported here are derived from the *logic of counterbalancing*: intensively supported candidates are most likely to gain strategic votes and least likely to lose them. Ceteris paribus, correlation between voter types and preference intensities makes strategic voting more welfare-increasing.

It seems fairly likely that setups with intensity correlation correspond more closely to real-world conditions than the uniform setups. This would be the case if some candidates were typically fairly tolerable to a large number of voters even if they had about the same number of supporters that put them first in their preference ordering (and some other candidates would have a narrower support base). The results thus provide a further dimension to the claim made by various authors that the Borda rule selects reasonable compromises: the utilitarian winner is one kind of compromise candidate.

Although no explicit comparison of different voting rules is given in this paper, there is good reason to claim that the Borda rule has two advantages over some other rules. First, strategic voting seems to be welfare-increasing even if the voters have unreliable information on other voters' preferences. Secondly, the welfare consequences of strategic voting are beneficial under this rule even if different types of players have heterogeneous behavioural dispositions or manipulative skills. Even if some voter types do not engage in strategic behaviour, strategic voting increases utilitarian efficiency. The Borda rule yields high utilitarian efficiencies even when some voter types engage in sincere behaviour because counter-balancing functions to some extent even at the level of the single voter type. Voters of the same type may have an incentive to increase

[17] See Dhillon and Mertens (1999) for an axiomatic defence of this kind of utilitarianism.

(situation 2) or decrease (situation 1) the Borda score of their second-best candidate, depending on their beliefs and preferences, because there are *two* different strategic situations they may face under the Borda rule.

Since even 'non-sincere manipulation' is welfare-increasing under the Borda rule, it could be concluded that the title of this paper is neither a joke nor a metaphor: the Borda rule is also intended for dishonest people.

Chapter 5

Unrealistic assumptions in rational choice theory

Abstract

The most common argument against the use of rational choice models outside economics is that they make unrealistic assumptions about individual behavior. We argue that whether the falsity of assumptions matters in a given model depends on which factors are explanatorily relevant. Since the explanatory factors may vary from application to application, effective criticism of economic model building should be based on model-specific arguments showing how the result really depends on the false assumptions. However, some modeling results in imperialistic applications are relatively robust with respect to unrealistic assumptions.

Keywords: Unrealistic assumptions; Economics Imperialism; Rational choice; As if; Robustness

5.1 Introduction

Economics has become an imperialistic science. Economic methods are increasingly used for explaining phenomena in fields that have traditionally been occupied by other sciences. The term 'rational choice theory' has come to denote theories that apply economics to new fields of research.[1] Economics imperialism is thus a matter of extending theories based on rational individuals into new areas. Yet even economists used to argue that the homo oeconomicus assumption should not be used to model human behavior outside the domain of market institutions. However applicable the self-interest assumption is in economic domains, it seems to give a particularly poor account of individual motivation in

[1] The traditional game and decision theories are, of course, also called theories of rational choice.

other areas. Have people just become generally more rational, or are market institutions invading new areas of social life? If neither is the case, the assumptions of strict rationality and self-interest do not seem to be realistic in imperialistic applications of economic methods.

The accusation of unrealism is reflected in the common reactions against the use of economic models outside their traditional domain; people are not really 'selfish', social relations are more fundamental than economic relations, and imperialism thus puts the cart before the horse (see e.g., Zafirowski 1999). We find these arguments wanting, since they misconstrue the explanatory properties of rational choice models. Our objective in this paper is to analyze how unrealistic assumptions may or may not matter for particular explanatory enterprises. Our claim is that behavioral and psychological assumptions may have different kinds of explanatory roles in different rational choice models. Some applications rely on substantive psychological assumptions, whereas others require certain kinds of behavioral patterns, and some get by with only scant reference to the structure of the situation. Therefore, the legitimacy of imperialistic applications of economic theorizing has to be assessed on a case by case basis. What follows is an attempt to provide criteria according to which such assessments can be made. Our aim is thus not to offer a sweeping defense of rational choice, but to pinpoint the possible weaknesses upon which *effective* criticism can be based. Since the explanatory properties of many rational choice models rely crucially on the *robustness* of the modeling result with respect to problematic modeling assumptions, we especially emphasize the value of explicit robustness analysis in such criticism. We will argue that if a result is demonstrably robust with respect to the unrealistic psychological or behavioral assumptions, the falsity does not matter.

Since the assumption of self-interest has been considered particularly unrealistic in non-economic domains, and since this has been explicitly put forward as an argument against economic applications in other social sciences, our arguments are most relevant in terms of discussing the legitimacy of economics imperialism. Nevertheless, what we say about unrealistic assumptions is, in principle, applicable to any social science, including economics itself.

The structure of the paper is as follows. We start by providing an account of the role of microfoundations and folk-psychology in rational choice theory in section 5.2. We note that a literal intentional or psychological reading of rational choice models requires the mediation of real intentional processes. We will then argue that since the assumptions of expected utility maximizing behavior and self-interest are merely templates for constructing empirically interpretable rational choice models, appraising the realisticness of these assumptions makes sense only if the appraisal concerns the arguments used to operationalize these assumptions into a substantial model. It is pointless to criticize or approve of these assumptions in the abstract because they are empirically empty. Since it is clear that not all rational choice models can be interpreted literally as accounts of intentional actions of agents, we proceed in the next section to discuss the various ways of interpreting as-if clauses that often accompany rational choice models. As-if clauses are usually thought to refer to the use of unrealistic assumptions. We argue that this need not be the case, since assumptions can be *psychologically unrealistic* in that they attribute implausible thought processes to the agents, yet at the same time be *behaviorally realistic* in the sense of correctly describing individual behavior. The possibility that a rational choice model is behaviorally realistic even though psychologically unrealistic is particularly relevant for models in which the structural constraints rather than individual psychological processes are the causally most relevant factors. In section 5.4 we discuss the way in which it is possible to analyze which factors are really relevant in a model by providing an account of robustness analysis. We will thus look at ways in which the validity of a modeling result may be insensitive to false assumptions, and argue for the importance of explicit robustness analysis by discussing some well-known examples from rational choice theory. One largely overlooked reason for the success of rational choice models is the fact that they often provide explicit frameworks for evaluating the normative acceptability of various institutional arrangements. Section 5.5 discusses the role of unrealistic assumptions in such comparative institutional analyses. The sixth section concludes the paper.

5.2 Operationalizing folk psychology

Those who use economic methods in new fields of research often justify their approach by arguing that their contribution consists of providing *microfoundations*.[2] Generally speaking, microfoundations give an account of the individual behavior of the agents underlying the aggregate result in the institutional context under study. Insofar as rational choice theorists and economists present explicit requests for microfoundations, and they often do, they seem to be in the business of making theories *more* rather than less realistic. The perceived problem with them clearly lies in the fact that they are so intimately coupled with rational choice that they, in turn, seem to be unrealistic. If rational choice is taken to be based on ordinary intentional psychology, the corresponding charge of psychological unrealism is usually that it is too exacting or too narrow a formalization: rational choice models either assume unrealistically sophisticated cognitive capacities or leave out important psychological non-goal-oriented causes of action. In this section we discuss the most common accusation concerning the lack of realism: the consistent pursuit of selfish interests is not psychologically realistic.

Why could we not provide irrational microfoundations? The sociologist Raymond Boudon (1998, 2003) argues that if a phenomenon can be shown to derive from individuals' rational choices, there is nothing left to be explained. It is enough to show that it can be understood in terms of rational choice theory. John Harsanyi (1982*a*, 1982*b*) goes even further in claiming that a normative

[2] However, the links between standard microeconomic consumer theory and methodological individualism in any substantial sense are strenuous. If there are any individuals to be found in micro-theory, virtually nothing follows from their utility maximization at the aggregate level. This is the content of the famous Sonnenschein-Mantel-Debreu results (Kirman 1989).

theory of rationality is necessary even for explanations of irrational behavior. Behavior is thus always either simply correct or deviant. If the latter, a further explanation is needed. This perceived explanatory asymmetry becomes understandable if one thinks of rational choice theory as essentially a *formalization* of folk psychology. What is meant by 'folk psychology' depends largely on the discursive context. Sometimes the term is used to refer only to the abstract interpretive *schema* of beliefs and desires, sometimes to all pre-theoretical interpretive *practices*, and sometimes to a fully fledged pseudo-scientific folk theory of human cognition.[3] However, here we limit our use of the term to refer to pre-theoretical psychology based on intentional states of belief and desire, although other kinds of mental states, such as emotions, are also often included in the concept.

Thus conceived, rational choice theory would be nothing more than a formal way of analyzing purposeful, intentional action. This view is explicitly endorsed by prominent rational-choice-oriented political scientists such as John Ferejohn and Gary Cox (1999), and by the sociologist James Coleman (1990).[4] The supposedly inherent intelligibility of intentional action would also provide a good reason for stopping the micro-foundations project at the level of individual behavior because there would be no more black boxes left to be opened (Boudon 2003). Intentionality would be the rock bottom of social inquiry. If we adopt this interpretation of rational choice as a formalization of folk psychology, the question to ask about imperialistic applications is whether they attribute credible intentional states and processes to the agents.

If economics were a formalization of folk psychology, it would seem natural to say that folk-psychological and decision-theoretical notions have the following structural similarities: probabilities correspond to degrees of belief and preferences to desires. However, it is not altogether evident that economists themselves accept a folk-psychological interpretation of rational choice. Consider some historical changes in their self-understanding that took place some half a century ago. Lionel Robbins (1952) and Frank Knight (1994[1935]) were the last major figures in mainstream economics to argue that it was based on psychological notions, which in turn were based on introspection. In contrast, modern decision theory was founded on the idea that people's preferences could be elicited by observing their *choices* in experimental settings using the reference-lottery technique or a similar procedure (see e.g., Hirshleifer and Riley 1992). This means that, in principle at least, constructing von Neumann-Morgenstern (vNM) utilities does not involve any mental attribution to the subjects. In fact, if the epistemic credibility of individual preferences in a given model is based on nothing else than folk-psychological reasoning, the model will

---

[3] We are indifferent between simulation theory and the theory theory of folk psychology, and in general wish to avoid commitment to specific philosophical positions concerning the nature of intentional states.

[4] Alexander Rosenberg (1992, see also 1980) argued some fifteen years ago that economics should be conceived of as a formalization of folk psychology. He also claimed that it was doomed to stagnation because we cannot obtain better and better information on people's desires and beliefs.

not be acceptable to economists. Consider economists' attitudes towards interpersonal comparisons of utilities. Introspection and empathetic identification are the primary means of making empirical interpersonal comparisons. However, economists strongly argue against the use of interpersonal comparisons of utilities in economic models because there are no choices comparable to those made in a reference lottery experiment from which we could construct a scale of interpersonally valid utility numbers.[5]

The formal machinery of expected utility theory and revealed preference theory form the content of *rationality as consistency* (or 'thin rationality') (Sen 1985). In principle, rational choice theory need not appeal to any kind of psychological factors because it relies on a purely formal account of consistent choice behavior. However, although rational choice models explain by invoking people's preferences and beliefs, the theorists hardly ever even try to obtain information on individuals' real preferences, conceived of as pure behavioral dispositions. Instead, economic applications of rational choice usually seem to rely on general intuitions and concepts (such as incentives) that are at least extensions of folk-psychological notions. Similarly, insofar as games are considered useful in describing, explaining or predicting some real-world phenomena, there has to be some way in which the payoffs are related to real people's payoffs (see Blackburn 1995; 1998). The basis for constructing a payoff structure for a game is usually the theorists' judgment, which in turn may well be based on identification with real subjects in some real situation (see Rubinstein 1991, Binmore 1994, pp. 98, 165). Here, knowledge of psychological factors such as the players' intentions and goals becomes important for determining their payoffs (Mueller 2004). Therefore, even though the theorists do not always admit it, *specific applications* of rational choice often rely on psychological assumptions, the credibility of which is crucial in ensuring the credibility of a specific model.

In contrast to formal consistency, *rationality as self-interest* (or 'thick rationality') (Sen 1985) provides a substantial account of what motivates real people. It is not best viewed as a *normative* account of what should motivate people. Indeed, as Sen (1987, p. 16) pointed out, it would be absurd to say that self-interest is a requirement of rationality. Rationality as self-interest is rather an *assumption* in empirical models that *specifies* what is assumed to motivate the individuals concerned. The assumption of self-interest takes many different forms. It is *operationalized*[6] in a fairly weak manner in the traditional microeconomic theory of market demand for commodities, because consumers

---

[5] Myerson (1985) makes this argument most forcefully, but it dates back at least to Roy Harrod (1938). Harrod also links it with the scientific respectability of economics. '[M]arshall says in the principles that the marginal utility of two pence is greater in the case of a poorer man than in that of a richer ... It may be urged that the economist hereby goes outside his proper "scientific" field ... Whether the nth unit of X has greater or less utility to a given individual may be made the subject of test. He can be given the choice. But there are no scientific means of deciding whether the nth of X has greater or less utility to individual P than the mth of Y has to another individual Q. The choice can never be put' (pp. 395-396).

[6] We use the term 'operationalization' as referring generally to the process of providing empirical content for the assumptions of self-interest and utility maximization. We do not mean to suggest that this is necessarily related with making theoretical concepts *measurable*.

are merely assumed to prefer having more rather than fewer (e.g., Hausman and McPherson 1996). Sometimes an economic rational choice model is able to provide an explanation or a prediction by merely describing the agents' economic incentives in a very broad way.[7] Examples of such models include the market for lemons (Akerlof 1970) and job-market signaling (Spence 1973). The plausibility of rational choice in economics is therefore, to a large extent, grounded on the relatively straightforward and non-controversial operationalization of thin rationality into thick rationality. Folk-psychological intuitions seem to be an important resource in assessing the realisticness of modeling assumptions for both advocates and critics of rational choice theory. The first crucial issue in the assessment of imperialistic applications is thus the psychological plausibility of the operationalization of self-interest.

What rational choice means in a given model depends crucially on the operationalization of self-interest, i.e. on the content of the preferences or payoffs and the structure of the choice situation. The assumptions of (expected) utility-maximizing behavior and self-interest, if abstracted from an institutional context, are merely *templates*, not proper targets for criticism or advocacy. More generally, the assumption of rationality does not imply much without auxiliary assumptions concerning the exact shape of the relevant individuals' utility functions and beliefs, and most importantly what they take to be the choice alternatives (Simon 1985, Kavka 1991). Consequently, non-specific criticism of these assumptions does not make much sense because few would be willing to seriously argue that there are no rational choice models with reasonably realistic assumptions. Hence, the arguments used in operationalizing the utility maximization or the homo oeconomicus assumption into an empirically specified hypothesis of individuals' behavior are the proper targets of critical discussion and model-independent empirical assessment.[8]

Let us now consider an example of a theory, Niskanen's (1971; 1975) model of the budget-maximizing bureaucrat, in order to illustrate the relevance of the arguments made above. Niskanen's model was designed to explain the growth of the public sector. The basic model is based on two important assumptions. First, bureaucrats are assumed to maximize the size of the budget, and second, the relationship between the bureau and the 'sponsor' is one of bilateral monopoly. The sponsor represents the relevant legislative committee, the whole legislature, or better yet, the whole population. The main conclusion is that the public bureau produces more than the sponsor wants.

This conclusion cannot be derived from the model if the bureaucrats do not maximize the budget. The credibility of the model is thus crucially dependent on the operationalization of self-interest. Furthermore, it is the weakness of such arguments that has led to the widespread suspicion that Niskanen's budget-maximization hypothesis is not particularly compelling. Assuming that bureaucrats maximize *utility* is not sufficient for predicting their behavior since

[7] Myerson (1999) characterizes the whole of economics as the science of incentives.

[8] In fact, such arguments have been subject to criticism even within economics. It was argued during the marginalist controversy in the 1940's that corporate profit maximization was not consistent with individual utility maximization by firm managers.

here utility maximization really means the rational pursuit of one's goals, and in order to predict what officials will do, we must know their goals (Downs 1967, p. 82). The bureaucrat in Niskanen's model is supposed to intentionally maximize the budget. In support of this assumption, Niskanen (1971, p. 38) postulated that a bureaucrat's 'salary, perquisites of office, public reputation, power [and] patronage' were all positively related to the size of the bureau. This is an example of an operationalization argument for a particular application of the self-interest assumption. Whether the desire for these goods overrides other, presumably more 'altruistic' goals is difficult to determine. However, there are also several empirically more tractable reasons why budget maximization is not a convincing operationalization of self-interest. First, the budget-maximizing hypothesis is only sustainable if it can be shown that the bureaucrats *can influence* the size of the budget in the first place (Udehn 1996, p. 75), but it is, in fact, the sponsor who holds the purse strings (Mueller 1989, p. 259). Secondly, the easiest way for top managers to increase their salaries is to be promoted to another bureau, and getting promotion may require slashing rather than maximizing the budget of the current bureau (Margolis 1975). This case clearly shows that assessing whether an account of action is adequate in a model also requires evaluating whether the institutional structure is specified in the correct way. Evaluating such institutional factors is thus part and parcel of the analyzing the empirical operationalizing assumptions.

If a rational choice model relies on substantial intentional attribution, the successful derivation of the explanandum from some set of reasonably operationalized preferences and a suitable solution concept is not sufficient for a successful explanation. Explanation relying on folk-psychological notions requires the mediation of practical reasoning as the explanatory mechanism. If, as is often the case, the idea that agents would consciously perform complex valuations between innumerable trade-offs is blatantly implausible, a literal intentional reading of rational choice is not possible. Rational choice theorists often justify attributions of seemingly unrealistic cognitive powers by claiming that individual errors in reasoning cancel out in the aggregate. The problem with this argument is that there is overwhelming empirical evidence suggesting that people have many kinds of cognitive *biases*, and that they therefore deviate *systematically* from the predictions given in a purely 'rational' model (see e.g., Camerer 1995 and Shafir and LeBoeuf 2002 for recent reviews). The validity of the canceling-out argument cannot be determined without explicitly evaluating whether the systematic irrationality implies systematic consequences in the model, although the burden of proof might be said to lie with the user of demonstrably unrealistic assumptions. Curiously enough, there are disappointingly few models that show how systematic violations of rational prescriptions affect the aggregate-level conclusion of a model outside the domain of traditional economics (but see Quattrone and Tversky 1988),[9] but the recent rise in popularity of behavioral economics shows that this may change in the near

[9] There are, however, some experimental results suggesting that cognitive biases do matter at the aggregate level (see e.g., Camerer 1995; 1998).

future.

An extreme case in point of a model that cannot be taken seriously when interpreted psychologically is the Beckerian theory of rational addiction, according to which present drug use is a rational investment decision in which the constantly increasing enjoyment provided by the drug consumption is weighed against the inevitable social and medical costs (Becker & Murphy 1988). In this case it is not even the cognitive limitations of the agents that make the model implausible, it is the obvious empirical misspecification of the causal factors in drug consumption, that renders it absurd (Rogeberg 2004). Calling every possible factor influencing the agents' behavior incentives or changes in the budget constraint is at best only metaphorical, and often confusing and misleading. Accordingly, it is often the case that the straightforward folk-psychological reading of a rational choice model is not the most sensible one in the first place, and that criticisms of psychological unrealism would therefore fall on deaf ears.

5.3 Behavioral realism and as-if methodology

Standard decision theory starts from the premise that if a person's preferences satisfy a set of axioms, his or her actions can be described *as if* he or she were maximizing a vNM utility function. Notice that it would be a category mistake to say that a person's actions could be described as maximizing a vNM utility function without the 'as-if' clause, since such a function is only a *representation* of a person's preferences (or choices). vNM utility functions are not unique, and can be determined only up to an affine transformation. This means that if a person's choices can be described as if she were maximizing a vNM utility function U, they could also be described as if she were maximizing the transformation V=aU+b (a>0). We might choose to assign a higher utility number to some choice alternative x under function U rather than under function V (U(x)>V(x)), but this obviously does not mean that the person has a stronger preference for x when his or her actions are described by U than when they are described by V.[10] This is why using the 'as-if' clause, when employed in an expected utility model, implies nothing about whether or not it is reasonable or realistic to assume that a person's actions can be described according to a utility maximization model. Whether it is realistic to use such a description depends merely on whether the person's choices satisfy the axioms.

On the other hand, the 'as-if' clause is often used in economic methodology specifically to refer to the adoption of unrealistic assumptions. Milton Friedman (1953) is famous for promoting as-if methodology. He argues that even though expert billiard players do not intentionally make complicated mathematical calculations for predicting the trajectories of the balls, their actions can be described as if they made such calculations using physical theory. To put this more precisely, *if we wish to explain the trajectories of the balls*, calculations provided by physical theory provide us with more useful information than the

[10] Luce and Raiffa's (1957) classic book remains one of the clearest expositions of these issues.

players' intentions do. The players' probably just think along the lines of, 'If I use light force with a little top-spin, I will have a good cue-ball position for the next shot over there.' On the other hand, accounting for the trajectories requires a minimal attribution of goals: the players must be assumed to desire to pot the balls.[11] It is perfectly possible, at least in principle, that the assumption that the players behave as if they had calculated the trajectories of the balls using physical theory is a realistic one. It may be realistic in the sense that it accurately accounts for the realized trajectories, but it is obviously unrealistic in terms of describing the players' mental states. Let us say that a theory containing an 'as-if' clause is *behaviorally realistic* if it allows for describing human behavior in a realistic way, and let us say that it is *psychologically (or intentionally) realistic* if the mental processes it evokes can be truthfully attributed to the agents.[12]

If a rational description of micro-behavior is at least roughly realistic in its behavioral assumptions, a model based on such a description could be used to explain or predict some interesting aspects of macro phenomena. As Uskali Mäki (1998) pointed out, whether the use of an as-if clause is truly instrumentalist depends on the explanatory set-up in question. Friedman's argument could thus be seen first and foremost as an argument against the necessity of finding out what the players' real thought processes are. It is an argument against 'verstehen' methodology, because it is based on the futility of finding out players' intentional states. Fritz Machlup puts it as follows:

> The 'extreme difficulty of calculating', the fact that 'it would be utterly impractical' to attempt to work out and ascertain the exact magnitudes of the variables which the theorist alleges to be significant, show merely that the *explanation* of an action must often include steps of reasoning which the acting individual himself does not *consciously* perform (because the action has become routine) and which perhaps he would never be *able* to perform in scientific exactness (because such exactness is not necessary in everyday life). To call, on these grounds, the theory 'invalid', 'unrealistic' or 'inapplicable' is to reveal failure to understand the basic methodological constitution of most social sciences. (Machlup 1946, p. 534)

For some purposes, the best description of people's behavior need not be in terms of their conscious intentional states. Machlup goes to great lengths to show that the profit-maximization assumption is, in fact, behaviorally realistic for the explanatory purposes of the theory of the firm, even though it is not psychologically realistic.

[11] Friedman also gives an example in which no goals need to be attributed to the 'agents'. The leaves of a tree position themselves as if they were seeking the maximum amount of sunlight.

[12] Our main aim is not to interpret what Friedman really said or meant. We just found it convenient to use his well-known views to introduce the distinction between behavioral and psychological realisticness.

Even if a rational choice model apparently attributes unrealistic psychological capabilities to the agents, this psychological unrealism may not matter if the resulting behavior is plausible, at least in the aggregate, for some other reason. The problem with arguments that promote the use of 'as-if' clauses for the purpose of making unrealistic assumptions about the intentional processes of individuals is that they provoke the need for additional arguments to find out the range of models or phenomena for which an 'as-if' clause is even behaviorally realistic. In the case of the billiard players, the assumption of trajectory-calculating behavior was further justified in the idea that the players would not be *expert* players if they did not hit the balls approximately as predicted by the physical theory. Friedman ultimately uses an evolutionary-selection argument to support the claim that *firms* always maximize profits but he does not present such an argument for *individual* rationality. Satz and Ferejohn (1994) suggest that rational choice theory is best *applicable* in circumstances in which selective pressures force *individuals* into utility-maximizing behavior. Even though they provide an otherwise structuralist account of rational choice theory, this particular argument seems to imply that players do have to behave rationally for rational choice theory to be applicable. We agree with Satz and Ferejohn that selective pressures are relevant for delineating the applicability of rational choice models, but the abstract argument in itself does not tell us when they are applicable and when they are not.[13]

Even though the idea that rational choice is formalized folk psychology seems to contradict the idea that rational choices are products of structural constraints, both are endorsed by many prominent rational-choice-oriented political scientists such as John Ferejohn (1975, with Fiorina; 1991; 2002) and Gary Cox (1999). However, accusations of incoherence are premature here, since we claim that both legitimizations of rational choice modeling may be valid given that different models explain different kinds of things in different ways. Although it may not make much sense to claim simultaneously that a particular model can be used because individual errors in *reasoning* cancel each other out, *and* that the structural constraints allow us to ignore psychological assumptions altogether, both kinds of claims can quite reasonably be made in different modeling situations. However, appeals to structural explanation via markets or selection ring hollow in cases in which there are no market institutions and no adequate selection pressures.

An example of a selection argument for maximizing behavior that is at least initially plausible is the market account of the supply of religious services, especially in such a little regulated and relatively diverse region as the US (see e.g., Finke 1997; Iannaccone 1997). Not every preacher, sect leader or cultist needs to be a cynical exploiter, nor indeed a perfectly honest but economically calculating entrepreneur. However, in an environment in which new 'religious firms' are constantly being established, and in which each one has to live on the donations of its members, the demand for religious services has to be met

[13]Money-pump arguments have often been coupled with selection pressures. See, however, Cubitt and Sugden (2001) for a critical account of such arguments.

as efficiently as possible for a particular church to survive.[14] In this case, talk of a market is not totally devoid of substance, and there indeed seems to be competition creating selection pressure that could structurally account for the apparently rational behavior of religious firms.

So far we have argued that the utility numbers in rational choice models are practically never actually elicited from choice behavior for the purposes of explaining particular real-world phenomena. The main reason for this is that it is notoriously difficult to obtain accurate and reliable information on individual utilities and beliefs, and practically impossible to use such information for directly predicting or explaining behavior in particular situations.[15] The operationalization of self-interest thus usually relies on folk-psychological intuitions and postulated institutional structure. However, these assumptions are empirical in the end, and should receive much more *model-independent* empirical attention than they usually do.[16] Nevertheless, unrealistic psychological or even behavioral assumptions sometimes do not need excuses in the first place, since they may not have any real explanatory role in a model. We explore this possibility further in the next section.

5.4 Explaining without preferences

If economic models are used beyond the traditional boundaries of economics, it would seem natural to assume that the assumption of self-interest is difficult to apply because people would seem to act in a more self-interested manner in markets than in matters of politics, family, crime, or religion. However, it is possible that it has virtually no explanatory role in these models regardless of the manner of operationalization. As noted above, one way to account for this is to say that it is really the structural constraints on people's behavior rather than self-interest that are the most crucial causal factors. The assumption of self-interest is not explanatorily important in a model if it can be replaced with another behavioral assumption without changing the analytical results. If it can thus be replaced, the model is *robust with respect to the behavioral assumption.*

Herbert Simon is famous for emphasizing the need to take the role of information into account in the study of individual behavioral dispositions and

[14] In this case, the more interesting results of the theory concern the instruments used to keep customers loyal to a particular firm and to avoid free-riding in the religious collective.

[15] Mosteller and Nogee (1951) were the first to try, and the first to fail, to do this.

[16] Many rational choice theorists subscribe to some version of the hypothetico-deductive view of theory testing. One version of this doctrine is usually attributed to Friedman (1953). He emphasizes that the realisticness of the assumptions does not matter because the empirical testing of predictions derived from a theory is the only ultimate arbiter for evaluating them, and that assumptions cannot be tested independently. Whatever plausibility this view has with respect to models or theories that really allow for deriving testable predictions, it is mere window-dressing when the theories are not in fact tested. We agree in broad terms (but only in broad terms, see e.g., Monroe (1997)) with Green and Shapiro's (1994) assessment that rational choice theorists have had only limited success in terms of empirically tested and verified theories. These issues concerning testing are relevant to the topic of this paper, but we cannot delve too deeply into them here because they are broad enough to deserve much more comprehensive treatment.

calculative capabilities. However, it is worth noting that Simon is commonly taken to entertain the view that explaining individual behavior is the goal of rational choice theory (e.g., Langlois 1986), while most rational choice theorists and economists explicitly argue that the theory is designed for explaining aggregate-level phenomena. It is generally agreed that satisficing is a more realistic theory of individual behavior than maximizing, but as long as the modeling results are not affected by this difference, the fact that maximizing is analytically more convenient suffices to explain why Simon's theory is not widely applied. Instead of having a major causal role, constraints on possible preferences are often more important in the *derivation* of the results. Correspondingly, the formal properties of preferences (e.g., transitivity) are often more important than the assumption of self-interest. This is because they ensure the analytical tractability of the equilibrium, which is seen as necessary for any understanding of the aggregate macrobehavior of interest.

In analytic model building, robustness has a primarily epistemic role in providing support for the claim that a *modeling result* (an equilibrium allocation or a dependency between variables derived from comparative statics) is not a mere artifact of particular modeling assumptions. Robustness in this epistemic sense is a measure of reliability with respect to erroneous assumptions and is therefore first and foremost a property of models, not of a phenomenon or a process (see Wimsatt 1981). In contrast, explanation requires dependence on rather than independence of the modeling result over some other features of the model, and that the dependence actually corresponds to a (causal) relation in the world (Woodward 2003). We are thus able to identify the explanatory relationships in a model or a family of models by examining how hypothetical changes in the values of the variables or parameters would change the analytical results, and what these changes would leave intact. For example, if in a particular model a dependency between the institutional setting and a resource allocation is robust with respect to various behavioral assumptions, the model could be used to structurally explain empirical phenomena (allocations) even if the behavioral assumptions are unrealistic.[17] As we will show, robust results often correspond to very general and resilient systemic properties of processes or mechanisms, but this empirical resilience has to be distinguished from robustness in the epistemic sense used here.

Let us now consider a few examples from economics and the social sciences in order to see the relevance of robustness considerations. It is often claimed that the predictions of neoclassical microeconomics are best borne out in situations of anonymous and essentially non-strategic market exchange. This good fit holds even in controlled behavioral experiments. However, this in itself does not yet show that actual self-regarding calculation or even consistent maximizing behavior is necessary for market outcomes. Gary Becker (1962) shows how the main results of standard demand theory can be derived even if people are assumed to try to stick to whatever choices they made previously or choose

[17]In a loose sense, the institutional setting "programs for" the allocation (Jackson & Pettit 1992). However, we prefer not to limit our discussion on robustness to cases in which the possible microbehaviors constitute the explanatorily relevant macrofeature.

randomly. Becker's insights are vividly demonstrated by Gode and Sunder's (1993) computerized market experiments, in which "zero-intelligence" computer sub-programs submit random bids and offers in a double auction resulting in aggregate allocative efficiency. In these exercises it is really the budget constraint rather than any behavioral assumption that is the explanatory factor responsible for the appearance of rational exchange.

Thomas Schelling's (1978) famous 'checkerboard' serves as a non-market example of a model that can explain macrophenomena with only very weak constraints on the agents' preferences. It is a paradigm of explanation according to 'the logic of the situation'. It can explain racial segregation even if the preferences concerning the racial make-up of one's neighborhood are extremely weak: for non-intended segregation to emerge, it is sufficient that people prefer not to live in a neighborhood in which the substantial majority of inhabitants belong to another race. Here the preferences do matter, since total indifference to the racial make-up of one's neighborhood would not lead to segregation. Nevertheless, it is the logic of the situation that is responsible for the amplification of even mild aversion towards being in the racial minority to clear-cut aggregate segregation. The segregation result comes about with a wide range of preferences for living in a neighborhood with at least some other members of the same race. The modeling result is thus quite, although not completely, robust with respect to behavioral assumptions.

Duverger's (1954) *law* states that the plurality rule with one elected representative per electoral district (first-past-the-post) leads to a 'two-party system'. More precisely, such electoral rules lead to a distribution of votes with two large parties in parliament. This result is robust with respect to different specifications of individual behavior because it may come about through the so-called *mechanical effect*: a larger-than-proportional share of the seats automatically goes to the two large parties under the plurality rule. Again, however, the strength of the result depends on individual behavior through the so-called *psychological effect* (see e.g., Riker 1982, Fey 1997). If voters who would vote for a small party realize that their candidate has no chance of winning, they may vote strategically for one of the major parties rather than waste their vote on a hopeless case. Such behavior obviously strengthens Duverger's law.

Anthony Downs (1957) argued that parties choose policies in order to obtain votes in elections, rather than obtain votes in order to formulate policies (p. 28). He showed that this followed from applying the assumption of self-interest to party politics. Party officials are assumed to be interested in obtaining power, i.e. in getting into government. Once they do so, they want to remain there. He then showed that if we assume that party positions can be represented as a one-dimensional continuum (from left to right), the two positions will converge in the middle of the distribution of voters (pp. 115-117). The reasoning behind this is that each party loses almost[18] no voters by moving toward the centre, but gains some votes from the other party.

[18] The qualifier 'almost' derives from the fact that the parties lose those who are too alienated to vote at all.

Since the fact that the major party positions are located close to each other was well-known before Downs, his main contribution was not in making this observation, but rather in showing that this would be the outcome of rational behavior on the part of the parties (p. 93). Although the model is so familiar by now that it seems obvious, Downs had to derive the consequences of rational choice so that we could see how the logic of party positions works. We can thus see, once again, that the assumption of self-interest does not imply all that much in itself.

Paul MacDonald (2003) argues that a scientific realist should insist on the realisticness of behavioral assumptions since realism assumes that real processes are at work in successful scientific endeavors. However, if the structural constraints are the primary explanatory factors in a rational choice model, and if the model is robust with respect to behavioral assumptions, its causally relevant assumptions do not concern them. It is thus perfectly possible to be a scientific realist and acknowledge the scientific value of models with unrealistic behavioral assumptions if the results are robust. MacDonald's argument would be plausible only if realists did not accept *any* kind of unrealistic assumptions. This, however, is just blatantly false about the realist position. Mäki (1989; 1992, a well-known realist engaged in the philosophy of economics, has emphasized the distinction between realism as an attribute of a theory's assumptions, and realism as a general philosophical doctrine.

This brings us to a related issue. As some authors have remarked, although it is possible to judge whether an assumption is intuitively realistic in isolation, it is not possible to determine in isolation whether the lack of realism matters for the explanatory purpose in question (cf. Friedman 1996). For example, an assumption may be realistic for one level of aggregation or explanatory purpose, but unrealistic for another (Levins 1993). Hence, in order to evaluate whether the problematic assumptions really matter for a particular explanatory purpose, judgments of psychological or behavioral realism should always be specified in terms of a *contrast*, i.e., an alternative psychological or behavioral assumption. Since an explicit robustness analysis always provides a contrast, analysis of a model's robustness with respect to the behavioral assumptions is a necessary part of all assessments of model validity. As with appeals to as-if utility maximizing, appeals to robustness should be backed up by additional arguments or, preferably, demonstrations of robustness. Since problematic assumptions are included in a model in order to ensure its analytic tractability in the first place, demonstrating that a modeling result is robust with respect to unrealistic behavioral assumptions is usually easier said than done. Robustness analysis is thus often exceedingly difficult, or downright impossible, to carry out with analytical models.

It would seem to follow from these considerations that proper assessment of whether behavioral assumptions are realistic requires a fairly thorough understanding of the particular model under investigation. For better or worse, vague appeals to the idea that assuming self-interest is intuitively unrealistic in matters of family, politics, and crime, for example, have not convinced and will not convince rational choice theorists, who have repeated ad nauseam that 'you cannot beat something with nothing'. If this metaphorical claim implies that a critic of rational choice assumptions is always responsible for presenting a fully-fledged alternative model, we cannot say we agree. It would be quite bizarre if the only acceptable way of criticizing the assumptions of a rational choice model would imply taking part in the rational choice enterprise. Nevertheless, the possibility of robustness with respect to a behavioral assumption implies that the opponent of rational choice ought to provide some kind of alternative account of what would happen in the situation in question if the problematic assumptions were changed.

For example, Akerlof and Yellen (1985) present a robustness analysis of some cases in economics in their well-known article. They show that in some models the punishment incurred by agents for irrational behavior is mild to the point of being non-existent, but the consequences of the model change dramatically due to such 'near rational' behavior. The reception given to this article shows that if a model demonstrably lacks robustness with respect to problematic behavioral assumptions, rational choice theorists will take the criticism seriously. However, we think that the burden of proof should again lie on those insisting on using blatantly unrealistic behavioral assumptions in the first place.

5.5 Unrealistic assumptions and comparative institutional analysis

Let us now discuss an altogether different class of models and arguments for their use. Our aim is to provide an account of the role of judgments concerning unrealistic assumptions in designing institutions. An important function of the social sciences is to provide guidance for decision-makers about how various institutions function. We believe that the success of rational choice theory is at least partly attributable to the fact that it has provided a framework for analyzing the *welfare implications* of various institutional arrangements. Such exercises are obviously valuable even though normative evaluation is not strictly speaking an empirical scientific endeavor.

The structure of the framework is the following. It is assumed that there is a set of individuals with a fixed set of preferences, i.e., they are described in terms of a *preference profile*.[19] Rational choice models are then focused on deriving the aggregate-level consequences or outcomes of the different behaviors that the various institutions induce. These outcomes may then be evaluated in terms of the preference profile. The framework is thus able to assist in institutional design by providing results that have the following form: institution X fares better than institution Y with respect to problem P because X satisfies individual preferences better than Y in situations relevant to P. Let us now consider how unrealistic assumptions concerning the preference profile (rather than the motivational basis) could be justified in such analysis by taking voting

[19] The preference profile is, roughly speaking, a collection of all individuals' preferences, i.e. it contains one preference ordering for each individual.

theory as an example.

It is often convenient to compare various different institutional arrangements by assuming that the individual preferences are randomly generated. For example, there is a tradition in voting theory to study the performance of different voting rules by randomly generating a profile of individual preferences. The voting rules are then evaluated on the basis of how often they select the best candidate, which is defined on the basis of those random preferences (see e.g., Merrill 1984, 1988). There seems to be complete unanimity among voting theorists that randomly generated preference profiles do not resemble real electorates even remotely. Nevertheless, this assumption is considered adequate for comparing *different* voting rules. It is obviously not adequate for comparing *voting qua institution* to other possible institutions such as markets or demand-revealing mechanisms. The idea behind this argument is fairly simple. It makes sense to use a random preference profile rather than one that endeavors to imitate real profiles because the analyst is mainly interested in comparing how different voting rules can cope with some particular problems such as path-dependence or excessive power of the agenda-setter, and random profiles are known to create the maximum amount of problems in voting rules (see e.g., Tsetlin, Regenwetter & Grofman 2003). Using a *deliberate* distortion of reality, a caricature (Gibbard & Varian 1978), allows for examining how different voting institutions deal with *particular* problems. Using unrealistic assumptions may thus have a reasonable methodological function even if we know how to describe reality in a more realistic way (i.e., if we know what kind of a distribution truthfully describes real electorates). It is also obviously necessary to keep the preference profile fixed in the different voting rules because otherwise the outcomes could not be compared.

The role of judgments concerning the realisticness of assumptions in institutional design is the following. Voting rules (and many other institutions) can be evaluated according to different criteria. Since the ultimate verdict on the acceptability of a given institution depends not merely on how the institution fares with respect to various criteria, but also on how these criteria are *weighed*, the results of a comparative institutional study using random profiles may be misleading if the selected criteria are not very important. It becomes necessary to assess the realisticness of the random profile assumption when we need to formulate an *overall judgment* concerning different voting rules in order to make policy decisions.

There is also an important argument for promoting interdisciplinary integration that is based on behavioral rather than formal consistency across various fields in the social sciences. Public choice theorists, in particular James Buchanan (1972; 1986, pp. 36-38; 1989, pp. 29, 64f), have argued that individuals should be assumed to act in a self-interested manner in analyses of political institutions, because otherwise the results from such analyses could not be compared to those that can be obtained through market institutions. Buchanan argues that consistency is particularly important in the normative comparison of market versus public organizations (Buchanan 1986, pp. 32-38; see also Diermeier and Krehbiel 2003). These arguments could be seen as applications of the general idea presented above: in order to compare institutional arrangements, we need to keep some admittedly unrealistic assumptions fixed across those institutions. Similarly, if the assumption of self-interest is unrealistic, but we design institutions assuming that individuals are self-interested, the institutions will be badly designed because they will be responsive to irrelevant issues (cf. e.g., Frey 1994, Hausman 1998).

5.6 Conclusions

Economics and rational choice theory have traditionally been viewed as epitomes of methodological individualism. The explanatory schema that would seem to follow from this conception is the following. The rational choice theorist begins with a set of preferences. He or she then formulates a model by specifying an institutional context and the individuals' possible strategies in this context. By solving the model for equilibrium he or she derives an explanation or prediction for some question of interest. This rough schema is otherwise correct, but it provides a misleading picture of the role of preferences in rational choice theory. The rational choice theorist does not begin with a set of preferences in the sense that he would go out and try to find out what the real preferences are. Neither does he start with data on preferences that someone else has collected. It is rather that, in order to explain an aggregate or macrolevel phenomenon, he postulates or argues that the preferences must have such and such properties because of folk-psychological or institutional considerations. Whether unrealistic assumptions concerning the properties of preferences really matter depends on where the explanatory power in the model in question resides.

Some imperialistic models offer explanations by aggregating interdependent intentional action. The crucial issue is the plausibility of the way in which self-interest is operationalized. Assessing the operationalization arguments includes formulating a judgment about the plausibility of the intentional states (beliefs and desires) attributed to the agents and of the psychological possibility of the practical reasoning required by the solution, and also of the way in which the institutional background is included in the model. If the rationality is alleged to be of the as-if kind, the crucial question to ask is whether the structural features allegedly guaranteeing the rationality of individual behavior are indeed in place. Finally, the result of an imperialistic model may be claimed to be fully or partly robust with respect to the unrealistic assumptions, and in such cases it usually corresponds to a very general and abstract systemic property. In any case, the most effective (theoretical) way of criticizing an imperialistic model would be to demonstrate that unrealistic psychological or institutional assumptions actually matter in terms of the conclusions, although we agree that the burden of proof should in principle lie upon those using the unrealistic assumptions in the first place. Explicit robustness analysis would thus be a valuable theoretical means of assessing the merits of economic models applied outside the traditional field of economics.

Chapter 6

Conclusion: A farewell to IIA

Arrow's Independence of Irrelevant Alternatives (IIA) has been under criticism for decades for not taking account of preference intensities. Computer-simulation results by Aki Lehtinen concerning strategic voting under various voting rules show that this intensity argument does not need to rest on mere intuition. Voters may express intensities by voting strategically, and that this has beneficial aggregate-level consequences: utilitarian efficiency is higher if voters engage in strategic behaviour than if they always vote sincerely. Strategic voting is thus unambiguously beneficial under a utilitarian evaluation of outcomes. What has been considered the main argument for IIA turns out to be one against it. This paper assesses the implications of these results for interpretations of Arrow's theorem and the Gibbard-Satterthwaite theorem in a discussion on the methodological and philosophical arguments concerning preference intensities and IIA.

Keywords: Strategic voting; IIA; Utilitarian winner; Observability; Strategy-proofness

6.1 Introduction

The normative and descriptive relevance of preference intensities and the normative validity of Arrow's (1963) Independence of Irrelevant Alternatives (IIA) have been under debate for decades in social choice theory. It has been argued since its inception that IIA does not take preference intensities into account.

A choice-theoretical definition of IIA is as follows. Let $\mathbf{p}$ and $\mathbf{p}'$ denote profiles of individual preferences: $\mathbf{p}$ assigns a preference ordering $\succ_i$ for each voter $i \in I : \mathbf{p} = (\succ_1, \succ_2, ..., \succ_N)$. Let $\mathbf{p}_{|Y}$ denote the restriction of the profile

$\mathbf{p}$ to the subset Y of X. Let $C(\succ, S)$ denote the social choice from profile $\mathbf{p}$ on S.

Independence of Irrelevant Alternatives: If for all $x, y \in S$ and all individuals,

$$\mathbf{p}_{|\{x,y\}} = \mathbf{p}'_{|\{x,y\}} \rightarrow C(\succ, S) = C(\succ', S). \tag{6.1}$$

In other words, if the two profiles $\mathbf{p}$ and $\mathbf{p}'$ rank each pair of alternatives in the same way, then the social choice from the two profiles should be the same.

Donald Saari (1998, 2001, 2003*a*) has recently argued that IIA is not normatively acceptable because voting rules that satisfy this condition fail to respect the rationality of voters. His proposal is to replace it with a condition called *binary intensity IIA* (see also Saari 1995). This requires that the aggregate ranking of each pair of alternatives is to be determined by each voter's relative ranking of that pair, and by the intensity of this ranking. The latter is determined by how many other alternatives are ranked between them. Saari argues that the Borda count (BC) satisfies this condition, and respects the rationality of voters by asking them to report a full preference ordering. Naturally, the BC does not satisfy IIA. Those who have not been willing to abandon IIA tend to emphasise its close link with strategic voting.[1] The BC is commonly considered to be highly manipulable.[2]

If there remain proponents of IIA, and if they view the debate related to it as an inevitable trade-off between rationality and intensities on the one hand, and susceptibility to strategic manipulation on the other, they have embarked on an enterprise that is doomed to failure. The simulation results in Lehtinen (2006*b*, forthcoming, 2007) suggest that utilitarian efficiency (the frequency with which the alternative with the highest sum of utility is selected) is higher if voters engage in strategic behaviour than if they always vote sincerely. Strategic voting is thus unambiguously beneficial under a utilitarian evaluation of outcomes. What has been considered the main argument for IIA thus turns out to be one against it. These results show that the traditional intensity argument against it does not need to rest on the mere intuition that it rules out intensity information. They illustrate how all of the voting rules studied take intensity information into account if and when IIA is violated through strategic voting, and *this has beneficial aggregate-level consequences.*

The discussion in this paper focuses on two interrelated topics concerning IIA and preference intensities. The first relates to the fact that Lehtinen's results, as well as the intensity arguments against IIA that were presented before Saari's contributions, were based on a cardinal notion of preference intensity.[3] Saari's notion of the *intensity level* upon which the binary intensity IIA is based is best characterised as an ordinal notion (Risse 2001, Dowding 2006). Given that Lehtinen's results are based on the cardinal notion, what they show is that cardinal intensities affect the results under all voting rules.

[1] See, e.g., McLean (1987, p. 154). Arrow (1983 [1977]*a*, p. 168) also makes this argument, and Saari (2001, pp. 45, 137) acknowledges it.

[2] Saari (1990*b*, 2003*b*) disagrees with this judgment, however.

[3] I mean contributions by Hildreth (1953), Rothenberg (1961, pp. 132-136), Coleman (1966), Campbell (1973), Ng (1979) and Mackay (1980), for example.

I will show that IIA is also violated in amendment agendas by analysing an example taken from Lehtinen (2007). This example is used to show that majority-rule agendas also take preference intensities into account if voters engage in strategic behaviour. It follows that Saari's arguments concerning the transitivity of preferences and the intensity level should not be understood as providing support for the Borda rule and against the majority rule, even though they are convincing *qua* arguments against IIA.

Secondly, I will draw on the methodological and philosophical implications of Lehtinen's results on strategic voting in my interpretation of Arrow's theorem and the Gibbard-Satterthwaite theorem. Hence I will discuss the methodological and philosophical arguments concerning preference intensities and IIA.

The traditional criticisms against preference intensities can be formulated in terms of three arguments for IIA. The *strategic-voting argument* states that strategic voting is to be avoided, and a voting rule that satisfies IIA precludes strategic voting. The *observability argument* states that since it is possible to observe preference orderings, but not preference intensities or interpersonal comparisons of utilities, allowable information must be restricted to preferences for pairs of alternatives, and this is what IIA does: 'Modern economic theory has insisted on the ordinal concept of utility; that is, only orderings can be observed, and therefore no measurement of utility independent of these orderings has any significance [...] The condition of IIA extends the requirement of observability one step farther.'(Arrow 1983 [1967], pp. 75-6).[4] The *epistemological-moral argument* against preference intensities and for IIA states that cardinal von Neumann-Morgenstern (vNM) utilities should not be used in social-welfare judgements because they "reflect only individuals' attitudes towards gambling" (Arrow 1951, p. 9-11).[5] The idea here is that vNM utilities are not appropriate in this context because they inevitably contain attitudes towards risk.[6]

I will respond to these arguments as follows. My main argument against the observability argument is a *tu quoque*: I will show that ordinal utility is not observable in voting contexts either. I will thus not attempt to show that preference intensities are observable, or that we have particularly precise information on interpersonal comparisons. I will rather establish that observability cannot be used as an argument for ordinal and against intensities in evaluating voting rules. The epistemological-moral argument suffers from a similar shortcoming: voting choices reflect attitudes towards risk under all voting rules whether we like it or not. A reasonable voting model should explicitly take this into account rather than try to avoid the problem by using only ordinal utilities. Furthermore, risk attitudes are not measured in terms of utility functions in Lehtinen's

[4] See also Arrow (1963, p. 110).

[5] Arrow no longer puts forward this argument in the second edition of Social Choice and Individual Values (1963). See also Rawls (1971, pp. 172, 323) and Pattanaik (1968).

[6] Arrow has also justified the IIA condition by referring to its 'intuitive appeal' (Arrow 1983[1952], p. 51), arguing that it has 'strong pragmatic justification' (Arrow 1983 [1967], pp. 70, 76), and that it restricts the available information to feasible outcomes (1983[1952], p. 51, 1983 [1967], p. 76, 1983 [1967], p. 164). The first two are not discussed here because they merely appeal to intuition and thus do not seem to be genuine arguments. The third is discussed in connection with the observability and strategic-voting arguments.

models, but are formalised merely in terms of voters' beliefs.

I will take it for granted that intensities of preference are intrinsically relevant for evaluating voting outcomes normatively. I believe that voting theorists agree with this judgment[7], and I will thus not attempt to argue for it. This being the case, rebutting the observability and epistemological-moral arguments should suffice to establish that voting outcomes ought to be evaluated on the basis of utilitarian criteria. Then, given a utilitarian evaluation of voting outcomes, strategic voting should be considered an argument against IIA rather than for it, because strategic voting typically increases utilitarian efficiency (or average utility) as compared to sincere voting under all commonly used voting rules. The main reason for this is that voters' behaviour depends on preference intensities when they vote strategically but not when they vote sincerely: the utilitarian winner is likely to get more strategic votes than other alternatives. Information on preference intensities can only be obtained through strategic voting (cf. Coleman 1966).

6.2 The observability argument

Social choice theory has been criticised from the outset for ignoring preference intensities. Dahl (1956, p. 90) provides a typical example of such criticisms:

> By making "most preferred" equivalent to "preferred by the most" we deliberately bypass a crucial problem: What if the minority prefers its alternative much more passionately than the majority prefers a contrary alternative? Does the majority principle still make sense?

Those who have opposed the use of preference intensities and vNM utilities in social-welfare judgements have based their criticism on epistemological considerations. Here are Arrow's reasons for not incorporating preference intensities into social choice theory.

> The oldest critique of social choice theory ... is that it disregards intensity of preference. Even with two alternatives, it would be argued that a majority with weak preferences should not necessarily prevail against a minority with strong feelings ... The problem in accepting this criticism is that of making it operational. Theoretically, is there any meaning to the interpersonal comparison of preference intensities? Practically, is there any way of measuring them, that is, is there any form of individual behavior from which the interpersonal comparisons can be inferred? (Arrow 1977*b*)

Arrow introduced IIA in order to impose an observational requirement on social choices. The idea was that the available information has to be restricted to ordinal utilities because preference orderings are observable but intensity

[7] But see Plott (1976, pp. 541-2).

is not. Indeed, he makes it perfectly clear that cardinal utilities (preference intensities) would be important for social choice and welfare if we could observe them directly (Arrow 1987).

Arrow (1973*a*) argued that 'In a voting context, the ordinalist-cardinalist controversy becomes irrelevant, for voting is intrinsically an ordinal comparison and no more' (see also Frohlich & Oppenheimer 1999). Strasnick (1976, p. 243) formulates the difficulty of observing preference intensity in a voting context as follows: 'There is no sense in which the magnitude or degree or intensity of a choice is observable in the choice itself'.

This, however, does not mean that voters' choices are unaffected by preference intensities. An example in which the outcomes *depend on preference intensities* even under a voting rule (the majority rule with an amendment agenda) in which voters may express a preference directly only for pairs of alternatives is given in section 6.4. It shows that voting is intrinsically an ordinal phenomenon only in the sense that voters can merely state whether one alternative is better than another in pair-wise contests. However, if voters engage in strategic behaviour, preference intensities inevitably reflect their choices, and *affect the outcomes* even under a rule that seemingly collects only ordinal information. Before discussing this example I will present a rudimentary version of the Enelow's (1981) model of strategic voting, and discuss Saari's arguments concerning IIA.

6.3 Does the majority rule lose information on rationality and preference intensities?

It used to be common to distinguish between different aspects of the IIA condition.[8] The *independence* (or irrelevance) *aspect* refers to the fact that the social ordering (or choice) between any two alternatives must depend only on individual preferences for those alternatives, and not on individual preferences for other irrelevant alternatives. The *ordering aspect* requires that the social ordering (or choice) of any two alternatives must be based only on individual orderings of these alternatives and on nothing else. This aspect explicitly rules out preference intensities.[9]

It is generally acknowledged that if relative intensities of preferences are somehow available then the ordering aspect of IIA need not be accepted, but the irrelevance aspect has been considered unassailable (Ng 1979, p. 115). However, it is not quite compelling either because the 'irrelevant' alternatives are not, strictly speaking, irrelevant. IIA does not distinguish between alternatives that are not even included in the set of available alternatives and those that

[8] See Sen (1970, p. 89) Mackay (1980, p. 79), and Kemp & Ng (1987).

[9] If IIA is formulated in such a way that it refers to cardinal-utility profiles, we end up with an impossibility result because cardinal utility without interpersonal comparisons does not make the impossibility result vanish (Sen 1970, Kalai & Schmeidler 1977). Accordingly, the standard view is that the most reasonable way to eschew Arrovian impossibility is to make interpersonal comparisons.

belong to that set but do not seem to be under explicit consideration at a given stage of voting. The truly irrelevant alternatives belong to the former set (cf. Hansson 1973, Bordes & Tideman 1991). Fortunately, voters' choices do take into account preferences for all the alternatives they consider. Even preferences for alternatives outside the pair for which they are voting at a given stage affect their decisions if they maximise expected utility, as in Enelow's (1981) model.

Saari's claims concerning IIA are perfectly justifiable. What is not so clear is whether these arguments can be used for defending the BC against the majority rule used in agenda voting, or any other voting rule for that matter. As Thomas Risse (2001, 2005) points out, arguments concerning IIA do not settle his dispute with Saari (2003*a*, 2006) concerning the BC and the Kemeny rule (see also Saari & Merlin 2000) because both violate IIA. Given, however, that Saari's target of criticism, at least previously (1995), was agenda voting and the majority rule, what I have to say about agendas may be relevant to this debate as well.

Saari's IIA criticisms cannot be viewed as arguments for the BC because IIA *also* is violated in agenda voting, and the majority rule in agenda voting takes cardinal as well as ordinal preference intensities into account. In order to illustrate this, let us consider Enelow's (1981) model of strategic voting under amendment agendas.

Let $X = \{x, y, z\}$ denote a set of available alternatives, U^i voter i's utility function, and $\succ_i$ voter i's preference relation defined on X. Table 6.1 shows the possible preference orderings and a normalisation convention for voters' utilities: v^i denotes voter i's *intensity* of preference.

type of voter						
t_1	t_2	t_3	t_4	t_5	t_6	U^i
x	y	z	x	y	z	$U^i(\cdot) = 1$
y	z	x	z	x	y	$U^i(\cdot) = v^i$
z	x	y	y	z	x	$U^i(\cdot) = 0$

Table 6.1: Voter types and utility normalisations with three alternatives

Alternatives are put to a sequence of pair-wise majority comparisons in *an amendment agenda*.[10] Two alternatives are put to a majority vote against each other in the first round of voting. The winner of this first contest is then put to vote against the third alternative in the second round. Figure 6.1 presents the three possible amendment agendas.

Voter i's subjective probability that a given alternative j beats another alternative k ($j, k \in X$) in a pair-wise second-round contest is denoted p^i_{jk}. In agenda A, and in the first round of voting, voters choose a branch in the voting tree by comparing expected utilities for lotteries $(x, z; p^i_{xz}, 1 - p^i_{xz})$ and $(y, z; p^i_{yz}, 1 - p^i_{yz})$. Note that merely formulating the voters' choice situation under incomplete information shows that they are making a choice not between the pair $\{x, y\}$, but rather between two lotteries that *also involve the third alternative z*. It follows immediately that their 'choice between x and y' in the first

[10] See Miller (1995) for a discussion on different agendas.

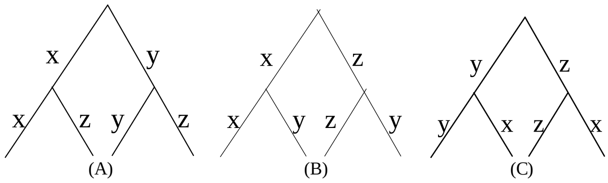

Figure 6.1: Amendment agendas with three alternatives

round *may* provide information concerning their preference intensity between this pair of alternatives. Expected-utility expressions need to be formulated in order to show this.

Maximising expected utility implies giving one's vote to the branch in the voting tree that has the greatest expected utility. A voter will vote for the left-hand branch under agenda (A) if

$$p^i_{xz} \cdot U^i(x) + (1 - p^i_{xz}) \cdot U^i(z) \geq p^i_{yz} \cdot U^i(y) + (1 - p^i_{yz}) \cdot U^i(z). \qquad (6.2)$$

Consider now voter types one and four. Both prefer x to y, but the preferences of type-four voters are ordinally more intensive because they separate the preferences between these alternatives with z by preferring x to z to y, whereas type-one voters prefer x to y to z.[11] Type-four voters have a dominant strategy to vote sincerely for x under agenda A. Applying utility normalisation for a type-one voter to equation 6.2 yields:

$$p^i_{xz} \cdot 1 + (1 - p^i_{xz}) \cdot 0 \geq p^i_{yz} \cdot v^i_1 + (1 - p^i_{yz}) \cdot 0, \qquad (6.3)$$

Type-one voters will thus vote strategically for the right-hand branch (y) if:

$$v^i_1 > \frac{p^i_{xz}}{p^i_{yz}}.$$

When they do, they are effectively expressing a cardinally strong intensity for x and y over z, and a cardinally weak intensity between x and y. Hence, they are able to express their preference intensity between x and y by deciding whether to vote strategically or not. In contrast, type-four voters never vote for y in the first round, and thereby reveal a strong intensity of preference for x over y. Voters thus express their ordinal and cardinal intensities under agenda voting, but they do this only in a probabilistic sense.

One might be willing to argue that since the BC always collects information on ordinal intensity, it should be preferred to majority voting under agendas. However, the normatively important intensity information is cardinal rather than ordinal. What is thus really relevant is the question of which voting rule

[11] If indifference is ruled out by assuming that $0 < v^i < 1$ for all voters, this intensity must also be cardinally stronger.

best reflects cardinal preference intensities on the aggregate level.[12] If the judgment that only aggregate-level cardinal information matters for normative evaluations is defensible, it is normatively irrelevant whether it is possible to obtain information on ordinal intensity from each person separately or not. From this perspective, trying to replace IIA with a condition like Saari's *binary intensity IIA,* or any condition that fails to take voters' behaviour explicitly into account, is not fully satisfactory as a criterion for the choice between various voting rules.

One possible reason why individual-level ordinal intensity information should be collected is that it is intrinsically related to rationality. It is simultaneously information concerning the transitivity of preferences. Saari's rationality argument against IIA could be formulated as follows. Since IIA restricts the relevant information to preferences for pairs of alternatives, a voting rule that satisfies it does not allow for taking into account connecting information between the different pairs. Thus, even though one condition for Arrow's theorem explicitly requires voters to have transitive preferences, this transitivity is trumped by IIA.[13]

Again, Saari's argument is entirely convincing as a criticism of IIA: if there were a voting rule that satisfied IIA, such a rule would lose the information on voters' rationality. There are no such rules, however, because voters *are rational* when they engage in strategic voting, and they will violate IIA under all voting rules if they maximise expected utility.[14] To put it differently, if voters are rational, it is not possible to distinguish between different voting rules by applying the rationality argument against IIA. Indeed, I presume that Saari agrees with me here because his point has been that voting rules that satisfy IIA are incapable of distinguishing the cyclic preferences of (non-existent) irrational individuals from cyclic preference profiles. If there is a rationality-related argument that could distinguish between different voting rules, it must concern the recognition of irrational voters from rational ones when not all voters are rational.

The important case is thus one in which at least some voters are irrational. It would be best to ignore their votes, but as this will probably be impossible under all voting rules, the task becomes one of trying to determine how these voters will affect the voting outcomes. One might argue that the BC forces voters to be rational because it requires them to provide a full ranking of the alternatives. It might have such edifying aspects, but it is just as plausible that irrational voters will provide truncated ballots. Nevertheless, there may be perfectly rational voters, rational in the sense of having transitive preferences, who also provide truncated ballots. If, for example, a voter prefers x strongly to y and z, and is not willing to distinguish between the latter two, he or she may just provide a Borda score for x and ignore the rest. If and when it is not

[12] Lehtinen (forthcoming a, b) provides results that are relevant to a comparison between the BC and agendas. I will not summarise those results here because my main goal is not to compare particular voting rules.

[13] In addition to previous references to Saari, see Saari and Sieberg (2001, 2004).

[14] Saari agreed in a private conversation (April 2006) that this characterisation of the relationship between IIA, rationality and strategic voting was apt.

possible to distinguish rational from irrational voters in practice, I conclude that the rationality argument is valid against IIA but cannot be used as an argument for BC against any other voting rule.

6.4 IIA is violated in amendment agendas

The Condorcet winner (the alternative that is preferred by a majority to all other alternatives) is always selected under the majority rule if the voters vote sincerely (Farquharson 1969) and if they vote strategically under complete information (McKelvey & Niemi 1978, Moulin 1979). It would thus seem that the Condorcet winner is observed since it will be the outcome under the majority rule. I will now use a simple example (from Lehtinen 2007) to show that this result does not hold under incomplete information (see also Ordeshook & Palfrey 1988). It follows that IIA is violated even under the majority rule and an amendment agenda when voters maximise their expected utility.

In fact, the idea that IIA is incompatible with expected utility maximisation has already been acknowledged by all those who have argued that it does not allow voters to express their intensities of preference or cardinal utilities (see in particular Hammond 1991*a*). Assume that the preferences of three voters A, B, and C, can be described as in the following table:

A	B	C
y (1)	y (1)	x (1)
x (0.9)	x (0.9)	z (0.9)
z (0)	z (0)	y (0)

Table 6.2: Example 1

The numbers in parentheses denote voters' utilities. The sum of utility for the utilitarian winner x is 0.9+0.9+1=2.8, and for the Condorcet winner y it is 1+1+0=2. Thus, x should be selected according to the utilitarian criterion instead of the Condorcet winner y. If all voters vote sincerely, the Condorcet winner y will beat the utilitarian winner x in the first round and z in the second round, and emerges as the final outcome.

Let us now see what would happen if the voters maximise expected utility under incomplete information. Assume that all three voters have identical beliefs such that $p_{yz} = 0.7$, and $p_{xz} = 0.9$. Voters A and B are of type five. They will vote strategically for x in the first round because $v_5 < \frac{p_{yz}}{p_{xz}}$ $(0.9 < \frac{0.7}{0.9} = 0.7778)$. Voter C has a weakly dominant strategy to vote for x in the first round of voting. Thus, x is the outcome if the voters maximise expected utility because it beats y in the first round and z in the second round. The utilitarian winner x is chosen if they maximise expected utility but the Condorcet winner y is chosen if they vote sincerely. A Condorcet winner is thus not necessarily chosen under the majority rule.[15]

[15] If satisfying IIA is desirable, selecting a Condorcet winner is desirable because 'IIA implies

Let us now proceed to show that IIA is violated in this example. Whether it is defined in terms of pairs of alternatives or in terms of a set of relevant alternatives[16] is not particularly important here, as long as it is formulated in choice-functional terms and the social choice function refers to *actually chosen* alternatives. Let us thus assume that $C(S)$ denotes a *choice* made by society in voting from a set of alternatives $S \subset X$. Arrow's (1963) treatment assumes that all voters vote sincerely so that each one chooses the alternative that he or she prefers the most. Let $C_i(S)$ denote individual i's choice from a set of alternatives S, and $\succ_i$his or her preference ordering, and let $n(j \succ k)$ denote the number of voters who prefer alternative j to k. Arrow requires that the individual choices fulfil equation (6.4):[17]

$$C_i(S) = \{x | x \in S : \forall y \in S : x \succ_i y\}. \tag{6.4}$$

The method of majority decision is defined by

$$\forall x, y \in S : C(S) = x \leftrightarrow \forall y \in S : n(x \succ y) > n(x \prec y). \tag{6.5}$$

A Condorcet winner (CW) is defined by

$$CW = \{x | x \in S : \forall y \in S : n(x \succ y) > n(x \prec y)\}. \tag{6.6}$$

A Condorcet winner is always chosen in the method of majority decision because (6.4) guarantees that the social choice according to (6.5) always selects it.

Consider now example 1. It is easy to see that IIA is violated: y was chosen when all voters voted sincerely, but x was chosen if some voted strategically. Two different outcomes emerged from a single preference profile from the two different behavioural assumptions. It is also possible, of course, to obtain two different outcomes from a single profile and only the behavioural assumption of expected utility maximisation if the voters' beliefs are different in two different cases.

The example is of importance for three reasons. First, it shows that attitudes towards risk and preference intensities will *inevitably* affect voting choices if voters maximise expected utility under incomplete information. Secondly, IIA is violated even under the majority rule if voters maximise expected utility and if there are at least three alternatives. It is well known that many commonly used voting rules (plurality, runoff, Borda, etc.) may fail to select a Condorcet winner. This means that insofar as Arrow's theorem is considered a theorem about voting rules, the IIA condition *is violated under all democratic voting rules* that consider at least three alternatives (cf. Hansson 1973).

Finally, ordinal utility is not observable either in the sense that the selected alternative need not be the Condorcet winner under the majority rule and amendment agendas. The sum-of-utility criterion has been criticised for

---

the Condorcet criterion' (Arrow 1997, p. 5). An argument against the plausibility of IIA is thus simultaneously an argument against the normative appeal of Condorcet winners.

[16] See Ray (1973), Sen (1986), and Bordes & Tideman (1991).

[17] The assumption that all preferences are strict is used here.

not being observable (e.g., Arrow 1973*b*). Preference orderings would be observable if the Condorcet winners were always selected under the majority rule, but this is not the case. The possibility of strategic voting thus undermines the observability argument. Therefore, preference orderings are not observable either, and observability is not a valid argument for ordinal utility and against intensities in a voting context. The claim that preference orderings are scientifically respectable because they can be observed is an invalid argument against intensities *in voting theory* even though this argument may have some weight in other contexts.

It would, of course, be easier to collect information on preference orderings than on intensities by other means than voting. We could, for example, simply ask the voters about their preference orderings. The problem with any procedure other than voting itself, however, is that insofar as the results are used for making decisions, the individuals have an incentive to misrepresent their preferences. If, on the other hand, the results are not used for making decisions, voters, particularly representatives in parliaments, have an incentive to misrepresent their preferences in order to give signals to their constituencies. Collecting information on preference orderings is thus easier than collecting information on preference intensities, but it is ultimately not possible to obtain fully reliable information on either of them.

It is not possible to prove the general claim that intensities will affect the results under all voting rules here. However, it is clear that insofar as an expected-utility model can be formulated for any voting rule, it can be shown that preference intensities will affect the outcomes under this rule. What follows from this is that if the epistemological-moral argument is to be effective against using intensities in voting theory, one has to deny that voting is characterised by decision-making under uncertainty. Surely, however, nobody is willing to argue that voters have complete information on other voters' preferences in an electorate of dozens, thousands or millions. Real-world voting is clearly characterised by decision-making under uncertainty, as Coleman (1966) argued long ago.

6.5 The epistemological-moral argument

Since I am arguing for a utilitarian evaluation of outcomes in voting theory, it would seem natural to take Harsanyi's (1953, 1955, 1977) theorems as a decision-theoretic justification for a utilitarian position. Harsanyi claims that the theorems show that von Neumann-Morgenstern (vNM) utilities represent preference intensities, and that they can be used to provide an argument for utilitarianism. I do not draw on these theorems because I fully accept the criticism that Harsanyi's utilitarianism is 'utilitarianism in name only'[18]: the theorems do not really provide an argument for utilitarianism.

[18] See Sen (1976, 1977, 1986), Weymark (1991), Roemer (1996, pp. 138-150), and Mongin (2001).

Arrow and Rawls first presented what I call the epistemological-moral argument as a criticism of Harsanyi's position. According to the argument, vNM utilities should not be used in social-welfare judgements because they inevitably contain morally irrelevant information on attitudes towards risk. The *moral* part is that attitudes towards risk are irrelevant to social-welfare judgments and they should therefore not be taken into account, and the *epistemological* part is that von Neumann-Morgenstern utility functions can only be constructed from choices involving risk.[19] Hence, attitudes towards risk inevitably affect social-welfare judgements if these judgements are based on von Neumann-Morgenstern utilities.

Harsanyi has persistently argued that vNM utility functions may be used for social-welfare judgements: they functions express a willingness to take risks in order to obtain some particular alternative (Harsanyi 1987). Hence, they express the relative intensity with which a person prefers one alternative to another (see also Harsanyi 1978, 1979, and Ng 1999).

Harsanyi (1992, pp. 682-684) argues that Arrow and Rawls confuse 'process utility' and 'outcome utility' (see also Harsanyi 1993). Process utility, or 'utility from gambling', refers to the enjoyment from playing a game that involves risk, while outcome utility relates to the prizes one may obtain. Harsanyi is right in that the reduction of the compound-lotteries axiom precludes process utilities and thereby 'utility from gambling'. The vNM theory thus rules out attitudes towards enjoyment from gambling by assumption. Harsanyi is also right in pointing out that outcome utilities are ethically important. His arguments could be taken to account for why we think preference intensities are morally relevant. The problem with his argument about process utility and outcome utility is that it does not really provide a response to the criticism: attitudes towards process utilities are not what a carefully stated epistemological-moral argument should be all about. Arrow (1983 [1973]*c*, p. 107), for example, argues that vNM utilities incorporate attitudes towards risk. The epistemological-moral argument also concerns attitudes towards risk that are related to voters' willingness to engage in strategic behaviour, not just attitudes towards enjoyment from gambling, and these attitudes are *also* irrelevant to social-welfare judgements.[20]

Arrow and Rawls' position is buttressed by a well-known decision-theoretical epistemological consideration: standard expected utility theory does not provide any way of distinguishing between the psychological sensations of diminishing marginal utility (or diminishing intensity of satisfaction) and risk aversion, if all we are given are a person's *choices* under uncertainty. Indeed, Harsanyi (1992, p. 685) admits this. Choices under risk do *reflect* preference intensities, just as he claims, but this argument does not change the fact that attitudes towards risk *also* affect these choices. Hence, while vNM utilities incorporate ethically relevant information concerning preference intensities, they also incorporate ethically irrelevant information concerning attitudes towards risk.

[19] See e.g., Alchian (1953), Baumol (1958), or Fishburn (1989).

[20] Here I am disregarding the entirely different question of whether the riskiness of the *choice alternatives* in an election should be taken into account.

Harsanyi has successfully shown that choices under uncertainty reflect preference intensities, and that these intensities are morally relevant. Nevertheless, the epistemological-moral argument remains valid because vNM utilities inevitably reflect morally irrelevant attitudes towards risk. However, this argument can be used against using preference intensities in social choice theory only if it is possible to collect reliable information on ordinal utilities that *do not* reflect attitudes towards risk. As shown in the previous section, this is not possible.

6.6 Conclusions

Social choice theorists have not been willing to abandon IIA mainly because it is closely related to excluding strategic voting. However, strategic voting is desirable rather than undesirable under most commonly used voting rules. The basic reason for this is that it reflects preference intensities, and sincere voting does not allow for this under most voting rules. This is why strategic voting should be taken as an argument against IIA rather than for it. By the same token, the strategy-proofness condition is also normatively questionable.

Arrow's impossibility result and the closely related theorems given by Gibbard (1973) and Satterthwaite (1975) are unassailable as deductive proofs. However, we should not be concerned about these results because their most crucial conditions are not justifiable. Fortunately, IIA and strategy-proofness are violated under all democratic voting rules, including the majority rule in agenda voting. Given that Arrow's theorem crucially depends on IIA, its importance is called into question. Saari (1995, p. 196) once formulated the meaning of the Arrow theorem as follows: it asserts that the ignored information is vital and that it is impossible to construct a procedure that systematically discards information on preference intensities. But why should anyone want such procedures in the first place?

None of the arguments discussed in this paper, the epistemological-moral, the observability, and the strategic-voting argument, is successful *as* an argument for using only preference orderings and Condorcet winners in voting theory. There seems to be no good reason for evaluating voting outcomes on the basis of Condorcet winners rather than utilitarian winners. Utilitarian winners are to be preferred on genuine ethical grounds, however, because they take preference intensities into account.

The above arguments thus give rise to three methodological conclusions. First, given that the three main arguments for IIA and against intensities fail, there is no reason to favour Condorcet winners over utilitarian winners in welfarist evaluations of voting rules. Secondly, the notion of cyclic preferences and the absence of a Condorcet winner have been given an all too prevalent role in voting theory. The possibility that the preferences are cyclic is only one among many factors that may influence voting outcomes. Beliefs, information and preference intensities are also important. Models that take into account only preference orderings provide a misleading picture of voting rules because such models are based on the false empirical assumption that voting is characterised by choice under certainty.

The third methodological conclusion is that the theory of strategic voting has not addressed the right questions. If strategic voting is beneficial under many commonly used voting rules, it does not seem very fruitful to seek strategy-proof voting mechanisms or to find out which voting rules are least susceptible to strategic voting. The relevant question concerns how much strategic voting increases (or perhaps decreases) utilitarian efficiency under various voting rules under different assumptions on the voters' willingness to take risks, preference intensities and interpersonal comparisons. There are significant differences between different voting rules in these respects.

:

Appendix A

Deriving the beliefs

The belief of a player who has obtained a signal s_i^ε will now be derived. In what follows the subscripts denoting the individuals (i), the candidates (j and k), and the simulated game (g) are omitted in order to make it easier to read the formulas. The signal (equation 2.9) may be written as

$$S = Q + R. \tag{A.1}$$

R is a random variable with mean zero and variance ε^2. Since variable Q is standard normal, the signal S can be viewed as a sum of two normally distributed random variables $Q \sim N(0,1)$ and $R \sim N(0,\varepsilon^2)$.

Let σ_Q $(= 1)$ and σ_R $(= \varepsilon)$ denote the standard deviations of Q and R, respectively. We will now derive a conditional distribution for the variable Q, $F(Q < 0|S = s_i^\varepsilon)$.

The density of Q is

$$f_Q(q) = \frac{1}{\sqrt{2\pi}\sigma_Q} \exp\left[-\frac{1}{2}\left(\frac{q}{\sigma_Q}\right)^2\right], \tag{A.2}$$

and the density of R is

$$f_R(r) = \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma_R}\right)^2\right]. \tag{A.3}$$

Now $q + r = s$ so that $r = s - q$. Let us now use

$$\begin{aligned} q &= x, \textit{ and} \\ r &= s - x. \end{aligned} \tag{A.4}$$

Since Q and R are two independent random variables, their joint density is given by the product of their densities (e.g., Casella & Berger 1990, p. 210)

$$\begin{aligned} f_{Q,R}(q,r) &= f_Q(q) f_R(r) \\ &= \frac{1}{2\pi}\frac{1}{\sigma_Q\sigma_R} \exp\left[-\frac{1}{2}\left(\frac{q}{\sigma_Q}\right)^2 - \frac{1}{2}\left(\frac{r}{\sigma_R}\right)^2\right]. \end{aligned} \tag{A.5}$$

Using (A.4) we get

$$= \frac{1}{2\pi}\frac{1}{\sigma_Q\sigma_R}\exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_Q}\right)^2 - \frac{1}{2}\left(\frac{s-x}{\sigma_R}\right)^2\right].$$

Let $D = \left[-\frac{1}{2}\left(\frac{x}{\sigma_Q}\right)^2 - \frac{1}{2}\left(\frac{s-x}{\sigma_R}\right)^2\right]$. This can be written as follows:

$$D = -\frac{1}{2}\left[\left(\frac{1}{\sigma_R^2} + \frac{1}{\sigma_Q^2}\right)x^2 - \frac{2sx}{\sigma_R^2} + \frac{s^2}{\sigma_R^2}\right].$$

Completing the square we have

$$D = -\frac{1}{2}\left[\left(\frac{1}{\sigma_R^2} + \frac{1}{\sigma_Q^2}\right)\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2 + \frac{s^2}{\sigma_R^2} - \frac{s^2}{\sigma_R^4}\right]. \tag{A.6}$$

Consider now the random variable $X = (Q|S = s)$. From $(A.6)$ we see that $X \sim N(\frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s, \sigma_X^2)$, where $\sigma_X^2 = \frac{1}{\left(\frac{1}{\sigma_R^2}+\frac{1}{\sigma_Q^2}\right)}$. The density function of X is of the following form: $f_X(x) =$

$$A\frac{1}{2\pi}\frac{1}{\sigma_Q\sigma_R}\exp\left[-\frac{1}{2}\left(\frac{s^2}{\sigma_R^2} - \frac{s^2}{\sigma_R^4}\right)\right]\exp\left[-\frac{1}{2\sigma_X^2}\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2\right]. \tag{A.7}$$

Now since $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$, we have

$A\frac{1}{2\pi}\frac{1}{\sigma_Q\sigma_R}\exp\left[-\frac{1}{2}\left(\frac{s^2}{\sigma_R^2} - \frac{s^2}{\sigma_R^4}\right)\right]\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma_X^2}\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2\right]dx = 1.$

and since

$$\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma_X^2}\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2\right]dx = \sqrt{2\pi}\sigma_X,$$

it is easy to see that

$$A = \frac{1}{\frac{1}{2\pi}\frac{1}{\sigma_Q\sigma_R}\exp\left[-\frac{1}{2}\left(\frac{s^2}{\sigma_R^2} - \frac{s^2}{\sigma_R^4}\right)\right]\sqrt{2\pi}\sigma_X},$$

so that

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{1}{2\sigma_X^2}\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2\right]. \tag{A.8}$$

The probability $p(Q < 0|S = s)$ is given by the cumulative distribution function of X:

$$F_X(x) = \int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}\sigma_X}\exp\left[-\frac{1}{2\sigma_X^2}\left(x - \frac{\sigma_Q^2}{\sigma_Q^2+\sigma_R^2}s\right)^2\right]dx. \tag{A.9}$$

Since $\sigma_X^2 = \frac{1}{\left(\frac{1}{\sigma_R^2}+\frac{1}{\sigma_Q^2}\right)} = \frac{1}{\left(\frac{\sigma_Q^2+\sigma_R^2}{\sigma_R^2\sigma_Q^2}\right)} = \frac{\sigma_Q^2\sigma_R^2}{\sigma_Q^2+\sigma_R^2}$, $\frac{1}{\sigma_X^2} = \frac{\sigma_Q^2+\sigma_R^2}{\sigma_Q^2\sigma_R^2}$, and $\frac{1}{\sigma_X} = \sqrt{\frac{\sigma_Q^2+\sigma_R^2}{\sigma_Q^2\sigma_R^2}}$, we can write equation (A.9) as follows: $F_X(x) =$

$$\frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_Q^2+\sigma_R^2}{\sigma_Q^2\sigma_R^2}}\int_{-\infty}^{0}\exp\left[-\frac{1}{2}\left(x\frac{\sqrt{\sigma_Q^2+\sigma_R^2}}{\sigma_Q\sigma_R}-\frac{1}{\sigma_Q\sigma_R}\frac{\sigma_Q^2}{\sqrt{\sigma_Q^2+\sigma_R^2}}s\right)^2\right]dx. \tag{A.10}$$

We will need to make two changes of variables in order to derive a functional form that can be used in computer simulations. Let $u = \sqrt{\sigma_Q^2+\sigma_R^2}x$, so that $dx = \frac{du}{\sqrt{\sigma_Q^2+\sigma_R^2}}$. When $x = -\infty$, $u = -\infty$, and when $x = 0, u = 0$. We have thus $F_X(x) =$

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\sigma_Q^2+\sigma_R^2}{\sigma_Q^2\sigma_R^2}}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}\times \\ & \int_{-\infty}^{0}\exp\left[-\frac{1}{2}\left(\frac{u}{\sigma_Q\sigma_R}-\frac{1}{\sigma_Q\sigma_R}\frac{\sigma_Q^2}{\sqrt{\sigma_Q^2+\sigma_R^2}}s\right)^2\right]du \\ = & \frac{1}{\sqrt{2\pi}\sigma_Q\sigma_R}\int_{-\infty}^{0}\exp\left[-\frac{1}{2}\left(\frac{u}{\sigma_Q\sigma_R}-\frac{1}{\sigma_Q\sigma_R}\frac{\sigma_Q^2}{\sqrt{\sigma_Q^2+\sigma_R^2}}s\right)^2\right]du. \end{aligned}$$

Now let $w = \frac{u}{\sigma_Q\sigma_R} - \frac{\sigma_Q}{\sigma_R}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}s$, so that $du = \sigma_Q\sigma_R dw$. When $u = -\infty$, $w = -\infty$, and when $u = 0, w = -\frac{\sigma_Q}{\sigma_R}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}s$, so that

$$F_X(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\sigma_Q}{\sigma_R}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}s}\exp\left[-\frac{1}{2}w^2\right]dw. \tag{A.11}$$

This is the cumulative distribution function of a standard normal random variable so that

$$F(Q<0|S=s) = \Phi\left(-\frac{\sigma_Q}{\sigma_R}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}s\right). \tag{A.12}$$

Now since $\sigma_Q = 1$, and $\sigma_R = \varepsilon$, $\frac{\sigma_Q}{\sigma_R}\frac{1}{\sqrt{\sigma_Q^2+\sigma_R^2}}s = \frac{s}{\varepsilon\sqrt{1+\varepsilon^2}}$. Player i's beliefs are given by

$$p_i(j,k) = 1 - \Phi\left(-\frac{1}{\varepsilon\sqrt{1+\varepsilon^2}}s_i^{\varepsilon}\right). \tag{A.13}$$

Appendix B

Deriving the variance of $\mathbf{S}^{jk}$

Let us start by deriving the variance of B^{jk}. It will then be easier to calculate the variance for an individual voter's Borda score. Therefore, let us define the difference in each individual's Borda score for candidates j and k, B_i^{jk}, as a function of the random variables $N_{1,i}^j, N_{2,i}^j, N_{1,i}^k$, and $N_{2,i}^k$:

$$B_i^{jk} = 2N_{1,i}^j + N_{2,i}^j - 2N_{1,i}^k - N_{2,i}^k. \tag{B.1}$$

According to the definition of variance,

$$Var\left(B_i^{jk}\right) = E[2N_{1,i}^j + N_{2,i}^j - 2N_{1,i}^k - N_{2,i}^k - E(2N_{1,i}^j + N_{2,i}^j - 2N_{1,i}^k - n_{2,i}^k)]^2.$$

The subscript i is dropped from the following expressions in order to avoid clutter. Since $E[2N_1^j + N_2^j - 2N_1^k - N_2^k] = 0$, $Var\left(B_i^{jk}\right) = E[2N_1^j + N_2^j - 2N_1^k - N_2^k]^2$. Expanding this expression gives $Var\left(B_i^{jk}\right) = E[4N_1^j N_2^j - 8N_1^j N_1^k - 4N_1^j N_2^k - 4N_1^k N_2^j + 4N_1^k N_2^k - 2N_2^j N_2^k + 4\left(N_1^j\right)^2 + 4\left(N_1^k\right)^2 + \left(N_2^j\right)^2 + \left(N_2^k\right)^2]$.

Since N_1^j and N_2^j mutually exclude each other (no voter can rank candidate j both first and second), $E\left[N_1^j N_2^j\right] = 0$. Similarly $E\left[N_1^j N_1^k\right] = 0$ (no voter can rank both j and k first). Similar reasoning shows that $E\left[N_1^k N_2^k\right] = 0$, $E\left[N_2^j N_2^k\right] = 0$. Hence

$$Var\left(B_i^{jk}\right) = E[-4N_1^j N_2^k - 4N_1^k N_2^j + 4\left(N_1^j\right)^2 + 4\left(N_1^k\right)^2 + \left(N_2^j\right)^2 + \left(N_2^k\right)^2].$$

The expected value of $N_1^j N_2^k$ is $\frac{1}{6}$, since this is the probability that a voter ranks j first and k second. Similar reasoning applies to $N_1^k N_2^j$. By definition, $E\left[\left(n_1^j\right)^2\right] = Var\left[n_1^j\right] + [E\left(n_1^j\right)]^2$. Since $E[n_1^j] = \frac{1}{3}$, $\left[E\left(n_1^j\right)\right]^2 = \left(\frac{1}{3}\right)^2 = \frac{1}{9}$, and since $Var\left[n_1^j\right] = \frac{1}{3}\frac{2}{3} = \frac{2}{9}$, $E\left[\left(n_1^j\right)^2\right] = \frac{1}{3}$. Similarly, $E\left[\left(n_1^k\right)^2\right] = \frac{1}{3}$, $E\left[\left(n_2^j\right)^2\right] = \frac{1}{3}$, and $E\left[\left(n_2^k\right)^2\right] = \frac{1}{3}$.

The variance of B_i^{jk} is thus

$$Var\left(B_i^{jk}\right) = E[-4\frac{1}{6} - 4\frac{1}{6} + [4\frac{1}{3} + 4\frac{1}{3} + \frac{1}{3} + \frac{1}{3}] = 2. \tag{B.2}$$

Since each individual type is independent of the others, the variance of B^{jk} is the sum of the variances of $B_i^{jk}, B_{i+1}^{jk}, ..., B_N^{jk}$. The variance of B^{jk} is thus $2N$.

The variance of R^{jk}, σ_R^2, is

$$\sigma_R^2 = Var\left(\varepsilon R^j - \varepsilon R^k\right) = \varepsilon^2 Var\left(R^j - R^k\right) =$$
$$\varepsilon^2 Var\left(R^j + R^k\right) = \varepsilon^2(1+1) = 2\varepsilon^2.$$ The variance of S^{jk} is thus

$$\sigma_S^2 = 2N + 2\varepsilon^2. \tag{B.3}$$

Appendix C

Excerpt: De welzijnsgevolgen van strategisch stemmen

Sociale keuze theorie werd gelanceerd toen Kenneth Arrow liet zien dat niet gelijktijdig kan worden voldaan aan een stel ogenschijnlijk redelijke voorwaarden voor een maatschappelijke welvaartsfunctie. Arrow's theorema wordt algemeen zo geïnterpreteerd dat er uit blijkt dat er iets mis is met alle stemregels. Het is gebaseerd op twee filosofische ideeën. Het eerste is dat interpersoonlijke nutsvergelijkingen nietszeggend zijn, en het tweede is dat uit individuele keuzen alleen voorkeurordeningen maar niet voorkeurintensiteiten kunnen worden afgeleid. Deze twee ideeën vormden een aanleiding voor de cruciale Onafhankelijkheid van Irrelevante Alternatieven (OIA) conditie. Die stelt dat de sociale keus alleen afhankelijk zou moeten zijn van individuele voorkeuren voor paren alternatieven. Sociale keuze theorie en de OIA-conditie zijn altijd bekritiseerd voor het niet behoorlijk in aanmerking nemen van voorkeurintensiteiten. Echter, Allan Gibbard en Mark Satterthwaite lieten in de jaren zeventig zien dat Arrow's voorwaarden logisch equivalent zijn met een stel voorwaarden, waaronder de zogeheten strategiebestendigheidsvoorwaarde. Een stemregel is strategiebestendig indien geen enkel individu een drijfveer heeft om zijn of haar voorkeuren verkeerd voor te stellen door strategisch te stemmen, dat wil zeggen door een stem te geven aan een kandidaat die niet het hoogst staat in iemands voorkeurordening. Aangezien strategisch stemmen algemeen wordt gezien als een onwenselijk fenomeen en strategiebestendigheid nauw verbonden is met OIA onderschrijven de meeste stemtheoretici nog altijd de normatieve geldigheid van OIA en daardoor Arrow's theorema.

Echter, de welzijnsgevolgen van strategisch stemmen zijn niet eerder nadrukkelijk bestudeerd binnen een formeel kader dat de verzorgingsstaat voorstaat. Dit Ph D proefschrift biedt een formeel kader waarbinnen de gevolgen van strategisch stemmen gedetailleerd geëvalueerd kunnen worden. Utilitaire doelmatigheid wordt gedefinieerd als het percentage simulatie runs waarin de kandidaat voor wie het totale nut het hoogste is, geselecteerd wordt. Gezegd wordt dat strategisch stemmen welvaartverhogend is als de utilitaire doelmatigheid hoger is bij strategisch gedrag dan onder oprecht gedrag. De belangrijkste uitkomst is dat strategisch stemmen welvaartverhogend is in alle stemregels die bestudeerd zijn onder een scala van redelijke onderstellingen omtrent de voorkeuren en opvattingen van stemmers. De reden is dat indien stemmers zich bezighouden met strategisch gedrag, zij ook hun voorkeurintensiteiten kunnen uiten maar onder oprecht gedrag hebben alleen voorkeurordeningen invloed op de uitkomst. De in dit proefschrift aangedragen modellen van strategisch gedrag laten zien hoe voorkeurintensiteiten invloed hebben op individuele keuzen door het formuleren van verwacht-nut condities.

Het mechanisme dat de verbinding levert tussen individueel gedrag en uitkomsten op macroniveau en dat er voor verantwoordelijk is dat strategisch stemmen welvaartverhogend is, wordt het tegenwicht vormen tegen strategische stemmen genoemd. In een electoraat met een groot aantal stemmers die hun beslissingen nemen op basis van onvolledige informatie hebben sommige individuen typisch een drijfveer om strategisch te stemmen voor een alternatief, zeg x, maar tegelijkertijd hebben anderen een drijfveer om precies hetzelfde alternatief x strategisch te verlaten door voor een ander alternatief y te stemmen. De voorwaarden voor strategisch stemmen op het punt van de structuur van de voorkeuren van stemmers en hun informatie verschillen in verschillende stemregels, maar zij wijzen er altijd op dat, ceteris paribus, een alternatief dat een hoog totaal nut heeft, waarschijnlijk veel strategische stemmen zal winnen en weinig zal verliezen. Dus is het waarschijnlijk dat weinig strategische stemmen tegen de utilitaire winnaar een tegenwicht vormen tegen veel stemmen vóór. Het tegenwicht vormen verklaart aldus hoe strategisch gedrag het individuele voorkeurintensiteiten mogelijk maakt de macroniveau uitkomsten op een voordelige manier te beïnvloeden.

Het resultaat, dat strategisch stemmen welvaartverhogend is, heeft twee fundamentele methodologische implicaties. Ten eerste betekent het dat Arrow's theorema en het Gibbard-Satterthwaite theorema niet echt betekenen dat alle stemregels gebrekkig zijn. Ten tweede dat, hoewel stemtheoretici voorheen geprobeerd hebben manieren uit te denken om strategisch stemmen te verhinderen, de belangrijke vraag eerder is, te zien welke stemregels het meest profiteren van strategisch stemmen.

Het proefschrift bestaat uit vier wetenschappelijke publikaties en een afsluitend essay. Het eerste essay biedt een model van onvolledige informatie dat gebaseerd is op statistische signaalextractie, en dat geschikt is voor computersimulatieonderzoek naar stemmen. Het epistemologisch basisidee hier is, dat het voor een onderzoeker niet noodzakelijk is exacte informatie te verkrijgen over de individuele opvattingen indien het individuele gedrag en daardoor de uitkomsten op macroniveau kunnen worden gekarakteriseerd met redelijke onderstellingen omtrent de gemiddelde kwaliteit van de informatie van de onderzoekers en hun gemiddelde mate van vertrouwen in hun informatiebronnen. Het belangrijkste voordeel van het gebruik van dit model vloeit voort uit het feit dat het mogelijk maakt een stel heterogene opvattingen te genereren die op een redelijke manier afhangen van de betrouwbaarheid van de signalen van de onderzoekers en hun vertrouwen daar in.

De essays twee en drie bestuderen de welzijnsgevolgen van strategisch stemmen in drie verschillende stemregels. Het kader wordt in zijn geheel in het tweede essay opgevoerd en toegepast in het derde. Het tweede essay biedt tevens een antwoord op kritiek dat het kader gebaseerd is op het maken van aanvechtbare interpersoonlijke nutsvergelijkingen. Getoond wordt dat strategisch stemmen welvaartverhogend is met alle mogelijke en ten minste minimaal redelijke interpersoonlijke vergelijkingen. Het laat zodoende zien dat het resultaat, dat strategisch stemmen welvaartverhogend is, krachtig is wat betreft verschillende interpersoonlijke vergelijkingen.

De veronderstelling van krachtig wordt in het vierde essay uitgewerkt. Het biedt een filosofisch kader voor het evalueren van onrealistische onderstellingen in rationele-keuzemodellen. De hoofdlijn is dat indien een model aantoonbaar krachtig is wat betreft zijn gedragsonderstellingen, hun onwaarheid het model niet op een ernstige manier ongeldig maakt.

Het afsluitend essay behandelt de diverse argumenten die zijn opgevoerd ten gunste van de OIA-conditie in het licht van de resultaten omtrent de welzijnsgevolgen van strategisch stemmen. Betoogd wordt dat de OIA-conditie in stemcontexten altijd wordt overtreden en dat dit niet beschouwd dient te worden als een negatief resultaat: er is geen dwingende reden om te voldoen aan OIA. Ook wordt betoogd dat Donald Saari's kritiek op OIA, hoewel volledig dwingend als kritiek op OIA, geen enkele stemregel in het bijzonder, zoals de Bordaregel, steunt.

Bibliography

Akerlof, George A. 1970. "The market for 'lemons': Quality uncertainty and the market mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Akerlof, George A. & Janet L. Yellen. 1985. "Can small deviations from rationality make significant differences to economic equilibria?" *American Economic Review* 75(4): 708–720.

Alchian, Armen A. 1953. "The meaning of utility measurement." *The American Economic Review* 43(1): 26–50.

Alvarez, R. Michael, Frederick J. Boehmke & Jonathan Nagler. 2006. "Strategic voting in british elections." *Electoral Studies* 25(1): 1–19.

Arrow, Kenneth J. 1951. *Social choice and individual values.* 1 ed. New York, London: Wiley, Chapman and Hall.

Arrow, Kenneth J. 1963. *Social choice and individual values.* 2 ed. New York: Wiley.

Arrow, Kenneth J. 1973*a*. Formal theories of social welfare. In *Dictionary of the history of ideas*, ed. Philip Wiener. Vol. 4 New York: Charles Scribner's Sons pp. 276–284.

Arrow, Kenneth j. 1973*b*. "Some ordinalist-utilitarian notes on RawlsŠ theory of justice." *Journal of Philosophy* 70: 245–263.

Arrow, Kenneth J. 1977*b*. "Current developments in theory of social choice." *Social Research* 44(4): 607–622.

Arrow, Kenneth J. 1983 [1967]. Values and collective decision making. In *Collected papers of Kenneth Arrow, Vol 1; Social Choice and Justice*, ed. Kenneth J. Arrow. Cambridge Massachusetts: Belknap Press of Harvard University Press chapter 4, pp. 59–77. Appeared originally in P. Laslett and W.G. Runciman, eds., Philosophy, Politics and Society 1967, Oxford, Basil Blackwell, pp. 215-232.

Arrow, Kenneth J. 1983 [1973]*c*. Some ordinalist-utilitarian notes on Rawls' theory of justice. In *Collected papers of Kenneth J. Arrow: social choice* *and justice*, ed. Kenneth J. Arrow. Belknap Press of Harvard University Press. Published originally in 1973, Journal of Philosophy, 70, pp. 245-263.

Arrow, Kenneth J. 1983 [1977]*a*. Current developments in theory of social choice. In *Collected Papers of Kenneth J. Arrow: social choice and justice*, ed. Kenneth J. Arrow. Belknap Press of Harvard University Press chapter 12, pp. 162–174. Published originally in Social Research, vol. 44, no. 4, pp. 607-622.

Arrow, Kenneth J. 1983[1952]. The principle of rationality in collective decisions. In *Collected Papers of Kenneth J. Arrow: social choice and justice*, ed. Kenneth J. Arrow. Belknap Press of Harvard University Press chapter 1.

Arrow, Kenneth J. 1987. Reflections on the essays. In *Arrow and the foundations of the theory of economic policy*, ed. G. Feiwel. U.K.: Macmillan pp. 727–734.

Arrow, Kenneth J. 1997. The functions of social choice theory. In *Social choice re-examined, Proceedings of the IEA Conference held at Schloss Hernstein, Berndorf, near Vienna.* Vol. 1 pp. 3–9.

Austen-Smith, David & Jeffrey Banks. 1999. *Positive political theory, collective Preference.* Vol. 1 Ann Arbor: The University of Michigan Press.

Banks, Jeffrey. 1985. "Sophisticated voting outcomes and agenda control." *Social Choice and Welfare* 1: 295–306.

Baumol, William J. 1958. "The cardinal utility which is ordinal." *The Economic Journal* 68(272): 665–672.

Becker, Gary S. 1962. "Irrational behavior and economic theory." *Journal of Political Economy* 70(1): 1–13.

Becker, Gary S. & Kevin M. Murphy. 1988. "A theory of rational addiction." *The journal of political economy* 96(4): 675–700.

Binmore, Ken. 1994. *Playing fair; game theory and the social contract.* Vol. 1 Cambridge Massachusetts: The MIT Press.

Black, Duncan. 1958. *The theory of committees and elections.* Cambridge: Cambridge University Press.

Black, Duncan. 1976. "Partial justification of the Borda count." *Public Choice* 28(28): 1–15.

Black, Jerome H. 1978. "The multicandidate calculus of voting: Application to Canadian federal elections." *American Journal of Political Science* 22(3): 609–638.

Blackburn, Simon. 1995. "Practical Tortoise Raising." *Mind* 104(416): 695–711.

Blackburn, Simon. 1998. *Ruling passions: a theory of practical reasoning.* Oxford; New York: Clarendon Press.

Blackorby, Charles & Walter Bossert. 2006. Interpersonal comparisons of well-being. In *Oxford Handbook of Political Economy*, ed. Barry R. Weingast & Donald A. Wittman. New York: Oxford University Press pp. 408–425.

Blais, Andre & Richard Nadeau. 1996. "Measuring strategic voting: a two-step procedure." *Electoral Studies* 15(1): 39–52.

Blin, Jean-Marie. 1976. "How relevant are 'irrelevant' alternatives?" *Theory and Decision* 7: 95–105.

Blin, Jean-Marie & Mark Satterthwaite. 1978. "Individual decisions and group decisions, the fundamental differences." *Journal of Public Economics* 10: 247–268.

Borch, Karl. 1975. "Probabilities of probabilities (a comment)." *Theory and Decision* 6: 155–159.

Bordes, Georges & Nikolaus Tideman. 1991. "Independence of irrelevant alternatives in the theory of voting." *Theory and Decision* 30(2): 163–186.

Boudon, Raymond. 1998. "Limitations of rational choice theory." *American Journal of Sociology* 104(3): 817–828.

Boudon, Raymond. 2003. "Beyond rational choice theory." *Annual Review of Sociology* 29: 1–21.

Buchanan, James M. 1972. Toward analysis of closed behavioral systems. In *The theory of public choice. Political applications of economics*, ed. James M. Buchanan & Robert D. Tollison. Ann Arbor: University of Michigan Press pp. 11–23.

Buchanan, James M. 1986. *Liberty, market, and state: Political economy in the 1980s.* New York: New York University Press.

Buchanan, James M. 1989. *Essays on the political economy.* Honolulu: University of Hawaii Press.

Buchanan, James M. & Yong J. Yoon. 2006. "All voting is strategic." *Public Choice* 129: 159–167.

Calvert, Randall & Robert Fenno. 1994. "Strategy and sophisticated voting in the senate." *The Journal of Politics* 56(2): 349–376.

Camerer, Colin. 1998. "Bounded rationality in individual decision making." *Experimental Economics* 1(2): 163–183.

Camerer, Colin F. 1995. Individual decision-making. In *Handbook of Experimental Economics*, ed. John Kagel & Alvin Roth. Princeton: Princeton University Press.

Campbell, Donald E. 1973. "Social choice and intensity of preference." *Journal of Political Economy* 81(1): 211–18.

Carlsson, Hans & Eric van Damme. 1993. "Global games and equilibrium selection." *Econometrica* 61(5): 989–1018.

Casella, George & Roger L. Berger. 1990. *Statistical inference*. Belmont, California: Duxbury Press.

Chen, Kong-Pin & Sheng-Zhang Yang. 2002. "Strategic voting in open primaries." *Public Choice* 112(1-2): 1–30.

Coleman, James S. 1990. *Foundations of social theory*. Cambridge: Harvard University Press.

Coleman, Jules. 1966. "The possibility of a social welfare function." *American Economic Review* 56: 1105–1122.

Cox, Gary W. 1997. *Making votes count: strategic coordination in the world's electoral systems*. Political economy of institutions and decisions Cambridge: Cambridge University Press.

Cox, Gary W. 1999. "The empirical content of rational choice theory: a reply to Green and Shapiro." *Journal of Theoretical Politics* 11(2): 147.

Cubitt, Robin P. & Robert Sugden. 2001. "On money pumps." *Games and Economic Behavior* 37(1): 121–160.

Dahl, Robert A. 1956. *A preface to democratic theory*. Charles R. Walgreen Foundation lectures Chicago: University of Chicago Press.

de Borda, Jean-Charles. 1995[1784]. On elections by ballot. In *Classics in Social Choice Theory*, ed. Iain Mclean & Arnold B. Urken. Ann Arbor: University of Michigan Press pp. 83–89.

de Finetti, Bruno. 1977. Probabilities of probabilities: A real problem or a misunderstanding? In *New directions in the application of Bayesian methods*, ed. A. Aykaç & C. Brumat. North Holland, Amsterdam: Elsevier.

Dhillon, Amrita & Jean-Francois Mertens. 1999. "Relative utilitarianism." *Econometrica* 67(3): 471.

Diermeier, Daniel & Keith Krehbiel. 2003. "Institutionalism as a methodology." *Journal of Theoretical Politics* 15(2): 123–144.

Dowding, Keith M. 2006. "Can populism be defended? William Riker, Gerry Mackie and the interpretation of democracy." *Government and Opposition* 41(3): 327–346.

Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper.

Downs, Anthony. 1967. *Inside bureaucracy.* Boston: Little, Brown.

Dummett, Michael. 1998. "The Borda count and agenda manipulation." *Social Choice and Welfare* 15(2): 289–296.

Duverger, Maurice. 1954. *Political parties: their organization and activity in the modern state.* New York: Wiley. transl. Barbara and Robert North.

Eckel, Catherine & Charles Holt. 1989. "Strategic voting in agenda-controlled committee experiments." *American Economic Review* 59(4): 1252–1263.

Ellsberg, Daniel. 1961. "Risk, ambiguity, and the Savage axioms." *The Quarterly Journal of Economics* 75(4): 643–669.

Elster, Jon. 1983. *Explaining technical change.* Cambridge: Cambridge University Press.

Enelow, James. 1981. "Saving amendments, killer amendments, and an expected utility theory of sophisticated voting." *The Journal of Politics* 43(4): 1062–1089.

Enelow, James & Melvin Hinich. 1983. "Voter expectations in multi-stage voting systems: an equilibrium result." *The American Journal of Political Science* 27(4): 820–827.

Farquharson, Robin. 1969. *Theory of voting.* UK: Hodge.

Favardin, Pierre, Dominique Lepelley & Jérôme Serais. 2002. "Borda rule, Copeland method and strategic manipulation." *Review of Economic Design* 7(2): 213–228.

Felsenthal, Dan S. 1990. *Topics in social choice: sophisticated voting, efficacy, and proportional representation.* New York: Praeger.

Felsenthal, Dan S. 1996. "Setting the record straight: a note on sophisticated voting under Borda's method." *Public Choice* 89(1-2): 17–25.

Ferejohn, John A. 1991. Rationality and interpretation: Parliamentary elections in early Stuart England. In *The Economic Approach to Politics: A Critical Reassessment of the Theory of Rational Action*, ed. Kristen R. Monroe. New York: HarperCollins pp. 279–305.

Ferejohn, John A. 2002. "Symposium on Explanations and Social Ontology 1: Rational choice theory and social explanation." *Economics and Philosophy* 18(2): 211–234.

Ferejohn, John A. & Morris P. Fiorina. 1975. "Purposive models of legislative behavior." *The American Economic Review, Papers and Proceedings* 65(2): 407–414.

Fey, Mark. 1997. "Stability and coordination in Duverger's law: a formal model of preelection polls and strategic voting." *American Political Science Review* 91(1): 135–147.

Finke, Roger. 1997. The consequences of religious competition: supply-side explanations for religious change. In *Rational Choice Theory and Religion - Summary and Assessment*, ed. Lawrence A. Young. New York: Routledge pp. 45–64.

Fishburn, Peter C. 1989. "Retrospective on the utility theory of von Neumann and Morgenstern." *Journal of Risk and Uncertainty* 2(2): 127–158.

Fleurbaey, Marc & Peter Hammond. 2004. Interpersonally comparable utility. In *Handbook of Utility Theory*, ed. Salvador Barber, Peter Hammond & Christian Seidl. Vol. 2 Dordrect: Kluwer Academic Publishers.

Frankel, David M., Stephen Morris & Ady Pauzner. 2003. "Equilibrium selection in global games with strategic complementarities." *Journal of Economic Theory* 108(1): 1–44.

Frey, Bruno S. 1994. "How intrinsic motivation is crowded out and in." *Rationality and Society* 6(3): 334–352.

Friedman, Jeffrey, ed. 1996. *The rational choice controversy : economic models of politics reconsidered*. New Haven, CT: Yale University Press.

Friedman, Milton. 1953. The methodology of positive economics. In *Essays in Positive Economics*. Chicago: University of Chicago Press pp. 3–43.

Frohlich, Norman & Joe A. Oppenheimer. 1999. Kenneth J. Arrow, welfare aggregation, and progress in political theory. In *Competition and cooperation: Conversations with Nobelists about economics and political science*, ed. E. Alt, James, Margaret Levi & Elinor Ostrom. New York: Russell Sage Foundation pp. 4–32.

García-Lapresta, José L. & Miguel Martínez-Panero. 2002. "Borda count versus approval voting: a fuzzy approach." *Public Choice* 112(1 - 2): 167–184.

Gehrlein, William. 2002. "Condorcet's paradox and the likelihood of its occurrence: different perspectives on balanced preferences." *Theory and Decision* 52(2): 171–199.

Gibbard, Allan. 1973. "Manipulation of voting schemes, a general result." *Econometrica* 41: 587–60.

Gibbard, Allan F. & Hal R. Varian. 1978. "Economic models." *Journal of Philosophy* 75(11): 664–677.

Gilboa, Itzhak. 1987. "Expected utility with purely subjective non-additive probabilities." *Journal of Mathematical Economics* 16(1): 65–88.

Gilmour, John B. 2001. "The Powell amendment voting cycle: an obituary." *Legislative Studies Quarterly* 26(2): 249–262.

Gode, Dhananjay K. & Shyam Sunder. 1993. "Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality." *Journal of Political Economy* 101(1): 119–37.

Good, Irving J. 1962. Subjective probability as a measure of a non-measurable set. In *Logic, Methodology and Philosophy of Science*, ed. Ernest Nagel, Patrick Suppes & Alfred Tarski. Vol. I Stanford, Calif.: Stanford University Press.

Gärdenfors, Peter & Nils-Erik Sahlin. 1982. "Unreliable probabilities, risk-taking, and decision-making." *Synthese* 53: 361–86.

Green, Donald P. & Ian Shapiro. 1994. *Pathologies of rational choice theory: A critique of applications in political science*. New Haven: Yale University Press.

Hammond, Peter J. 1991*a*. "Independence of irrelevant interpersonal comparisons." *Social Choice and Welfare* 8(1): 1–19.

Hammond, Peter J. 1991*b*. Interpersonal comparisons of utility: Why and how they are and should be made. In *Interpersonal comparisons of well-being*, ed. Jon Elster & John E. Roemer. Cambridge: Cambridge University Press pp. 200–254.

Hansson, Bengt. 1973. "The independence condition in the theory of social choice." *Theory and Decision* 4: 25–49.

Hansson, Bengt. 1988. Risk aversion as a problem of conjoint measurement. In *Decision, Probability, and Utility*, ed. Peter Gärdernfors & Nils-Eric Sahlin. Cambridge: Cambridge University Press pp. 136–158.

Harrod, Roy. 1938. "Scope and method of economics." *Economic Journal* 48: 383–412.

Harsanyi, John C. 1953. "Cardinal utility in welfare economics and in the theory of risk-taking." *Journal of Political Economy* 61: 434–435.

Harsanyi, John C. 1955. "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility." *The Journal of Political Economy* 63(4): 309–321.

Harsanyi, John C. 1967-68. "Games with incomplete information played by 'Bayesian' players, parts I, II, and III." *Management Science* 14: 159–182, 320–334, and 486–502.

Harsanyi, John C. 1977. *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge Eng. ; New York: Cambridge University Press.

Harsanyi, John C. 1978. "Bayesian decision-theory and utilitarian ethics." *American Economic Review* 68(2): 223–228.

Harsanyi, John C. 1979. "Bayesian decision-theory, rule utilitarianism, and Arrow impossibility theorem." *Theory and Decision* 11(3): 289–317.

Harsanyi, John C. 1982*a*. "Rejoinder to professors Kadane and Larkey." *Management Science* 28(2): 124–125.

Harsanyi, John C. 1982*b*. "Subjective probability and the theory of games: Comments on Kadane and Larkey's paper." *Management Science* 28(2): 120–124.

Harsanyi, John C. 1987. Von Neumann-Morgenstern utilities, risk taking, and welfare. In *Arrow and the ascent of modern economic theory*, ed. George Feiwel. Oxford: Macmillan pp. 545–558.

Harsanyi, John C. 1992. Game and decision theoretic models in ethics. In *Handbook of game theory with economic applications*, ed. Robert Aumann & Sergiu Hart. Vol. 1 Amsterdam, London and Tokyo: North-Holland pp. 669–707.

Harsanyi, John C. 1993. Normative validity and meaning of von Neumann-Morgenstern utilities. In *Frontiers of game theory*, ed. Ken Binmore, Alan P. Kirman & Piero Tani. London: The MIT Press pp. 307–320.

Harsanyi, John C. 1995. "A new theory of equilibrium selection for games with incomplete information." *Games and Economic Behavior* 10(2): 318–332.

Hausman, Daniel. 1998. Rationality and knavery. In *Game Theory, Experience, Rationality*, ed. Werner Leinfellner & Köhler Eckehart. Dordrecht: Kluwer Academic Publishers pp. 67–79.

Hausman, Daniel M. & Michael S. McPherson. 1996. *Economic analysis and moral philosophy.* Cambridge surveys of economic literature Cambridge England ; New York, NY: Cambridge University Press.

Heckelman, Jac C. 2003. "Probabilistic Borda rule voting." *Social Choice and Welfare* 21(3): 455–468.

Hildreth, Clifford. 1953. "Alternative conditions for social orderings." *Econometrica* 21: 81–94.

Hirshleifer, Jack & John G. Riley. 1992. *The analytics of uncertainty and information.* Cambridge ; New York: Cambridge University Press.

Hoffman, D.T. 1982. "A model for strategic voting." *SIAM Journal on Applied Mathematics* 42(4): 751–761.

Iannaccone, Laurence R. 1997. Rational choice: framework for the scientific study of religion. In *Rational Choice Theory and Religion - Summary and Assessment*, ed. Lawrence A. Young. New York: Routledge pp. 25–44.

Jackson, Frank & Philip Pettit. 1992. Structural explanation in social theory. In *Reduction, Explanation, and Realism*, ed. David Charles & Kathleen Lennon. Oxford: Clarendon Press pp. 97–131.

Jung, Joon Pyo. 1987. Condorcet consistent binary agendas under incomplete information. In *Models of strategic choice in politics*, ed. Peter Ordeshook. Ann Arbor: University of Michigan Press pp. 315–341.

Kalai, Ehud & David Schmeidler. 1977. "Aggregation procedure for cardinal preferences: A formulation and proof of Samuelson's impossibility conjecture." *Econometrica: Journal of the Econometric Society* 45(6): 1431–1438.

Kavka, Gregory. 1991. Rational maximizing in economic theories of politics. In *The Economic Approach to Politics: A Critical Reassessment of the Theory of Rational Action*, ed. Kristen R. Monroe. New York: HarperCollins pp. 371–385.

Kelly, Jerry S. 1988. *Social choice theory: an introduction*. Berlin, New York: Springer-Verlag. Jerry S. Kelly.; Includes index.

Kemp, Murray C. & Yew-Kwang Ng. 1987. Arrow's independence condition and the Bergson-Samuelson tradition. In *Arrow and the foundations of the theory of economic policy*, ed. R. Feiwel, George. New York University Press pp. 223–241.

Kirman, Alan P. 1989. "The intrinsic limits of modern economic theory: The emperor has no clothes." *The Economic Journal* 99(395, Supplement: Conference Papers): 126–139.

Knight, Frank. 1994[1935]. Economics and human action. In *The Philosophy of Economics: An Anthology*, ed. Daniel Hausman. Cambridge: Cambridge University Press.

Kolm, Serge-Christophe. 1996. *Modern theories of justice*. Cambridge, MA: MIT Press.

Krehbiel, Keith & Douglas Rivers. 1990. "Sophisticated voting in congress: A reconsideration." *The Journal of Politics* 52(2): 548–578.

Langlois, Richard N. 1986. Rationality, institutions, and explanation. In *Economics as a process: Essays in the new institutional economics*, ed. Richard N. Langlois. Cambridge; New York and Melbourne: Cambridge University Press pp. 225–55.

Lehtinen, Aki. 2002. "The effect of the number of alternatives and voters on whether strategic voting is welfare-increasing in amendment agendas." mimeo.

Lehtinen, Aki. 2006*a*. "Signal extraction for simulated games with a large number of players." *Computational Statistics and Data Analysis* 50: 2495–2507.

Lehtinen, Aki. 2006*b*. "The welfare consequences of strategic voting in plurality and runoff rules." Unpublished manuscript, University of Helsinki, Department of social and moral philosophy.
URL: *http://www.valt.helsinki.fi/staff/alehtine/pandr.pdf*

Lehtinen, Aki. 2007. "The welfare consequences of strategic voting in two commonly used parliamentary agendas." *Theory and Decision* 63(1): 1–40.

Lehtinen, Aki. forthcoming. "The Borda rule is also intended for dishonest men." *Public Choice* .

Lepelley, Dominique & Boniface Mbih. 1994. "The vulnerability of four social choice functions to coalitional manipulation of preferences." *Social Choice and Welfare* 11: 256–265.

Lepelley, Dominique & Fabrice Valognes. 2003. "Voting rules, manipulability and social homogeneity." *Public Choice* 116(1-2): 165–184.

Levins, Richard. 1993. "A response to Orzack and Sober: Formal analysis and the fluidity of science." *quarterly Review of Biology* 68(4): 547–555.

Luce, R.D. & Howard Raiffa. 1957. *Games and decisions; introduction and critical survey.* New York: Wiley.

Ludwin, William G. 1978. "Strategic voting and the Borda method." *Public Choice* 33(1): 85–90.

MacDonald, Paul K. 2003. "Useful fiction or miracle maker: The competing epistemological foundations of rational choice theory." *American Political Science Review* 97(4): 551–565.

Machlup, Fritz. 1946. "Marginal analysis and empirical research." *American Economic Review* 36(4): 519–554.

Mackay, Alfred. 1980. *Arrow's theorem: The paradox of social choice, a case study in the philosophy of economics.* London: Yale University Press.

Mackie, Gerry. 2003. *Democracy defended.* Contemporary political theory Cambridge: Cambridge University Press. Gerry Mackie.; Includes bibliographical references.

Marchant, Thierry. 2000. "Does the Borda rule provide more than a ranking?" *Social Choice and Welfare* 17(3): 381–391.

Margolis, Julius. 1975. "Bureaucrats and politicians: Comment." *Journal of Law and Economics* 18(3): 645–59.

Marschak, Jacob. 1975. "Personal probabilities of probabilities." *Theory and Decision* 6: 121–153.

McKelvey, Richard. 1976. "Intransitivities in multidimensional voting models and some implications for agenda institutions." *Journal of Economic Theory* 12: 472–482.

McKelvey, Richard. 1990. Game theoretic models of voting in multidimensional issue spaces. In *Game Theory and Applications*, ed. T. Ichiichi, Abraham Neyman & Y. Tauman. San Diego: Academic Press.

McKelvey, Richard D. & Peter C. Ordeshook. 1972. A general theory of the calculus of voting. In *Mathematical Applications in Political Science*, ed. James F. Herndon & Joseph L. Bernd. Charlottesville: University Press of Virginia pp. 32–78.

McKelvey, Richard & Richard Niemi. 1978. "A multistage game representation of sophisticated voting for binary procedures." *Journal of Economic Theory* 18: 1–22.

McLean, Iain. 1987. *Public choice: an introduction.* New York: Basil Blackwell.

Merrill, Samuel III. 1981*a*. "Strategic decisions under one-stage multi-candidate voting systems." *Public Choice* 36(1): 115–134.

Merrill, Samuel III. 1981*b*. Strategic voting in multicandidate elections under uncertainty and under risk. In *Power, Voting, and Voting Power*, ed. Manfred J. Holler. Würzburg: Physica-Verlag pp. 179–187.

Merrill, Samuel III. 1984. "A comparison of efficiency of multicandidate electoral systems." *American Journal of Political Science* 28(1): 23–48.

Merrill, Samuel III. 1988. *Making multicandidate elections more democratic.* New Jersey: Princeton University Press.

Miller, Nicholas R. 1977. "Graph-theoretical approaches to the theory of voting." *American Journal of Political Science* 21(4): 769–803.

Miller, Nicholas R. 1980. "A new solution set for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting." *American Journal of Political Science* 24: 68–96.

Miller, Nicholas R. 1995. *Committees, agendas, and voting.* Vol. 58 of *Fundamentals of pure and applied economics* Chur, Switzerland: Harwood Academic Publishers.

Mäki, Uskali. 1989. "On the problem of realism in economics." *Ricerche Economiche* 43(1-2): 176–198.

Mäki, Uskali. 1992. "Friedman and realism." *Research in the history of economic thought and methodology* 10: 171–195.

Mäki, Uskali. 1998. As if. In *The Handbook of Economic Methodology*, ed. John B. Davis, Wade Hands & Uskali Mäki. Cheltenham: Edward Elgar pp. 25–27.

Mongin, Philippe. 2001. "The impartial observer theorem of social ethics." *Economics and Philosophy* 17(2): 147–179.

Monroe, Kristen R. 1997. *Contemporary empirical political theory*. Berkeley: University of California Press.

Morris, Stephen & Hyun Song Shin. 2003. Global games: theory and applications. In *Advances in economics and econometrics*, ed. Lee Hansen Mathias Dewatripont & S. Turnovsky. Cambridge: Cambridge University Press pp. 56–114.

Mosteller, Frederick & Philip Nogee. 1951. "An experimental measurement of utility." *The Journal of Political Economy* 59(5): 371–404.

Moulin, Hervé. 1979. "Dominance solvable voting schemes." *Econometrica* 47(6): 1337–1352.

Mueller, Dennis C. 1989. *Public choice*. 2 ed. Cambridge: Cambridge University Press.

Mueller, Dennis C. 2004. "Models of man: Neoclassical, behavioural, and evolutionary." *Politics, Philosophy and Economics* 3(1): 59–76.

Myatt, David P. 2007. "On the theory of strategic voting." *Review of Economic Studies* 74(1): 255–281.

Myatt, David P. & Stephen D. Fisher. 2002. "Everything is uncertain and uncertainty is everything: strategic voting in simple plurality elections." University of Oxford, discussion paper no. 115.

Myerson, Roger B. 1985. Bayesian equilibrium and incentive-compatibility: an introduction. In *Social Goals and Social Organization: Essays in Memory of Elisha Pazner*, ed. Leonid Hurwicz, David Schmeidler & Hugo Sonnenschein. Cambridge University Press pp. 229–259.

Myerson, Roger B. 1999. "Nash equilibrium and the history of economic theory." *Journal of Economic Literature* 37(3): 1067–1082.

Myerson, Roger B. & Robert J. Weber. 1993. "A theory of voting equilibria." *American Political Science Review* 87(1): 102–114.

Ng, Yew-Kwang. 1979. *Welfare economics : introduction and development of basic concepts*. London: Macmillan.

Ng, Yew-Kwang. 1999. "Utility, informed preference, or happiness: Following Harsanyi's argument to its logical conclusion." *Social Choice and Welfare* 16(2): 197–216.

Niskanen, William A. 1971. *Bureaucracy and representative government.* Chicago: Aldine, Atherton.

Niskanen, William A. 1975. "Bureaucrats and Politicians." *Journal of Law and Economics* 18(3): 617–43.

Ordeshook, Peter. 1986. *Game theory and political theory: An introduction.* Cambridge: Cambridge University Press.

Ordeshook, Peter & Thomas Palfrey. 1988. "Agendas, strategic voting, and signaling with incomplete information." *American Journal of Political Science* 32: 441–466.

Ordeshook, Peter & Thomas Schwartz. 1987. "Agendas and the control of political outcomes." *American Political Science Review* 81: 179–200.

Pattanaik, Prasanta K. 1968. "Risk, impersonality, and the social welfare function." *Journal of Political Economy* 76(6): 1152–1169.

Pattanaik, Prasanta K. 2002. Positional rules of collective decision-making. In *Handbook of Social Choice and Welfare*, ed. Kenneth J. Arrow, Amartya K. Sen & Kotaro Suzumura. Vol. 1 Amsterdam, London and New York: Elsevier Science, North-Holland pp. 361–94.

Plott, Charles R. 1976. "Axiomatic social choice theory: An overview and interpretation." *American Journal of Political Science* 20(3): 511–596.

Quattrone, George A. & Amos Tversky. 1988. "Contrasting rational and psychological analyses of political choice." *American Political Science Review* 82(3): 719–736.

Rabin, Matthew. 2000. "Risk aversion and expected-utility theory: a calibration theorem." *Econometrica* 68(5): 1281–1292.

Rawls, John. 1971. *A theory of justice.* Cambridge, Mass.: Belknap Press of Harvard University Press.

Ray, Paramesh. 1973. "Independence of irrelevant alternatives." *Econometrica* 41(5): 987–91.

Riker, William H. 1982. "The two-party system and Duverger's law: An essay on the history of political science." *American Political Science Review* 76(4): 753–766.

Risse, Mathias E. 2001. "Arrow's theorem, indeterminacy, and multiplicity reconsidered." *Ethics* 111(4): 706–734.

Risse, Mathias E. 2004. "Arguing for majority rule." *Journal of Political Philosophy* 12(1): 41–64.

Risse, Mathias E. 2005. "Why the count de Borda cannot beat the marquis de Condorcet." *Social Choice and Welfare* 25(1): 95–113.

Robbins, Lionel. 1938. "Interpersonal comparisons of utility: A comment." *Economic Journal* 48(192): 635–641.

Robbins, Lionel. 1952. *An Essay on the Nature and Significance of Economic Science*. 2 ed. London: Macmillan.

Roemer, John E. 1996. *Theories of distributive justice*. Cambridge, Mass: Harvard University Press.

Rogeberg, Ole. 2004. "Taking absurd theories seriously: Economics and the case of rational addiction theories." *Philosophy of science* 71(3): 263–285.

Rosenberg, Alexander. 1980. "A sceptical history of economic theory." *Theory and Decision* 12: 79–93.

Rosenberg, Alexander. 1992. *Economics - Mathematical Politics or Science of Diminishing Returns?* Chicago: University of Chicago Press.

Rothenberg, Jerome. 1961. *The measurement of social welfare*. Englewood Cliffs, N.J.: Prentice-Hall.

Rubinstein, Ariel. 1991. "Comments on the interpretation of game theory." *Econometrica* 59(4): 909–924.

Saari, Donald G. 1990*a*. "The Borda dictionary." *Social Choice and Welfare* 7: 279–317.

Saari, Donald G. 1990*b*. "Susceptibility to manipulation." *Public Choice* 64(1): 21–41.

Saari, Donald G. 1995. Inner consistency or not inner consistency: A reformulation is the answer. In *Social Choice, Welfare, and Ethics*, ed. William A. Barnett, Hervé Moulin, Maurice Salles & Norman Schofield. Cambridge: Cambridge University Press pp. 187–212.

Saari, Donald G. 1998. "Connecting and resolving Sen's and Arrow's theorems." *Social Choice and Welfare* 15(2): 239–261.

Saari, Donald G. 2001. *Decisions and elections: Explaining the unexpected*. Cambridge: Cambridge University Press.

Saari, Donald G. 2003*a*. "Capturing the 'Will of the People'." *Ethics* 113(2): 333–349.

Saari, Donald G. 2003*b*. "Unsettling aspects of voting theory." *Economic Theory* 22(3): 529–555.

Saari, Donald G. 2006. "Which is better: the Condorcet or Borda Winner?" *Social Choice and Welfare* 26(1): 107–129.

Saari, Donald G. & Katri K. Sieberg. 2001. "The sum of the parts can violate the whole." *The American Political Science Review* 95(2): 415–433.

Saari, Donald G. & Katri K. Sieberg. 2004. "Are partwise comparisons reliable?" *Research in Engineering Design* 15: 62–71.

Saari, Donald G. & Vincent R. Merlin. 2000. "A geometric examination of Kemeny's rule." *Social Choice and Welfare* 17: 403–438.

Satterthwaite, Mark. 1975. "Strategy-proofness and Arrows conditions: Existence and correspondence theorems for voting procedures and social welfare functions." *Journal of Economic Theory* 10: 187–217.

Satz, Debra & John A. Ferejohn. 1994. "Rational choice and social theory." *Journal of Philosophy* 91(2): 71–87.

Savage, Leonard J. 1954. *The foundations of statistics*. Wiley publications in statistics New York: Wiley.

Schelling, Thomas C. 1978. *Micromotives and macrobehavior*. New York: Norton.

Schmeidler, David. 1989. "Subjective probability and expected utility without additivity." *Econometrica* 57(3): 571–587.

Schofield, Norman. 1978. "Instability of simple dynamic games." *Review of Economic Studies* 45: 575–594.

Sen, Amartya K. 1970. *Collective choice and social welfare*. Vol. 5 of *Mathematical economics texts* San Francisco: Holden-Day.

Sen, Amartya K. 1976. "Welfare inequalities and Rawlsian axiomatics." *Theory and Decision* 7(4): 243–262.

Sen, Amartya K. 1977. Non-linear social welfare functions: a reply to professor Harsanyi. In *Foundational Problems in the Special Sciences*, ed. R Butts & J. Hintikka. Dordrecht: Reidel pp. 297–302.

Sen, Amartya K. 1985. "Rationality and uncertainty." *Theory and Decision* 18: 109–127.

Sen, Amartya K. 1986. Social choice theory. In *Handbook of mathematical economics*, ed. Kenneth Arrow & Michael Intriligator. Vol. 3 Amsterdam: Elsevier Science publishers B.V. pp. 1073–1181.

Sen, Amartya K. 1987. *On ethics and economics*. The Royer lectures Oxford Oxfordshire ; New York, NY, USA: B. Blackwell. Amartya Sen.; Bibliography: p. [90]-122.; Includes indexes.

Shafir, Eldar & Robyn A. LeBoeuf. 2002. "Rationality." *Annual Review of Psychology* 53(1): 491–517.

Shepsle, Kenneth & Barry Weingast. 1984. "Uncovered sets and sophisticated voting outcomes with implications for agenda institutions." *American Journal of Political Science* 28: 27–59.

Shepsle, Kenneth & Barry Weingast. 1994. "Positive theories of congressional institutions." *Legislative Studies Quarterly* 19(2): 149–179.

Simon, Herbert A. 1985. "Human nature in politics: The dialogue of psychology with political science." *American Political Science Review* 79(2): 293–304.

Skyrms, Brian. 1990. *The dynamics of rational deliberation.* Cambridge, Mass.: Harvard University Press.

Sloth, Birgitte. 1993. "The theory of voting and equilibria in noncooperative games." *Games and Economic Behaviour* 5: 152–169.

Smith, David A. 1999. "Manipulability measures of common social choice functions." *Social Choice and Welfare* 16(4): 639–661.

Spence, Michael. 1973. "Job market signaling." *Quarterly Journal of Economics* 87(3): 355–374.

Strasnick, Steven. 1976. "The problem of social choice: Arrow to Rawls." *Philosophy and Public Affairs* 5(3): 241–273.

Stratmann, Thomas. 1997. Logrolling. In *Perspectives on public choice: A handbook*, ed. Dennis C. Mueller. Cambridge: Cambridge University Press pp. 322–341.

Taylor, Alan D. 2005. *Social choice and the mathematics of manipulation.* New York: Cambridge University Press.

Tsetlin, Ilia, Michel Regenwetter & Bernard Grofman. 2003. "The impartial culture maximizes the probability of majority cycles." *Social Choice and Welfare* 21(3): 387–398.

Udehn, Lars. 1996. *The limits of public choice: a sociological critique of the economic theory of politics.* New York: Routledge.

Van Hees, Martin & Keith Dowding. forthcoming. "In praise of manipulation." *British Journal of Political Science* . available at http://med.ces.free.fr/pdf/vanHees.pdf.

Volden, Craig. 1998. "Sophisticated voting in supermajoritarian settings." *Journal of Politics* 60(1): 149–173.

Weymark, John A. 1991. A reconsideration of the Harsanyi-Sen debate on utilitarianism. In *Interpersonal comparisons of well-being*, ed. Jon Elster & John Roemer. Cambridge: Cambridge University Press.

Wilkerson, John D. 1999. “"Killer" amendments in congress.” *American Political Science Review* 93(3): 535–552.

Wimsatt, William C. 1981. Robustness, Reliability and Overdetermination. In *Scientific inquiry and the social sciences*, ed. Marilynn B. Brewer & Barry E. Collins. San Francisco: Jossey-Bass pp. 124–163.

Woodward, James. 2003. *Making things happen: a theory of causal explanation*. Oxford studies in philosophy of science Oxford and New York: Oxford University Press.

Zafirowski, Milan. 1999. “Unification of sociological theory by the rational choice model: conceiving the relationship between economics and sociology.” *Sociology* 33(3): 495–514.