Technische Universität München

Fakultät für Informatik
Lehrstuhl für Wirtschaftsinformatik und Entscheidungstheorie

Generating Insights in Social Choice Theory via Computer-aided Methods

Christian Thorsten Geist

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r):		Prof. Dr. Francisco Javier Esparza Estaun
Prüfer der Dissertation:	1.	Prof. Dr. Felix Brandt
	2.	Prof. Tobias Nipkow, Ph.D.

Die Dissertation wurde am 07.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 02.08.2016 angenommen.

GENERATING INSIGHTS IN SOCIAL CHOICE THEORY VIA COMPUTER-AIDED METHODS

CHRISTIAN GEIST

Christian Geist: *Generating Insights in Social Choice Theory via Computer-aided Methods*, © September 2016.

E-MAIL:
geist@in.tum.de

This thesis was typeset based on a tasteful style designed by Hans Georg Seedig. It uses LaTeX and the ClassicThesis style by André Miede, combined with the ArsClassica package by Lorenzo Pantieri and some minor modifications (mostly by Hans Georg Seedig). The text is set in Palatino with math in Euler, both due to Hermann Zapf. Headlines are set in Iwona by Janusz M. Nowacki, the monospace font is Bera Mono designed by Bitstream, Inc. Most of the graphics were created using Ti*k*Z by Till Tantau.

ABSTRACT

The rise of *computational social choice*, a relatively new research field, indicates that a computer-science perspective on social choice theory is fruitful and can contribute to our understanding of collective decision making. In this thesis, we show how substantial insights in social choice theory can be achieved via computer-aided methods that are based on powerful solving techniques, such as SAT and SMT (satisfiability modulo theories).

Our contribution is twofold: first and foremost, we provide a range of such computer-aided methods for the domain of social choice theory. Most importantly, this includes techniques for computer-aided theorem proving as initially proposed by Tang and Lin [2009]. But also the generation of (counter-)examples, the design of solving-based algorithms, and experimental analyses can significantly benefit from the use of constraint solvers.

The second contribution of this thesis lies in the numerous results for social choice theory that we obtained with the aforementioned methods. We resolve open problems regarding different notions of strategic manipulation for set-valued and probabilistic social choice, analyze structural properties of majority graphs and preference profiles, and take a practical perspective on issues in voting and preference aggregation. For instance, we generalize a set of existing theorems by proving that weak notions of efficiency and strategyproofness are incompatible in probabilistic social choice.

PUBLICATIONS

This cumulative (publication-based) thesis contains the following publications and working papers in their original, published form:*

COMPUTER-AIDED THEOREM PROVING

[1] Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565-602, 2016 (with F. Brandt).†

[2] Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 18–24. AAAI press, 2015 (with F. Brandl, F. Brandt, and J. Hofbauer).

[3] Optimal bounds for the no-show paradox via SAT solving. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 314–322. IFAAMAS, 2016 (with F. Brandt and D. Peters).

[4] Proving the incompatibility of efficiency and strategyproofness via SMT solving. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 116–122. AAAI Press, 2016 (with F. Brandl and F. Brandt).

SOLVING-BASED ALGORITHMS

[5] A note on the McKelvey uncovered set and Pareto optimality. *Social Choice and Welfare*, 46(1):81–91, 2016 (with F. Brandt and P. Harrenstein).

[6] Identifying k-majority digraphs via SAT solving. In *Proceedings of the 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2014 (with F. Brandt and H. G. Seedig).

[7] Finding preference profiles of Condorcet dimension k via SAT. Technical report, http://arxiv.org/abs/1402.4303, 2014.

* Supplementary page numbers have been added for convenience.

† An earlier version of this paper appeared in the Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 1193–1200. IFAAMAS, 2014.

[8] Analyzing the practical relevance of voting paradoxes via Ehrhart theory, computer simulations, and empirical data. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 385–393. IFAAMAS, 2016 (with F. Brandt and M. Strobel).

[9] Pnyx: A powerful and user-friendly tool for preference aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1915–1916. IFAAMAS, 2015 (with F. Brandt and G. Chabin).

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude towards my supervisor Felix Brandt. All the way through this thesis project I could not have imagined better supervision: always approachable, giving knowledgable and actionable advice, co-creating results and articles, and yet offering me space and a stage to develop and progress, Felix really made a difference to my experience in academia. Furthermore, he is truly committed and provides a great working environment, which forms the basis for a happy and successful team, whom—consisting of Florian Brandl, Markus Brill, Johannes Hofbauer, and Hans Georg Seedig—I would also like to thank for many engaging discussions and the good times we had at TUM as well as at multiple conferences. Big thanks also go to

- my friends, co-authors, and colleagues in the community, among them particularly Haris Aziz, Umberto Grandi, Zhen Hao, Paul Harrenstein, Nick Mattei, and Dominik Peters for numerous helpful discussions and fun encounters,
- Mate Soos and Valentin Mayer-Eichberger for getting me started with SAT solving, Jasmin Christian Blanchette for assistance with Isabelle/HOL, Alberto Griggio and Mohammad Mehdi Pourhashem Kallehbasti for practical guidance with MathSAT and z3, respectively, and Susanne Page for showing me ideas and techniques from category theory,
- my experienced mentors in the community, in particular, Ulle Endriss for introducing me to computational social choice, Michel Le Breton for his caring advice, ideas, and challenges, and Bill Zwicker for his very supportive and helpful comments,
- my students Michael Bay, Elisabeth Brändle, Guillaume Chabin, and Martin Strobel for being part of the team, their good work and contribution, and the fun we had,
- the participants and organizers of the much celebrated doctoral school in Estoril in 2010 for keeping me enthusiastic about coming back to social choice and academia,
- McKinsey & Company, Inc. for giving me the chance to take such a long leave of absence for working on this thesis,
- and, last but not least, my friends and my family, including my wife, parents, brother, grandparents, parents-in-law for always being there for me and supporting me, and my wife and son, in particular, for putting a smile on my face every single day!

Thank you all for this highly enjoyable time in academia!

FUNDING SOURCES My work was supported by Deutsche Forschungsgemeinschaft under grant BR 2312/9-1. I also gratefully acknowledge travel support by the TUM Graduate School (TUM-GS), the COST Action IC1205 on Computational Social Choice, the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), and the 12th Meeting of the Society for Social Choice and Welfare.

CONTENTS

Part I

INTRODUCTION TO AND SUMMARY OF CONTRIBUTIONS

1 INTRODUCTION

> "The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that [...] when there are disputes among persons, we can simply say: Let us calculate, without further ado, to see who is right."
>
> — G. W. Leibniz, 1686

The impact of computers, and artificial intelligence in particular, in our everyday lifes is undebatable. We experience self-driving cars, autonomous robots, automated online assistants, computer-aided interpretation of medical images, and many more supporting systems that facilitate otherwise tedious or complex tasks. In formal sciences, however, successful applications of automated reasoning are much less frequent. Other than a few lighthouse theorems, such as the four color theorem, few results have been proved with computer-aided methods and, to date, most applications of mechanized reasoners are to be found in industrial applications, most prominently hardware and software design.

In pure mathematics and related disciplines, successful applications of automated theorem provers have mostly concentrated on the verification of existing results and proofs thereof. Corresponding systems, as powerful as they are, in addition, usually have to be operated by specialists. The task of *theorem discovery*, i.e., the search for *novel* results, is mostly carried out manually and without machine-support.

Yet, we will see that social choice theory has three characteristics that make it well-suited for computer-aided reasoning: it uses the axiomatic method,[1] it is concerned with combinatorial structures, and its main concepts can be defined based on rather elementary mathematical notions.

The goal of this thesis project, thus, is to broaden and deepen the basis for applying computer-aided methods to theoretical economics, and social choice theory in particular. Our main tools for this task are satisfiability (SAT) solvers, i.e., pieces of software that apply powerful heuristics to decide whether a given propositional formula has a satisfying assignment or not. But we also apply other solving paradigms, such as satisfiability modulo theories (SMT), answer set programming (ASP), and integer programming (IP). In order to make use

1 I.e., it defines economical and social requirements in mathematically precise terms and then reasons about them deductively.

of these general problem solving tools, we need to translate instances of problems from social choice theory into the muss less expressive languages accepted by these solvers. If we manage to do so, however, we are rewarded with an easily adaptable, automated assistant that allows for "testing" of finite conjectures (and variants thereof with minimal effort by simply replacing or altering some axioms). In many cases these finite results carry over to arbitrarily large instances by easy-to-prove inductive lemmas, which then yields novel results of full generality instantaneously.

But let us now have a brief glimpse at the subject matter under consideration before we return to methodological aspects in Section 1.2 and Chapter 3 in particular.

1.1 SOCIAL CHOICE THEORY AND COMPUTATIONAL SOCIAL CHOICE

Social choice theory is a truly interdisciplinary field with contributions by mathematicians, economists, political scientists, and, more recently, computer scientists. The discipline is concerned with the analysis of *preference aggregation* and *collective decision making*, both of which occur in a multitude of forms. Most apparent are political elections, in which a set of *voters* select one or many representatives. Much more frequent, however, are everyday decisions of groups (without monetary payments), say colleagues deciding for a lunch location or friends deciding on which movie to watch together. And then there is an abundance of applications in computer science where *agents* (for instance, robots or distributed software) have to agree on joint plans and actions.[2] Another example is the aggregation of results of different search engines (by so-called metasearch engines), which is also often viewed as a preference aggregation problem.

What all these applications have in common is that they require an aggregation mechanism, a function that takes as input the preferences of all agents and outputs a collective choice.[3] Unfortunately, striking *impossibility results*, such as *Arrow's theorem* and the *Gibbard-Satterthwaite theorem*, state that even very basic conditions any aggregation mechanism should intuitively satisfy are incompatible with each other. While this might seem to indicate that no reasonable mechanism for collective decision making exists—and, certainly, there is no perfect mechanism for any application—a multitude of

2 In the following (except the original publications) we limit the reference to "voters" as much as possible in favor of the slightly more general term "agent".

3 Depending on the setting the collective choice may be a ranking, single winner, set of winners, lottery over winners, etc.

mechanisms has evolved, each of which satisfies a different set of properties.

As an example, a property that has received particular interest and is generally well-understood is the notion of *strategyproofness*. Strategyproofness postulates that, even in the case that an agent has full knowledge about the other agents' preferences, he cannot benefit from manipulating the election, i.e., from misrepresenting his true preferences. Unfortunately, the *Gibbard-Satterthwaite theorem* says that only dictatorships can be strategyproof.[4] In order to escape this impossibility, three major approaches have been suggested in the literature for social choice [see, e.g., Moulin, 1980],[5] out of which we concentrate on (1.) and (2.) in this thesis:

1. allowing for *sets* of alternatives as outcomes rather than just single winners (*set-valued* social choice, cf. Section 2.2),

2. allowing for *probability distributions* as outcomes (*randomized*, or *probabilistic* social choice, cf. Section 2.3), and

3. restricting the domain of preferences (most prominently to single-peaked preferences, which guarantees the existence of a Condorcet winner and hence allows for strategyproof rules as observed already by Dummett and Farquharson [1961]).

While each of these approaches generally allows for strategyproof aggregation mechanisms, options (1.) and (2.) entail that the notion of strategyproofness is no longer uniquely defined. In these settings, the exact meaning of strategyproofness heavily relies on how preferences over individual alternatives are extended to sets of, or even lotteries over these alternatives. An analysis of the different notions of strategyproofness and whether they allow for desirable aggregation functions is a key concern of social choice theory and, furthermore, a main point of departure for many of our results (reflected, e.g., in all of our publications that fall into the category of computer-aided theorem proving, i.e., Publications [1], [2], [3], and [4]).[6]

In contrast to classical social choice theory, *computational social choice* is a relatively new discipline that has its roots in contributing to social choice by taking a computational perspective, i.e., the perspective of computer science. For instance, regarding the notion of strategyproofness, Bartholdi, III et al. [1989] show in a seminal article how computational complexity can be used as a shield against manipulations, thereby providing voting rules that are strategyproof *in practice* despite the theoretical possibility of a manipulation. Quite soon,

4 Unless all but at most 2 alternatives are completely ignored, in which case a simple decision by majority works.

5 Monetary side-payments would also form an option, but these are usually not considered a part of social choice theory.

6 Publications [2] and [3] are concerned with a variant of strategyproofness called *participation*.

however, also the opposite direction became popular and is now an integral part of computational social choice: importing concepts from social choice theory into computer science, and multi-agent systems in particular. Since the 2000s we can observe a very active and stable community with contributions in voting theory, ressource allocation, fair division, coalition formation, and judgement aggregation, just to name a few. For an extensive account of the area of computational social choice, including a historical perspective, see the recent books by Brandt et al. [2016a] and Rothe [2015], and, in particular, the introductory chapter of the former [Brandt et al., 2016b].

Our own work clearly falls into the domain of computational social choice as it applies computational tools to generate insights in social choice theory. Yet, the particular methods applied have only been brought to social choice relatively recently, as we will argue when discussing *methodologically related* work in Section 1.2. For overviews of related work from a social choice perspective, the reader is—given the multitude of different aspects for which we were able to gain insights via computer-aided methods—referred to Chapter 4 and to the respective original publications referenced therein.

1.2 COMPUTER-AIDED RESULTS IN SOCIAL CHOICE THEORY

While time and again there have been contributions to social choice theory that in some way used computers to generate insights in social choice theory, computer-aided theorem proving was only brought to this field in 2008 by Lin and Tang; a journal version of their work appeared one year later [Tang and Lin, 2009]. In the method they introduced, they reduce well-known impossibility results, such as Arrow's theorem, to finite instances, which can then be checked by a satisfiability (SAT) solver [for a general introduction to satisfiability, see, e.g., Biere et al., 2009]. We extended this method in previous work to a fully-automatic search algorithm for impossibility theorems in the relatively simple context of preference relations over sets of alternatives [Geist and Endriss, 2011].

While those contributions mark the point of departure for this thesis project, in recent years similar solving-based techniques have proven to be quite effective for other problems in economics, too. Examples are the work by Fréchette et al. [2016], in which SAT solvers are used for the development and execution of the FCC's reverse spectrum auction, and recent results by Drummond et al. [2015], who solve stable matching problems via SAT solving. Again closer to computer-aided theorem proving and discovery is the article by Tang and Lin [2011], who apply SAT solving to identify classes of two-player games with unique pure Nash equilibrium payoffs. In another recent paper, Caminati et al. [2015] verified combinatorial Vickrey auctions via higher-order theorem provers.

In some respect, our approach also bears similarities to *automated mechanism design* [see, e.g., Conitzer and Sandholm, 2002], where desirable properties are encoded as constraints, too, but mechanisms are computed to fit specific problem instances (rather than being applicable generally). In a similar spirit, Mennle and Seuken [2015] run linear programs in order to compute optimal (randomized) choice mechanisms that satisfy approximate versions of strategyproofness and efficiency.[7]

Also related, but directed more towards formalizing and verifying *existing* results and proofs thereof, is a body of work on logical formalizations of important theorems in social choice theory, most prominently, *Arrow's theorem* [see, e.g., Nipkow, 2009, Grandi and Endriss, 2013, Cinà and Endriss, 2015].

During the course of this project the results obtained by computer-aided theorem proving have already found some attention in the social choice community [Chatterjee and Sen, 2014] and further popularizing work is underway [Kerber et al., 2015]. It still, however, appears to be a long way from the current systems—that have to be operated by expert users or programmers—to an automatic proof assistant for social choice theory which is powerful enough to discover new results and which, at the same time, is easy to use. We elaborate on this point a little more in Chapter 8.

1.3 OUTLINE AND CONTRIBUTION OF THIS THESIS

With this thesis we contribute in two ways to the state of the art: first, by providing methods for computer-aided analysis in social choice and potentially other areas of (computational) economics, and, second, by answering questions from those fields that remained open after "manual" efforts to solve them. In particular, we concentrate on finding new results (rather than verifying existing theorems), and especially those that are unlikely to have been found without the help of computers.

A few selected highlights of such novel results are:

- we prove for set-valued social choice that Pareto optimality is incompatible with Fishburn-strategyproofness and -participation, respectively,

7 The specific choices of these properties allowed for the use of linear programming rather than the more expressive framework of satisfiability modulo theories (SMT), which we had to apply in Publication [4].

- we extract complex but *human-readable* proofs of these theorems from the SAT solver's output,[8]
- we leverage SMT solving to strengthen a set of existing results for *randomized* social choice mechanisms by showing that a weak form of strategyproofness stands in conflict with a corresponding form of efficiency, and
- we find optimal bounds and an elegant, computer-generated proof for the fact that Condorcet-consistency and participation are incompatible.

While methodologically the core of our work concerns computer-aided proving of impossibility results and extracting human-readable proofs for these results, there are also three other ways, in which we have used computers to generate insights in social choice theory:

1. by applying different solving techniques, such as SAT, ASP, and LP, to generate (minimal) counterexamples for conjectures,
2. by providing solving-based algorithms for computational problems in social choice that deliver state-of-the-art performance while being quite flexible, and
3. by simply using computational tools to deal with practical concerns regarding preference aggregation and election methods.[9]

The remainder of Part I is structured as follows. In Chapter 2, we set up two basic models for social choice. We then, in Chapter 3, use these models to exemplify our main methodological contributions using different solving techniques. In Chapter 4, we summarize our findings and describe how the individual contributions are connected to each other. In addition, we give some details that had to be omitted from the original publications.

Part II carries the core of this thesis, the original publications. In Part III we conclude by briefly discussing the presented methods and by describing some ideas for future work.

8 To the best of our knowledge this is the first time that human-readable proofs have been constructed via SAT solving.

9 For instance, by providing an online tool for preference aggregation that is both powerful and easy to use.

2 MAIN SETTINGS AND MATHEMATICAL MODELS

This chapter describes two basic models of social choice, which we then formalize exemplarily as SAT/SMT instances in Chapter 3. Variants of these models are widely used and we also build upon and extend them in our publications as needed, for instance by variable agendas or electorates, or by allowing for ties in the individual preferences.

The two models differ by the class of outcomes of their respective aggregation mechanisms: while in set-valued social choice the outcomes are *sets* of best alternatives (for which then eventually ties have to be broken by other means), the outcomes in probabilistic social choice are lotteries (i.e., a probability distributions) over alternatives. The latter can not only be used to identify a single winner by means of randomization, but also allows for an interpretation as fractional allocations (e.g., for splitting time shares or monetary budgets).

2.1 FOUNDATIONS

Both models share some mathematical foundations, which we briefly define in this section. The following Sections 2.2 and 2.3 then introduce some of the specifics of each model.

Let A be a set of m *alternatives* and $N = \{1, \dots, n\}$ be a set of *agents*. Each agent $i \in N$ is equipped with a *preference relation* $\succsim_i$ over the alternatives, which is postulated to be *linear*.[10] The interpretation of $x \succsim_i y$ is that agent i values alternative x at least as much as alternative y. We write $\succ_i$ for the strict part of $\succsim_i$, i.e., $x \succ_i y$ if $x \succsim_i y$ but not $y \succsim_i x$. Note that, due to anti-symmetry, the only difference between $\succsim_i$ and $\succ_i$ is that $\succsim_i$ is reflexive while $\succ_i$ is not. The set of all preference relations is denoted by $\mathcal{R}$. As we will be concerned with social choice, i.e., collective decisions by the set of *all* agents, we define a *preference profile* to be an n-tuple $R = (\succsim_1, \dots, \succsim_n) \in \mathcal{R}^N$ of preference relations.

alternatives
agents
preference relation
preference profile

The central objects of consideration are aggregation functions $f : \mathcal{R}^N \to O$, which map preference profiles to outcomes from a set O. Depending on the setting, the set O differs. In the simplest model, *resolute social choice*, O simply is the set of all alternatives A and the aggregation functions (then also called *voting rules*) yield a single win-

resolute social choice
voting rules

10 A linear (sometimes called *strict*) preference relation is complete, transitive, and anti-symmetric. In some publications, we also consider *weak* preferences, which only satisfy the former two properties, i.e., they allow for ties between alternatives.

ner. Unfortunately, this setting is prone to sweeping impossibilities like the Gibbard-Satterthwaite theorem [Gibbard, 1973, Satterthwaite, 1975] and even basic fairness conditions such as anonymity and neutrality cannot be guaranteed to be compatible [cf. Moulin, 1983, Chapter 2]. Hence, one has to resort to more general sets of outcomes. In the following, we consider the cases of $O = 2^A$ (set-valued social choice), where 2^A stands for the set of all subsets of A, and $O = \Delta(A)$ (probabilistic social choice), where $\Delta(A)$ denotes the set of all lotteries over A, i.e., functions $p : A \to \mathbb{R}$ such that $p(x) \geqslant 0$ for all $x \in A$ and $\sum_{x \in A} p(x) = 1$.

lottery

Many axioms have been put forward to describe desirable properties of such aggregation functions. Resistance to strategic manipulation (so-called *strategyproofness*) and efficiency (e.g., in the form of *Pareto optimality*) are two such properties that have been particularly well-studied, and which will be presented in the following sections for set-valued and for probabilistic social choice.

Before we get to these specific properties, however, we briefly introduce some even more basic properties, which we require for reviewing our main results in Chapter 4. Anonymity and neutrality, for instance, are common fairness properties, which require that an aggregation function treats all agents (or all alternatives, respectively) equally. In formal terms, $f(R) = f(\sigma(R))$ for all permutations $\sigma \colon N \to N$ over agents (anonymity), and $f(R)(x) = f(\pi(R))(\pi(x))$ for all permutations $\pi \colon A \to A$ and $x \in A$ (neutrality).[11]

anonymity
neutrality

Many of our result will be valid for the restricted class of *majoritarian* aggregation functions, i.e., functions that are neutral and only depend on the *(pairwise) majority relation* R_M, which is defined as

(pairwise) majority relation

$$x \; R_M \; y \text{ if and only if } |\{i \in N \mid x \succsim_i y\} \geqslant |\{i \in N \mid y \succsim_i x\}.$$

In line with conventional notation, we denote by P_M the strict part of the majority relation R_M, i.e., $a \; P_M \; b$ if $a \; R_M \; b$ but not $b \; R_M \; a$.

An aggregation function f then formally is *majoritarian* if it is neutral and $f(R) = F(R')$ whenever $R_M = R'_M$. Note that the majority relation can be suitably represented as a graph $T = (A, R_M)$, which, in the case of an odd number of voters with strict preferences, is a *tournament*, and a *weak tournament* otherwise.[12]

majoritarian
tournament
weak tournament

Based on the majority relation, one also easily defines the notion of a *Condorcet winner*, an obvious choice for a "best" alternative, which, however need not exist: an alternative x is a *Condorcet winner* of a given preference profile R if $x \; P_M \; y$ for all $y \in A \setminus \{x\}$. An aggregation function that uniquely selects a Condorcet winner whenever it exists is called a *Condorcet extension*.

Condorcet winner
Condorcet extension

11 The permutations σ and π are extended to preference profiles in the usual way. For example, $\pi(R)$ is the preference profile obtained from π by replacing $\succsim_i$ with $\succsim_i^\pi$ for every $i \in N$, where $\pi(x) \succsim_i^\pi \pi(y)$ if and only if $x \succsim_i y$.

12 A tournament (graph) is a directed graph in which each pair of distinct nodes is connected by a single directed edge. In a weak tournament (graph) two directed edges in opposite directions are also allowed between any pair of nodes.

2.2 SET-VALUED SOCIAL CHOICE

Allowing for multiple tied winners, set-valued social choice separates the problem of breaking ties from the aggregation procedure. An aggregation function, then called a *social choice function* (SCF), maps preference profiles to non-empty sets of alternatives, the winners. The resolute case is embedded by SCFs that are single-valued for all preference profiles, i.e., $|f(R)| = 1$ for all $R \in \mathcal{R}^N$, in which case we also speak of *resolute* SCFs.

social choice function

Many axioms have been put forward to describe desirable properties of SCFs. Resistance to strategic manipulation (so-called *strategyproofness*) and efficiency (e.g., in the form of *Pareto optimality*) are two such properties that have been particularly well-studied. As three major publication of this thesis ([1], [4], and [5]) answer open questions regarding these two properties, and since, additionally, the case study in Section 3.1.1 is based on strategyproofness, we define these concepts here.

2.2.1 Strategyproofness

> "My scheme is intended only for honest men."
>
> J.-C. de Borda, 18th century

Rather than starting with the formal definition, let us consider an example of a preference profile R and assume a winner is to be determined by Borda's rule f_{Borda},[13] which has been named after J.-C. de Borda:

3	3	2
a	c	b
b	a	c
c	b	a
d	d	d

R

3	3	1	**1**
a	c	b	$\mathbf{c}$
b	a	c	$\mathbf{b}$
c	b	a	$\mathbf{d}$
d	d	d	$\mathbf{a}$

R'

In the example of preference profile R, the (single) winner is alternative a, i.e., $f_{\text{Borda}}(R) = \{a\}$ (scores are $a : 17, b : 15, c : 16, d : 0$). If, however, one of the agents i with truthful preferences $b \succ_i c \succ_i a \succ_i d$ misrepresents these preferences by expressing $c \succ_i b \succ_i d \succ_i a$, we arrive at the new profile R', for which $f_{\text{Borda}}(R') = \{c\}$ (scores are $a : 16, b : 14, c : 17, d : 1$)—an outcome preferred by the manipulating agent i.

13 Definition of Borda's rule: the alternatives receive points from each agent depending on their positions in the corresponding preference relations; for each agent the top-ranked alternative gets $m-1$ points, the next one $m-2$ points, and so on (down to 0 points for the bottom-ranked alternative). The alternatives with the highest accumulated score win.

Of course, we would like to avoid examples like this and have SCFs that are resistant to this kind of strategic manipulation. That, at least for the resolute case, this is not generally possible without making one agent a dictator, i.e., an agent who always gets his most preferred alternative, was proven independently by Gibbard [1973] and Satterthwaite [1975] in the famous Gibbard-Satterthwaite theorem. Moving to set-valued social choice, and thereby relaxing the assumption of resoluteness, one can get positive results. On the other hand, then the notion of strategyproofness is no longer unique because agents need to compare sets of winners rather than single alternatives.

dictatorship

Gibbard-Satterthwaite theorem

Such comparisons are usually achieved by defining a suitable *set extension* $\mathcal{E}$, which extends the individual preferences $\succsim_i$ to preferences $\succsim_i^{\mathcal{E}}$ over sets of alternatives. The definition of strategyproofness then becomes dependent on the extension $\mathcal{E}$:

set extension

Definition 2.1

$\mathcal{E}$-manipulable

Let $\mathcal{E}$ be a set extension. An SCF f is *$\mathcal{E}$-manipulable* by agent i if there exist preference profiles R and R' with $\succsim_j = \succsim_j'$ for all $j \neq i$ such that $f(R')$ is (strictly) $\mathcal{E}$-preferred to $f(R)$ by agent i, i.e.,

$$f(R') \succ_i^{\mathcal{E}} f(R).$$

$\mathcal{E}$-strategyproof

An SCF is called *$\mathcal{E}$-strategyproof* if it is not $\mathcal{E}$-manipulable.

Note the strict preference $f(R') \succ_i^{\mathcal{E}} f(R)$, which is equivalent to $f(R') \succsim_i^{\mathcal{E}} f(R)$ and not $f(R) \succsim_i^{\mathcal{E}} f(R')$.[14]

optimist extension

pessimist extension

Two simple set extensions are the *optimist* (O) and the *pessimist* (P) extension , in which agents compare sets simply by considering their *best* and *worst* elements, respectively. In formal terms, given individual preferences $\succsim_i$, a set of alternatives X is (weakly) optimist-preferred to another set Y (short: $X \succsim_i^O Y$) if there exists $x \in X$ such that $x \succsim y$ for all $y \in Y$. Analogously, X is (weakly) pessimist-preferred to Y (short: $X \succsim_i^P Y$) if there exists $y \in Y$ such that $x \succsim y$ for all $x \in X$.

As an example, consider the preferences $a \succ b \succ c$. We then, for instance, have $\{a, c\} \succ^O \{b\}$ and $\{b\} \succ^P \{a, c\}$.

While these two extensions play an important role in the case study in Section 3.1.1, our publications are mostly concerned with two other extensions that have particular natural interpretations and are due to Kelly [1977] and Fishburn [1972a] . For definitions and discussions of these additional extensions, the reader is referred to Publications [1] and [2].

Kelly extension

Fishburn extension

14 There are other notions of strategyproofness which require the truthful outcome to be weakly preferred to the manipulated one. For *complete* set extensions, this is equivalent to the presented notion. For *incomplete* set extensions, however, requiring a weak preference is quite demanding.

2.2.2 Pareto Optimality

Economic efficiency is captured in a rather mild form in the notion of Pareto optimality: if there is an alternative y which everyone prefers weakly, and at least one agent prefers strictly to alternative x (i.e., y Pareto dominates x), then x should not be selected.[15] If no such alternative exists, x is called Pareto optimal. Formally, we denote by $PO(R)$ the set of all alternatives that are Pareto optimal in R and define for set-valued social choice:

DEFINITION 2.2

An SCF f is *Pareto optimal* if it only selects Pareto optimal alternatives, i.e., if $f(R) \subseteq PO(R)$ for all preference profiles R.

Pareto optimality

In the examples R and R' above, for instance, alternative d is Pareto dominated by alternatives b and c. Hence, d must not be selected as a winner by any Pareto optimal SCF.

2.3 PROBABILISTIC SOCIAL CHOICE

While the roots of probabilistic principles for social choice date back as far as to ancient Athens [Headlam, 1933], these ideas have received surprisingly little attention after the first formal studies of probabilistic social choice int the 60s and 70s [see, e.g., Zeckhauser, 1969, Fishburn, 1972b, Intriligator, 1973, Gibbard, 1977]. Recently, however, probabilistic social choice is gaining interest again with a range of contributions both from social choice [see, e.g., Ehlers et al., 2002, Bogomolnaia et al., 2005, Chatterji et al., 2014, Brandl et al., 2016a] and political science [see, e.g., Goodwin, 2005, Dowlen, 2009, Stone, 2011]. One reason for its increasing popularity might be that it is *not* prone to many classical impossibility theorems, including the previously mentioned ones (Gibbard-Satterthwaite and Duggan-Schwartz).

In fact, Gibbard [1977] proved that *random dictatorships* are the only strategyproof and (ex post) efficient probabilistic mechanism. Although the name might suggest that these are undesirable aggregation functions, at least their uniform version (in which an agent is picked uniformly at random and gets to decide upon the outcome) is considered quite fair for many settings and is even frequently used in real-life, for instance when items need to be allocated among a set of agents.

random dictatorship

Formally, the aggregation functions for probabilistic social choice, so-called *social decision schemes* (SDS), are defined as functions $f : \mathcal{R}^N \to \Delta(A)$, where (as defined before) $\Delta(A)$ stands for the set of all lotteries over A.

social decision scheme

Properties such as strategyproofness and efficiency then depend upon the way how agents compare lotteries to one another given their

15 For strict individual preferences, this is equivalent to $y \succ_i x$ for all $i \in N$.

preferences over alternatives [for an extensive treatment see Aziz et al., 2014]. In Publication [4], we consider a natural way of extending individual preferences to lotteries based on *stochastic dominance (SD)*: given individual preferences $\succsim_i$, a lottery p is *SD*-preferred to another lottery q (short: $p \succsim_i^{SD} q$) if for every alternative x, lottery p is at least as likely as lottery q to yield an alternative at least as good as x. Formally,

lottery extension

stochastic dominance

$$p \succsim_i^{SD} q \text{ if and only if } \sum_{y \succsim_i x} p(y) \geqslant \sum_{y \succsim_i x} q(y) \text{ for all } x \in A.$$

As a shorthand, we write $\succ_i^{SD}$ for the strict part of $\succsim_i^{SD}$, i.e., $a \succ_i^{SD} b$ if $a \succsim_i^{SD} b$ but not $b \succsim_i^{SD} a$.

In contrast to the previously considered optimist and pessimist extensions, the *SD*-extension is not complete, i.e., there are lotteries that cannot be compared by this extension. Consider, for example, the preferences $a \succ b \succ c$, which imply

$$(1/2\, a + 1/2\, c) \succ_i^{SD} (1/4\, a + 1/4\, b + 1/2\, c),$$

but leave the lotteries $1/2\, a + 1/2\, c$ and b incomparable to each other (recall that the support of the former is O-preferred to the latter, and the support of the latter is P-preferred to the former).

The notion of $\mathcal{E}$-strategyproofness for set extensions $\mathcal{E}$ can be applied in the same fashion for any lottery extension $\mathcal{L}$. With $\mathcal{L} = SD$ this gives the usual definition of (weak) *SD*-strategyproofness.[16]

$\mathcal{L}$-strategyproofness

For efficiency one can extend the notion of Pareto optimality for arbitrary lottery extensions. An SDS f then is $\mathcal{L}$-efficient if it only returns $\mathcal{L}$-optimal lotteries, i.e., lotteries that are Pareto optimal with respect to $\mathcal{L}$. For $\mathcal{L} = SD$, this means that f may never return a lottery p for which we can find another lottery q with $q \succsim_i^{SD} p$ for all $i \in N$, and $q \succ_i^{SD} p$ for at least one $i \in N$.[17]

$\mathcal{L}$-efficiency

Interestingly, both these notions, (weak) *SD*-strategyproofness and *SD*-efficiency, naturally correspond to the setting when agents are equipped with utility functions (rather than just ordinal preferences): it can be shown that $p \succsim_i^{SD} q$ if and only if p yields at least as much expected utility as q according to *all* consistent utility representations of $\succsim_i$.

utility functions

16 *Strong* *SD*-strategyproofness would require a potential manipulator to always (weakly) prefer the truthful outcome to any manipulated outcome, cf. Footnote 14.

17 It appears to be a specific feature of probabilistic social choice that efficiency notions are defined based on preference extensions. In particular, in some results (see e.g., Publication [4]), both efficiency and strategyproofness are even based on the *same* extension. While one could theoretically also define efficiency notions based on *set* extensions in *set-valued* social choice (where usually only Pareto optimality is considered), such efficiency notions seem to be either too weak (e.g., Fishburn, Kelly), too strong (e.g., pessimist), or incomparable to Pareto optimality (e.g., optimist). For instance, our main result regarding Fishburn-strategyproofness, Theorem 4.1, cannot be weakened from Pareto optimality to Fishburn-efficiency since a majoritarian SCF known as the *top cycle* satisfies Fishburn-strategyproofness [Sanver and Zwicker, 2012] and Fishburn-efficiency [Aziz et al., 2015].

How problems from probabilistic social choice involving *SD*-strategyproofness and *SD*-efficiency can, despite the infinite space of lotteries, be encoded and solved via computer-aided methods (as SMT instances) is briefly touched upon in Sections 3.2 and 4.1; for a more detailled account, see Publication [4].

3 METHODOLOGY

In this chapter, we describe the basic ideas of the computer-aided methods we use in this thesis to solve problems in social choice theory. As our core results were obtained via an inductive approach coupled with SAT solving, we mostly concentrate on this specific approach and exemplify it for the celebrated Duggan-Schwartz theorem.[18] Afterwards we also briefly review the other solving techniques applied in this thesis.

3.1 COMPUTER-AIDED THEOREM PROVING VIA SAT SOLVING

As mentioned in the introduction, the general idea of proving theorems in social choice theory by reducing them to an extensive base case, which can then be solved on a computer, is attributed to Tang and Lin [2009]. We build upon this idea and extend it in two ways: we transfer it to the more complex realms of set-valued and probabilistic social choice, and we provide a method for extracting human-readable proofs. The settings of set-valued and probabilistic social choice require more advanced techniques for solving the base cases as otherwise the search space rapidly becomes too large. As an illustration, consider the sizes of the search spaces in Publications [1] and [2] given in Table 3.1, which clearly show that any attempt at the problem via exhaustive search would be doomed to fail.

Alternatives (m)	4	5	6	7
Publication [1] (odd n, strict pref.)	50,625	$\sim 10^{18}$	$\sim 10^{101}$	$\sim 10^{959}$
Publication [2] (arbitrary n, strict or weak pref.)	$\sim 10^{49}$	$\sim 10^{868}$	$\sim 10^{38,650}$	$\sim 10^{4,506,953}$

Table 3.1: Sizes of the search spaces of our problems, i.e., number of different majoritarian SCFs, depending on the number of alternatives and the settings considered.

Despite the more advanced encoding that we developed, the overall method of proving the results remains unchanged. An induction step,

18 This case study has not been published before and, in particular, is not contained in any of the publications that are part of this thesis.

which is usually easy to prove, reduces the desired theorems to *finite* instances. Depending on the number of variable parameters, more than one such reduction lemma may be required; for instance, one for the number of agents and another one for the number of alternatives.

Most work, however, needs to be invested into finding a suitable representation of this finite version of the theorem in some language that allows for automatic verification. One may ask why we chose the relatively restricted language of propositional logic in conjunction with SAT solvers for this task, even though many of the problems we consider involve statements that are naturally formulated in second-order logic. The answer is that there usually is a trade-off between expressivity of the input language and the ability of corresponding solvers to *automatically* solve problem instances. SAT solvers proved to be particularly effective for our problems from social choice theory. In contrast, experiments with formulations in—the much more natural choice of—higher-order logic unfortunately failed because the corresponding automated theorem provers did not terminate in any reasonable amount of time. These observations reflect similar findings by Grandi and Endriss [2013], who tried to verify existing impossibility results with theorem provers for first-order logic, and Lange et al. [2013], who verify Vickrey's theorem by providing extensive guidance to four interactive theorem provers.

The high-level architecture that we employed for proving base cases and constructing counterexamples is depicted in Figure 3.1. First, the setting and axioms under consideration are encoded into a suitable representation, e.g., conjunctive normal form (CNF) for the case of SAT solving. This procedure might require additional tools, such as an LP solver or the graph isomorphism program NAUTY [McKay and Piperno, 2013], for instance, to determine if a specific lottery is efficient or if two tournaments are isomorphic,[19] respectively. Then the encoded problem is fed to a SAT or SMT solver, whose output then gets translated back—based on the encoding chosen—into human-readable format (e.g., a tabular representation of an SCF that satisfies the given axioms).

The next subsection zooms in on the encoder and explains, at the example of the Duggan-Schwartz theorem, how encodings in propositional logic of statements that are naturally formulated in higher-order logic can be constructed. This case study illustrates in a simple way (and yet being completely disjoint from) the work included in Publications [1], [2], [3], and [4]. In particular, it does not require the advanced optimization techniques that we describe in those publications.

19 Isomorphic tournaments are important when encoding neutrality.

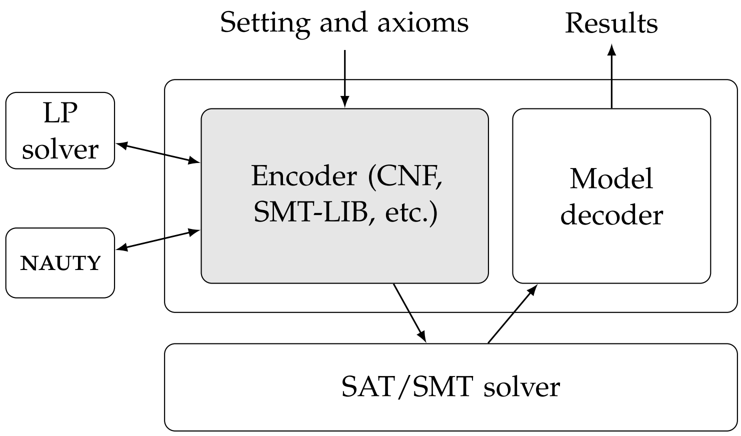


Figure 3.1: High-level system architecture as used to prove finite instances of theorems and to compute counterexamples

3.1.1 Case Study: the Duggan-Schwartz Theorem

The Duggan-Schwartz theorem [Duggan and Schwartz, 2000] is the most prominent generalization of the famous Gibbard-Satterthwaite theorem [Gibbard, 1973, Satterthwaite, 1975]. While the Gibbard-Satterthwaite theorem says that no *resolute* SCF can be both strategyproof and non-dictatorial, the Duggan-Schwartz theorem drops the restriction of resoluteness at the cost of introducing the strong notions of O- *and* P-strategyproofness as well as a strengthened version of non-dictatorship: the condition of not having a nominator. A *nominator* is an agent whose most preferred alternative is always selected as a winner (but not necessarily uniquely). In addition, for every alternative there must be a preference profile for which the alternative is selected as the unique winner—a very mild condition called *non-imposition*, which, for instance, is much weaker than Pareto optimality.

nominator

non-imposition

Formally, the statement of the Duggan-Schwartz theorem then is as follows [for more details and a direct proof, see e.g., Taylor, 2005].[20]

Duggan-Schwartz theorem

Theorem 3.1 *(Duggan and Schwartz, 2000)*
Let $m \geqslant 3$ and $n \geqslant 2$. Then any non-imposed SCF that is O- and P-strategyproof must have a nominator.

Reduction to a Finite Instance

In order to reduce the Duggan-Schwartz theorem to a finite instance, one can apply two inductive lemmas, one for the number of alternatives and one for the number of agents. In the case of the Duggan-Schwartz theorem, these lemmas are:

20 The theorem also holds for $n = 1$. Yet we state the theorem for $n \geqslant 2$ in favor of a more instructive exposition. Also note that an inductive lemma would likely require $n \geqslant 2$ [cf. Tang and Lin, 2009]. For $m \leqslant 2$, majority rule satisfies the axioms.

Lemma 3.2

Let $m \geqslant 3$ and $n \geqslant 2$. If f is a non-imposed SCF for $\mathbf{m+1}$ **alternatives** and n agents that is O- and P-strategyproof and has no nominator, then also for $\mathbf{m}$ **alternatives** and n agents there is an SCF f' that satisfies the same properties.

Lemma 3.3

Let $m \geqslant 3$ and $n \geqslant 2$. If f is a non-imposed SCF for m alternatives and $\mathbf{n+1}$ **agents** that is O- and P-strategyproof and has no nominator, then also for m alternatives and $\mathbf{n}$ **agents** there is an SCF f' that satisfies the same properties.

In many instances (e.g., Publications [1], [2], and [4]), such inductive lemmas are rather easy to proof. This is not the case for the Duggan-Schwartz theorem: while Theorem 3.1 itself, of course, implies Lemmas 3.2 and 3.3, we are not aware of a simpler proof. Since, for the induction on the number of alternatives m, only the condition of not having a nominator is problematic, and in order to still survey the basic proof technique here, we strengthen non-nominatorship to anonymity and present a proof for the corresponding, slightly weaker Lemma 3.4. Unfortunately, the induction on the number of agents n does—even with anonymity instead of non-nominatorship—not succeed with the standard techniques of cloning a fixed agent, cloning an agent based on his preferences, or adding a constant agent.

In the proof of Lemma 3.4 we occasionally apply an observation by Taylor [2005, p. 82] that any non-imposed and P-strategyproof SCF f may only choose from *top sets*. A *top set* $X_R \subset A$ of a preference profile R is a set of alternatives such that every agent prefers any $x \in X_R$ over any $y \notin X_R$. If a non-imposed and P-strategyproof SCF f chooses an alternative outside the top set, i.e., if there is $x \in f(R)$ with $x \notin X_R$, then we could construct a P-manipulation instance by iteratively, i.e., agent by agent, moving from R to a profile R' with $f(R') = \{y\}$ for some $y \in X_R$: the profile R' exists by non-imposition and at some stage the worst alternative for some manipulating agent would increase since $y \succ_i x$ for all $i \in N$.

top set

Lemma 3.4

Let $m \geqslant 3$ and $n \geqslant 2$. If f is a non-imposed, anonymous SCF for $\mathbf{m+1}$ **alternatives** and n agents that is O- and P-strategyproof, then also for $\mathbf{m}$ **alternatives** and n agents there is an SCF f' that satisfies the same properties.

Proof. Let f be a non-imposed, anonymous SCF for $m+1$ alternatives and n agents that is O- and P-strategyproof. We define f_e to be the restriction of f to m alternatives based on preference profiles in which alternative $e \in A$ is ranked last by all agents. Formally,

$$f_e(R) := f(R^{+e}),$$

where R^{+e} is the preference profile obtained from R by adding a new alternative e in the last position for all agents. This restriction of f is a well-defined SCF since alternative e is not contained in a top set and, hence, cannot be contained in $f(R^{+e})$.

We now need to show that for some alternative e the restriction f_e is a non-imposed, anonymous, O- and P-strategyproof SCF. Since this turns out to work for any $e \in A$, we just pick e arbitrarily.

- **Non-imposition:** Consider a preference profile R with all agents agreeing on some alternative x as their top choice. Then $\{x\}$ is a top set also in R^{+e} and hence $f_e(R) = f(R^{+e}) = \{x\}$.

- **Anonymity:** The fact that f_e is an anonymous SCF carries over from f directly. To see this, let R be a preference profile and $\pi : N \to N$ be a permutation of agents. Then
$$f_e(R) = f(R^{+e}) = f(\pi(R^{+e})) = f(\pi(R)^{+e}) = f_e(\pi(R))$$
since f is anonymous and since the operations of renaming agents and adding e at the bottom are independent of each other.

- **O- and P-strategyproofness:** Let $\mathcal{E}$ be the optimist (O) or pessimist (P) set extension.[21] Assume for a contradiction that f_e is not $\mathcal{E}$-strategyproof. Then there exist preference profiles R and R' and an agent $i \in N$ such that $f_e(R') \succ_i^{\mathcal{E}} f_e(R)$. But since $f_e(R') = f(R'^{+e})$ and $f_e(R) = f(R^{+e})$, we get
$$f(R'^{+e}) \succ_{i'}^{\mathcal{E}} f(R^{+e}),$$
where $\succ_{i'}$ stands for the preferences of agent i in R^{+e}. This contradicts $\mathcal{E}$-strategyproofness of f. □

By applying Lemmas 3.2 and 3.3, we can reduce the Duggan-Schwartz theorem to a finite instance, which then only concerns 3 alternatives and 2 agents.

LEMMA 3.5
Let $m = 3$ and $n = 2$. Then any non-imposed SCF that is O- and P-strategyproof must have a nominator.

Next, we encode this finite instance in propositional logic and prove it using a SAT solver.

Encoding of the Base Case

Our encoding of Lemma 3.5 is a basic version (and resembles parts) of what we used in Publications [1] and [2]. The idea is to encode an arbitrary SCF via Boolean variables $c_{R,X}$—one variable for each pair of a preference profile R and a non-empty set of alternatives X. Each

21 In fact, strategyproofness can be shown for a large class of set extensions $\mathcal{E}$ (see Publication [1] for details) rather than just the specific two considered here.

such variable $c_{R,X}$ then stands for $f(R) = X$, i.e., the choice set X being assigned to preference profile R.[22] In order to model Lemma 3.5, every (implicit and explicit) property of f now needs to be encoded as a constraint using only Boolean operators on these variables. After encoding these properties we then employ a SAT solver to verify that this combination of properties is unsatisfiable, proving the desired result, Lemma 3.5.

We have the following relevant properties:

- Functionality of the SCF f (Func),
- $\mathcal{E}$-strategyproofness (for $\mathcal{E} = O, P$) ($\mathcal{E}$-SP),
- Non-imposition (NImp), and
- Non-nominator (NNom).

Let us start with (Func), which has a straightforward encoding in propositional logic. The property says that for each preference profile R there is *exactly one* set X such that the variable $c_{R,X}$ is set to true.[23] Formally,

$$\begin{aligned} &(\forall R)\,((\exists X)\ c_{R,X} \wedge (\forall Y, Z)\ Y \neq Z \rightarrow \neg(c_{R,Y} \wedge c_{R,Z})) \\ &\equiv \bigwedge_R \left(\left(\bigvee_X c_{R,X} \right) \wedge \bigwedge_{Y \neq Z} (\neg c_{R,Y} \vee \neg c_{R,Z}) \right). \end{aligned} \tag{Func}$$

Note that we omit the domains of the quantifiers for the sake of brevity whenever they are clear from the notation or context.

Two simple tricks are applied to reach the purely propositional statement in (Func). First, all universal and existential quantifiers are replaceable by conjunctions and disjunctions, respectively, because of the finite universes they quantify over. Second, conditions which are not propositional, such as $Y \neq Z$, can be easily taken care of by a computer program when instantiating the constraints.

22 An encoding based on variables $c_{R,x}$ (with intended meaning $x \in f(R)$) would also be possible but is less efficient for the notion of strategyproofness.

23 There exist less direct, but more efficient encodings of the "at least one" and "at most one" properties [see, e.g., Hoelldobler and Nguyen, 2013]. In our encoding, however, the functionality of the SCF is never a performance critical axiom, which is why we stick to the simpler formulation.

The second of these tricks also plays a major role when it comes to $\mathcal{E}$-strategyproofness (for any set extension $\mathcal{E}$), one of the more complex, but still manageable conditions to be encoded:

$$\begin{aligned}
&(\forall R, i, \succsim) \neg \left(f(R^{i \mapsto \succsim}) \succ_i^{\mathcal{E}} f(R) \right) \\
&\equiv (\forall R, i, \succsim, X, Y) \left(\left(f(R^{i \mapsto \succsim} = X) \wedge f(R) = Y \right) \to \neg X \succ_i^{\mathcal{E}} Y \right) \\
&\equiv (\forall R, i, \succsim, X, Y) \left(X \succ_i^{\mathcal{E}} Y \to \neg \left(f(R^{i \mapsto \succsim}) = X \wedge f(R) = Y \right) \right) \\
&\equiv (\forall R, i, \succsim, X, Y) \left(X \succ_i^{\mathcal{E}} Y \to \left(f(R^{i \mapsto \succsim}) \neq X \vee f(R) \neq Y \right) \right) \\
&\equiv \bigwedge_R \bigwedge_i \bigwedge_{\succsim} \bigwedge_{\substack{X,Y \\ X \succ_i^{\mathcal{E}} Y}} \left(\neg c_{R^{i \mapsto \succsim}, X} \vee \neg c_{R,Y} \right), \qquad (\mathcal{E}\text{-SP})
\end{aligned}$$

where the symbol $R^{i \mapsto \succsim}$ stands for the preference profile obtained from R by replacing agent i's preferences with $\succsim$. This, just like $X \succ_i^{\mathcal{E}} Y$, is easily computable by a computer program while instantiating the formula.

The properties of non-imposition (NImp) and non-nominator (NNom) are more straightforward to encode again; the latter condition, however, requires transformation into conjunctive normal form (CNF),[24] which is efficiently achieved by replacing the term $\bigwedge_{X \ni top_{R,i}} \neg c_{R,X}$ by helper variables $h_{R,i}$ [Tseitin's encoding, see, e.g., Tseitin, 1983]. Note that $top_{R,i}$ is used to denote the top-ranked alternative according to agent i's preference relation (within the preference profile R).

$$\begin{aligned}
&(\forall x)(\exists R) f(R) = \{x\} \\
&\equiv \bigwedge_x \bigvee_R c_{R,\{x\}} \qquad (\text{NImp})
\end{aligned}$$

$$\begin{aligned}
&\neg (\exists i)(\forall R)\, top_{R,i} \in f(R) \\
&\equiv (\forall i)(\exists R)\, top_{R,i} \notin f(R) \\
&\equiv (\forall i)(\exists R)(\forall X) \left(top_{R,i} \in X \to f(R) \neq X \right) \\
&\equiv \bigwedge_i \bigvee_R \bigwedge_{X \ni top_{R,i}} \neg c_{R,X} \\
&\equiv \bigwedge_i \bigvee_R h_{R,i} \qquad (\text{NNom})
\end{aligned}$$

As indicated already, in the last step we replace the term $\bigwedge_{X \ni top_{R,i}} \neg c_{R,X}$ by helper variables $h_{R,i}$ in order to arrive at a for-

24 Practically all SAT solvers expect the input formula to be in CNF.

mula in CNF. This entails the following definitions that also have to be included in the encoding:

$$\begin{aligned}
&(\forall R, i)\left[h_{R,i} \longleftrightarrow \left(\bigwedge_{X \ni top_{R,i}} \neg c_{R,X}\right)\right] \\
\equiv\ &(\forall R, i)\left[\left(\neg h_{R,i} \vee \left(\bigwedge_{X \ni top_{R,i}} \neg c_{R,X}\right)\right)\right. \\
&\qquad\left. \wedge \left(h_{R,i} \vee \neg\left(\bigwedge_{X \ni top_{R,i}} \neg c_{R,X}\right)\right)\right] \\
\equiv\ &(\forall R, i)\left[\bigwedge_{X \ni top_{R,i}} (\neg h_{R,i} \vee \neg c_{R,X})\right. \\
&\qquad\left. \wedge \left(h_{R,i} \vee \bigvee_{X \ni top_{R,i}} c_{R,X}\right)\right].
\end{aligned} \tag{H}$$

Putting together the constraints (Func), ($\mathcal{E}$-SP) for $\mathcal{E} \in \{O, P\}$, (NImp), (NNom), and (H) for $m = 3$ and $n = 2$, we arrive at a propositional formula in CNF (with 324 variables and 13253 clauses) that current SAT solvers can detect to be unsatisfiable within less than one second.

3.2 FURTHER SOLVING TECHNIQUES APPLIED IN THIS THESIS

Apart from SAT, which takes the most prominent role in this thesis project, we also applied other solving methodologies to gain insights in social choice theory:

- linear programming (LP), applied as an auxiliary tool for identifying inefficient lotteries (Publication [4]) and for calculating the bipartisan set (Publications [1] and [2]),
- integer programming (IP), to show the non-existence of a preference profile satisfying specific conditions *regardless of the number of agents* (Publication [5]),
- answer set programming (ASP), as a more expressive alternative to SAT for questions regarding k-majority digraphs (Publication [6]), also applied as an auxiliary tool for finding completions of incompletely specified preference relations (Publication [5]), and

- satisfiability modulo theories (SMT), a powerful methodology enabling the treatment of the infinite domain of lotteries over alternatives (Publication [4]).

While linear and integer programming are well-known and established paradigms, in this section, we briefly introduce the concept of ASP solving and describe exemplarily how an axiom from *probabilistic* social choice can be encoded as an SMT instance.

Answer Set Programming (ASP)

ASP [see Gebser et al., 2012, for a comprehensive introduction] is a relatively new logic programming paradigm that offers a compact and expressive modelling language in conjunction with state-of-the-art solving performance. Indeed, in our experiments in Publication [6], we could hardly find a performance gap between current SAT and ASP solvers despite the richer modelling language.

There are two key differences between SAT and ASP solving: first, ASP works with a different logic, called *stable model semantics*, in which statements are only considered true if they are provable, and, second, ASP solving comprises an initial grounding step, in which first-order statements are automatically converted to propositional statements. While, for the user, the former is mostly a design choice, the latter—together with the availability of arithmetic operations and predicates—significantly reduces the effort required for modelling (i.e., encoding) a given problem. On the flipside, the somewhat unusual semantics might make ASP less accessible compared to classical techniques, such as SAT solving.

Satisfiability Modulo Theories (SMT)

SMT instances can be viewed as extensions of SAT instances, in which predicates from a given *theory* replace simple propositional, i.e., Boolean, variables. In other words, SMT instances are Boolean combinations of statements from a given theory. As an example, consider the SAT instance $p_1 \wedge (p_2 \vee p_3)$ in comparison to the SMT instance (for the theory of linear arithmetic) $x < 5 + y \wedge (y \leqslant -2 \vee x \geqslant 1)$.

There are SMT solvers for very different theories ranging from specific data types (such as arrays or bitvectors) to different versions of arithmetic (integer vs. real, linear vs. non-linear, etc). Similar to SAT, also for SMT there is an active community that runs annual competitions [Barrett et al., 2013]. In practice, SMT solvers are typically used as backends for verification tasks, such as the verification of software.

We apply SMT solvers in Publication [4] to model problems from probabilistic social choice. Doing the same via SAT appears hopeless because one would have to discretize the space of lotteries over alternatives. In the theory of (quantifier-free) *linear real arithmetic*, on the other hand, such lotteries are easily expressible. The corresponding (highly non-trivial) encoding is presented in Publication [4]; here we only display an SMT-formalization of a toy axiom saying that a lottery p should *not* be (weakly) *SD*-preferred to a lottery q by an agent with preferences $\succsim_i$, i.e., $p \not\succsim_i^{SD} q$.

For this purpose, we define variables p_x and q_x for all $x \in A$, and make sure these actually denote lotteries over A:

$$\sum_{x \in A} r_x = 1 \wedge \bigwedge_{x \in A} r_x \geqslant 0 \quad \text{for } r \in \{p, q\}.$$

Then the forbidden *SD*-preference can be encoded in a straightforward way as:

$$\begin{aligned}
& p \not\succsim_i^{SD} q \\
\equiv & \neg \left[(\forall x \in A) \sum_{y \succsim_i x} p(y) \geqslant \sum_{y \succsim_i x} q(y) \right] \\
\equiv & (\exists x \in A) \sum_{y \succsim_i x} p(y) < \sum_{y \succsim_i x} q(y) \\
\equiv & \bigvee_{x \in A} \sum_{y \succsim_i x} p_y < \sum_{y \succsim_i x} q_y .
\end{aligned}$$

If we instantiate this axiom for $A = \{a, b, c, d\}$ and the individual preferences $a \succ_i b \succ_i c \succ_i d$ we, for instance, get

$$\begin{aligned}
& \bigvee_{x \in A} \sum_{y \succsim_i x} p_y < \sum_{y \succsim_i x} q_y \\
\equiv & (p_a < q_a) \vee \\
& (p_a + p_b < q_a + q_b) \vee \\
& (p_a + p_b + p_c < q_a + q_b + q_c) ,
\end{aligned}$$

where we have omitted the two final sums as they are both trivially equal to 1.

Note that, in this form, the axiom cannot be modelled as a linear program since linear programs do not allow disjunctions. As an SMT axiom, however, it is perfectly well-formed.

Encodings of other axioms are similar in style, but require more evolved techniques and additional insights; for *SD*-strategyproofness and *SD*-efficiency, as well as neutrality and anonymity (which are encoded implicitly), these techniques and insights are presented in Publication [4].

4 SUMMARY OF PUBLICATIONS AND ADDITIONAL RESULTS

This chapter provides an overview of the results that we obtained during this thesis project. The corresponding original publications are then included in full in Part II.

For the purpose of summarizing the publications, we classify them into three categories based on their methodological contribution. This classification is reflected in the Sections 4.1 to 4.3, in each of which we briefly review the respective articles and describe how they are connected. The methodological categories are as follows.

First and most importantly, in Section 4.1 we inductively prove novel results about the incompatibility of seemingly basic conditions such as strategyproofness and economic efficiency. As explained earlier, the main components of the respective proofs are situated in the respective induction bases and are established by means of automated solving techniques (SAT and SMT). We also describe in detail how *human-readable* (or at least human-verifiable) proofs of these computer-aided results can be extracted using our approach.

Second, it turns out that solving-based algorithms (in particular, based on SAT, ASP, and LP) perform well for computational tasks in social choice. Besides being interesting in their own right, these algorithms can also be applied not only to generate counterexamples but also to prove the correctness and minimality of these counterexamples. Corresponding advances and results are described in Section 4.2.

And, finally, as a third category, we explain in Section 4.3 how computers can assist for some practical concerns of social choice.

4.1 COMPUTER-AIDED THEOREM PROVING

With Publications [1], [2], [3], and [4], extensions to the computer-aided approach to theorem proving by Tang and Lin [2009] (which has also been exemplified in Section 3.1) lie at the core of this thesis. The extensions are increasingly sophisticated from paper to paper in order to capture more and more complex settings and axioms. In particular, the setting progresses from set-valued social choice ([1], [2], [3]) to probabilistic social choice ([4]).

Finding Strategyproof Social Choice Functions via SAT Solving

Publication [1] and its encoding (which captures notions of strategyproofness for *set-valued* social choice) are to be seen as the first extension to the general approach. Thus, the article also offers the most extensive treatment of the method itself and how it extends previous work. The novel technique of proof extraction is described in most detail here, too. Publications [2], [3], and [4] then subsequently build on the methodological insights from this paper.

Publication [1] offers a range of the results regarding the popular notions of Kelly- and Fishburn-strategyproofness (cf. Section 2.2.1), which are all obtained by applying the computer-aided method to differing axioms or settings, thereby showcasing the universality and flexibility of the approach. The publication's main result, however, clearly stands out:

Fishburn-strategyproofness

THEOREM 4.1 *(Brandt and G., 2016)*
For any number of alternatives $m \geqslant 5$ there is no majoritarian SCF that satisfies Fishburn-strategyproofness and Pareto optimality.

This theorem confirms the suspicion that Fishburn-strategyproofness may only be satisfied by rather indiscriminating SCFs such as the top cycle [Feldman, 1979, Brandt and Brill, 2011, Sanver and Zwicker, 2012] and can be shown to even hold without the assumption of neutrality.[25]

pairwise

It was noticed by Peters [2016] that Theorem 4.1 can even be extended to *pairwise* SCFs,[26] based upon the computer-aided results contained in Publication [1].[27]

The second key result of Publication [1] demonstrates that also possibility results can be achieved with the computer-aided method. While previous contributions [see, e.g., Brandt, 2015] suggested that the bipartisan set (*BP*), an attractive majoritarian SCF, might be the finest majoritarian SCF satisfying Kelly-strategyproofness, we find an even finer such SCF.[28] This SCF, however, fails to satisfy other natural conditions (e.g., composition-consistency), which is why we

25 Every majoritarian SCF has to be neutral according to our definitions.

26 An SCF is *pairwise* if its outcomes only depend upon the pairwise majority margins of the preference profiles. In other words, the SCF may depend upon the *weighted* majority relations.

27 In Remark 1 of Publication [1] we state that Theorem 4.1 also holds for a weaker variant of Fishburn-strategyproofness, in which the manipulator is only allowed to swap two adjacent alternatives. This observation stands in conflict with the existence of a pairwise, Fishburn-strategyproof, and Pareto optimal SCF (since otherwise we could define a *majoritarian* SCF that satisfies Pareto optimality and this weaker variant of Fishburn-strategyproofness based upon the weighted tournaments with only weights 1).

28 An SCF f is *finer* than (or, a *refinement* of) another SCF g if $f(R) \subseteq g(R)$ for all preference profiles R.

phrase the result negatively and view it mostly as an insight into the notion of Kelly-strategyproofness.

THEOREM 4.2 *(Brandt and G., 2016)*
> There exists a majoritarian Condorcet extension that refines *BP* and is still Kelly-strategyproof. As a consequence, *BP* is not even *a* finest majoritarian Condorcet extension satisfying Kelly-strategyproofness.

Kelly-strategyproofness

We furthermore analyze whether *BP* can be characterized by composition-consistency in conjunction with further natural properties. While we are not able to resolve this question completely, we are able to provide a set of insightful results on finite domains that might guide the search for such a characterization in the future.

For these additional results, we required an encoding of the property of composition-consistency, which relates outcomes for tournaments of *different* sizes (which is not possible in our original formalization). A technique for encoding this property had to be omitted from the paper and is presented in Section 4.4.1. In addition, we also conduct a study of the discriminatory power of Kelly-strategyproof SCFs and report on our findings before we turn to the novel technique of proof extraction.

Proof extraction enables us to construct a human-readable proof from a certificate of unsatisfiability, which is generated automatically by the SAT solver. In more detail, the SAT solver extracts a minimal unsatisfiable set (MUS) of clauses, i.e., a set of clauses that is unsatisfiable but any of its subsets is satisfiable. Given our encoding, such an MUS represents a list of "ingredients" for the proof of the result under consideration, which can then serve as the basis for a human-readable proof. We find a small MUS with only 16 clauses and decode it (with machine support) into a human-readable proof for the main result that occupies approximately two pages and, furthermore, carries information on the (so far implicit) number of agents required.

proof extraction

minimal unsatisfiable set

Strategic Abstention Based on Preference Extensions: Positive Results and Computer-generated Impossibilities

Publication [2] and [3] build upon the successful encoding and techniques from Publication [1] and extend it to the notion of participation. An SCF f satisfies *participation*, if no agent can benefit from abstaining the election, or, formally, if there is no preference profile R and agent i such that

participation

$$f(R_{-i}) \succ_i^{\mathcal{E}} f(R),$$

where R_{-i} stands for the preference profile obtained by removing agent i from R. Note that this also requires enriching the mathematical model for SCFs by *variable electorates*.

In practice, extending the framework to participation for *majoritarian* SCFs (Publication [2]) means to also allow for majority relations *with ties* (as one has to deal with variable electorates and, thus, cannot assume an odd number of voters anymore). In conjunction with allowing *weak* individual preferences, this seemingly innocent extension leads to a massive blow-up of the corresponding search spaces as Table 4.1 shows.

$\|A\|$	Publication [1]	Publication [2]
3	49	823,543
4	50,625	$\sim 2.5 \cdot 10^{49}$
5	$\sim 7.9 \cdot 10^{17}$	$\sim 9.4 \cdot 10^{867}$
6	$\sim 5.8 \cdot 10^{100}$	$\sim 6.8 \cdot 10^{38649}$

Table 4.1: Number of different majoritarian SCFs. While in Publication [1] we could assume an odd number of agents with strict preferences, the concept of participation requires us to deal with variable electorates, and therefore weak majority relations, in Publication [2].

Yet, by again using canonical representations of the majority relations (with respect to neutrality) and by proving lemmas which are similar to the ones in Publication [1], we are able to obtain the following main result.[29] It is presented in the paper together with a human-readable proof, which was extracted with the tools from Publication [1].

Fishburn-participation

THEOREM 4.3 *(Brandl, Brandt, G., and Hofbauer, 2015)*
There is no majoritarian and Pareto optimal SCF that satisfies Fishburn-participation if $|A| \geqslant 4$.

By simple adjustments one also gets that this impossibility still holds for strict preferences, but then requires at least five alternatives. A proof for this additional result could also theoretically be extracted from an MUS, which, however, is a tedious task given that it contains 124 instances of abstaining agents (compared to 10 such instances for Theorem 4.3). Furthermore, not many additional insights are to be expected from this proof by (massive) case distinction.

Optimal Bounds for the No-Show Paradox via SAT Solving

no-show paradox

In contrast to the previous publication, the computer-generated proofs in Publication [3] actually provide new insights into the problem: they exhibit a symmetric structure, which had not been exploited in any of the similar manual proofs before. The theorems in this publication are multiple stronger versions of the famous *no-show paradox* by Moulin [1988]: for each Condorcet extension there is

29 Further results include insights regarding the cardinality of choice sets returned by SCFs that satisfy participation. These are reflected in Section 5.2 of Publication [2].

a situation in which an agent is better off by abstaining the election if there are at least 4 alternatives and 25 agents.

The high number of agents required for this statement clearly stands out, especially when compared to similar results. One might even suspect that there are Condorcet extensions that satisfy participation for any purpose involving only a small number of agents. We answer these questions by providing tight bounds not only for the resolute case, but also for set-valued and probabilistic social choice. The main results for resolute rules, which were all obtained via SAT solving are:

Theorem 4.4 *(Brandt, G., and Peters, 2016)*
There is no Condorcet extension that satisfies participation for $m \geqslant 4$ and $n \geqslant 12$.

Theorem 4.5 *(Brandt, G., and Peters, 2016)*
There is a Condorcet extension f that satisfies participation for $m = 4$ and $n \leqslant 11$. Moreover, f is pairwise, Pareto optimal, and a refinement of the top cycle.

Methodologically, the main advances of this paper are the transfer to pairwise voting rules (i.e., rules that only depend on the anonymized comparisons between pairs of alternatives) and an incremental proving technique, the latter of which we gratefully attribute to D. Peters. This incremental technique uses insights from computer-generated proofs of *weaker* statements to heuristically search for a proof of the desired statement, which would not be tractable directly because of the immensely large search space.

Proving the Incompatibility of Efficiency and Strategyproofness via SMT Solving

The induction-based computer-aided proving methods of this thesis culminate in Publication [4] where many of the previously developed techniques and an advanced encoding in the language of SMT (with linear real arithmetic as its theory) are applied to prove a conjecture by Aziz et al. [2013]. The conjecture forms a common generalization of important known results from social choice [Aziz et al., 2013, 2014, Brandl et al., 2016b] and strengthens statements that were shown within the domain of assignment [Zhou, 1990, Bogomolnaia and Moulin, 2001, Katta and Sethuraman, 2006]. The result is stated in Theorem 4.6; for formal definitions of these notions of strategyproofness and efficiency, which are based on utility functions, the reader is referred to Publication [4]. We will, however, see that they are equivalent to *SD*-strategyproofness and -efficiency, respectively.

satisfiability modulo theories

THEOREM 4.6 *(Brandl, Brandt, and G., 2016)*
If $m \geqslant 4$ and $n \geqslant 4$, there is no anonymous and neutral social decision scheme (SDS) that satisfies efficiency and strategyproofness.

encoding infinite domains

Fortunately, the inductive lemma to reduce Theorem 4.6 to its base case of $m = n = 4$ has a straightforward proof. However, compared to previous contributions, two main challenges of infinite domains have to be overcome when modelling the base case:

1. preferences are modelled via utility functions and
2. the outcomes of the aggregation procedure are lotteries,

both of which do, when interpreted naïvely, not admit a finite representation.

The first of these challenges is solved analytically whereas the latter is treated by technical means. Utility functions (1.) are handled by resorting to the concept of *stochastic dominance*, which allows representing both, strategyproofness and efficiency, by only considering ordinal preferences (i.e., preference relations) rather than full utility representations. This leads to the following simplified version of Theorem 4.6:

THEOREM 4.7 *(Brandl, Brandt, and G., 2016)*
If $m \geqslant 4$ and $n \geqslant 4$, there is no anonymous and neutral social decision scheme (SDS) that satisfies *SD*-efficiency and *SD*-strategyproofness.

SD-efficiency
SD-strategyproofness

The challenge of representing lotteries (2.), on the other hand, can be transferred to the SMT solver. As described in some more detail in Section 3.2, it is an advantage of SMT over SAT that statements from an underlying theory (here: the theory of quantifier-free linear real arithmetic) can take the place of purely boolean variables. This enables modelling an unknown SDS by m real variables for each preference profile R (variable $p_{R,x}$ represents the probability assigned to alternative x at profile R).

Other (less significant) methodological challenges include

- representing anonymity and neutrality without quantifying over all permutations,
- encoding efficiency without quantifying over all lotteries,
- finding a domain large enough for the impossibility but small enough to be solvable, and
- gaining confidence in the result through a human-verifiable encoding.

Resolving these challenges again involves a mixture of computer-aided and manual effort, details of which are to be found in Publication [4].

In addition to what is stated in the paper, Eberl [2016] was able to completely verify the results of Publication [4] within the ISABELLE proof assistant and even produced (in a semi-automatical way) a very complex, but in principle human-readable proof of the main result.

4.2 SOLVING-BASED ALGORITHMS

In Publications [6] and [7] we develop algorithms for computational problems in social choice theory and use them to improve our understanding of the notions of k-majority digraphs and Condorcet winning sets, respectively. These algorithms have in common that they are based on solving methodologies (here: SAT and ASP), use corresponding solvers as backends, and compute preference profiles that satisfy certain properties. The main idea is simple: encode an unknown preference profile together with desirable properties as a satisfiability problem and let the SAT solver compute whether such a profile exists.[30] Interestingly, these solving-based algorithms outperform existing tailor-made approaches (to the specific questions considered).

In Publication [5] we apply these and similar techniques to gain insights (mostly by means of computing counterexamples) into the connection between the McKelvey uncovered set and the notion of Pareto optimality. Further applications are:

- an experimental analysis of how many agents it takes to obtain the majority relations of real-world and generated preference profiles (Publication [6]), and
- finding a minimal preference profile with Condorcet dimension 3 (Publication [7]).

A Note on the McKelvey Uncovered Set and Pareto Optimality

Even though the main theorem of Publication [5] is obtained manually, computer-aided methods (which are backed by solving methodologies) establish the boundaries of this result. To this end, we consider the two most natural extensions of the main theorem and show, by computing suitable counterexamples, that neither of them holds.

Let us first review the main result before we turn to its potential extensions. If we assume that only the pairwise majority relation of a preference profile R is known, one may pose the question of which

30 In some sense this is similar to the base case analysis in Chapter 4.1, where we encoded aggregation functions rather than preference profiles.

alternatives are guaranteed to be Pareto optimal (cf. Definition 2.2). Theorem 4.8 implies that these are exactly the alternatives in the *McKelvey uncovered set* $UC(R)$ [Bordes, 1983, McKelvey, 1986].[31,32] In addition, Theorem 4.8 establishes the existence of a preference profile in which precisely the alternatives of the McKelvey uncovered set are Pareto optimal. Formally, the statement is as follows (recall that $PO(R)$ denotes the set of all alternatives that are Pareto optimal in R).

McKelvey uncovered set

THEOREM 4.8 *(Brandt, G., and Harrenstein, 2016)*
For every preference profile R, there is another preference profile R' with the same majority relation as R such that

$$UC(R) = PO(R').$$

As a consequence, the McKelvey uncovered set can be characterized as the coarsest majoritarian SCF that satisfies Pareto optimality:

COROLLARY 4.9 *(Brandt, G., and Harrenstein, 2016)*
A majoritarian SCF f is Pareto optimal iff it is a refinement of the McKelvey uncovered set (i.e., $f(R) \subseteq UC(R)$ for all preference profiles R).

Note that the proof of Theorem 4.8 crucially relies on the assumption of variable electorates (i.e., R might have a different number of agents compared to R'). Using computer-aided methods, which are based on the algorithms from Publication [6], we show that this is not simply a deficiency of the proof, but actually cannot be avoided:

variable electorates

PROPOSITION 4.10 *(Brandt, G., and Harrenstein, 2016)*
There is a preference profile R such that

$$UC(R) \neq PO(R')$$

for all preference profiles R' with the same number of agents and the same majority relation as R.

We obtain the corresponding counterexample by exhaustively iterating through tournaments (i.e., majority relations for odd numbers of agents) and checking for each covering edge whether it can be turned into a Pareto edge with the same number of agents. Hence, the provided counterexample actually establishes the stronger statement that turning even just an individual covering edge into a Pareto edge might imply a change in the number of agents. This implies

31 The McKelvey uncovered set is a generalization to *weak* tournaments of the standard uncovered set. Among the commonly described generalizations it appears to be the most appealing one [see, e.g., Dutta and Laslier, 1999, Brandt and Fischer, 2008]. For an overview of the theory of covering relations and uncovered sets, see the comprehensive survey paper by Duggan [2013].

32 Note that the uncovered set is a concept based only on the majority relation, whereas Pareto optimality requires information from the underlying preference profile.

that, for arbitrary *fixed* sizes of the electorate, the McKelvey uncovered set *cannot* be characterized via Pareto optimality (as the set of necessarily Pareto optimal alternatives).

A further natural extension of Theorem 4.8 would be to show equivalence of the Pareto *relation* and the McKelvey covering *relation* (rather than just the corresponding sets of alternatives). This time, we apply *answer set programming* (ASP) to a specific tournament in order to generate an *integer program* (IP), which then establishes the following proposition.

Proposition 4.11 *(Brandt, G., and Harrenstein, 2016)*
There is a preference profile R such that, for no preference profile R' with the same majority relation as R, the Pareto relation of R' agrees with the covering relation of R.

For the proof of Proposition 4.11 an IP is needed rather than the SAT-based approach from Publication [6], since with latter approach one can only show the non-existence of suitable preference profiles for *fixed* (small) numbers of agents whereas with an IP one can ask for a solution with an *arbitrary* number of agents.[33]

Identifying k-majority Digraphs via SAT Solving

The techniques developed in Publication [6] form the basis of similar algorithms which we apply in the other publications of this section.

The research problem is to develop a solving-based algorithm that determines, given a (weak) tournament T and an integer k, whether a preference profile R exists which contains exactly k agents and which has T as its majority relation. If such a profile exists, T is also referred to as a k-*majority digraph* and we say that T is inducible by k agents. k-*majority digraph*

In Publication [6], we present such algorithms backed by SAT and ASP solving, and apply them to improve our understanding of k-majority digraphs in exhaustive, empirical and stochastic experiments. We were surprised by the significant performance improvement over traditional approaches (for instance, based upon a characterization of 3-majority digraphs by Brandt et al. [2013]). As an example for $k = 3$, the SAT-based algorithm solves many randomly sampled instances with up to 100 alternatives (corresponding to a search space of roughly 10^{473} anonymous preference profiles) within seconds, whereas the traditional approach requires 20 minutes for 8 alternatives already. k-*inducibility*

Our two main findings regarding k-majority digraphs are:

33 Even though it is not a problem here, in general one has to be careful about the numerical stability of solutions when dealing with IP. Many SMT solvers are safe from these potential issues (as they use exact arithmetic at the cost of lower computational efficiency).

- All tournaments with up to 7 nodes are inducible by 3 agents (also shown independently by Eggermont et al. [2013] and confirming a conjecture by Shepardson and Tovey [2009]).

- All of the millions of tournaments we randomly sampled (by means of different commonly used preference models) are inducible by 5 agents (and 8 agents for the case of weak tournaments). This is surprising, given a theoretical result by Stearns [1959], which guarantees the existence of tournaments which *cannot* be induced by k agents, for any $k \in \mathbb{N}$.

While the encoding presented in this publication was optimized for performance, there are other, more flexible (and slightly less well-performing) encodings, one of which we lay out in Section 4.4.2. Generally, however, even more performant encodings can easily be extended by additional constraints. This includes forcing a particular set of edges to be Pareto edges, or ensuring that no small Condorcet winning sets exist, which we made use of in Publication [5] and Publication [7], respectively.

Finding preference profiles of Condorcet dimension k via SAT

Condorcet winning set

Elkind et al. [2011] introduce the notion of Condorcet winning sets of a preference profile R as a set-valued generalization of the concept of a Condorcet winner. For each alternative x outside a *Condorcet winning set* C, there is a majority of agents such that each member of this majority finds at least one alternative within C more desirable than x.

Condorcet dimension

Elkind et al. then also define the notion of the *Condorcet dimension* $\dim_C(R)$ of a preference profile R as the size of the smallest Condorcet winning set it admits. Profiles with a Condorcet winner hence have Condorcet dimension 1 and high dimensions appear to indicate very diverse (and hence difficult to aggregate) preferences.

The short Publication [7] serves the purpose of providing a minimal example of a preference profile R which has Condorcet dimension $\dim_C(R) = 3$. We find that such a profile contains 6 agents and alternatives, which improves previously known examples in both the number of agents and the number of alternatives.[34]

Methodologically, the publication can be viewed as an extension to Publication [6] in that a very similar approach leads to success. The difference lies in an additional (complex) constraint for the non-existence of small Condorcet winning sets, which then replaces the constraint for majority relations. Despite the efficient approach, it is

34 The profile provided by Elkind et al. [2011] consists of 15 agents and alternatives. Other examples were of sizes $(7, 21)$, $(8, 13)$, $(11, 11)$, and $(12, 12)$ (agents, alternatives).

still open whether a preference profile with Condorcet dimension 4 exists.

4.3 PRACTICAL CONTRIBUTIONS

In contrast to the other publications in this thesis, Publications [8] and [9] do *not* make extensive use of solving methodologies.[35] Yet, they heavily rely on computers to solve *practical* concerns of social choice. While Publication [8] is a theoretical contribution to a practical problem (evaluating the probabilities with which certain voting paradoxes occur), Publication [9] has a purely applied focus and provides an online tool for practical preference aggregation among human agents.

Analyzing the Practical Relevance of Voting Paradoxes via Ehrhart Theory, Computer Simulations, and Empirical Data

The axiomatic method of social choice theory judges the quality of voting rules based on whether these rules satisfy certain desirable properties, so-called *axioms*. Failure to satisfy a natural axiom, such as *not selecting the Condorcet loser*, is often also referred to as a *voting paradox*. Classical social choice then does not distinguish whether a paradox occurs at a *single* preference profile or at many (or even all) preference profiles.

In Publication [8] we take a much more practical perspective and analyze *how often* two common paradoxes, the Condorcet loser paradox (CLP) and the agenda contraction paradox (ACP), occur for certain voting rules and under common assumptions about the distribution of preferences. While for some specific settings tailor-made approaches exist and can be carried out purely on analytical grounds, computers can contribute significantly to answering these types of questions flexibly, quickly and with high precision. Interestingly, even the somewhat universal, analytical approach via *Ehrhart theory*, which yields *exact* probabilities, requires expensive computations that would not be feasible without computer support or powerful software.

Condorcet loser paradox

agenda contraction paradox

Ehrhart theory

Our main findings are:

1. Despite being viewed as a major flaw of some Condorcet extensions, the CLP only occurs with negligible probabilities and hence is of no practical relevance.

2. The ACP, on the other hand, frequently occurs under various distributional assumptions about the agents' preferences. The extent to which it is real threat, however, strongly depends on

[35] Applications of solvers are limited to the computation of choice rules (Young's, Dodgson's, and Kemeny's rule via an IP, and maximal lotteries via an LP).

the voting rule, the underlying distribution of preferences, and, somewhat surprisingly, the parity of the number of agents.

Pnyx: a Powerful and User-friendly Tool for Preference Aggregation

The web-based tool Pnyx (pnyx.dss.in.tum.de) is the result of an effort to provide the advanced methods studied in social choice to a broader audience.[36] Under the hood, the tool also makes use of linear and integer programming for the computation of maximal lotteries and Kemeny's rule, respectively. These techniques, however, are rather straightforward and well-known.

Thus, the main focus of the project, which had Publication [9] as a side result, was to provide suitable aggregation rules to users regardless of their knowledge about such rules. The user only has to define input and output formats, which are then automatically mapped to a suitable aggregation function. In addition, the tool supports the whole process from setting up a poll, to eliciting and aggregating preferences, to the communication of the results to agents. A screencast explaining the core functionalities is provided at vimeo.com/118576213.

4.4 DETAILS OMITTED FROM ORIGINAL PUBLICATIONS

This section describes some results that had to be omitted from the corresponding original publications (mostly due to space constraints).

4.4.1 Encoding of Composition-consistency

In Publication [1] we list a few results in Section 4.1 that involve the property of composition-consistency, an invariance condition with respect to cloning of alternatives. Since this property requires us to reason about tournaments of different sizes—a non-trivial extension of the encoding presented in Publication [1]—we present here how this property can be encoded in our framework.

component

A *component* C of a tournament $T = (A, R_M)$ is a subset of the alternatives in which all alternatives are indistinguishable by their relationship to outside alternatives, i.e., for all $x \in C$ we either have $\{x\}\ R_M\ A \setminus C$ or $A \setminus C\ R_M\ \{x\}$, where $X\ R_M\ Y$ denotes $x\ R_M\ y$ for all $x \in X$, $y \in Y$. A set of pairwise disjoint components $\{C_1, \dots, C_k\}$

36 The development of user-friendly tools for social choice that enable user to run advanced methods has, with the exception of spliddit.org, received surprisingly little attention so far.

such that $A = \bigcup_{i=1}^{k} C_i$ is then called a *decomposition* of T, and each decomposition implicitly defines a *summary* $\widetilde{T} := (\{1, \ldots, k\}, \widetilde{R_M})$ by

decomposition
summary

$$i \; \widetilde{R_M} \; j \text{ if and only if } C_i \; R_M \; C_j.$$

In slight abuse of notation, we sometimes let the symbol $\widetilde{T}$ refer to the set $\{1, \ldots, k\}$ only.

The property of *composition-consistency* [Laffond et al., 1996] postulates that whatever a majoritarian SCF f selects on smaller tournaments must be preserved by $f(T)$ with respect to any decomposition $\{C_1, \ldots, C_k\}$ of T with summary $\widetilde{T}$, i.e.,

composition-consistency

$$f(T) = \bigcup_{i \in f(\widetilde{T})} f(T|_{C_i}). \tag{CC}$$

An encoding of composition-consistency can thus be defined by instantiating the statement (CC) for any decomposition $\{C_1, \ldots, C_k\}$ of T with summary $\widetilde{T}$. A major difference of composition-consistency compared to previously implemented axioms is that it links tournaments of different sizes, whereas previous axioms only had implications for tournaments of a fixed size. Thus, in order to prove a statement for $|A| = m$ one has to instantiate all axioms for $1, \ldots, m$ alternatives. Because of the exponential growth of the involved objects, fortunately, this does not increase the running time by much, compared to working with m alternatives only.

Note that $\{C_1, \ldots, C_k\}$ forms a partition of A and hence we can encode (CC) as

$$\begin{aligned}
& f(T) = \bigcup_{i \in f(\widetilde{T})} f(T|_{C_i}) \\
\equiv & \left(\forall i \in \widetilde{T}\right) (\forall x \in C_i) \left[x \in f(T) \longleftrightarrow \left(i \in f(\widetilde{T}) \wedge x \in f(T|_{C_i})\right)\right] \\
\equiv & \bigwedge_{i \in \widetilde{T}} \bigwedge_{x \in C_i} \left[m_{T,x} \to \left(m_{\widetilde{T},i} \wedge m_{T|_{C_i},x}\right) \wedge \right. \\
& \qquad\qquad \left. \left(m_{\widetilde{T},i} \wedge m_{T|_{C_i},x}\right) \to m_{T,x}\right] \\
\equiv & \bigwedge_{i \in \widetilde{T}} \bigwedge_{x \in C_i} \left[\left(\neg m_{T,x} \vee m_{\widetilde{T},i}\right) \wedge \left(\neg m_{T,x} \vee m_{T|_{C_i},x}\right) \wedge \right. \\
& \qquad\qquad \left. \neg m_{\widetilde{T},i} \vee \neg m_{T|_{C_i},x} \vee m_{T,x}\right],
\end{aligned}$$

where $m_{T,x}$ are variable symbols representing $x \in f(T)$. These can be defined in a straightforward manner via the existing variable symbols

$c_{T,X}$ (which stand for $f(T) = X$) by setting, for each tournament T and non-empty set X:

$$\begin{aligned}
c_{T,X} \longleftrightarrow &\left(\bigwedge_{x\in X} m_{T,x} \wedge \bigwedge_{y\notin X} \neg m_{T,y}\right)\\
\equiv &\left[\neg c_{T,X} \vee \left(\bigwedge_{x\in X} m_{T,x} \wedge \bigwedge_{y\notin X} \neg m_{T,y}\right)\right] \wedge\\
&\left[c_{T,X} \vee \left(\bigvee_{x\in X} \neg m_{T,x} \vee \bigvee_{y\notin X} m_{T,y}\right)\right]\\
\equiv &\left[\bigwedge_{x\in X} (\neg c_{T,X} \vee m_{T,x}) \wedge \bigwedge_{y\notin X} (\neg c_{T,X} \vee \neg m_{T,y})\right] \wedge\\
&\left[c_{T,X} \vee \bigvee_{x\in X} \neg m_{T,x} \vee \bigvee_{y\notin X} m_{T,y}\right].
\end{aligned}$$

As a practical comment, one may note that, by over-approximation, it suffices for an impossibility if partially instantiated axioms already lead to an unsatisfiable formula (cf. Publication [1], Section 3.2.3). We make use of this fact in Publication [1] and only instantiate composition-consistency for specific decompositions rather than for any decomposition (of which there can be many). All reported results already hold when only considering the decomposition as represented by the first level of the tournament's unique *decomposition tree* [for a definition see, e.g., Brandt et al., 2011, Section 3].

4.4.2 More Flexible Encodings of k-majority Digraphs

While the tools presented in Publication [6] for the question of k-majority digraphs are already quite flexible, the concrete encodings of *majority implications* and *indifference implications* (which ensure that the right majority relation is implemented by the desired preference profile) implicitly assume anti-symmetry of individual preferences. This assumption, however, is no longer valid in the setting of *weak* individual preferences, i.e., preferences with ties.

weak preferences

Fortunately, there are more general encodings of these two properties which come only at the cost of a rather small performance loss. We present two such encodings with increasing generality.

The basic idea of the first encoding is to partition, for each pair of alternatives a and b, the set of agents into three subsets: those with $a \succ_i b$ (we call this set $I \subseteq N$), those with $b \succ_i a$ (called $J \subseteq N$), and those with $a \sim_i b$ (implicitly given by $N \setminus (I \cup J)$). One can then compare these three subsets of N with respect to their relative size in order to determine the majority situation between a and b.

The majority implications of a majority edge from a to b can, hence, be encoded as (with variables $p_{i,x,y}$ standing for $x \succsim_i y$):

majority implications

$$
\begin{aligned}
&(\exists I, J)\ \big[|I| > |J| \wedge I \cap J = \emptyset \wedge \\
&\qquad\qquad (\forall i \in I) a \succ_i b \wedge (\forall j \in J) b \succ_j a \wedge (\forall k \notin I \cup J) a \sim_k b\big] \\
&\equiv \bigvee_{\substack{I \cap J = \emptyset \\ |I| > |J|}} \left[\bigwedge_{i \in I} a \succ_i b \wedge \bigwedge_{j \in J} b \succ_j a \wedge \bigwedge_{k \notin I \cup J} a \sim_k b \right] \\
&\equiv \bigvee_{\substack{I \cap J = \emptyset \\ |I| > |J|}} \left[\bigwedge_{i \in I} (p_{i,a,b} \wedge \neg p_{i,b,a}) \wedge \bigwedge_{j \in J} (p_{j,b,a} \wedge \neg p_{j,a,b}) \wedge \right. \\
&\qquad\qquad \left. \bigwedge_{k \notin I \cup J} (p_{k,a,b} \wedge p_{k,b,a}) \right].
\end{aligned}
$$

In order to convert this formula into CNF, one, for instance, applies Tseitin's transformation (cf. Section 3.1.1) and replaces the long conjunction by an auxiliary variable. For the indifference implications, one just needs to replace $|I| > |J|$ by $|I| = |J|$ in the outmost disjunction.

indifference implications

If one wants to allow for even more general (e.g., incomplete) individual preferences, one can resort to an even more flexible (but again less performant) encoding for each majority edge from a to b:

incomplete preferences

$$
\begin{aligned}
&|\{i \in N \mid a \succsim_i b\}| > |\{j \in N \mid b \succsim_j a\}| \\
&\equiv (\exists I \subseteq N)\ [((\forall i \in I)\ a \succsim_i b) \wedge \\
&\qquad\qquad (\forall J \subseteq N)\ (((\forall j \in J)\ b \succsim_j a) \to |J| < |I|)] \\
&\equiv \bigvee_{I \subseteq N} \left[\bigwedge_{i \in I} a \succsim_i b \wedge \bigwedge_{J \subseteq N} \left(\left(\bigwedge_{j \in J} b \succsim_j a \right) \to |J| < |I| \right) \right] \\
&\equiv \bigvee_{I \subseteq N} \left[\bigwedge_{i \in I} a \succsim_i b \wedge \bigwedge_{J \subseteq N} \left(|J| \geqslant |I| \to \left(\bigvee_{j \in J} \neg b \succsim_j a \right) \right) \right] \\
&\equiv \bigvee_{I \subseteq N} \left[\bigwedge_{i \in I} p_{i,a,b} \wedge \bigwedge_{\substack{J \subseteq N \\ |J| \geqslant |I|}} \bigvee_{j \in J} \neg p_{j,b,a} \right],
\end{aligned}
$$

where the term in square parentheses can, as before, be replaced by helper variables to reach a formulation in CNF.

Part II

PUBLICATIONS

5 COMPUTER-AIDED THEOREM PROVING

[1] Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565-602, 2016 (with F. Brandt).[37]

[2] Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 18–24. AAAI press, 2015 (with F. Brandl, F. Brandt, and J. Hofbauer).

[3] Optimal bounds for the no-show paradox via SAT solving. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 314–322. IFAAMAS, 2016 (with F. Brandt and D. Peters).

[4] Proving the incompatibility of efficiency and strategyproofness via SMT solving. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 116–122. AAAI Press, 2016 (with F. Brandl and F. Brandt).

37 An earlier version of this paper appeared in the Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 1193–1200. IFAAMAS, 2014.

FINDING STRATEGYPROOF SOCIAL CHOICE FUNCTIONS VIA SAT SOLVING [1]

Peer-reviewed Journal Paper

Authors: F. Brandt and C. Geist (alphabetical author ordering)

In: Journal of Artificial Intelligence Research, 55:565–602, 2016

Abstract: A promising direction in computational social choice is to address research problems using computer-aided proving techniques. In particular with SAT solvers, this approach has been shown to be viable not only for proving classic impossibility theorems such as Arrow's Theorem but also for finding new impossibilities in the context of preference extensions. In this paper, we demonstrate that these computer-aided techniques can also be applied to improve our understanding of strategyproof irresolute social choice functions. These functions, however, requires a more evolved encoding as otherwise the search space rapidly becomes much too large. Our contribution is two-fold: We present an efficient encoding for translating such problems to SAT and leverage this encoding to prove new results about strategyproofness with respect to Kelly's and Fishburn's preference extensions. For example, we show that no Pareto-optimal majoritarian social choice function satisfies Fishburn-strategyproofness. Furthermore, we explain how human-readable proofs of such results can be extracted from minimal unsatisfiable cores of the corresponding SAT formulas.

Contribution of thesis author: Methodology, results, implementation, literature review, mathematical model, presentation, project and paper management

Journal of Artificial Intelligence Research 55 (2016) 565-602 Submitted 08/15; published 03/16

Finding Strategyproof Social Choice Functions via SAT Solving

Felix Brandt BRANDTF@IN.TUM.DE
Technical University of Munich (TUM)
Munich, Germany

Christian Geist GEIST@IN.TUM.DE
Technical University of Munich (TUM)
Munich, Germany

Abstract

A promising direction in computational social choice is to address research problems using computer-aided proving techniques. In particular with SAT solvers, this approach has been shown to be viable not only for proving classic impossibility theorems such as Arrow's Theorem but also for finding new impossibilities in the context of preference extensions. In this paper, we demonstrate that these computer-aided techniques can also be applied to improve our understanding of strategyproof irresolute social choice functions. These functions, however, require a more evolved encoding as otherwise the search space rapidly becomes much too large. Our contribution is two-fold: We present an efficient encoding for translating such problems to SAT and leverage this encoding to prove new results about strategyproofness with respect to Kelly's and Fishburn's preference extensions. For example, we show that no Pareto-optimal majoritarian social choice function satisfies Fishburn-strategyproofness. Furthermore, we explain how human-readable proofs of such results can be extracted from minimal unsatisfiable cores of the corresponding SAT formulas.

1. Introduction

Ever since the famous Four Color Problem was solved using a computer-assisted approach, it has been clear that computers can contribute significantly not only to verifying existing but also to finding and proving new results. Due to its rigorous axiomatic foundation, *social choice theory* appears to be a field in which computer-aided theorem proving is a particularly promising line of research. Perhaps the best known result in this context stems from Tang and Lin (2009), who reduce well-known impossibility results such as Arrow's theorem to finite instances, which can then be checked by a satisfiability (SAT) solver (see, e.g., Biere, Heule, van Maaren, & Walsh, 2009). Geist and Endriss (2011) were able to extend this method to a fully-automatic search algorithm for impossibility theorems in the context of preference relations over sets of alternatives. In this paper, we apply these techniques to improve our understanding of strategyproofness in the context of set-valued, or so-called irresolute, social choice functions. These types of problems, however, are more complex and require an evolved encoding as otherwise the search space rapidly becomes too large. Table 1 illustrates how quickly the number of involved objects grows and that, as a result, exhaustive search is doomed to fail.

Alternatives	4	5	6	7
Choice sets	15	31	63	127
Tournaments	64	1,024	32,768	$\sim 2 \cdot 10^6$
Canonical tournaments	4	12	56	456
Majoritarian SCFs	$\mathbf{50,625}$	$\sim \mathbf{10^{18}}$	$\sim \mathbf{10^{101}}$	$\sim \mathbf{10^{959}}$

Table 1: Number of objects involved in problems with irresolute majoritarian SCFs

Our contribution is two-fold. On the one hand, we provide an extended framework of SAT-based computer-aided theorem proving techniques for statements in social choice theory and related research areas. Despite its complexity, this framework allows for the extraction of human-readable proofs, which eliminates the need for extensive (and difficult) verification of the underlying techniques. On the other hand, rather than only reproducing existing results, we solve some open problems, which are of independent interest, in the context of irresolute strategyproof social choice functions. These results are unlikely to have been found without the help of computers, which further strengthens the importance of the approach.

The results obtained by computer-aided theorem proving have already found attention in the social choice community (Chatterjee & Sen, 2014) and similar techniques have proven to be quite effective for other problems in economics, too. Examples are the ongoing work by Fréchette, Newman, and Leyton-Brown (2016) in which SAT solvers are used for the development and execution of the FCC's upcoming reverse spectrum auction, recent results by Drummond, Perrault, and Bacchus (2015) who solve stable matching problems via SAT solving, as well as work by Tang and Lin (2011) who apply SAT solving to discover classes of two-player games with unique pure Nash equilibrium payoffs. In another recent paper, Caminati, Kerber, Lange, and Rowat (2015) verified combinatorial Vickrey auctions via higher-order theorem provers. In some respect, our approach bears similarities to *automated mechanism design* (see, e.g., Conitzer & Sandholm, 2002), where desirable properties are encoded and mechanisms are computed to fit specific problem instances. There is also a body of work on logical formalizations of important theorems in social choice theory, most prominently, *Arrow's Theorem* (see, e.g., Nipkow, 2009; Grandi & Endriss, 2013; Cinà & Endriss, 2015), which has been directed more towards formalizing and verifying existing results.

Given the universality of the SAT-based method and its ease of adaptation (e.g., "testing" of similar conjectures with minimal effort by simply replacing or altering some axioms), we expect these and similar techniques to be applicable to other open problems in social choice theory and related research areas in the future. Results for different variants of the no-show paradox (Brandl, Brandt, Geist, & Hofbauer, 2015; Brandt, Geist, & Peters, 2016c) support this hypothesis. It should be noted, however, that—at least currently—an expert user or programmer is required to operate these systems. An interesting question that remains is whether it is possible to develop an automatic proof assistant that allows researchers to quickly test hypotheses on small domains without giving up too much generality and efficiency.

Let us now turn towards the social choice theoretic results. Formally, a social choice function (SCF) is defined as a function that maps individual preferences over a set of alternatives to a set of socially most-preferred alternatives. An SCF is strategyproof if no agent can obtain a more preferred outcome by misrepresenting his preferences. It is well-known from the Gibbard-Satterthwaite theorem that, when restricting attention to SCFs that always return a single alternative, only trivial SCFs can be strategyproof. The assumption of single-valuedness, however, has been criticized for being unreasonably restrictive (see, e.g., Gärdenfors, 1976; Kelly, 1977; Taylor, 2005; Barberà, 2010). A proper definition of strategyproofness for the more general setting of irresolute SCFs requires the specification of preferences over *sets* of alternatives. Rather than asking the agents to specify their preferences over all sets (which requires exponential space and would be bound to various rationality constraints), it is typically assumed that preferences over single alternatives can be extended to preferences over sets. Of course, there are various ways how to extend preferences to sets (see, e.g., Gärdenfors, 1979; Duggan & Schwartz, 2000; Taylor, 2005), each of which leads to a different class of strategyproof SCFs. A function that yields a preference relation over subsets of alternatives when given a preference relation over single alternatives is called a *set extension* or *preference extension*. In this paper, we focus on two set extensions attributed to Kelly (1977) and Fishburn (1972),[1] which have been shown to arise uniquely under very natural assumptions (Gärdenfors, 1979; Erdamar & Sanver, 2009; see also Section 2.2 of this paper).

While strategyproofness for Kelly's extension (henceforth *Kelly-strategyproofness*) is known to be a rather restrictive condition (Kelly, 1977; Barberà, 1977; Nehring, 2000), some SCFs such as the Pareto rule, the omninomination rule, the top cycle, the uncovered set, the minimal covering set, and the bipartisan set were shown to be Kelly-strategyproof (Brandt, 2015). Interestingly, the more prominent of these SCFs are *majoritarian*, i.e., they are based on the pairwise majority relation only and can be ordered with respect to set inclusion. These results suggest that the bipartisan set may be the finest Kelly-strategyproof majoritarian SCF. In this paper, we show that this is not the case by automatically generating a Kelly-strategyproof SCF that is strictly contained in the bipartisan set. Brandt (2015) furthermore showed that, under a mild condition, Kelly-strategyproofness carries over to coarsenings of an SCF. Thus, finding inclusion-minimal Kelly-strategyproof SCFs is of particular interest. We address this problem by automating the search for these functions in small domains and report on our findings.

Existing results suggest that the more demanding notion of *Fishburn-strategyproofness* may only be satisfied by rather indiscriminating SCFs such as the top cycle (Feldman, 1979; Brandt & Brill, 2011; Sanver & Zwicker, 2012).[2] Using our computer-aided proving technique, we are able to confirm this suspicion by proving that, within the domain of majoritarian SCFs, Fishburn-strategyproofness is incompatible with Pareto-optimality. In order to achieve this impossibility, we manually prove a novel characterization of Pareto-optimal ma-

1. Gärdenfors (1979) attributed this extension to Fishburn because it is the weakest extension that satisfies a certain set of axioms proposed by Fishburn (1972). Some authors, however, refer to it as the Gärdenfors extension, a term which we reserve for the extension due to Gärdenfors (1976) himself.
2. The negative result by Ching and Zhou (2002) uses Fishburn's extension but a much stronger notion of strategyproofness.

joritarian SCFs and an induction step, which allows us to generalize the computer-generated impossibility to larger numbers of alternatives.

The paper is structured as follows. In Section 2, we present the general mathematical framework that we use throughout this paper and introduce the new condition of *tournament-strategyproofness*, which we show to be equivalent to standard strategyproofness for majoritarian SCFs. In Section 3, we describe our computer-aided proving method and explain how to encode the main questions of this paper as SAT problems. We also describe optimization techniques and other features of the approach. In Section 4, we report on our main findings—an impossibility and a possibility result—and discuss possible extensions and their limits. In Section 5, our novel approach to proof extraction from these computer-generated results is presented. We provide a human-readable proof of our main result that can be verified without the help of computers. Finally, in Section 6 we wrap up our work and give an outlook on further research directions.

2. Mathematical Framework of Strategyproofness

In this section, we provide the terminology and notation required for our results and introduce notions of strategyproofness for majoritarian SCFs that allow us to abstract away any reference to preference profiles.

2.1 Social Choice Functions

Let $N = \{1, \dots, n\}$ be a set of at least three voters with preferences over a finite set A of m alternatives. For convenience, we assume that n is odd, which entails that the pairwise majority relation is antisymmetric. The preferences of each voter $i \in N$ are represented by a complete, antisymmetric, and transitive *preference relation* $R_i \subseteq A \times A$. The interpretation of $(x, y) \in R_i$, usually denoted by $x \; R_i \; y$, is that voter i values alternative x at least as much as alternative y. The set of all preference relations over A will be denoted by $\mathcal{R}(A)$. The set of *preference profiles*, i.e., finite vectors of preference relations, is then given by $\mathcal{R}^*(A)$. The typical element of $\mathcal{R}^*(A)$ will be $R = (R_1, \dots, R_n)$. In accordance with conventional notation, we write P_i for the strict part of R_i, i.e., $x \; P_i \; y$ if $x \; R_i \; y$ but not $y \; R_i \; x$. Note that the only difference between R_i and P_i is that R_i is reflexive while P_i is not. In order to improve readability, we write $R_i : x, y, z$ as a shorthand for $x \; P_i \; y \; P_i \; z$. In a preference profile, the *weight* of an ordered pair of alternatives $w_R(x, y)$ is defined as the majority margin $|\{i \in N \mid x \; R_i \; y\}| - |\{i \in N \mid y \; R_i \; x\}|$.

Our central objects of study are *social choice functions*, i.e., functions that map the individual preferences of the voters to a nonempty set of socially preferred alternatives.

Definition 1. A *social choice function (SCF)* is a function $f : \mathcal{R}^*(A) \to 2^A \setminus \emptyset$.

An SCF is *resolute* if $|f(R)| = 1$ for all $R \in \mathcal{R}^*(A)$, otherwise it is *irresolute*.

We restrict our attention to *majoritarian* SCFs, or *tournament solutions*, which are defined using the *majority relation*. The majority relation R_M of a preference profile R is the relation on $A \times A$ defined by

$$(x, y) \in R_M \text{ if and only if } w_R(x, y) \geq 0,$$

for all alternatives $x, y \in A$. An SCF f is said to be *majoritarian* if it is neutral[3] and its outcome only depends on the majority relation, i.e., $f(R) = f(R')$ whenever $R_M = R'_M$. As before, we write P_M for the strict part of R_M, i.e., $a \; P_M \; b$ if $a \; R_M \; b$ but not $b \; R_M \; a$.

An alternative x is called a *Condorcet winner* in R if $x \; P_M \; y$ for all $y \in A \setminus \{x\}$. In other words, a Condorcet winner is a "best" alternative with respect to the majority relation and it seems natural that majoritarian SCFs should select a Condorcet winner. Unfortunately, such clear-cut winners do not exist in general and a variety of so-called *Condorcet extensions*, i.e., SCFs that uniquely return a Condorcet winner whenever one exists but differ in their treatment of the remaining cases, have been proposed in the literature. In this paper, we consider the following majoritarian Condorcet extensions (see, e.g., Laslier, 1997; Brandt, Brill, & Harrenstein, 2016a, for more information).

Top Cycle Define a *dominant set* to be a non-empty set of alternatives $D \subseteq A$ such that for any alternative $x \in D$ and $y \in A \setminus D$ we have $x \; P_M \; y$. The *top cycle TC* (also known as *weak closure maximality*, *GETCHA*, or the *Smith set*) is defined as the (unique) inclusion-minimal dominant subset of A.[4]

Uncovered Set Let C denote the *covering relation* on $A \times A$, i.e., $x \; C \; y$ ("x covers y") if and only if $x \; P_M \; y$ and, for all $z \in A$, $y \; P_M \; z$ implies $y \; P_M \; z$. The *uncovered set UC* contains those alternatives that are not covered according to C, i.e., $UC(R) = \{x \in A \mid y \; C \; x \text{ for no } y \in A\}$.

Bipartisan Set Consider the symmetric two-player zero-sum game in which the set of actions for both players is given by A and payoffs are defined as follows. Suppose the first player chooses a and the second player chooses b. Then the payoff for the first player is 1 if $a \; P_M \; b$, -1 if $b \; P_M \; a$, and 0 otherwise. The *bipartisan set BP* contains all alternatives that are played with positive probability in the unique Nash equilibrium of this game.

An SCF f is called a *refinement* of another SCF g if $f(R) \subseteq g(R)$ for all preference profiles $R \in \mathcal{R}^*(A)$. In short, we write $f \subseteq g$ in this case. It can be shown for the above that $BP \subseteq UC \subseteq TC$ (see, e.g., Laslier, 1997).

For our main result, we define the well-known notion of *Pareto-optimality*: an SCF f is Pareto-optimal if it never selects any *Pareto-dominated* alternative $x \in A$, i.e., $x \notin f(R)$ whenever there exists $y \in A$ such that $y \; P_i \; x$ for all $i \in N$.

2.2 Strategyproofness

Although our investigation of strategyproof SCFs is universal in the sense that it can be applied to any set extension, in this paper we will concentrate on two well-known set extensions attributed to Kelly (1977) and Fishburn (1972).[5] These two set extensions

3. Neutrality postulates that for any permutation π of the alternatives A the SCF produces the "same" outcome (modulo the permutation). See also Section 3.1.1.
4. It is easily seen that the set of dominant sets is ordered with respect to set inclusion and therefore admits a unique minimal element. Assume for a contradiction that two dominant sets $X, Y \subseteq A$ are not contained in each other. Then, there exists $x \in X \setminus Y$ and $y \in Y \setminus X$. The definition of dominant sets requires that $x \; P_M \; y$ and $y \; P_M \; x$, a contradiction.
5. Another natural and well-known set extension by Gärdenfors leads to an even stronger notion of strategyproofness, which cannot be satisfied by any interesting majoritarian SCF (Brandt & Brill, 2011). Note

are defined as follows: Let R_i be a preference relation over A and $X, Y \subseteq A$ two nonempty subsets of A.

$$X \; R_i^{\mathbf{K}} \; Y \text{ if and only if } x \; R_i \; y \text{ for all } x \in X \text{ and all } y \in Y. \qquad \text{(Kelly, 1977)}$$

One interpretation of this extension is that voters are completely unaware of the mechanism (e.g., a lottery) that will be used to pick the winning alternative (Gärdenfors, 1979; Erdamar & Sanver, 2009). In other words, it contains exactly the pairwise comparisons which voters can make without knowledge of the mechanism (e.g., $\{a, b\} \; R_i^{\mathbf{K}} \; \{c\}$ if $a \; P_i \; b \; P_i \; c$).

$$X \; R_i^{\mathbf{F}} \; Y \text{ if and only if all of the following three conditions are satisfied:}$$

$$\begin{aligned} & x \; R_i \; y \text{ for all } x \in X \setminus Y \text{ and } y \in X \cap Y, \\ & y \; R_i \; z \text{ for all } y \in X \cap Y \text{ and } z \in Y \setminus X, \text{ and} \qquad \text{(Fishburn, 1972)} \\ & x \; R_i \; z \text{ for all } x \in X \setminus Y \text{ and } z \in Y \setminus X. \end{aligned}$$

For this extension one may assume the winning alternative to be picked by a lottery according to some underlying *a priori* distribution that voters are not aware of (Ching & Zhou, 2002). Alternatively, the existence of a chairman who breaks ties according to a linear, but unknown, preference relation also rationalizes this preference extension (Erdamar & Sanver, 2009). For both of these interpretations, the extension describes exactly the conclusions a voter who is aware of the tie-breaking method can draw (e.g., $\{a, b\} \; R_i^{\mathbf{F}} \; \{b, c\}$ if $a \; P_i \; b \; P_i \; c$, which does not hold for Kelly's extension $R_i^{\mathbf{K}}$).

It is easy to see that $X \; R_i^{\mathbf{K}} \; Y$ implies $X \; R_i^{\mathbf{F}} \; Y$ for any pair of sets $X, Y \subseteq A$.

As we plan to prove a few results for entire classes of set extensions, we call a set extension $\mathcal{E}$ *independent of irrelevant alternatives (IIA)* if its comparison of two sets X and Y only depends on the restriction of individual preferences to $X \cup Y$. Formally, $\mathcal{E}$ satisfies IIA if for all pairs of preference relations R_i, R_i' and nonempty sets $X, Y \subseteq A$ such that $R_i\,|_{X \cup Y} = R_i'\,|_{X \cup Y}$ it holds that

$$X \; R_i^{\mathcal{E}} \; Y \text{ if and only if } X \; R_i'^{\mathcal{E}} \; Y.$$

This is a very mild and natural condition, which is satisfied by the previously mentioned set extensions and any other major set extension from the literature we are aware of.

Based on any set extension $\mathcal{E}$, we can state a corresponding notion of $P^{\mathcal{E}}$-strategyproofness for irresolute SCFs. Note that in contrast to some related papers (e.g., Ching & Zhou, 2002; Sato, 2008), we interpret preference extensions as fully specified (incomplete) preference relations rather than minimal conditions on set preferences.

Again, we write $P_i^{\mathcal{E}}$ for the asymmetric part of $R_i^{\mathcal{E}}$, for any set extension $\mathcal{E}$.

Definition 2. Let $\mathcal{E}$ be a set extension. An SCF f is $P^{\mathcal{E}}$*-manipulable* by voter i if there exist preference profiles R and R' with $R_j = R_j'$ for all $j \neq i$ such that $f(R')$ is $\mathcal{E}$-preferred to $f(R)$ by voter i, i.e.,

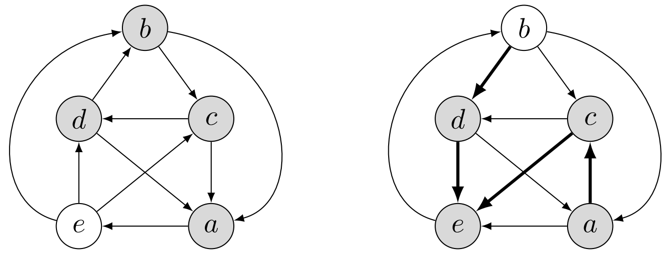
$$f(R') \; P_i^{\mathcal{E}} \; f(R).$$

An SCF is called $P^{\mathcal{E}}$*-strategyproof* if it is not $P^{\mathcal{E}}$-manipulable.

that our negative result for Fishburn-strategyproofness trivially carries over to such more demanding set extensions.

1	2,3	4	5	6	7
e	d	d	c	b	b
c	a	e	e	c	a
a	e	b	b	a	c
d	b	c	d	e	e
b	c	a	a	d	d

(a) A preference profile R



(b) The corresponding (strict) majority relation P_M

(c) The manipulated (strict) majority relation P'_M when the first agent submits b, a, c, d, e as his preferences. All edges that have been impacted by this change are depicted in bold.

Figure 1: Let the choice sets be as indicated by shaded nodes; this example is taken from the proof of Theorem 3 (cf. Section 5.1.3). The first agent in R can $P^{\mathbf{F}}$-manipulate by submitting b, a, c, d, e as his preferences (since $f(R') = \{a, c, d, e\}\ P_1^{\mathbf{F}}\ \{a, b, c, d\} = f(R)$), but this does not constitute a $P^{\mathbf{K}}$-manipulation (since $\{a, b, c, d\}$ and $\{a, c, d, e\}$ are incomparable according to the Kelly-extension).

It follows from the observation on set extensions above that $P^{\mathbf{F}}$-strategyproofness implies $P^{\mathbf{K}}$-strategyproofness. An example illustrating both notions of strategyproofness is shown in Figure 1.

Of the above SCFs, TC is $P^{\mathbf{F}}$-strategyproof, BP is $P^{\mathbf{K}}$- but not $P^{\mathbf{F}}$-strategyproof, whereas UC was only known to satisfy $P^{\mathbf{K}}$-strategyproofness (Brandt & Brill, 2011; Brandt, 2015).

2.3 Tournament-Strategyproofness

In order to allow for a more efficient encoding, we would like to omit references to preference profiles and replace them with a more succinct representation with the same expressive power. For majoritarian SCFs, the natural choice is to use the (strict) majority relation, which, for an odd number of voters, can be represented by a *tournament*:

A *tournament* is an asymmetric and complete binary relation on the set of alternatives A.[6] We can thus view majoritarian SCFs as functions defined on tournaments rather than preference profiles, and, in slight abuse of notation,[7] write $f(T)$ instead of $f(R)$ with $T = P_M$ being the strict part of the majority relation of R. We, furthermore, denote by $T \setminus T' := \{e \in T : e \notin T'\}$ the *edge difference* of two tournaments T and T'.

For our encoding to be efficient, it will be important to formalize the notion of strategyproofness using only references to tournaments rather than preference profiles. The

6. Note that tournaments can be defined by their edge set only. Since there is exactly one edge between any pair of vertices, the vertex set can be derived from the edge set.

7. It may be noted that, while majoritarian SCFs map from profiles (with an arbitrary, but fixed number of voters) to sets of alternatives, their interpretation via tournaments abstracts away the reference to individual voters. This has implications for Theorems 1 and 3, which depend upon the presence of a sufficient number of voters. We discuss the required number of voters in Section 5.2.

following definition serves this purpose and will be shown to be equivalent to the standard notion of strategyproofness for majoritarian SCFs.

Definition 3. A majoritarian SCF f is said to be $P^{\mathcal{E}}$*-tournament-manipulable* if there exist tournaments T, T' and a preference relation $R_\mu \supseteq T \setminus T'$ such that

$$f(T') \; P^{\mathcal{E}}_{\mu} \; f(T).$$

A majoritarian SCF is called $P^{\mathcal{E}}$*-tournament-strategyproof* if it is not $P^{\mathcal{E}}$-tournament-manipulable.

Theorem 1. *A majoritarian SCF is* $P^{\mathcal{E}}$*-strategyproof if and only if it is* $P^{\mathcal{E}}$*-tournament-strategyproof.*

Proof. We show that a majoritarian SCF is $P^{\mathcal{E}}$-manipulable if and only if it is $P^{\mathcal{E}}$-tournament-manipulable.

For the direction from left to right, let f be a $P^{\mathcal{E}}$-manipulable majoritarian SCF. Then there exist preference profiles R, R' and an integer j with $R_i = R'_i$ for all $i \neq j$ such that $f(R') \; P^{\mathcal{E}}_j \; f(R)$. Define tournaments $T := P_M$ and $T' := P'_M$ as the strict majority relations of R and R', respectively. Since R and R' only differ for voter j, it follows that $T \setminus T' \subseteq R_j$, i.e., all edges that are reversed from T to T' must have been in R_j. Thus, with $R_\mu := R_j$, we get that f is tournament-manipulable.

For the converse, let f be a $P^{\mathcal{E}}$-tournament-manipulable majoritarian SCF. The SCF f then admits a manipulation instance, i.e., there are two tournaments T, T' and a preference relation $R_\mu \supseteq T \setminus T'$ such that $f(T') \; P^{\mathcal{E}}_{\mu} \; f(T)$.

As in the proof of McGarvey's Theorem (McGarvey, 1953), we construct a preference profile $R^- = (R_1, \ldots, R_{n-1})$ which has $T \cap T'$ as the strict part P^-_M of its majority relation: we start from an empty profile and, for each strict edge $(a, b) \in T \cap T'$, add two voters $i_{a,b}$ and $j_{a,b}$ with the preferences

$$R_{i_{a,b}} : a, b, x_1, \ldots, x_{m-2} \text{ and } R_{j_{a,b}} : x_{m-2}, \ldots, x_1, a, b, \text{ respectively.}$$

Here $x_1, \ldots, x_{m-2}$ denotes an arbitrary enumeration of the $m - 2$ alternatives in $A \setminus \{a, b\}$.

It then holds for the weights $w_{R^-}(a, b)$ of all edges $(a, b) \in T$ that

$$w_{R^-}(a, b) = \begin{cases} 2 & \text{if } (a, b) \in T \cap T' \\ 0 & \text{if } (a, b) \in T \setminus T'. \end{cases}$$

Note that the number of voters $n - 1$ in R^- has to be even (and at most $m^2 - m - 2$). By adding R_μ as the n-th voter, we get to a profile $R := (R^-, R_\mu)$ with an odd number of voters as required. Then $w_R(a, b) \geq 1$ for all edges $(a, b) \in T$ and, thus, R has T as its (strict) majority relation. The second profile R' can be defined to contain the same first $n - 1$ voters from R and the reversed preference $\overline{R_\mu}$ as the n-th voter (i.e., $R' := (R^-, \overline{R_\mu})$).[8] The profile R' then has T' as its (strict) majority relation (since $w_{R'}(a, b) = -1$ for all

8. Immunity to manipulation by *reversing preferences* has been considered by Sanver and Zwicker (2012) under the name of *half-way monotonicity.* Our proof entails that (weak) half-way monotonicity is equivalent to strategyproofness for majoritarian SCFs.

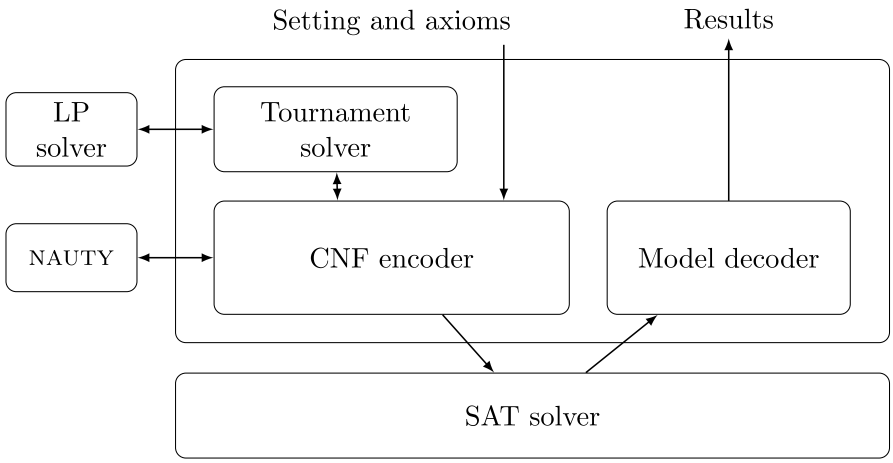


Figure 2: High-level system architecture

edges $(a, b) \in T \setminus T'$ and the weights of all edges in $T \cap T'$ are at least 1 again), which completes the manipulation instance. I.e., we have found preference profiles R, R' which only differ for voter n (who has "truthful" preferences R_μ) and for which it holds that $f(R') = f(T')\ P_\mu^{\mathcal{E}}\ f(T) = f(R)$. $\square$

3. Methodology

The method applied in this paper is similar to and yet more powerful than the ones presented by Tang and Lin (2009) and Geist and Endriss (2011). Rather than translating the whole problem naïvely to SAT, a more evolved approach, which resolves a large degree of freedom already during the encoding of the problem, is employed. This approach is comparable to the way SMT (satisfiability modulo theories) solving works: At the core there is a SAT solver; certain aspects of the problem, however, are dealt with in a separate theory solving unit which accepts a richer language and makes use of specific domain knowledge (Biere et al., 2009, ch. 26). The general idea, however, remains to encode the problem into a language suitable for SAT solving and to apply a SAT solver as an efficient and universal problem solving machine.

While desirable, using existing tools for higher-order formalizations directly rather than our specific approach, unfortunately, is not an option. For instance, a formalization of strategyproof majoritarian SCFs in higher-order logic (HOL) as accepted by Nitpick (Blanchette & Nipkow, 2010) is straightforward, highly flexible, and well-readable, but only successful for proofs and counterexamples involving up to three alternatives before the search space is exceeded.[9] An optimized formalization, which we derived together with the author of Nitpick (at the cost of reduced readability and flexibility), extends the performance to four alternatives, which turns out to be just too low for our results.

9. On the other hand, the strict formalization required for Nitpick helped to identify a formally inaccurate definition of Fishburn-strategyproofness by Gärdenfors (1979) (which was later repeated by other authors).

Concretely, our approach is the following (see also the high-level architecture in Figure 2): for a given domain size n we want to check whether there exists a majoritarian SCF f that satisfies a set of axioms (e.g., $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality). We then encode the setting as well as the given axioms as a propositional formula (SAT instance) and let a SAT solver decide whether this formula has a satisfying assignment. If it has a satisfying assignment, we can decode it into a concrete instance of a majoritarian SCF f which satisfies the required properties. If the formula is unsatisfiable, we know that no such function f exists.

As we will see, depending on the problem, some preparatory tasks have to be solved before the actual encoding: (i) sets, tournaments, and preference relations are enumerated; (ii) isomorphisms between tournaments are determined using the tool nauty (McKay & Piperno, 2013); (iii) choice sets for specific SCFs are computed (e.g., via matrix multiplication for UC and linear programming for BP).

In the following, we describe in more detail how the general setting of majoritarian SCFs as well as desirable properties, such as strategyproofness, can be encoded as a SAT problem in CNF (conjunctive normal form).[10] First, we describe our initial encoding, which is expressive enough to encode all required properties, but allows for small domain sizes of (depending on the axioms) at most four to five alternatives only. Second, we explain how this encoding can be optimized to increase the overall performance by orders of magnitude such that larger instances of up to seven alternatives are solvable.

3.1 Initial Encoding

By design, SAT solvers operate on propositional logic. A direct and naïve propositional encoding of the problem would, however, require a huge number of propositional variables since many higher-order concepts are involved (e.g., sets of alternatives, preference relations over sets as well as over alternatives, and functions from tuples of such relations to sets). In our approach, we use only one type of variable to encode SCFs. The variables are of the form $c_{T,X}$ with T being a tournament and X being a set of alternatives.[11] The semantics of these variables are that $c_{T,X}$ if and only if $f(T) = X$, i.e., the majoritarian SCF f selects the set of alternatives X as the choice set for any preference profile with (strict) majority relation T. In total, this gives us a high but manageable number of $2^{\frac{m(m-1)}{2}} \cdot 2^m = 2^{\frac{m(m+1)}{2}}$ variables in the initial encoding.

An encoding with variables $c_{T,x}$ for alternatives x rather than sets would require less variable symbols. This encoding, however, leads to much more complexity in the generated clauses, which more than offsets these savings. This imbalance is best exhibited in the encoding of strategyproofness where statements are always made for pairs of outcomes (i.e., sets of alternatives). Each occurrence of $c_{T,X}$ could be replaced by $\bigwedge_{x \in X} c_{T,x} \wedge \bigwedge_{y \notin X} \neg c_{T,y}$. But since this formula then contains a conjunction within a disjunction, which is not possible

10. Converting an arbitrary propositional formula naïvely to CNF can lead to an exponential blow-up in the length of the formula. There are, however, well-known efficient techniques (e.g., Tseitin's encoding, see Tseitin, 1983) to avoid this at the cost of introducing linearly many auxiliary variables. We apply these techniques manually when needed.

11. In all algorithms, a subroutine $\mathrm{c}(T, X)$ will take care of the compact enumeration of variables. Since we know in advance how many tournaments and non-empty subsets there are, we can simply use a standard enumeration method for pairs of objects.

in CNF, either expansion (and therefore an exponential blow-up) or replacement (e.g., by a helper variable $c_{T,X} \leftrightarrow \bigwedge_{x\in X} c_{T,x}$) would be required.

The following two subsections demonstrate the initial encoding of both contextual as well as explicit axioms to CNF.

3.1.1 Context Axioms

Apart from the explicit axioms, which we are going to describe in the next subsection, there are further axioms that need to be considered in order to fully model the context of majoritarian SCFs. For this purpose, an arbitrary function that maps tournaments to non-empty sets of its vertices will be called a *tournament choice function.* Using our initial encoding three axioms are introduced, which will ensure that functionality of the tournament choice function and neutrality are respected (making it a tournament solution): (1) functionality, (2) canonical isomorphism equality, and (3) the orbit condition.

The first axiom ensures that the relational encoding of f by variables $c_{T,X}$ indeed models a function rather than an arbitrary relation, i.e., for each tournament T there is exactly one set X such that the variable $c_{T,X}$ is set to true. In formal terms this can be written as

$$\begin{aligned} &(\forall T)\,((\exists X)\ c_{T,X} \wedge (\forall Y,Z)\ Y \neq Z \rightarrow \neg(c_{T,Y} \wedge c_{T,Z})) \\ \equiv\ &\bigwedge_{T}\left(\left(\bigvee_{X} c_{T,X}\right) \wedge \bigwedge_{Y\neq Z}(\neg c_{T,Y} \vee \neg c_{T,Z})\right). \end{aligned} \tag{1}$$

As an illustrative example, the corresponding simple pseudo-code for generating the CNF file can be found in Appendix B.

The second and third axiom together constitute neutrality of the tournament choice function f, which, formally, can be written as

$$\pi(f(T)) = f(\pi(T)) \text{ for all tournaments } T \text{ and permutations } \pi : A \rightarrow A.$$

A direct encoding of this neutrality axiom, however, would be tedious due to the quantification over all permutations. In addition, our reformulation as *canonical isomorphism equality* and *orbit condition* enables a substantial optimization of the encoding as we will see in Section 3.2. We require further observations in order to precisely state these two axioms.

We use the well-known fact that graph isomorphisms define an equivalence relation on the set of all tournaments.[12] For each equivalence class, pick a representative as the *canonical tournament* of this class. For any tournament T, we then have a unique canonical representation (denoted by $T_{\mathfrak{c}}$). We also pick one of the potentially many isomorphisms from $T_{\mathfrak{c}}$ to T as the *canonical isomorphism* of T and denote it by π_T.[13] This allows us to formulate the axiom of *canonical isomorphism equality.*

Definition 4. A tournament choice function f satisfies *canonical isomorphism equality* if

$$f(T) = \pi_T(f(T_{\mathfrak{c}})) \text{ for all tournaments } T. \tag{2}$$

12. Two tournaments T and T' are isomorphic if there is a permutation $\pi : A \rightarrow A$ such that $\pi(T) = T'$.

13. In practice, the tool nauty will automatically compute canonical representations for both tournaments and isomorphisms.

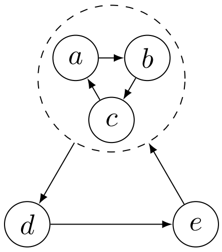


Figure 3: The orbits of this tournament are $\mathcal{O}_T = \{\{a,b,c\},\{d\},\{e\}\}$. A corresponding automorphism would be $\alpha = \begin{pmatrix} a & b & c & d & e \\ b & c & a & d & e \end{pmatrix}$. $C := \{a,b,c\}$ represents a component in the sense that for all of its elements $x \in C$ it holds that $x \; P_M \; d$ and $e \; P_M \; x$.

For the last of the three context axioms, the definition of an *orbit* should be clarified. The orbits of a tournament T are equivalence classes of alternatives according to the following equivalence relation: two alternatives a, b are considered equivalent if and only if there is an *automorphism* $\alpha : A \to A$ which maps a to b, i.e., for which $\alpha(a) = b$. The set of orbits of a tournament T is denoted by $\mathcal{O}_T$. An example can be found in Figure 3.

Definition 5. A tournament choice function f satisfies the *orbit condition* if

$$O \subseteq f(T_\mathfrak{c}) \;\text{ or }\; O \cap f(T_\mathfrak{c}) = \emptyset \tag{3}$$

for all canonical tournaments $T_\mathfrak{c}$ and their orbits $O \in \mathcal{O}_{T_\mathfrak{c}}$.

It can be shown that for any tournament choice function, neutrality is equivalent to the conjunction of the orbit condition and canonical isomorphism equality, or equivalently, that the class of tournament choice functions satisfying the orbit condition and canonical isomorphism equality is equal to the class of tournament solutions. We formalize this statement in Lemma 1. The proof of Lemma 1 is based on standard arguments from category theory and is presented in Appendix A.

Lemma 1. *For any tournament choice function, neutrality is equivalent to the conjunction of the orbit condition and canonical isomorphism equality.*

3.1.2 Explicit Axioms

Many axioms can be efficiently encoded in our proposed encoding language. In this section we present the main conditions required to achieve the results in Section 4. Clearly, the most important one is strategyproofness. In formal terms, $P^\mathcal{E}$-tournament-strategyproofness can be written as

$$\begin{aligned} &(\forall T, T', R_\mu \supseteq T \setminus T') \; \neg \left(f(T') \; P_\mu^\mathcal{E} \; f(T) \right) \\ \equiv &\bigwedge_T \bigwedge_{T'} \bigwedge_{R_\mu \supseteq T \setminus T'} \bigwedge_{Y P_\mu^\mathcal{E} X} (\neg c_{T,X} \vee \neg c_{T,Y}) \end{aligned} \tag{4}$$

where T, T' are tournaments, R_μ is a preference relation, and X, Y are non-empty subsets of A. The algorithmic encoding of strategyproofness is omitted here since we present an optimized version in Section 3.2.

Another property of SCFs that will play an important role in our results is the one of being a refinement of another (known) SCF g. Fortunately, this can easily be encoded using our framework:

$$\begin{aligned} &(\forall T)(\exists X \subseteq g(T))\ f(T) = X \\ \equiv &\bigwedge_{T} \bigvee_{X \subseteq g(T)} c_{T,X}. \end{aligned} \tag{5}$$

If we desire that the resulting SCF f is different from g (for instance, to obtain a *strict* refinement in conjunction with Axiom (5)), we encode the additional clause:

$$\begin{aligned} &(\exists T)\ f(T) \neq g(T) \\ \equiv &\bigvee_{T} \neg c_{T,g(T)}. \end{aligned} \tag{6}$$

Finally, even properties regarding the cardinalities of choice sets can be encoded. The following axiom—stating that $|f(T)| < |g(T)|$ for at least one tournament T—will, for instance, be useful in Section 4.1.1 when searching for SCFs that return small choice sets:

$$\begin{aligned} &(\exists T)(\exists X)\ |X| < |g(T)| \wedge f(T) = X \\ \equiv &\bigvee_{T} \bigvee_{\substack{X \\ |X|<|g(T)|}} c_{T,X}. \end{aligned} \tag{7}$$

3.2 Optimized Encoding for Improved Performance

In order to efficiently solve instances of more than four alternatives, we need to streamline our initial encoding without weakening its logical and expressiv power. In this section, we present the three optimization techniques we found most effective.

3.2.1 Obvious Redundancy Elimination

A straightforward first step is to reduce the obvious redundancy within the axioms. As an example, consider the axiom of strategyproofness, where—in order to determine whether an outcome $Y = f(T')$ is preferred to an outcome $X = f(T)$—we consider *all* preference relations $R_\mu \supseteq T \setminus T'$. It suffices, however, if we stop after finding the first such preference relation with $Y\ P_\mu^{\mathcal{E}}\ X$ because then we already know that not both $Y = f(T')$ and $X = f(T)$ can be true.

Similarly, in many axioms, we can exclude considering symmetric pairs of objects (e.g., for functionality of the tournament choice function, there is no need to consider both pairs of sets (X, Y) and (Y, X)).

3.2.2 Canonical Tournaments

The main efficiency gain can be achieved by making use of the canonical isomorphism equality (see Section 3.1.1) during encoding. Recall that this condition states that for any tournament T the choice set $f(T)$ can be determined from the choice set $f(T_\mathfrak{c})$ of the corresponding *canonical tournament* $T_\mathfrak{c}$ by applying the respective canonical isomorphism π_T.

```
foreach Canonical tournament T_c do
    foreach Tournament T' do
        R_{T_c\T'} ← {R_μ | R_μ is a preference relation and R_μ ⊇ T_c \ T'};
        foreach Set X do
            foreach Set Y do
                boolean found ← false;
                foreach R_μ ∈ R_{T_c\T'} do
                    if ¬found ∧ setExt(R_μ, E).prefers(Y, X) then
                        variable_not(c(T_c, X));
                        variable_not(c(T'_c, π_{T'}^{-1}(Y)));
                        newClause();
                        found ← true;
```

Algorithm 1: $P^{\mathcal{E}}$-tournament-strategyproofness (optimized)

Therefore, it suffices to formulate the axioms on a single representative of each equivalence class of tournaments, in our case, the canonical tournament. The magnitudes in Table 1 illustrate that this formulation dramatically reduces the required number of variables, the size of the CNF formula, and the time required for encoding it.

In particular, in all axioms we can replace any outer quantifier $\forall T$ by a quantifier $\forall T_\mathfrak{c}$ that ranges over canonical tournaments only.[14] In the case of strategyproofness, however, there is a second tournament T' for which the restriction to canonical tournaments is potentially not strong enough to capture the full power of the axiom. We therefore keep T' as an arbitrary tournament but make sure that we only need variable symbols $c_{T'_\mathfrak{c},Y}$ for canonical tournaments in our CNF encoding. This can be achieved through the canonical isomorphism $\pi_{T'}$ since by Condition (2), $f(T') = Y$ if and only if $f(T'_\mathfrak{c}) = \pi_{T'}^{-1}(Y)$. The optimized encoding is shown in Algorithm 1.

Furthermore, since we no longer make any statements within the CNF formula about non-canonical tournaments, the canonical isomorphism equality condition becomes an "empty" condition and, thus, can be dropped from the encoding.

3.2.3 Approximation through Logically Related Properties

Approximation is a standard tool in SAT/SMT which can speed up the solving process. For instance, over-approximation can help find unsatisfiable instances faster by only solving parts of the full problem description in CNF. If this partial CNF formula is found to be unsatisfiable, any superset will also trivially be unsatisfiable. Since common manipulation instances in the literature require only one edge in a tournament to be reversed, one can, for instance, use over-approximation in the form of *single-edge-strategyproofness*, a slightly *weaker* variant of (tournament-)strategyproofness with $|T \setminus T'| = 1$.[15]

14. The tool nauty is capable of enumerating such non-isomorphic (i.e., canonical) tournaments.

15. While it was not obvious whether this condition is actually strictly weaker than tournament-strategyproofness, we identified Pareto-optimal SCFs that are Kelly-single-edge-strategyproof but not Kelly-tournament-strategyproof (cf. Section 4.1.1).

If the solver returns that there is no single-edge-strategyproof SCF that satisfies some set of properties Γ, we know immediately that there is also no strategyproof SCF that satisfies Γ. We used this form of approximation to prove the results in Remark 2.[16]

In a similar fashion, one can also apply *logically simpler* conditions, such as the ones by Brandt and Brill (2011), that are slightly stronger or weaker than $P^{\mathcal{E}}$-strategyproofness for specific set extensions $\mathcal{E}$ in order to *logically* under- or over-approximate problems, respectively. While these logically simpler conditions can help to further improve encoding and solving times, none of them were required to obtain the results presented in this paper.

Another way to over-approximate our problems is to restrict the domain of the SCF (e.g., by random sampling), which we explore in somewhat more detail when extracting small proofs in Section 5.1.1.

3.3 Finding Refinements through Incremental Solving

In order to obtain results for most refined (i.e., inclusion-minimal) or otherwise minimal SCFs, it will be important to also produce this property to the SAT solver in a satisfactory way. Generally, since the task of a SAT solver is to generate only one satisfying assignment, it does not necessarily output the finest SCF to satisfy a given set of properties. Through iterated or incremental solving, however, we can force the SAT solver to generate progressively finer or simply different SCFs that satisfy a set of desired properties.[17] For refinements, this can be achieved by adding clauses which encode that the desired SCF must be (strictly) finer than previously found solution (see, e.g., the formulation in Section 3.1.2). When the finest SCF with the desired properties has been found, adding these clauses leads to an unsatisfiable formula, which the SAT solver detects and therefore verifies the minimality of the solution.

With this final solving step, we have the main tools at hand required for our results, the most significant ones of which we describe in the next section.

4. Results and Discussion

Here we present our two main findings:

- There exists a strict refinement of *BP* which is $P^{\mathbf{K}}$-strategyproof (Theorem 2).
- For majoritarian SCFs with $m \geq 5$, $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality are incompatible (Theorem 3). For $m < 5$, *UC* satisfies $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality.

Further minor results are mentioned in the discussions proceeding the proofs and in Section 4.2.1.

16. While for $m = 7$ approximation was required to reach the result, it also enabled a speed-up for smaller instances: the running time for $m = 6$, for example, was reduced from almost five hours to three minutes.
17. Note that finding a refinement of an SCF is *not* equivalent to finding a smaller/minimal model in the SAT sense; in our encoding all assignments have the same number of satisfied variables.

4.1 Minimal Kelly-Strategyproof SCFs

Brandt (2015) showed that every coarsening f of a $P^{\mathbf{K}}$-strategyproof SCF f' is $P^{\mathbf{K}}$-strategyproof if $f(R) = f'(R)$ whenever $|f'(R)| = 1$. Thus, it is an interesting question to identify finest (or *inclusion-minimal*) $P^{\mathbf{K}}$-strategyproof SCFs.

While previous results suggested that BP could be a—or even *the*—finest majoritarian SCF which satisfies $P^{\mathbf{K}}$-strategyproofness, we first provide a counterexample to these assertions using $m = 5$ alternatives, and second show that also for larger domain sizes there exist majoritarian refinements of BP that are still $P^{\mathbf{K}}$-strategyproof and return significantly smaller choice sets than BP.

Theorem 2. *There exists a majoritarian Condorcet extension that refines BP and is still $P^{\mathbf{K}}$-strategyproof. As a consequence, BP is not even* a *finest majoritarian Condorcet extension satisfying $P^{\mathbf{K}}$-strategyproofness.*

Proof. Within seconds our implementation finds a satisfying assignment for $m = 5$ and the encoding of the explicit axioms *refinement of BP* (implies *Condorcet extension*) and $P^{\mathbf{K}}$-*strategyproofness*. The corresponding majoritarian SCF can be decoded from the assignment and is defined like BP with the exception depicted in Figure 4. $\square$

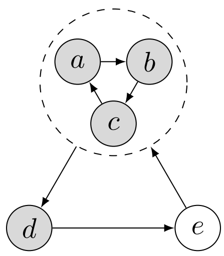


Figure 4: Tournament on which a $P^{\mathbf{K}}$-strategyproof refinement of BP is possible. $C := \{a, b, c\}$ represents a component in the sense that for all of its elements $x \in C$ it holds that $x\ P_M\ d$ and $e\ P_M\ x$. While BP chooses the whole set A on this tournament, the refined solution selects $\{a, b, c, d\}$ only.

Using the technique described in Section 3.3, we furthermore confirmed that the obtained SCF is the only refinement of BP on five alternatives which is still $P^{\mathbf{K}}$-strategyproof. Note, however, that it does not satisfy the (natural, but strong) property of *composition-consistency* (see, e.g., Laslier, 1997). Thus, it remains open whether BP might be characterized as *an*—or even *the*—inclusion-minimal, $P^{\mathbf{K}}$-strategyproof, and composition-consistent majoritarian SCF.[18]

While we were not able to resolve this open problem completely, we proved the following statements by extending our approach to also cover composition-consistency. BP is *an* inclusion-minimal, $P^{\mathbf{K}}$-strategyproof, and composition-consistent majoritarian SCF

18. Although already on the domain of up to five alternatives there are further inclusion-minimal, $P^{\mathbf{K}}$-strategyproof, and composition-consistent Condorcet extensions, which we could find using the computer-aided method, these counterexamples might not extend to larger domains.

for $m \leq 5$.[19] For $m \leq 7$, BP is *an* inclusion-minimal majoritarian SCF satisfying *set-monotonicity*[20] and composition-consistency. While this result might extend to larger instances, it only holds for at most 5 alternatives that these properties *uniquely* characterize BP.

If we, however, drop composition-consistency again, we can find multiple inclusion-minimal majoritarian SCFs that are refinements of BP and still $P^{\mathbf{K}}$-strategyproof. Interestingly, some of these SCFs turn out to be more discriminating than others in the sense that on average they yield significantly smaller choice sets. In the following section we are going to search for such discriminating SCFs and analyze the average size of their respective choice sets.

4.1.1 Finding Discriminating Kelly-Strategyproof SCFs

Many $P^{\mathbf{K}}$-strategyproof tournament solutions have been criticized for not being discriminating enough. It is known, for instance, that in large random tournaments, TC and UC select all alternatives with probability approaching 1 (Scott & Fey, 2012), while BP selects exactly half of the alternatives on average for any fixed number of alternatives (Fisher & Reeves, 1995). More discriminating tournament solutions, on the other hand, such as the *Copeland*, *Markov*, and *Slater* rules violate $P^{\mathbf{K}}$-strategyproofness. Using the computer-aided approach, we search for the most discriminating majoritarian SCFs that satisfy $P^{\mathbf{K}}$-strategyproofness. Though this is in the spirit of automated mechanism design (see, e.g., Conitzer & Sandholm, 2002), we apply these techniques mostly to improve our understanding of $P^{\mathbf{K}}$–strategyproofness and related axioms rather than to propose the generated tournament solutions for actual use.

As a measure for the discriminating power of majoritarian SCFs, we use the *average relative size* $\text{avg}(f)$ of the choice sets returned by an SCF f. Formally we define

$$\text{avg}(f) := \frac{1}{|A| \cdot |\mathcal{T}|} \sum_{T \in \mathcal{T}} |f(T)|,$$

where $\mathcal{T}$ is the set of all labeled tournaments on $|A| = m$ alternatives. We call an SCF f more discriminating than another SCF g if $\text{avg}(f) < \text{avg}(g)$. Given a set of axioms Γ, we try to find a *most discriminating* SCF f (i.e., with the minimal value for $\text{avg}(f)$) such that f satisfies the axioms in Γ.

While in theory it would be possible to just encode the relevant axioms and then enumerate all SCFs with the required properties by incrementally applying Axiom (6), the number of such SCFs is usually much too large. If we instead refine the initial solution further and further by applying Axioms (5) and (6) as indicated in Section 3.3, we will find an inclusion-minimal SCF, but not necessarily a most discriminating SCF f. We thus proceed via Algorithm 2, which is guaranteed to find a most discriminating SCF f without enumerating *all* candidates of SCFs. The algorithm starts by constructing an initial candidate of an SCF which satisfies the required axioms, iteratively refines it as much as possible (via the conjunction of Axioms (5) and (6)), and then encodes an additional axiom stating

19. For $m = 6$ we can already find a refinement with the same properties.
20. Set-monotonicity postulates that the choice set is invariant under the weakening of unchosen alternatives; it implies $P^{\mathbf{K}}$-strategyproofness (Brandt, 2015).

that all future solutions must yield a choice set with strictly smaller cardinality for at least one tournament T (Axiom (7)). The algorithm then repeats the refinement and encoding process until no further solution can be found. Since Axiom (7) is a necessary condition for $\text{avg}(f) < \text{avg}(g)$, we can be sure that a finest SCF f is returned.

```
SCF  smallestSolution ← null;
CNF  minimalRequirements ← encodeAxioms();
minimalRequirements ← preprocess(minimalRequirements); // optional
while isSatisfiable(minimalRequirements) do
    CNF currentRequirements ← minimalRequirements;
    SCF currentSolution ← solve(currentRequirements);
    while canBeRefined(currentSolution) do
        Append Axioms (5) and (6) to currentRequirements with g = currentSolution;
        currentSolution ← solve(currentRequirements);
    // an inclusion-minimal solution has been found
    if avgSize(currentSolution) < avgSize(smallestSolution) then
        smallestSolution ← currentSolution;
    Append Axiom (7) to minimalRequirements with g = currentSolution;
return smallestSolution;
```

Algorithm 2: A search algorithm to find a cardinality-minimal SCF f (i.e., with minimal value for $\text{avg}(f)$) that satisfies a given set of axioms. As a reminder, Axioms (5) and (6) encode a strict refinement of g; Axiom (7) encodes $|f(T)| < |g(T)|$ for some tournament T.

Preprocessing is generally optional in Algorithm 2; for $m = 6$ we, however, had to use unit propagation in order to reduce the size of the resulting SAT instance.[21] Note that the optimization techniques as described in Section 3.2 (in particular, canonical tournaments) can also be applied here.

The results of our analysis are exhibited in Figure 5. While on up to four alternatives all axioms under consideration lead to the same minimal size of $\text{avg}(f)$, on larger domains, $P^{\mathbf{K}}$-strategyproofness allows for smaller choice sets than BP (e.g., 45% instead of 50% of the alternatives for $m = 6$). Interestingly, the gap between BP and these more discriminating SCFs that satisfy $P^{\mathbf{K}}$-strategyproofness is not extraordinarily large; in particular, moving from $P^{\mathbf{K}}$-strategyproofness to $P^{\mathbf{K}}$-single-edge-strategyproofness allows for a more sizable reduction of $\text{avg}(f)$. For the related property of Kelly-participation, Brandl et al. (2015) remarked that the average size of choice sets can be reduced by almost 50% compared to BP, which supports the intuition that participation is a “weaker” property than strategyproofness (even though logically the two are independent).

BP and set-monotonicity yield the exact same values of $\text{avg}(f)$ for $m \leq 6$, which is somewhat surprising as we found SCFs that are not coarsenings of BP and are yet set-monotonic on this domain size. These SCFs, however, have no set-monotonic refinements that are more discriminating than BP. Interestingly, this does not generalize to larger

21. For the case of Kelly-strategyproofness, unit propagation and deletion of duplicate clauses reduced the CNF formula from about 600 million to just below three million clauses.

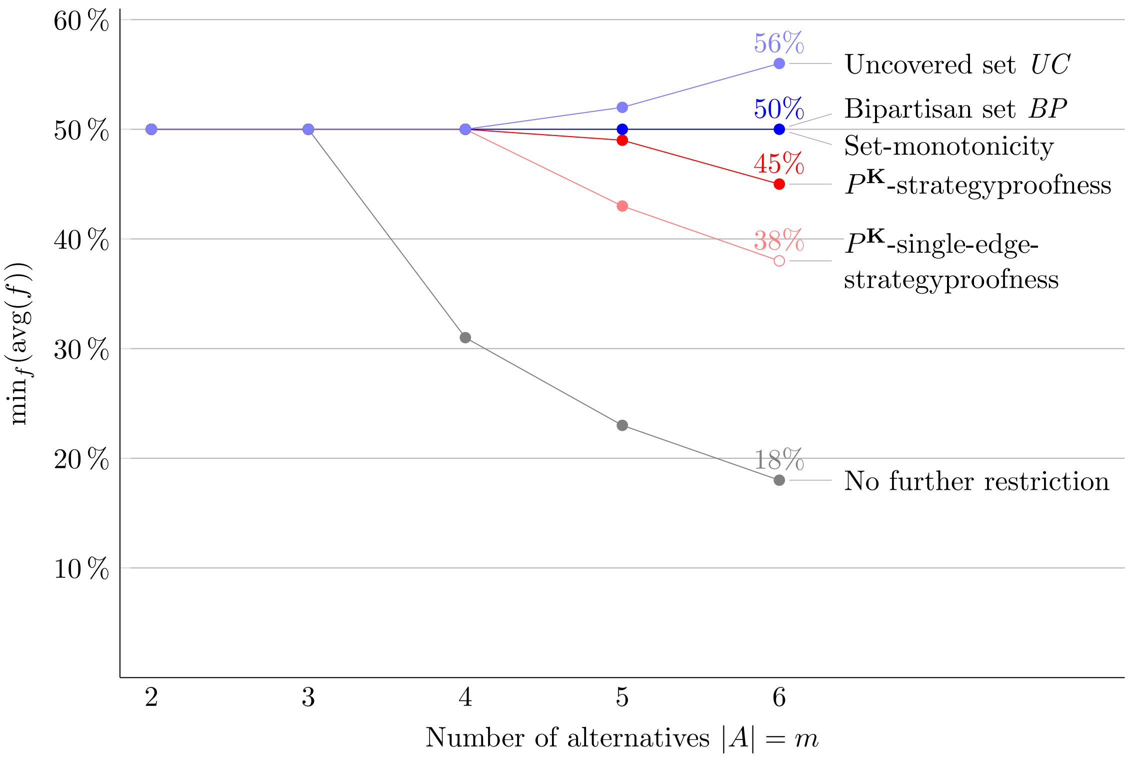


Figure 5: A comparison of the minimal values (rounded) of avg(f) for majoritarian, Pareto-optimal SCFs f that satisfy the given axioms (e.g., $P^{\mathbf{K}}$-strategyproofness). Interestingly, the values for *set-monotonicity* are identical to the ones for BP. Non-solid dots represent upper bounds, i.e., cases where we could only compute an SCF f with this value of avg(f) but have no guarantee that it is indeed minimal.

domains since we found a most discriminating majoritarian SCF f for $m = 7$ that satisfies set-monotonicity and Pareto optimality while only selecting 49.73% of the alternatives on average.

As more demanding axioms usually lead to larger choice sets (for instance, the SCF that always returns all alternatives trivially satisfies many axioms), one might view the minimal value of $\text{avg}(f)$ as an attempt to "quantify" the strength of an axiom. We leave a more detailed study of such a quantification as future work.

4.2 Incompatibility of Fishburn-Strategyproofness and Pareto-Optimality

In order to prove our main result on the incompatibility of Pareto-optimality and $P^{\mathbf{F}}$-strategyproofness we first show the following lemma, which establishes that, for majoritarian SCFs, the notion of Pareto-optimality is equivalent to being a refinement of the uncovered set (UC).[22]

Lemma 2. *A majoritarian SCF f is Pareto-optimal if and only if it is a refinement of UC.*

Proof. It is well-known, and was already observed by Fishburn (1977), that UC is Pareto-optimal, which implies that all its refinements are also Pareto-optimal.

For the direction from left to right, let f be a Pareto-optimal majoritarian SCF and T an arbitrary tournament. It suffices to show that $f(T)$ can never contain a covered alternative (since then $f(T) \subseteq UC(T)$ contains uncovered alternatives only). So let b be an alternative that is covered by another alternative a. We are going to construct a preference profile R which has T as its (strict) majority relation and in which b is Pareto-dominated by a. Together with the Pareto-optimality of f this implies that $b \notin f(T)$. We use a variant of the well-known construction by McGarvey (1953), but for triples rather than pairs of alternatives. Note that for each voter we need to ensure that he strictly prefers a to b in order to obtain the desired Pareto-dominance of a over b. Starting with an empty profile, for each alternative $x \notin \{a, b\}$ we add two voters R_{x_1}, R_{x_2} to the profile. These two voters are defined depending on how x is ranked relative to a and b in order to establish the edges between a, x and b, x. Note that since $x\ T\ a$ implies $x\ T\ b$ (because of $a\ C\ b$), edge (a, b) cannot be contained in a three-cycle with x and, thus, forms a transitive triple with x.

- Case 1: $x\ T\ a$ (implies $x\ T\ b$)
  $R_{x_1}:\quad x, a, b, v_1, \dots, v_{m-3};\qquad R_{x_2}:\quad v_{m-3}, \dots, v_1, x, a, b$
- Case 2a: $a\ T\ x$ and $x\ T\ b$
  $R_{x_1}:\quad a, x, b, v_1, \dots, v_{m-3};\qquad R_{x_2}:\quad v_{m-3}, \dots, v_1, a, x, b$
- Case 2b: $a\ T\ x$ and $b\ T\ x$
  $R_{x_1}:\quad a, b, x, v_1, \dots, v_{m-3};\qquad R_{x_2}:\quad v_{m-3}, \dots, v_1, a, b, x$

Here $v_1, \dots, v_{m-3}$ denotes an arbitrary enumeration of the $m-3$ alternatives in $A \setminus \{a, b, x\}$.

In all cases, the two voters cancel out each other for all pairwise comparisons other than (a, b), (x, a) and (x, b). For each of the remaining edges $(y, z) \in T$ (with $\{y, z\} \cap \{a, b\} = \emptyset$)

22. A stronger version of this lemma was shown by Brandt, Geist, and Harrenstein (2016b).

we further add two voters (now even closer to the construction by McGarvey.)

$$R_{(y,z)_1}: \quad y, z, a, b, v_1, \ldots, v_{m-4} \quad \text{and}$$
$$R_{(y,z)_2}: \quad v_{m-4}, \ldots, v_1, a, b, y, z,$$

which together establish edge (y, z), reinforce (a, b) and cancel otherwise. Note that in order to achieve an odd number of voters, an arbitrary voter can be added without changing the majority relation (as all edges had a weight of at least two so far). This completes the construction of a preference profile R which has T as its (strict) majority relation and in which b is Pareto-dominated by a. $\square$

To establish the full result (which does not admit a proof by counterexample as in Theorem 2) we—similarly to previous approaches—make use of an inductive argument.

Lemma 3. *For any set extension $\mathcal{E}$ that satisfies IIA, if there exists a majoritarian SCF f for $m+1$ alternatives that is $P^{\mathcal{E}}$-strategyproof and Pareto-optimal, then there also exists a majoritarian SCF f' for just m alternatives that satisfies these two properties.*

Proof. Let $f \subseteq UC$ be a majoritarian SCF for $m + 1 \geq 2$ alternatives that is $P^{\mathcal{E}}$-strategyproof. Then we define f_e to be the restriction of f to m alternatives based on tournaments in which alternative e is a Condorcet loser, i.e., an alternative x for which $(y, x) \in T$ for all $y \in A \setminus \{x\}$. In formal terms, define

$$f_e(T) := f(T^{+e}),$$

where T^{+e} is the tournament obtained from T by adding an alternative e as a Condorcet loser. This restriction of f is a well-defined SCF since alternative e cannot be contained in $f(T^{+e}) \subseteq UC(T^{+e}) = UC(T)$, where the last equation follows from the simple observation that the covering relation is unaffected by deleting Condorcet losers.

We now need to show that for some alternative e the restriction f_e is a majoritarian SCF that is $P^{\mathcal{E}}$-strategyproof and Pareto-optimal. Since this holds for any $e \in A$, we just pick e arbitrarily.

- **Majoritarian:** The fact that f_e is a majoritarian SCF carries over trivially from f.
- **$P^{\mathcal{E}}$-strategyproofness:** Assume for a contradiction that f_e is not $P^{\mathcal{E}}$-strategyproof. Then, by Theorem 1 there exist tournaments T and T' on m alternatives such that $f_e(T')\ P^{\mathcal{E}}_{\mu}\ f_e(T)$ with $R_\mu \supseteq T \setminus T'$. But since $f_e(T') = f(T'^{+e})$ and $f_e(T) = f(T^{+e})$ (and by the fact that $\mathcal{E}$ satisfies IIA), we get

  $$f(T'^{+e})\ P^{\mathcal{E}}_{\mu}\ f(T^{+e}),$$

  which contradicts $P^{\mathcal{E}}$-tournament-strategyproofness of f (as the two tournaments T'^{+e} and T^{+e} form a manipulation instance), and thus $P^{\mathcal{E}}$-strategyproofness.
- **Pareto-optimality:** By Lemma 2, this is equivalent to being a refinement of UC. Thus, let T be an arbitrary tournament on m alternatives and consider the following chain of set inclusions, which proves that $f_e \subseteq UC$:

  $$f_e(T) = f(T^{+e}) \subseteq UC(T^{+e}) = UC(T).$$

□

By virtue of Lemma 3 it now suffices to check the claim for the restricted domain of $m = 5$, which we do in the following lemma.

Lemma 4. *For exactly five alternatives (i.e., $m = 5$) there is no majoritarian SCF f that satisfies $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality.*

Proof. This base case of $m = 5$ alternatives was verified using our computer-aided approach, i.e., we checked that, with $|A| = 5$ alternatives, there is no satisfying assignment for an encoding of $P^{\mathbf{F}}$-tournament-strategyproofness (cf. Theorem 1) and being a refinement of UC (cf. Lemma 2), which the SAT solver confirmed within seconds. A human-readable proof of this claim has been extracted from the computer-aided approach and is presented in Section 5.1.2. □

Finally, this paper's main result regarding $P^{\mathbf{F}}$-strategyproofness follows directly from Lemmas 3 and 4.

Theorem 3. *For any number of alternatives $m \geq 5$ there is no majoritarian SCF f that satisfies $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality.*

Proof. We prove the statement inductively. The base case of $m = 5$ is covered by Lemma 4. For the induction step, we apply the contrapositive of Lemma 3 with $\mathcal{E} := \mathbf{F}$, which directly yields the desired results. □

While the number of voters required for this impossibility has been kept implicit so far, an upper bound of at most $m^2 - m - 1 = 19$ voters can be derived from the construction in the proof of Theorem 1. In Section 5 we will see, however, that a human-readable proof of Theorem 3 can be extracted, which only requires seven voters.

As a consequence of Theorem 3, virtually all common tournament solutions—except the top cycle (see Remark 2)—fail to be $P^{\mathbf{F}}$-strategyproof.

4.2.1 Remarks

Before we turn towards the technique of proof extraction, let us discuss some further insights regarding Theorem 3, which have been, to a large extent, enabled by the universality of the presented method.

Remark 1 (Strengthenings). It can be shown with the computer-aided method that Theorem 3 holds even without the assumption of neutrality. Since then, however, the optimizations based on canonical tournaments can no longer be used, extracted proofs (cf. Section 5) are much more complex and we therefore decided to present the result with neutrality here.[23]

The theorem can be further strengthened by additionally only requiring $P^{\mathbf{F}}$-single-edge-strategyproofness (cf. Section 3.2) or an even weaker variant of $P^{\mathbf{F}}$-strategyproofness where the manipulator is only allowed to swap two adjacent alternatives (see, e.g., Sato, 2013).

23. In addition, running times are much longer, which, however, is not a major concern given that not many conjectures had to be tested for this result.

Remark 2 (The Top Cycle TC). Note that Theorem 3 is not in conflict with the fact that TC is $P^{\mathbf{F}}$-strategyproof, as, for $m \geq 4$ alternatives, TC is strictly coarser than UC and therefore not Pareto-optimal. Possibly, TC is even the finest majoritarian Condorcet extension that satisfies $P^{\mathbf{F}}$-strategyproofness for $m \geq 5$. We were able to verify this for $5 \leq m \leq 7$ using our computer program. In the case of four alternatives, UC is a strict refinement of TC and (as our method shows) still $P^{\mathbf{F}}$-strategyproof. For $m = 8$ the time and space requirements appear to be prohibitive; already for $m = 7$ (despite all optimizations and approximations) encoding and solving the problem takes almost 24 hours, while for $m = 6$ it runs in about three minutes. It is not obvious whether an inductive argument can extend these verified instances to larger numbers of alternatives (as, for instance, such an induction step would require at least five alternatives).

Remark 3 (Other Preference Extensions). An advantage of the computer-aided approach is its universality. We can, for instance, very easily adapt the implementation to check set extensions other than the ones by Kelly and Fishburn.

Interestingly, our main result only relies on a small fraction of the power of the Fishburn extension: it suffices to only compare disjoint sets and sets that are contained in one another. In formal terms, the following set extension suffices for the impossibility:

$$X \; R_i^{\mathbf{F}^-} \; Y \text{ if and only if } \begin{cases} X \; R_i^{\mathbf{K}} \; Y & \text{when } X \cap Y = \emptyset, \\ X \; R_i^{\mathbf{F}} \; Y & \text{when } X \subseteq Y \text{ or } Y \subseteq X, \\ \bot & \text{otherwise.} \end{cases}$$

Actually, it would even suffice to only compare sets X and Y such that $|X \cap Y| \leq 3$.

We also checked a strengthening of the Fishburn extension: a voter prefers a set X to a set Y if X is better than Y under both optimistic and pessimistic expectations. Formally, $X \; R_i^{\mathbf{OP}} \; Y$ if and only if

$$\begin{aligned} &x \; R_i \; y \text{ for all } x \in X \text{ and some } y \in Y, \text{ and} \\ &y \; R_i \; x \text{ for all } y \in Y \text{ and some } x \in X. \end{aligned}$$

This extension is a weakening of both the optimistic and the pessimistic notions of strategyproofness in the Duggan-Schwartz Theorem (Duggan & Schwartz, 2000). In the majoritarian setting, $P^{\mathbf{OP}}$-strategyproofness leads to an analogous impossibility as in Theorem 3 for $m \geq 4$ already.

Remark 4 (Generality of Lemma 3). Note that the proofs of the individual properties within the inductive proof of Lemma 3 do only rely on the definition of f_e and stand independently of each other. Furthermore, it may be noted that Lemma 3 can even be shown for refinements of arbitrary majoritarian SCFs g whose choice set $g(T)$ does not shrink when Condorcet losers are removed from T (rather than Pareto-optimal majoritarian SCFs).

5. Proof Extraction

A major concern regarding computer-aided proofs is the difficulty of checking their correctness. While our implementation correctly confirmed a number of existing results and this can be considered as testing, some doubts about the correctness of new results naturally remain. Most SAT solvers offer some kind of proof trace, which can be checked by third-party-software. This, however, does not guarantee correctness of the encoding but only confirms the unsatisfiability of the corresponding CNF formula.

In this section, we show how *human-readable* proofs of the desired statements can be extracted from our approach, which can then be verified just as any manual mathematical proof. The general idea of this proof extraction technique lies in finding and analyzing a *minimal unsatisfiable core* (also referred to as a *minimal unsatisfiable set* (MUS)) of the SAT instance. An unsatisfiable core of a CNF formula is a subset of clauses that is already unsatisfiable by itself. If any subset of clauses of the unsatisfiable core is satisfiable, then the core is called minimal. In our case, the minimal unsatisfiable core contains information about the concrete instances of axioms that have to be employed to obtain an impossibility (e.g., manipulation instances, applications of Pareto optimality, etc). This information can be extracted in a straightforward way and reveals the structure and arguments of the proof.

We exemplify this technique in Section 5.1, in which we extract a human-readable proof of our main result (Theorem 3). In Section 5.2 we additionally enrich this proof by a set of *minimal* corresponding preference profiles, which then shows that the result of Theorem 3 holds for any setting with at least seven voters.

In general, extracting human-readable proofs serves two separate purposes. On the one hand, a human-readable proof can significantly raise confidence in the correctness of the results, basically by making verification of the *approach* obsolete since now the *results* themselves are directly verifiable. On the other hand, the extracted proofs sometimes provide additional insight into the problems via their arguments and structure. In our case, the number of voters required for the impossibility would not have been (easily) accessible directly.

5.1 A Human-Readable Proof of Theorem 3

In order to extract a human-readable proof of Theorem 3, or actually its main ingredient Lemma 4, we have to follow a series of three steps:

1. Obtain a suitable MUS of the CNF formula that encodes a $P^{\mathbf{F}}$-tournament-strategyproof refinement of UC on five alternatives
2. Decode the MUS into a human-readable format
3. Interpret the human-readable MUS to obtain a human-readable proof

While the first two steps are computer-aided and can be largely automated, step three requires some manual effort.

5.1.1 Obtaining a Suitable MUS of the CNF Formula

Extracting a minimal unsatisfiable core is a feature offered by a range of SAT solvers. In this paper, we use PicoMUS (part of PicoSAT, Biere, 2008) for this job.[24] It should be

24. Compiled with trace support in order to use core extraction in addition to clause selector variables. This significantly improves the size of the resulting MUS.

noted, however, that while an MUS is inclusion-minimal, it does not necessarily represent a *smallest* unsatisfiable set (i.e., with a minimal number of clauses or variables).[25]

As the number of clauses turned out to be a good proxy for proof complexity and length, we tried to find an MUS with a small number of clauses. When run on the complete, optimized SAT encoding as described in Section 3.2, PicoMUS returns an MUS with 55 clauses. This is already a massive reduction compared to more than three million clauses in the original problem instance, but we found an even smaller MUS with only *16 clauses* by randomly sampling sets of tournaments to be used instead of the full domain of all tournaments when generating our problem files. Another heuristic approach of considering "neighborhoods" of single tournaments (for instance, all tournaments that can be reached by changing at most two edges in the transitive tournament) yielded a less significant improvement with a total of 25 clauses.

While it seems natural that larger domains are generally better as they lead to the required impossibility more often than smaller domains, larger domains actually tend towards larger proofs and even miss very small proofs. For instance, for the domain size $s = 200$ (consisting of s labeled tournaments) no proof smaller than 18 clauses was found, while the same number of runs with $s = 50$ produced four proofs with just 16 clauses each.[26]

Therefore, in our setting, a medium-sized domain ($s = 50$ or $s = 100$ in our experiments) appears to be best suited. The complete results of running time and proof size analysis given different domain sizes s can be obtained from Figures 8 and 9 in Appendix C.

5.1.2 Decoding the MUS into a Human-Readable Format

The next step is to make the obtained MUS more accessible to humans. To this end, we first (automatically) add comments to the original CNF for each manipulation clause during its creation, and then select those comments that belong to clauses in the MUS. The comments contain witnesses for the manipulation instances found, i.e., information about the original tournament T, the manipulated tournament T', the respective choice sets $f(T)$ and $f(T')$, and the original preferences of the manipulator R_μ (compare Definition 3). Furthermore, any variable symbol can easily be decoded into the tournament and choice set it represents, which is helpful in particular for all non-manipulation clauses (orbit condition and Pareto-optimality).

The result of this step is presented in Figure 6, where each tournament is represented by a lower triangular representation of its adjacency matrix (see the proof of Lemma 4 in Section 5.1.3 for graphical representations).

5.1.3 Interpreting the MUS and Obtaining a Human-Readable Proof

From the witnessed MUS it is just a small step to a textual, human-readable proof. With a bit of practice, one can quickly understand the structure of the proof: it starts from the orbit condition in the first line and the refinement condition in the last line, which each

25. While the tool CAMUS by Liffiton and Sakallah (2008) is theoretically capable of finding a *smallest* MUS (with a minimal number of clauses), it did not terminate in a reasonable amount of time on our very large CNF instances.
26. In addition, medium-sized domains are more efficient regarding their running time per generated proof, which admittedly plays only a minor, but still important role given that the total running time for large domains is about 20 hours.

```
p cnf 341 16
218 231 232 233 234 247 248 0
−202 −330 0
c T: 1111111111 −> [e]; T': 1011100111 −> [d,e]; P_i: b,d,c,e,a
−233 −202 0
c T: 1101100111 −> [e]; T': 0010100111 −> [a]; P_i: b,c,d,a,e
−234 −202 0
c T: 1101100111 −> [a,e]; T': 0010100111 −> [a]; P_i: b,c,d,a,e
−218 −218 0
c T: 1101100111 −> [a]; T': 1001000100 −> [e]; P_i: e,c,a,d,b
−232 −232 0
c T: 1101100111 −> [a,b,c,d]; T': 1001000100 −> [a,c,d,e]; P_i: e,c,a,d,b
−248 −338 0
c T: 1101100111 −> [a,b,c,d,e]; T': 1100100101 −> [b,c,e]; P_i: b,e,c,d,a
−231 −202 0
c T: 1101100111 −> [b,c,d]; T': 1111111111 −> [e]; P_i: a,e,b,c,d
−247 −202 0
c T: 1101100111 −> [b,c,d,e]; T': 1111111111 −> [e]; P_i: a,e,b,c,d
−314 −314 0
c T: 1100101110 −> [c]; T': 1100100101 −> [e]; P_i: b,d,e,a,c
c T: 1100101110 −> [c]; T': 1100110110 −> [b]; P_i: b,c,d,e,a
−318 −318 0
c T: 1100101110 −> [d]; T': 1100100101 −> [b]; P_i: b,d,e,a,c
c T: 1100101110 −> [d]; T': 1100110110 −> [a]; P_i: b,c,e,a,d
−322 −322 0
c T: 1100101110 −> [c,d]; T': 1100110110 −> [a,b]; P_i: b,e,a,c,d
−326 −326 0
c T: 1100101110 −> [e]; T': 1100110110 −> [d]; P_i: b,c,d,e,a
−334 −202 0
c T: 1100101110 −> [d,e]; T': 1001111010 −> [d]; P_i: c,a,d,e,b
202 0
314 318 322 326 330 334 338 0
```

Figure 6: A version of the extracted MUS, in which all manipulation instances (here: binary clauses) have been decoded into a human-readable format: two mappings of tournaments (original T and manipulated T') to choice sets and the truthful preferences of the manipulator P_μ. This information covers all variables and thus suffices to also decode the remaining clauses.

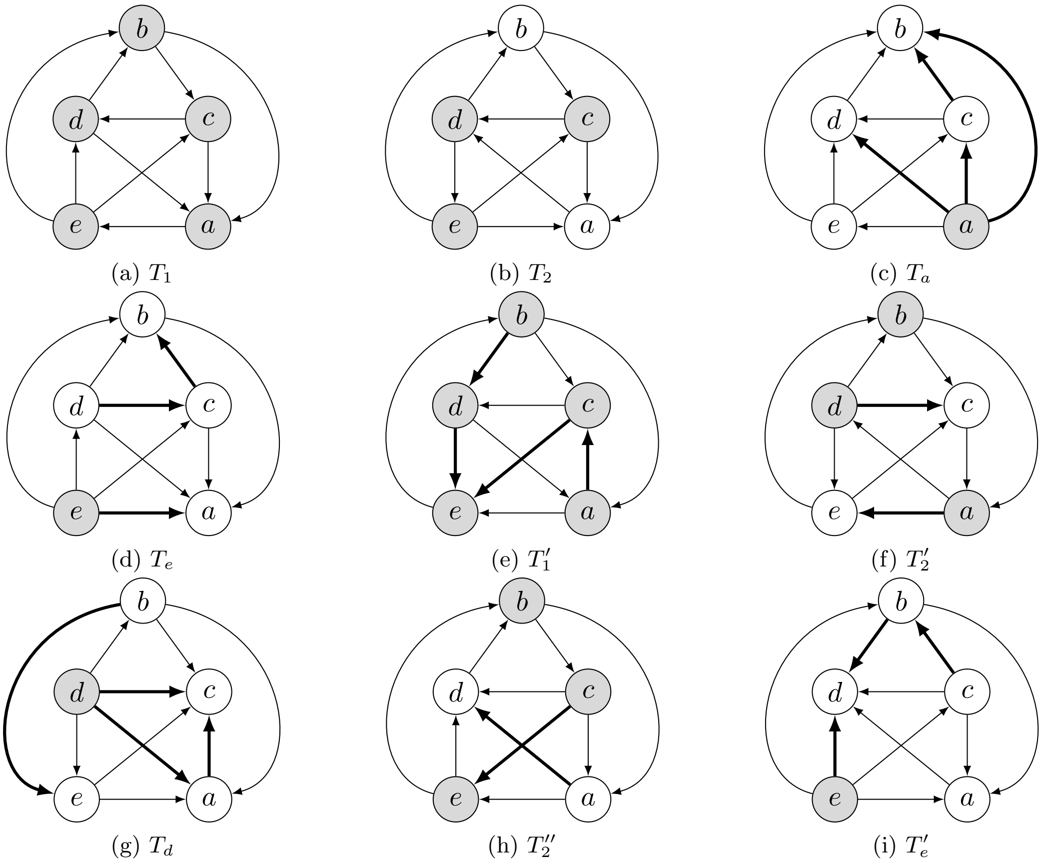
Truthful choice	Manipulated choice	Manipulator's preferences
$f(T_1) = \begin{cases} \{e\} & \lightning \\ \{a\} \cup \{e\} & \lightning \\ \{b,c,d\} & \lightning \\ \{b,c,d\} \cup \{e\} & \lightning \\ \{a\} & \lightning \\ \{a\} \cup \{b,c,d\} & \lightning \end{cases}$	$f(T_a) \subseteq UC(T_a) = \{a\}$ $f(T_e) \subseteq UC(T_e) = \{e\}$ $f(T_1') = \begin{cases} \{e\} & T_1' \text{ is iso-} \\ \{a,c,d,e\} & \text{morphic}^{27} \text{ to } T_1 \end{cases}$	b,c,d,a,e a,e,b,c,d e,c,a,d,b
$f(T_2) = \begin{cases} \{c\} & \lightning \\ \{d\} & \lightning \\ \{c,d\} & \lightning \\ \{e\} & \lightning \\ \{d,e\} & \lightning \end{cases}$	$f(T_2') = \begin{cases} \{b\} & \\ \{a\} & T_2' \text{ is iso-} \\ \{a,b\} & \text{morphic}^{27} \text{ to } T_2 \\ \{d\} & \end{cases}$ $f(T_d) \subseteq UC(T_d) = \{d\}$	b,e,a,c,d b,c,d,e,a c,a,d,e,b
$f(T_e') \subseteq UC(T_e') = \{e\}$	$f(T_2) = \{c,e\} \quad \lightning^{28}$	a,c,b,e,d

Table 2: Set of manipulation instances (one per line) to conclude that $f(T_1) = A = \{a,b,c,d,e\}$ and $f(T_2) = \{c,d,e\}$. Each of the truthful choices considered here leads to a $P^{\mathbf{F}}$-tournament-manipulation instance (a contradiction to the assumption of $P^{\mathbf{F}}$-tournament-strategyproofness). The tournaments are defined in Figure 7.

leave some (limited) possibilities for respective choice sets, and then excludes all possible choices one after another by suitable manipulation instances. The full proof runs as follows.

Proof of Lemma 4. For a contradiction, let f be a majoritarian SCF on $A = \{a,b,c,d,e\}$ that satisfies $P^{\mathbf{F}}$-strategyproofness and Pareto-optimality. Recall that, by Theorem 1, f is $P^{\mathbf{F}}$-tournament-strategyproof, too, and by Lemma 2 it has to be a refinement of UC (i.e., $f \subseteq UC$). Let furthermore T_1 and T_2 be the tournaments depicted in Figure 7. We proceed in three steps: first, we show that $f(T_1) = UC(T_1) = A$. Second, we argue that $f(T_2) = UC(T_2) = \{c,d,e\}$. And last, we prove that these two insights actually forms the basis of a manipulation instance, which leads to the desired contradiction.

Let us start with $f(T_1) = UC(T_1) = A$. First, note that since the alternatives $\{b,c,d\}$ form an orbit we know that either $\{b,c,d\} \subseteq f(T_1)$ or $\{b,c,d\} \cap f(T_1) = \emptyset$ (cf. Definition 5). We are going to exclude all remaining choice sets through $P^{\mathbf{F}}$-tournament-manipulation instances. As a first example, suppose $f(T_1) = \{e\}$. Then a voter with individual preferences $P_\mu : b,c,d,a,e$ could reverse the edges (b,a) and (b,c) in T_1 such that a transitive tournament T_a with Condorcet winner a results (which needs to be uniquely selected by f since $f \subseteq UC$). Since, however, $\{a\} \; P_\mu^{\mathbf{F}} \; \{e\}$, this contradicts $P^{\mathbf{F}}$-tournament-strategyproofness. The same example also works to exclude $f(T_1) = \{a,e\}$. Note how these arguments correspond to lines 5 to 8 of the extracted MUS in Figure 6. The (analogous) manipulation instances for all possible choice sets other than $A = \{a,b,c,d,e\}$ are given in Table 2 and Figure 7.



(a) T_1 (b) T_2 (c) T_a

(d) T_e (e) T_1' (f) T_2'

(g) T_d (h) T_2'' (i) T_e'

Figure 7: Tournaments which are required in the proof of Lemma 4. The uncovered sets are marked in grey; edges that have been (for T_e': will be) reversed by the manipulating voter (cf. Table 2) are depicted as thick edges. Note the proof would also succeed with less edge reversals in T_a, T_e, T_d, and T_e' (such that these tournaments only have Condorcet winners rather than being transitive). These transitive tournaments are isomorphic, however, and thus can be succinctly represented as the single clause 202 in the extracted MUS.

For $f(T_2) = UC(T_2) = \{c, d, e\}$, first observe that $f(T_2) \subseteq UC(T_2) = \{c, d, e\}$ and hence we only need to exclude any strict subset of $\{c, d, e\}$. Again we proceed by giving a possible manipulation instance for each of those subsets. The complete list is to be found in Table 2 and Figure 7. Observe how the last line in Table 2 excludes $f(T_2) = \{c, e\}$ by considering it as the *manipulated* choice for the (known) truthful choice $f(T_e') \subseteq UC(T_e') = \{e\}$.

As a last step, we provide a manipulation instance based on $f(T_1) = A$ and $f(T_2) = \{c, d, e\}$. For this, first observe that by renaming the alternatives we get $f(T_2'') = \{b, c, e\}$ and so the manipulation instance results from a voter with preferences $P_\mu' : b, e, c, d, a$. This

27. The isomorphisms are $\pi_1 = \begin{pmatrix} a & b & c & d & e \\ b & e & c & d & a \end{pmatrix}$ and $\pi_2 = \begin{pmatrix} a & b & c & d & e \\ d & c & a & e & b \end{pmatrix}$, respectively.

28. The SAT solver actually returned an isomorphic copy of this instance, which we restructured to improve readability.

voter can reverse the edges (d, a) and (e, c) in T_1 to create T_2'' and obtain the $P^{\mathbf{F}}$-preferred outcome $\{b, c, e\}$, a contradiction to the $P^{\mathbf{F}}$-strategyproofness of f. $\square$

Note that actually only the manipulation instance with $f(T_1) = \{a\} \cup \{b, c, d\}$ and $f(T_1') = \{a, c, d, e\}$ requires the Fishburn-extension; for the other instances the Kelly-extension suffices.

5.2 Number of Voters Required

In the previous parts of the paper we have taken advantage of the fact that our condition of *tournament-strategyproofness* abstracted away any reference to voters. It is interesting to ask, however, how many voters are at least required for the obtained impossibility of Theorem 3 to hold. The construction in the proof of Theorem 1 gives an implicit upper bound of $m^2 - m - 1 = 19$ voters, but this can be further improved to *seven* voters.

By slightly modifying the techniques described by Brandt, Geist, and Seedig (2014), we were able to (automatically) construct *minimal* preference profiles for all steps in Proof 5.1.3. While Brandt et al. (2014) provided a SAT-formulation of whether a given majority relation can be induced by a given number of voters, we extended this framework to include axioms for manipulation instances. In more detail, we re-used the axioms for *linear preferences* and majority implications, but added axioms for the *truthful preferences of the manipulator* and *majority implications for the manipulated profile.*

The profiles that we generated for all steps in the proof of Lemma 4 in Section 5.1.3 are given in Appendix D. The largest of these profiles contains seven voters, and all other profiles can easily be extended to seven voters by adding pairs of voters with opposite preferences. While this observation shows that seven is the smallest number of voters which can be achieved with our extracted proof, it remains open whether, by another proof, the number of voters can be further reduced below seven.

6. Conclusion

We have extended and applied computer-aided theorem proving based on SAT solving to extensively analyze Kelly- and Fishburn-strategyproof majoritarian SCFs. This has led to a range of results, both positive and negative. An important novel contribution of our work is the ability to extract a human-readable proof from negative SAT instances. This eliminates the need to verify the computer-aided method since impossibility results can directly be checked based on their human-readable proofs. Based on the ease of adaptation of the proposed method, we anticipate further insights to spring from the overall approach in the future. Apart from simply applying our system to further investigate strategyproofness, other potential applications related to our line of work include:

Unrestricted SCFs In order to reduce complexity, we have studied *majoritarian* SCFs only. The framework, however, is applicable in the same way to general SCFs, which "operate" on full preference profiles (rather than majority relations). The challenge then is to find a suitable representation of such preference profiles and potentially corresponding inductive arguments on the number of voters.

Further axioms Some preliminary experiments suggest that our technique can easily be applied to a range of properties other than strategyproofness, these deserve further investigation. In many cases it suffices to just formalize and implement the additional axioms. Of particular interest could be such properties that link the behavior of SCFs for different domain sizes. As initial steps in this direction, we were able to extend the approach to cover the property of *participation* (Brandl et al., 2015; Brandt et al., 2016c) as well as a weak version of *composition-consistency* (cf. Section 4.1).

Smallest number of voters required As mentioned in Section 5.2, Theorem 3 holds for any number of voters $n \geq 7$, but it is not known whether this number is minimal. One could adapt proof extraction as presented in Section 5 to search for a smallest proof in the number of voters, rather than in the number of clauses, to settle this question.

Generalization of the inductive argument It appears reasonable to investigate whether the inductive argument of Lemma 3 can be further generalized to a whole class of properties/axioms, ideally based on their logical form. As in the work of Geist and Endriss (2011), this would then enable an automated search for further theorems about SCFs.

Apart from these concrete ideas, applications of the general approach can be envisioned in many areas of theoretical economics.

Acknowledgments

This material is based upon work supported by Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/9-1. The paper benefitted from discussions at the COST Action Meeting IC1205 on Computational Social Choice (Maastricht, 2014), the 13th International Conference on Autonomous Agents and Multiagent Systems (Paris, 2014), the 5th International Workshop on Computational Social Choice (Pittsburgh, 2014), and the Dagstuhl Seminar on Computational Social Choice: Theory and Applications (Dagstuhl, 2015). The authors in particular thank Jasmin Christian Blanchette, Markus Brill, Hans Georg Seedig, and Bill Zwicker for helpful discussions and their support, and three anonymous reviewers for their valuable comments and suggestions to improve the paper.

Appendix A. Proof of Lemma 1

We first show that the orbit condition is equivalent to a statement about automorphisms:

Lemma 5. *Let f be a tournament choice function. Then the following statement is equivalent to the orbit condition:*

$$\alpha(f(T_\mathfrak{c})) = f(T_\mathfrak{c}) \text{ for all canonical tournaments } T_\mathfrak{c} \text{ and their automorphisms } \alpha. \quad (8)$$

Proof. Let f be a tournament choice function and $T_\mathfrak{c}$ a canonical tournament. For the direction from left to right, let furthermore $O \in \mathcal{O}_{T_\mathfrak{c}}$ an orbit on $T_\mathfrak{c}$. Now pick two alternatives $a, b \in O$. We show that either both alternatives are chosen by f or neither one is. Since a and b are in the same orbit, there must be an automorphism α on $T_\mathfrak{c}$ for which $\alpha(a) = b$. Observe that $a \in f(T_\mathfrak{c})$ if and only if $b \in \alpha(f(T_\mathfrak{c}))$ if and only if $b \in f(T_\mathfrak{c})$, where the last step is an application of Condition (8).

For the converse, let α be an automorphism on $T_{\mathfrak{c}}$, pick an arbitrary alternative $a \in A$ and consider its inverse image $\alpha^{-1}(a) =: b$. Since a and b are in the same orbit, it holds by the orbit condition that $a \in f(T_{\mathfrak{c}})$ if and only if $b \in f(T_{\mathfrak{c}})$. Furthermore, as $\alpha(b) = a$ we get that $a \in f(T_{\mathfrak{c}})$ if and only if $a \in \alpha(f(T_{\mathfrak{c}}))$. Thus, $f(T_{\mathfrak{c}}) = \alpha(f(T_{\mathfrak{c}}))$, which is what we wanted to prove. □

Next we prove a general statement about how to split any isomorphism into a canonical isomorphism and an automorphism.

Lemma 6. *Any isomorphism $\pi : T_{\mathfrak{c}} \to T$ can be decomposed into the canonical isomorphism π_T and an automorphism $\alpha : T_{\mathfrak{c}} \to T_{\mathfrak{c}}$. I.e., for any isomorphism $\pi : T_{\mathfrak{c}} \to T$ there is an automorphism $\alpha : T_{\mathfrak{c}} \to T_{\mathfrak{c}}$ such that $\pi = \pi_T \circ \alpha$.*

Proof. Define $\alpha : T_{\mathfrak{c}} \to T_{\mathfrak{c}}$ by setting $\alpha := \pi_T^{-1} \circ \pi$. Since inverses and compositions of isomorphisms are themselves isomorphisms, it follows directly that α is an automorphism. Furthermore, $\pi_T \circ \alpha = \pi_T \circ \left(\pi_T^{-1} \circ \pi\right) = \left(\pi_T \circ \pi_T^{-1}\right) \circ \pi = \pi$. □

Lemmas 5 and 6 together can then be used to prove Lemma 1:

Lemma 1. *For any tournament choice function, neutrality is equivalent to the conjunction of the orbit condition and canonical isomorphism equality.*

Proof. Let f be a tournament choice function and first note that by Lemma 5 we might use Condition (8) rather than the orbit condition. Therefore, the direction from left to right is trivially true.

For the direction from right to left, we first only show that canonical isomorphism equality (2) together with Condition (8) implies neutrality *for canonical tournaments*: So let $T_{\mathfrak{c}}$ be a canonical tournament, π a permutation and define $T' := \pi(T_{\mathfrak{c}})$. By Lemma 6, we can decompose the isomorphism $\pi : T_{\mathfrak{c}} \to T'$ such that $\pi = \pi'_T \circ \alpha$ for some automorphism α on $T_{\mathfrak{c}}$. Then the following chain of equalities holds, which proves the claim for canonical tournaments:

$$f(\pi(T_{\mathfrak{c}})) \;=\; f(T') \;\overset{(2)}{=}\; \pi_{T'}(f(T'_{\mathfrak{c}})) \;=\; \pi_{T'}(f(T_{\mathfrak{c}})) \;\overset{(8)}{=}\; \pi_{T'}(\alpha(f(T_{\mathfrak{c}}))) \;=\; \pi(f(T_{\mathfrak{c}})). \quad (9)$$

For arbitrary tournaments T and permutations π, we write T as $\pi_T(T_{\mathfrak{c}})$ and obtain

$$f(\pi(T)) = f(\pi(\pi_T(T_{\mathfrak{c}}))) = f((\pi \circ \pi_T)(T_{\mathfrak{c}})),$$

which, since $T_{\mathfrak{c}}$ is canonical, is equal to

$$(\pi \circ \pi_T)(f(T_{\mathfrak{c}}))) = \pi(\pi_T(f(T_{\mathfrak{c}}))) \overset{(2)}{=} \pi(f(T))$$

by Condition (9). This finishes the proof. □

Appendix B. Pseudo-Code for Encoding

We present (as an illustrative example) the simple pseudo-code of Algorithm 3 to generate the CNF form of Axiom 1 (functionality of the tournament choice function; cf. Section 3.1.1).

```
foreach Tournament T do
    foreach Set X do
        variable(c(T, X));
    newClause();
    foreach Set Y do
        foreach Set Z ≠ Y do
            variable_not(c(T, Y));
            variable_not(c(T, Z));
            newClause();
```

Algorithm 3: Functionality of the tournament choice function

Appendix C. MUS Search Analysis (Running Time and Size of MUS)

In this appendix, we present the complete results of the running time (Figure 8) and MUS size (measured in number of clauses; Figure 9) analyses given different sizes s of randomly sampled domains. In our setting, sizes of $s = 50$ or $s = 100$ appear to offer good results both in terms of running time and actually finding small proofs.

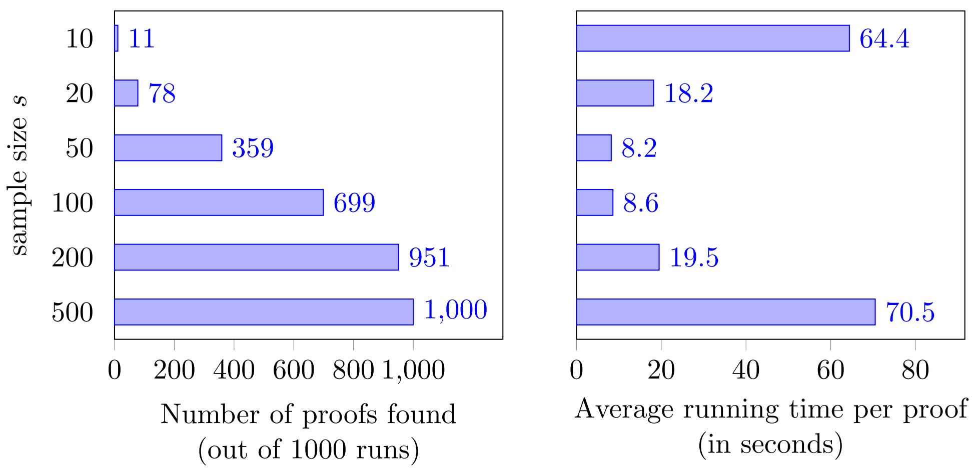


Figure 8: Number of unsatisfiable instances (i.e., proofs found) and running time results under heuristics with different numbers s of sampled tournaments (labeled, 1000 runs).

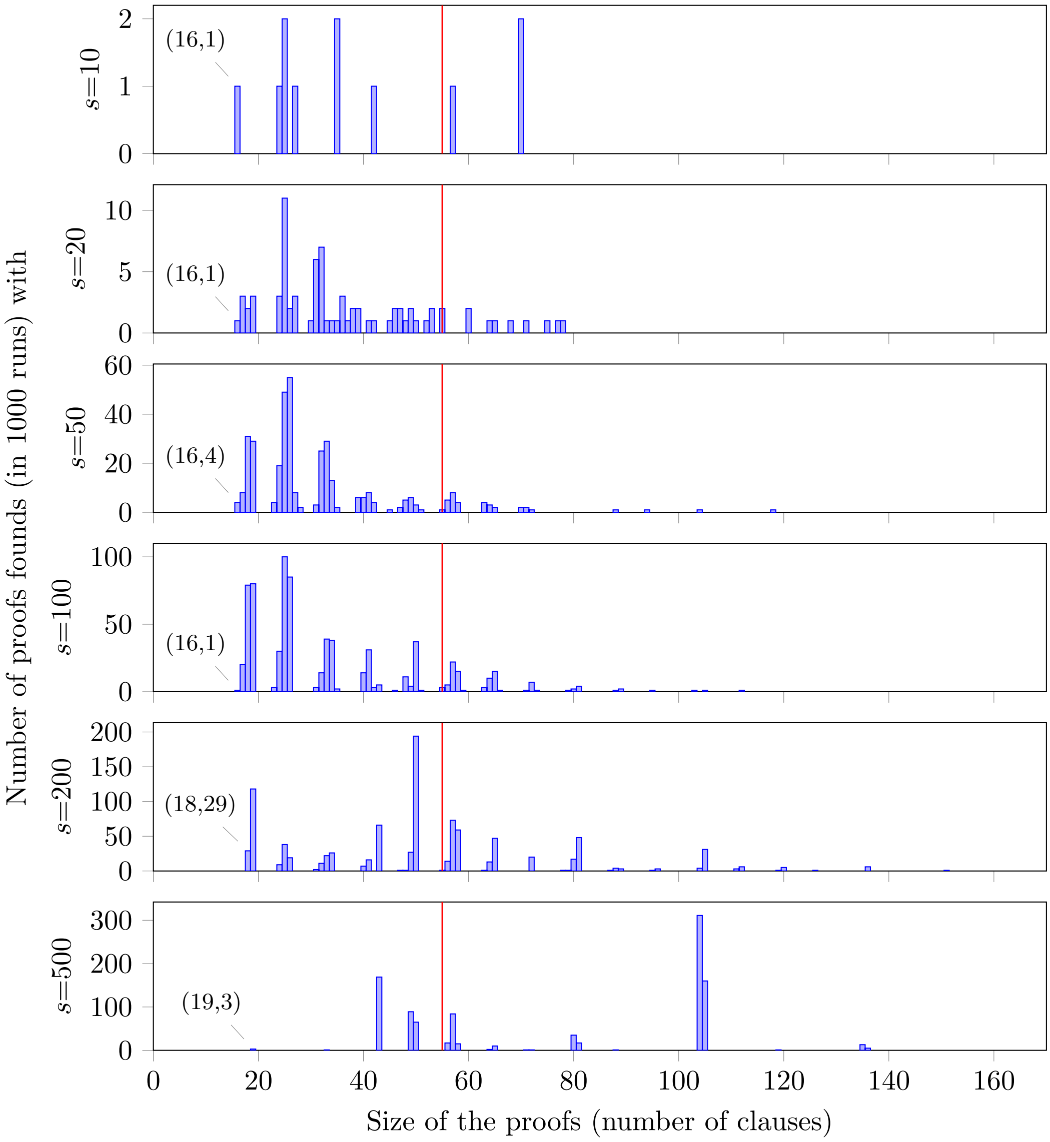


Figure 9: The sizes of MUSes (proofs) under heuristics with different numbers s of sampled tournaments (labeled). The size of the MUS obtained from running on the full domain is indicated by a red line. For improved readability, the size and multiplicity of the smallest MUS is explicitly listed.

Appendix D. Profiles for the Extracted Proof of Theorem 3

Here we display the MUS of Figure 6 enriched with *minimal* preference profiles for each step in the proof of Theorem 3. The profiles were generated and checked for minimality on a computer (and using a SAT solver) in less than a second each.

```
p cnf 341 16
218 231 232 233 234 247 248 0
Agent 0: b,c,d,a,e
Agent 1: a,e,c,d,b
Agent 2: e,d,b,c,a

-202 -330 0
c T: 1111111111 -> [e]; T': 1011100111 -> [d,e]; P_i: b,d,c,e,a
Agent 0: b,d,c,e,a
Agent 1: c,d,a,e,b
Agent 2: e,c,d,b,a
Agent 3: a,e,d,c,b
Agent 4: e,d,b,a,c
Manipulated preferences of agent 0: b,a,c,d,e

-233 -202 0
c T: 1101100111 -> [e]; T': 0010100111 -> [a]; P_i: b,c,d,a,e
-234 -202 0
c T: 1101100111 -> [a,e]; T': 0010100111 -> [a]; P_i: b,c,d,a,e
Agent 0: b,c,d,a,e
Agent 1: a,e,d,b,c
Agent 2: e,c,d,b,a
Manipulated preferences of agent 0: a,c,b,d,e

-218 -218 0
c T: 1101100111 -> [a]; T': 1001000100 -> [e]; P_i: e,c,a,d,b
-232 -232 0
c T: 1101100111 -> [a,b,c,d]; T': 1001000100 -> [a,c,d,e]; P_i: e,c,a,d,b
Agent 0: e,c,a,d,b
Agent 1: d,a,e,b,c
Agent 2: d,a,e,b,c
Agent 3: d,e,b,c,a
Agent 4: c,e,b,d,a
Agent 5: b,c,a,e,d
Agent 6: b,a,c,e,d
Manipulated preferences of agent 0: b,a,c,d,e

-248 -338 0
c T: 1101100111 -> [a, b, c, d, e]; T': 1100100101 -> [b, c, e]; P_i: 1 > 4 >
    2 > 3 > 0
Agent 0: 1 > 4 > 2 > 3 > 0
Agent 1: 2 > 3 > 0 > 4 > 1
Agent 2: 2 > 4 > 3 > 1 > 0
Agent 3: 0 > 4 > 3 > 1 > 2
Agent 4: 1 > 0 > 4 > 2 > 3
Manipulated preferences of agent 0:
1 > 2 > 0 > 4 > 3

-231 -202 0
c T: 1101100111 -> [b, c, d]; T': 1111111111 -> [e]; P_i: 0 > 4 > 1 > 2 > 3
```

```
−247 −202 0
c T: 1101100111 −> [b, c, d, e]; T': 1111111111 −> [e]; P_i: 0 > 4 > 1 > 2 > 3
Agent 0: 0 > 4 > 1 > 2 > 3
Agent 1: 3 > 1 > 2 > 0 > 4
Agent 2: 4 > 2 > 3 > 1 > 0
Manipulated preferences of agent 0:
4 > 0 > 3 > 2 > 1

−314 −314 0
c T: 1100101110 −> [c]; T': 1100100101 −> [e]; P_i: 1 > 3 > 4 > 0 > 2
Agent 0: 1 > 3 > 4 > 0 > 2
Agent 1: 4 > 3 > 1 > 2 > 0
Agent 2: 4 > 1 > 2 > 0 > 3
Agent 3: 2 > 0 > 3 > 4 > 1
Agent 4: 2 > 0 > 3 > 4 > 1
Manipulated preferences of agent 0:
1 > 2 > 0 > 4 > 3
c T: 1100101110 −> [c]; T': 1100110110 −> [b]; P_i: 1 > 2 > 3 > 4 > 0
Agent 0: 1 > 2 > 3 > 4 > 0
Agent 1: 0 > 3 > 4 > 2 > 1
Agent 2: 0 > 4 > 2 > 3 > 1
Agent 3: 4 > 1 > 2 > 0 > 3
Agent 4: 3 > 4 > 1 > 2 > 0
Manipulated preferences of agent 0:
3 > 1 > 2 > 0 > 4

−318 −318 0
c T: 1100101110 −> [d]; T': 1100100101 −> [b]; P_i: 1 > 3 > 4 > 0 > 2
c T: 1100101110 −> [d]; T': 1100110110 −> [a]; P_i: 1 > 2 > 4 > 0 > 3
Agent 0: 1 > 2 > 4 > 0 > 3
Agent 1: 3 > 4 > 1 > 2 > 0
Agent 2: 4 > 0 > 2 > 3 > 1
Agent 3: 2 > 0 > 3 > 4 > 1
Agent 4: 1 > 0 > 3 > 4 > 2
Manipulated preferences of agent 0:
3 > 1 > 2 > 0 > 4

−322 −322 0
c T: 1100101110 −> [c, d]; T': 1100110110 −> [a, b]; P_i: 1 > 4 > 0 > 2 > 3
Agent 0: 1 > 4 > 0 > 2 > 3
Agent 1: 2 > 0 > 3 > 4 > 1
Agent 2: 3 > 4 > 1 > 2 > 0
Manipulated preferences of agent 0:
1 > 0 > 3 > 4 > 2

−326 −326 0
c T: 1100101110 −> [e]; T': 1100110110 −> [d]; P_i: 1 > 2 > 3 > 4 > 0
Agent 0: 1 > 2 > 3 > 4 > 0
Agent 1: 0 > 3 > 4 > 2 > 1
Agent 2: 0 > 4 > 2 > 3 > 1
Agent 3: 4 > 1 > 2 > 0 > 3
Agent 4: 3 > 4 > 1 > 2 > 0
Manipulated preferences of agent 0:
3 > 1 > 2 > 0 > 4
```

```
-334 -202 0
c T: 1100101110 -> [d, e]; T': 1001111010 -> [d]; P_i: 2 > 0 > 3 > 4 > 1
Agent 0: 2 > 0 > 3 > 4 > 1
Agent 1: 1 > 4 > 0 > 2 > 3
Agent 2: 3 > 4 > 1 > 2 > 0
Manipulated preferences of agent 0:
3 > 1 > 0 > 4 > 2

202 0
Agent 0: 4 > 3 > 2 > 1 > 0

314 318 322 326 330 334 338 0
Agent 0: 2 > 0 > 3 > 4 > 1
Agent 1: 3 > 4 > 1 > 2 > 0
Agent 2: 4 > 1 > 2 > 0 > 3
```

References

Barberà, S. (1977). Manipulation of social decision functions. *Journal of Economic Theory*, *15*(2), 266–278.

Barberà, S. (2010). Strategy-proof social choice. In Arrow, K. J., Sen, A. K., & Suzumura, K. (Eds.), *Handbook of Social Choice and Welfare*, Vol. 2, chap. 25, pp. 731–832. Elsevier.

Biere, A. (2008). PicoSAT essentials. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)*, *4*, 75–79.

Biere, A., Heule, M., van Maaren, H., & Walsh, T. (Eds.). (2009). *Handbook of Satisfiability*, Vol. 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.

Blanchette, J. C., & Nipkow, T. (2010). Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proceedings of the First International Conference on Interactive Theorem Proving*, pp. 131–146. Springer.

Brandl, F., Brandt, F., Geist, C., & Hofbauer, J. (2015). Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 18–24. AAAI Press.

Brandt, F. (2015). Set-monotonicity implies Kelly-strategyproofness. *Social Choice and Welfare*, *45*(4), 793–804.

Brandt, F., & Brill, M. (2011). Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pp. 136–142. ACM Press.

Brandt, F., Brill, M., & Harrenstein, P. (2016a). Tournament solutions. In Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.), *Handbook of Computational Social Choice*, chap. 3. Cambridge University Press.

Brandt, F., Geist, C., & Harrenstein, P. (2016b). A note on the McKelvey uncovered set and Pareto optimality. *Social Choice and Welfare*, *46*(1), 81–91.

Brandt, F., Geist, C., & Peters, D. (2016c). Optimal bounds for the no-show paradox via SAT solving. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 314–322. IFAAMAS.

Brandt, F., Geist, C., & Seedig, H. G. (2014). Identifying k-majority digraphs via SAT solving. In *Proceedings of the 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*.

Caminati, M. B., Kerber, M., Lange, C., & Rowat, C. (2015). Sound auction specification and implementation. In *Proceedings of the 16th ACM Conference on Economics and Computation (ACM-EC)*, pp. 547–564. ACM Press.

Chatterjee, S., & Sen, A. (2014). Automated reasoning in social choice theory — some remarks. *Mathematics in Computer Science*, *8*(1), 5–10.

Ching, S., & Zhou, L. (2002). Multi-valued strategy-proof social choice rules. *Social Choice and Welfare*, *19*(3), 569–580.

Cinà, G., & Endriss, U. (2015). A syntactic proof of Arrow's theorem in a modal logic of social choice functions. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pp. 1009–1017. IFAAMAS.

Conitzer, V., & Sandholm, T. (2002). Complexity of mechanism design. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 103–110.

Drummond, J., Perrault, A., & Bacchus, F. (2015). SAT is an effective and complete method for solving stable matching problems with couples. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 518–525. AAAI Press.

Duggan, J., & Schwartz, T. (2000). Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare*, *17*(1), 85–93.

Erdamar, B., & Sanver, M. R. (2009). Choosers as extension axioms. *Theory and Decision*, *67*(4), 375–384.

Feldman, A. (1979). Manipulation and the Pareto rule. *Journal of Economic Theory*, *21*, 473–482.

Fishburn, P. C. (1972). Even-chance lotteries in social choice theory. *Theory and Decision*, *3*(1), 18–40.

Fishburn, P. C. (1977). Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, *33*(3), 469–489.

Fisher, D. C., & Reeves, R. B. (1995). Optimal strategies for random tournament games. *Linear Algebra and its Applications*, *217*, 83–85.

Fréchette, A., Newman, N., & Leyton-Brown, K. (2016). Solving the station repacking problem. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press.

Gärdenfors, P. (1976). Manipulation of social choice functions. *Journal of Economic Theory*, *13*(2), 217–228.

Gärdenfors, P. (1979). On definitions of manipulation of social choice functions. In Laffont, J. J. (Ed.), *Aggregation and Revelation of Preferences*. North-Holland.

Geist, C., & Endriss, U. (2011). Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, *40*, 143–174.

Grandi, U., & Endriss, U. (2013). First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, *42*(4), 595–618.

Kelly, J. S. (1977). Strategy-proofness and social choice functions without single-valuedness. *Econometrica*, *45*(2), 439–446.

Laslier, J.-F. (1997). *Tournament Solutions and Majority Voting*. Springer-Verlag.

Liffiton, M. H., & Sakallah, K. A. (2008). Algorithms for computing minimal unsatisfiable subsets of constraints. *Journal of Automated Reasoning*, *40*(1), 1–33.

McGarvey, D. C. (1953). A theorem on the construction of voting paradoxes. *Econometrica*, *21*(4), 608–610.

McKay, B. D., & Piperno, A. (2013). Practical graph isomorphism, II. *Journal of Symbolic Computation*.

Nehring, K. (2000). Monotonicity implies generalized strategy-proofness for correspondences. *Social Choice and Welfare*, *17*(2), 367–375.

Nipkow, T. (2009). Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automatated Reasoning*, *43*, 289–304.

Sanver, M. R., & Zwicker, W. S. (2012). Monotonicity properties and their adaption to irresolute social choice rules. *Social Choice and Welfare*, *39*(2–3), 371–398.

Sato, S. (2008). On strategy-proof social choice correspondences. *Social Choice and Welfare*, *31*, 331–343.

Sato, S. (2013). A sufficient condition for the equivalence of strategy-proofness and non-manipulability by preferences adjacent to the sincere one. *Journal of Economic Theory*, *148*, 259–278.

Scott, A., & Fey, M. (2012). The minimal covering set in large tournaments. *Social Choice and Welfare*, *38*(1), 1–9.

Tang, P., & Lin, F. (2009). Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, *173*(11), 1041–1053.

Tang, P., & Lin, F. (2011). Discovering theorems in game theory: Two-person games with unique pure nash equilibrium payoffs. *Artificial Intelligence*, *175*(14–15), 2010–2020.

Taylor, A. D. (2005). *Social Choice and the Mathematics of Manipulation*. Cambridge University Press.

Tseitin, G. S. (1983). On the complexity of derivation in propositional calculus. In *Automation of Reasoning*, pp. 466–483. Springer.

STRATEGIC ABSTENTION BASED ON PREFERENCE EXTENSIONS: POSITIVE RESULTS AND COMPUTER-GENERATED IMPOSSIBILITIES [2]

Peer-reviewed Conference Paper

Authors: F. Brandl, F. Brandt, C. Geist, and J. Hofbauer

In: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), pages 18–24. AAAI press, 2015

Abstract: Voting rules are powerful tools that allow multiple agents to aggregate their preferences in order to reach joint decisions. A common flaw of some voting rules, known as the *no-show paradox*, is that agents may obtain a more preferred outcome by abstaining from an election. We study strategic abstention for set-valued voting rules based on Kelly's and Fishburn's preference extensions. Our contribution is twofold. First, we show that, whenever there are at least five alternatives, *every* Pareto-optimal majoritarian voting rule suffers from the no-show paradox with respect to Fishburn's extension. This is achieved by reducing the statement to a finite—yet very large—problem, which is encoded as a formula in propositional logic and then shown to be unsatisfiable by a SAT solver. We also provide a human-readable proof which we extracted from a minimal unsatisfiable core of the formula. Secondly, we prove that every voting rule that satisfies two natural conditions cannot be manipulated by strategic abstention with respect to Kelly's extension. We conclude by giving examples of well-known Pareto-optimal majoritarian voting rules that meet these requirements.

Contribution of thesis author: Computer-aided results and methods; presentation of these; joint project and paper management

Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Strategic Abstention Based on Preference Extensions: Positive Results and Computer-Generated Impossibilities

Florian Brandl Felix Brandt Christian Geist Johannes Hofbauer
Technische Universität München
85748 Garching bei München, Germany
{brandlfl,brandtf,geist,hofbauej}@in.tum.de

Abstract

Voting rules are powerful tools that allow multiple agents to aggregate their preferences in order to reach joint decisions. A common flaw of some voting rules, known as the *no-show paradox*, is that agents may obtain a more preferred outcome by abstaining from an election. We study strategic abstention for set-valued voting rules based on Kelly's and Fishburn's preference extensions. Our contribution is twofold. First, we show that, whenever there are at least five alternatives, *every* Pareto-optimal majoritarian voting rule suffers from the no-show paradox with respect to Fishburn's extension. This is achieved by reducing the statement to a finite—yet very large—problem, which is encoded as a formula in propositional logic and then shown to be unsatisfiable by a SAT solver. We also provide a human-readable proof which we extracted from a minimal unsatisfiable core of the formula. Secondly, we prove that every voting rule that satisfies two natural conditions cannot be manipulated by strategic abstention with respect to Kelly's extension. We conclude by giving examples of well-known Pareto-optimal majoritarian voting rules that meet these requirements.

1 Introduction

Whenever a group of multiple agents aims at reaching a joint decision in a fair and satisfactory way, they need to aggregate their (possibly conflicting) preferences. Voting rules are studied in detail in social choice theory and are coming under increasing scrutiny from computer scientists who are interested in their computational properties or want to utilize them in computational multiagent systems [Brandt *et al.*, 2013].

A common flaw of many such rules, first observed by Fishburn and Brams [1983], who called it the *no-show paradox*, is that agents may obtain a more preferred outcome by abstaining from an election. Following Moulin [1988], a voting rule is said to satisfy *participation* if it is immune to the no-show paradox. Moulin has shown that all resolute, i.e., single-valued, scoring rules (such as Borda's rule) satisfy participation while all resolute Condorcet extensions suffer from the no-show paradox. Condorcet extensions comprise a large class of voting rules that satisfy otherwise rather desirable properties.

In this paper, we study participation for *irresolute*, i.e., set-valued, *social choice functions (SCFs)*. A proper definition of participation for irresolute SCFs requires the specification of preferences over *sets* of alternatives. Rather than asking the agents to specify their preferences over all subsets (which would be bound to various rationality constraints and require exponential space), it is typically assumed that the preferences over single alternatives can be extended to preferences over sets. Of course, there are various ways how to extend preferences to sets (see, e.g., [Gärdenfors, 1979; Barberà *et al.*, 2004]), each of which leads to a different version of participation. A function that yields a (possibly incomplete) preference relation over subsets of alternatives when given a preference relation over single alternatives is called a *preference extension*. In this paper, we focus on two common preference extensions due to Kelly [1977] and Fishburn [1972], both of which arise under natural assumptions about the agents' knowledge of the tie-breaking mechanism that eventually picks a single alternative from the choice set (see, e.g., [Gärdenfors, 1979; Ching and Zhou, 2002; Sanver and Zwicker, 2012; Brandt and Brill, 2011; Brandt, 2015]). Kelly's extension, for example, can be motivated by assuming that the agents possess no information whatsoever about the tie-breaking mechanism. A common interpretation of Fishburn's extension, on the other hand, is that ties are broken according to the unknown preferences of a chairman. Since Fishburn's extension is a refinement of Kelly's extension it follows that Fishburn-participation is stronger than Kelly-participation. The idea pursued in this paper is to exploit the uncertainty of the agents about the tie-breaking mechanism in order to prevent strategic abstention. Our two main results are as follows.

- Whenever there are at least four alternatives, Pareto-optimality and Fishburn-participation are incompatible in the context of majoritarian SCFs. When there are at least five alternatives, this even holds for strict preferences.
- Every SCF that satisfies set-monotonicity and independence of indifferent voters satisfies Kelly-participation. Every set-monotonic majoritarian SCF satisfies Kelly-participation when preferences are strict.

The first result is obtained using computer-aided theorem proving techniques. In particular, we reduce the statement to a finite—yet very large—problem, which is encoded as a formula in propositional logic and then shown to be unsatisfiable by a SAT solver. We also provide a human-readable proof for this result, which we extracted from a minimal unsatisfiable core of the SAT formula.

The conditions for the second result are easy to check and satisfied by a small number of well-studied SCFs, including Pareto-optimal majoritarian SCFs. In contrast to Moulin's negative result for resolute SCFs, there *are* appealing Condorcet extensions that satisfy Kelly-participation.

Our negative result holds even for strict preferences while our positive result holds even for weak preferences. The latter is somewhat surprising and stands in sharp contrast to the related finding that no Condorcet extension satisfies Kelly-*strategyproofness* when preferences are weak (recall that an SCF is strategyproof if no agent can obtain a more preferred outcome by misrepresenting his preferences) [Brandt, 2015].

Participation is similar to, but logically independent from, strategyproofness. Manipulation by abstention is arguably a more severe problem than manipulation by misrepresentation for two reasons. First, agents might not be able to find a beneficial misrepresentation. It was shown in various papers that the corresponding computational problem can be intractable (see, e.g., [Faliszewski *et al.*, 2010]). Finding a successful manipulation by strategic abstention, on the other hand, is never harder than computing the outcome of the respective SCF. Secondly, one could argue that agents will not lie about their preferences because this is considered immoral (Borda famously exclaimed "my scheme is intended only for honest men"), while strategic abstention is deemed acceptable.[1]

2 Related Work

The problem of strategic abstention for irresolute SCFs has been addressed by Pérez [2001], Jimeno *et al.* [2009], and Brandt [2015]. Pérez [2001] examined the situation where an agent can cause his most preferred alternative to be excluded from the choice set when joining an electorate and showed that almost all Condorcet extensions suffer from this paradox. Jimeno *et al.* [2009], on the other hand, proved that manipulation by abstention is possible for most Condorcet extensions when agents compare sets according to an optimistic, pessimistic, or lexicographic extension. They mentioned the study of participation in the context of weak preferences and Fishburn's extension as interesting research directions for future work. Both of these questions are addressed in our paper. Brandt [2015] investigated strategyproofness for Kelly's extension and gave a simple argument connecting strategyproofness and participation. Brandl *et al.* [2015] studied participation for probabilistic social choice functions and, among other results, proposed functions where a participating agent is always *strictly* better off (unless he already obtains a most-preferred outcome). Abstention in slightly different contexts than the one studied in this paper has recently also caught the attention of computer scientists working on voting equilibria and campaigning (see, e.g., [Desmedt and Elkind, 2010; Baumeister *et al.*, 2012]).

The computer-aided techniques in this paper have been inspired by Tang and Lin [2009], who reduced well-known impossibility results for *resolute* SCFs—such as Arrow's theorem—to finite instances, which can then be checked by a SAT solver. Geist and Endriss [2011] extended this method to a fully-automatic search algorithm for impossibility theorems in the context of preference relations over sets of alternatives. More recently, Brandt and Geist [2014] proved both impossibility and possibility results regarding the strategyproofness of irresolute SCFs using this computer-aided approach. We strongly build on their methodology and extended it to cover the notion of participation, which—as we will see—requires a more advanced framework.

Also for other problems in economics the application of SAT solvers has proven to be quite effective. A prominent example is the ongoing work by Leyton-Brown [2014] in which SAT solvers are used for the development and execution of the FCC's upcoming reverse spectrum auction.

In some respect, our approach also bears some similarities to *automated mechanism design* (see, e.g., [Conitzer and Sandholm, 2002]), where desirable properties are encoded and mechanisms are computed to fit specific problem instances.

3 Preliminaries

Let A be a finite set of alternatives and $\mathbb{N}$ a countable set of agents of which we will consider finite subsets $N \subseteq \mathbb{N}$. Therefore, let $\mathcal{F}(\mathbb{N})$ denote the set of all finite and non-empty subsets of $\mathbb{N}$. A *(weak) preference relation* is a complete, reflexive, and transitive binary relation on A. The preference relation of agent i is denoted by $\succsim_i$. The set of all preference relations is denoted by $\mathcal{R}$. We write $\succ_i$ for the strict part of $\succsim_i$, i.e., $x \succ_i y$ if $x \succsim_i y$ but not $y \succsim_i x$, and $\sim_i$ for the indifference part of $\succsim_i$, i.e., $x \sim_i y$ if $x \succsim_i y$ and $y \succsim_i x$. A preference relation $\succsim_i$ is called *strict* if it additionally is anti-symmetric, i.e., $x \succ_i y$ or $y \succ_i x$ for all distinct alternatives x, y. We will compactly represent a preference relation as a comma-separated list with all alternatives among which an agent is indifferent placed in a set. For example $x \succ_i y \sim_i z$ is represented by $\succsim_i\colon x, \{y, z\}$.

A *preference profile* R is a function from a set of agents N to the set of preference relations $\mathcal{R}$. The set of all preference profiles is denoted by $\mathcal{R}^{\mathcal{F}(\mathbb{N})}$. For a preference profile $R \in \mathcal{R}^N$ and two agents $i \in N$, $j \in \mathbb{N}$, we define

$$R_{-i} = R \setminus \{(i, \succsim_i)\} \quad \text{and} \quad R_{+j} = R \cup \{(j, \succsim_j)\}.$$

The *majority relation* of R is denoted by $\succsim(R)$, where

$$x \succsim(R)\, y \text{ iff } |\{i \in N\colon x \succsim_i y\}| \geq |\{i \in N\colon y \succsim_i x\}|.$$

Its strict part is denoted by $\succ(R)$ and its indifference part by $\sim(R)$. An alternative x is a *Condorcet winner* in R if $x \succ(R)\, y$ for all $y \in A \setminus \{x\}$.

Our central objects of study are *social choice functions (SCFs)*, i.e., functions that map a preference profile to a set of alternatives. Formally, an SCF is a function

$$f\colon \mathcal{R}^{\mathcal{F}(\mathbb{N})} \to 2^A \setminus \emptyset.$$

[1]Alternatively, one could also argue that manipulation by misrepresentation is more critical because agents are tempted to act immorally, which is a valid, but different, concern.

Two minimal fairness conditions for SCFs are anonymity and neutrality. An SCF is *anonymous* if the outcome does not depend on the identities of the agents and *neutral* if it is symmetric with respect to alternatives. An SCF f is *majoritarian* (or a neutral C1 function) if it is neutral and for all $R, R' \in \mathcal{R}^{\mathcal{F}(\mathbb{N})}$, $f(R) = f(R')$ whenever $\succ(R) = \succ(R')$. Even the seemingly narrow class of majoritarian SCFs contains a variety of interesting functions (sometimes called *tournament solutions*). Examples include Copeland's rule, the top cycle, and the uncovered set (see, e.g., [Brandt *et al.*, 2015]). These functions usually also happen to be *Condorcet extensions*, i.e., SCFs that uniquely return a Condorcet winner whenever one exists.

Next we introduce a very weak variable electorate condition which requires that a completely indifferent agent does not affect the outcome. An SCF f satisfies *independence of indifferent voters* (IIV) if

$$f(R) = f(R_{+i}) \text{ for all } R \in \mathcal{R}^{\mathcal{F}(\mathbb{N})},$$

where i is an agent who is indifferent between all alternatives, i.e., $x \sim_i y$ for all $x, y \in A$. It is easy to see that every majoritarian SCF satisfies anonymity, neutrality, and IIV.

We say that R' is an f-improvement over R if alternatives that are chosen by f in R are not weakened from R to R', i.e., for all $x \in f(R)$, $y \in A$, and $i \in N$, $x \succsim_i y$ implies $x \succsim'_i y$ and $y \succsim'_i x$ implies $y \succsim_i x$. An SCF f satisfies *set-monotonicity* if

$$f(R) = f(R') \text{ whenever } R' \text{ is an } f\text{-improvement over } R.$$

The two preference extensions we consider in this paper are Kelly's extension and Fishburn's extension. For all $X, Y \subseteq A$ and $\succsim_i \in \mathcal{R}$,

$$X \succsim_i^K Y \text{ iff } x \succsim_i y \text{ for all } x \in X, y \in Y, \text{ and} \qquad \text{(Kelly)}$$

$$X \succsim_i^F Y \text{ iff } X \backslash Y \succsim_i^K Y \text{ and } X \succsim_i^K Y \backslash X. \qquad \text{(Fishburn)}$$

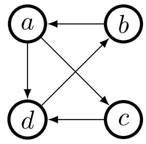
The strict part of these relations will be denoted by $\succ_i^K$ and $\succ_i^F$, respectively. It follows from the definitions that Fishburn's extension is a refinement of Kelly's extension, i.e., $\succsim_i^K \subseteq \succsim_i^F$ for every $\succsim_i \in \mathcal{R}$. In the interest of space, we refer to Section 1 (and the references therein) for justifications of these extensions.

With the preference extensions at hand, we can now formally define participation and strategyproofness. An SCF f is *Kelly-manipulable by strategic abstention* if there exists a preference profile $R \in \mathcal{R}^N$ with $N \in \mathcal{F}(\mathbb{N})$ and an agent $i \in N$ such that $f(R_{-i}) \succ_i^K f(R)$. An SCF f is *Kelly-manipulable* if there exist preference profiles $R, R' \in \mathcal{R}^N$, and an agent $i \in N$, such that $\succsim_j = \succsim'_j$ for all agents $j \neq i$ and $f(R') \succ_i^K f(R)$. f is said to satisfy *Kelly-participation* or *Kelly-strategyproofness* if it is not Kelly-manipulable by strategic abstention or Kelly-manipulable, respectively. Fishburn-participation and Fishburn-strategyproofness are defined analogously.

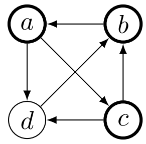
The following example illustrates the definitions of Kelly-participation and Fishburn-participation. Consider the preference profile R with six agents and four alternatives depicted below. The numbers on top of each column denote the identities of the agents with the respective preference relation.

1	2	3, 4	5, 6
c	d	a	b
d	b	c	a
b	a	d	c
a	c	b	d

R



$\succ(R)$



$\succ(R_{-6})$

The profile R induces the majority relation $\succsim(R)$ with its strict part $\succ(R)$. A well-studied majoritarian SCF is the *bipartisan set* [Laffond *et al.*, 1993; Dutta and Laslier, 1999]. The bipartisan set of R is $\{a, b, c, d\}$. If agent 6 leaves the electorate, we obtain the profile R_{-6}, which induces the majority relation $\succsim(R_{-6})$ whose bipartisan set is $\{a, b, c\}$. Observe that $\{a, b, c\} \succ_6^F \{a, b, c, d\}$, i.e., agent 6 can obtain a preferred outcome according to Fishburn's extension by abstaining from the election. Hence, the bipartisan set does not satisfy Fishburn-participation. However, $\{a, b, c\} \succ_6^K \{a, b, c, d\}$ does not hold and, hence, agent 6 cannot manipulate by abstaining according to Kelly's extension. In general, the bipartisan set satisfies Kelly-participation because it satisfies set-monotonicity and IIV (cf. Theorem 3).

We will relate participation and strategyproofness to *Pareto-optimality* in the subsequent sections. An alternative x is said to be *Pareto-dominated* (in $R \in \mathcal{R}^N$) by another alternative y if $y \succsim_i x$ for all $i \in N$ and there exists $j \in N$ such that $y \succ_j x$. Whenever there is no $y \in A$ that Pareto-dominates x, x is called *Pareto-optimal*. The *Pareto rule (PO)* is defined as the SCF that selects all Pareto-optimal alternatives.

4 Computer-aided Theorem Proving

For some of our results, we are going to make use of the computer-aided proving methodology described by Brandt and Geist [2014]. The main idea is to prove statements by encoding a finite instance as a satisfiability problem, which can be solved by a computer using a SAT solver, and providing a (simple) reduction argument, which extends this result to arbitrary domain sizes. We extend their framework to also cater for indifferences in the majority relations, which is an important requirement for being able to deal with the notion of participation: if an agent with at least one strict preference abstains the election, the corresponding majority relation might already contain indifferences.

Note that the introduction of majority ties significantly increases the size of the search space (see Table 1), which makes any type of exhaustive search even less feasible. Apart from being able to treat such large search spaces, another major advantage of the computer-aided approach is that many similar conjectures and hypotheses (here, e.g., statements about other preference extensions) can be checked quickly using the same framework.

In the coming subsections, we are going to explain our extension and some core features of the computer-aided method; for details of the original approach, however, the reader is referred to Brandt and Geist [2014].

$\lvert A\rvert$	Brandt and Geist [2014]	This paper
3	49	**823,543**
4	50,625	$\sim \mathbf{2.5 \cdot 10^{49}}$
5	$\sim 7.9 \cdot 10^{17}$	$\sim \mathbf{9.4 \cdot 10^{867}}$
6	$\sim 5.8 \cdot 10^{100}$	$\sim \mathbf{6.8 \cdot 10^{38649}}$

Table 1: Number of different majoritarian SCFs. While Brandt and Geist [2014] could assume an odd number of agents with strict preferences, participation requires us to deal with variable electorates, and therefore weak majority relations.

4.1 Encoding Participation

At the core of the computer-aided approach lies an encoding of the problem to be solved as a SAT instance. For this, all axioms involved need to be stated in propositional logic. We take over the formalization of the optimized encoding by Brandt and Geist [2014], which contains the following relevant axioms: *functionality of the choice function*, the *orbit condition*, and *Pareto-optimality*. Pareto-optimality is encoded as being a refinement of the uncovered set. What remains is to encode the notion of participation. While this encoding turns out to be similar to the one of strategyproofness defined by Brandt and Geist [2014], it is more complex and not straightforward. In particular, it requires a novel condition that is equivalent to participation for majoritarian SCFs, which we are going to call *majority-participation*.

We are going to identify preference profiles with their corresponding majority relations (i.e., we write $f(\succsim)$ instead of $f(R)$). Moreover, the *inverse* of $\succsim$ is denoted by $\precsim$.

Definition 1. A majoritarian SCF f is *Fishburn-majority-manipulable by strategic abstention* if there exist majority relations $\succsim$, $\succsim'$ and a preference relation $\succsim_\mu \in \mathcal{R}$ such that $f(\succsim') \succ^F_\mu f(\succsim)$, with

$$\succ \cap \prec' \;=\; \emptyset, \tag{1}$$
$$(\succsim \setminus \succsim') \cup (\precsim' \setminus \precsim) \;\subseteq\; \succ_\mu \text{ , and} \tag{2}$$
$$\sim \cap \sim' \;\subseteq\; \sim_\mu. \tag{3}$$

If the agents' preferences are required to be strict, it additionally has to hold that either $\succsim$ or $\succsim'$ is anti-symmetric. A majoritarian SCF f satisfies *Fishburn-majority-participation* if it is not Fishburn-majority-manipulable by strategic abstention.

Conditions (1) to (3) can intuitively be phrased as follows: (1) no strict relationship may be reversed between $\succ$ and $\succ'$, (2) $\succsim_\mu$ has to be in line with the changes from $\succsim$ to $\succsim'$, and (3) majority ties that occur in both majority relations must be reflected by an indifference in $\succsim_\mu$.

In the following lemma, we show that, for majoritarian SCFs, the condition of Fishburn-majority-manipulability corresponds to an abstaining agent with preferences $\succsim_\mu$ who thereby obtains a preferred outcome.[2]

[2]Note that both the definition of majority-participation and Lemma 1 are independent of a specific preference extension, and thus also applicable to, e.g., Kelly's extension.

Lemma 1. *A majoritarian SCF satisfies Fishburn-participation if and only if it satisfies Fishburn-majority-participation.*

Proof. Due to space restrictions we will only provide a short proof sketch here, the full proof is available from the authors upon request.

In general, we show that for every preference profile R that allows for a Fishburn-manipulation by abstention by agent μ, the two majority graphs $\succsim(R)$ and $\succsim(R_{-\mu})$ together with $\succsim_\mu$ satisfy all required conditions. In return, whenever we have two majority relations $\succsim$, $\succsim'$ and a preference relation $\succsim_\mu$ with the properties stated in Definition 1, we can assign integer weights to all pairs of alternatives and, by Debord [1987], use these to determine a preference profile R' that induces the majority relation $\succsim'$. Together with $R = R'_{+\mu}$ we obtain $\succsim(R') = \succsim'$, $\succsim(R) = \succsim$ and thus $f(R') \succ^F_\mu f(R)$ for majoritarian SCFs f. □

Fishburn-majority-participation can then be encoded in propositional logic (with variables $f_{\succsim,X}$ representing $f(\succsim) = X$) as the following simple transformation shows:

$$\neg\left(f(\succsim') \succ^F_\mu f(\succsim)\right) \equiv \bigwedge_{Y \succ^F_\mu X} (\neg f_{\succsim,X} \vee \neg f_{\succsim,Y})$$

for all majority relations $\succsim$, $\succsim'$ and preference relations $\succsim_\mu$ satisfying conditions (1) to (3).

4.2 Proof Extraction

A very interesting feature of the approach by Brandt and Geist [2014] is the possibility to extract human-readable proofs from an unsatisfiability result by the SAT solver. This is done by computing a minimal unsatisfiable set (MUS), an inclusion-minimal set of clauses that is still unsatisfiable.[3] This MUS can then, assisted by our encoder/decoder program, be read and transformed into a standard human-readable proof. Different proofs can be found by varying the MUS extractor or by encoding the problem for different subdomains, such as neighborhoods of a set of profiles or randomly sampled subdomains, respectively.

5 Results and Discussion

In general, participation and strategyproofness are not logically related. However, extending an observation by Brandt [2015], it can be shown that strategyproofness implies participation under certain conditions. The proof of this statement is omitted due to space restrictions.

Lemma 2. *Consider an arbitrary preference extension. Every SCF that satisfies IIV and strategyproofness satisfies participation. When preferences are strict, every majoritarian SCF that satisfies strategyproofness satisfies participation.*

As a consequence of Lemma 2, some positive results for Kelly-strategyproofness and Fishburn-strategyproofness carry over to participation. We will complement these results by impossibility theorems for Fishburn-participation and a positive result for Kelly-participation, which specifically does not hold for Kelly-strategyproofness.

[3]We used PICOMUS, which is part of the PICOSAT distribution [Biere, 2008].

5.1 Fishburn-participation

It turns out that Pareto-optimality is incompatible with Fishburn-participation in majoritarian social choice. The corresponding Theorems 1 and 2 and their proofs were obtained using the computer-aided method laid out in Section 4. In order to simplify the original proofs, which were found by the computer, we first state a lemma, which offers further insights into the possible choices of majoritarian SCFs that satisfy Fishburn-participation and Pareto-optimality.

To state Lemma 3 we introduce some additional notation: an alternative x (McKelvey) covers an alternative y if x is at least as good as y compared to every other alternative. Formally, x *covers* y if $x \succ y$ and, for all $z \in A$, both $y \succsim z$ implies $x \succsim z$, and $z \succsim x$ implies $z \succsim y$. The *uncovered set* of $\succsim$, denoted $UC(\succsim)$, is the set of all alternatives that are not covered by any other alternative. By definition, UC is a majoritarian SCF.

Brandt *et al.* [2014] have shown that every majoritarian and Pareto-optimal SCF selects a subset of the (McKelvey) uncovered set. We show that an SCF that additionally satisfies Fishburn-participation furthermore only depends on the majority relation between alternatives in the uncovered set.

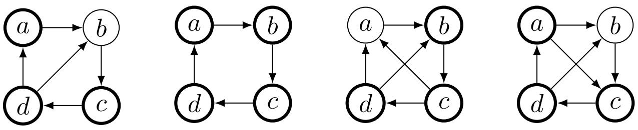
Lemma 3. *Let f be a majoritarian and Pareto-optimal SCF that satisfies Fishburn-participation. Let R, R' be preferences profiles such that $\succsim(R)|_{UC(\succsim(R))} = \succsim(R')|_{UC(\succsim(R'))}$. Then $f(R) = f(R') \subseteq UC(\succsim(R))$.*

The proof of Lemma 3 is omitted due to space restrictions.[4]

Now, let us turn to our impossibility theorems. The computer found these impossibilities even without using Lemma 3. However, the formalization of the lemma allowed the SAT solver to find smaller proofs and makes the human-readable proofs more intuitive.

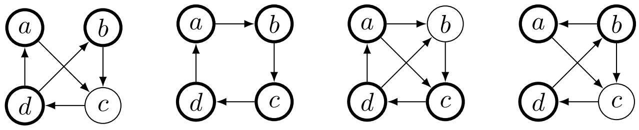
Theorem 1. *There is no majoritarian and Pareto-optimal SCF that satisfies Fishburn-participation if $|A| \geq 4$.*

Proof. Let f be a majoritarian and Pareto-optimal SCF satisfying Fishburn-participation. We first prove the statement for $A = \{a, b, c, d\}$ and reason about the outcome of f for some specific majority relations. Throughout this proof, we are going to make extensive use of Lemma 1, which allows us to apply Fishburn-majority-participation instead of regular Fishburn-participation. Intuitively, the proof strategy is to alter the majority relations $\succsim$, $\succsim'$, and $\succsim''$ as depicted below by letting varying agents *join* some underlying electorate, which will exclude certain choices of f (by an application of Fishburn-majority-participation), until we reach a contradiction. In the figures of the strict part of the majority relations we highlight alternatives that have to be chosen by f with a thick border.



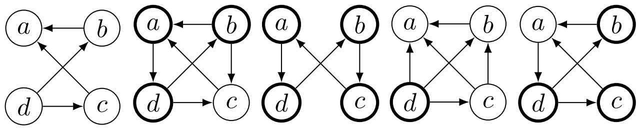
$\succ$ $\succ_{+1}$ $\succ_{+2}$ $\succ_{+3}$

First consider $\succsim$. Adding an agent with preferences $\succsim_1 : \{a, b, c\}, d$ possibly yields $\succsim_{+1}$ where, due to symmetry, $f(\succsim_{+1}) = \{a, b, c, d\}$. As f satisfies Fishburn-participation, nothing that is strictly preferred to $\{a, b, c, d\}$ according to $\succsim_1^F$ can be chosen in $\succsim$. Thus, $d \in f(\succsim)$. Adding another agent with $\succsim_1 : b, \{a, c, d\}$ may also lead to $\succsim_{+1}$. Hence $f(\succsim) \neq \{b, d\}, \{a, b, d\}, \{b, c, d\}$.



$\succ'$ $\succ'_{+1}$ $\succ'_{+2}$ $\succ'_{+3}$

In a similar fashion, we obtain—step by step and using the majority relations depicted—that $f(\succsim) = \{a, c, d\}$, $f(\succsim') = \{a, b, d\}$, and finally $f(\succsim'') = \emptyset$, a contradiction. The details of these cases have to be omitted due to space constraints and are available from the authors upon request.



$\succ''$ $\succ''_{+1}$ $\succ''_{+2}$ $\succ''_{+3}$ $\succ''_{+4}$

Now let $|A| \geq 5$. It follows from Lemma 3 that the choice of f does not depend on covered alternatives. Hence the statement follows by extending the majority relations depicted above to A such that all alternatives but a, b, c, d are covered. $\square$

We could verify with our computer-aided approach that this impossibility still holds for strict preferences when there are at least 5 alternatives.

Theorem 2. *There is no majoritarian and Pareto-optimal SCF that satisfies Fishburn-participation if $|A| \geq 5$, even when preferences are strict.*

The shortest proof of Theorem 2 that we were able to extract from our computer-aided approach still uses 124 different instances of manipulation by abstention. The proof of Theorem 1, by comparison, consists of 10 such instances. As a consequence, the complete proof of Theorem 2 has to be omitted here; a computer-generated version, however, is available from the authors upon request. Theorems 1 and 2 are both tight in the sense that, whenever there are less than four or five alternatives, respectively, there exists an SCF that satisfies the desired properties.

An interesting question is whether these impossibilities also extend to other preference extensions. Given the computer-aided approach, this can be easily checked by simply replacing the preference extension in the SAT encoder. For instance, it turns out that the impossibility of Theorem 1 still holds if we consider a coarsening of Fishburn's extension which can only compare sets that are contained in each other. Kelly's extension on the other hand does not lead to an impossibility for $|A| \leq 5$, which will be confirmed more generally in the next section.

[4]Lemma 3 can be strengthened in various respects. It also holds for the *iterated* uncovered set, all preference extensions satisfying some mild conditions, and probabilisitic social choice functions.

Extension	Property	Strict preferences	Weak preferences
Kelly	Participation	$\oplus$ All set-monotonic SCFs that satisfy IIV (Theorem 3)	
	Strategy-proofness	$\oplus$ All set-monotonic SCFs[b]	$\ominus$ No Condorcet extension[b]
Fishburn	Participation	$\ominus$ No majoritarian & Pareto-optimal SCF ($\lvert A\rvert \geq 5$) (Theorem 2)	$\ominus$ No majoritarian & Pareto-optimal SCF ($\lvert A\rvert \geq 4$) (Theorem 1)
		$\oplus$ Few undiscriminating SCFs, e.g., $COND$[c], TC[g], and PO[e] (Lemma 2), and all scoring rules[f]	
	Strategy-proofness	$\ominus$ No majoritarian & Pareto-optimal SCF ($\lvert A\rvert \geq 5$)[d]	$\ominus$ No pairwise & Pareto-optimal SCF ($\lvert A\rvert \geq 4$)[a]
		$\oplus$ Few undiscriminating SCFs, e.g., $COND$[c], TC[g], and PO[e]	

Table 2: Overview of results for participation and strategyproofness with respect to Kelly's and Fishburn's extension and strict/weak preferences. The symbol $\oplus$ marks sufficient conditions while $\ominus$ marks impossibility results. [a]: follows from a result by Aziz *et al.* [2014] about probabilistic social choice functions, [b]: Brandt [2015], [c]: Brandt and Brill [2011], [d]: Brandt and Geist [2014], [e]: Feldman [1979], [f]: Moulin [1988], [g]: Sanver and Zwicker [2012]

5.2 Kelly-participation

Theorems 1 and 2 are sweeping impossibilities within the domain of majoritarian SCFs. For Kelly's extension, we obtain a much more positive result that covers attractive majoritarian and non-majoritarian SCFs. Brandt [2015] has shown that set-monotonicity implies Kelly-strategyproofness for strict preferences and that no Condorcet extension is Kelly-strategyproof when preferences are weak. We prove that set-monotonicity (and the very mild IIV axiom) imply Kelly-participation even for weak preferences. We have thus found natural examples of SCFs that violate Kelly-strategyproofness but satisfy Kelly-participation.[5]

Theorem 3. *Let f be an SCF that satisfies IIV and set-monotonicity. Then f satisfies Kelly-participation.*

Proof. Let f be an SCF that satisfies IIV and set-monotonicity. Assume for contradiction that f does not satisfy Kelly-participation. Hence there exist a preference profile R and an agent i such that $f(R_{-i}) \succ_i^K f(R)$. Let $X = f(R), Y = f(R_{-i})$, and $Z = A \setminus (X \cup Y)$. By definition of $Y \succ_i^K X$ we have that $x \sim_i y$ for all $x, y \in X \cap Y$.

We define a new preference relation $\succsim_i'$ in which all alternatives in Y are tied for the first place, followed by all alternatives in $X \setminus Y$ as they are ordered in $\succsim_i$, and all remaining alternatives in one indifference class at the bottom of the ranking. Formally,

$$\succsim_{i'} = (Y \times A) \cup \succsim_i|_X \cup (A \times Z).$$

Let i'' be an agent who is indifferent between all alternatives, i.e., $x \sim_{i''} y$ for all $x, y \in A$. Since f satisfies IIV we have that $f(R_{-i+i''}) = f(R_{-i})$.

By definition, $R_{-i+i'}$ is an f-improvement over both R and $R_{-i+i''}$. Hence, set-monotonicity implies that $f(R_{-i+i'}) = f(R)$ and $f(R_{-i+i'}) = f(R_{-i+i''})$. In summary, we obtain

$$f(R_{-i+i'}) = f(R_{-i+i''}) = f(R_{-i}) \succ_i^K f(R) = f(R_{-i+i'}),$$

which is a contradiction. □

[5] It is easily seen that the proof of Theorem 3 straightforwardly extends to *group*-participation, i.e., no group of agents can obtain a unanimously more preferred outcome by abstaining.

Two rather undiscriminating SCFs that satisfy both IIV and set-monotonicity are the *Pareto rule* and the *omninomination rule* (which returns all alternatives that are ranked first by at least one agent). Majoritarian SCFs satisfy IIV by definition and there are several appealing majoritarian SCFs that satisfy set-monotonicity, among those for instance the *top cycle*, the *minimal covering set*, and the *bipartisan set* (see, e.g., [Brandt, 2015; Brandt *et al.*, 2015]). These majoritarian SCFs are sometimes criticized for not being discriminating enough. The computer-aided approach described in this paper can be used to find more discriminating SCFs that still satisfy Kelly-participation. We thus found a refinement of the bipartisan set that, for $|A| = 5$, selects only 1.43 alternatives on average, and satisfies Kelly-participation. For comparison, the bipartisan set (the smallest previously known majoritarian SCF satisfying Kelly-participation) yields 2.68 alternatives on average.

6 Conclusion

Previous results have indicated a conflict between strategic non-manipulability and Condorcet-consistency [Moulin, 1988; Pérez, 2001; Jimeno *et al.*, 2009; Brandt, 2015]. For example, Moulin [1988] has shown that no resolute Condorcet extension satisfies participation and Brandt [2015] has shown that no irresolute Condorcet extension satisfies Kelly-strategyproofness. Theorem 3 addresses an intermediate question and finds that—perhaps surprisingly—there *are* attractive Condorcet extensions that satisfy Kelly-participation, even when preferences are weak. On the other hand, we have presented elaborate computer-generated impossibilities (Theorems 1 and 2), which show that these encouraging results break down once preferences are extended by the more refined Fishburn extension. These findings improve our understanding of which behavioral assumptions allow for aggregation functions that are immune to strategic abstention.

An overview of the main results of this paper and how they relate to other related results is given in Table 2.

References

[Aziz *et al.*, 2014] H. Aziz, F. Brandl, and F. Brandt. On the incompatibility of efficiency and strategyproofness in randomized social choice. In *Proc. of 28th AAAI Conference*, pages 545–551. AAAI Press, 2014.

[Barberà *et al.*, 2004] S. Barberà, W. Bossert, and P. K. Pattanaik. Ranking sets of objects. In S. Barberà, P. J. Hammond, and C. Seidl, editors, *Handbook of Utility Theory*, volume II, chapter 17, pages 893–977. Kluwer Academic Publishers, 2004.

[Baumeister *et al.*, 2012] D. Baumeister, P. Faliszewski, J. Lang, and J. Rothe. Campaigns for lazy voters: truncated ballots. In *Proc. of 11th AAMAS Conference*, pages 577–584. IFAAMAS, 2012.

[Biere, 2008] A. Biere. PicoSAT essentials. *Journal on Satisfiability, Boolean Modeling and Computation (JSAT)*, 4:75–79, 2008.

[Brandl *et al.*, 2015] F. Brandl, F. Brandt, and J. Hofbauer. Incentives for participation and abstention in probabilistic social choice. In *Proc. of 14th AAMAS Conference*. IFAAMAS, 2015. Forthcoming.

[Brandt and Brill, 2011] F. Brandt and M. Brill. Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Proc. of 13th TARK Conference*, pages 136–142. ACM Press, 2011.

[Brandt and Geist, 2014] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. In *Proc. of 13th AAMAS Conference*, pages 1193–1200. IFAAMAS, 2014.

[Brandt *et al.*, 2013] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiß, editor, *Multiagent Systems*, chapter 6, pages 213–283. MIT Press, 2nd edition, 2013.

[Brandt *et al.*, 2014] F. Brandt, C. Geist, and P. Harrenstein. A note on the McKelvey uncovered set and Pareto optimality. 2014. Working Paper.

[Brandt *et al.*, 2015] F. Brandt, M. Brill, and P. Harrenstein. Tournament solutions. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 3. Cambridge University Press, 2015. Forthcoming.

[Brandt, 2015] F. Brandt. Set-monotonicity implies Kelly-strategyproofness. *Social Choice and Welfare*, 2015. Forthcoming.

[Ching and Zhou, 2002] S. Ching and L. Zhou. Multi-valued strategy-proof social choice rules. *Social Choice and Welfare*, 19:569–580, 2002.

[Conitzer and Sandholm, 2002] V. Conitzer and T. Sandholm. Complexity of mechanism design. In *Proc. of 18th UAI Conference*, pages 103–110, 2002.

[Debord, 1987] B. Debord. Caractérisation des matrices des préférences nettes et méthodes d'agrégation associées. *Mathématiques et sciences humaines*, 97:5–17, 1987.

[Desmedt and Elkind, 2010] Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *Proc. of 11th ACM-EC Conference*, pages 347–356. ACM Press, 2010.

[Dutta and Laslier, 1999] B. Dutta and J.-F. Laslier. Comparison functions and choice correspondences. *Social Choice and Welfare*, 16(4):513–532, 1999.

[Faliszewski *et al.*, 2010] P. Faliszewski, E. Hemaspaandra, and L. Hemaspaandra. Using complexity to protect elections. *Communications of the ACM*, 53(11):74–82, 2010.

[Feldman, 1979] A. Feldman. Manipulation and the Pareto rule. *Journal of Economic Theory*, 21:473–482, 1979.

[Fishburn and Brams, 1983] P. C. Fishburn and S. J. Brams. Paradoxes of preferential voting. *Mathematics Magazine*, 56(4):207–214, 1983.

[Fishburn, 1972] P. C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, 3:18–40, 1972.

[Gärdenfors, 1979] P. Gärdenfors. On definitions of manipulation of social choice functions. In J. J. Laffont, editor, *Aggregation and Revelation of Preferences*. North-Holland, 1979.

[Geist and Endriss, 2011] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[Jimeno *et al.*, 2009] J. L. Jimeno, J. Pérez, and E. García. An extension of the Moulin No Show Paradox for voting correspondences. *Social Choice and Welfare*, 33(3):343–459, 2009.

[Kelly, 1977] J. S. Kelly. Strategy-proofness and social choice functions without single-valuedness. *Econometrica*, 45(2):439–446, 1977.

[Laffond *et al.*, 1993] G. Laffond, J.-F. Laslier, and M. Le Breton. The bipartisan set of a tournament game. *Games and Economic Behavior*, 5(1):182–201, 1993.

[Leyton-Brown, 2014] K. Leyton-Brown. Pragmatic algorithmic game theory. *Keynote address at the ACM-EC Conference*, June 2014.

[Moulin, 1988] H. Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45:53–64, 1988.

[Pérez, 2001] J. Pérez. The strong no show paradoxes are a common flaw in Condorcet voting correspondences. *Social Choice and Welfare*, 18(3):601–616, 2001.

[Sanver and Zwicker, 2012] M. R. Sanver and W. S. Zwicker. Monotonicity properties and their adaption to irresolute social choice rules. *Social Choice and Welfare*, 39(2–3):371–398, 2012.

[Tang and Lin, 2009] P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

OPTIMAL BOUNDS FOR THE NO-SHOW PARADOX VIA SAT SOLVING [3]

Peer-reviewed Conference Paper

Authors: F. Brandt, C. Geist, and D. Peters

In: Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 314–322. IFAAMAS, 2016

Abstract: Voting rules are powerful tools that allow multiple agents to aggregate their preferences in order to reach joint decisions. Perhaps one of the most important desirable properties in this context is *Condorcet-consistency*, which requires that a voting rule should return an alternative that is preferred to any other alternative by some majority of voters. Another desirable property is *participation*, which requires that no voter should be worse off by joining an electorate. A seminal result in social choice theory by Moulin has shown that Condorcet-consistency and participation are incompatible whenever there are at least 4 alternatives and 25 voters. We leverage SAT solving to obtain an elegant human-readable proof of Moulin's result that requires only 12 voters. Moreover, the SAT solver is able to construct a Condorcet-consistent voting rule that satisfies participation as well as a number of other desirable properties for up to 11 voters, proving the optimality of the above bound. We also obtain tight results for set-valued and probabilistic voting rules, which complement and significantly improve existing theorems.

Contribution of thesis author: project management and guidance (in particular regarding computer aided-methods); idea for incremental domain construction and proof trees; mathematical model; joint paper management

Optimal Bounds for the No-Show Paradox via SAT Solving

Felix Brandt
TU München
Munich, Germany
brandtf@in.tum.de

Christian Geist
TU München
Munich, Germany
geist@in.tum.de

Dominik Peters
University of Oxford
Oxford, UK
dominik.peters@cs.ox.ac.uk

ABSTRACT

Voting rules allow multiple agents to aggregate their preferences in order to reach joint decisions. Perhaps one of the most important desirable properties in this context is *Condorcet-consistency*, which requires that a voting rule should return an alternative that is preferred to any other alternative by some majority of voters. Another desirable property is *participation*, which requires that no voter should be worse off by joining an electorate. A seminal result in social choice theory by Moulin [28] has shown that Condorcet-consistency and participation are incompatible whenever there are at least 4 alternatives and 25 voters. We leverage SAT solving to obtain an elegant human-readable proof of Moulin's result that requires only 12 voters. Moreover, the SAT solver is able to construct a Condorcet-consistent voting rule that satisfies participation as well as a number of other desirable properties for up to 11 voters, proving the optimality of the above bound. We also obtain tight results for set-valued and probabilistic voting rules, which complement and significantly improve existing theorems.

Keywords

Computer-aided theorem proving; social choice theory; SAT; no-show paradox; participation; Condorcet

1. INTRODUCTION

Whenever a group of autonomous software agents or robots needs to decide on a joint course of action in a fair and satisfactory way, they need to aggregate their preferences. A common way to achieve this is to use voting rules. Voting rules are studied in detail in social choice theory and are coming under increasing scrutiny from computer scientists who are interested in their computational properties or want to utilize them in multiagent systems (see, e.g., [31, 9]).

In social choice theory, voting rules are usually compared using desirable properties (so-called axioms) that they may or may not satisfy. There are a number of well-known impossibility theorems—among which Arrow's impossibility is arguably the most famous—which state that certain axioms are incompatible with each other. These results, which show the non-existence of voting rules that satisfy a given set of axioms, are important because they clearly define the boundary of what can be achieved at all. This applies to the explicitly stated axioms as well as implicit ones such as boundaries on the number of voters or alternatives. For instance, if there are only two alternatives, Arrow's theorem does not apply and there are many voting rules, including majority rule, that satisfy the conditions used in Arrow's theorem. One impossibility that requires unusually high bounds on the number of voters and alternatives is Moulin's *no-show paradox* [28], which states that the axioms of Condorcet-consistency and participation are incompatible whenever there are at least 4 alternatives and 25 voters. Moulin proves that the bound on the number of alternatives is tight by showing that the maximin voting rule (with lexicographic tie-breaking) satisfies the desired properties when there are at most 3 alternatives. The tightness of the more restrictive condition on the number of voters was left open, however. The goal of this paper is to give tight bounds on the number of voters required for Moulin's theorem and related theorems that appear in the literature. To achieve this, we encode these problems as formulas in propositional logic and then use SAT solvers to decide their satisfiability and extract minimal unsatisfiable sets (MUSes) in the case of unsatisfiability. This approach is based on previous work by Tang and Lin [35], Geist and Endriss [20], Brandt and Geist [8], and Brandl et al. [4]. However, it turned out that a straightforward application of this methodology is insufficient to deal with the magnitude of the problems we considered. Several novel techniques were necessary to achieve our results. In particular, we extracted knowledge from computer-generated proofs of weaker statements and then used this information to guide the search for proofs of more general statements.

As mentioned above, Moulin's theorem uses the axioms of Condorcet-consistency and participation. Condorcet-consistency goes back to one of the most influential notions in social choice theory, namely that of a *Condorcet winner*. A Condorcet winner is an alternative that is preferred to any other alternative by a majority of voters. The Marquis de Condorcet, after whom this concept is named, argued that, whenever a Condorcet winner exists, it should be elected [15]. A voting rule satisfying this condition is called *Condorcet-consistent*. Apart from the intuitive appeal of this condition, Condorcet-consistent rules are more robust to changes in the of feasible alternatives and less susceptible to strategic manipulation than other voting rules (such as Borda's rule) (see, e.g., [11, 14]). While the desirability of Condorcet-consistency—as that of any other axiom—

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

has been subject to criticism, many scholars agree that it is very appealing—if not indispensable—and a large part of the social choice literature deals exclusively with Condorcet-consistent voting rules (e.g., [17, 25, 9]). *Participation* was first considered by Fishburn and Brams [18] and requires that no voter should be worse off by joining an electorate, or—alternatively—that no voter should benefit by abstaining from an election. The desirability of this axiom in any context with voluntary participation is evident. All the more surprisingly, Fishburn and Brams have shown that single transferable vote (STV), a common voting rule, violates participation and referred to this phenomenon as the *no-show paradox*. Moulin [28], perhaps even more startlingly, proved that no Condorcet-consistent voting rule satisfies participation when there are at least 25 voters.

We leverage SAT solving to obtain an elegant human-readable proof of Moulin's result that requires only 12 voters. While computer-aided solving techniques allow us to tackle difficult combinatorial problems, they usually do not provide additional insight into these problems. Somewhat surprisingly, the computer-aided proofs we found possess a certain kind of symmetry that has not been exploited in previous proofs. Moreover, the SAT solver is able to construct a Condorcet-consistent voting rule that satisfies participation as well as a number of other desirable properties for up to 11 voters, proving the optimality of the above bound. This computer-generated voting rule is compatible with the maximin voting rule in 99.8% of all cases and, in contrast to maximin, only selects alternatives from the top cycle. As a practical consequence of our theorem, strategic abstention need not be a concern for Condorcet-consistent voting rules when there are at most 4 alternatives and 11 voters, for instance when voting in a committee. We also use our techniques to provide optimal bounds for related results about set-valued and probabilistic voting rules [22, 36]. In particular, we give a tight bound of 17 voters for the optimistic preference extension, 14 voters for the pessimistic extension, and 12 voters for the stochastic dominance preference extension. These results are substantial improvements of previous results. For example, the previous statement for the pessimistic extension requires an additional axiom, at least 5 alternatives, and at least 971 voters [22]. Our results are summarized in Table 1.

2. RELATED WORK

The no-show paradox was first observed by Fishburn and Brams [18] for the STV voting rule. Ray [30] and Lepelley and Merlin [26] investigate how frequently this phenomenon occurs in practice. The main theorem addressed in this paper is due to Moulin [28] and requires at least 25 voters. This bound was recently brought down to 21 voters by Kardel [23]. Simplified proofs of Moulin's theorem are given by Schulze [33] and Smith [34]. Holzman [21] and Sanver and Zwicker [32] strengthen Moulin's theorem by weakening Condorcet-consistency and participation, respectively. Jimeno et al. [22] prove variants of Moulin's theorem for set-valued voting rules based on the optimistic and the pessimistic preference extension. Pérez [29] defines a weaker notion of participation in the context of set-valued voting rules and shows that all common Condorcet extensions except the maximin rule and Young's rule violate this property. Pérez notes that "a practical question, which has not been dealt with here, refers to the number of candidates and voters that are necessary to invoke the studied paradoxes" ([29], p. 614).

When assuming that voters have incomplete preferences over sets or lotteries, participation and Condorcet-consistency can be satisfies simultaneously and positive results for common Condorcet-consistent voting rules have been obtained by Brandt [7] and Brandl et al. [4, 5, 6]. Abstention in slightly different contexts than the one studied in this paper recently caught the attention of computer scientists working on voting equilibria and campaigning [16, 1].

The computer-aided techniques in this paper are inspired by Tang and Lin [35], who reduced well-known impossibility results from social choice theory—such as Arrow's theorem—to finite instances, which can then be checked by a SAT solver. This methodology has been extended and applied to new problems by Geist and Endriss [20], Brandt and Geist [8], and Brandl et al. [4]. The results obtained by computer-aided theorem proving have already found attention in the social choice community [12]. More generally, SAT solvers have also proven to be quite effective for other problems in economics. A prominent example is the ongoing work by Fréchette et al. [19] in which SAT solvers are used for the development and execution of the FCC's upcoming reverse spectrum auction. In some respects, our approach also bears some similarities to *automated mechanism design* (see, e.g., [13]), where desirable properties are encoded and mechanisms are computed to fit specific problem instances.

3. PRELIMINARIES

Let A be a set of m alternatives and $\mathcal{N}$ be a set of n voters. Whether no-show paradoxes occur depends on the exact values of m and n. By $\mathcal{E}(\mathcal{N}) := 2^{\mathcal{N}} \setminus \{\emptyset\}$ we denote the set of *electorates*, i.e., non-empty subsets of $\mathcal{N}$. For many of our results, we will take $A = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$, and we use the labels $\mathtt{x}, \mathtt{y}$ for arbitrary elements of A.

A *(strict) preference relation* is a complete, antisymmetric, and transitive binary relation on A. The preference relation of voter i is denoted by $\succcurlyeq_i$. The set of all preference relations over A is denoted by $\mathcal{R}$. For brevity, we denote by $\mathtt{abcd}$ the preference relation $\mathtt{a} \succcurlyeq_i \mathtt{b} \succcurlyeq_i \mathtt{c} \succcurlyeq_i \mathtt{d}$, eliding the identity of voter i, and similarly for other preferences.

A *preference profile* R is a function from an electorate $N \in \mathcal{E}(\mathcal{N})$ to the set of preference relations $\mathcal{R}$. The set of all preference profiles is thus given by $\mathcal{R}^{\mathcal{E}(\mathcal{N})}$. For the sake of adding and deleting voters, we define for any preference profile $R \in \mathcal{R}^N$ with $(i, \succcurlyeq_i) \in R$, and $j \in \mathcal{N} \setminus N$, $\succcurlyeq_j \in \mathcal{R}$

$$R - i := R \setminus \{(i, \succcurlyeq_i)\}, \quad R + (j, \succcurlyeq_j) := R \cup \{(j, \succcurlyeq_j)\}.$$

If the identity of the voter is clear or irrelevant we sometimes, in slight abuse of notation, refer to $R - i$ by $R - \succcurlyeq_i$, and write $R + \succcurlyeq_j$ instead of $R + (j, \succcurlyeq_j)$. If k voters with the same preferences $\succcurlyeq_i$ are to be added or removed, we write $R + k \cdot \succcurlyeq_i$ and $R - k \cdot \succcurlyeq_i$, respectively.

The *majority margin* of R is the map $g_R \colon A \times A \to \mathbb{Z}$ with $g_R(\mathtt{x}, \mathtt{y}) = |\{i \in N \mid \mathtt{x} \succcurlyeq_i \mathtt{y}\}| - |\{i \in N \mid \mathtt{y} \succcurlyeq_i \mathtt{x}\}|$. The majority margin can be viewed as the adjacency matrix of a *weighted tournament* T_R. We say that a preference profile R *induces* the weighted tournament T_R.

An alternative $\mathtt{x}$ is called *Condorcet winner* if it wins against any other alternative in a majority contest, i.e., if $g_R(\mathtt{x}, \mathtt{y}) > 0$ for all $\mathtt{y} \in A \setminus \{\mathtt{x}\}$. Conversely, an alternative $\mathtt{x}$ is a *Condorcet loser* if $g_R(\mathtt{x}, \mathtt{y}) < 0$ for all $\mathtt{y} \in A \setminus \{\mathtt{x}\}$.

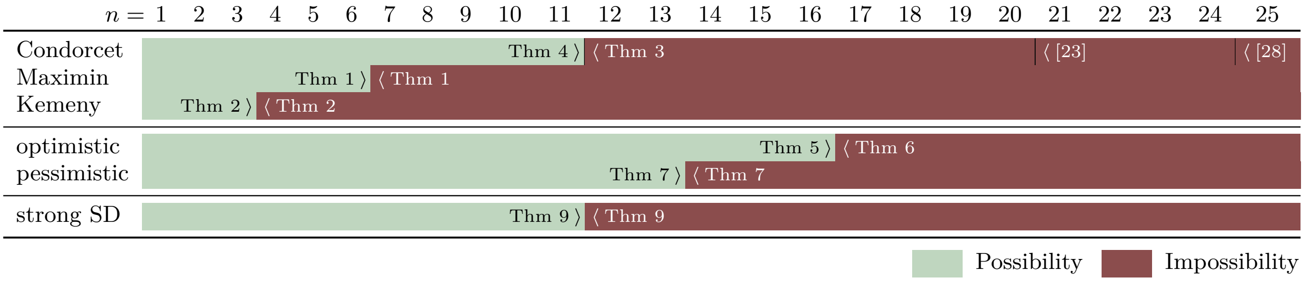


Table 1: Bounds on the number of voters for which Condorcet extensions can satisfy participation. Green cells indicate the existence of a Condorcet extension satisfying participation (for $m = 4$). Red cells indicate that no Condorcet extension satisfies participation (for $m \geqslant 4$).

Our central object of study are *voting rules*, i.e., functions that assigns every preference profile a socially preferred alternative. Thus, a voting rule is a function $f\colon \mathcal{R}^{\mathcal{E}(\mathcal{N})} \to A$. In this paper, we study voting rules that satisfy *Condorcet-consistency* and *participation*.

Definition 1. *A* Condorcet extension *is a voting rule that selects the Condorcet winner whenever it exists. Thus, f is a Condorcet extension if for every preference profile R that admits a Condorcet winner* $\mathtt{x}$, *we have $f(R) = \mathtt{x}$. We say that f is* Condorcet-consistent*.*

Definition 2. *A voting rule f satisfies* participation *if all voters always weakly prefer voting to not voting, i.e., if $f(R) \succcurlyeq_i f(R-i)$ for all $R \in \mathcal{R}^N$ and $i \in N$ with $N \in \mathcal{E}(\mathcal{N})$.*

Equivalently, participation requires that for all preference profiles R not including voter j, we have $f(R+\succcurlyeq_j) \succcurlyeq_j f(R)$.

4. MAXIMIN AND KEMENY'S RULE

The proofs of both positive and negative results to come were obtained through automated techniques that we describe in Section 5. To become familiar with the kind of arguments produced in this way, we will now study a more restricted setting which is of independent interest.

Specifically, let us consider voting rules that select winners in accordance with the popular *maximin* and *Kemeny* rules. For a preference profile R, an alternative $\mathtt{x}$ is a *maximin winner* if it maximizes $\min_{\mathtt{y}\in A\setminus\{\mathtt{x}\}} g_R(\mathtt{x},\mathtt{y})$; thus, $\mathtt{x}$ never gets defeated too badly in pairwise comparisons. An alternative $\mathtt{x}$ is a *Kemeny winner* if it is ranked first in some *Kemeny ranking*. A Kemeny ranking is a preference relation $\succcurlyeq_{\mathrm{K}} \in \mathcal{R}$ maximizing agreement with voters' individual preferences, i.e., it maximizes the quantity $\sum_{i\in N} |\succcurlyeq_{\mathrm{K}} \cap \succcurlyeq_i|$.

We call a voting rule a *maximin extension* (resp. *Kemeny extension*) if it always selects a maximin winner (resp. Kemeny winner). Since a Condorcet winner, if it exists, is always the unique maximin and Kemeny winner of a preference profile, any such voting rule is also a Condorcet extension. We can now prove an easy version of Moulin's theorem for these more restricted voting rules.

To this end, we first prove a useful lemma allowing us to extend impossibility proofs for 4 alternatives to also apply if there are more than 4 alternatives. Its proof gives a first hint on how Condorcet-consistency and participation interact.

Lemma 1. *Suppose that f is a Condorcet extension satisfying participation. Let R be a preference profile and $B \subsetneq A$ a set of* bad *alternatives such that each voter ranks every $\mathtt{x} \in B$ below every $\mathtt{y} \in A \setminus B$. Then $f(R) \notin B$.*

Proof. By induction on the number of voters $|N|$ in R. If R consists of a single voter i, then, since f is a Condorcet extension, $f(R)$ must return i's top choice, which is not bad. If R consists of at least 2 voters, and $i \in N$, then by participation $f(R) \succcurlyeq_i f(R-i)$. If $f(R)$ were bad, then so would be $f(R-i)$, contradicting the inductive hypothesis. □

The following computer-aided proofs, just like the more complicated proofs to follow, can be understood solely by carefully examining the corresponding 'proof diagram'. An arrow such as $R \xrightarrow{+\,\mathtt{abcd}} R'$ indicates that profile R' is obtained from R by adding a voter $\mathtt{abcd}$, and is read as "if one of the bold green alternatives (here $\mathtt{ab}$) is selected at R, then one of them is selected at R'" (by participation). A circled node a indicates a profile admitting Condorcet winner $\mathtt{a}$, although in the proofs of Theorems 1 and 2, we use it to refer to maximin and Kemeny winners, respectively.

Theorem 1. *There is no maximin extension that satisfies participation for $m \geqslant 4$ and $n \geqslant 7$. (For $m = 4$ and $n \leqslant 6$, such a maximin extension exists.)*

Proof. Let f be a maximin extension which satisfies participation. Consider the following 6-voter profile R:

1	2	2	1
a	b	c	d
b	d	a	c
d	c	b	a
c	a	d	b

R

$R \xrightarrow{+\,\mathtt{abcd}}$ c $\qquad R \xrightarrow{+\,\mathtt{dcba}}$ b

Suppose $f(R) \in \{\mathtt{a},\mathtt{b}\}$. Adding an $\mathtt{abcd}$ vote leads to a weighted tournament in which alternative $\mathtt{c}$ is the unique maximin winner. But this contradicts participation since the added voter would benefit from abstaining the election.

Symmetrically, if $f(R) \in \{\mathtt{c},\mathtt{d}\}$, then adding a $\mathtt{dcba}$ vote leads to a weighted tournament in which $\mathtt{b}$ is the maximin winner, again contradicting participation. The symmetry of the argument is due to an automorphism of R, namely the relabelling of alternatives according to $\mathtt{abcd} \mapsto \mathtt{dcba}$.

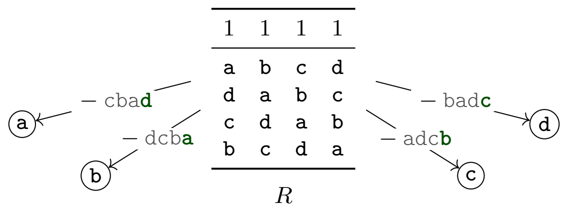
If $m > 4$, we add new bad alternatives $\mathtt{x}_1, \mathtt{x}_2, \ldots, \mathtt{x}_{m-4}$ to the bottom of R and of the additional voters. By Lemma 1, f chooses from $\{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}$ at each step, completing the proof.

The existence result for $n \leqslant 6$ is established by the methods described in Section 5. □

For 3 alternatives, Moulin [28] proved that the voting rule that chooses the lexicographically first maximin winner satisfies participation. Theorem 1 shows that this is not the case for 4 alternatives, even if there are only 7 voters and no matter how we pick among maximin winners.

THEOREM 2. *There is no Kemeny extension that satisfies participation for $m \geqslant 4$ and $n \geqslant 4$. (For $m = 4$ and $n \leqslant 3$, such a Kemeny extension exists.)*

PROOF. Let f be a Kemeny extension which satisfies participation. Consider the following 4-voter profile R:



R

Suppose $f(R) = \mathtt{d}$. Then removing cbad from R yields a weighted tournament in which a is the (unique) Kemeny winner, which contradicts participation. Analogously, we can exclude the other three possible outcomes for R by letting a voter abstain, which always leads to a unique Kemeny winner and a contradiction with participation. The arguments are identical because R is completely symmetric in the sense that for any pair of alternatives x and y, there is an automorphism of R that maps x to y.

Just like for Theorem 1, if $m > 4$, we add new bad alternatives $\mathtt{x}_1, \mathtt{x}_2, \ldots, \mathtt{x}_{m-4}$ to the bottom of R and of the additional voters. By Lemma 1, f chooses from $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ at each step, completing the proof. □

One remarkable and unexpected aspect of the computer-aided proofs above is that their simplicity is due to automorphisms of the underlying preference profiles. Similar automorphisms will also be used in the proofs of the stronger theorems in Sections 6, 7, and 8. We emphasize that these symmetries are not hard-coded into our problem specification and, to the best of our knowledge, have not been exploited in previous proofs of similar statements.

5. METHOD: SAT SOLVING FOR COMPUTER-AIDED PROOFS

The bounds in this paper were obtained with the aid of a computer. In this section, we describe the method that we employed. The main tool in our approach is an encoding of our problems into propositional logic. We then use SAT solvers to decide whether (in a chosen setting) there exists a Condorcet extension satisfying participation. If the answer is yes, the solver returns an explicit such voting rule. If the answer is no, we use a process called *MUS extraction* to find a short certificate of this fact which can be translated into a human-readable proof. By successively proving stronger theorems and using the insights obtained through MUS extraction, we arrived at results as presented in their full generality in this paper.

5.1 SAT Encoding

"For n voters and 4 alternatives, is there a voting rule f that satisfies Condorcet-consistency and participation?"

A natural encoding of this question into propositional logic proceeds like this: Generate all profiles over 4 alternatives with at most n voters. For each such profile R, introduce 4 propositional variables $x_{R,\mathtt{a}}$, $x_{R,\mathtt{b}}$, $x_{R,\mathtt{c}}$, $x_{R,\mathtt{d}}$, where the intended meaning of $x_{R,\mathtt{a}}$ is

$$x_{R,\mathtt{a}} \text{ is set true} \iff f(R) = \mathtt{a}.$$

We then add clauses requiring that for each profile R, $f(R)$ takes *exactly one* value, and we add clauses requiring f to be Condorcet-consistent and satisfy participation.

Sadly, the encoding sketched above is not tractable for the values of n that we are interested in: the number of variables and clauses used grows as $\Theta(24^n)$, because there are $4! = 24$ possible preference relations over 4 alternatives and thus 24^n profiles with n voters. For $n = 7$, this leads to more than 400 billion variables, and for $n = 15$ we exceed 10^{22} variables.

To escape this combinatorial explosion, we will temporarily restrict our attention to *pairwise* voting rules. This means that we assign an outcome alternative $f(T)$ to every weighted tournament T. We then define a voting rule that assigns the outcome $f(T_R)$ to each preference profile R, where T_R is the weighted tournament induced by R.

The number of tournaments induced by profiles with n voters grows much slower than the number of profiles—our computer enumeration suggests a growth of order about 1.5^n. This much more manageable (yet still exponential) growth allows us to consider problem instances up to $n \approx 16$ which turns out to be *just* enough.

Other than referring to (weighted) tournaments instead of profiles, our encoding into logic now proceeds exactly like before. For each tournament T, we introduce the variables $x_{T,\mathtt{a}}$, $x_{T,\mathtt{b}}$, $x_{T,\mathtt{c}}$, $x_{T,\mathtt{d}}$ and define the formulas

$$\textsf{non-empty}_T := x_{T,\mathtt{a}} \vee x_{T,\mathtt{b}} \vee x_{T,\mathtt{c}} \vee x_{T,\mathtt{d}}$$

$$\textsf{mutex}_T := \bigwedge_{\mathtt{x} \neq \mathtt{y}} (\neg x_{T,\mathtt{x}} \vee \neg x_{T,\mathtt{y}})$$

With our intended interpretation of the variables $x_{T,\mathtt{x}}$, all models of $\bigwedge_T \textsf{non-empty}_T \wedge \textsf{mutex}_T$ are functions from tournaments into $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$. (The word mutex abbreviates 'mutual exclusion' and corresponds to the voting rule selecting a unique winner.)

Since we are interested in voting rules that satisfy participation, we also need to encode this property. To this end, let $T = T_R$ be a tournament induced by R and let $\succcurlyeq$ be a preference relation. Define $T + \succcurlyeq := T_{R+\succcurlyeq}$. (The tournament $T + \succcurlyeq$ is independent of the choice of R.) We define

$$\textsf{participation}_{T,\succcurlyeq} := \bigwedge_{\mathtt{x}} \left(x_{T,\mathtt{x}} \to \bigvee_{\mathtt{y} \succcurlyeq \mathtt{x}} x_{T+\succcurlyeq,\mathtt{y}} \right).$$

Requiring f to be Condorcet-consistent is straightforward: if tournament T admits b as the Condorcet winner, we add

$$\textsf{condorcet}_T := \neg x_{T,\mathtt{a}} \wedge x_{T,\mathtt{b}} \wedge \neg x_{T,\mathtt{c}} \wedge \neg x_{T,\mathtt{d}},$$

and we add similar formulas for each tournament that admits a Condorcet winner. Then the models of the conjunction of all the non-empty, mutex, participation, and condorcet formulas are precisely the pairwise voting rules satisfying Condorcet-consistency and participation.

By adapting the condorcet formulas, we can impose more stringent conditions on f—this is how our results for maximin and Kemeny extensions are obtained. We can also use this to exclude Pareto-dominated alternatives, and to require f to always pick from the top cycle.

For some purposes it will be useful not to include the mutex clauses in our final formula. Models of this formula then correspond to *set-valued* voting rules that satisfy participation interpreted according to the optimistic preference extension. See Section 7 for results in this setting.

5.2 SAT Solving and MUS Extraction

The formulas we have obtained above are all given in *conjunctive normal form* (CNF), and thus can be evaluated without further transformations by any off-the-shelf SAT solver. In order to physically produce a CNF formula as described, we employ a straightforward Python script that performs a breadth-first search to discover all weighted tournaments with up to n voters (see Algorithm 1 for a schematic overview of the program). The script outputs a CNF formula in the standard DIMACS format, and also outputs a file that, for each variable $x_{T,\mathtt{x}}$, records the tournament T and alternative $\mathtt{x}$ it denotes. This is necessary because the DIMACS format uses uninformative variable descriptors (consecutive integers) and memorizing variable meanings allows us to interpret the output of the SAT solver.

Algorithm 1 Generate formula for up to n voters

$T_0 \leftarrow$ {weighted tournament on $\{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}$ with all edges having weight 0}.
for $k = 1, \dots, n$ **do**
  $T_k \leftarrow \emptyset$
  for $T \in T_{k-1}$ **do**
    for $\succcurlyeq \in \mathcal{R}$ **do**
      Calculate $T' := T + \succcurlyeq$
      Add T' to T_k
      Write $\mathsf{non\text{-}empty}_{T'}, \mathsf{mutex}_{T'}, \mathsf{condorcet}_{T'}$
      Write $\mathsf{participation}_{T,\succcurlyeq}$

As an example, the output formula for $n = 15$ in DIMACS format has a size of about 7 GB and uses 50 million variables and 2 billion clauses, taking 6.5 hours to write. Plingeling [3], a popular SAT solver that we used for all results in this paper, solves this formula in 50 minutes of wall clock time, half of which is spent parsing the formula.

In case a given instance is satisfiable, the solver returns a satisfying assignment, giving us an existence proof and a concrete example for a voting rule satisfying participation (and any further requirements imposed). In case a given instance in unsatisfiable, we would like to have short certificate of this fact as well. One possibility for this is having the SAT solver output a resolution proof (in DRUP format, say). This yields a machine-checkable proof, but has two major drawbacks: the generated proofs can be uncomfortably large [24], and they do not give human-readable insights about *why* the formula is unsatisfiable.

We handle this problem by computing a *minimal unsatisfiable subset (MUS)* of the unsatisfiable CNF formula. An MUS is a subset of the clauses of the original formula which itself is unsatisfiable, and is minimally so: removing any clause from it yields a satisfiable formula. We used the tools MUSer2 [2] and MARCO [27] to extract MUSes. If an unsatisfiable formula admits a very small MUS, it is often possible to obtain a human-readable proof of unsatisfiability from it [8, 4].

Note that for purposes of extracting human-readable proofs, it is desirable for the MUS to be as small as possible, and also to refer to as few different tournaments as possible. The first issue can be addressed by running the MUS extractor repeatedly, instructing it to order clauses randomly (note that clause sets of different cardinalities can be minimally unsatisfiable with respect to set inclusion); similarly, we can use MARCO to enumerate all MUSes and look for small ones. The second issue can be addressed by computing a *group MUS*: here, we partition the clauses of the CNF formula into *groups*, and we are looking for a minimal set of groups that together are unsatisfiable. In our case, the clauses referring to a given tournament T form a group. In practice, finding a group MUS first and then finding a standard (clause-level) MUS within the group MUS yielded sets of size much smaller than MUSes returned without the intermediate group-step (often by a factor of 10).

To translate an MUS into a human-readable proof, we created another program that visualized the MUS in a convenient form.[1] Indeed, this program outputs the 'proof diagrams' like Figure 1 that appear throughout this paper (though we re-typeset them). We think that interpreting these diagrams is quite natural (and is perhaps even easier than reading a textual translation). More importantly, the automatically produced graphs allowed us to quickly judge the quality of an extracted MUS.

5.3 Incremental Proof Discovery

The SAT encoding described in Section 5.1 only concerns pairwise voting rules, yet none of the (negative) results in this paper require or use this assumption. This is the product of multiple rounds of generating and evaluating SAT formulas, extracting MUSes, and using the insights generated by this as 'educated guesses' to solve more general problems.

Following the process as described so far led to a proof that for 4 alternatives and 12 voters, there is no pairwise Condorcet extension that satisfies participation. That proof used the assumption of pairwiseness, i.e., it assumed that the voting rule returns the same alternative on profiles inducing the same weighted tournament. However, intriguingly, the preference profiles mentioned in the proof did not contain all $4! = 24$ possible preference relations over $\{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}$. In fact, it only used 10 of the possible orders. Further, each profile included $R_0 = \{\mathtt{abdc}, \mathtt{bdca}, \mathtt{cabd}, \mathtt{dcab}\}$ as a subprofile. As we argued at the start of Section 5.1, it is intractable to search over the entire space of preference profiles. On the other hand, it is much easier to merely search over all extensions of R_0 that contain at most $n = 12$ voters and only contain copies of the 10 orders previously identified. The SAT formula produced by doing exactly this turned out to be unsatisfiable, and a small MUS extracted from it gave rise to Theorem 3.

The proof of Theorem 6 for 17 voters was obtained by running Algorithm 1 with T_0 initialized to the weighted tournament induced by the initial profile R used in the proof of Theorem 3. Before finding this tournament, we tried various other tournaments as T_0, including ones featuring in Moulin's original proof, and ones occurring at other steps in the proof of Theorem 3, but R turned out to give the best

[1] Roughly, the visualization program proceeds by drawing an edge for every $\mathsf{participation}_{T,\succcurlyeq}$ clause that occurs in the MUS, and marks the nodes for which $\mathsf{condorcet}_T$ clauses appear in the MUS.

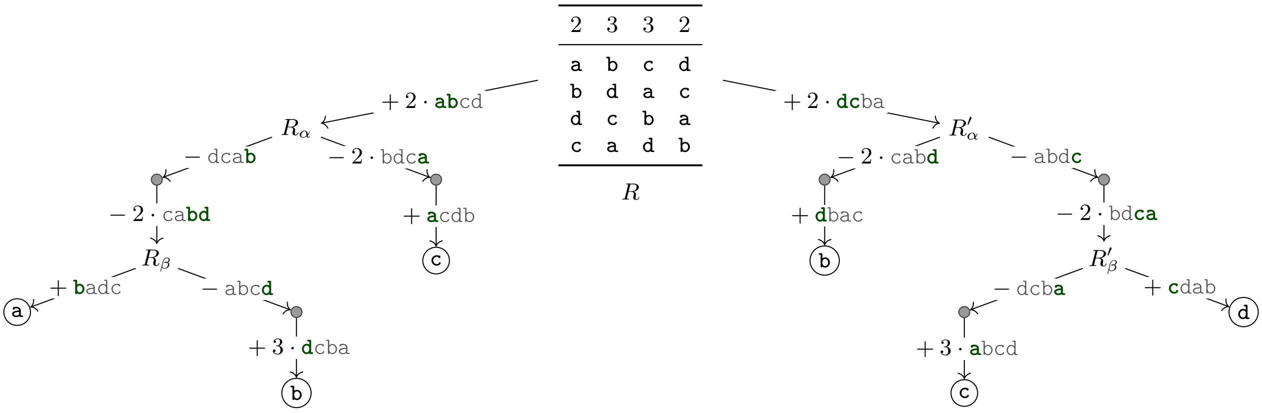


Figure 1: Computer-aided proof of Theorem 3 in graphical form, showing that there is no Condorcet extension that satisfies participation for $m \geqslant 4$ and $n \geqslant 12$. See Section 4 for an explanation of how to read this diagram.

(and indeed a tight) bound, and additionally exhibits a lot of symmetry that was also present in the MUS we extracted.

6. MAIN RESULT

We are now in a position to state and prove our main claim that Condorcet extensions cannot avoid the no-show paradox for 12 or more voters (when there are at least 4 alternatives), and that this result is optimal.

Theorem 3. *There is no Condorcet extension that satisfies participation for $m \geqslant 4$ and $n \geqslant 12$.*

Proof. The proof follows the structure depicted in Figure 1. Let R be the preference profile shown there.

Since R remains fixed after relabelling alternatives according to $\mathtt{abcd} \mapsto \mathtt{dcba}$, we may assume without loss of generality that $f(R) \in \{\mathtt{a}, \mathtt{b}\}$. (An explicit proof in case $f(R) \in \{\mathtt{c}, \mathtt{d}\}$ is indicated in Figure 1.)

By participation, it follows from $f(R) \in \{\mathtt{a}, \mathtt{b}\}$ that also $f(R_\alpha := R + 2 \cdot \mathtt{abcd}) \in \{\mathtt{a}, \mathtt{b}\}$ since the voters with preferences abcd cannot be worse off by joining the electorate. If $f(R_\alpha) = \mathtt{a}$, again by participation, removing 2 voters with preferences bdca does not change the winning alternative (so $f(R_\alpha - 2 \cdot \mathtt{bdca}) = \mathtt{a}$), and neither does adding acdb, so $f(R_\alpha - 2 \cdot \mathtt{bdca} + \mathtt{acdb}) = \mathtt{a}$, which, however, is in conflict with $R_\alpha - 2 \cdot \mathtt{bdca} + \mathtt{acdb}$ having a Condorcet winner, c.

Thus we must have $f(R_\alpha) = \mathtt{b}$, which implies that $f(R_\alpha - \mathtt{dcab}) = \mathtt{b}$, and thus $f(R_\beta := R_\alpha - \mathtt{dcab} - 2 \cdot \mathtt{cabd}) \in \{\mathtt{b}, \mathtt{d}\}$.

We again proceed by cases: If $f(R_\beta) = \mathtt{b}$, we can add a voter badc to obtain a profile with Condorcet winner a, which contradicts participation. But then, if $f(R_\beta) = \mathtt{d}$, we get that $f(R_\beta - \mathtt{abcd}) = \mathtt{d}$ and, by another application of participation, that $f(R_\beta - \mathtt{abcd} + 3 \cdot \mathtt{dcba}) = \mathtt{d}$ in contrast to the existence of Condorcet winner b, a contradiction.

If $m > 4$, we add bad alternatives $\mathtt{x}_1, \mathtt{x}_2, \ldots, \mathtt{x}_{m-4}$ to the bottom of R and all other voters. By Lemma 1, f chooses from $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}\}$ at each step, completing the proof. □

The following result establishes that our bound on the number of voters is tight. A very useful feature of our computer-aided approach is that we can easily add additional requirements for the desired voting rule. Two common requirements for voting rules are that they should only return alternatives that are *Pareto-optimal* and contained in the *top cycle* (also known as the *Smith set*) (see, e.g., [17]).

Theorem 4. *There is a Condorcet extension f that satisfies participation for $m = 4$ and $n \leqslant 11$. Moreover, f is pairwise, Pareto-optimal, and a refinement of the top cycle.*

The Condorcet extension f is given as a look-up table, which is derived from the output of a SAT solver. The look-up table lists all $1,204,215$ weighted tournaments inducible by up to 11 voters and assigns each an output alternative (see Figure 2 for an excerpt). The relevant text file has a size of 28 MB (gzipped 4.5 MB) and is available as part of an arXiv version of this paper [10].

Comparing this voting rule with known voting rules, it turns out that it picks a maximin winner in 99.8% and a Kemeny winner in 98% of all weighted tournaments. Moreover, the rule agrees with the maximin rule with lexicographic tie-breaking on 95% of weighted tournaments. The similarity with the maximin rule is interesting insofar as a well-documented flaw of the maximin rule is that it fails to be a refinement of the top cycle (and may even return Condorcet losers). Our computer-generated rule always picks from the top cycle while it remains very close to the maximin rule.

80% of the considered weighted tournaments admit a Condorcet winner, which uniquely determines the output of the rule; this can be used to reduce the size of the look-up table.

7. SET-VALUED VOTING RULES

A drawback of voting rules, as we defined them so far, is that that the requirement to always return a single winner is in conflict with basic fairness conditions, namely anonymity and neutrality. A large part of the social choice literature therefore deals with set-valued voting rules, where ties are usually assumed to be eventually broken by some tie-breaking mechanism.

A *set-valued voting rule* (sometimes known as a voting *correspondence* or as an *irresolute* voting rule) is a function $F \colon \mathcal{R}^{\mathcal{E}(\mathcal{N})} \to 2^A \setminus \{\emptyset\}$ that assigns each preference profile R a non-empty set of alternatives. The function F is a *(set-valued) Condorcet extension* if for every preference profile R that admits a Condorcet winner x, we have $F(R) = \{\mathtt{x}\}$.

Following the work of Pérez [29] and Jimeno et al. [22], we seek to study the occurrence of the no-show paradox in

```
a,#1,(1,1,1,1,1,1)          a,#11,(9,11,3,9,1,-9)
a,#1,(1,1,1,1,1,-1)         a,#11,(11,9,3,7,1,-9)
a,#1,(1,1,1,-1,1,1)         c,#11,(5,-9,-1,-11,-1,7)
a,#1,(1,1,1,-1,-1,1)        c,#11,(5,-9,-1,-11,-1,5)
a,#1,(1,1,1,1,-1,-1)        c,#11,(3,-11,-1,-9,1,7)
a,#1,(1,1,1,-1,-1,-1)       c,#11,(3,-11,-3,-9,1,7)
b,#1,(-1,1,1,1,1,1)         c,#11,(3,-11,-3,-11,-1,7)
b,#1,(-1,1,1,1,1,-1)        b,#11,(-1,3,-5,-3,5,-3)
b,#1,(-1,-1,1,1,1,1)        b,#11,(-3,3,-7,-3,5,-3)
b,#1,(-1,-1,-1,1,1,1)       b,#11,(-3,1,-7,-3,5,-3)
b,#1,(-1,1,-1,1,1,-1)       c,#11,(-3,1,-5,-5,5,-1)
b,#1,(-1,-1,-1,1,1,-1)      a,#11,(3,7,11,-3,9,11)
c,#1,(1,-1,1,-1,1,1)        a,#11,(3,7,11,-3,9,9)
c,#1,(1,-1,1,-1,-1,1)       a,#11,(3,7,11,-5,9,11)
```

Figure 2: Excerpt of look-up table giving a pairwise Condorcet extension satisfying participation for $n \leqslant 11$ voters (from Theorem 4). Each row lists a weighted tournament as $(g_R(\mathtt{a},\mathtt{b}), g_R(\mathtt{a},\mathtt{c}), g_R(\mathtt{a},\mathtt{d}), g_R(\mathtt{b},\mathtt{c}), g_R(\mathtt{b},\mathtt{d}), g_R(\mathtt{c},\mathtt{d}))$ with a chosen alternative, and with the number of voters inducing the tournament.

this setting. To do so, we need to define appropriate notions of participation, and for this we will need to specify agents' preferences over *sets* of alternatives. Here, we use the *optimistic* and *pessimistic preference extensions*. An optimist prefers sets with better most-preferred alternative, while a pessimist prefers sets with better least-preferred alternative. For example, if $U = \{\mathtt{a},\mathtt{b},\mathtt{d}\}$ and $V = \{\mathtt{b},\mathtt{c}\}$, then an optimist with preferences abcd prefers U to V, while a pessimist prefers V to U. With these notions, we extend the participation property to set-valued voting rules.

Definition 3. *A set-valued voting rule F satisfies* optimistic participation *if* $\max_{\succcurlyeq_i} F(R + \succcurlyeq_i) \succcurlyeq_i \max_{\succcurlyeq_i} F(R)$.
A set-valued voting rule F satisfies pessimistic participation *if* $\min_{\succcurlyeq_i} F(R) \succcurlyeq_i \min_{\succcurlyeq_i} F(R - i)$.

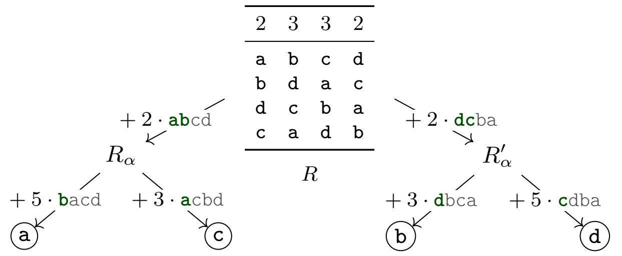
A set-valued voting rule F is called *resolute* if it always selects a single alternative, so that for all R we have $|F(R)| = 1$. A (single-valued) voting rule f is naturally identified with a resolute set-valued voting rule F; if f satisfies participation, then this F satisfies both optimistic and pessimistic participation. Hence, by Theorem 4, there is a (resolute) set-valued Condorcet extension F that satisfies both optimistic and pessimistic participation. However, there might be hope that allowing voting rules to be irresolute also allows for participation to be attainable for more voters, and indeed this is the case.

Theorem 5. *There is a set-valued Condorcet extension F that satisfies optimistic participation for $m = 4$ and $n \leqslant 16$, and also is Pareto-optimal and a refinement of the top cycle.*

The SAT solver indicates that no such set-valued voting rule is pairwise. Theorem 5 is optimal in the sense that optimistic participation cannot be achieved if we allow for one more voter.

Theorem 6. *There is no set-valued Condorcet extension that satisfies optimistic participation for $m \geqslant 4$ and $n \geqslant 17$.*

Proof. Let F be such a function, and consider the following 10-voter profile R:



Suppose that either $\mathtt{a} \in F(R)$ or $\mathtt{b} \in F(R)$. (The case of $\mathtt{c} \in F(R)$ or $\mathtt{d} \in F(R)$ is symmetric.) Then let $R_\alpha := R + 2 \cdot \mathtt{abcd}$. By optimistic participation, we then have either $\mathtt{a} \in F(R_\alpha)$ or $\mathtt{b} \in F(R_\alpha)$. If we had $\mathtt{a} \in F(R_\alpha)$, then also $\mathtt{a} \in F(R_\alpha + 3 \cdot \mathtt{acbd})$ but this profile has Condorcet winner c, and if $\mathtt{b} \in F(R_\alpha)$ then also $\mathtt{b} \in F(R_\alpha + 5 \cdot \mathtt{bacd})$ but this profile has Condorcet winner a. This is a contradiction.

This argument extends to more than 4 alternatives by appealing to a set-valued analogue of Lemma 1. □

Inspecting Moulin's original proof [28] shows that it also establishes an impossibility for optimistic participation (for 25 voters). Apparently unaware of this, Jimeno et al. [22] explicitly establish such a result for 27 voters and 5 alternatives. It is worth observing that each step of the proof of Theorem 6 involves *adding* voters to the current profile, and we never remove voters. In light of Definition 3, this is the reason why the proof establishes a result for optimistic participation. If we restrict ourselves to deleting voters, we obtain a result for pessimistic participation.

Theorem 7. *There is no set-valued Condorcet extension that satisfies pessimistic participation for $m \geqslant 4$ and $n \geqslant 14$. On the other hand, for $m = 4$ and $n \leqslant 13$, there exists such a set-valued voting rule.*

Proof Sketch. The proof has a similar structure to the proof of Theorem 3, displayed in Figure 1. The initial profile of this proof is $R + 2 \cdot \mathtt{abcd} + 2 \cdot \mathtt{dcba}$, taking R to be the profile of Figure 1. We further replace proof steps in which voters are added by similar ones where voters are deleted, and invoke pessimistic participation at each such step to obtain a contradiction. □

This result strengthens a result of Jimeno et al. [22], who show that for $m \geqslant 5$ no set-valued Condorcet extension satisfying a property called "weak translation invariance" can also satisfy pessimistic participation. Our proof does not need the extra assumption, already works for $m = 4$ alternatives, and uses just 14 instead of 971 voters.[2]

As previously observed, adding voters in our impossibility proofs corresponds to optimistic participation, while removing voters corresponds to pessimistic participation. In the proof of Theorem 3, we use both operations, which allows for a tighter bound of just 12 voters. In the set-valued setting, we can formulate this result in a slightly stronger way.

[2]The large number of voters is due to several applications of the "weak translation invariance" axiom, each of which adds $5! = 120$ voters to the preference profile under consideration.

Theorem 8. *There is no set-valued Condorcet extension that satisfies optimistic and pessimistic participation simultaneously for $m \geqslant 4$ and $n \geqslant 12$. On the other hand, for $m = 4$ and $n \leqslant 11$ such a set-valued rule exists (and also is Pareto-optimal and a refinement of the top cycle).*

Proof. Use the proof of Theorem 3, invoking optimistic participation for edges labelled with the addition of a voter (+), and invoking pessimistic participation for edges labelled with removal of a voter (−). On the other hand, the (single-valued) voting rule of Theorem 4 clearly satisfies both optimistic and pessimistic participation. □

The preference extension combining the optimistic and pessimistic preference extension is also known as the *Egli-Milner extension*.

8. PROBABILISTIC VOTING RULES

A *probabilistic voting rule* (also known as a *social decision scheme*) assigns to each preference profile R a probability distribution (or *lottery*) over A. Thus, a probabilistic voting rule might assign a fair coin flip between a and b as the outcome of an election.

Formally, let $\Delta(A) = \{\mathbf{p} : A \to [0,1] : \sum_{\mathtt{x}\in A} \mathbf{p}(\mathtt{x}) = 1\}$ be the set of lotteries over A; a lottery $\mathbf{p} \in \Delta(A)$ assigns probability $\mathbf{p}(\mathtt{x})$ to alternative x. A probabilistic voting rule f is a function $f : \mathcal{R}^{\mathcal{E}(\mathcal{N})} \to \Delta(A)$. In this context, we say that f is a *Condorcet extension* if $f(R)$ puts probability 1 on the Condorcet winner of R whenever it exists: if R admits x as the Condorcet winner, then $f(R)(\mathtt{x}) = 1$.

As in the set-valued case, we need a notion of comparing outcomes in order to extend the definition of participation. Here, we use the concept of *stochastic dominance (SD)*.

Definition 4. *Let $\succcurlyeq \in \mathcal{R}$ be a preference relation over A, and let $\mathbf{p},\mathbf{q} \in \Delta(A)$ be lotteries. Then $\mathbf{p}$ is (weakly) SD-preferred over $\mathbf{q}$ by $\succcurlyeq$ if for each alternative x, we have*

$$\textstyle\sum_{\mathtt{y}\succcurlyeq\mathtt{x}} \mathbf{p}(y) \geqslant \sum_{\mathtt{y}\succcurlyeq\mathtt{x}} \mathbf{q}(y).$$

In this case, we write $\mathbf{p} \succcurlyeq^{\mathrm{SD}} \mathbf{q}$.

For example, the lottery $\frac{2}{3}\mathtt{a} + \frac{1}{3}\mathtt{c}$ is SD-preferred to the lottery $\frac{1}{3}\mathtt{a} + \frac{1}{3}\mathtt{b} + \frac{1}{3}\mathtt{c}$ by a voter with preferences abcd. A voter with preferences bacd will feel the other way around. The main appeal of stochastic dominance stems from the following equivalence: $\mathbf{p} \succcurlyeq^{\mathrm{SD}} \mathbf{q}$ if and only if $\mathbf{p}$ yields at least as much von-Neumann-Morgenstern utility as $\mathbf{q}$ under any utility function that is consistent with the ordinal preferences $\succcurlyeq$. Using this notion of comparing lotteries, we can define participation analogously to the previous settings.

Definition 5. *A probabilistic voting rule f satisfies* strong SD-participation *if $f(R) \succcurlyeq_i^{\mathrm{SD}} f(R-i)$ for all $R \in \mathcal{R}^N$ and $i \in N$ with $N \in \mathcal{E}(\mathcal{N})$.*

Any (single-valued) voting rule f can be seen as a probabilistic voting rule that puts probability 1 on its chosen alternative. If f satisfies participation, then this derived probabilistic voting rule is easily seen to satisfy strong SD-participation. Hence Theorem 4 gives us a probabilistic Condorcet extension that satisfies strong SD-participation for $n \leqslant 11$ voters and $m = 4$ alternatives.

We now establish a connection between strong SD-participation and the set-valued notions of participation that we considered in Section 7. This connection will allow us to translate the impossibility results we obtained there to the probabilistic setting. To set up this connection, let us define the *support* of a lottery $\mathbf{p} \in \Delta(A)$ to be $\mathrm{supp}(\mathbf{p}) := \{\mathtt{x} \in A : \mathbf{p}(\mathtt{x}) > 0\}$.

Proposition 1. *Let f be a probabilistic voting rule satisfying strong SD-participation. Let $F = \mathrm{supp} \circ f$ be the support of f, i.e., $F(R) = \mathrm{supp}(f(R))$ for all profiles R. Then F satisfies both optimistic and pessimistic participation.*

Proof. We only verify optimistic participation; the pessimistic case is similar. Let R be a preference profile with electorate $N \in \mathcal{E}(\mathcal{N})$, and let $i \in \mathcal{N} \setminus N$ be a voter with preferences $\succcurlyeq_i$. Let $\mathtt{x} = \max_{\succcurlyeq_i} F(R)$, and let $U = \{\mathtt{y} : \mathtt{y} \succcurlyeq_i \mathtt{x}\}$. We need to show that $\max_{\succcurlyeq_i} F(R + \succcurlyeq_i) \succcurlyeq_i \mathtt{x}$, by finding an alternative $\mathtt{y} \in U$ that is in the support of $f(R + \succcurlyeq_i)$.

But since f satisfies strong SD-participation, we have

$$\textstyle\sum_{\mathtt{y}\in U} f(R + \succcurlyeq_i)(\mathtt{y}) \geqslant \sum_{\mathtt{y}\in U} f(R)(\mathtt{y}) > 0,$$

where the strict inequality follows from the definition of the support and of x. Hence some alternative from U is in the support of $f(R + \succcurlyeq_i)$, as required. □

Putting these results together with the impossibility result of Theorem 8, we obtain the following.

Theorem 9. *There is no probabilistic Condorcet extension that satisfies strong SD-participation for $n \geqslant 12$ and $m \geqslant 4$. On the other hand, for $m = 4$ and $n \leqslant 11$, such a probabilistic voting rule exists.*

Theorem 9 resolves an open problem mentioned by Brandl et al. [5, Sec. 6].

9. CONCLUSIONS AND FUTURE WORK

We have given tight results delineating in which situations no-show paradoxes must occur. As such, our results nicely complement recent advances to satisfy Condorcet-consistency and participation by exploiting uncertainties of the voters about their preferences or about the voting rule's tie-breaking mechanism [4, 5, 6].

Due to unmanageable branching factors when there are 5 alternatives (and hence 5! = 120 possible preference relations), we were unable to check using our approach whether no-show paradoxes occur with even less voters when the number of alternatives grows. It would be interesting to gain a deeper understanding of the computer-generated Condorcet extension that satisfies participation for up to 11 voters. So far, we only know that it (slightly) differs from all Condorcet extensions that are usually considered in the literature. As a first step, it would be desirable to obtain a representation of this rule that is more concise than a look-up table.

Another interesting topic for future research is to find optimal bounds for a variant of the no-show paradox due to Sanver and Zwicker [32], in which participation is weakened to half-way monotonicity. Their proof requires 702 voters.

Acknowledgments

Christian Geist is supported by Deutsche Forschungsgemeinschaft under grant BR 2312/9-1. Dominik Peters is supported by EPSRC. Part of this work was conducted while Dominik Peters visited TUM, supported by the COST Action IC1205 on Computational Social Choice.

REFERENCES

[1] D. Baumeister, P. Faliszewski, J. Lang, and J. Rothe. Campaigns for lazy voters: truncated ballots. In *Proc. of 11th AAMAS Conference*, pages 577–584. IFAAMAS, 2012.

[2] A. Belov and J. Marques-Silva. MUSer2: An efficient MUS extractor. *Journal on Satisfiability, Boolean Modeling and Computation*, 8:123–128, 2012.

[3] A. Biere. Lingeling, Plingeling and Treengeling entering the SAT competition 2013. In *Proc. of SAT Competition 2013*, pages 51–52, 2013.

[4] F. Brandl, F. Brandt, C. Geist, and J. Hofbauer. Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proc. of 24th IJCAI*, pages 18–24. AAAI Press, 2015.

[5] F. Brandl, F. Brandt, and J. Hofbauer. Incentives for participation and abstention in probabilistic social choice. In *Proc. of 14th AAMAS Conference*, pages 1411–1419. IFAAMAS, 2015.

[6] F. Brandl, F. Brandt, and J. Hofbauer. Welfare maximization entices participation. 2015. Working Paper.

[7] F. Brandt. Set-monotonicity implies Kelly-strategyproofness. *Social Choice and Welfare*, 45(4): 793–804, 2015.

[8] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565–602, 2016.

[9] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[10] F. Brandt, C. Geist, and D. Peters. Optimal bounds for the no-show paradox via SAT solving. Technical report, http://arxiv.org/abs/1602.08063, 2016.

[11] D. E. Campbell and J. S. Kelly. A strategy-proofness characterization of majority rule. *Economic Theory*, 22(3):557–568, 2003.

[12] S. Chatterjee and A. Sen. Automated reasoning in social choice theory — some remarks. *Mathematics in Computer Science*, 8(1):5–10, 2014.

[13] V. Conitzer and T. Sandholm. Complexity of mechanism design. In *Proc. of 18th UAI Conference*, pages 103–110, 2002.

[14] P. Dasgupta and E. Maskin. On the robustness of majority rule. *Journal of the European Economic Association*, 6(5):949–973, 2008.

[15] M. de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, 1785. Facsimile published in 1972 by Chelsea Publishing Company, New York.

[16] Y. Desmedt and E. Elkind. Equilibria of plurality voting with abstentions. In *Proc. of 11th ACM-EC Conference*, pages 347–356. ACM Press, 2010.

[17] P. C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

[18] P. C. Fishburn and S. J. Brams. Paradoxes of preferential voting. *Mathematics Magazine*, 56(4):207–214, 1983.

[19] A. Fréchette, N. Newman, and K. Leyton-Brown. Solving the station repacking problem. In *Proc. of 30th AAAI Conference*. AAAI Press, 2016.

[20] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[21] R. Holzman. To vote or not to vote: What is the quota? *Discrete Applied Mathematics*, 22(2):133–141, 1988.

[22] J. L. Jimeno, J. Pérez, and E. García. An extension of the Moulin No Show Paradox for voting correspondences. *Social Choice and Welfare*, 33(3):343–459, 2009.

[23] K. Kardel. Participation and group-participation in social choice theory. Master's thesis, Technische Universität München, 2014.

[24] B. Konev and A. Lisitsa. A SAT attack on the Erdős discrepancy conjecture. In *Theory and Applications of Satisfiability Testing–SAT 2014*, pages 219–226. Springer, 2014.

[25] J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer, 1997.

[26] D. Lepelley and V. Merlin. Scoring run-off paradoxes for variable electorates. *Economic Theory*, 17(1):53–80, 2000.

[27] M. H. Liffiton, A. Previti, A. Malik, and J. Marques-Silva. Fast, flexible MUS enumeration. *Constraints*, pages 1–28, 2015.

[28] H. Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45:53–64, 1988.

[29] J. Pérez. The strong no show paradoxes are a common flaw in Condorcet voting correspondences. *Social Choice and Welfare*, 18(3):601–616, 2001.

[30] D. Ray. On the practical possibility of a 'no show paradox' under the single transferable vote. *Mathematical Social Sciences*, 11(2):183–189, 1986.

[31] J. Rothe, editor. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer, 2015.

[32] M. R. Sanver and W. S. Zwicker. One-way monotonicity as a form of strategy-proofness. *International Journal of Game Theory*, 38(4):553–574, 2009.

[33] M. Schulze. Condorcet and Participation. Election-Methods mailing list. http://lists.electorama.com/pipermail/election-methods-electorama.com/2003-October/011042.html, 2003. (Accessed: 7 Nov 2015).

[34] W. D. Smith. "Participation failure" is forced in Condorcet methods with at least 4 candidates. The Center for Range Voting. http://rangevoting.org/CondPF.html, 2007. (Accessed: 7 Nov 2015).

[35] P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[36] D. R. Woodall. Monotonicity of single-seat preferential election rules. *Discrete Applied Mathematics*, 77 (1):81–98, 1997.

PROVING THE INCOMPATIBILITY OF EFFICIENCY AND STRATEGYPROOFNESS VIA SMT SOLVING [4]

Peer-reviewed Conference Paper

Authors: F. Brandl, F. Brandt, and C. Geist

In: Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI), pages 116–122. AAAI Press, 2016

Abstract: Two important requirements when aggregating the preferences of multiple agents are that the outcome should be economically efficient and the aggregation mechanism should not be manipulable. In this paper, we provide a computer-aided proof of a sweeping impossibility using these two conditions for randomized aggregation mechanisms. More precisely, we show that every efficient aggregation mechanism can be manipulated for *all* expected utility representations of the agents' preferences. This settles a conjecture by Aziz *et al.* and strengthens a number of existing theorems, including statements that were shown within the special domain of assignment. Our proof is obtained by formulating the claim as a satisfiability problem over predicates from real-valued arithmetic, which is then checked using an SMT (satisfiability modulo theories) solver. To the best of our knowledge, this is the first application of SMT solvers in computational social choice.

Contribution of thesis author: Computer-aided methods and results; literature review (computer-aided methods); project and paper management

Proving the Incompatibility of Efficiency and Strategyproofness via SMT Solving

Florian Brandl, Felix Brandt, Christian Geist
Technische Universität München
{brandlfl,brandtf,geist}@in.tum.de

Abstract

Two important requirements when aggregating the preferences of multiple agents are that the outcome should be economically efficient and the aggregation mechanism should not be manipulable. In this paper, we provide a computer-aided proof of a sweeping impossibility using these two conditions for randomized aggregation mechanisms. More precisely, we show that every efficient aggregation mechanism can be manipulated for *all* expected utility representations of the agents' preferences. This settles a conjecture by Aziz *et al.* [2013b] and strengthens a number of existing theorems, including statements that were shown within the special domain of assignment. Our proof is obtained by formulating the claim as a satisfiability problem over predicates from real-valued arithmetic, which is then checked using an SMT (satisfiability modulo theories) solver. To the best of our knowledge, this is the first application of SMT solvers in computational social choice.

1 Introduction

Models and results from microeconomic theory, in particular from game theory and social choice, have proven to be very valuable when reasoning about computational multiagent systems. Two fundamental notions in this context are efficiency—no agent can be made better off without making another one worse off—and strategyproofness—no agent can obtain a more preferred outcome by manipulating his preferences. Gibbard [1973] and Satterthwaite [1975] have shown that every strategyproof social choice function is either dictatorial or imposing. Hence, strategyproofness can only be achieved at the cost of discriminating among the agents or among the alternatives. One natural possibility to restore fairness, which is particularly popular in computer science, is to allow for randomization. Functions that map a profile of individual preferences to a probability distribution over alternatives (a so-called *lottery*) are known as *social decision schemes (SDSs)*.

Generalizing his previous result, Gibbard [1977] proved that the only strategyproof and *ex post* efficient social decision schemes are randomizations over dictatorships. Gibbard's notion of strategyproofness requires that no agent is better off by manipulating his preferences for *some* expected utility representation of the agents' ordinal preferences. This condition is quite demanding because an SDS may be deemed manipulable just because it can be manipulated for a contrived and highly unlikely utility representation. In this paper, we adopt a weaker notion of strategyproofness, first used by Postlewaite and Schmeidler [1986] and popularized by Bogomolnaia and Moulin [2001]. This notion requires that no agent should be better off by manipulating his preferences for *all* expected utility representations of the agents' preferences. At the same time, we use a stronger notion of efficiency than Gibbard [1977]. This notion is defined in analogy to our notion of strategyproofness and requires that no agent can be made better off for *all* utility representations of the agents' preferences, without making another one worse off for *some* utility representation. This type of efficiency was introduced by Bogomolnaia and Moulin [2001] and is also known as ordinal efficiency or SD-efficiency where SD stands for stochastic dominance.

Our main result establishes that no anonymous and neutral SDS satisfies efficiency and strategyproofness. This settles a conjecture by Aziz *et al.* [2013b] and generalizes theorems by Aziz *et al.* [2013b], Aziz *et al.* [2014], and Brandl *et al.* [2016b]. It also strengthens related statements by Zhou [1990], Bogomolnaia and Moulin [2001], and Katta and Sethuraman [2006], which were shown within the special domain of assignment.

Our proof of this theorem heavily relies on computer-aided solving techniques. Some of these have already been applied in computational social choice, where, due to the rigorous axiomatic foundation, computer-aided theorem proving appears to be a particularly promising line of research. Perhaps the best known result in this context stems from Tang and Lin [2009], who reduce well-known impossibility results, such as Arrow's theorem, to finite instances, which can then be checked by a Boolean satisfiability (SAT) solver. Their work has sparked a number of contributions which, besides using this general idea for more complex settings or axioms, focus on proving *novel* results [Geist and Endriss, 2011; Brandl *et al.*, 2015; Brandt *et al.*, 2016; Brandt and Geist, 2016].

In this paper, we go beyond the SAT-based techniques of previous contributions by designing an SMT (satisfiability modulo theories) encoding that captures axioms for *randomized* social choice. SMT can be viewed as an enriched form of the satisfiability problem (SAT) where Boolean variables are replaced by statements from a *theory*, such as specific data types or arithmetics. Similar to SAT, there is a range of SMT solvers developed by an active community that runs annual competitions [Barrett *et al.*, 2013]. Typically, SMT solvers are used as backends for verification tasks such as the verification of software. To capture axioms about lotteries, we use the theory of (quantifier-free) linear real arithmetic. Solving this version of SMT can be seen as an extension to *linear programming* in which arbitrary Boolean operators are allowed to connect (in-)equalities.

We follow the idea of Brandt and Geist [2016] and extract a *minimal unsatisfiable set (MUS)* of constraints in order to verify our result. Despite its relatively complex 94 (non-trivial) constraints, the MUS enables manual and computer-aided verification of the encoding, and, hence, releases any need to verify our program for generating it.

2 The Model

Let A be a finite set of m alternatives and $N = \{1, \dots, n\}$ a set of agents. A *(weak) preference relation* is a complete and transitive binary relation on A. The preference relation reported by agent i is denoted by $\succsim_i$, and the set of all preference relations by $\mathcal{R}$. In accordance with conventional notation, we write $\succ_i$ for the strict part of $\succsim_i$, i.e., $x \succ_i y$ if $x \succsim_i y$ but not $y \succsim_i x$, and $\sim_i$ for the indifference part of $\succsim_i$, i.e., $x \sim_i y$ if $x \succsim_i y$ and $y \succsim_i x$. A preference relation $\succsim_i$ is *linear* if $x \succ_i y$ or $y \succ_i x$ for all distinct alternatives $x, y \in A$. We will compactly represent a preference relation as a comma-separated list with all alternatives among which an agent is indifferent placed in a set. For example, $x \succ_i y \sim_i z$ is represented by $\succsim_i\colon x, \{y, z\}$. A *preference profile* $R = (\succsim_1, \dots, \succsim_n)$ is an n-tuple containing a preference relation $\succsim_i$ for each agent $i \in N$. The set of all preference profiles is thus given by $\mathcal{R}^N$. For a given $R \in \mathcal{R}^N$ and $\succsim \in \mathcal{R}$, $R^{i \mapsto \succsim}$ denotes a preference profile identical to R except that $\succsim_i$ is replaced with $\succsim$, i.e., $R^{i \mapsto \succsim} = R \setminus \{(i, \succsim_i)\} \cup \{(i, \succsim)\}$.

2.1 Social Decision Schemes

Our central objects of study are social decision schemes: functions that map a preference profile to a *lottery (or probability distribution)* over the alternatives. The set of all lotteries over A is denoted by $\Delta(A)$, i.e., $\Delta(A) = \{p \in \mathbb{R}^A_{\geq 0}\colon \sum_{x \in A} p(x) = 1\}$, where $p(x)$ is the probability that p assigns to x. Then, formally, a *social decision scheme (SDS)* is a function $f\colon \mathcal{R}^N \to \Delta(A)$. By $\text{supp}(p)$ we denote the *support* of a lottery $p \in \Delta(A)$, i.e., the set of all alternatives to which p assigns positive probability. Two common minimal fairness conditions for SDSs are anonymity and neutrality, i.e., symmetry with respect to agents and alternatives, respectively. Formally, *anonymity* requires that $f(R) = f(R \circ \sigma)$ for all $R \in \mathcal{R}^N$ and permutations $\sigma\colon N \to N$ over agents. *Neutrality*, on the other hand, is defined via permutations over alternatives. An SDS f is *neutral* if $f(R)(x) = f(\pi(R))(\pi(x))$ for all $R \in \mathcal{R}^N$, permutations $\pi\colon A \to A$, and $x \in A$.[1]

2.2 Efficiency and Strategyproofness

Many important properties of SDSs, such as efficiency and strategyproofness, require us to reason about the preferences that agents have over lotteries. This is commonly achieved by assuming that in a preference profile R every agent i, in addition to this preference relation $\succsim_i$, is equipped with a von Neumann-Morgenstern (vNM) *utility function* $u_i^R\colon A \to \mathbb{R}$. By definition, a utility function u_i^R has to be consistent with the ordinal preferences, i.e., for all $x, y \in A$, $u_i^R(x) \geq u_i^R(y)$ iff $x \succsim_i y$. A *utility representation* u then associates with each preference profile R an n-tuple $(u_1^R, \dots, u_n^R)$ of such utility functions. Whenever the preference profile R is clear from the context, the superscript will be omitted and we write u_i instead of the more cumbersome u_i^R.

Given a utility function u_i, agent i prefers lottery p to lottery q iff the expected utility for p is at least as high as that of q. With slight abuse of notation the domain of utility functions can be extended in the canonical way to $\Delta(A)$ by letting

$$u_i(p) = \sum_{x \in A} p(x) u_i(x).$$

It is straightforward to define efficiency and strategyproofness using expected utility. For a given utility representation u and a preference profile R, a lottery p *u-(Pareto-)dominates* another lottery q if

$$u_i(p) \geq u_i(q) \text{ for all } i \in N, \text{ and}$$
$$u_i(p) > u_i(q) \text{ for some } i \in N.$$

An SDS f is *u-efficient* if it never returns u-dominated lotteries, i.e., for all $R \in \mathcal{R}^N$, $f(R)$ is not u-dominated. The notion of u-strategyproofness can be defined analogously: for a given utility representation u, an SDS can be *u-manipulated* if there are $R \in \mathcal{R}^N$, $i \in N$, and $\succsim \in \mathcal{R}$ such that

$$u_i^R(f(R^{i \mapsto \succsim})) > u_i^R(f(R)).$$

An SDS is *u-strategyproof* if it cannot be u-manipulated.

The assumption that the vNM utility functions of all agents (and thus their complete preferences *over lotteries*) are known is quite unrealistic. Often even the agents themselves are uncertain about their preferences over lotteries and only know their ordinal preferences over alternatives.[2] A natural way to model this uncertainty is to leave the utility functions unspecified and instead *quantify over all utility functions* that are consistent with the agents' ordinal preferences. This model leads to much weaker notions of efficiency and strategyproofness.

[1] $\pi(R)$ is the preference profile obtained from π by replacing $\succsim_i$ with $\succsim_i^\pi$ for every $i \in N$, where $\pi(x) \succsim_i^\pi \pi(y)$ if and only if $x \succsim_i y$.

[2] When assuming that all agents possess vNM utility functions, these utility functions could be taken as inputs for the aggregation function. Such aggregation functions are called *cardinal decision schemes* (see, e.g., [Dutta *et al.*, 2007]). In addition to the fact that concrete vNM utility functions are typically unavailable, their representation may require infinite space.

Definition 1. An SDS is *efficient* if it never returns a lottery that is u-dominated for all utility representations u.

As mentioned in the introduction, this notion of efficiency is also known as *ordinal efficiency* or *SD-efficiency* (see, e.g., [Bogomolnaia and Moulin, 2001; Aziz *et al.*, 2014; 2015]). The relationship to stochastic dominance will be discussed in more detail in Section 4.2.

Example 1. For illustration consider $A = \{a, b, c, d\}$ and the preference profile $R = (\succsim_1, \dots, \succsim_4)$,

$$\succsim_1 : \{a, c\}, \{b, d\}, \qquad \succsim_2 : \{b, d\}, \{a, c\},$$
$$\succsim_3 : \{a, d\}, b, c, \qquad \succsim_4 : \{b, c\}, a, d$$

Observe that the lottery $7/24\, a + 7/24\, b + 5/24\, c + 5/24\, d$, which is returned by the well-known SDS *random serial dictatorship (RSD)*, is u-dominated by $1/2\, a + 1/2\, b$ for every utility representation u. Hence, any SDS that returns this lottery for the profile R would not be efficient. On the other hand, the lottery $1/2\, a + 1/2\, b$ is not u-dominated, which can, for instance, be checked via linear programming (see Lemma 4).

We can also define a weak notion of strategyproofness in analogy to our notion of efficiency.

Definition 2. An SDS is *strategyproof* if it cannot be u-manipulated for all utility representations u.

Alternatively, there is a stronger version of strategyproofness by Gibbard [1977], in which an SDS should not be u-manipulable for *some* utility representation u.

For more information concerning the relationship between sets of possible utility functions and preference extensions, such as stochastic dominance, the reader is referred to Aziz *et al.* [2015].

3 The Result

Our main result shows that efficiency and strategyproofness are incompatible with basic fairness properties. Aziz *et al.* [2013b] raised the question whether there exists an anonymous, efficient, and strategyproof SDS. When additionally requiring neutrality, we can answer this question in the negative.

Theorem 1. *If $m \geq 4$ and $n \geq 4$, there is no anonymous and neutral SDS that satisfies efficiency and strategyproofness.*

The proof of Theorem 1, which heavily relies on computer-aided solving techniques, is discussed in Section 4. Let us first discuss the independence of the axioms and relate the result to existing theorems. *RSD* satisfies all axioms *except efficiency*; another SDS known as *maximal lotteries* satisfies all axioms *except strategyproofness* (cf. [Aziz *et al.*, 2013b]). Serial dictatorship, the deterministic version of *RSD*, satisfies neutrality, efficiency, and strategyproofness *but violates anonymity*. It is unknown whether Theorem 1 still holds when dropping the assumption of neutrality. Our proof, however, only requires a technical weakening of neutrality (cf. Section 4.1).

3.1 Related Results for Social Choice

Our result generalizes several existing results and is closely related to a number of results in subdomains of social choice. Aziz *et al.* [2013b] proved a weak version of Theorem 1 for the rather restricted class of majoritarian SDSs, i.e., SDSs whose outcome may only depend on the pairwise majority relation. This statement has later been generalized by Aziz *et al.* [2014] to all SDSs whose outcome only depends on the *weighted* majority relation. More recently, Brandl *et al.* [2016b] have shown that while random dictatorship is efficient and strategyproof on the domain of linear preferences, it cannot be extended to the full domain of weak preferences without violating at least one of these properties. Their theorem, which also assumes anonymity and neutrality, is a direct consequence of Theorem 1. Other impossibility results have been obtained for stronger notions of efficiency and strategyproofness, which weakens the corresponding statements [Aziz *et al.*, 2014].

3.2 Related Results for Assignment

A subdomain of social choice that has been thoroughly studied in the literature is the assignment (aka house allocation or two-sided matching with one-sided preferences) domain. An assignment problem can be associated with a social choice problem by letting the set of alternatives be the set of deterministic allocations and postulating that agents are indifferent among all allocations in which they receive the same object (see, e.g., [Aziz *et al.*, 2013a]).[3] Thus, impossibility results for the assignment setting can be interpreted as impossibility results for the social choice setting because they even hold in a smaller domain.

In the following we discuss impossibility results in the assignment domain which, if interpreted for the social choice domain, can be seen as weaker versions of Theorem 1 because they are based on stronger notions of efficiency or strategyproofness or require additional properties. In a very influential paper, Bogomolnaia and Moulin [2001] have shown that no randomized assignment mechanism satisfies both efficiency and a strong notion of strategyproofness while treating all agents equally. The underlying notion of strategyproofness is identical to the one used by Gibbard [1977] and prescribes that the SDS cannot be u-manipulated for *some* utility representation u. The result by Bogomolnaia and Moulin even holds when preferences over objects are linear. (Nevertheless, when transferred to the social choice domain, the preferences over allocations will contain ties.) In a related paper, Katta and Sethuraman [2006] proved that no assignment mechanism satisfies efficiency, strategyproofness, and envy-freeness for the full domain of preferences.

Settling a conjecture by Gale [1987], Zhou [1990] showed that no cardinal assignment mechanism satisfies u-efficiency and u-strategyproofness while treating all agents equally.[4] The relationship between Zhou's result and Theorem 1 is not obvious because Zhou's theorem concerns cardinal mechanisms, i.e., functions that take a utility profile rather than a

[3]Note that this transformation turns assignment problems with linear preferences over k objects into social choice problems with non-linear preferences over $k!$ allocations.

[4]The theorem by Zhou only requires that agents with the same utility function receive the same amount of utility but not necessarily the same assignment. Gale's original conjecture assumed equal treatment of equals.

preference profile as input. However, every cardinal assignment mechanism can be associated with an ordinal assignment mechanism. Hence, Theorem 1 implies that there is no anonymous, neutral, u-efficient, and u-strategyproof cardinal decision scheme.

4 Proving the Result

In this section, we first reduce the statement of Theorem 1 to the case of $m = 4$ and $n = 4$, which we then prove via SMT solving. We present an encoding for any finite instance of Theorem 1 as an SMT problem in the logic of (quantifier-free) linear real arithmetic (QF_LRA). For compatibility with different SMT solvers our encoding adheres to the SMT-LIB standard [Barrett *et al.*, 2010]. In total, we are going to design the following four types of SMT constraints: lottery definitions (Lottery), the orbit condition which models a part of neutrality (Orbit), strategyproofness (SP), and efficiency (Efficiency). Other conditions such as anonymity are taken care of by the representation of preference profiles.

We then, first, apply an SMT solver to show that this set of constraints for the case of $m = 4$ and $n = 4$ is unsatisfiable, i.e., no SDS f with the desired properties exists. Second, we explain how the output of the solver can be used to obtain a human-verifiable proof of this result.

But let us start with the reduction lemma before we turn to the concrete encoding in the following subsections.

Lemma 1. *If there is an anonymous and neutral SDS f that satisfies efficiency and strategyproofness for $|A| = m$ alternatives and $|N| = n$ agents then we can also find an SDS f' defined for $m' \leq m$ alternatives and $n' \leq n$ agents that satisfies the same properties.*

Proof. Let f be an anonymous and neutral SDS that satisfies efficiency and strategyproofness for m alternatives and n agents. We define a projection f' of f onto $A' \subseteq A, |A'| = m' \leq m$ and $N' = \{1, 2, \ldots, n'\} \subseteq N, n' \leq n$ that satisfies all required properties:

For every preference profile R' on A' and N', let $f'(R') = f(R)$, where R is defined by the following conditions:

$$\succsim_i \cap (A' \times A') = \succsim'_i \text{ for all } i \in N', \tag{1}$$

$$x \succ_i y \text{ for all } x \in A', y \in A \setminus A' \text{ and } i \in N', \tag{2}$$

$$y \sim_i z \text{ for all } y, z \in A \setminus A' \text{ and } i \in N', \text{ and} \tag{3}$$

$$y \sim_i z \text{ for all } y, z \in A \text{ and } i \in N \setminus N'. \tag{4}$$

Informally, by (1) agents in N' have the same preferences over alternatives from A' in R and R'. Moreover, by (2) they like every alternative in A' strictly better than every alternative not in A' and by (3) they are indifferent between all alternatives not in A'. Finally, by (4) all agents in $N \setminus N'$ are completely indifferent. With these conditions, R is uniquely specified given R', and only lotteries p with $\mathrm{supp}(p) \subseteq A'$ are efficient in R. Thus, f' is well-defined and it is left to show that f' inherits the relevant properties from f. The SDS f' is anonymous since f is anonymous and agents in N can only differ by their preferences over A'. Neutrality follows as f is neutral and all agents are indifferent between all alternatives not in A'. Efficiency is satisfied by f' since f is efficient and the same set of lotteries is efficient in R and R'. Finally, f' is strategyproof because f is strategyproof and the outcomes of f' under the two profiles R' and $(R')^{i \mapsto \succsim'}$ are equal to the outcomes of f under the two (extended) profiles R and $R^{i \mapsto \succsim}$, respectively. □

4.1 Framework, Anonymity, and Neutrality

For a given number of agents n and set of alternatives A, we encode an arbitrary SDS $f\colon \mathcal{R}^N \to \Delta(A)$ by a set of real-valued variables $p_{R,x}$ with $R \in \mathcal{R}^N$ and $x \in A$. Each $p_{R,x}$ then represents the probability with which alternative x is selected for profile R, i.e., $p_{R,x} = f(R)(x)$.

This encoding of lotteries leads to the first simple constraints for our SMT encoding, which ensure that for each preference profile R the corresponding variables $p_{R,x}$, $x \in A$ indeed encode a lottery:

$$\begin{aligned} &\sum_{x \in A} p_{R,x} = 1 \text{ for all } R \in \mathcal{R}^N, \text{ and} \\ &p_{R,x} \geq 0 \text{ for all } R \in \mathcal{R}^N \text{ and } x \in A. \end{aligned} \tag{Lottery}$$

We are now going to argue that, in conjunction with anonymity and neutrality (see Section 2), it suffices to consider these constraints for a subset of preference profiles. This is because, in contrast to the other axioms, we directly incorporate anonymity and neutrality into the structure of the encoding rather than formulating them as actual constraints. Similar to the construction involving canonical tournament representations by Brandt and Geist [2016], we model anonymity and neutrality by computing for each preference profile $R \in \mathcal{R}^N$ a *canonical representation* $R_\mathfrak{c} \in \mathcal{R}^N$ with respect to these properties. In this representation, two preference profiles R and R' are equal (i.e., $R_\mathfrak{c} = R'_\mathfrak{c}$) iff one can be transformed into the other by renaming the agents and alternatives. Equivalently, $R_\mathfrak{c} = R'_\mathfrak{c}$ iff, for every anonymous and neutral SDS f, the lotteries $f(R)$ and $f(R')$ are equal (modulo the renaming of the alternatives).

The SMT constraints and SMT variables are then instantiated only for these canonical representations $\mathcal{R}^N_\mathfrak{c} \subseteq \mathcal{R}^N$. Apart from enabling an encoding of anonymous and neutral SDSs without any explicit reference to permutations, this also offers a substantial performance gain compared to considering the full domain $\mathcal{R}^N$ of (non-anonymous and non-neutral) preference profiles.

Technically, we compute the canonical representation $R_\mathfrak{c}$ as follows: Let $R = (\succsim_1, \ldots, \succsim_n) \in \mathcal{R}^N$ be a preference profile. First, we identify R with a function $r : \mathcal{R} \to \mathbb{N}$, which we call *anonymous preference profile*, and which counts the number of agents with a certain preference relation, i.e., $r(\succsim) = |\{i \in N \mid \succsim_i = \succsim\}|$, thereby ignoring the identity of the agents. This representation fully captures anonymity.

To additionally enforce neutrality, we had to resort to a computationally demanding, naive solution: given r, we compute all anonymous preference profiles $\pi(r)$ that can be achieved via a permutation $\pi\colon A \to A$, and, among those profiles, choose the one $\pi_{\text{lexmin}}(r)$ with lexicographically minimal values (for some fixed ordering of preference relations). For the canonical representation $R_\mathfrak{c}$ we then pick any preference profile $R \in \mathcal{R}^N$ which agrees with $\pi_{\text{lexmin}}(r)$, for instance, by again using the same fixed ordering of preference relations. Fortunately, this approach is still feasible for the small numbers of alternatives with which we are dealing.

While this representation of preference profiles does not completely capture neutrality—the *orbit condition* (see [Brandt and Geist, 2016]) is missing—this weaker version suffices to prove the impossibility. In favor of simpler proofs we, however, include the simple constraints corresponding to a randomized version of the orbit condition.

In our context, an *orbit* O of a preference profile R is an equivalence class of alternatives. Two alternatives $x, y \in A$ are considered equivalent if $\pi(x) = y$ for some permutation $\pi\colon A \to A$ that maps the anonymous preference profile associated with R to itself (i.e., π is an automorphism of the anonymous preference profile). In such a situation, every anonymous and neutral SDS has to assign equal probabilities to x and y. We hence require that, for each orbit $O \in \mathcal{O}_R$ of a (canonical) profile R, the probabilities $p_{R,x}$ are equal for all alternatives $x \in O$. As an SMT constraint, this reads

$$p_{R,x} = p_{R,y} \tag{Orbit}$$

for all $R \in \mathcal{R}^N_{\mathfrak{c}}$, $O \in \mathcal{O}_R$, and $x, y \in O$.

Example 2. Consider the anonymous preference profile r based on R from Example 1 and the permutation

$$\pi = \begin{pmatrix} a & b & c & d \\ b & a & d & c \end{pmatrix}.$$

As $\pi(r) = r$ (and since no other non-trivial permutation has this property) the set of orbits of R is $\mathcal{O}_R = \{\{a, b\}, \{c, d\}\}$.

4.2 Stochastic Dominance

In order to avoid quantifying over utility functions, we leverage well-known representations of efficiency and strategyproofness via *stochastic dominance (SD)* (cf. [Bogomolnaia and Moulin, 2001; McLennan, 2002; Aziz *et al.*, 2015]). A lottery p *stochastically dominates* a lottery q for an agent i (short: $p \succsim_i^{SD} q$) if for every alternative x, lottery p is at least as likely as lottery q to yield an alternative at least as good as x. Formally,

$$p \succsim_i^{SD} q \text{ iff } \sum_{y \succsim_i x} p(y) \geq \sum_{y \succsim_i x} q(y) \text{ for all } x \in A.$$

When $p \succsim_i^{SD} q$ and not $q \succsim_i^{SD} p$ we write $p \succ_i^{SD} q$.

As an example, consider the preference relation $\succsim_i\colon a, b, c$. We then have that

$$({}^2\!/\!_3\, a + {}^1\!/\!_3\, c) \succ_i^{SD} ({}^1\!/\!_3\, a + {}^1\!/\!_3\, b + {}^1\!/\!_3\, c)$$

while ${}^2\!/\!_3\, a + {}^1\!/\!_3\, c$ and b are incomparable according to stochastic dominance.

Lemma 2. *A lottery p SD-dominates another lottery q for an agent i iff $u_i^{\succsim_i}(p) \geq u_i^{\succsim_i}(q)$ for every utility function $u_i^{\succsim_i}$. As a consequence,*

1. *an SDS f is efficient iff, for all $R \in \mathcal{R}^N$, there is no lottery p such that $p \succsim_i^{SD} f(R)$ for all $i \in N$ and $p \succ_i^{SD} f(R)$ for some $i \in N$, and*
2. *an SDS f is manipulable iff there exist a preference profile R, an agent i, and a preference relation $\succsim$ such that $f(R^{i \mapsto \succsim}) \succ_i^{SD} f(R)$.*

In words, Lemma 2 shows that an SDS f is efficient if and only if $f(R)$ is Pareto-efficient with respect to stochastic dominance for all preference profiles R. Secondly, f is manipulable if and only if some agent can misrepresent his preferences to obtain a lottery that he prefers to the lottery obtained by sincere voting with respect to stochastic dominance.

Encoding Strategyproofness

Starting from the above equivalence, encoding strategyproofness as an SMT constraint is now a much simpler task. For each (canonical) preference profile $R \in \mathcal{R}^N_{\mathfrak{c}}$, agent $i \in N$, and preference relation $\succsim \in \mathcal{R}$, we encode that the manipulated outcome $f(R^{i \mapsto \succsim})$ is not SD-preferred to the truthful outcome $f(R)$ by agent i:

$$\begin{aligned}
&\neg\left(f(R^{i \mapsto \succsim}) \succ_i^{SD} f(R)\right) \\
\equiv\, & f(R^{i \mapsto \succsim}) \not\succsim_i^{SD} f(R) \vee f(R) \succsim_i^{SD} f(R^{i \mapsto \succsim}) \\
\equiv & \left((\exists x \in A) \sum_{y \succsim_i x} f(R^{i \mapsto \succsim})(y) < \sum_{y \succsim_i x} f(R)(y)\right) \vee \\
& \left((\forall x \in A) \sum_{y \succsim_i x} f(R^{i \mapsto \succsim})(y) \overset{(*)}{\leq} \sum_{y \succsim_i x} f(R)(y)\right) \\
\equiv & \left(\bigvee_{x \in A} \sum_{y \succsim_i x} p_{(R^{i \mapsto \succsim})_{\mathfrak{c}}, \pi_{\mathfrak{c}}^{R^{i \mapsto \succsim}}(y)} < \sum_{y \succsim_i x} p_{R,y}\right) \vee \\
& \left(\bigwedge_{x \in A} \sum_{y \succsim_i x} p_{(R^{i \mapsto \succsim})_{\mathfrak{c}}, \pi_{\mathfrak{c}}^{R^{i \mapsto \succsim}}(y)} \overset{(**)}{=} \sum_{y \succsim_i x} p_{R,y}\right),
\end{aligned} \tag{SP}$$

where $\pi_{\mathfrak{c}}^{R^{i \mapsto \succsim}}$ stands for a permutation of alternatives that (together with a potential renaming of alternatives) leads from $R^{i \mapsto \succsim}$ to $(R^{i \mapsto \succsim})_{\mathfrak{c}}$. The inequality $(*)$ can be replaced by the equality $(**)$ since the case of at least one strict inequality is captured by the corresponding disjunctive condition one line above.

Encoding Efficiency

While Lemma 2 helps to formulate efficiency as an SMT axiom it is not yet sufficient since a quantification over the set of all lotteries $\Delta(A)$ remains. In order to get rid of this quantifier, we apply two lemmas by Aziz *et al.* [2015]. The first lemma states that efficiency of a lottery only depends on its support. The second lemma shows that deciding whether a lottery is efficient reduces to solving a linear program; for this statement we include a (slightly simplified) proof in favor of a self-contained presentation.

Lemma 3 (Aziz *et al.*, 2015)**.** *A lottery $p \in \Delta(A)$ is efficient iff every lottery $p' \in \Delta(A)$ with $\mathrm{supp}(p') \subseteq \mathrm{supp}(p)$ is efficient.*

Lemma 4 (Aziz *et al.*, 2015)**.** *Whether a lottery $p \in \Delta(A)$ is efficient for a given preference profile R can be computed in polynomial time by solving a linear program.*

Proof. Given the equivalence from Lemma 2, a lottery p is easily seen to be efficient iff the optimal objective value of the following linear program is zero (since then there is no lottery q that SD-dominates p):

$$\max_{q,r} \sum_{i \in N} \sum_{x \in A} r_{i,x} \quad \text{subject to}$$

$$\sum_{y \succsim_i x} q_y - r_{i,x} = \sum_{y \succsim_i x} p_y \text{ for all } x \in A, i \in N,$$

$$\sum_{x \in A} q_x = 1, \ q_x \geq 0 \text{ for all } x \in A,$$

$$r_{i,x} \geq 0 \text{ for all } x \in A, i \in N.$$

□

Recall that an SDS is efficient if it never returns a dominated lottery. By Lemma 3, this is equivalent to never returning a lottery with *inefficient support*. To capture this, we encode, for each (canonical) preference profile $R \in \mathcal{R}_{\mathfrak{c}}^N$, that the probability for at least one alternative in every (inclusion-minimal) inefficient support $I_R \subseteq A$ is zero:

$$\bigvee_{x \in I_R} p_{R,x} = 0. \qquad \text{(Efficiency)}$$

4.3 Restricted Domains

Since RSD (cf. Section 3) is known to satisfy both strategyproofness as well as efficiency for up to 3 alternatives, a search for an impossibility has to start at $m = 4$ alternatives. For $n = 3$ agents, the encoding is solved as satisfiable; for $n = 4$, an encoding of the full domain, unfortunately, becomes prohibitively large. Hence, for $m = 4$ and $n = 4$, one has to carefully optimize the domain under consideration, on the one hand, to include a sufficient number of profiles for a successful proof, and, on the other hand, not to include too many profiles, which would prevent the solver from terminating within a reasonable amount of time.

The following incremental strategy was found to be successful. We start with a specific profile R, from which we only consider sequences of potential manipulations as long as (in each step) the manipulated individual preferences are not too distinct from the truthful preferences. To this end, we measure the magnitude of manipulations by the Kendall tau distance τ, which counts pairwise disagreements between R_i and R'_i (see also [Sato, 2013]). A change in the individual preferences of an agent will be called a *k-manipulation* if $\tau(R_i, R'_i) \leq k$. Then, for example, strategically swapping two alternatives is a 2-manipulation, and breaking or introducing a tie between two alternatives is a 1-manipulation.

On the domain which starts from the preference profile R from Example 1 and allows sequences of $(1, 2, 1, 2)$-manipulations, we were able to prove the result within a few minutes of running-time.[5,6] On similar, but smaller domains (e.g., $(1, 2, 2)$) the axioms are still compatible.

4.4 Verification of Correctness

For verification of the result, one would ideally construct a human-readable proof from the output of the SMT solver. While the approach described by Brandt and Geist [2016] for SAT solving—of finding a *minimal unsatisfiable set (MUS)* of constraints, i.e., an inclusion-minimal set of constraints such that this set is still unsatisfiable—is theoretically also applicable to SMT solving, it is less clear how these "proof ingredients" have to be combined.[7] The proof object that z3 can produce, which also contains information of how the MUS constraints have to be combined, unfortunately, is too long and complicated for humans to parse.

Hence two aspects of our approach still deserve verification: the correctness of the constraints in the MUS and the unsatisfiability of the MUS. In addition to manual inspection of the constraints and some sanity-checks,[8] we have certified in Isabelle/HOL that all constraints logically follow from the original axioms presented in Section 2. This also releases any need to verify our program for generating the constraints. The unsatisfiability of the MUS, on the other hand, has been verified by the solvers CVC4, MathSAT, Yices2, z3, and even by the Isabelle/HOL kernel.

Furthermore, based on the MUS, a proof of Theorem 1 which no longer relies on SMT solving has been created in Isabelle/HOL. This proof, however, is tedious to verify by hand since it is rather large (more than 500 lines of code) and offers little insight.

5 Conclusion

In this paper, we have leveraged computer-aided solving techniques to prove a sweeping impossibility for randomized aggregation mechanisms.

It seems unlikely that this proof would have been found without the help of computers because manual proofs of significantly weaker statements already turned out to be quite complex. Nevertheless, now that the theorem has been established, our computer-aided methods may guide the search for related, perhaps even stronger statements that allow for more intuitive proofs and provide more insights.

Generally speaking, we believe that SMT solving is applicable to a wide range of problems in randomized social choice. In particular, extending our result to the special domain of assignment (see Section 3.2) is desirable as this would strengthen a number of existing theorems. Other interesting questions are whether the impossibility still holds when weakening strategyproofness even further to BD-strategyproofness (see, e.g., [Aziz *et al.*, 2014]) or when omitting neutrality.

[5]The SMT solver MathSAT [Cimatti *et al.*, 2013] terminates quickly within less than 3 minutes with the suggested competition settings, whereas z3 [de Moura and Bjørner, 2008] requires some additional configuration, but then also supports core extraction within the same time frame.

[6]Showing the result on this domain implies a slightly stronger statement where strategyproofness only applies to "small" lies (of at most Kendall tau distance 2).

[7]Here we have an MUS of 94 constraints, not counting the (trivial) lottery definitions. This MUS, annotated with e.g., the 47 required canonical preference profiles, is available as part of an arXiv version of this paper [Brandl *et al.*, 2016a].

[8]Such as running solvers on multiple variants of the encoding which represent known theorems. This way, we reproduced (amongst others) the results by Bogomolnaia and Moulin [2001] and Katta and Sethuraman [2006], as well as the possibility result for $m < 4$.

6 Acknowledgments

This material is based upon work supported by the Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/10-1 and the TUM Institute for Advanced Study through a Hans Fischer Senior Fellowship. The authors also thank Manuel Eberl for the extensive verification work in Isabelle/HOL, Alberto Griggio and Mohammad Mehdi Pourhashem Kallehbasti for guidance on how to most effectively use MathSAT and z3, respectively, and three anonymous reviewers for their helpful comments.

References

[Aziz *et al.*, 2013a] H. Aziz, F. Brandt, and M. Brill. The computational complexity of random serial dictatorship. *Economics Letters*, 121(3):341–345, 2013.

[Aziz *et al.*, 2013b] H. Aziz, F. Brandt, and M. Brill. On the tradeoff between economic efficiency and strategyproofness in randomized social choice. In *Proc. of 12th AAMAS Conference*, pages 455–462. IFAAMAS, 2013.

[Aziz *et al.*, 2014] H. Aziz, F. Brandl, and F. Brandt. On the incompatibility of efficiency and strategyproofness in randomized social choice. In *Proc. of 28th AAAI Conference*, pages 545–551. AAAI Press, 2014.

[Aziz *et al.*, 2015] H. Aziz, F. Brandl, and F. Brandt. Universal Pareto dominance and welfare for plausible utility functions. *Journal of Mathematical Economics*, 60:123–133, 2015.

[Barrett *et al.*, 2010] C. Barrett, A. Stump, and C. Tinelli. The SMT-LIB Standard: Version 2.0. In *Proc. of 8th International Workshop on Satisfiability Modulo Theories*, 2010.

[Barrett *et al.*, 2013] C. Barrett, M. Deters, L. de Moura, A. Oliveras, and A. Stump. 6 years of SMT-COMP. *Journal of Automated Reasoning*, 50(3):243–277, 2013.

[Bogomolnaia and Moulin, 2001] A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328, 2001.

[Brandl *et al.*, 2015] F. Brandl, F. Brandt, C. Geist, and J. Hofbauer. Strategic abstention based on preference extensions: Positive results and computer-generated impossibilities. In *Proc. of 24th IJCAI*, pages 18–24. AAAI Press, 2015.

[Brandl *et al.*, 2016a] F. Brandl, F. Brandt, and C. Geist. Proving the incompatibility of efficiency and strategyproofness via SMT solving. Technical report, http://arxiv.org/abs/1604.05692, 2016.

[Brandl *et al.*, 2016b] F. Brandl, F. Brandt, and W. Suksompong. The impossibility of extending random dictatorship to weak preferences. *Economics Letters*, 141:44–47, 2016.

[Brandt and Geist, 2016] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565–602, 2016.

[Brandt *et al.*, 2016] F. Brandt, C. Geist, and D. Peters. Optimal bounds for the no-show paradox via SAT solving. In *Proc. of 15th AAMAS Conference*, pages 314–322. IFAAMAS, 2016.

[Cimatti *et al.*, 2013] A. Cimatti, A. Griggio, B. Schaafsma, and R. Sebastiani. The MathSAT5 SMT Solver. In *Proceedings of TACAS*, volume 7795 of *LNCS*, pages 93–107. Springer, 2013.

[de Moura and Bjørner, 2008] L. de Moura and N. Bjørner. Z3: An efficient SMT solver. In *Proceedings of TACAS*, volume 7795 of *LNCS*, pages 337–340. Springer, 2008.

[Dutta *et al.*, 2007] B. Dutta, H. Peters, and A. Sen. Strategy-proof cardinal decision schemes. *Social Choice and Welfare*, 28(1):163–179, 2007.

[Gale, 1987] D. Gale. College course assignments and optimal lotteries, 1987. Mimeo.

[Geist and Endriss, 2011] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[Gibbard, 1973] A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

[Gibbard, 1977] A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

[Katta and Sethuraman, 2006] A.-K. Katta and J. Sethuraman. A solution to the random assignment problem on the full preference domain. *Journal of Economic Theory*, 131(1):231–250, 2006.

[McLennan, 2002] A. McLennan. Ordinal efficiency and the polyhedral separating hyperplane theorem. *Journal of Economic Theory*, 105(2):435–449, 2002.

[Postlewaite and Schmeidler, 1986] A. Postlewaite and D. Schmeidler. Strategic behaviour and a notion of ex ante efficiency in a voting model. *Social Choice and Welfare*, 3(1):37–49, 1986.

[Sato, 2013] S. Sato. Strategy-proofness and the reluctance to make large lies: the case of weak orders. *Social Choice and Welfare*, 40(2):479–494, 2013.

[Satterthwaite, 1975] M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

[Tang and Lin, 2009] P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[Zhou, 1990] L. Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.

SOLVING-BASED ALGORITHMS

[5] A note on the McKelvey uncovered set and Pareto optimality. *Social Choice and Welfare*, 46(1):81–91, 2016 (with F. Brandt and P. Harrenstein).

[6] Identifying k-majority digraphs via SAT solving. In *Proceedings of the 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2014 (with F. Brandt and H. G. Seedig).

[7] Finding preference profiles of Condorcet dimension k via SAT. Technical report, http://arxiv.org/abs/1402.4303, 2014.

A NOTE ON THE MCKELVEY UNCOVERED SET AND PARETO OPTIMALITY [5]

Peer-reviewed Journal Paper

Authors: F. Brandt, C. Geist, and P. Harrenstein

In: Social Choice and Welfare, 46(1):81–91, 2016

Abstract: We consider the notion of Pareto optimality under the assumption that only the pairwise majority relation is known and show that the set of necessarily Pareto optimal alternatives coincides with the McKelvey uncovered set. As a consequence, the McKelvey uncovered set constitutes the coarsest Pareto optimal majoritarian social choice function. Moreover, every majority relation is induced by a preference profile in which the uncovered alternatives precisely coincide with the Pareto optimal ones. We furthermore discuss the structure of the McKelvey covering relation and the McKelvey uncovered set.

Contribution of thesis author: Computer-aided methods; results (except for proof of Theorem 1); mathematical model; literature review; project and paper management

Soc Choice Welf (2016) 46:81–91
DOI 10.1007/s00355-015-0904-5

ORIGINAL PAPER

A note on the McKelvey uncovered set and Pareto optimality

Felix Brandt[1] · **Christian Geist**[1] · **Paul Harrenstein**[2]

Received: 22 August 2014 / Accepted: 8 July 2015 / Published online: 11 August 2015

Abstract We consider the notion of Pareto optimality under the assumption that only the pairwise majority relation is known and show that the set of necessarily Pareto optimal alternatives coincides with the McKelvey uncovered set. As a consequence, the McKelvey uncovered set constitutes the coarsest Pareto optimal majoritarian social choice function. Moreover, every majority relation is induced by a preference profile in which the uncovered alternatives precisely coincide with the Pareto optimal ones. We furthermore discuss the structure of the McKelvey covering relation and the McKelvey uncovered set.

1 Introduction

Let A be a finite set of m alternatives. The preferences of an agent i over these alternatives are represented by a complete, transitive, and antisymmetric *preference*

This material is based on work supported by Deutsche Forschungsgemeinschaft under Grants BR 2312/7-2 and BR 2312/9-1. Paul Harrenstein is supported by the ERC under Advanced Grant 291528 ("RACE"). The paper has benefitted from discussions at the 12th Meeting of the Society for Social Choice and Welfare in Boston (June 2014) as well as the Dagstuhl Seminar on Computational Social Choice (June 2015). The authors furthermore thank Jean-François Laslier and Hans Georg Seedig for helpful discussions and technical support.

✉ Felix Brandt
brandtf@in.tum.de

Christian Geist
geist@in.tum.de

Paul Harrenstein
paul.harrenstein@cs.ox.ac.uk

1 Technische Universität München, Munich, Germany

2 University of Oxford, Oxford, UK

relation $R_i \subseteq A \times A$.[1] The interpretation of $(x, y) \in R_i$, usually denoted by $x \ R_i \ y$, is that agent i values alternative x at least as much as alternative y. In accordance with conventional notation, we write P_i for the strict part of R_i, i.e., $x \ P_i \ y$ if $x \ R_i \ y$ but not $y \ R_i \ x$. Since R_i is antisymmetric, $x \ P_i \ y$ iff $x \ R_i \ y$ and $x \neq y$. A *preference profile* R is a finite vector of pairs which associate each agent i with its corresponding preference relation R_i. For a given preference profile, N_R denotes the set of agents represented in R. In particular, we do not assume a fixed number of agents. For convenience, we furthermore define $n_R(x, y) := |\{i \in N_R \mid x \ R_i \ y\}|$. The relations $R_{\mathbf{Maj}}$ and $R_{\mathbf{Par}}$ of a given preference profile R are then given by

$$\begin{aligned} x \ R_{\mathbf{Maj}} \ y \quad &\text{iff} \quad n_R(x, y) \geq n_R(y, x), \quad \text{and} \\ x \ R_{\mathbf{Par}} \ y \quad &\text{iff} \quad n_R(x, y) = |N_R|. \end{aligned}$$

By convention, again $P_{\mathbf{Maj}}$ denotes the strict part of $R_{\mathbf{Maj}}$ and $P_{\mathbf{Par}}$ the strict part of $R_{\mathbf{Par}}$. Thus, $x \ P_{\mathbf{Par}} \ y$ iff $x \ R_i \ y$ for all i and $x \ P_j \ y$ for some j. The relation $R_{\mathbf{Maj}}$ will be called the *majority relation* and $P_{\mathbf{Par}}$ the *Pareto relation* of R, respectively. Note that, by definition, $R_{\mathbf{Maj}}$ is complete whereas $P_{\mathbf{Par}}$ is transitive and antisymmetric.

We say that a profile R is *consistent* with the majority relation $R'_{\mathbf{Maj}}$ of another profile R' if $R_{\mathbf{Maj}} = R'_{\mathbf{Maj}}$. In this case, R is also said to be consistent with R'.

An alternative $x \in A$ is called a *Condorcet winner* of a given (majority) relation $R_{\mathbf{Maj}}$ if it strictly dominates all other alternatives, i.e., if $x \ P_{\mathbf{Maj}} \ y$ for all $y \in A \setminus \{x\}$.

Given a preference profile R, an alternative x is said to be *Pareto optimal* whenever there is no alternative y with $y \ P_{\mathbf{Par}} \ x$, i.e., if x is a maximal element of $R_{\mathbf{Par}}$.[2]

A *social choice function* (*SCF*) f associates with every preference profile R over a set A of alternatives a nonempty subset $f(R)$ of A. An SCF f is called a *refinement* of another SCF g if $f(R) \subseteq g(R)$ for all preference profiles R. In short, we write $f \subseteq g$ in this case and also say that g is *coarser* than f. An example of an SCF is the *Pareto set PO*, which selects the alternatives that are Pareto optimal, i.e.,

$$PO(R) = \{x \in A \mid y \ P_{\mathbf{Par}} \ x \text{ for no } y \in A\}$$

An SCF f is called *Pareto optimal* if it is a refinement of the Pareto set, i.e., if $f \subseteq PO$.

We will restrict attention to so-called *majoritarian* SCFs, whose outcomes only depend on the majority relation, i.e., to SCFs f such that, for all preference profiles R, R',

$$R_{\mathbf{Maj}} = R'_{\mathbf{Maj}} \quad \text{implies} \quad f(R) = f(R').$$

[1] Antisymmetry is not required for any of our results to hold. In fact, Theorem 1 is even stronger when also assuming antisymmetric individual preferences (since this only increases the difficulty of constructing a suitable preference profile).

[2] Some authors call this *strong* Pareto optimality. In contrast, an alternative x would be *weakly* Pareto optimal if there is no alternative y with $y \ P_i \ x$ for all $i \in N_R$. In the case of antisymmetric preferences, the two notions coincide.

An interesting class of majoritarian SCFs are defined using certain refinements of the majority relation called *covering relations*. For a given covering relation, the *uncovered set* contains those alternatives that are not covered by any other alternative. For a comprehensive overview of the theory of covering relations and uncovered sets, see Duggan (2013). A range of varying definitions of the covering relation exist, all of which coincide when restricted to antisymmetric majority relations. We will be concerned with what Duggan refers to as *McKelvey* covering (Bordes 1983; McKelvey 1986).

In order to define the McKelvey covering relation, we need to introduce the notions of *strict* and *weak dominators* of a given alternative. The strict dominators $P_{\mathbf{Maj}}(x)$ of an alternative $x \in A$ are defined by the set of all alternatives $y \in A$ that are strictly majority preferred to x, i.e., $P_{\mathbf{Maj}}(x) := \{y \in A \mid y \; P_{\mathbf{Maj}} \; x\}$. Analogously, the weak dominators $R_{\mathbf{Maj}}(x)$ of an alternative $x \in A$ are defined as the set of all alternatives $y \in A$ that are weakly majority preferred to x, i.e., $R_{\mathbf{Maj}}(x) := \{y \in A \mid y \; R_{\mathbf{Maj}} \; x\}$

Let $C_{R_{\mathbf{Maj}}}$ denote the *(McKelvey) covering relation*, i.e., for any pair of alternatives x, y, the relation $x \; C_{R_{\mathbf{Maj}}} \; y$ holds iff each of the following three conditions is satisfied:

(i) $x \; P_{\mathbf{Maj}} \; y$,
(ii) $P_{\mathbf{Maj}}(x) \subseteq P_{\mathbf{Maj}}(y)$, and
(iii) $R_{\mathbf{Maj}}(x) \subseteq R_{\mathbf{Maj}}(y)$.

As can easily be seen, $C_{R_{\mathbf{Maj}}}$ is transitive and asymmetric. The *(McKelvey) uncovered set UC* is then defined as

$$UC(R) := \left\{y \in A \mid x \; C_{R_{\mathbf{Maj}}} \; y \text{ for no } x \in A\right\}.$$

Alternative, but equivalent, definitions of the McKelvey uncovered set were used by Dutta and Laslier (1999), Peris and Subiza (1999), and Brandt and Fischer (2008).

It is also well-known (see, e.g., Duggan 2013) that the McKelvey uncovered set can be characterized as the set of alternatives that dominate every other alternative in at most two steps (of which at most one may be a tie). Formally, the McKelvey uncovered set then consists of all alternatives $x \in A$ such that for all $y \in A \backslash \{x\}$ at least one of the following three conditions is satisfied:

(i) $x \; R_{\mathbf{Maj}} \; y$,
(ii) there is a $z \in A$ such that $x \; R_{\mathbf{Maj}} \; z \; P_{\mathbf{Maj}} \; y$, or
(iii) there is a $z \in A$ such that $x \; P_{\mathbf{Maj}} \; z \; R_{\mathbf{Maj}} \; y$.

For brevity, we will omit any reference to McKelvey in the following and just write "covering" and "uncovered set."

2 The structure of the McKelvey covering relation

In this section we consider the structural properties of the *covering relation* and observe that any transitive and asymmetric relation can be obtained as the covering relation of some preference profile.

Proposition 1 *Let $Q \subseteq A \times A$ be a binary relation. Then the following are equivalent:*

(i) *Q is transitive and asymmetric.*
(ii) *There exists a preference profile R such that $C_{R_{\mathbf{Maj}}} = Q$.*

Proof The implication from (ii) to (i) is immediate, as for every preference profile R the covering relation $C_{R_{\mathbf{Maj}}}$ is transitive and asymmetric. For the other direction, assume that Q is transitive and asymmetric and let R be a preference profile with $P_{\mathbf{Maj}} = Q$. By virtue of McGarvey's Theorem (McGarvey 1953) we know such a preference profile exists. It then remains to observe that $(x, y) \in C_{R_{\mathbf{Maj}}}$ iff $(x, y) \in Q$. The "if"-direction is immediate. For the "only if"-direction, consider an arbitrary edge $(x, y) \in Q$. By construction of R, also $x \; P_{\mathbf{Maj}} \; y$ and, by transitivity of $Q = P_{\mathbf{Maj}}$, we obtain $P_{\mathbf{Maj}}(x) \subseteq P_{\mathbf{Maj}}(y)$. To see that also $R_{\mathbf{Maj}}(x) \subseteq R_{\mathbf{Maj}}(y)$, consider an arbitrary $z \in R_M(x)$ and assume for contradiction that $y \; P_{\mathbf{Maj}} \; z$. Again by transitivity of $P_{\mathbf{Maj}}$, then $x \; P_{\mathbf{Maj}} \; z$, a contradiction. It follows that $z \in R_{\mathbf{Maj}}(y)$. □

Interestingly, the implication from (i) to (ii) does not hold if $R_{\mathbf{Maj}}$ is required to be antisymmetric (e.g., when the number of agents is odd). We are not aware of a non-trivial characterization of potential covering relations for this case.[3]

A result analogous to Proposition 1 was shown by Dushnik and Miller (1941) for the Pareto relation. They proved that for any transitive and asymmetric relation $Q \subseteq A \times A$ there exists a preference profile R such that $P_{\mathbf{Par}} = Q$. Note that this does not imply Proposition 1 as the majority relation outside Q may be very different in the preference profile R instantiating Q as the Pareto relation, and a preference profile R' instantiating Q as the covering relation. In our proof, in particular, all edges outside Q are required to be majority ties.

Regarding the internal structure of the *uncovered set*, it was already shown by Moulin (1986) that any complete binary relation without a (non-trivial) Condorcet winner is the majority relation between uncovered alternatives for some preference profile. While Moulin (1986) proved this result for tournaments only, the argument can easily be adapted to cover our setting, in which majority ties are allowed.

3 The McKelvey uncovered set and the Pareto set are majority-equivalent

In this section, we consider the relationship between the uncovered set and the Pareto set for profiles that yield the same majority relation (i.e., *consistent* preference profiles).

Our main result shows that for every preference profile R, we can find another preference profile R' consistent with $R_{\mathbf{Maj}}$ such that the uncovered set of R and the Pareto set of R' coincide. This has a number of consequences. First, if we assume that only the majority relation is known, the uncovered set not only coincides with the Pareto optimal alternatives for some consistent preference profile, but also consists of precisely those alternatives that are Pareto optimal for *every* consistent preference profile. Moreover, there exists a consistent profile in which *all* covered alternatives

[3] Consider, for instance, the simple case of $A = \{a, b, c\}$ and $Q = \{(a, b), (a, c)\}$, which is easily seen not to be the covering relation for any preference profile.

are Pareto dominated. Secondly, the theorem implies that the uncovered set can be characterized as the coarsest majoritarian Pareto optimal SCF.

Theorem 1 *For every preference profile R, there is another preference profile R' with $R_{\mathbf{Maj}} = R'_{\mathbf{Maj}}$ such that*

$$UC(R) = PO(R').$$

Proof Consider an arbitrary preference profile R. Then, for every alternative $y \notin UC(R)$ there is some $x \in UC(R)$ such that $x\ C_{R_{\mathbf{Maj}}}\ y$. Thus, let $UC(R) = \{x_1, \dots, x_\ell\}$ and associate with every $x_i \in UC(R)$ a (possibly empty) set $f(x_k) \subseteq \{x' \in A \mid x\ C_{R_{\mathbf{Maj}}}\ x'\}$ such that

$$f(x_k) \cap f(x_{k'}) = \emptyset \quad \text{for} \quad 1 \le k < k' \le \ell \quad \text{and} \quad f(x_1) \cup \dots \cup f(x_\ell) = A \backslash UC(R).$$

Define the relation $F \subseteq C_{R_{\mathbf{Maj}}}$ such that, for all alternatives x and y,

$$x\ F\ y \quad \text{iff} \quad x \in UC(R) \quad \text{and} \quad y \in f(x).$$

We construct a preference profile R' such that $R'_{\mathbf{Maj}} = R_{\mathbf{Maj}}$ and $R'_{\mathbf{Par}} = F$. It can readily be appreciated that then $UC(R) = PO(R')$, as desired.

For notational convenience we denote $f(x_k)$ by X_k. For any subset Y of alternatives, by $\overrightarrow{Y}$ and $\overleftarrow{Y}$ we denote an enumeration of Y and its inverse, respectively, i.e., if $\overrightarrow{Y} = y_1, \dots, y_k$ then $\overleftarrow{Y} = y_k, \dots, y_1$.

First, we generate the relation F. To this end, we introduce two agents, i_F and j_F with preferences given by the following two sequences:

$$i_F: x_1, \overrightarrow{X}_1, \dots, x_\ell, \overrightarrow{X}_\ell,$$
$$j_F: x_\ell, \overleftarrow{X}_\ell, \dots, x_1, \overleftarrow{X}_1.$$

Letting $R^F = (R_{i_F}, R_{j_F})$, we thus have $R^F_{\mathbf{Par}} = F$.

Furthermore, for every pair (v, w) in $R_{\mathbf{Maj}}$ that is not contained in F we also introduce two additional agents i_{vw} and j_{vw}. We distinguish five cases and denote by R^{vw} the profile $(R_{i_{vw}}, R_{j_{vw}})$. Without loss of generality and for notational convenience, we will usually assume that $v^+ = x_1$ and $w^+ = x_2$.

Case 1 There is a $v^+ \in UC(R)$ such that $v, w \in f(v^+)$. Let $V = f(v^+) \backslash \{v, w\}$. For an illustration see Fig. 1a. Then, define the preferences of i_{vw} and j_{vw} by the following lists:

$$i_{vw}: v^+, v, w, \overrightarrow{V}, x_2, \overrightarrow{X}_2, \dots, x_\ell, \overrightarrow{X}_\ell,$$
$$j_{vw}: x_\ell, \overleftarrow{X}_\ell, \dots, x_2, \overleftarrow{X}_2, v^+, \overleftarrow{V}, v, w.$$

Observe that in this case, we have $P^{vw}_{\mathbf{Maj}} = F \cup \{(v, w)\}$.

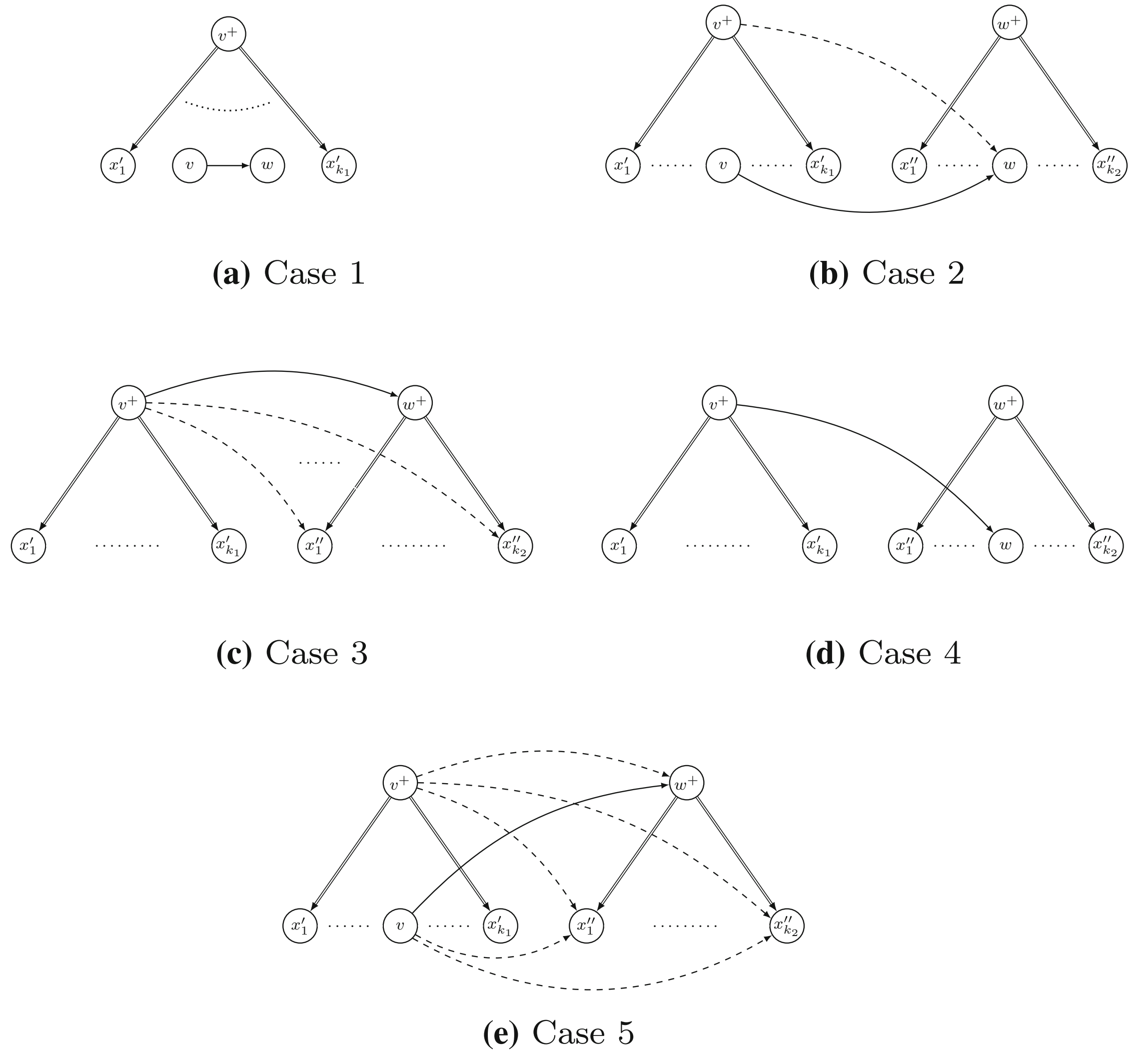


Fig. 1 The five cases distinguished in the Proof of Theorem 1. The relation F is indicated by *double arrows*

Case 2 There are $v^+, w^+ \in UC(R)$ with $v \in f(v^+)$ and $w \in f(w^+)$. Let $V = f(v^+)\backslash\{v\}$ and $W = f(w^+)\backslash\{w\}$. Observe that in this case we have that $v^+ \; C_{R_{\mathbf{Maj}}} \; v$ for R and, hence, also $v^+ \; P_{\mathbf{Maj}} \; w$. For an illustration see Fig. 1b. In this case define the preferences of i_{vw} and j_{vw} as

$$\begin{aligned} i_{vw}&: v^+, \vec{V}, v, w^+, w, \vec{W}, x_3, \vec{X}_3, \ldots, x_\ell, \vec{X}_\ell,\\ j_{vw}&: x_\ell, \overleftarrow{X}_\ell, \ldots, x_3, \overleftarrow{X}_3, w^+, \overleftarrow{W}, v^+, v, w, \overleftarrow{V}. \end{aligned}$$

It thus follows that $P^{vw}_{\mathbf{Maj}} = F \cup \{(v, w), (v^+, w)\}$.

Case 3 There are $v^+, w^+ \in UC(R)$ with $v = v^+$ and $w = w^+$. Let $V = f(v^+)$ and $W = f(w^+)$. Observe that in this case we have for R that $w^+ \; C_{R_{\mathbf{Maj}}} \; x$ and, hence, also $v^+ \; P_{\mathbf{Maj}} \; x$ for all $x \in W$. The situation is depicted in Fig. 1c. Now define

$$\begin{aligned} i_{vw}&: v^+, \vec{V}, w^+, \vec{W}, x_3, \vec{X}_3, \ldots, x_\ell, \vec{X}_\ell,\\ j_{vw}&: x_\ell, \overleftarrow{X}_\ell, \ldots, x_3, \overleftarrow{X}_3, v^+, w^+, \overleftarrow{W}, \overleftarrow{V}. \end{aligned}$$

Observe that now $P^{vw}_{\mathbf{Maj}} = F \cup \{(v^+, w^+)\} \cup \{(v^+, x) \mid x \in W\}$.

Case 4 There are $v^+, w^+ \in UC(R)$ with $v = v^+$ and $w \in f(w^+)$. Let $V = f(v^+)$ and $W = f(w^+) \backslash \{w\}$. The situation is depicted in Fig. 1d. In this case define the preferences of i_{vw} and j_{vw} as follows:

$$i_{vw} : v^+, \overrightarrow{V}, w^+, w, \overrightarrow{W}, x_3, \overrightarrow{X}_3, \ldots, x_\ell, \overrightarrow{X}_\ell,$$
$$j_{vw} : x_\ell, \overleftarrow{X}_\ell, \ldots, x_3, \overleftarrow{X}_3, w^+, \overleftarrow{W}, v^+, w, \overleftarrow{V}.$$

Accordingly, we also have $P^{vw}_{\mathbf{Maj}} = F \cup \{(v^+, w)\}$.

Case 5 There are $v^+, w^+ \in UC(R)$ with $v \in f(v^+)$ and $w = w^+$. Let $V = f(v^+) \backslash \{v\}$ and $W = f(w^+)$. Observe that in this case we have for R that $v^+ \; C_{R_{\mathbf{Maj}}} \; v$ and $w^+ \; C_{R_{\mathbf{Maj}}} \; x$ for all $x \in W$. As a consequence, also $v^+ \; P_{\mathbf{Maj}} \; w^+$, and thus $v^+ \; P_{\mathbf{Maj}} \; x$ and $v \; P_{\mathbf{Maj}} \; x$ for all $x \in W$. The situation is depicted in Fig. 1e. In this case define:

$$i_{vw} : v^+, \overrightarrow{V}, v, w^+, \overrightarrow{W}, x_3, \overrightarrow{X}_3, \ldots, x_\ell, \overrightarrow{X}_\ell,$$
$$j_{vw} : x_\ell, \overleftarrow{X}_\ell, \ldots, x_3, \overleftarrow{X}_3, v^+, v, w^+, \overleftarrow{W}, \overleftarrow{V}.$$

Hence, $P^{vw}_{\mathbf{Maj}} = F \cup (\{v^+, v\} \times (W \cup \{w^+\}))$.

Let $R_{\mathbf{Maj}} \backslash F$ be given by $\{(v_1, w_1), \ldots, (v_p, w_p)\}$ and consider the profile

$$R' = (R_1, \ldots, R_{2p+2}) = (R_{i_F}, R_{j_F}, R_{i_{v_1 w_1}}, \ldots, R_{i_{v_p w_p}}).$$

Some reflection reveals that $R'_{\mathbf{Maj}} = R_{\mathbf{Maj}}$. Also observe that $F \subseteq R_i$ for all $1 \le i \le 2p + 2$. Moreover, $x \; P_{i_F} \; y$ iff $y \; P_{j_F} \; x$, for all $(x, y) \notin F$. Hence, $R'_{\mathbf{Par}} = F$. Since, $x \notin UC(R)$ iff $y \; F \; x$ for some $y \in UC(R)$, it follows that $PO(R') = UC(R)$, which concludes the proof. □

For the next corollary we additionally need the (easy-to-prove) fact that the uncovered set UC is Pareto optimal, which to the best of our knowledge was first mentioned by Bordes (1983). We formalize it in the following lemma for the sake of completeness. Note that this result relies on the definition of the majority relation via *simple majority rule* (see, e.g., Gaertner 2009, p. 39). If majorities are defined via (the less common) *absolute* majority rule, then the uncovered set only satisfies *weak* Pareto optimality (Duggan 2013).

Lemma 1 (Bordes 1983) *The Pareto relation is a subrelation of the McKelvey covering relation. Hence, $UC \subseteq PO$.*

Proof Let R be a preference profile and $x, y \in A$ alternatives such that $x \; P_{\mathbf{Par}} \; y$. To show that then also $x \; C_{R_{\mathbf{Maj}}} \; y$, first suppose $z \; P_{\mathbf{Maj}} \; x$ for some alternative $z \in A$. Since individual preferences are assumed to be transitive it follows that $z \; P_{\mathbf{Maj}} \; y$. The case of $z \; R_{\mathbf{Maj}} \; x$ (implying $z \; R_{\mathbf{Maj}} \; y$) is analogous and $x \; P_{\mathbf{Maj}} \; y$ is an immediate consequence of $x \; P_{\mathbf{Par}} \; y$. □

The following corollary, which follows from Theorem 1 and Lemma 1, provides a characterization of the McKelvey uncovered set via Pareto optimality.[4]

Corollary 1 *A majoritarian SCF f is Pareto optimal iff $f \subseteq UC$. Consequently, the McKelvey uncovered set is the coarsest Pareto optimal majoritarian SCF.*

Proof Lemma 1 establishes that the McKelvey uncovered set UC (and any refinement of it) is Pareto optimal. To prove the other direction by contraposition, consider an arbitrary majoritarian SCF f such that $f \nsubseteq UC$. Then, there is a profile R and an alternative a such that $a \in f(R)$ and $a \notin UC(R)$. By Theorem 1, some preference profile R' exists such that $R_{\mathbf{Maj}} = R'_{\mathbf{Maj}}$ and $UC(R) = PO(R')$. Having assumed f to be majoritarian, $f(R') = f(R)$ and hence $a \in f(R')$. It also follows that $a \notin PO(R')$ and we may conclude that f is not Pareto optimal. $\square$

4 Potential extensions

We consider two natural extensions of Theorem 1 and show that neither of them holds.

4.1 Constant number of agents

The proof of Theorem 1 crucially depends on the assumption of a variable electorate since R' usually has a different number of agents than R. The fact that the same result cannot be achieved with a constant number of agents is exhibited by the following minimal, computer-generated example.[5] The majority relation $R_{\mathbf{Maj}}$ depicted in Fig. 2 contains a unique covering edge $a\ C_{R_{\mathbf{Maj}}}\ b$ and can be realized by 3 agents.[6] If we, however, require b to be Pareto dominated, i.e., $x\ \ P_{\mathbf{Par}}\ \ b$ for *at least one* of the alternatives $x\ \ P_{\mathbf{Maj}}\ \ b$, then the minimal number of agents required to realize $R_{\mathbf{Maj}}$ (together with the additional requirement of $x\ \ P_{\mathbf{Par}}\ \ b$) rises to 5, which has been verified on a computer.[7]

This example also shows that Corollary 1 does not in general hold for all constant electorates. To see this, let the number of agents be fixed at three and define an SCF f such that for all preference profiles R',

$$f(R') = \begin{cases} A & \text{if } R' \text{ is consistent with } R_{\mathbf{Maj}}, \\ UC(R') & \text{otherwise,} \end{cases}$$

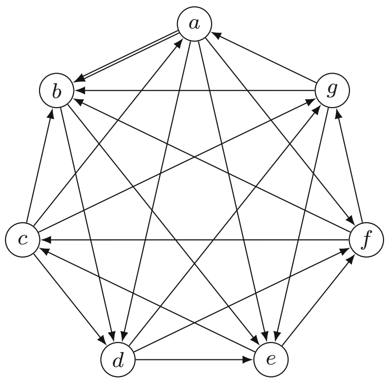
where $R_{\mathbf{Maj}}$ is the majority relation in Fig. 2. Thus, f is a coarsening of UC but, due to the electorate being fixed at three agents, f is still Pareto optimal.

[4] Corollary 1 also entails an analogous weaker result for the special case of tournaments (i.e., antisymmetric majority relations $R_{\mathbf{Maj}}$), which was used as a Lemma by Brandt and Geist (2014).

[5] The whole example was obtained from and proved minimal by an automated computer search based on the method developed by Brandt et al. (2014).

[6] In fact, any tournament of size 7 can be realized by 3 agents.

[7] This even holds when individual preferences are allowed to be *weak* orders.



1	1	1
e	c	b
f	a	d
c	d	g
g	f	a
a	g	e
b	b	f
d	e	c

(a) majority relation $R_{\mathbf{Maj}}$ **(b)** a consistent preference profile R

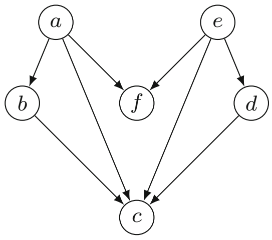
Fig. 2 A minimal example of a majority relation $R_{\mathbf{Maj}}$ for which the transition from $b \notin UC(R)$ to $b \notin PO(R)$ increases the minimal number of agents required to realize this relation

4.2 Equivalence of the covering relation and the Pareto relation

With Theorem 1 in mind, it is a natural question whether, given a preference profile R, one can even obtain a consistent preference profile R' (i.e., $R_{\mathbf{Maj}} = R'_{\mathbf{Maj}}$) in which the Pareto *relation* coincides with the covering relation (i.e., $P'_{\mathbf{Par}} = C_{R_{\mathbf{Maj}}}$). Unfortunately, the answer to this question is negative as the following counterexample shows.

Consider the preference profile R and the corresponding majority relation $R_{\mathbf{Maj}}$ in Fig. 3 and note that all strict majority edges are also covering edges, i.e., $P_{\mathbf{Maj}} = C_{R_{\mathbf{Maj}}}$. Therefore, constructing a consistent preference profile R' with $P'_{\mathbf{Par}} = C_{R_{\mathbf{Maj}}}$, in this particular example, means to construct a consistent preference profile R' such that $P'_{\mathbf{Par}} = C_{R_{\mathbf{Maj}}} = P_{\mathbf{Maj}}$. We now show—using computer-aided solving techniques—that such a profile does not exist.

For $P'_{\mathbf{Par}} = P_{\mathbf{Maj}}$ to hold, R' may only contain agents with individual preferences R'_i that respect all given strict majority edges as Pareto edges, i.e., for which $P'_i \supseteq P_{\mathbf{Maj}}$. This will stand in conflict with being able to maintain the majority ties in $R'_{\mathbf{Maj}} = R_{\mathbf{Maj}}$. We used an ASP program (Answer Set Programming, a declarative problem-solving paradigm; see, e.g., Gebser et al. (2011, 2012); in our case the packaged grounder/solver CLASP) to compute all 22 candidates R''_i for complete, antisymmetric, and transitive preference relations such that $P''_i \supseteq P_{\mathbf{Maj}}$. (One such ordering is e, d, a, f, b, c.) We would now have to construct a preference profile using only these 22 candidates for individual preference relations as building blocks. A small IP (Integer Program, see, e.g., Schrijver (1986); the performance of the open-source solver LPSOLVE satisfied our requirements) suffices to show that this is not possible: it contains 22 integer-valued variables $x_{R''_i}$, which denote, for each of the 22 preference relations, how many agents with this particular preference relation are contained in R'. Furthermore, it contains one constraint per indifference edge in the majority relation (of which there are 7, namely (a, d), (a, e), (b, d), (b, e), (b, f), (c, f), and (d, f)). Each of these constraints postulates for one indifference edge (y, z) that $0 = \sum_{R''_i} s((y, z), R''_i) \cdot x_{R''_i}$, where $s((y, z), R''_i) = 1$ if $(y, z) \in R''_i$, and



1	1	1	1
a	a	e	e
b	f	d	d
e	b	a	a
c	e	f	b
f	d	b	c
d	c	c	f

(a) majority relation $R_{\mathbf{Maj}}$, whose strict part $P_{\mathbf{Maj}}$ coincides with the covering relation $C_{R_{\mathbf{Maj}}}$

(b) a consistent preference profile R

Fig. 3 An example of a majority relation $R_{\mathbf{Maj}}$ whose covering relation $C_{R_{\mathbf{Maj}}}$ cannot be obtained as the Pareto relation of a consistent preference profile

$s((y, z), R_i'') = -1$ otherwise. With the additional constraints that $x_{R_i''} \geq 0$ for all R_i'' and $\sum_{R_i''} x_{R_i''} \geq 1$ the IP solver returns that there are no feasible solutions to this problem, which completes the proof.[8]

References

Bordes G (1983) On the possibility of reasonable consistent majoritarian choice: Some positive results. J Econ Theory 31(1):122–132

Brandt F, Fischer F (2008) Computing the minimal covering set. Math Soc Sci 56(2):254–268

Brandt F, Geist C (2014) Finding strategyproof social choice functions via SAT solving. In: Proceedings of the 13th international conference on autonomous agents and multi-agent systems (AAMAS), pp 1193–1200. IFAAMAS

Brandt F, Geist C, Seedig GH (2014) Identifying k-majority digraphs via SAT solving. In: Proceedings of the 1st AAMAS workshop on exploring beyond the worst case in computational social choice (EXPLORE)

Duggan J (2013) Uncovered sets. Soc Choice Welf 41(3):489–535

Dushnik B, Miller EW (1941) Partially ordered sets. Am J Math 63(3):600–610

Dutta B, Laslier J-F (1999) Comparison functions and choice correspondences. Soc Choice Welf 16(4):513–532

Gaertner W (2009) A primer in social choice theory: revised edition. In: LSE perspectives in economic analysis. Oxford University Press, Oxford

Gebser M, Kaminski R, Kaufmann B, Ostrowski M, Schaub T, Schneider M (2011) Potassco: the Potsdam answer set solving collection. AI Commun 24(2):107–124

Gebser M, Kaminski R, Kaufmann B, Schaub T (2012) Answer set solving in practice. Synth Lect Artif Intell Mach Learn 6(3):1–238

McGarvey DC (1953) A theorem on the construction of voting paradoxes. Econometrica 21(4):608–610

McKelvey RD (1986) Covering, dominance, and institution-free properties of social choice. Am J Polit Sci 30(2):283–314

Moulin H (1986) Choosing from a tournament. Soc Choice Welf 3(4):271–291

8 The same counterexample also applies to the case of weak individual orders (i.e., without the assumption of antisymmetry of R_i). In this case there are 256 instead of 22 candidates for individual preferences relations, and additional constraints for the Pareto edges are required, which makes the resulting IP significantly larger, but still solvable within less than one second.

Peris JE, Subiza B (1999) Condorcet choice correspondences for weak tournaments. Soc Choice Welf 16(2):217–231

Schrijver A (1986) Theory of linear and integer programming. Wiley, New York

IDENTIFYING k-MAJORITY DIGRAPHS VIA SAT SOLVING [6]

Peer-reviewed Conference Paper

Authors: F. Brandt, C. Geist, and H. G. Seedig

In: Proceedings of the 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE), 2014

Abstract: Many voting rules—including single-valued, set-valued, and probabilistic rules—*only* take into account the majority digraph. The contribution of this paper is twofold. First, we provide a surprisingly efficient implementation for computing the minimal number of voters that is required to induce a given digraph. This implementation relies on an encoding of the problem as a Boolean satisfiability (SAT) problem which is then solved by a SAT solver. Secondly, we experimentally evaluate how many voters are required to induce the majority digraphs of real-world and generated preference profiles. Our results are based on datasets from the PrefLib library and preferences generated using stochastic models such as impartial culture, impartial anonymous culture, Mallows mixtures, and spatial models. It turns out that all tournaments checked in these experiments can be induced by at most five voters whereas all other digraphs can be induced by at most eight voters. We also confirm a conjecture by Shepardson and Tovey by verifying that all tournaments with less than eight vertices can be induced by three voters.

Contribution of thesis author: SAT-based encoding and algorithms (including optimization via tournament components); ASP-based extension; joint project and paper management

Identifying k-Majority Digraphs via SAT Solving

Felix Brandt
Institut für Informatik
TU München
brandtf@in.tum.de

Christian Geist
Institut für Informatik
TU München
geist@in.tum.de

Hans Georg Seedig
Institut für Informatik
TU München
seedigh@in.tum.de

ABSTRACT

Many voting rules—including single-valued, set-valued, and probabilistic rules—*only* take into account the majority digraph. The contribution of this paper is twofold. First, we provide a surprisingly efficient implementation for computing the minimal number of voters that is required to induce a given digraph. This implementation relies on an encoding of the problem as a Boolean satisfiability (SAT) problem which is then solved by a SAT solver. Secondly, we experimentally evaluate how many voters are required to induce the majority digraphs of real-world and generated preference profiles. Our results are based on datasets from the PrefLib library and preferences generated using stochastic models such as impartial culture, impartial anonymous culture, Mallows mixtures, and spatial models. It turns out that all tournaments checked in these experiments can be induced by at most five voters whereas all other digraphs can be induced by at most eight voters. We also confirm a conjecture by Shepardson and Tovey by verifying that all tournaments with less than eight vertices can be induced by three voters.

1. INTRODUCTION

Perhaps one of the most natural ways to aggregate binary preferences from individual voters to a group of voters is *simple majority rule*, which prescribes that one alternative is socially preferred to another whenever a majority of voters prefers the former to the latter. Majority rule has an intuitive appeal to democratic principles, is easy to understand and—most importantly—satisfies some attractive formal properties [25]. Moreover, almost all common voting rules coincide with majority rule in the two-alternative case. It would therefore seem that the existence of a majority of individuals preferring alternative x to alternative y signifies something fundamental and generic about the group's preferences over x and y. Indeed, many voting rules—including single-valued, set-valued, and probabilistic rules—*only* take into account the majority digraph.

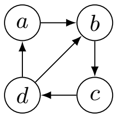
The central role of majority rule establishes an interesting connection between voting theory and graph theory. The earliest (and most fundamental) result in this context is *McGarvey's theorem*, which states that, given sufficiently many voters with linear preferences, *every* digraph may be induced by the majority rule [26]. In this paper, we will be concerned with the *minimal* number of voters $v(G)$ required to induce a given digraph G.

McGarvey's original construction requires two voters for each edge of the digraph, thus showing that $v(G) \leq 2\binom{n}{2}$ where n is the number of vertices of G. Consequently, this implied that the minimal number of voters $v(n)$ required to induce *any* digraph on n vertices is in $O(n^2)$. This bound was subsequently improved by Stearns [32], who showed that $v(n) = \Omega(n/\log n)$. Erdős and Moser [11] non-constructively provided a matching upper bound by proving that $v(n) = \Theta(n/\log n)$. More recently, Fiol [13] showed that $v(G) \leq n - \log n + 1$.

A digraph is a *k-majority digraph* if it can be induced by k voters. Interestingly, surprisingly little is known about the structure of k-majority digraphs. Dushnik and Miller [10] gave a complete characterization of 2-majority digraphs and Yannakakis [34] showed that the characterizing properties can be verified in polynomial time. Brandt et al. [8] provided a similar characterization for 3-majority digraphs. However, the computational complexity of checking whether a given digraph is a 3-majority digraph remains open. For the special case of *tournaments*, i.e., asymmetric and complete digraphs, Alon et al. [1] showed that the domination number of k-majority tournaments is bounded whereas Milans et al. [27] showed that every k-majority tournament contains a transitive subtournament whose size is at least polynomial in n.

The contribution of this paper is twofold. First, we provide a practical implementation for computing $v(G)$ for a given digraph G by encoding the problem as a Boolean satisfiability (SAT) problem which is then solved by a SAT solver. This technique turns out to be surprisingly efficient and easily outperforms an implementation for 3-majority digraphs based on the graph-theoretic characterization by Brandt et al. [8]. Secondly, we experimentally evaluate how many voters are required to induce the majority relations of real-world and generated preference profiles. Our results are based on datasets from the PrefLib library and preferences generated using stochastic models such as impartial culture, impartial anonymous culture, Mallows-ϕ, and spatial models. It turns out that all tournaments checked in these experiments are 5-majority tournaments whereas all other checked digraphs are 8-majority digraphs. Among other things, this shows that perhaps $v(G)$ itself may be used as a parameter to govern the generation of realistic preference profiles. We also confirm a conjecture by Shepardson and Tovey [31] by verifying that all tournaments with less than eight vertices are 3-majority digraphs.

Appears at: 1st Workshop on Exploring Beyond the Worst Case in Computational Social Choice. Held as part of the 13th International Conference on Autonomous Agents and Multiagent Systems. May 6th, 2014. Paris, France.



(a) digraph G

2	1	1	1	1
c	a	d	a	b
d	b	a	d	c
a	c	b	b	d
b	d	c	c	a

(b) preference profile R

Figure 1: A smallest 6-majority digraph with a minimal inducing preference profile.

2. PRELIMINARIES

Let A be a set of n alternatives and $K = \{1, \dots, k\}$ a set of voters. The preferences of voter $i \in K$ are represented by a linear (i.e., reflexive, complete, transitive, and antisymmetric) *preference relation* $R_i \subseteq A \times A$. The interpretation of $(a, b) \in R_i$, usually denoted by $a\ R_i\ b$, is that voter i values alternative a at least as much as alternative b. A *preference profile* $R = (R_1, \dots, R_k)$ is a k-tuple containing a preference relation R_i for each agent $i \in K$. For a preference profile R and two alternatives $a, b \in A$, the *majority margin* $g_R(a, b)$ is defined as the difference between the number of voters who prefer a to b and the number of voters who prefer b to a, i.e.,

$$g_R(a, b) = |\{i \in K \mid a\ R_i\ b\}| - |\{i \in K \mid b\ R_i\ a\}|.$$

Thus, $g_R(b, a) = -g_R(a, b)$ for all $a, b \in A$.

The *majority relation* $\succ_R$ of a given preference profile is defined as

$$a \succ_R b \quad \text{iff} \quad g_R(a, b) > 0.$$

Every majority relation $\succ_R$ is fully represented by a digraph G and we say that R *induces* G. If R has k voters, we say that G is *k-inducible*, or, equivalently, that G is a *k-majority digraph.*

If a digraph is complete, which is always the case if the number of voters is odd, we speak of a tournament $T = (A, \succ)$.

For any digraph G, by $v(G)$ we denote the minimal number of voters k such that G is a k-majority digraph. Occasionally, we will call this number the *voter complexity* of G.[1]

Example 1. *Consider the digraph G depicted on the left of Figure 1. We found that G is not 4-inducible. It cannot be 5-inducible either, because it is not a tournament as there is no strict relation between a and c. The profile R on the right of Figure 1, however, induces G and therefore G is a 6-majority digraph (or, equivalently, $v(G) = 6$). It turns out that G is a smallest digraph (in terms of the number of nodes) with voter complexity larger than 5.*

In this work, we address the computational problem of computing the voter complexity. To this end, we define the problem of checking whether for a given digraph G there exists a preference profile with k voters that induces G, i.e., whether G is a k-majority digraph.

[1]This complexity measure of digraphs can also be interpreted as a complexity measure for preference profiles. The voter complexity of a given preference profile is then simply defined as the voter complexity of the induced majority graph.

Name: Check-k-Majority
Instance: A digraph G and a positive integer k.
Question: Is G a k-majority digraph?

Note that the following two simple observations reduce the candidates for $v(G)$ to odd and even numbers, respectively, depending on whether G is a complete digraph or not.

Observation 1. *For all tournaments T, the voter complexity $v(T)$ is odd.*

Proof. Assume $v(T) = k$ was even. Then there exists a preference profile R with k voters that induces T. Since k is even, the majority margin must be even for every pair of alternatives and can furthermore never be 0 as T is a tournament. Therefore, removing any single voter from R gives a profile R' with just $k - 1$ voters that still induces T, a contradiction. $\square$

Observation 2. *For all incomplete digraphs G, the voter complexity $v(G)$ is even. It even holds that G is no k-majority digraph for k odd.*

Proof. This follows directly from the fact that for all preference profiles R with an odd number of voters k, the majority relation $\succ_R$ is complete and anti-symmetric (as no majority ties can occur). $\square$

3. METHODOLOGY

The number of objects potentially involved in the Check-k-Majority problem are given in Table 1. It is immediately clear that a naïve algorithm will not solve the problem in a satisfactory manner. This section describes our algorithmic efforts to solve this problem for reasonably large instances.

3.1 Translation to propositional logic (SAT)

In order to answer Check-k-Majority, we follow a similar approach as Tang and Lin [33], Geist and Endriss [16], and Brandt and Geist [4]: we translate the problem to propositional logic (on a computer) and use state-of-the-art SAT solvers to find a solution. At a glance, the overall solving steps are shown in Algorithm 1.

Generally speaking, the problem at hand can be understood as the problem of finding a preference profile that satisfies certain conditions—here: inducing a given digraph. Thus, a satisfying instance of the propositional formula to be designed should represent a preference profile. To capture this, a surprisingly simple formalization involving just one type of variable suffices: in our encoding the boolean variable $r_{i,a,b}$ represents $a\ R_i\ b$, i.e., voter i ranking alternative a at least as high as alternative b. As it turns out, this one variable type also suffices for the additional condition of inducing the given digraph.

In more detail, the following three conditions/axioms need to be formalized:

1. All k voters have linear orders over the n alternatives as their preferences (short: linear preferences)

2. For each majority edge $x \succ y$ in the digraph, a majority of voters needs to prefer x over y (short: majority implications)

Preference profiles	$n=4$	$n=5$	$n=10$	$n=25$	$n=50$
$k=1$	24	120	$\sim 3.6 \cdot 10^{6}$	$\sim 1.6 \cdot 10^{25}$	$\sim 3.0 \cdot 10^{64}$
$k=3$	13,824	$\sim 1.7 \cdot 10^{6}$	$\sim 4.8 \cdot 10^{19}$	$\sim 3.7 \cdot 10^{75}$	$\sim 2.8 \cdot 10^{193}$
$k=5$	$\sim 8.0 \cdot 10^{6}$	$\sim 2.5 \cdot 10^{10}$	$\sim 6.3 \cdot 10^{32}$	$\sim 9.0 \cdot 10^{125}$	$\sim 2.6 \cdot 10^{322}$
Tournaments (unlabeled)	4	12	$\sim 9.7 \cdot 10^{6}$	$\sim 1.3 \cdot 10^{65}$	$\sim 1.9 \cdot 10^{305}$

Table 1: Number of objects involved in the CHECK-k-MAJORITY problem for one, three, and five voters.

```
Input: digraph (A, ≻), positive integer k
Output: whether (A, ≻) is a k-majority digraph
/* Encoding of problem in CNF                    */
File cnfFile;
foreach voter i do
    cnfFile += Encoder.reflexivePreferences(i);
    cnfFile += Encoder.completePreferences(i);
    cnfFile += Encoder.transitivePreferences(i);
    cnfFile += Encoder.antisymmetricPreferences(i);
cnfFile += Encoder.majorityImplications((A, ≻));
if ≻ is not complete then
    cnfFile +=
    Encoder.indifferenceImplications((A, ≻));
/* SAT solving                                   */
satisfiable = SATsolver.solve(cnfFile);
if instance is satisfiable then
    return true;
else
    return false
```

Algorithm 1: SAT-CHECK-k-MAJORITY

3. For each missing edge ($x \not\succ y$ and $y \not\succ x$) in the digraph, *exactly* half the voters need to prefer x over y (short: indifference implications)[2]

For the first axiom, we encode reflexivity, completeness, transitivity, and anti-symmetry of the relation R_i for all voters i. The complete translation to CNF (conjunctive normal form, the established standard input format for SAT solvers) is given exemplarily for the case of transitivity; the other axioms are converted analogously.

In formal terms transitivity can be written as

$$\begin{aligned}
& (\forall i)(\forall x,y,z)\,(x\ R_i\ y \wedge y\ R_i\ z \rightarrow x\ R_i\ z) \\
\equiv\ & (\forall i)(\forall x,y,z)\,(r_{i,x,y} \wedge r_{i,y,z} \rightarrow r_{i,x,z}) \\
\equiv\ & \bigwedge_i \bigwedge_{x,y,z} (\neg\,(r_{i,x,y} \wedge r_{i,y,z}) \vee r_{i,x,z}) \\
\equiv\ & \bigwedge_i \bigwedge_{x,y,z} (\neg r_{i,x,y} \vee \neg r_{i,y,z} \vee r_{i,x,z})\,,
\end{aligned}$$

which then translates to the pseudo code in Algorithm 2 for generating the CNF file. The key in the translation of the inherently higher order axioms to propositional logic is (as pointed out by Geist and Endriss [16] already) that because of finite domains, all quantifiers can be replaced by finite conjunctions or disjunctions, respectively.

In all algorithms, a subroutine $r(i,x,y)$ takes care of the compact enumeration of variables.[3]

```
foreach Voter i do
    foreach Alternative x do
        foreach Alternative y do
            foreach Alternative z do
                variable_not(r(i, x, y));
                variable_not(r(i, y, z));
                variable(r(i, x, z));
                newClause;
```

Algorithm 2: Encoding of transitivity of individual preferences

The axioms of majority and indifference implications can be formalized in a similar fashion. We describe the translation for the majority implications here; the procedure for the indifference implications (needed for incomplete digraphs) is analogous again. In the following, we denote the smallest number of voters required for a positive majority margin by $m(k) := \lfloor k \cdot \frac{1}{2} \rfloor + 1$. Note that then, because of anti-symmetry of the individual preferences, for $x \succ y$ it suffices that there exists a set of $m(k)$ many voters who prefer x to y. In formal terms:

$$\begin{aligned}
& (\forall x,y)\,(x \succ y \rightarrow |\{i \mid x\ R_i\ y\}| > |\{i \mid y\ R_i\ x\}|) \\
\equiv\ & (\forall x,y)\,(x \succ y \rightarrow |\{i \mid x\ R_i\ y\}| \geq s(n)) \\
\equiv\ & (\forall x,y)\,(x \succ y \rightarrow \\
& \qquad (\exists M \subseteq K)|M| = m(k) \wedge (\forall i \in M) x\ R_i\ y) \\
\equiv\ & \bigwedge_{x \succ y} \bigvee_{|M|=m(k)} \bigwedge_{i \in M} r_{i,x,y}.
\end{aligned}$$

In order to avoid an exponential blow-up when converting this formula to CNF, variable replacement (a standard procedure also known as Tseitin transformation) is applied. In our case, we replaced $\bigwedge_{i \in M} r_{i,x,y}$ by new variables of the form $h_{M,x,y}$ and introduced the following defining clauses:

$$\begin{aligned}
& \bigwedge_M \bigwedge_{x,y} \left(h_{M,x,y} \rightarrow \bigwedge_{i \in M} r_{i,x,y} \right) \\
\equiv\ & \bigwedge_M \bigwedge_{x,y} \left(\neg h_{M,x,y} \vee \bigwedge_{i \in M} r_{i,x,y} \right) \\
\equiv\ & \bigwedge_M \bigwedge_{x,y} \bigwedge_{i \in M} (\neg h_{M,x,y} \vee r_{i,x,y})\,.
\end{aligned}$$

In this case, the helper variables even have an intuitive meaning as $h_{M,x,y}$ enforces that all the voters $i \in M$ prefer alternative x over alternative y.

[2] Note that this axiom is only required for incomplete digraphs.

[3] The DIMACS CNF format only allows for integer names of variables. But since we know in advance how many voters and alternatives there are, we can simply use a standard enumeration method for tuples of objects.

Note that the conditions like $|M| = m(k)$ can easily be fulfilled during generation of the corresponding CNF formula on a computer. For enumerating all subsets of voters of a given size we, for instance, used Gosper's Hack [18]. The corresponding pseudo code for majority implications can be found in Algorithm 3.

```
foreach Pair of alternatives x ≻ y do
    foreach M ⊆ K, |M| = m(k) do
        variable(h(M, x, y));
    newClause;
/* start of helper variable definition */
foreach Pair of alternatives x ≻ y do
    foreach M ⊆ K, |M| = m(k) do
        foreach i ∈ M do
            variable_not(h(M, x, y));
            variable(r(i, x, y));
            newClause;
```

Algorithm 3: Encoding of majority implications

This encoding leads to a total of $k \cdot n^2 + \binom{k}{m(k)} \cdot n^2 = n^2 \cdot \left(k + \binom{k}{m(k)}\right)$ variables for the case of tournaments and $n^2 \cdot \left(k + \binom{k}{m(k)} + \binom{k}{k/2}\right)$ variables for incomplete digraphs.The number of clauses is equal to $k \cdot (n^3 + n^2) + \frac{n^2-n}{2} \cdot \left(1 + \binom{k}{m(k)} \cdot m(k)\right)$ and at most $k \cdot (n^3 + n^2) + (n^2 - n) \cdot \left(1 + \binom{k}{k/2} \cdot \frac{k}{2}\right)$ for tournaments and incomplete digraphs, respectively.

With all axioms formalized in propositional logic, we are now ready to analyze arbitrary digraphs G for their voter complexity $v(G)$. Before we do so, however, we describe an optimization technique for tournament graphs, which, for certain instances, speeds up the computation significantly.

3.2 Optimized computation for tournaments via components

An important structural property in the context of tournaments is whether a tournament admits a non-trivial decomposition. Brandt et al. [7] show that this decomposition allows for a recursive computation of certain concepts, which is particularly helpful if the original computation is costly for large instances.[4] We are now going to prove that a similar optimization can be carried out for the computation of the voter complexity $v(T)$ of a given tournament T. In particular, we show that the voter complexity of a tournament is equal to the maximum of the voter complexities of its components and the corresponding summary.

In formal terms, a non-empty subset B of A is a *component* of a tournament $T = (A, \succ)$ if for all $a \in A \setminus B$ either $B \succ a$ or $a \succ B$, where $B \succ a$ stands for $(\forall b \in B) b \succ a$. A *decomposition* of T is a set of pairwise disjoint components $\{B_1, \ldots, B_p\}$ of T such that $A = \bigcup_{j=1}^{p} B_j$. The decomposition is *proper* if $p > 1$ and not all B_j are singletons. Every tournament admits a decomposition that is minimal in a well-defined sense [20]. Given a particular decomposition $\tilde{B} = \{B_1, \ldots, B_p\}$, the *summary* of T with respect to $\tilde{B}$ is defined as the tournament $T_B = (\{1, \ldots, p\}, \tilde{\succ})$ on the individual components rather than the alternatives, i.e.,

$$q \mathrel{\tilde{\succ}} r \quad \text{if and only if} \quad B_q \succ B_r.$$

Each component B_j (including A) naturally induces a subtournament T_{B_j} which is the summary of $T|_{B_j}$ with respect to its minimal decomposition.

The following lemma then enables the recursive computation of $v(T)$ along the component structure of T:

Lemma 1. *Let T be a tournament and $\tilde{B} = \{B_1, \ldots, B_p\}$ a decomposition of T. Then*

$$v(T) = \max_j \{v(T_{B_j}), v(T_B)\}.$$

Proof. Let R be a minimal profile inducing T. Then, $R|_{B_j}$ induces T_{B_j} for every B_j establishing $v(T) \geq v(T_{B_j})$. That $v(T) \geq v(T_B)$ holds is also easy to see by considering a variant of R in which from each component all but one node are arbitrarily chosen and removed. The remaining profile then induces T_B. For the other direction, let $v'(T) = \max_j\{v(T_{B_j}), v(T_B)\}$. We know, by Observation 1, that $v(T')$ and every $v(T_{B_j})$ is odd, as these are all tournaments. Each T_{B_j} (and T_B) has a minimal profile R^j (and R, respectively). We can add pairs of voters with opposing preferences to each profile without changing its majority relation. This way, we get profiles R'^j (and R') that still induce T_{B_j} (or T_B) but now all have the same number of voters $v'(T)$. Now, create a new profile $\hat{R}$ from R' in which R^j_i replaced alternative j as a segment in R'_i for each voter i and every alternative j as in [19]. It is easy to check that $\hat{R}$ has $v'(T)$ voters and still induces T, i.e., $v(T) \geq v'(T) = v(T_{B_j})$. $\square$

We have implemented this optimization and found that many real-world majority digraphs exhibit proper decompositions, speeding up the computation of SAT-Check-k-Majority.

3.3 Data sources and method of analysis

In the preference library PrefLib [23], scholars have contributed data sets from real world scenarios ranging from preferences over movies or sushi via Formula 1 championship results to real election data. Accordingly, the number of voters whose preferences originally induced these data sets vary heavily between 4 and 44000. At the time of writing, PrefLib contained 354 tournaments induced from pairwise majority comparisons as well as 185 incomplete majority digraphs.

Additionally, we consider *stochastic* models to generate tournaments of a given size n. Many different models for linear preferences (or orderings) have been considered in the literature. We refer the interested reader to [9, 22, 24, 6]. In this work, we decided to examine tournaments generated with five different stochastic models.

In the *uniform random tournament model*, the same probability is assigned to each *labeled* tournament of size n, i.e.,

$$\Pr(T) = \frac{1}{2^{\binom{n}{2}}} \text{ for each } T \text{ with } |T| = n.$$

In all of the remaining models, we sample preference profiles and work with the tournament induced by the majority relation. In accordance with [6], we generated profiles with 51 voters.

[4] As Brandt et al. [7] point out, the decomposition of a tournament can be computed in linear time.

The *impartial culture model* (IC) is the most widely-studied model for individual preferences in social choice. It assumes that every possible preference ordering has the same probability of $\frac{1}{n!}$. If we add anonymity by having indistinguishable voters, the set of profiles is partitioned into equivalence classes. In the *impartial anonymous culture* (IAC), each of these equivalence classes is chosen with equal probability.

In *Mallows-ϕ model* [21], the distance to a reference ranking is measured by means of the Kendall-tau distance which counts the number of pairwise disagreements. Let R_0 be the reference ranking. Then, the Kendall-tau distance of a preference ranking R to R_0 is

$$\tau(R, R_0) = \binom{n}{2} - |R \cap R_0| \,.$$

According to the model, this induces the probability of a voter having R as his preferences to be

$$\Pr(R) = \frac{\phi^{\tau(R,R_0)}}{C}$$

where C is a normalization constant and $\phi \in (0, 1]$ is a dispersion parameter. Small values for ϕ put most of the probability on rankings very close to R_0 whereas for $\phi = 1$ the model coincides with IC.

A very different kind of model is the *spatial model*. Here, alternatives and voters are uniformly at random placed in a multi-dimensional space and the voters' preferences are determined by the (Euclidian) distanced to the alternatives. The spatial model has played an important role in political and social choice theory where the dimensions are interpreted as different aspects or properties of the alternatives (see, e.g., [28, 2]).

4. RESULTS

All experiments were run on a Intel Core i5, 2.66GHz (quad-core) machine with 12 GB RAM using the SAT solver PLINGELING [3].

4.1 Exhaustive analysis

We generated all tournaments with up to 10 alternatives and found that all of these are 5-inducible. In fact, all tournaments of size up to seven are even 3-inducible, confirming a conjecture by Shepardson and Tovey [31]. They also showed that there exist tournaments of size 8 that are not 3-inducible. We find that the exact number of such tournaments is 96 (out of 6880).

Brandt and Seedig [5] presented a highly structured tournament on 24 alternatives that serves as the current minimal counterexample to a now disproved conjecture by Schwartz [30] in social choice theory. We found it to be a 5-majority tournament, implying that the negative theoretical consequences of the counterexample already hold for scenarios with only 5 voters (and at least 24 alternatives).

4.2 Empirical analysis

Among the tournaments in PrefLib, 58 are 3-inducible. Out of the two largest tournaments in the data set with 240 and 242 alternatives, respectively, the first is a 5-majority tournament while on the second the SAT solver did not terminate within one day. The remaining tournaments are transitive and thus 1-inducible. Therefore, all checkable tournaments in PrefLib are inducible by only 5 voters.

For the non-complete majority digraphs in PrefLib, we found that the indifference constraints which are imposed on missing edges change the picture. Not only does it negatively affect the running time of SAT-Check-k-Majority in comparison to tournaments which made us restrict our attention to instances with at most 40 alternatives, but it also seems to result in higher voter complexities of up to 8 among the 85 feasible instances. However, given that the number of voters in the profiles that originally induced these majority digraphs are often in the hundreds or thousands, we still consider these low voter complexities.

n	uniform	IC	IAC	Mallows-ϕ ($\phi = 0.95$)	spatial (dim = 2)
3	1.40	1.13	1.13	1.13	1.00
5	3.00	1.67	2.13	1.33	1.13
7	3.00	2.67	2.67	2.47	1.33
9	3.13	3.00	3.00	2.67	1.60
11	3.93	3.07	3.00	2.87	2.33
13	4.80	3.07	3.20	2.93	2.53
15	5.00	3.27	3.40	3.00	2.67
17	5.00	3.40	3.80	2.93	2.80
19	5.00	4.27	4.20	3.00	2.80
21	5.00	4.47	4.33	3.00	2.87

Table 2: Average voter complexity in tournaments generated by stochastic (preference) models. The given values are averaged over 30 samples each.

4.3 Stochastic analysis

For up to 21 alternatives, we sampled preference profiles (each consisting of 51 voters[5]) from the aforementioned stochastic models and examined the corresponding majority graphs for their voter complexity using SAT-Check-k-Majority. The average complexities over 30 instances of each size are shown in Table 2. We see that the unbiased models (IC, IAC, uniform) tend to induce digraphs with higher voter complexity. We encountered no tournament that was not a 5-majority tournament.[6]

4.4 Runtime analysis

A characterization by Brandt et al. [8] of 3-majority digraphs allows for a straightforward algorithm, which is expected to have a much better running time than any naïve implementation enumerating all preference profiles (also compare Table 1). The characterization is given in Lemma 2 below, as is the corresponding algorithm 2-Partition-Check-3-Majority (Algorithm 4). Besides enumerating all 2-partitions of the majority relations, the only non-trivial part is to check whether a relation has a transitive reorientation. This can be done efficiently using an algorithm by Pnueli et al. [29].

We compared the running times of 2-Partition-Check-3-Majority with the ones of our implementation via SAT

[5]In another study [6], this size turned out to be sufficiently large to discriminate the different underlying stochastic models.

[6]Our efforts also included checking more than 8 million uniform random tournaments with 12 alternatives.

as described in Section 3.1 (see also Algorithm 1).[7]

Surprisingly, it turns out that—even though it is much more universal—SAT-Check-3-Majority offers significantly better running times. Preliminary data is displayed in Table 3. Note that in addition to being more efficient, SAT-Check-k-Majority is even able to return a preference profile with k voters that induces the given digraph (without the need for additional computations).

Further runtimes, which exhibit the practical power of our SAT approach (and its limits), can be obtained from Table 4.

Lemma 2 (Brandt et al.). *A digraph $(A, \succ)$ is a 3-majority digraph if and only if $\succ$ is complete and there are disjoint sets $\succ_1, \succ_2$ with $\succ = (\succ_1 \cup \succ_2)$ such that*

- *$(A, \succ_1)$ is a 2-majority digraph and*
- *$\succ_2$ is acyclic.*

Whether $(A, \succ_1)$ is a 2-majority digraph can efficiently be checked [34] via the following characterization by Dushnik and Miller [10]:

Lemma 3 (Dushnik and Miller). *A digraph $(A, \succ)$ is a 2-majority digraph if and only if*

- *$\succ$ is transitive and*
- *there exists a reorientation of $\big((A \times A) \setminus (\succ \cup \succ^{-1})\big)$ that is transitive and asymmetric.*

Input: digraph $(A, \succ)$
Output: whether $(A, \succ)$ is a 3-majority digraph
if $\succ$ *is complete* **then**
  foreach 2-*partition* $\{\succ_1, \succ_2\}$ *of* $\succ$ **do**
    if $\succ_1$ *is transitive and* $\succ_2$ *is acyclic and* $\succ_2$ *has a transitive reorientation* **then**
      return true;
    else
      return false;
else
  return false;

Algorithm 4: 2-Partition-Check-3-Majority

5. OUTLOOK AND FUTURE WORK

The following two insights of this work have been most surprising to us.

- First, our SAT-based implementation significantly outperforms the best direct algorithm known to us, while at the same time being much more flexible and powerful.[8]
- Second, the voter complexity of any majority digraph we could analyze does not exceed five for tournaments, and eight for incomplete digraphs, respectively.

[7]As a programming language Java was used in both cases.
[8]In the sense that it can also solve instances for $k \geq 3$.

Both of these points offer many directions for future work. Our implementation might be useful to find concrete tournaments that are not k-inducible, a problem that has occupied graph theorists. For example, the order of the smallest tournament that is not 5-inducible is currently unknown. Analytical results by Alon et al. [1], Graham and Spencer [17], and Fidler [12] can be used to narrow down the search for such tournaments. Preliminary results suggest that quadratic residue tournaments are good candidates for tournaments that can only be induced by a large number of voters. We intend to further pursue this direction in future work.

As other solving techniques are concerned, a natural choice for the problem at hand are techniques that can handle cardinality constraints natively (rather than encoding them in SAT/CNF as we did). ASP (answer set programming, see, e.g., Gebser et al. [15]) is an example of such a technique. We were able to obtain preliminary results using an ASP formulation of the problem (see Figure 2) and a corresponding solver (clasp with grounder Gringo [14]). While due to its richer problem description language (which also includes cardinality constraints) the formalization is much more compact than the corresponding SAT/CNF formulation, interestingly, performance appears to be similar or even slightly worse compared to current SAT solvers. Other solvers with cardinality constraints, however, might lead to different performance results.

Our approach can also be used to treat a range of related problems and questions. For instance, one could define natural variants of the notion of k-majority digraphs such as voters having weak (i.e., ties are allowed) or even incomplete preferences. Because of the high flexibility of our SAT formalization, one can easily apply the same method to analyze these related concepts and questions.[9] Even weighted majority graphs, i.e., graphs which carry the majority margin as weights on edges, can be analyzed regarding their voter complexity by slightly adapting our SAT or ASP encodings.

REFERENCES

[1] N. Alon, G. Brightwell, H. A. Kierstead, A. V. Kostochka, and P. Winkler. Dominating setse in k-majority tournaments. *Journal of Combinatorial Theory Series B*, 96:374–387, 2006.

[2] D. Austen-Smith and J. S. Banks. *Positive Political Theory I: Collective Preference.* University of Michigan Press, 2000.

[3] A. Biere. Lingeling, plingeling and treengeling entering the sat competition 2013. In A. Balint, A. Belov, M. Heule, and M. Järvisalo, editors, *Proceedings of the SAT Competition 2013*, pages 51–52, 2013.

[4] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. In *Proceedings of the 13th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. IFAAMAS, 2014. Forthcoming.

[5] F. Brandt and H. G. Seedig. A tournament of order 24 with two disjoint TEQ-retentive sets. Technical report, http://arxiv.org/abs/1302.5592, 2013.

[9]For the two suggested variants, deleting axioms from the formalization suffices.

algorithm	5	6	7	8	9	10	20	50	100
SAT	< 0.1s	< 0.1s	< 0.1s	< 0.1s	< 0.1s	< 0.1s	0.1s	1.5s	12.5s
2-Partition	< 0.1s	0.1s	2s	1200s	—	—	—	—	—

Table 3: Runtime comparison of the SAT implementation for $k = 3$ and 2-Partition-Check-3-Majority for complete digraphs (tournaments) of different sizes n with a cutoff time of one hour.

$n \backslash k$	3	4	5	6	7	8	9	10	11	12
3	.04	.04	.03	.04	.04	.04	.04	.05	.08	.10
4	.03	.04	.03	.04	.04	.04	.05	.07	.10	.18
5	.03	.04	.03	.04	.06	.05	.06	.09	.16	.35
6	.03	.04	.04	.04	.05	.06	.08	.12	.27	.63
7	.04	.04	.04	.05	.05	.07	.10	.17	.45	1.10
8	.04	.05	.05	.05	.07	.08	.13	.23	.69	1.80
9	.04	.05	.05	.64	.07	.10	.17	.33	1.06	2.83
10	.05	.05	.06	.67	.09	.12	.23	.46	1.56	4.25
11	.06	.06	.06	1.92	.10	.14	.30	.63	2.22	6.37
12	.06	.07	.07	3.35	.12	.19	.40	.85	3.18	8.48
13	.07	.07	.09	3.93	.15	.27	.52	1.16	4.44	12.30
14	.07	.09	.10	4.15	.18	.36	.64	1.51	5.99	16.84
15	.08	.10	.13	3.89	.21	.88	.79	2.22	7.67	—
16	.09	.11	.14	4.12	.25	4.55	.99	2.90	9.80	—
17	.10	.12	.19	4.41	.29	7.15	1.23	4.69	12.48	—
18	.11	.14	.23	4.76	.35	17.51	1.53	8.25	15.97	—
19	.12	.15	.35	4.97	.43	—	1.80	—	19.99	—
20	.13	.17	.54	5.04	.47	—	2.21	—	—	—
21	.14	.18	5.87	6.15	.63	—	2.71	—	—	—
22	.16	.20	11.07	5.43	.96	—	3.24	—	—	—
23	.17	.23	18.95	5.76	1.57	—	4.12	—	—	—
24	.20	.26	—	5.87	2.56	—	4.60	—	—	—
25	.22	.29	—	6.12	4.21	—	5.85	—	—	—

Table 4: Runtime in seconds of SAT-Check-k-Majority for different number of alternatives and different number of voters k when average runtimes did not exceed 20 seconds. For this table, averages were taken over 5 samples from the uniform random tournament model.

[6] F. Brandt and H. G. Seedig. On the discriminative power of tournament solutions. In *Proceedings of the 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2014. Forthcoming.

[7] F. Brandt, M. Brill, and H. G. Seedig. On the fixed-parameter tractability of composition-consistent tournament solutions. In T. Walsh, editor, *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 85–90. AAAI Press, 2011.

[8] F. Brandt, P. Harrenstein, K. Kardel, and H. G. Seedig. It only takes a few: On the hardness of voting with a constant number of agents. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 375–382. IFAAMAS, 2013.

[9] D. E. Critchlow, M. A. Fligner, and J. S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35:294–318, 1991.

[10] B. Dushnik and E. W. Miller. Partially ordered sets. *American Journal of Mathematics*, 63(3):600–610, 1941.

[11] P. Erdős and L. Moser. On the representation of directed graphs as unions of orderings. *Publ. Math. Inst. Hungar. Acad. Sci.*, 9:125–132, 1964.

[12] D. Fidler. A recurrence for bounds on dominating sets in k-majority tournaments. *The Electronic Journal of Combinatorics*, 18(1), 2011.

[13] M. A. Fiol. A note on the voting problem. *Stochastica*, XIII-1:155–158, 1992.

[14] M. Gebser, R. Kaminski, B. Kaufmann, M. Ostrowski, T. Schaub, and M. Schneider. Potassco: The Potsdam answer set solving collection. *AI Communications*, 24 (2):107–124, 2011.

[15] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub. Answer set solving in practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(3):1–238, 2012.

[16] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[17] R. L. Graham and J. H. Spencer. A constructive solution to a tournament problem. *Canadian Mathematical Bulletin*, 14(1):45–48, 1971.

```
%Alternatives
#const m=19.
alt(1..m).

%Number of voters
#const n=6.
voter(1..n).

#const simple_majority=4.

#const indifference_majority=3.

%Completeness and Antisymmetry
1{r(I,X,Y);r(I,Y,X)}1 :- voter(I), alt(X;Y), X!=Y.

%Reflexivity
r(I,X,X) :- voter(I), alt(X).

%Transitivity
r(I,X,Z) :- r(I,X,Y), r(I,Y,Z).

%Majority implications
simple_majority{r(I,X,Y):voter(I)} :- g(X,Y), X!=Y.

%Indifference implications
indifference_majority{r(I,X,Y):voter(I)}
indifference_majority :- i(X,Y), X!=Y.

g(1,2).
g(1,3).
g(1,4).
# (...) graph encoding mostly ommitted
g(19,14).
g(19,18).


#show r/3.
```

Figure 2: Problem description in ASP for $k = 6$ and a majority digraph with $n = 19$ nodes. Parts of the majority graph have been omitted to increase readability.

[18] D. E. Knuth. *Combinatorial Algorithms*, volume 4A, part 1 of *The Art of Computer Programming*. Addison-Wesley, 2011.

[19] G. Laffond, J. Lainé, and J.-F. Laslier. Composition-consistent tournament solutions and social choice functions. *Social Choice and Welfare*, 13:75–93, 1996.

[20] J.-F. Laslier. *Tournament Solutions and Majority Voting*. Springer-Verlag, 1997.

[21] C. L. Mallows. Non-null ranking models. *Biometrika*, 44(1/2):114–130, 1957.

[22] J. I. Marden. *Analyzing and Modeling Rank Data*. Number 64 in Monographs on Statistics and Applied Probability. Chapman & Hall, 1995.

[23] N. Mattei and T. Walsh. PrefLib: A library for preference data http://www.preflib.org. In P. Perny, M. Pirlot, and A. Tsoukiàs, editors, *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, volume 8176 of *Lecture Notes in Computer Science (LNCS)*, pages 259–270. Springer, 2013.

[24] N. Mattei, J. Forshee, and J. Goldsmith. An empirical study of voting rules and manipulation with large datasets. In *Proceedings of the 4th International Workshop on Computational Social Choice (COMSOC)*, 2012.

[25] K. May. A set of independent, necessary and sufficient conditions for simple majority decisions. *Econometrica*, 20:680–684, 1952.

[26] D. C. McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, 21(4):608–610, 1953.

[27] K. G. Milans, D. H. Schreiber, and D. B. West. Acyclic sets in k-majority tournaments. *The Electronic Journal of Combinatorics*, 18(1), 2011.

[28] P. Ordeshook. The spatial analysis of elections and committees: four decades of research. Mimeo, 1993.

[29] A. Pnueli, A. Lempel, and S. Even. Transitive orientation of graphs and identification of permutation graphs. *Canadian Journal of Mathematics*, 23:160–175, 1971.

[30] T. Schwartz. Cyclic tournaments and cooperative majority voting: A solution. *Social Choice and Welfare*, 7 (1):19–29, 1990.

[31] D. Shepardson and C. A. Tovey. Smallest tournament not realizable by $\frac{2}{3}$-majority voting. *Social Choice and Welfare*, 33(3):495–503, 2009.

[32] R. Stearns. The voting problem. *American Mathematical Monthly*, 66(9):761–763, 1959.

[33] P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

[34] M. Yannakakis. The complexity of the partial order dimension problem. *SIAM Journal on Algebraic and Discrete Methods*, 3(3):351–358, 1982.

FINDING PREFERENCE PROFILES OF CONDORCET DIMENSION k VIA SAT [7]

Technical Report

Author: C. Geist

In: http://arxiv.org/abs/1402.4303, 2014

Abstract: Condorcet winning sets are a set-valued generalization of the well-known concept of a Condorcet winner. As supersets of Condorcet winning sets are always Condorcet winning sets themselves, an interesting property of preference profiles is the size of the smallest Condorcet winning set they admit. This smallest size is called the Condorcet dimension of a preference profile. Since little is known about profiles that have a certain Condorcet dimension, we show in this paper how the problem of finding a preference profile that has a given Condorcet dimension can be encoded as a satisfiability problem and solved by a SAT solver. Initial results include a minimal example of a preference profile of Condorcet dimension 3, improving previously known examples both in terms of the number of agents as well as alternatives. Due to the high complexity of such problems it remains open whether a preference profile of Condorcet dimension 4 exists.

Finding Preference Profiles of Condorcet Dimension k via SAT

Christian Geist
Technische Universität München
Munich, Germany
geist@in.tum.de

ABSTRACT

Condorcet winning sets are a set-valued generalization of the well-known concept of a Condorcet winner. As supersets of Condorcet winning sets are always Condorcet winning sets themselves, an interesting property of preference profiles is the size of the smallest Condorcet winning set they admit. This smallest size is called the Condorcet dimension of a preference profile. Since little is known about profiles that have a certain Condorcet dimension, we show in this paper how the problem of finding a preference profile that has a given Condorcet dimension can be encoded as a satisfiability problem and solved by a SAT solver. Initial results include a minimal example of a preference profile of Condorcet dimension 3, improving previously known examples both in terms of the number of agents as well as alternatives. Due to the high complexity of such problems it remains open whether a preference profile of Condorcet dimension 4 exists.

1. INTRODUCTION

The contribution of this paper is twofold. Firstly, we provide a practical implementation for finding a preference profile for a given Condorcet dimension by encoding the problem as a boolean satisfiability (SAT) problem [2], which is then solved by a SAT solver. This technique has proven useful for a range of other problems in social choice theory (see, e.g., [9, 7, 4, 3]) and can easily be adapted. For instance, only little needs to be altered in order answer similar questions for dominating sets rather than Condorcet winning sets. Secondly, we give an answer to an open question by Elkind et al. [5] and provide a minimal example of a preference profile of Condorcet dimension 3, which we computed using our implementation. This profile involves 6 alternatives and agents only, improving the size of previous examples both in terms of agents and alternatives.[1] The formalization in SAT turns out to be efficient enough, not only to discover this particular profile of Condorcet dimension 3, but also to show its minimality.

[1]For instance, the example in Elkind et al. [5] required 15 alternatives and agents.

Appears in: *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), Lomuscio, Scerri, Bazzan, Huhns (eds.), May, 5–9, 2014, Paris, France.*

2. PRELIMINARIES

Let A be a set of m alternatives and $N = \{1, \dots, n\}$ a set of agents. The preferences of agent $i \in N$ are represented by a linear (i.e., reflexive, complete, transitive, and antisymmetric) *preference relation* $R_i \subseteq A \times A$. The interpretation of $(a, b) \in R_i$, usually denoted by $a\ R_i\ b$, is that agent i values alternative a at least as much as alternative b. A *preference profile* $R = (R_1, \dots, R_n)$ is an n-tuple containing a preference relation R_i for each agent $i \in N$.

Let R be a preference profile. As introduced by Elkind et al. [5], we now define the notion of a Condorcet winning set through an underlying covering relation between sets of alternatives and alternatives: A set of alternatives X θ-covers an alternative y (short: $X \succ_R^\theta y$) if

$$|\{i \in N \mid \exists x \in X \text{ such that } x\ R_i\ y\}| > \theta n.$$

A set of alternatives X is called a *Condorcet winning set* if for each alternative $y \notin X$ the set X $\frac{1}{2}$-covers y. The set of all Condorcet winning sets of R will be denoted by $\mathcal{C}(R)$. The *Condorcet dimension* $\dim_C(R)$ is defined as the size of the smallest Condorcet winning set the profile R admits, i.e.,

$$\dim_C(R) := \min\{k \in \mathbb{N} \mid k = |S| \text{ and } S \in \mathcal{C}(R)\}.$$

EXAMPLE 1. *Consider the preference profile R depicted in Figure 1. As R does not have a Condorcet winner* $\dim_C(R) \geq 2$. *It can easily be checked that* $\{a, b\}$ *(like any other two-element set in this case) is a Condorcet winning set of R and, thus,* $\dim_C(R) = 2$.

1	1	1
a	b	c
b	c	a
c	a	b

Figure 1: A preference profile of Condorcet dimension 2.

In this work, we address the computational problem of finding a preference profile of a given Condorcet dimension. To this end, we define the problem of checking whether for a given number of agents n and alternatives m there exists a preference profile R with $\dim_C(R) = k$.

Name: CHECK-CONDORCET-DIMENSION-k
Instance: A pair of natural numbers n and m.
Question: Does there exist a preference profile R with n agents and m alternatives that has Condorcet dimension of at least k?

Preference profiles	$n=3$	$n=5$	$n=6$	$n=7$	$n=10$	$n=15$
$m=5$	$\sim 1.7 \cdot 10^{6}$	$\sim 2.5 \cdot 10^{10}$	$\sim 3.0 \cdot 10^{12}$	$\sim 3.6 \cdot 10^{14}$	$\sim 6.2 \cdot 10^{20}$	$\sim 1.5 \cdot 10^{31}$
$m=6$	$\sim 3.7 \cdot 10^{8}$	$\sim 1.9 \cdot 10^{14}$	$\sim \mathbf{1.4 \cdot 10^{17}}$	$\sim 1.0 \cdot 10^{20}$	$\sim 3.7 \cdot 10^{28}$	$\sim 7.2 \cdot 10^{42}$
$m=7$	$\sim 1.3 \cdot 10^{11}$	$\sim 3.3 \cdot 10^{18}$	$\sim 1.6 \cdot 10^{22}$	$\sim 8.3 \cdot 10^{25}$	$\sim 1.1 \cdot 10^{37}$	$\sim 3.4 \cdot 10^{55}$
$m=10$	$\sim 4.8 \cdot 10^{19}$	$\sim 6.3 \cdot 10^{32}$	$\sim 2.3 \cdot 10^{39}$	$\sim 8.3 \cdot 10^{45}$	$\sim 4.0 \cdot 10^{65}$	$\sim 2.5 \cdot 10^{98}$

Table 1: Number of objects involved in the CHECK-CONDORCET-DIMENSION-3 problem. For $k=3$ the subsets of size 2 are the candidates for Condorcet winning sets.

Note that the following simple observation can be used to prune the search space in terms of the number of alternatives.

OBSERVATION 1. *If there is a preference profile R of Condorcet dimension $\dim_C(R)$ involving m alternatives, then there is also one of the same dimension involving $m+1$ alternatives.*

PROOF. Let R be a preference profile on a set of m alternatives A with $\dim_C(R)$. We need to construct a preference profile R' on a set of $m+1$ alternatives $A' = A \cup \{a'\}$ with $a' \notin A$ such that $\dim_C(R') = \dim_C(R)$. For each i, define $R'_i := R_i \cup \{(x, a') \mid x \in A\}$, i.e., add a' in the last place of agent i's preference ordering. It is then immediately clear that $\mathcal{C}(R) \subseteq \mathcal{C}(R')$, which establishes $\dim_C(R) \geq \dim_C(R')$. On the other hand, if we assume $\dim_C(R) > \dim_C(R')$, then there exist a Condorcet winning set S' for R' of size $k := |S'| < \dim_C(R)$. This set, however, must–by the construction of R'–also be a Condorcet winning set for R; a contradiction. □

3. METHODOLOGY

The number of objects potentially involved in the CHECK-CONDORCET-DIMENSION-k problem are given in Table 1 for $k=3$. It is immediately clear that a naïve algorithm will not solve the problem in a satisfactory manner. This section describes our algorithmic efforts to solve this problem for reasonably large instances.

3.1 Translation to propositional logic (SAT)

In order to solve the problem CHECK-CONDORCET-DIMENSION-k for arbitrary $k \in \mathbb{N}$, we follow a similar approach as Brandt et al. [4]: we translate the problem to propositional logic (on a computer) and use state-of-the-art SAT solvers to find a solution. At a glance, the overall solving steps are shown in Algorithm 1.

Generally speaking, the problem at hand can be understood as the problem of finding a preference profile that satisfies certain conditions—here: having a Condorcet dimension of at least k). Thus, a satisfying instance of the propositional formula to be designed should represent a preference profile. To capture this, a formalization based on two types of variables suffices. The boolean variable $r_{i,a,b}$ represents $a\ R_i\ b$, i.e., agent i ranking alternative a at least as high as alternative b; and the variable $c_{S,y}$ stands for the set S covering alternative y.

In more detail, the following conditions/axioms need to be formalized:[2]

[2]The further axiom for neutrality is not required for correctness, but speeds up the solving process. It is discussed in Section 3.2.

Input: positive integers n and m
Output: whether there exists a preference profile R with n agents and m alternatives and $\dim_C(R) \geq k$

```
/* Encoding of problem in CNF                         */
File cnfFile;
foreach agent i do
    cnfFile += Encoder.reflexivePreferences(i);
    cnfFile += Encoder.completePreferences(i);
    cnfFile += Encoder.transitivePreferences(i);
    cnfFile += Encoder.antisymmetricPreferences(i);
foreach set S ⊆ A with |S| = k − 1 do
    cnfFile += Encoder.noCondorcetWinningSet(S);
/* Symmetry breaking                                  */
cnfFile += Encoder.neutrality();
/* SAT solving                                        */
satisfiable = SATsolver.solve(cnfFile);
if instance is satisfiable then
    return true;
else
    return false
```

Algorithm 1: SAT-CHECK-CONDORCET-DIMENSION-k

1. All n agents have linear orders over the m alternatives as their preferences (short: linear preferences)
2. For each set $S \subseteq A$ with $|S| = k-1$, it is not the case that S is a Condorcet winning set (short: no Condorcet set)

For the first axiom, we encode reflexivity, completeness, transitivity, and anti-symmetry of the relation R_i for all agents i. The complete translation to CNF (conjunctive normal form, the established standard input format for SAT solvers) is given exemplarily for the case of transitivity; the other axioms are converted analogously.

In formal terms transitivity can be written as

$$\begin{aligned} & (\forall i)(\forall x,y,z)\,(x\ R_i\ y \wedge y\ R_i\ z \rightarrow x\ R_i\ z) \\ \equiv\ & (\forall i)(\forall x,y,z)\,(r_{i,x,y} \wedge r_{i,y,z} \rightarrow r_{i,x,z}) \\ \equiv\ & \bigwedge_i \bigwedge_{x,y,z} (\neg\,(r_{i,x,y} \wedge r_{i,y,z}) \vee r_{i,x,z}) \\ \equiv\ & \bigwedge_i \bigwedge_{x,y,z} (\neg r_{i,x,y} \vee \neg r_{i,y,z} \vee r_{i,x,z})\,, \end{aligned}$$

which then translates to the pseudo code in Algorithm 2 for generating the CNF file. The key in the translation of the inherently higher order axioms to propositional logic is (as pointed out by Geist and Endriss [7] already) that because of finite domains, all quantifiers can be replaced by finite conjunctions or disjunctions, respectively.

In all algorithms, a subroutine $\mathrm{r}(i,x,y)$ takes care of the

foreach *agent* i **do**
  foreach *alternative* x **do**
    foreach *alternative* y **do**
      foreach *alternative* z **do**
        variable_not(r(i, x, y));
        variable_not(r(i, y, z));
        variable(r(i, x, z));
        newClause;

Algorithm 2: Encoding of transitivity of individual preferences

compact enumeration of variables.[3]

The axiom "no Condorcet set" can be formalized in a similar fashion, but requires further subroutines to avoid an exponential blow-up of the size of the formula in CNF. In short, the axiom can be written as

$$
\begin{aligned}
& (\forall S \subseteq A)\,(|S| = k-1 \rightarrow S \notin \mathcal{C}(R)) \\
\equiv\ & (\forall S \subseteq A)\left(|S| = k-1 \rightarrow (\exists y \notin X) S \not\succ_R^\theta y\right) \\
\equiv\ & \bigwedge_{\substack{S\subseteq A\\ |S|=k-1}} \bigvee_{y\notin X} \neg c_{S,y}.
\end{aligned}
$$

It remains as part of this axiom to define a sufficient condition for $S \succ_R^\theta y$. In the following, we denote the smallest number of agents required for a strict θ-majority by $m(n) := \lfloor \theta k \rfloor + 1$. In formal terms, we write for each set $S \subseteq A$ with $|S| = k-1$ and each alternative $y \notin X$:

$$
\begin{aligned}
& S \succ_R^\theta y \leftarrow ((\exists M \subseteq N)|M| = m(n) \wedge \\
& \qquad (\forall i \in M)(\exists x \in S) x\ R_i\ y) \\
\equiv\ & S \succ_R^\theta y \vee ((\forall M \subseteq N)|M| = m(n) \rightarrow \\
& \qquad (\exists i \in M)(\forall x \in S) \neg x\ R_i\ y) \\
\equiv\ & c_{S,y} \vee \left(\bigwedge_{\substack{M\subseteq N\\ |M|=m(n)}} \bigvee_{i\in M} \bigwedge_{x\in S} \neg r_{i,x,y} \right).
\end{aligned}
$$

In order to avoid an exponential blow-up when converting this formula to CNF, variable replacement (a standard procedure also known as Tseitin transformation) is applied. In our case, we replaced $\bigwedge_{x\in S} \neg r_{i,x,y}$ by new variables of the form $h_{S,y,i}$ and introduced the following defining clauses:[4]

$$
\begin{aligned}
& \bigwedge_{\substack{S\subseteq A\\ |S|=k-1}} \bigwedge_{y\in A} \bigwedge_{i\in N} \left(h_{S,y,i} \rightarrow \bigwedge_{x\in S} \neg r_{i,x,y} \right) \\
\equiv\ & \bigwedge_{\substack{S\subseteq A\\ |S|=k-1}} \bigwedge_{y\in A} \bigwedge_{i\in N} \left(\neg h_{S,y,i} \vee \bigwedge_{x\in S} \neg r_{i,x,y} \right) \\
\equiv\ & \bigwedge_{\substack{S\subseteq A\\ |S|=k-1}} \bigwedge_{y\in A} \bigwedge_{i\in N} \bigwedge_{x\in S} (\neg h_{S,y,i} \vee \neg r_{i,x,y}).
\end{aligned}
$$

[3]The DIMACS CNF format only allows for integer names of variables. But since we know in advance how many agents and alternatives there are, we can simply use a standard enumeration method for tuples of objects.

[4]Note that one direction of the standard bi-implication suffices here.

In this case, the helper variables even have an intuitive meaning as $h_{S,y,i}$ enforces that for no alternative $x \in S$ it is the case that agent i prefers alternative y over alternative x, i.e., agent i does not contribute to S θ-covering y.

Note that the conditions like $|S| = k-1$ can easily be fulfilled during generation of the corresponding CNF formula on a computer. For enumerating all subsets of alternatives of a given size we, for instance, used Gosper's Hack [8].

The corresponding pseudo code for the "no Condorcet set" axiom can be found in Algorithm 3.

foreach *set* $S \subseteq A$ *with* $|S| = k-1$ **do**
  foreach *alternative* $y \notin S$ **do**
    variable_not(c(S, y));
  newClause;
  `/* Definition of variable` $c_{S,y}$ `*/`
  foreach *set* $M \subseteq N$ *with* $|M| = m(n)$ **do**
    variable(c(S, y));
    foreach *agent* $i \in M$ **do**
      variable(h(S, y, i));
    newClause;
  `/* Definition of auxiliary variable` $h_{S,y,i}$ `*/`
  foreach *agent* $i \in N$ **do**
    foreach $x \in S$ **do**
      variable_not(r(i, x, y));
      variable_not(h(S, y, i));
      newClause;

Algorithm 3: Encoding of the axiom "no Condorcet set"

With all axioms formalized in propositional logic, we are now ready to search for preference profiles R of Condorcet dimension $\dim_C(R) \geq k$. Before we do so, however, we describe a (standard) optimization technique called symmetry breaking, which speeds up the solving process of the SAT solver.

3.2 Optimized computation

Observe that from a given example of a preference profile R with $\dim_C(R) \geq k$ we can always generate further examples simply by permuting the (names of the) alternatives. One could say that all positive witnesses to the SAT-Check-Condorcet-Dimension-k problem are invariant under permutations of the alternatives. Therefore, we implemented a standard technique in SAT solving called symmetry breaking; here in the form of setting agent 1's preferences to a fixed preference ordering, for instance to lexicographic preferences. This trims the search space for the SAT solver and therefore reduces the runtime of the solving process. An encoding can be achieved simply by adding a subformula of the form

$$\bigwedge_{x<y} r(n_1, x, y),$$

which sets the first agents preferences to lexicographic ordering.

4. INITIAL RESULTS

All computations were run on a Intel Core i5, 2.66GHz (quad-core) machine with 12 GB RAM using the SAT solver plingeling [1].

1	1	1	1	1	1
a	b	c	d	e	f
b	c	d	f	a	e
c	e	f	b	d	a
d	d	a	e	b	b
e	f	e	a	c	c
f	a	b	c	f	d

Figure 2: A smallest preference profile of Condorcet dimension 3 (with $n = 6$ agents $m = 6$ alternatives).

When called with the parameters $n = m = 6$, our implementation of SAT-CHECK-CONDORCET-DIMENSION-k returns the preference profile R^{dim3} within about one second. R^{dim3} is a smallest preference profile of Condorcet dimension 3 and is shown in Figure 2.[5]

Furthermore, it turns out that this preference profile is a smallest profile of Condorcet dimension 3. All strictly smaller profiles (i.e., with less agents and at most as many alternatives, or with less alternatives and at most as many agents) can be shown to have a Condorcet dimension of at most 2 via SAT-CHECK-CONDORCET-DIMENSION-3.[6]

An overview of further (preliminary) results can be found in Table 2.

```
Model (decoding of satisfying assignment) found:
Agent 0: 0 > 1 > 2 > 3 > 4 > 5
Agent 1: 2 > 3 > 5 > 0 > 4 > 1
Agent 2: 5 > 4 > 0 > 1 > 2 > 3
Agent 3: 3 > 5 > 1 > 4 > 0 > 2
Agent 4: 4 > 0 > 3 > 1 > 2 > 5
Agent 5: 1 > 2 > 4 > 3 > 5 > 0
does not have a Condorcet winning set of size 2
(6 agents and 6 alternatives).
Witnesses:
{0, 1} does not cover alternative(s): 5
{0, 2} does not cover alternative(s): 4
{1, 2} does not cover alternative(s): 0
{0, 3} does not cover alternative(s): 4
{1, 3} does not cover alternative(s): 0
{2, 3} does not cover alternative(s): 0
{0, 4} does not cover alternative(s): 5
{1, 4} does not cover alternative(s): 5
{2, 4} does not cover alternative(s): 1
{3, 4} does not cover alternative(s): 2
{0, 5} does not cover alternative(s): 3
{1, 5} does not cover alternative(s): 3
{2, 5} does not cover alternative(s): 1
{3, 5} does not cover alternative(s): 2
{4, 5} does not cover alternative(s): 3
```

Figure 3: Output of SAT-CHECK-CONDORCET-DIMENSION-3 for $n = 6$ agents and $m = 6$ alternatives.

[5]The witnesses for all sets $S \subseteq A$ with $|S| = 2$ not being Condorcet winning sets are also returned by SAT-CHECK-CONDORCET-DIMENSION-3 and can be obtained from the output in Figure 3. That there is a larger set (e.g., $\{a, b, c\}$) which forms a Condorcet winning set can easily be confirmed manually (or by calling SAT-CHECK-CONDORCET-DIMENSION-4).

[6]The running time to check all cases again is only a few seconds.

$m \backslash n$	1	2	3	4	5	6	7	8	9	10	11	12
1	–	–	–	–	–	–	–	–	–	–	–	–
2	–	–	–	–	–	–	–	–	–	–	–	–
3	–	–	–	–	–	–	–	–	–	–	–	–
4	–	–	–	–	–	–	–	–	–	–	–	–
5	–	–	–	–	–	–	–	–	–	–	–	–
6	–	–	–	–	–	+	–		–	–	–	+
7	–	–	–	–	–	+	–		–	+		+
8	–	–	–		–	+				+		+
9	–	–	–			+				+		+
10	–	–	–			+				+		+

Table 2: Preliminary collection of results obtained with SAT-CHECK-CONDORCET-DIMENSION-3 for different numbers of alternatives m and voters n. A plus (+) stands for a preference profile found; a minus (–) for the fact that all preference profiles have a Condorcet winning set of size 2.

5. OUTLOOK AND FUTURE WORK

Our implementation might be useful to find preference profiles of Condorcet dimension 4, a problem that has been raised by Elkind et al. [5]. Even though with the current formalization the solving process did not terminate within a reasonable amount of time, we intend to further pursue this direction in future work. Adding further symmetry breaking clauses (which make use of anonymity in addition to neutrality) could be a first step in this direction.

Furthermore, one could extend the notion of Condorcet dimension to other individual preferences, e.g., with agents having weak (i.e., ties are allowed) or even incomplete preferences. Because of the high flexibility of our SAT formalization, one can easily apply the same method to analyze these related concepts and questions.[7]

A formalization with other solving techniques, e.g., ASP [6], might be another way to achieve the desired performance.

Acknowledgments

This material is based upon work supported by Deutsche Forschungsgemeinschaft under grant BR 2312/9-1. The author thanks Felix Brandt and Hans Georg Seedig for helpful discussions and their support.

REFERENCES

[1] A. Biere. Lingeling, Plingeling and Treengeling entering the SAT competition 2013. In *Proceedings of the SAT Competition 2013*, pages 51–52, 2013.

[2] A. Biere, M. Heule, H. van Maaren, and T. Walsh, editors. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009.

[3] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. *Journal of Artificial Intelligence Research*, 55:565–602, 2016.

[4] F. Brandt, C. Geist, and H. G. Seedig. Identifying k-majority digraphs via SAT solving. In *Proceedings of the 1st AAMAS Workshop on Exploring Beyond*

[7]For the two suggested variants, deleting axioms from the formalization suffices.

the Worst Case in Computational Social Choice (EXPLORE), 2014.

[5] E. Elkind, J. Lang, and A. Saffidine. Choosing collectively optimal sets of alternatives based on the condorcet criterion. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 186–191. AAAI Press, 2011.

[6] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub. Answer set solving in practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(3):1–238, 2012.

[7] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40: 143–174, 2011.

[8] D. E. Knuth. *Combinatorial Algorithms*, volume 4A, part 1 of *The Art of Computer Programming*. Addison-Wesley, 2011.

[9] P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

7 PRACTICAL CONTRIBUTIONS

[8] Analyzing the practical relevance of voting paradoxes via Ehrhart theory, computer simulations, and empirical data. In *Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 385–393. IFAAMAS, 2016 (with F. Brandt and M. Strobel).

[9] Pnyx: A powerful and user-friendly tool for preference aggregation. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1915–1916. IFAAMAS, 2015 (with F. Brandt and G. Chabin).

ANALYZING THE PRACTICAL RELEVANCE OF VOTING PARADOXES VIA EHRHART THEORY, COMPUTER SIMULATIONS, AND EMPIRICAL DATA [8]

Peer-reviewed Conference Paper

Authors: F. Brandt, C. Geist, and M. Strobel

In: Proceedings of the 15th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 385–393. IFAAMAS, 2016

Abstract: Results from social choice theory are regularly used to argue about collective decision making in computational multiagent systems. A large part of the social choice literature studies voting paradoxes in which seemingly mild properties are violated by common voting rules. In this paper, we investigate the likelihood of the Condorcet Loser Paradox (CLP) and the Agenda Contraction Paradox (ACP) using Ehrhart theory, computer simulations, and empirical data. We present the first analytical results for the CLP on four alternatives and show that our experimental results, which go well beyond four alternatives, are in almost perfect congruence with the analytical results. It turns out that the CLP—which is often cited as a major flaw of some Condorcet extensions such as Dodgson's rule, Young's rule, and MaxiMin—is of no practical relevance. The ACP, on the other hand, frequently occurs under various distributional assumptions about the voters' preferences. The extent to which it is real threat, however, strongly depends on the voting rule, the underlying distribution of preferences, and, somewhat surprisingly, the parity of the number of voters.

Contribution of thesis author: Project management and guidance; partial implementation; presentation; paper management

Analyzing the Practical Relevance of Voting Paradoxes via Ehrhart Theory, Computer Simulations, and Empirical Data

Felix Brandt
TU München
Munich, Germany
brandtf@in.tum.de

Christian Geist
TU München
Munich, Germany
geist@in.tum.de

Martin Strobel
TU München
Munich, Germany
martin.r.strobel@gmail.com

ABSTRACT

Results from social choice theory are increasingly used to argue about collective decision making in computational multiagent systems. A large part of the social choice literature studies voting paradoxes in which seemingly mild properties are violated by common voting rules. In this paper, we investigate the likelihood of the Condorcet Loser Paradox (CLP) and the Agenda Contraction Paradox (ACP) using Ehrhart theory, computer simulations, and empirical data. We present the first analytical results for the CLP on four alternatives and show that our experimental results, which go well beyond four alternatives, are in almost perfect congruence with the analytical results. It turns out that the CLP—which is often cited as a major flaw of some Condorcet extensions such as Dodgson's rule, Young's rule, and MaxiMin—is of no practical relevance. The ACP, on the other hand, frequently occurs under various distributional assumptions about the voters' preferences. The extent to which it is real threat, however, strongly depends on the voting rule, the underlying distribution of preferences, and, somewhat surprisingly, the parity of the number of voters.

Keywords

Social choice theory; voting; Ehrhart theory

1. INTRODUCTION

Results from social choice theory are increasingly used to argue about collective decision making in computational multiagent systems (see, e.g., [12, 8, 30, 9]). A large part of the social choice literature studies voting paradoxes in which seemingly mild properties are violated by common voting rules. Moreover, there are a number of sweeping impossibilities, which entail that there exists no "optimal" voting rule that avoids all paradoxes. As a consequence, much of the research in social choice theory is concerned with whether a paradox can appear for a given voting rule or not. However, it turns out that some paradoxes—while possible in principle—will almost never occur in practice.

An extreme example of this phenomenon was recently revealed for the voting rule *TEQ*. Due to its unwieldy recursive definition, it was unknown for more than 20 years whether *TEQ* satisfies any of a number of very basic desirable properties. In 2011, Brandt et al. [7] have shown that *TEQ* violates all of these properties. However, their proof is non-constructive and only shows the existence of astronomically large counterexamples requiring about 10^{136} alternatives. While there are smaller computer-generated counterexamples [4], computer experiments have shown that these counterexamples are *extremely* rare and that *TEQ* satisfies the desirable properties for all practical purposes [6]. These findings motivated us to provide analytical, experimental, and empirical justifications for such statements.

In this paper, we study two voting paradoxes. The first is the well-known Condorcet loser paradox (CLP), which occurs when a voting rule selects the Condorcet loser, an alternative that loses against every other alternative in pairwise majority contests. Perhaps surprisingly, this paradox affects some Condorcet extensions, i.e., voting rules that are guaranteed to select an alternative that *wins* against every other alternative in pairwise majority contests. Common affected Condorcet extensions are Dodgson's rule, Young's rule, and MaxiMin [18]. The second paradox, called agenda contraction paradox (ACP), occurs when removing losing alternatives changes the set of winners. There are only few voting rules that do not suffer from this paradox, one of them being the essential set. In fact, all common voting rules that violate the CLP also violate the ACP.

In principle, quantitative results on voting paradoxes can be obtained via three different approaches. The analytical approach uses theoretical models to quantify paradoxes based on certain assumptions about the voters' preferences. Analytical results usually tend to be quite hard to obtain and are limited to simple—and often unrealistic—assumptions. The experimental approach uses computer simulations based on underlying stochastic models of how the preference profiles are distributed. Experimental results have less general validity than analytical results, but can be obtained for arbitrary distributions of preferences. Finally, the empirical approach is based on evaluating real-world data to analyze how frequently paradoxes actually occur or how frequently they would have occurred if certain voting rules had been used for the given preferences. Unfortunately, only very limited real-world data for elections is available.

Our main results are as follows.

Using Ehrhart theory, we compute upper bounds for the CLP as well as the exact probabilities under which the CLP occurs for MaxiMin when there are four alternatives and preferences are distributed according to the Impartial Anonymous Culture (IAC) distribution. This approach also

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

yields the exact limit probabilities (for CLP and ACP) when the number of voters goes to infinity. To the best of our knowledge, these are the first analytical results for the CLP on four alternatives (which is the minimal number of alternatives for which the voting rules we consider exhibit the CLP).

For both the CLP and the ACP, we throughly analyze a variety of other settings with more alternatives and other stochastic preference models using computer simulations. For those settings in which the analytical approach is also feasible, our results are in almost perfect congruence with the analytical results. This is strong evidence for the accuracy of our simulation results.

It turns out that the CLP—which is often cited as a major flaw of some Condorcet extensions—is of no practical relevance. The maximum probability under all preference models we studied is 2.2% (for MaxiMin, three voters, four alternatives, and IAC). In more realistic settings, it is much lower. For Dodgson's rule, it never exceeds 0.01%. We did not find any occurrence of the paradox in real-world data, neither in the PrefLib library [25] nor in millions of elections based on data from the Netflix Prize [3].

The ACP, on the other hand, frequently occurs under various distributional assumptions about the voters' preferences. The extent to which it is real threat, however, strongly depends on the voting rule, the underlying distribution of preferences, and the parity of the number of voters. If the number of voters is much larger than the number of alternatives, less discriminating voting rules seem to fare better than more discriminating ones. For example, when there are 1,000 voters and four alternatives, the probability for the ACP under Copeland's rule and IAC is 9% while it occurs with a probability of 33% for Borda's rule. When there are fewer voters, the parity of the number of voters plays a surprisingly strong role. For example, if there are 6 alternatives, the ACP probability for Copeland's rule is 44% for 50 voters, but only 26% for 51 voters. These results are in line with the empirical data we analyzed.

2. RELATED WORK

There is a huge body of research on the quantitive study of voting paradoxes. Gehrlein [19] focusses on the non-existence of Condorcet winners, arguably the most studied voting paradox. Gehrlein and Lepelley [20], on the other hand, provide an overview of many paradoxes and, in particular, analyze the influence of group coherence. In addition, Gehrlein and Lepelley [20] survey different tools and techniques that have been applied over the years for the quantitive study of voting paradoxes.

The analytical study of voting paradoxes under the assumption of IAC is most effectively done via Ehrhart theory, which goes back to the year 1962 and the French mathematician Eugène Ehrhart [16]. Interestingly, parts of these results have been reinvented (in the context of social choice) by Huang and Chua [22] in 2000, before Ehrhart's original work was independently rediscovered for social choice by Wilson and Pritchard [34] and Lepelley et al. [23] more than forty years later.

Current research on the probability of voting paradoxes under IAC is based on algorithms that build upon Ehrhart's results, such as the algorithm developed by Barvinok [2]. For many years, these approaches were limited to cases with three or fewer alternatives. Recent advances in software tools and mathematical modeling enabled the study of elections with four alternatives. Bruns and Söger [10] and Schürmann [31] provide such results for Condorcet's paradox, the Condorcet efficiency of plurality and the similarity between plurality and plurality with runoff. Schürmann [31] further shows how symmetries in the formulation of the paradoxes can be exploited to facilitate the corresponding computations.

For the CLP (sometimes also referred to as "Borda's paradox") many quantitive results are known [20], which are, however, limited to simple voting rules and scoring rules in particular. These results also include some empirical evidence for the paradox under plurality ([20], p.15) and suggest that it is an unlikely yet possible problem in practice. Interestingly, the CLP for *Condorcet extensions* has—to the best of our knowledge—only been considered by Plassmann and Tideman [28]. However, they restrict their analysis to the three-alternative case and find that the CLP never occurs, which is unsurprising since provably four alternatives are required for the Condorcet extensions they considered.

The ACP appears to have received less attention in the quantitative literature on voting paradoxes. Some limit probabilities for scoring rules were obtained by Gehrlein and Fishburn (see [20], p. 282–284). Fishburn [17] experimentally studied a variant of this paradox called "winner turns loser paradox" for Borda's rule under Impartial Culture. For Condorcet extensions, Plassmann and Tideman [28] considered another variant of the ACP under a spatial model, but again limit their experiments to three alternatives. These few results already seem to indicate that the ACP might occur even under realistic assumptions. However, there are no results for more than three alternatives, Condorcet extensions, and the ACP in its full generality.

The preference models we consider (such as IC, IAC, and the Mallows-ϕ model) have also found widespread acceptance for the experimental analysis of voting rules within the multiagent systems and AI community (see, e.g., [1, 5, 21, 27]).

3. MODELS AND DEFINITIONS

Let A be a set of m alternatives and $N = \{1, \ldots, n\}$ a set of voters. Each voter is equipped with a *(strict) preference relation* $\succ_i$, i.e., a connex,[1] transitive, and asymmetric binary relation on A. We read $x \succ_i y$ as voter i (strictly) preferring alternative x to alternative y.

A *(preference) profile* (or an *election*) is an n-tuple of preference relations and will be denoted by $R := (\succ_1, \ldots, \succ_n)$. We will sometimes consider the restriction of $\succ_i$ to a subset of alternatives $B \subseteq A$, called an *agenda*. Such a restriction will be denoted by $R|_B := (\succ_1|_B, \ldots, \succ_n|_B)$.

3.1 Stochastic Preference Models

In this paper we consider five of the most common stochastic preference models. These models vary in their degree of realism. Impartial culture (IC) and impartial anonymous culture (IAC), for example, are usually considered as rather unrealistic. However, the simplicity of these models enables

[1]A binary relation $\succ_i$ on A is connex if $x \succ_i y \vee y \succ_i x$ for all $x \neq y \in A$. One may alternatively define $\succ_i$ as the irreflexive component of a complete, antisymmetric, and transitive relation $\succcurlyeq_i$.

the use of analytical tools that cannot be applied to the other models. IC and IAC typically yield higher probabilities for paradoxes than other preference models and can therefore be seen as worst-case estimates (see, e.g., [29]). We only give informal definitions here; for more extensive treatments see, e.g., Critchlow et al. [14] and Marden [24].

Impartial culture The most widely-studied distribution is the so-called *impartial culture (IC)*, under which every possible preference relation has the same probability of $\frac{1}{m!}$. Thus, every preference profile is equally likely to occur.

Impartial anonymous culture In contrast to IC the *impartial anonymous culture (IAC)* is not based on the probabilities of individual preferences but on the probabilities of whole profiles. Under IAC one assumes that each possible *anonymous* preference profile on n voters is equally likely to occur. A more formal definition is given in Section 4.1.

Mallows-ϕ model In *Mallows-ϕ model*, the distance to a reference ranking (or ground truth) is measured by means of the Kendall-tau distance[2] and a parameter ϕ is used to indicate the dispersion. The case of $\phi = 1$ means absolute dispersion and coincides with IC, the case $\phi = 0$ corresponds to no dispersion and every voter always picks the "true" ranking. We chose $\phi = 0.8$ to simulate voters with relatively bad estimates, which leads to situations in which paradoxes are more likely to occur.

Pólya-Eggenberger urn model In the Pólya-Eggenberger *urn model*, each possible preference relation is represented by a ball in an urn from which individual preferences are drawn. After each draw, the chosen ball is put back and $\alpha \in \mathbb{N}_0$ new balls of the same kind are added to the urn. While the urn model subsumes both impartial culture ($\alpha = 0$) and impartial anonymous culture ($\alpha = 1$), we set $\alpha = 10$ to obtain a reasonably realistic interdependence of individual preferences.

Spatial model In the *spatial model* alternatives and agents are placed in a multi-dimensional space uniformly at random and the agents' preferences are then determined by the Euclidean distances to the alternatives (closer alternatives are preferred to more distant ones). The spatial model is considered particularly realistic in political science where the dimensions are interpreted as different aspects of the alternatives [32]. We chose the simple case of two dimensions for our analysis.[3]

3.2 Voting Rules

A *voting rule* is a function f that maps a preference profile to a non-empty *set of winners*.

For a preference profile R, let $g_{xy} := |\{i \in N : x \succ_i y\}| - |\{i \in N : y \succ_i x\}|$ denote the *majority margin* of x against y. A very influential concept in social choice is the notion of a Condorcet winner, an alternative that wins against any other alternative in a pairwise majority contest. Alternative x is a *Condorcet winner (CW)* of a profile R if $g_{xy} > 0$ for all $y \in A \setminus \{x\}$. Conversely, alternative x is a *Condorcet loser (CL)* if $g_{yx} > 0$ for all $y \in A \setminus \{x\}$. Neither CWs nor CLs necessarily exist, but whenever they do they are unique. A voting rule f is called a *Condorcet extension* if $f(R) = \{x\}$ whenever x is the CW in R.

In the following paragraphs we briefly introduce the voting rules considered in this paper.

Borda's Rule Under *Borda's rule* each alternative receives from 0 to $|A| - 1$ points from each voter (depending on the position the alternative is ranked in). The alternatives with highest accumulated score win.

MaxiMin The *MaxiMin* rule is only concerned with the highest defeat of each alternative in a pairwise majority contest. It yields all alternatives as winners which have the maximal value of $\min_{y \in A} g_{xy}$.

Young's Rule *Young's rule* yields all alternatives that can be made a CW by removing a minimal number of voters.

Dodgson's Rule *Dodgson's rule* selects all alternatives that can be made a CW by a minimal number of pairwise swaps of adjacent alternatives in the individual preference relations.

Essential Set Consider the symmetric two-player zero-sum game G given by the skew-symmetric matrix with entries g_{xy} for all pairs of alternatives x, y. The *essential set* is the set of all alternatives that are played with positive probability in some mixed Nash equilibrium of G.[4]

Except for Borda's rule, all presented voting rules are in fact Condorcet extensions. While Borda's rule, MaxiMin, and the essential set can be computed efficiently, Young's rule and Dodgson's rule have been shown to be complete for parallel access to NP. The essential set is one of the few voting rules that do suffer from neither the CLP nor the ACP, and is merely included as a reference. For more formal definitions and computational properties of these rules, we refer to Brandt et al. [9].

3.3 Voting Paradoxes

In this paper we focus on two voting paradoxes whose occurrence can be determined given a voting rule f and a preference profile R.

Let f be a voting rule. Formally, a *(voting) paradox* is a characteristic function that maps a preference profile to 0 or 1. In the latter case, we say the paradox *occurs* for voting rule f at profile R.

The *Condorcet Loser Paradox (CLP)* occurs when a voting rule selects the CL as a winner.

DEFINITION 1. *Given a voting rule f the* Condorcet loser paradox *CLP_f is defined as*

$$CLP_f(R) = \begin{cases} 1 & \text{if } f(R) \text{ contains a CL} \\ 0 & \text{otherwise.} \end{cases}$$

The *agenda contraction paradox (ACP)* occurs when reducing the set of alternatives, by eliminating unchosen alternatives, influences the outcome of an election.

DEFINITION 2. *Given a voting rule f the* agenda contraction paradox *ACP_f is a paradox defined as*

$$ACP_f(R) = \begin{cases} 1 & \text{if } f(R|_B) \neq f(R) \text{ for some } B \supseteq f(R) \\ 0 & \text{otherwise.} \end{cases}$$

[2]The Kendall-tau distance counts the number of pairwise disagreements.

[3]In a related study, Brandt and Seedig [5] have found that the number of dimensions does not seem to have a large impact on the results as long as it is at least two.

[4]These mixed equilibria are also known as *maximal lotteries* in probabilistic social choice.

4. QUANTIFYING VOTING PARADOXES

In this section we present the three general approaches for quantifying voting paradoxes: the analytical approach via Ehrhart theory, the experimental approach via computer simulations, and the empirical approach via real-world data.

4.1 Exact Analysis via Ehrhart Theory

Anonymous preference profiles only count the number of voters for each of the $m!$ possible rankings on m alternatives. An anonymous preference profile can hence be viewed as an integer point in a space of $d := m!$ dimensions. Formally, the set $S_{m,n}$ of anonymous preference profiles on m alternatives with n voters can be identified with the set of all integer points $z = (z_1, \ldots, z_{m!}) \in \mathbb{Z}^{m!}$ which satisfy

$$z_i \geq 0 \text{ for all } i \in \{1, \ldots, m!\}, \text{ and } \sum_{i=1}^{m!} z_i = n.$$

Under IAC each anonymous preference profile is assumed to be equally likely to occur. Hence, in order to determine the probability of a paradox *under IAC* it is enough to compute the number of points belonging to preference profiles in which the paradox occurs and compare them to the total number of points in $S_{m,n}$, which is known to be $|S_{m,n}| = \binom{m!+n-1}{m!-1}$.[5]

In this framework, many paradoxes X can be described with the help of linear constraints, i.e., the set of points belonging to the event can be described with the help of (in)equalities, a polytope. For variable n, this approach then describes a *dilated polytope* $P_n = nP := \{n\vec{x} \colon \vec{x} \in P\}$. Hence, we know that the probability of a paradox X_n under IAC is given by:

$$\mathbb{P}(X_n) = \frac{|nP \cap \mathbb{Z}^d|}{|S_{m,n}|}.$$

and we can determine the probability of (many) voting paradoxes under IAC by evaluating the function $L(P, n) := |nP \cap \mathbb{Z}^d|$, which describes the number of integer points inside the dilation nP. This can be done with the help of Ehrhart theory. Ehrhart [16] was the first to show that $L(P, n)$ can be described by special functions, called quasi- or Ehrhart-polynomials. A function $f : \mathbb{Z} \to \mathbb{Q}$ is a *quasi-polynomial* of degree d and period q if there exists a list of q polynomials $f_i \colon \mathbb{Z} \to \mathbb{Q}$ $(0 \leq i < q)$ of degree d such that $f(n) = f_i(n)$ if $n \equiv i \mod q$.

Quasi-polynomials can be determined with the help of computer programs such as LattE (De Loera et al. [15]) or Normaliz (Bruns et al. [11]). Unfortunately, the computation of our quasi-polynomials is computationally very demanding, especially because the dimension of the polytopes grows super-exponentially in the number of alternatives. This limits analytical results under IAC to rather small numbers of alternatives. To the best of our knowledge, Normaliz is the only program which is able to compute polytopes corresponding to elections with up to *four* alternatives. And even Normaliz is not always able to compute the whole quasi-polynomial, but sometimes we had to resort to computing the leading coefficients only of the polynomial, which fortunately suffices for determining the limit probability of a paradox when the number of voters goes to infinity.

An overview of our analytical findings obtained in this way is provided in Table 1.

Paradox	Voting rule(s)	Result
CLP	Condorcet extensions	upper bound ($\forall n \in \mathbb{N}$)
	MaxiMin	probability ($\forall n \in \mathbb{N}$)
	Tideman's rule	limit prob. ($n \to \infty$)
ACP	MaxiMin	limit prob. ($n \to \infty$)

Table 1: Theoretical results obtained via Ehrhart theory (for four alternatives and under IAC)

Finding a Quasi-polynomial for MaxiMin

As an example for the method just described, we consider the $\text{CLP}_{\text{MaxiMin}}$ in four-alternative elections under IAC, the probabilities of which can be computed from a quasi-polynomial with degree 23 and a period of 5,040.[6]

In order to determine the polynomial, we first need to describe the corresponding polytope with equalities and inequalities. Recall the definition of MaxiMin from Section 3.2:

$$f_{\text{MaxiMin}}(R) := \arg\max_{x \in A} \min_{y \in A} g_{xy}.$$

For $\text{CLP}_{\text{MaxiMin}}(R) = 1$ the CL of R has to have the lowest highest defeat. Formally, there is $x \in A$ such that for all $y \in A \backslash \{x\}$,

$$g_{yx} > 0, \text{ and} \tag{1}$$

$$\max_{z \in A \backslash \{x\}} g_{zx} \leq \max_{z \in A \backslash \{y\}} g_{zy}. \tag{2}$$

Now let $A = \{a, b, c, d\}$ and assume $x = d$. We then have that $g_{ad}, g_{bd}, g_{cd} > 0$, which implies $\max_{z \in A \backslash \{d\}} g_{zd} > 0$. Furthermore,

$$\max_{z \in A \backslash \{y\}} g_{zy} > 0 \text{ for all } y \in \{a, b, c\},$$

from which it follows that either $g_{ab}, g_{bc}, g_{ca} > 0$ or $g_{ba}, g_{cb}, g_{ac} > 0$. In both cases there is a majority cycle between a, b, and c. Due to symmetry we can choose one direction of the cycle arbitrarily and assume $g_{ab}, g_{bc}, g_{ca} > 0$. Then,

$$\max_{z \in A \backslash \{a\}} g_{za} = g_{ca}, \max_{z \in A \backslash \{b\}} g_{zb} = g_{ab}, \text{ and } \max_{x \in A \backslash \{c\}} g_{zc} = g_{ac}.$$

Condition (1) is already represented in the form of linear inequalities. In order to model condition (2) we determine $\max_{z \in A \backslash \{d\}} g_{zd}$ and distinguish cases for the seven possible outcomes. The inequalities for the case $\max_{z \in A \backslash \{d\}} g_{zd} = \{g_{ad}\}$ are

$$g_{ad} - g_{bd} > 0 \qquad \text{and} \qquad g_{ad} - g_{cd} > 0.$$

[5] For most preference models other than IAC this approach does not work. While for specific combinations of (simple) distributions and voting rules there are some highly tailor-made computations in the literature (cf. Section 2), these have to be redesigned for each individual setting.

[6] In theory, the analysis can be adapted to also cover more complex rules (e.g., Dodgson's and Young's rule, which involve solving an ILP). It is unclear, however, how one would translate their definitions to linear inequalities.

Condition (2) furthermore yields

$$g_{ca} - g_{ad} \geq 0, \qquad g_{ab} - g_{ad} \geq 0, \qquad \text{and} \qquad g_{bc} - g_{ad} \geq 0.$$

Each case belongs to a different polytope and the polytopes are pairwise distinct, so we can compute each quasi-polynomial separately and later combine them to one. To get the final polynomial we have to multiply by eight for the four different possible choices of a CL and the two possible directions of the majority cycle. This then enables us to efficiently evaluate the *exact* probabilities for any number of voters. The results are depicted in Figure 2. The leading coefficient of the quasi-polynomial can also be used to determine the limit probability which is given by

$$\begin{aligned} &\mathbb{P}(\mathrm{CLP}_{\mathrm{MaxiMin}} = 1 \mid m = 4, n \to \infty) \\ &= 8 \cdot \frac{485052253637930099}{6443662124777472000000} \approx \underline{\underline{0.06\%}}. \end{aligned}$$

4.2 Experimental Analysis

As we will see, simulating elections with the help of computers is a viable way of achieving very good approximations for the probabilities we are looking for. It even turns out that the results of our simulations are almost indistinguishable from the theoretical result obtained via Ehrhart theory (with the exception of the limit case, which cannot be realized via simulations).

More specifically, the experimental approach works as follows: a profile source creates random preference profiles according to a specific preference model. The profiles are then used to compute the winner(s) according to a given voting rule and to determine if the paradox occurs. Any such experiment is carried out for each pair of n and m and repeated frequently. In many cases in which we covered a wide range of voters, we did not consider every possible value of n but, more economically, only simulated the values: 1–30, 49–51, 99–101, 199–201, 499–501, 999–1,001.

In contrast to many other studies, we are concerned about the statistical significance of our experimental results. Thus, we also computed 99%-confidence intervals for each data point we generated. To this end, we used the `binofit` function in MATLAB which is based on the standard approach by Clopper and Pearson [13]. It shows that, based on our sampling rate of 10^5 and 10^6, respectively, the 99%-confidence intervals are pleasantly small. Hence, even though they are depicted in all of the figures throughout this paper, sometimes it can be difficult to recognize them.

4.3 Empirical Analysis

The most valuable quantification of voting paradoxes would be their actual frequency in real-world elections. As mentioned before, real-world election data is generally relatively sparse, incomplete, and inaccurate. This makes empirical research on this topic rather difficult. Otherwise, the empirical approach strongly resembles the experimental approach.

For this paper we used two sources of empirical data. First, we used the 314 profiles with strict order preferences from the PREFLIB library [25]. Second, we had access to the 54,650 preference profiles over four alternatives without a CW which belong to the roughly 11 million four-alternative elections which Mattei et al. [26] derived from the Netflix Prize data [3]. Non-existence of Condorcet winners is a prerequisite for the paradoxes we study.

5. CONDORCET LOSER PARADOX

In this section we present our findings on the CLP. We conclude that—even though the CLP is possible in principle—it is so unlikely that it cannot be used as a serious argument against any of the Condorcet extensions we considered.

5.1 An Upper Bound

Before analyzing the CLP for concrete voting rules, we discuss an upper bound valid for all Condorcet extensions. For a Condorcet extension to choose the CL a profile obviously has to satisfy two conditions. First, there has to exist a CL in the profile. Second, no CW may exist in the profile. In the case of four-alternative elections—which is the first interesting case—we can compute the quasi-polynomial via Ehrhart theory and hence know the exact probabilities for any number of voters. The derivation and presentation of the quasi-polynomial, which has degree 23 and contains 24 polynomials, is omitted due to space constraints. The resulting probabilities for up to 1 000 voters—and a comparison with the results of an experimental analysis—can be obtained from Figure 1. The value of the limit probability is approximately 8%.

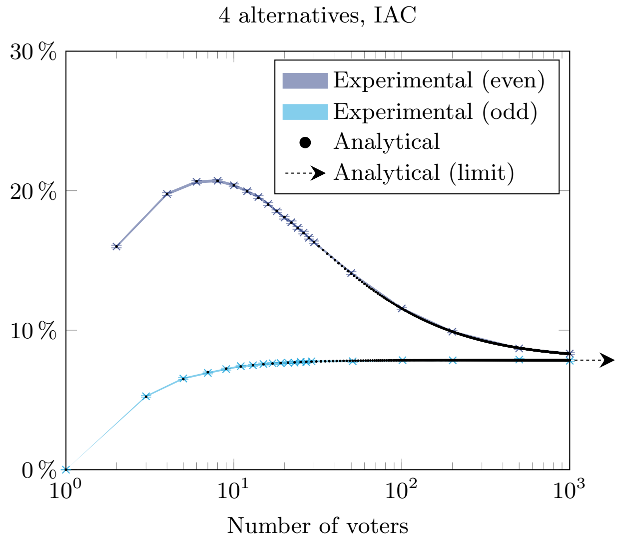


Figure 1: Probability of the event that a Condorcet extension could choose a CL in four-alternative elections under IAC

Especially for small even numbers of voters, where the probability is around 20%, the upper bound is too high to discard the CLP for Condorcet extensions altogether, and even the limit probability of 8% is relatively large. Also, for an increasing number of alternatives this problem does not vanish (for elections with 50 and 51 one voters and up to 100 alternatives the probabilities range between 5% and 25%).[7]

Note that differences between odd and even number of voters were to be expected since even numbers allow for majority ties, which have significant consequences for the paradoxes; this effect decreases for larger electorates. In the

[7]These upper bounds turn out to be relatively independent from the underlying preference distribution (among the models we considered, cf. Section 5.3).

specific case under consideration, the upper bound is generally higher for an even number of voters because the much higher likelihood of not having a CW more than counterbalances the lower likelihood of having a CL.

5.2 Results under IAC

Despite the high upper bounds from the previous section, the picture is quite clear for concrete Condorcet extensions: even under IAC, the risk of the considered Condorcet extensions selecting the CL is very low, as shown in Figure 2 and Figure 3 for four-alternative elections. The highest probability was found for CLP$_{\text{MaxiMin}}$ with 2.2% for three voters (CLP$_{\text{Young}}$ with about 0.9%). The limit probability of CLP$_{\text{MaxiMin}}$, with 0.06% is so low that for sufficiently large electorates it would occur in only one out of 10,000 elections. The same seems to hold for the limit probability for CLP$_{\text{Young}}$. The probability of CLP$_{\text{Dodgson}}$ is even significantly lower, with a maximum of about 0.01% in elections with 9,999 voters. We could determine the limit probability of 0.01% only for an approximation of Dodgson's rule by Tideman [33], which seems to be close to that for Dodgson's rule, based on our experimental data.

When increasing the number of alternatives the probabilities drop even further. For elections with more than ten alternatives they reach a negligibly small level of less than 0.005% for all considered rules and in no simulations with twelve or more alternatives we could find any occurrence of the paradox.

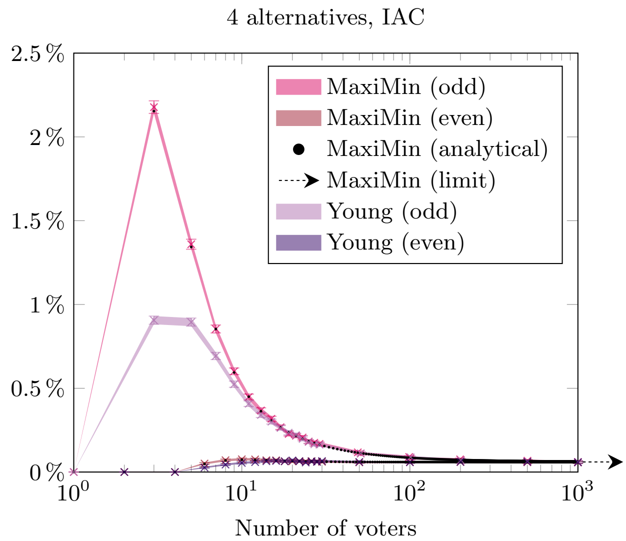


Figure 2: Comparison between CLP probabilities for MaxiMin and Young's rule under IAC in four-alternative elections

5.3 Results under Other Preference Models

Figure 4, as one would expect, shows that under more realistic assumptions the probability of the CLP decreases further in four-alternative elections with 50/51 voters, with the highest probability occurring under the unrealistic assumption of IC and the lowest probability under what may be the most realistic model in many settings, the spatial

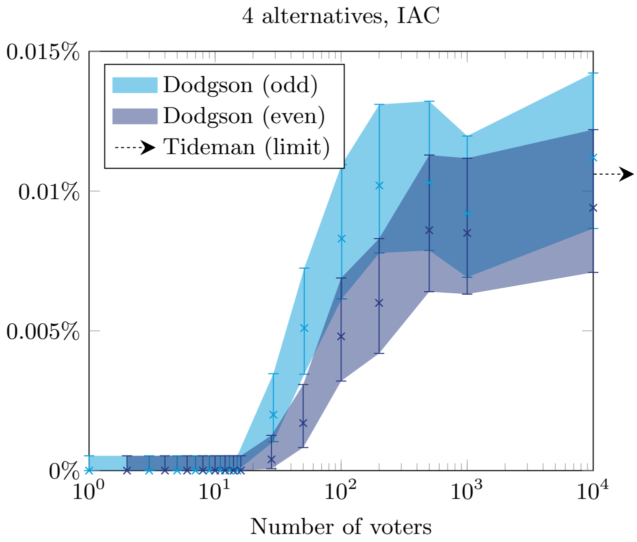


Figure 3: CLP probabilities for Dodgson's rule under IAC in four-alternative elections; the seemingly large confidence intervals are due to the small scale of the graph.

model. In our experiments, Dodgson's rule *never* selected a CL in the spatial model.

Similarly, we could not find any occurrence of the CLP in real-world data, which may be considered the strongest evidence that the CLP virtually never materializes in practice.[8]

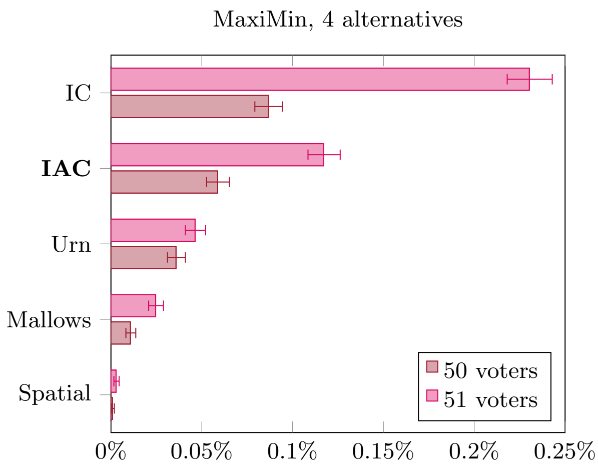


Figure 4: CLP probabilities in four-alternative elections for varying preference models and MaxiMin

[8]We tested 314 preference profiles with strict orders from the PrefLib library as well as the roughly 11 million four-alternative elections which Mattei et al. [26] derived from the Netflix Prize data. While about 54,000 of those elections were susceptible to the CLP, it never occurred under the rules we considered in this paper. In contrast, under plurality it already occurred in twelve out of the 314 PrefLib-instances.

6. AGENDA CONTRACTION PARADOX

Recall that the *agenda contraction paradox (ACP)* occurs when a reduced set of alternatives (created by the unavailability of losing alternatives) influences the outcome of an election. For many cases, it may be considered a generalization of the CLP as the following argument shows. Suppose the CL x is uniquely selected by a voting rule which implements majority rule on two-alternative choice sets. Then restricting A to $\{x, y\}$ for some alternative $y \neq x$ yields the new winner y (since $g_{xy} > 0$).

As we will see, the ACP is much more of a practical problem than the CLP. The picture, however, is not black and white. Whether or not it is a serious threat depends on the voting rule, the underlying preference distribution, and on the parity of the number of voters.

6.1 Varying Voting Rules

The ACP probability strongly varies for different voting rules (see Figure 5). Borda's rule generally exhibits the worst behavior of the rules studied, with probabilities of up to 56%, and with 34% for large electorates with 1,000 voters. In contrast, Copeland's rule is quite robust to the ACP for large electorates (with only about 8% occurrence probability for 1,000 voters).[9]

The reason for this gap between Borda's and Copeland's rule appears to be two-fold: First, Condorcet extensions are safe from this paradox as long as a CW exists; Borda's rule, by contrast, is not. Second, the discriminatory power of voting rules (i.e., their ability to select small winning sets) strongly supports the paradox. As soon as a single majority-dominated alternative is selected, the ACP has to occur. For large numbers of voters, this is in line with Copeland's rule being least discriminating among those evaluated. The essential set is among the most discriminating known voting rules immune to the ACP, but presumably less discriminating than Copeland's rule.

The behavior of MaxiMin is almost identical to that of Young's and Dodgson's rule. Confirming our approximate "limit" results of 1,000 voters, we were able to analytically compute the limit probability for MaxiMin as $\frac{331}{2048} \approx 16\%$. This is in perfect congruence with the (rounded) values for MaxiMin, Young's rule, and Dodgson's rule.

It should also be noted that with fewer than 100 voters, the parity of the number of voters plays a major role. For even numbers, significantly higher probabilities arise (which is particularly true for Copeland's rule, see above). At least part of this can be explained by a reduced probability for CWs in these cases.

For more alternatives (see the right-hand side of Figure 5), the relative behavior remains vastly unchanged with probabilities further increasing to values larger than 40% to 80% (mostly since the likelihood of a CW decreases roughly at the same rate).

6.2 Varying Preference Models

Figure 6 extends the analysis of the previous section by additionally considering preference models beyond IAC. The overall picture regarding the different rules remains the same. For large electorates Copeland's rule outperforms the other rules, whereas Borda's rule performs worst.

Regarding the different preference models, three classes emerge from Figure 6.

First, for Mallows-ϕ we observe probabilities that are vanishing with increased numbers of voters. Under the spatial model this is true as well, with the surprising exception of Borda's rule, for which the picture looks completely different and the probability does not go below 20% in the spatial model. Presumably, this can be explained by Borda's inability to select the CW in this setting, a hypothesis that deserves further study, however. On the contrary, the other rules appear to be benefitting from the fact that the existence of a CW becomes very likely under models with high voter interdependence.

Second, as expected, the assumption of IC serves as an upper bound for all other preference models. The results for IAC are not much lower, fostering the impression that IAC could also be an unrealistic upper bound.

Third, the urn model yields much lower values compared to IAC and IC. The absolute numbers, however, are still beyond acceptable levels (between 4% and 23% for 1,000 voters).

The findings in the empirical data corroborate our experimental findings. In PrefLib the ACP occurs 17 times for Borda, three times for Copeland and exactly once for MaxiMin as well as Young's and Dodgson's rule. In the Netflix data set, where the number of voters is at least 350, Copeland performs much better than the other Condorcet extensions (4,400 compared to 18,470 occurrences for the other Condorcet extensions). Borda's rule virtually always suffers from the ACP on this data set: there are 54,620 instances of ACPs already when considering profiles that do not have a CW (there are 54,650 of such).

7. CONCLUSION

We investigated the likelihood of the CLP and the ACP using Ehrhart theory, computer simulations, and empirical data. The CLP is often cited as a major flaw of some Condorcet extensions such as Dodgson's rule, Young's rule, and MaxiMin. For example, Fishburn regards Condorcet extensions that suffer from the CLP (specifically referring to the three rules mentioned above) as " 'dubious' extensions of the basic Condorcet criterion" ([18], p. 480).[10] While this is intelligible from a theoretical point of view, our results have shown that the CLP is of virtually no practical concern. The ACP, on the other hand, frequently occurs under various distributional assumptions about the voters' preferences. The extent to which it is real threat, however, strongly depends on the voting rule, the underlying distribution of preferences, and, surprisingly, the parity of the number of voters. Our main quantitative results for the worst case are summarized in Table 2. Potential future work includes the analysis of other voting paradoxes (such as monotonicity failures or the no-show paradox) and other rules (such as Nanson's rule or Black's rule).

Acknowledgments

Christian Geist was supported by Deutsche Forschungsgemeinschaft under grant BR 2312/9-1. The authors also thank Nicholas Mattei for providing the Netflix Prize data and Christof Söger for his guidance regarding Normaliz.

[9]For small *even* numbers of voters, Copeland's rule also frequently fails agenda contraction, which is also visible in Figure 6 and explains the seemingly high values in Table 2.

[10]Fishburn [18] actually analyzes violations of "Smith's Condorcet principle", which is weaker than the CLP.

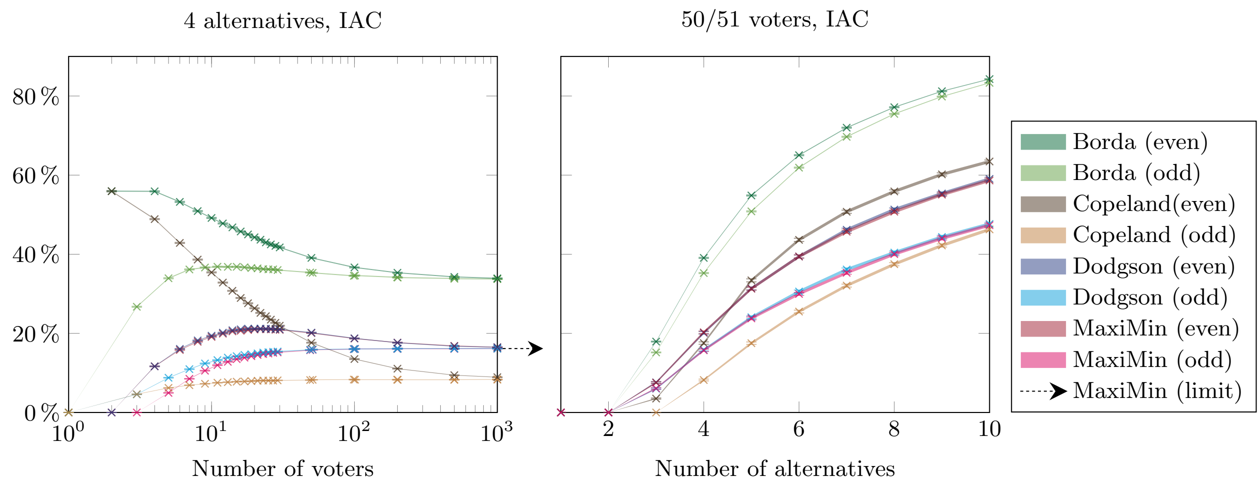


Figure 5: Comparison between ACP probabilities for different voting rules under IAC

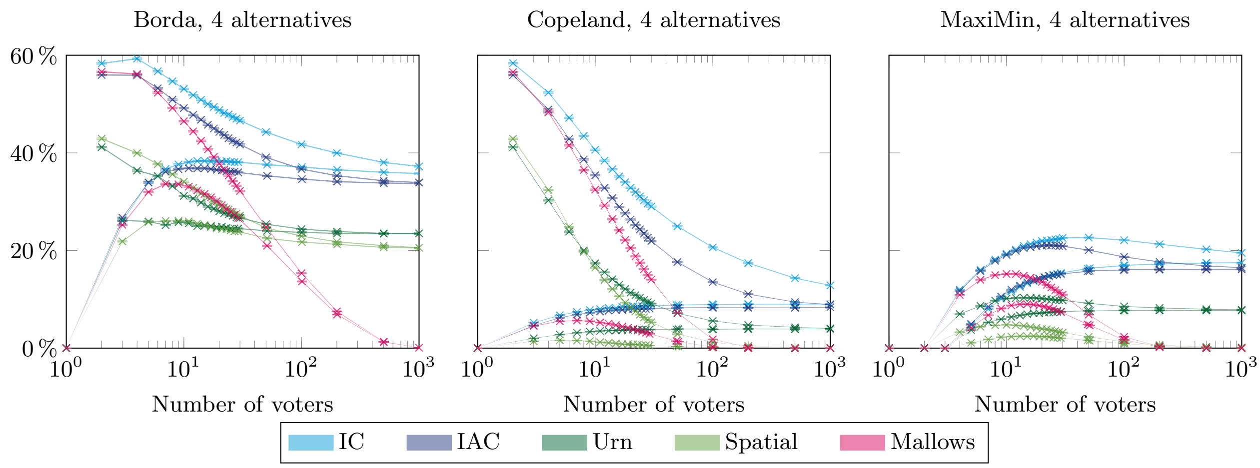


Figure 6: Comparison between $\text{ACP}_{\text{Borda}}$, $\text{ACP}_{\text{MaxiMin}}$ and $\text{ACP}_{\text{Copeland}}$ for varying preference models in four-alternative elections; the values of $\text{ACP}_{\text{Young}}$ and $\text{ACP}_{\text{Dodgson}}$ are omitted since they strongly resemble the ones of $\text{ACP}_{\text{MaxiMin}}$.

Paradox	Condorcet loser paradox (CLP)			Agenda contraction paradox (ACP)		
Model	IAC		IC	IAC		IC
n	$\{1,\dots,1000\}$	$\{50,51\}$	$\{50,51\}$	$\{1,\dots,1000\}$	$\{50,51\}$	$\{1,\dots,1000\}$
m	4	$\{1,\dots,10\}$	4	4	$\{1,\dots,10\}$	4
Essential set	0%	0%	0%	0%	0%	0%
Borda	0%	0%	0%	56%	84%	59%
Copeland	0%	0%	0%	56%	63%	58%
Dodgson	0.01%	0.005%	0.005%	21%	59%	23%
Young	1%	0.15%	0.25%	21%	59%	23%
MaxiMin	2.2%	0.15%	0.25%	21%	59%	23%

Table 2: Rounded maximal CLP and ACP probabilities which occurred during our simulations

REFERENCES

[1] H. Aziz, S. Gaspers, N. Mattei, N. Narodytska, and T. Walsh. Ties matter: Complexity of manipulation when tie-breaking with a random vote. In *Proc. of 27th AAAI Conference*, pages 74–80. AAAI Press, 2013.

[2] A. I. Barvinok. A polynomial time algorithm for counting integral points in polyhedra when the dimension is fixed. *Mathematics of Operations Research*, 19(4):769–779, 1994.

[3] J. Bennett and S. Lanning. The Netflix prize. In *Proceedings of KDD Cup and Workshop*, 2007.

[4] F. Brandt and H. G. Seedig. A tournament of order 24 with two disjoint TEQ-retentive sets. Technical report, `http://arxiv.org/abs/1302.5592`, 2013.

[5] F. Brandt and H. G. Seedig. On the discriminative power of tournament solutions. In *Proc. of 1st AAMAS Workshop on Exploring Beyond the Worst Case in Computational Social Choice (EXPLORE)*, 2014.

[6] F. Brandt, F. Fischer, P. Harrenstein, and M. Mair. A computational analysis of the tournament equilibrium set. In *Proc. of 23rd AAAI Conference*, pages 38–43. AAAI Press, 2008.

[7] F. Brandt, M. Chudnovsky, I. Kim, G. Liu, S. Norin, A. Scott, P. Seymour, and S. Thomassé. A counterexample to a conjecture of Schwartz. *Social Choice and Welfare*, 40:739–743, 2013.

[8] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiß, editor, *Multiagent Systems*, chapter 6, pages 213–283. MIT Press, 2nd edition, 2013.

[9] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[10] W. Bruns and C. Söger. The computation of generalized Ehrhart series in Normaliz. *Journal of Symbolic Computation*, 68(2):75–86, 2015. Effective Methods in Algebraic Geometry.

[11] W. Bruns, B. Ichim, T. Römer, R. Sieg, and C. Söger. Normaliz. Algorithms for rational cones and affine monoids. Available at `https://www.normaliz.uni-osnabrueck.de`.

[12] Y. Chevaleyre, U. Endriss, J. Lang, and N. Maudet. A short introduction to computational social choice. In *Proc. of 33rd Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, volume 4362 of *LNCS*, pages 51–69. Springer, 2007.

[13] C. J. Clopper and E. S. Pearson. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413, 1934.

[14] D. E. Critchlow, M. A. Fligner, and J. S. Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35:294–318, 1991.

[15] J. A. De Loera, R. Hemmecke, J. Tauzer, and R. Yoshida. Effective lattice point counting in rational convex polytopes. *Journal of Symbolic Computation*, 38(4):1273–1302, 2004.

[16] E. Ehrhart. Sur les polyedres rationnels homothetiques á n dimensions. *Compte-Rendus de l'académie des Sciences*, 254(4):616, 1962.

[17] P. C. Fishburn. Paradoxes of voting. *The American Political Science Review*, 68(2):537–546, 1974.

[18] P. C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

[19] W. V. Gehrlein. *Condorcet's Paradox*. Springer, 2006.

[20] W. V. Gehrlein and D. Lepelley. *Voting paradoxes and group coherence: the Condorcet efficiency of voting rules*. Springer-Verlag, 2010.

[21] J. Goldsmith, J. Lang, N. Mattei, and P. Perny. Voting with rank dependent scoring rules. In *Proc. of 28th AAAI Conference*, pages 698–704. AAAI Press, 2014.

[22] H. C. Huang and V. C. H. Chua. Analytical representation of probabilities under the IAC condition. *Social Choice and Welfare*, 17(1):143–155, 2000.

[23] D. Lepelley, A. Louichi, and H. Smaoui. On Ehrhart polynomials and probability calculations in voting theory. *Social Choice and Welfare*, 30(3):363–383, 2008.

[24] J. I. Marden. *Analyzing and Modeling Rank Data*. Number 64 in Monographs on Statistics and Applied Probability. Chapman & Hall, 1995.

[25] N. Mattei and T. Walsh. PrefLib: A library for preference data. In *Proc. of 3rd ADT*, volume 8176 of *LNCS*, pages 259–270. Springer, 2013.

[26] N. Mattei, J. Forshee, and J. Goldsmith. An empirical study of voting rules and manipulation with large datasets. In *Proc. of 4th COMSOC*, 2012.

[27] J. Oren, Y. Filmus, and C. Boutilier. Efficient vote elicitation under candidate uncertainty. In *Proc. of 23rd IJCAI*, pages 309–316. AAAI Press, 2015.

[28] F. Plassmann and N. Tideman. How frequently do different voting rules encounter voting paradoxes in three-candidate elections? *Social Choice and Welfare*, 42(1):31–75, 2014.

[29] M. Regenwetter, B. Grofman, A. A. J. Marley, and I. M. Tsetlin. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press, 2006.

[30] J. Rothe, editor. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer, 2015.

[31] A. Schürmann. Exploiting polyhedral symmetries in social choice. *Social Choice and Welfare*, 40(4):1097–1110, 2013.

[32] N. Tideman and F. Plassmann. Modeling the outcomes of vote-casting in actual elections. In D. S. Felsenthal and M. Machover, editors, *Electoral Systems: Paradoxes, Assumptions, and Procedures*, Studies in Choice and Welfare. Springer, 2012.

[33] T. N. Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987.

[34] M. C. Wilson and G. Pritchard. Probability calculations under the IAC hypothesis. *Mathematical Social Sciences*, 54(3):244–256, 2007.

PNYX: A POWERFUL AND USER-FRIENDLY TOOL FOR PREFERENCE AGGREGATION [9]

Peer-reviewed Conference Paper (Demonstration)

Authors: G. Chabin, F. Brandt, and C. Geist

In: Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pages 1915–1916. IFAAMAS, 2015

Abstract: *Pnyx* is an easy-to-use and entirely web-based tool for preference aggregation that does not require any prior knowledge about social choice theory. The tool is named after a hill in Athens called Pnyx, which was the official meeting place of the Athenian democratic assembly and is therefore known as one of the earliest sites in the creation of democracy. Pnyx is available at pnyx.dss.in.tum.de.

Contribution of thesis author: Project management and guidance; requirements management; partial implementation; practical and theoretical presentation; paper management

Pnyx: a Powerful and User-friendly Tool for Preference Aggregation

(Demonstration)

Felix Brandt
TU München
Munich, Germany
brandtf@in.tum.de

Guillaume Chabin
TU München
Munich, Germany
guillaume.chabin@tum.de

Christian Geist
TU München
Munich, Germany
geist@in.tum.de

ABSTRACT

Pnyx is an easy-to-use and entirely web-based tool for preference aggregation that does not require any prior knowledge about social choice theory. The tool is named after a hill in Athens called Pnyx, which was the official meeting place of the Athenian democratic assembly and is therefore known as one of the earliest sites in the creation of democracy. Pnyx is available at pnyx.dss.in.tum.de.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Types of Systems—*Decision Support*; J.4 [**Social and Behavioral Sciences**]: *Economics*

Keywords

Preference aggregation; collective decision making; social choice theory; voting; web services

1. INTRODUCTION

Preference aggregation and collective decision making are common tasks in real life and computational multiagent systems. Apart from the well-known example of political elections, there are a number of more mundane settings that require the aggregation of preferences. This includes joint decisions such as where to have lunch or a company retreat, or how to make best use of a scarce resource such as a meeting room.

While computational social choice, a novel field of study in the area of multiagent systems, has added a significant amount of insights about the algorithmic properties of voting rules, the development of *practical* and *user-friendly* IT-support for the problem of preference aggregation has found little attention. (A rare exception is the fair allocation platform `spliddit.org`, which was launched in November 2014.) Apparently, there is a practical need for such tools and we would like to contribute to fulfilling this need for the domain of preference aggregation.

Often preference aggregation is conducted using inferior aggregation methods such as plurality rule and using unsuitable tools such as `doodle.com` (which was originally intended for *scheduling* a joint activity). The goal of this work is to build a tool for preference aggregation that takes into account theoretical insights from social choice theory and does so without expecting the user to possess any knowledge of the underlying mechanisms. More concretely, we have created a first version of a web-based and user-friendly application that supports the whole process of collective decision making from setting up the poll/election to the communication of the aggregated outcome to participants. The user who sets up a poll only has to select the desired input (e.g., preference relations, sets of approved alternatives, or simply most-preferred alternatives) and output (e.g., single alternatives, rankings of alternatives, or lotteries over alternatives), and the tool then automatically selects the most appropriate aggregation method. There is also support for periodic polls (for instance, lunch polls that are conducted daily before lunchtime), for which users only need to update their preferences if desired.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

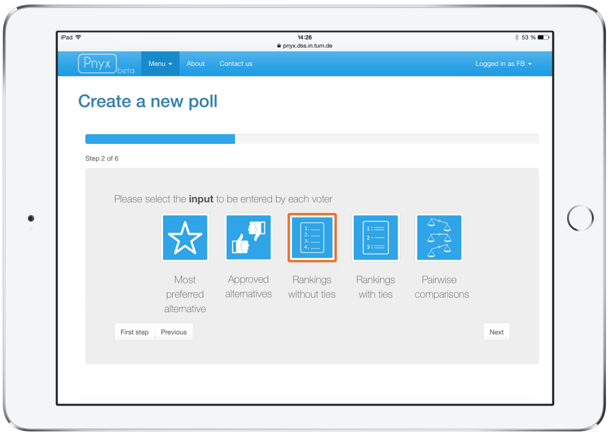


Figure 1: Screenshot of our web-based tool for preference aggregation called Pnyx.

2. PRACTICAL PERSPECTIVE

The overall visual impression and users' workflows are probably best described in a 5-minute screencast, which can be watched at `vimeo.com/118576213`. Figure 1 shows a screenshot of the creation of a new poll.

Implementation Pnyx is an entirely web-based application that was developed in *Python*, with the core of the application being built using the web framework *django*. We tried to rely on further development frameworks and open-source packages as much as possible. For the front-end, these are the HTML, CSS, and JS framework *Bootstrap* as well as the JavaScript libraries *jQuery* and *jQuery UI*. For the back-end, the aggregation engine partially relies on linear and integer programming, respectively,[1] and makes use of further Python packages for scientific computing (*NumPy*, *PuLP*). The underlying database structure is currently supported by *SQLite*.

3. THEORETICAL PERSPECTIVE

Pnyx is based on three preference aggregation rules: Borda's rule, Fishburn's rule, and Kemeny's rule. Which of these mechanisms is selected depends on the *output type* chosen by the poll creator:[2]

Single alternative: Borda's rule Borda's rule is a simple scoring rule that is particularly intuitive when preferences are linear orders. When there are m alternatives, each voter assigns a score of $m - 1$ to his most-preferred alternative, $m-2$ to his second most-preferred alternatives, etc. The alternative with the highest accumulated score wins. We consider a natural extension of Borda's rule to arbitrary binary relations where the score each voter assigns to alternative x is the number of alternatives that x is preferred to minus the number of alternatives that are preferred to x.

Lottery over alternatives: Fishburn's rule The rule that we call Fishburn's rule here was first proposed by Kreweras and studied in much more detail by Fishburn [3]. The rule returns a so-called *maximal lottery*, i.e., a probability distribution over the alternatives that is weakly preferred to any other such probability distribution. Maximal lotteries are equivalent to mixed maximin strategies (or Nash equilibria) of the symmetric zero-sum game given by the pairwise majority margins, which allows us to use linear programming for their computation. For more details on Fishburn's rule, we refer to Brandl et al. [1].

Ranking of alternatives: Kemeny's rule Kemeny's rule [4] is an aggregation rule which returns a ranking of the alternatives that maximizes pairwise agreement, i.e., a ranking in which as many pairwise comparisons as possible coincide with the preferences of the voters. Alternatively, Kemeny's rule can be characterized using maximum likelihood estimation or a consistency property for electorates of variable size [5]. We implemented the NP-hard problem of finding a Kemeny ranking via integer programming.

All of Pnyx' rules belong to Fishburn's class of C2 functions [2]. There is a long and ongoing debate in social choice theory about the advantages and disadvantages of certain aggregation rules. While it has become increasingly clear that there is no optimal rule for all purposes, there is strong evidence that some rules are inferior to others in terms of desirable axiomatic properties that have been proposed in the literature. We tried our best to preselect three rules that, in our view, represent a decent compromise between various of these properties.

[1]Currently implemented via *GLPK*

[2]Note that, when *inputs* are simply given as unique most-preferred alternatives or sets of approved alternatives (dichotomous preferences), all three rules coincide with the well-known *plurality rule* and *approval voting rule*, respectively.

As *inputs*, Pnyx supports five different choices of individual preference types:

Most preferred alternative Each voter can only select a unique most-preferred alternative among all alternatives. With these individual preferences, all three aggregation rules coincide with *plurality rule*.

Dichotomous preferences Each voter may approve an arbitrary number of alternatives and automatically disapproves the remaining ones. There is no distinction between alternatives within the set of approved or non-approved alternatives, respectively. With these individual preferences, all three aggregation rules coincide with *approval voting*.

Linear order Each voter has to provide a ranking of the alternatives *without* ties.

Complete preorder This input format is a generalization of *linear orders* that allows ties between alternatives.

Complete binary relation This is the most general form of preferences supported by Pnyx. Voters may specify each individual pairwise comparison among alternatives. By default, indifference between any pair is assumed. Note that transitivity of the preferences is no longer required.

4. CONCLUSION AND OUTLOOK

While the system is fully operative with its core features already, there are further ideas of how it could be extended:

Verification of randomness A particularly challenging problem for probabilistic methods will be to conduct randomizations in a user-verifiable way. To this end, we intend to review and employ cryptographic protocols developed in the e-voting and cryptography communities.

Anonymous preference collection There is an increasing demand for practical preference collection, which, for instance, the *PrefLib* library aims to satisfy. Pnyx could contribute to such a library by means of anonymized preference data.[3]

Acknowledgments

This work was partially supported by Deutsche Forschungsgemeinschaft under grants BR 2312/7-2 and BR 2312/9-1. Pnyx has been presented at the Workshop on Challenges in Algorithmic Social Choice (Bad Belzig, Germany) in October 2014. The authors thank two anonymous reviewers for their helpful comments.

REFERENCES

[1] F. Brandl, F. Brandt, and H. G. Seedig. Consistent probabilistic social choice. 2015. Working Paper.

[2] P. C. Fishburn. Condorcet social choice functions. *SIAM Journal on Applied Mathematics*, 33(3):469–489, 1977.

[3] P. C. Fishburn. Probabilistic social choice based on simple voting comparisons. *Review of Economic Studies*, 51 (4):683–692, 1984.

[4] J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88:577–591, 1959.

[5] H. P. Young. Optimal voting rules. *Journal of Economic Perspectives*, 9(1):51–64, 1995.

[3]PrefLib's data format is already supported as an export option for poll administrators.

Part III

DISCUSSION AND CONCLUSION

8 DISCUSSION OF THE PRESENTED METHODS

Despite the success cases and new insights generated in this thesis, of course, the presented approach also holds some challenges, which we consider here before we conclude in the next chapter by offering some ideas for future work. We take a critical stance and discuss the approach with respect to its usability, its verifiability, and its limitations.

8.1 USABILITY OF THE TOOLS

The question of whether the presented methods can be applied by non-specialists is a particularly valid concern. Clearly, in their current state without general tool support, applying solving methodologies to social choice theory remains a task for expert users with programming skills. While this does not impact the overall power of the approach, it obviously limits the degree to which it can be broadly used by any researcher in social choice.

But even if we envision user-friendly tools that help formalizing concepts of social choice theory in the languages of solvers, doubts remain that this will change the game. In our experience, the design of efficient encodings has to follow the requirements of—and needs to be optimized for—the concrete problem. Hence, no one-fits-all encoding appears to exist that could be the basis for a general toolset. In a way, this appears to also be an issue for very advanced and general proof assistants with highly expressive input languages, such as the Isabelle system [Nipkow et al., 2002]: while many problems can be easily and intuitively formalized even by untrained mathematicians, the ability of these systems to discover *new* results is rather limited due to the high complexity of the *general* problem.

Yet, some basic toolsets to assist *expert users* when formalizing concepts from social choice are certainly desirable and should be achievable based on the commonalities of existing contributions. It remains an interesting question to which extend such tools can take the role of an automatic proof assistant which allows researchers to quickly test hypotheses on small domains without giving up too much generality and efficiency.

8.2 VERIFIABILITY OF THE RESULTS

As for any mathematical result or piece of software, also for our theorems and tools there is the important question of how correctness can be ensured. Compared to theorems with classical proofs, however, many people appear to be particularly skeptical of computer-aided and computer-generated results; probably because of the lack of a classical certificate (human-readable proof) and since the "thinking" is invisibly carried out by a machine rather than a human.

If we leave the manually proven lemmas out of the discussion, we are left with two aspects of our approach that might be felt to deserve additional verification:

1. the correctness of the encoding (i.e., "Does the encoding actually represent the given problem or question?"), and

2. the correctness of the solver output (i.e., "Is the answer we got from the solver correct given the encoding?")

As we have argued in Publications [1], [2], and [3], extracting a human-readable proof from the computer-aided approach is an elegant solution to both problems since it reduces (1.) and (2.) to verifying a human-readable proof, a task mathematicians and economist alike appear to be comfortable with. By providing such a human-readable proof—even if the encoding and/or the solver was incorrect—we still arrive at a correct result with a verifiable certificate.

As some results in Publication [2] and the main result in Publication [4] exhibit, this shortcut, however, is not always feasible as proofs become too complex to be decoded by humans. In this case, (2.) can usually be answered much more easily than (1.). A simple measure to increase the confidence in (2.), i.e., that the problem encoding is solved correctly, is to simply run a set of different solvers on the same problem. For the case of SAT solving, there are even standardized formats for proof traces of unsatisfiable instances[38] [e.g., DRUP, see Heule et al., 2013] that can—despite being highly non-accessible to humans—be verified by third-party software. For SMT-based results, another potential way out is the technique of proof reconstruction [see, e.g., Böhme and Weber, 2010]: based on the output of a solver one constructs (with machine support) a proof in another language (e.g., higher-order logic (HOL)) that can then be verified by a proof assistant (e.g., the Isabelle system). In fact, supported by the Isabelle system, Eberl [2016] was able, not only to verify the results of Publication [4], but also to semi-automatically produce a very complex, but in principle human-readable proof of the main result. This way, also for Publication [4] the need to verify (1.), the correctness of the encoding, has been lifted.

38 For satisfiable instances, a satisfying assignment is an efficiently verifiable certificate.

But even in the absence of a human-readable proof, there are still a few methods that one can apply to gain confidence in (1.). Whereas verification of the tailor-made encoding software is mostly a theoretical possibility, we found a combination of manual code review and a special kind of testing most effective. The latter is based on encoding *known* variants of the result to see whether the solver behaves in the expected way: in the case of Publication [4] (when Eberl's proof had not yet been developed), we, for instance, reproduced in the very same framework the results by Bogomolnaia and Moulin [2001] and Katta and Sethuraman [2006], as well as the corresponding possibility result for $m < 4$. Previously, similar checks had also been carried out for Publications [1] and [2] before the technique of proof extraction was available. Code review, on the other hand, can be applied against the encoding software (mostly eliminating typos), but more importantly also against the encoding itself (if it has been suitably annotated by the encoding software). The full encoding will usually be too large to be verified by hand, but an MUS can often be of manageable size (e.g., 16 and 94 clauses for the main results in Publications [1] and [4], respectively).

8.3 LIMITATIONS OF THE APPROACH

The vision that Tang [2010] had when he invented the basis for the methods applied in this thesis was computer-aided *theorem discovery*, which in his words includes two aspects: "to come up with reasonable conjectures automatically" and "to prove or disprove the conjectures automatically". Our previous work on the relatively simple subject matter of set extensions deals with both these components [Geist and Endriss, 2011]. For the more complex settings in this thesis, however, we had to concentrate on the latter aspect and come up with reasonable conjectures manually. Based on this experience, we believe that this is the main role that computer-aided methods will play in the near future. Key for successful application will then be close collaboration between subject matter experts (who formulate the questions and provide theoretical tools) and experts on the method (who answer the questions with the help of machine support). This enables not only quick testing of conjectures, but also helps exploring similar statements as well as limits of the hypotheses. When applied interactively, such collaboration might even guide the search for new results in cases where the conjectures are not clearly formulated yet, for instance by quickly providing counterexamples to some ideas.[39]

39 This idea also manifests in the counterexample generator NITPICK [Blanchette and Nipkow, 2010], which is part of the ISABELLE system. Unfortunately, its high flexibility implied that its performance did not suffice for our special purposes (cf. Publication [1]).

Regarding the types of theorems that can be proven with the presented approach, there neither is an obvious classification nor are there strict limiting factors that are easily recognizable. As a good proxy of what can be treated with the approach, one might be tempted to look at the degree to which infinite domains are concerned. But even this is no strict criterion as we show in Publication [4], in which two infinite domains are overcome not only by theoretical lemmas, but also by more expressive solving frameworks (cf. Section 4.1).

Based on the scope of existing results, proving impossibilities rather than positive results, on the other hand, appears to be at least somewhat of a restriction. Interestingly, however, this is a restriction not of the finite computer-aided part, but of the inductive lemmas extending finite results to full generality. While it certainly is possible to prove positive *finite* results with the presented methods (see, e.g., Publications [1], [3], and [5]),[40] it remains highly unclear how an inductive step could be proven that extends, say a characterization result from a finite to arbitrary numbers of alternatives.

40 For instance, for a characterization result one shows that (i) a given function is a model for the axioms (SAT), and (ii) that the axioms are incompatible with an additional constraint saying that for at least one profile one must deviate from the given function (UNSAT).

9 OUTLOOK AND FUTURE WORK

In this thesis and beyond, the application of computer-aided methods has lead to new insights for a range of questions in social choice theory that are of independent interest to the social choice community and unlikely to have been found without the help of computers. Apart from their main application in computer-aided theorem proving, we have demonstrated the usefulness of computer-aided methods also for computational purposes (e.g., computing counterexamples) and for some practical concerns of social choice.

Given the universality of the presented methods and their ease of adaptation (e.g., "testing" of similar conjectures with minimal effort by replacing or altering some axioms), we anticipate these and similar techniques to yield further insights and solve other open problems in social choice theory and related research areas in the future. The breadth of results obtained so far supports this hypothesis.

A few concrete ideas for future work regarding specific problems from social choice have already been mentioned in the corresponding publications. Therefore, here we want to concentrate on more general future challenges with respect to applying and further developing the computer-aided methods of this project. We list those ideas first which we believe to be easier to achieve.

APPLICATIONS IN OTHER AREAS OF ECONOMIC THEORY As indicated in the introduction already, Kerber et al. [2015] have started to apply mechanized reasoning to the domain of auctions [Caminati et al., 2015], an area in which there appears to be much more potential for such methods. But there are more areas of economic theory in which we see value for the presented computer-aided methods.

The domain of assignment is a particularly promising example: not only has it remained mostly untouched by these methods so far, but also can it be viewed as a subdomain of social choice (cf. Publication [4]). Hence, modeling problems from the assignment domain should—despite some novel axioms that only make sense in assignment, such as envy-freeness—not pose major challenges. Solving these problems, however, could turn out to be harder since, at least with the naïve translation, the number of alternatives increases super-exponentially when moving from assignment to social choice.

Beyond these two domains, for example,

- cooperative and non-cooperative game theory (cf. the contribution by Tang and Lin [2011]),

- judgement aggregation (with its inherent link to propositional logic, the language of SAT solvers), and
- argumentation theory (see, e.g., the work by R. Booth and Rahwan [2014], in which three-valued logics are applied to directed graphs, reminiscent of tournament solutions),

offer settings in which computer-aided methods could be promising tools for solving open problems.

TOOL-SUPPORT FOR RESEARCHERS On the practical side, it will probably be a key success factor for wider acceptance of the presented methods to have user-friendly tools in place that can facilitate the encoding and solving process. We discussed the limitations of such tools in Section 8.1. Yet we believe that support in the form of, e.g., automatic encoders with a richer input language than SAT/SMT or experimenter tools (based on already formalized settings), could be achievable.

LOGIC-BASED CLASSIFICATION AND GENERAL INDUCTION STEPS For all future applications of the described methods, it would be desirable to have an (as simple as possible) characterization in logical terms of which problems can be treated with these methods. If this characterization could be complemented with general inductive theorems which reduces any problem of a certain logical form to a finite instance, the door to automated theorem discovery would be wide open again. Our previous work on preference extensions [Geist and Endriss, 2011] marks a first small step in this direction, but so far it is entirely unclear how such results can be obtained for more complex settings, such as the ones considered in this thesis.

BIBLIOGRAPHY

H. Aziz, F. Brandt, and M. Brill. On the tradeoff between economic efficiency and strategyproofness in randomized social choice. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 455–462. IFAAMAS, 2013.

H. Aziz, F. Brandl, and F. Brandt. On the incompatibility of efficiency and strategyproofness in randomized social choice. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pages 545–551. AAAI Press, 2014.

H. Aziz, F. Brandl, and F. Brandt. Universal Pareto dominance and welfare for plausible utility functions. *Journal of Mathematical Economics*, 60:123–133, 2015.

C. Barrett, M. Deters, L. de Moura, A. Oliveras, and A. Stump. 6 years of SMT-COMP. *Journal of Automated Reasoning*, 50(3):243–277, 2013.

J. Bartholdi, III, C. A. Tovey, and M. A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6 (3):227–241, 1989.

A. Biere, M. Heule, H. van Maaren, and T. Walsh, editors. *Handbook of Satisfiability*, volume 185 of *Frontiers in Artificial Intelligence and Applications*. IOS Press, 2009.

J. C. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In *Proceedings of the First International Conference on Interactive Theorem Proving*, pages 131–146. Springer, 2010.

A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328, 2001.

A. Bogomolnaia, H. Moulin, and R. Stong. Collective choice under dichotomous preferences. *Journal of Economic Theory*, 122(2):165–184, 2005.

S. Böhme and T. Weber. Fast LCF-style proof reconstruction for Z3. In M. Kaufmann and L. C. Paulson, editors, *Proceedings of the First International Conference on Interactive Theorem Proving (ITP 2010)*, pages 179–194. Springer-Verlag, 2010.

G. Bordes. On the possibility of reasonable consistent majoritarian choice: Some positive results. *Journal of Economic Theory*, 31(1):122–132, 1983.

F. Brandl, F. Brandt, and H. G. Seedig. Consistent probabilistic social choice. *Econometrica*, 84(5):1839–1880, 2016a.

F. Brandl, F. Brandt, and W. Suksompong. The impossibility of extending random dictatorship to weak preferences. *Economics Letters*, 141: 44–47, 2016b.

F. Brandt. Set-monotonicity implies Kelly-strategyproofness. *Social Choice and Welfare*, 45(4):793–804, 2015.

F. Brandt and M. Brill. Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, pages 136–142. ACM Press, 2011.

F. Brandt and F. Fischer. Computing the minimal covering set. *Mathematical Social Sciences*, 56(2):254–268, 2008.

F. Brandt, M. Brill, and H. G. Seedig. On the fixed-parameter tractability of composition-consistent tournament solutions. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 85–90. AAAI Press, 2011.

F. Brandt, P. Harrenstein, K. Kardel, and H. G. Seedig. It only takes a few: On the hardness of voting with a constant number of agents. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 375–382. IFAAMAS, 2013.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016a.

F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia. Introduction to computational social choice. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 1. Cambridge University Press, 2016b.

M. B. Caminati, M. Kerber, C. Lange, and C. Rowat. Sound auction specification and implementation. In *Proceedings of the 16th ACM Conference on Economics and Computation (ACM-EC)*, pages 547–564. ACM Press, 2015.

S. Chatterjee and A. Sen. Automated reasoning in social choice theory — some remarks. *Mathematics in Computer Science*, 8(1):5–10, 2014.

S. Chatterji, A. Sen, and H. Zeng. Random dictatorship domains. *Games and Economic Behavior*, 86:212–236, 2014.

G. Cinà and U. Endriss. A syntactic proof of Arrow's theorem in a modal logic of social choice functions. In *Proceedings of the 14th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 1009–1017. IFAAMAS, 2015.

V. Conitzer and T. Sandholm. Complexity of mechanism design. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 103–110, 2002.

O. Dowlen. Sorting out sortition: A perspective on the random selection of political officers. *Political Studies*, 57(2):298–315, 2009.

J. Drummond, A. Perrault, and F. Bacchus. SAT is an effective and complete method for solving stable matching problems with couples. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 518–525. AAAI Press, 2015.

J. Duggan. Uncovered sets. *Social Choice and Welfare*, 41(3):489–535, 2013.

J. Duggan and T. Schwartz. Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare*, 17(1):85–93, 2000.

M. Dummett and R. Farquharson. Stability in voting. *Econometrica*, 29(1):33–43, 1961.

B. Dutta and J.-F. Laslier. Comparison functions and choice correspondences. *Social Choice and Welfare*, 16(4):513–532, 1999.

M. Eberl. A formal proof of the incompatibility of SD-efficiency and SD-strategy-proofness. Bachelor's thesis, Technische Universität München, 2016.

C. Eggermont, C. Hurkens, and G. J. Woeginger. Realizing small tournaments through few permutations. *Acta Cybernetica*, 21(2):267–271, 2013.

L. Ehlers, H. Peters, and T. Storcken. Strategy-proof probabilistic decision schemes for one-dimensional single-peaked preferences. *Journal of Economic Theory*, 105(2):408–434, 2002.

E. Elkind, J. Lang, and A. Saffidine. Choosing collectively optimal sets of alternatives based on the condorcet criterion. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 186–191. AAAI Press, 2011.

A. Feldman. Manipulation and the Pareto rule. *Journal of Economic Theory*, 21:473–482, 1979.

P. C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, 3(1):18–40, 1972a.

P. C. Fishburn. Lotteries and social choices. *Journal of Economic Theory*, 5(2):189–207, 1972b.

A. Fréchette, N. Newman, and K. Leyton-Brown. Solving the station repacking problem. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2016.

M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub. Answer set solving in practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(3):1–238, 2012.

C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

A. Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587–601, 1973.

A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

B. Goodwin. *Justice by Lottery*. University of Chicago Press, 2005. 1st edition 1992.

U. Grandi and U. Endriss. First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 42(4):595–618, 2013.

J. W. Headlam. *Election by Lot at Athens*. Cambridge University Press, 1933.

M. J. H. Heule, W.A. Hunt, Jr., and N. Wetzler. Trimming while checking clausal proofs. In *Proceedings of the 13th Conference on Formal Methods in Computer-Aided Design (FMCAD)*, pages 181–188, 2013.

S. Hoelldobler and V. H. Nguyen. On SAT-encodings of the at-most-one constraint. In G. Katsirelos and C.-G. Quimper, editors, *Proceedings of the the 12th International Workshop on Constraint Modelling and Reformulation (ModRef 2013)*, pages 1–17, 2013.

M. D. Intriligator. A probabilistic model of social choice. *Review of Economic Studies*, 40(4):553–560, 1973.

A.-K. Katta and J. Sethuraman. A solution to the random assignment problem on the full preference domain. *Journal of Economic Theory*, 131(1):231–250, 2006.

J. S. Kelly. Strategy-proofness and social choice functions without single-valuedness. *Econometrica*, 45(2):439–446, 1977.

M. Kerber, C. Lange, and C. Rowat. An introduction to mechanized reasoning. Working paper, 2015.

G. Laffond, J. Lainé, and J.-F. Laslier. Composition-consistent tournament solutions and social choice functions. *Social Choice and Welfare*, 13(1):75–93, 1996.

C. Lange, M. B. Caminati, M. Kerber, T. Mossakowski, C. Rowat, M. Wenzel, and W. Wintersteiger. A qualitative comparison of the suitability of four theorem provers for basic auction theory. In J. Carette, D. Aspinall, C. Lange, P. Sojka, and W. Windsteiger, editors, *Intelligent Computer Mathematics*, volume 7961 of *Lecture Notes in Computer Science (LNCS)*, pages 200–215. Springer-Verlag, 2013.

F. Lin and P. Tang. Computer-aided proofs of Arrow's and other impossibility theorems. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2008.

B. McKay and A. Piperno. *nauty and traces User's Guide*, 2.5 edition, 2013.

R. D. McKelvey. Covering, dominance, and institution-free properties of social choice. *American Journal of Political Science*, 30(2):283–314, 1986.

T. Mennle and S. Seuken. The Pareto frontier for random mechanisms. Mimeo, 2015.

H. Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

H. Moulin. *The Strategy of Social Choice*. North-Holland, 1983.

H. Moulin. Condorcet's principle implies the no show paradox. *Journal of Economic Theory*, 45:53–64, 1988.

T. Nipkow. Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automatated Reasoning*, 43:289–304, 2009.

T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL – A Proof Assistant for Higher-Order Logic*, volume 2283 of *Lecture Notes in Computer Science (LNCS)*. Springer-Verlag, 2002.

D. Peters. Personal communication, 2016.

E. Awad R. Booth and I. Rahwan. Interval methods for judgment aggregation in argumentation. In *Proceedings of the 5th International Workshop on Computational Social Choice (COMSOC)*, 2014.

J. Rothe, editor. *Economics and Computation: An Introduction to Algorithmic Game Theory, Computational Social Choice, and Fair Division*. Springer, 2015.

M. R. Sanver and W. S. Zwicker. Monotonicity properties and their adaption to irresolute social choice rules. *Social Choice and Welfare*, 39(2–3):371–398, 2012.

M. A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

D. Shepardson and C. A. Tovey. Smallest tournament not realizable by $\frac{2}{3}$-majority voting. *Social Choice and Welfare*, 33(3):495–503, 2009.

R. Stearns. The voting problem. *American Mathematical Monthly*, 66 (9):761–763, 1959.

P. Stone. *The Luck of the Draw: The Role of Lotteries in Decision Making*. Oxford University Press, 2011.

P. Tang. *Computer-aided Theorem Discovery - A New Adventure and its Application to Economic Theory*. PhD thesis, The Hong Kong University of Science and Technology (HKUST), 2010.

P. Tang and F. Lin. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence*, 173(11):1041–1053, 2009.

P. Tang and F. Lin. Discovering theorems in game theory: Two-person games with unique pure nash equilibrium payoffs. *Artificial Intelligence*, 175(14–15):2010–2020, 2011.

A. D. Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.

G. S. Tseitin. On the complexity of derivation in propositional calculus. In *Automation of Reasoning*, pages 466–483. Springer, 1983.

R. Zeckhauser. Majority rule with lotteries on alternatives. *Quarterly Journal of Economics*, 83(4):696–703, 1969.

L. Zhou. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135, 1990.