Studies in Choice and Welfare

Sascha Kurz
Nicola Maaser
Alexander Mayer *Editors*

Advances in Collective Decision Making

Interdisciplinary Perspectives for the 21st Century

Springer

Studies in Choice and Welfare

Studies in Choice and Welfare is a book series dedicated to the ethical and positive aspects of welfare economics and choice theory. Topics comprise individual choice and preference theory, social choice and voting theory (normative, positive and strategic sides) as well as all aspects of welfare theory (Pareto optimality; welfare criteria; fairness, justice and equity; externalities; public goods; optimal taxation; incentives in public decision making; cost-benefit analysis, etc.). All titles in the book series are peer-reviewed.

Sascha Kurz · Nicola Maaser · Alexander Mayer
Editors

Advances in Collective Decision Making

Interdisciplinary Perspectives for the 21st Century

Springer

Editors
Sascha Kurz
Institute of Mathematics
University of Bayreuth
Bayreuth, Germany

Alexander Mayer
Department of Economics
University of Bayreuth
Bayreuth, Germany

Nicola Maaser
Department of Economics and Business Economics
Aarhus University
Aarhus, Denmark

ISSN 1614-0311 ISSN 2197-8530 (electronic)
Studies in Choice and Welfare
ISBN 978-3-031-21695-4 ISBN 978-3-031-21696-1 (eBook)
https://doi.org/10.1007/978-3-031-21696-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Stefan Napel
on the occasion of his 50th birthday.

Preface

Collective decisions affect our economic, political, and social lives almost daily. They are not only important in large-scale elections, but also in small groups such as management boards, government cabinets, panels of judges, expert committees, job hiring committees, and multinational organizations. Applications range widely, from analyzing the complicated institutional rules employed by the European Union to the responsibility-based distribution of cartel damages or the design of webpage rankings.

Over the past 20 years, Stefan Napel has helped to improve our understanding of the interdisciplinary links of collective decision-making. He has published more than 50 papers, many of them in leading academic journals. His keen mind and pursuit of new knowledge have been an inspiration for all of us. But most of all, we are pleased that over the years he has become more than a co-author and an academic teacher: We are proud to call him our friend. His 50th birthday provides a perfect opportunity to review recent contributions to the long history of collective decision-making research, to highlight the interdisciplinary aspect of the discipline, and to look ahead to its promising future by pointing to unanswered questions that can only be resolved through collaborative efforts. This intention is also summarized by the volume's subtitle *Interdisciplinary Perspectives for the 21st Century:* a comprehensive look at current research by economists, mathematicians, computer scientists, philosophers, and political scientists on the design and implications of collective decisions.

Bayreuth, Germany	Sascha Kurz
Aarhus, Denmark	Nicola Maaser
Bayreuth, Germany	Alexander Mayer
July 2022	

Acknowledgements We are very grateful to Federico Valenciano for drawing the portrait of Stefan Napel and to all the authors and referees who have contributed to this book. We also thank the editors-in-chief, Marc Fleurbaey and Maurice Salles, for the opportunity to publish this volume in the *Studies in Choice and Welfare* series and our editors and project coordinators at Springer, Johannes Glaeser and Vijay Kumar Selvaraj, for their encouragement and advice.

Contents

Contributors

Maher Jakob Abou Zeid European Commission, Brussels, Belgium

Federica Alberti University of Portsmouth, Portsmouth, UK

Fuad Aleskerov HSE University, Moscow, Russia;
Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

Encarnación Algaba Matemática Aplicada II and Instituto de Matemáticas de la Universidad de Sevilla, Sevilla, Spain

Ahmad Awde University of Bourgogne Franche-Comté, Besançon, France

Markus Brill University of Warwick, Coventry, UK

André Casajus Leipzig Graduate School of Management, Leipzig, Germany;
Dr. Hops Craft Beer Bar, Leipzig, Germany

Vincent Conitzer Carnegie Mellon University, Pittsburgh, USA;
University of Oxford, Oxford, UK

Mostapha Diss University of Bourgogne Franche-Comté, Besançon, France;
University Mohamed VI Polytechnic, Africa Institute for Research in Economics and Social Sciences (AIRESS), Rabat, Morocco

Josep Freixas Departament de Matemàtiques i Escola Politècnica Superior d'Enginyeria de Manresa, Universitat Politècnica de Catalunya, Manresa, Spain

Werner Güth Max Planck Institute for Research on Collective Goods, Bonn, Germany

Herbert Hamers CentER, Department of Econometrics and Operations Research and TIAS Business School, Tilburg University, Tilburg, The Netherlands

Manfred J. Holler University of Hamburg, Germany and CCR-Munich, Hamburg, Germany

Madeleine O. Hosli Leiden University, Leiden, The Netherlands

Frank Huettner SKK Graduate School of Business, Sungkyunkwan University, Seoul, Korea

Alexander Ivanov HSE University, Moscow, Russia

Eric Kamwa Université des Antilles, Schoelcher Cedex, Martinique, France

Serguei Kaniovski Austrian Institute of Economic Research (WIFO), Vienna, Austria

Daniel Karabekyan HSE University, Moscow, Russia

Dominik Karos Center for Mathematical Economics, Bielefeld, Germany

Werner Kirsch Fakultät für Mathematik und Informatik, FernUniversität Hagen, Hagen, Germany

Christian Klamler Institute of Public Economics, University of Graz, Graz, Austria

Hartmut Kliemt Justus-Liebig-Universität Giessen, Giessen, Germany

Sascha Kurz Department of Mathematics, University of Bayreuth, Bayreuth, Germany

Jo Leinen European Parliament, Strasbourg, France

Nicola Maaser Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark

Alexander Mayer Department of Economics, University of Bayreuth, Bayreuth, Germany

Maria Montero School of Economics, University of Nottingham, Nottingham, UK

Hannu Nurmi Department of Philosophy, Contemporary History and Political Science, University of Turku, Turku, Finland

Montserrat Pons Departament de Matemàtiques i Escola Politècnica Superior d'Enginyeria de Manresa, Universitat Politècnica de Catalunya, Manresa, Spain

Arash Pourebrahimi Leiden University, Leiden, The Netherlands

Andrea Prieto Matemática Aplicada II, Escuela Superior de Ingenieros, Camino de los Descubrimientos, Sevilla, Spain

Friedrich Pukelsheim Institute for Mathematics, University of Augsburg, Augsburg, Germany

Clemens Puppe Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany;
Higher School of Economics (HSE), Perm Campus, Moscow, Russian Federation;
Balliol College, Oxford University, Oxford, UK

Julien Yves Rolland University of Bourgogne Franche-Comté, Besançon, France

Jana Rollmann Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany

Olivier Roy Department of Philosophy, University of Bayreuth, Bayreuth, Germany

Florian Rupp Kutaisi International University, Georgia and CCR-Munich, Kutaisi, Georgia

Alejandro Saavedra-Nieves Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

David Stadelmann University of Bayreuth, Bayreuth, Germany

Abdelmonaim Tlidi University of Abdelmalek Essaâdi, FSJES of Tetouan, Tetouan, Morocco

Kei Tsutsui University of Bath, Bath, UK

Marina Uzunova John Stuart Mill College, Vrije Universiteit, Amsterdam, The Netherlands

René van den Brink School of Business and Economics and Tinbergen Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Gerard van der Laan School of Business and Economics and Tinbergen Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Peter van Roozendaal Netherlands Court of Audit, Leiden, The Netherlands

Valeri Vasil'ev Sobolev Institute of Mathematics, Novosibirsk, Russia

Vyacheslav Yakuba HSE University, Moscow, Russia;
Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

Alexander Zaigraev Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

Introduction

Sascha Kurz, Nicola Maaser, and Alexander Mayer

After long being dominated by the social sciences, the analysis of collective decision-making has evolved into a thoroughly multidisciplinary field of research. It now attracts not only interest in economics, political science, and psychology, but also in applied mathematics, computer science, philosophy, and even biology.

One important strand of literature is rooted in modern social choice theory, a field pioneered by Duncan Black and Nobel laureates Kenneth Arrow and Amartya Sen (see, e.g., Arrow et al., 2002, 2011). This tradition is broadly concerned with the aggregation of individual preferences and information into group decisions that realize some exogenous social goal such as welfare maximization or a notion of fairness. It has developed dynamic offshoots such as, e.g., computational social choice, which applies techniques originally developed in computer science to social choice mechanisms (such as determining the computational complexity of manipulation in elections) and concepts from social choice to computing (such as the design of webpage rankings). For an accessible and comprehensive overview of the rapidly growing field of computational social choice, see Brandt et al. (2016). In contrast to the classical economic view that markets are merely a device to coordinate and allocate the resources of heterogeneous agents based on their individual (and usually diverging) preferences, methods of collective decision-making focus on coordination mechanisms that do not follow a market logic to define social objectives based

S. Kurz
Department of Mathematics, University of Bayreuth, Bayreuth, Germany
e-mail: sascha.kurz@uni-bayreuth.de

N. Maaser (✉)
Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark
e-mail: nmaaser@econ.au.dk

A. Mayer
Department of Economics, University of Bayreuth, Bayreuth, Germany
e-mail: alexander.mayer@uni-bayreuth.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_1

on individual preferences. Typical topics include, but are not limited to, the analysis of voting procedures, fair division, matching problems, and more recently, modern recommender systems, interactive democracy, and deliberation.

A second strand of literature, pursued in particular by political scientists and economists, relates collective choices to the incentives of the individuals. One way to do so is to use bargaining models that apply non-cooperative game theory. This approach allows for the inclusion of institutional details to generate empirically testable hypotheses about, e.g., legislative and parliamentary policy outcomes from a comparative perspective. Approaches from organization theory allow predictions about how the structure of a group affects its decision-making performance. For example, are ethnically diverse bureaucracies better at public service delivery? Is there an optimal committee size to achieve well-informed collective decisions?

The two traditions described above increasingly intertwine, taking on board ideas from other disciplines such as philosophy and evolutionary biology (see, e.g., Seeley 2010, for a study on collective decision-making and "voting" among honeybees). Moreover, many topics that are best addressed with the tools and perspectives of multiple disciplines have only recently received greater attention from the scientific community (e.g., multi-winner elections, interactive democracy, big data). We see each of the above as examples of timely topics holding new challenges that require interdisciplinary efforts, using foundational, methodological, empirical, and experimental tools and approaches.

This volume is a collection of twenty-one invited peer-reviewed contributions from researchers across disciplines. We have grouped the articles into five parts, but our choice of taxonomy should not disguise the often very large overlap between these parts.

Part I, *Social Choice*, is devoted to the underpinnings and recent developments in social choice theory. Hannu Nurmi sets the scene with his chapter "Building Bridges Over the Great Divide". He discusses some methods aiming to reconcile Borda's and Condorcet's winning intuitions in the theory of voting. He begins with a brief summary of the advantages and disadvantages of binary and positional voting rules. He then reviews in some detail Black's, Nanson's, and Dodgson's rules as well as relatively recently introduced methods based on the super-covering relation over the candidate set. These are evaluated in terms of some well-known choice-theoretic criteria.

In the chapter "Social Unacceptability for Simple Voting Procedures", Ahmad Awde, Mostapha Diss, Eric Kamwa, Julien Yves Rolland, and Abdelmonaim Tlidi investigate the notion of "socially unacceptable" candidates from a computational perspective. Since the existence of socially unacceptable candidates is not always guaranteed, they determine the probability that such candidates exist given the number of running candidates and the size of the electorate. Moreover, they evaluate how often prominent voting procedures can elect a socially unacceptable candidate.

In "Probability of Majority Inversion with Three States and Interval Preferences", Serguei Kaniovski and Alexander Zaigraev examine the probability of majority inversion in a two-stage electoral process with three states. They extend May's model to state-specific general interval preferences and population weights and investigate the effects of variation in population weights and the effects of variance and bias in preferences on the inversion probability. The numerical sensitivity analysis for the probability of inversion is conducted using an exact formula under general interval preferences in the three states with different population weights.

Markus Brill and Vincent Conitzer study the incentives of candidates to run in an election in the chapter "Strategic Voting and Strategic Candidacy". Most work on this topic assumes that strategizing only takes place among candidates, whereas voters vote truthfully. Markus Brill and Vincent Conitzer extend the analysis to include strategic behavior on the part of the voters. They also study cases where only candidates or only voters are strategic.

In the chapter "Meta-Agreement and Rational Single-Peaked Preferences", Olivier Roy and Maher Jakob Abou Zeid revisit the claim that rationality requires participants in deliberation to form single-peaked preferences once they have reached meta-agreements. They provide two different arguments that cast doubts on this claim and show that to the extent that deliberation fosters the formation of meta-agreements and single-peaked preferences, the bridge between these two notions might not be solely a matter of rational preference formation.

In "On the Individual and Coalitional Manipulability of q-Paretian Social Choice Rules", Fuad Aleskerov, Alexander Ivanov, Daniel Karabekyan, and Vyacheslav Yakuba study the degree of individual and coalitional manipulability of q-Paretian social rules. For the cases of three, four, and five alternatives and for the agent numbers ranging from three to one hundred, they use computer modeling to calculate various well-known manipulability indices.

Part II, *Weighted Voting*, comprises four chapters that deal with weighted voting decisions. In the chapter "Efficiency, Decisiveness, and Success in Weighted Voting Systems: Collective Behavior and Voting Measures", Werner Kirsch defines and investigates a large class of voting measures with respect to their efficiency and success in weighted voting systems. This class can be characterized as those voting measures which are invariant under permuting the voters and which allow a natural extension to an arbitrary number of voters. The class includes the prominent Penrose-Banzhaf and Shapley-Shubik measures.

The chapter of Josep Freixas and Montserrat Pons, "All Power Structures are Achievable in Basic Weighted Games", proves that for each achievable hierarchy of power for weighted games there is a handy weighted game fulfilling three desirable properties. A representation of this type is ideal for the design of a weighted game with a given hierarchy. Moreover, they show that the subclass of weighted games with these properties is considerably smaller than the class of weighted games.

In the chapter "Bargaining in Legislatures: A New Donation Paradox", Maria Montero considers a model of legislative bargaining and shows that it is possible for a player to donate part of its proposing probability to another player and be better off as a result. Thus, even though actually being selected to propose is always valuable ex post, having a higher probability of being proposer may be harmful in equilibrium.

In their chapter "Egalitarian Collective Decisions as Good Corporate Governance", Federica Alberti, Werner Güth, Hartmut Kliemt, and Kei Tsutsui translate substantive normative premises of stakeholder value approaches into operational axioms that characterize a class of collective decision mechanisms. They challenge stakeholder theorists and critics of shareholder value approaches, who may find the implied characterization unattractive, to come up with alternative collective decision mechanisms or a modified set of values.

Part III, *Interpretation and Measurement of Power*, revisits some classical power indices and applies them in very different scenarios. It starts with the chapter by Frank Huettner and Dominik Karos "Liability Situations with Successive Tortfeasors". They consider successive torts—i.e., torts that involve a causality chain—and show that simple and intuitive principles, which are well known in tort law, uniquely define an allocation scheme that reflects tortfeasors' responsibility. They show that this scheme incentivizes agents to exhibit a certain level of care, creating an efficient prevention of accidents. Then, Huettner and Karos describe the unique rule according to which a liability situation has to be adjusted after a partial settlement such that incentives to settle early are created.

In the chapter "Solidarity and Fair Taxation in TU Games", André Casajus considers an analytic formulation of the class of efficient, linear, and symmetric values for TU games that rests on the linear representation of TU games by unanimity games. Unlike most of the other formulae for this class, his formula allows for an economic interpretation in terms of taxing the Shapley payoffs of unanimity games. He identifies those parameters for which the values behave in an economically sound way. That is, he indicates requirements on fair taxation in TU games by which solidarity among players is expressed.

Encarnación Algaba, Andrea Prieto, Alejandro Saavedra-Nieves, and Herbert Hamers introduce in their chapter "Analyzing the Zerkani Network with the Owen Value" a new centrality measure based on the Owen value to rank members in covert networks. In particular, they consider the Zerkani network responsible for the Paris attacks of November 2015 and the Brussels attack of March 2016. They consider two different appropriate cooperative games defined on the Zerkani network and calculate for each game the Owen value. This provides a ranking of the members in the Zerkani network.

In "The Power of Closeness in a Network", Manfred Holler and Florian Rupp consider the question of whether it is profitable for a weaker player to be closely linked to a strong (i.e., powerful) player and whether it is more beneficial to a powerful player to be closely linked with a weak player than with a strong player. They demonstrate that power (as based on the Public Good index) is a non-local concept indicating that strong players form a "hot-region" about the strongest player. They present an easy-to-perform algorithm for the computer-based determination of the Public Good index on networks that allows to study voting power in small networks.

In the chapter "Political Power on a Line Graph", René van den Brink, Gerard van der Laan, Marina Uzunova, and Valeri Vasil'ev consider situations of majority voting, where the players are ordered on a line graph based, e.g., on ideological or political preferences over various policy dimensions. Motivated by the observation that a number of solutions for line graph games (interpreted as power indices) either are not core stable, or do not reward intermediate veto players, they axiomatically characterize two alternative indices that are core stable and reward all veto players.

Part IV, *EU*, consists of four papers that analyze decision-making in the European Union from different perspectives. In "Double Proportionality for the European Parliament: The Tandem System", Jo Leinen and Friedrich Pukelsheim consider the doubly proportional electoral system for the European Parliament. This tandem system offers a forum for europarties to contest an election with power, visibility and influence. They show that the system satisfies the "one person, one vote" principle and the principle of degressive representation. Moreover, it respects the EU's subsidiarity principle.

In "Explaining Contestation: Votes in the Council of the European Union", Arash Pourebrahimi, Peter van Roozendaal, and Madeleine Hosli study voting behavior in the Council of the European Union using more than 1229 legislative decisions taken in the Council from 2010 to 2021. They investigate the impact of different independent variables on member states' voting behavior: net contributions to the EU budget, voting power, left-right policy positions, and finally, the distance of a member state's ideological position from the position of the winning coalition under the qualified majority voting rule.

In their chapter "Codecision in Context Revisited: The Implications of Brexit", Nicola Maaser and Alexander Mayer analyze the implications of the UK's leave from the EU for the distribution of power between the Council of the EU and the European Parliament and within the Council under the EU's codecision procedure. They model the codecision procedure as a bargaining game between the Parliament and the Council under various a priori preference assumptions. They find that the withdrawal of the UK has no significant effect on the power distribution between the Parliament and the Council and that it is mainly the large member states that benefit from the UK's leave.

Finally, Part V, *Field Experiments and Quasi-Experiments*, collects experimental approaches to voting and political decisions. Experimental methods have been rapidly developing in economics and political science over the past few years, especially in the field of electoral behavior. Christian Klamler's chapter "Proximity-Based Preferences and Their Implications Based on Data from the Styrian Parliamentary Elections in 2019" applies a proximity-based approach using exit poll data. He shows that (i) a single-peaked model does not perfectly fit the data, and (ii) declared preferences and proximity-based preferences differ significantly (which indicates that other factors do play a role in determining voters' preferences). Klamler then specifies the actual impact of those other factors on the election results.

In the chapter "Participation in Voting Over Budget Allocations: A Field Experiment", Clemens Puppe and Jana Rollmann study the effect on the participation rate of employing different voting rules (in particular, the mean rule and the median rule). They report the results of a field experiment in which subjects could allocate money to fund two different public projects. Their results shed important light on the use of different voting rules in the context of budget allocation in practice.

In "The Office Makes the Politician", David Stadelmann explores behavioral changes regarding the political representation of voters by leveraging data from the two Houses of Parliament in Switzerland. He provides evidence that is consistent with the existence of an incentive effect of the office itself which acts on politicians to fulfill public expectations. Such an incentive effect, termed a "Thomas Becket incentive", would be complementary to the established relevance of elections as a selection and incentive device.

We hope that this volume can fulfill its ambitious purpose and give a comprehensive overview of the interdisciplinary state-of-the-art research on collective decision-making.

References

Arrow, K. J., Sen, A. K., & Suzumura, K. (2002). *Handbook of social choice and welfare* (Vol. 1). Elsevier.

Arrow, K. J., Sen, A. K., & Suzumura, K. (2011). *Handbook of social choice and welfare* (Vol. 2). Elsevier.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (2016). *Handbook of computational social choice*. Cambridge University Press.

Seeley, T. D. (2010). *Honeybee democracy*. Princeton University Press.

Social Choice

Building Bridges Over the Great Divide

Hannu Nurmi

Abstract We discuss some methods aiming to reconcile Borda's and Condorcet's winning intuitions in the theory of voting. We begin with a brief summary of the advantages and disadvantages of binary and positional voting rules. We then review in some detail Black's, Nanson's and Dodgson's rules as well as the relatively recently introduced methods based on supercovering relation over the candidate set. These are evaluated in terms of some well-known choice–theoretic criteria.

1 Introduction

The theory of voting is largely based on the social choice theory. Much of the theory deals with various norms and their compatibility. When rules are compared with each other, the norms are typically invoked to prioritize one rule over the other. Occasionally, the rules are viewed as instruments for attaining specific goals of the voters. In their recent work, the honoree of this volume together with his collaborators have pursued this strategic approach to rule selection (Kurz et al., 2020; Mayer and Napel, 2020; Kurz et al., 2021). The underlying assumption of this work is that the participants in institution design are interested in maximizing their influence over the decision making outcomes and—to the extent various rules tend to favor actors of different sizes—the actors form their preference over the rules accordingly. Napel and his associates have thus opened a new perspective to voting rule selection, one that combines voting power with social choice. While I commend these authors for their fresh ideas and am quite optimistic about the prospects of the approach, I will here focus on the more traditional way of looking at rule selection problems.

One of the perennial themes—dating back to the later eighteenth century, the "Golden Age of Social Choice" (McLean and Urken, 1995)—in the theory of voting is the discrepancy between positional and binary voting rules. In the former class, the winners are determined on the basis of the positions of alternatives in the preference

H. Nurmi (✉)
Department of Philosophy, Contemporary History and Political Science, University of Turku, Turku, Finland
e-mail: hnurmi@utu.fi

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_2

rankings of voters, whereas in the latter, the performance in pairwise comparisons is of paramount importance in selecting the winners. The basic intuition behind the binary winners is that whichever alternative defeats all the others in pairwise comparisons ought to be the winner. The obvious drawback of this intuition is that it is not always applicable. Over time, various rules were invented, some of them specifically aimed at guaranteeing that a winning set of alternatives always be found while making sure that this set reduces to a singleton consisting of the alternative that beats all others in pairwise comparisons—should such an alternative be found—is elected. Such an alternative is called the Condorcet winner. Accordingly, the rules based on this principle are called Condorcet extensions and the principle itself the Condorcet principle.

The positional rules (a.k.a. scoring rules) are not based on a common overarching principle. Rather, they constitute a heterogeneous class of aggregation methods. The best-known rules in this class are the plurality voting and the Borda count. The former elects the alternative ranked first by more voters than any other alternative, while the latter assigns points to m alternatives so that a voter gives $m - 1$ points to an alternative ranked first, $m - 2$ to an alternative ranked second, etc. The sum of points given by all voters to alternative x give the Borda score of x. The alternative with the largest Borda score is the winner, and the order of magnitude of the scores indicates the Borda ranking. Obviously, the plurality voting utilizes only a very limited amount of the information given by the individual rankings, whereas the Borda count takes into account all preference positions.

The discrepancy between the Condorcet principle and Borda count (and implicitly also between the former and plurality voting) was established very early by Marquis de Condorcet (McLean and Urken, 1995). Over the past few decades, the Condorcet principle has often been elevated to the status of one of the most important social choice desiderata (Riker, 1982; McLean, 1991; Felsenthal and Machover, 1992). This view has by no means been universally adopted. Notably, Saari has presented a strong defense for the Borda count (Saari, 1995, 2003). Both intuitions have virtues and drawbacks some of which will be briefly discussed in the next section. However, our main focus is on rules that aim to reconcile the two intuitions in voting rules by invoking positional information while not losing sight of the Condorcet winner intuition.

2 The Main Pros and Cons of Binary and Positional Rules

The main virtues of binary rules are:

1. Simplicity for the voters: only pairwise comparisons of alternatives are called for.
2. They aim at electing 'obvious' winners: Condorcet winners.
3. When a Condorcet winner exists, it is the winner in all subsets of alternatives as well.

Table 1 Condorcet winner C has the smallest number of first ranks

4 voters	3 voters	2 voters
A	B	D
C	C	C
B	A	B
D	D	A

Table 2 Fishburn's example

1 voter	1 voter	1 voter	1 voter	1 voter	1 voter	1 voter
w	w	w	a	a	a	b
a	a	a	b	b	w	w
b	b	b	w	w	b	a

4. When a Condorcet winner exists, the rules are positively involved (Saari, 1995) or, equivalently, invulnerable to the P-TOP paradox (Felsenthal and Tideman, 2013).[1]

   As defined by Saari (1995, p. 216):

   > A procedure is positively involved if when c_j is the selected outcome from a profile and when a group of voters, all of the same voter type with c_j top-ranked, join the group, c_j remains the selected candidate.

Simplicity, the first virtue of binary rules, is obvious, but begs the question of why the voters are in general supposed to be endowed with complete and transitive preference relations over the alternatives if the information contained in the preference rankings is not utilized in social choices. In fact, all pairwise comparisons could be implemented even in cases of cyclic preference relations. Similarly, the second virtue is debatable. There are situations where the Condorcet winner is intuitively less plausible than a plurality one as shown in Table 1 where the Condorcet winner C is ranked first by no voters, while the plurality winner A is ranked first by nearly a half of the electorate. In an effort to discredit the Borda count Marquis de Condorcet came up with an example which—in somewhat counterproductive fashion—may be used to undermine the general plausibility of the Condorcet winner. The example is discussed by Black (1958, pp. 176–177). A somewhat simplified version devised by (Fishburn, 1974, p. 544) is reproduced in Table 2.

In Table 2 w is the Condorcet winner, but a seems more plausible on the grounds that it is placed first by equally many voters as w, and second by strictly more voters than w and last by only one voter, while w is ranked last by two voters. Thus, a positionally dominates (weakly) w, the Condorcet winner.

[1] Pérez's invulnerability to the positive strong no-show paradox (Pérez, 2001, p. 602) refers to the same property.

Table 3 Max-min elects the absolute loser D

Profile					Pairwise matrix					
2 voters	3 voters	3 voters	1 voter	2 voters		A	B	C	D	row min.
D	D	C	B	A	A	–	4	8	6	4
A	B	B	A	C	B	7	–	4	6	4
C	A	A	C	B	C	3	7	–	6	3
B	C	D	D	D	D	5	5	5	–	5

The third virtue follows directly from the definition of the Condorcet winner. It is a distinguishing property of Condorcet extensions. The fourth virtue pertains to the invulnerability of the P-TOP paradox. This occurs in a situation where a winning candidate x becomes a non-winner if a group of voters all ranking x first joins the electorate, *ceteris paribus*. When a Condorcet winner exists and is therefore elected in the original profile, it also remains the Condorcet winner in the augmented profile and is by definition elected in the latter profile as well. Hence, the P-TOP paradox cannot inflict the Condorcet extensions in Condorcet domains, i.e., in settings where a Condorcet winner exists. As will be seen shortly, though, this invulnerability does not extend to all Condorcet extensions outside the Condorcet domains (Richelson, 1978, p. 174).

The virtues are counterbalanced by some important flaws:

1. The Condorcet winner may not be first ranked by any voter.
2. There are Condorcet extensions that (in the absence of a Condorcet winner) may elect a Condorcet loser, i.e., a candidate that would be defeated by all other alternatives in pairwise majority comparisons.
3. Some Condorcet extensions may even elect an absolute loser, i.e., a candidate ranked last by at least 50% of the voters.
4. They may behave strangely under adding or removing of Condorcet components.
5. The Condorcet extensions are vulnerable to the no-show paradox (Moulin, 1988).

The first possibility is demonstrated by Table 1. The second and third one are shown in Table 3 where the rankings of 11 voters over four candidates are exhibited. Since in Table 3, D is ranked last by six voters, it is the absolute loser. The pairwise comparison matrix on the right-hand part of the table represents the votes given to the candidate represented by the row when confronted with the candidate represented by the column. The minimum entries on each row are written in the right-most column. The max-min rule elects the candidate with the largest minimum support in pairwise comparisons. In this case, it is D, the absolute loser. Similar example can be constructed for Dodgson's rule, i.e., a method that elects the candidate that—given a preference profile—is closest to being the Condorcet winner with closeness measured by the binary inversion metric (Nurmi, 2004, p. 10).

The fourth flaw refers to the possibility that—starting from a profile with a Condorcet winner—an addition or removal of a set of voters whose preference form a

Condorcet cycle (A beats B, B beats C and C beats A with identical margins) may replace the original Condorcet winner with another one. So, while it is well known that the Borda count and other positional procedures are not robust under adding or removing candidates, the Condorcet extensions are not robust under adding or removing of *sets of voters* with preferences constituting a perfect tie (Saari, 1995; Nurmi, 1999).

A serious drawback of the Condorcet extensions is their vulnerability to the no-show paradox (Moulin, 1988). Although discussed by election practitioners more than a hundred years ago (Meredith, 1913), the paradox was not systematically analyzed until early 1980s by Fishburn and Brams (1983). It is a nonmonotonicity-related property that refers to a counterintuitive response of a procedure to certain types of changes in the electorate. To wit, it may happen that a group of voters with identical preference rankings will bring about their worst outcome by voting, whereas by abstaining, *ceteris paribus*, a preferable outcome (for them) would ensue. This possibility is called the no-show paradox by Fishburn and Brams. It is also known as the P-BOT paradox (Felsenthal and Tideman, 2013) or the negative strong no-show paradox (Pérez, 2001). To quote Fishburn and Brams (1983, p. 207), this paradox occurs when

> [t]he addition of identical ballots with candidate x ranked last may change the winner from another candidate to x.

It should be emphasized that the no-show paradox is a possibility that pertains to electorates that change in size, to be distinguished from the standard monotonicity properties which are related to changes in a fixed electorate.

3 Some Attempts to Reconcile Binary and Positional Intuitions

One would expect that the Condorcet extensions lose much of their practical appeal if the very core concept of them, the Condorcet winner, turns out to be defective as a general solution. Yet, many a voting theorist has over the past decades believed in the tenability of those methods. Hence, many Condorcet-inspired methods have been invented, some of them aiming at building bridges over the binary-positional divide. We now turn to some of these methods.

3.1 Black's Rule

The most straightforward attempt to reconcile Borda' and Condorcet's solutions was proposed by Black in his *magnum opus* (Black, 1958). The rule is simple enough: given a profile of rankings, elect the Condorcet winner if one exists, otherwise elect the Borda winner. It has the two first-mentioned virtues of Condorcet extensions,

Table 4 Black's rule is vulnerable to the P-TOP

3 voters	3 voters	4 voters	3 voters	1 voter
D	E	C	D	E
E	A	D	E	B
A	C	E	B	A
B	B	A	C	D
C	D	B	A	C

Table 5 Black's rule is vulnerable to P-BOT

4 voters	3 voters		1 voter
B	C		A
C	A	+	B
A	B		C

but fails on the last two in case the Condorcet winner does not exist in the profile at hand. This has been discussed under truncated point-total paradox by Fishburn (1974, p. 538) and generalized by Saari (2001, p. 36). This failure is due to there being no Condorcet winner at the outset whereby the Borda count kicks in. Should there be a Condorcet winner, it would, by definition, remain one is all subsets it belongs to. Black's rule is also vulnerable to the P-TOP paradox, again provided that no Condorcet winner exists in the profile under study. This appears to be first discovered by Richelson (1978, p. 174). Table 4 gives an example (Felsenthal and Nurmi, 2017, pp. 74–75). It involves 14 voters and five candidates.[2]

As there is no Condorcet winner in this profile, Black's rule yields the Borda winner, E. Suppose now that this electorate is augmented with two voters with ranking EBADC. In the augmented profile, Black's rule results in the Condorcet winner D. Hence, we have an instance of the P-TOP paradox.

As all Condorcet extensions, also Black's rule is vulnerable to the no-show paradox as defined by Fishburn and Brams. Table 5 demonstrates that Black's rule may lead to an instance of the no-show paradox.[3] In the 7-voter profile on the left-hand side, there is a Condorcet and, hence, a Black winner, viz. B. Suppose now that one voter with the ABC ranking joins the electorate (perhaps eager to improve the chances of his/her first-ranked A *vis-à-vis* his/her next to last ranked B). The ensuing outcome under Black's rule now becomes C, the last-ranked candidate of the entrant voter. So by voting this, voter makes the outcome the worst possible one from his/her point of view.

[2] Two much simpler examples each involving only three candidates and nine voters are presented in Brandt et al. (2022) in the context of reinforcement and no-show (Sect. 4.7).

[3] This example somewhat simplifies the one given in Felsenthal and Nurmi (2017, p. 75). A similar result is obtained in Brandt et al. (2022).

Table 6 Nanson's rule and a Condorcet component

2 voters	1 voter		2 voters	2 voters	2 voters
A	B		A	B	C
B	C	+	C	A	B
C	A		B	C	A

3.2 Nanson's and Baldwin's Rules

About a hundred years after the publication of Condorcet's major work on voting theory (Condorcet, 1785), E. J. Nanson came up with an ingenious proposal to combine Condorcet's and Borda's winner intuitions without switching between two different methods, as Black chose to suggest (Nanson, 1883). Nanson observed a connection between the Condorcet winner and the Borda count: while the Condorcet winner may not always coincide with the candidate with the largest Borda score, its Borda score cannot be the smallest, either. Thus, if one eliminates the alternative with the smallest Borda score, one can rest assured that the Condorcet winner (if one exists) remains among the non-eliminated candidates. Since the Condorcet winner—by definition—defeats all the other candidates, it remains one in all subsets of the original candidate set. Thus, repeating the elimination process guarantees that the eventual Condorcet winner remains among the non-eliminated ones. Furthermore, Nanson noticed that the elimination process can be sped up without jeopardizing the choice of a Condorcet winner. To wit, he noticed that a Condorcet winner can never have an average or smaller Borda score. Hence, by eliminating on each round those candidates with an average or smaller Borda score, one can avoid eliminating an eventual Condorcet winner.

Given a preference profile, Nanson's rule singles out as the winner the candidate that survives the elimination process based on Borda scores. So, in contrast to Black's rule which relies on both Cordorcet's and Borda's intuitions, Nanson's rule resorts to only one principle, the Borda count. As Borda noted in his presentation in late 18th century (de Borda, 1781), his method can be implemented via pairwise comparisons of candidates. Thus, Nanson's rule is equally simple for the voter as the other pairwise comparison methods. The crucial difference is that in the determination of Borda results the *size* of support for candidates in pairwise comparisons is essential, whereas in most other pairwise methods the essential information is whether or not a candidate defeats another candidate in a pairwise contest. By design, Nanson's rule also elects the Condorcet winner when one exists.

Even though the Borda scores of the candidates in a Condorcet component are identical, adding or removing such a Condorcet component can change the outcome under Nanson's rule. This is shown in Table 6.

In the 3-voter profile on the left, there is a (strong) Condorcet and, therefore, a Nanson winner, A (after C and then B have been eliminated). The 6- voter profile on the right constitutes a Condorcet component where each candidate is ranked first,

Table 7 Baldwin versus Nanson

3 voters	3 voters	1 voter
A	C	B
B	A	C
C	B	A

second and third by equally many voters. Yet, in the combined 9-voter profile, B emerges as the Nanson winner (after C and B have been eliminated).) Nanson's rule is vulnerable to both P-TOP and P-BOT paradoxes (Felsenthal and Nurmi, 2017, pp. 58–60).[4]

Baldwin's method was apparently intended as a practical modification of Nanson's rule (Baldwin, 1926). It is also a Borda elimination rule, but instead of eliminating on each round all candidates with the average or smaller Borda score, Baldwin's rule eliminates the candidate(s) with the minimum Borda score.[5] Thus computing the winner takes in general more—often far more—eliminations rounds. The feature that Nanson regarded most important, viz. that the eventual Condorcet winner not be eliminated, characterizes Baldwin's rule as well since obviously the Condorcet winner cannot have the smallest Borda score at any stage of the process. Baldwin's and Nanson's are, however, distinct rules as can be seen in Table 7. As A is the only candidate with strictly larger than average Borda score, it is elected under Nanson's rule. By the same token, B with the smallest average is eliminated under Baldwin's rule, whereupon C beats A and gets elected using the latter rule.

While not always ending up with the same outcomes in all profiles, Baldwin's and Nanson's rules by and large satisfy the same desiderata of social choice rules.

3.3 Dodgson's Rule

Dodgson's rule emphasizes the plausibility of the Condorcet winner as a solution concept to the degree that it determines the winner in terms of 'closeness' to the Condorcet winner: whichever candidate needs the smallest number of preference changes in the electorate to become the Condorcet winner is the winner (Black, 1958, pp. 222–234).[6] It is the determination of closeness that invokes positional information

[4] Brandt et al. provide simpler examples of these and other paradoxes (Brandt et al., 2022).

[5] Perhaps a better-known term for Baldwin's rule is Borda elimination rule (BER)(Smaoui et al., 2016).

[6] Associating C. L. Dodgson (a.k.a. Lewis Carrol) with the rule described in this subsection has plausibly been called into question by, e.g., Fishburn (1977, pp. 474–475). Indeed, what is known as the Dodgson rule is just one of several proposed by him (Black, 1958). Also Tideman has doubts about the plausibility of associating Dodgson with this rule (Tideman, 1987). See Brandt (2009). Keeping these caveats in mind we shall, however, conform to the standard usage of the concept of Dodgson's rule.

Table 8 Dodgson winner is not the winner is all subsets of candidates

10 voters	7 voters	1 voter	7 voters	4 voters
D	B	B	C	D
A	C	A	A	C
B	A	C	B	A
C	D	D	D	B

in computing the Dodgson winner. In short, one determines the minimum number of pairwise preference switches required to make any given candidate the Condorcet winner. We call this minimum the Dodgson score of the candidate in a given profile. The candidate with the smallest Dodgson score is the winner. Obviously, if a Condorcet winner exists in a profile, it requires zero switches to become one, while all other candidates need more switches. Hence, Dodgson's rule is a Condorcet extension.

Despite being a Condorcet extension, Dodgson's rule cannot be implemented solely on the basis of information obtained from pairwise comparisons. A full preference ranking is called for to determine the minimum number of binary switches required to make any given candidate a Condorcet winner. Thus, the simplicity of voter input characterizing many Condorcet extensions is not present in Dodgson's rule. However, the other virtues of pairwise procedures mentioned above apply to this rule. With a Condorcet winner absent in a profile, all these virtues evaporate. Table 8 is reproduced from Nurmi (2004, p. 10) where it was used to demonstrate that the rule can elect an absolute loser, i.e., a candidate ranked last by more than 50% of the voters. This profile can also be used to show that Dodgson's winner is not necessarily the winner in all subsets of the candidate set. To wit, while D is the Dodgson winner in the Table 8 profile, it is not the winner in the subset consisting of A, C and D; in the latter, C is the Condorcet and hence Dogdson winner.

When a Condorcet winner exists in a profile, Dodgson's rule cannot lead to a P-TOP paradox. In the absence of a Condorcet winner, on the other hand, an instance of the P-TOP paradox can occur. Table 9 reproduced from Felsenthal and Nurmi (2017, p. 55) illustrates. This is admittedly a massive electorate voting on a large set of candidates, but the point is that Dodgson's rule is not in principle exempt from the P-TOP paradox. In the 100-voter profile on left side, there is no Condorcet winner, but A becomes one after the smallest number of pairwise preference switches between adjacent candidates. Adding now 10 voters with the preference ranking $ABECD$, *ceteris paribus*, leads to a profile where B requires less preference switches between adjacent candidates than A or any other candidate to become the Condorcet winner.

Dodgson's rule can also lead to the P-BOT paradox, even in relatively small electorates. Table 10 illustrates. In the 9-voter profile on the left-hand side, B is the (strong) Condorcet winner and thus also the Dodgson winner. If we augment the electorate with three voters each with the $ADBC$ ranking, C emerges as the Dodgson winner.

Table 9 Dodgson's rule and the P-TOP paradox (Felsenthal and Nurmi, 2017, p. 55)

42 voters	26 voters	21 voters	11 voters	10 voters
B	A	E	E	A
A	E	D	A	B
C	C	B	B	E
D	B	A	D	C
E	D	C	C	D

Table 10 Dodgson's rule and the P-BOT paradox (Felsenthal and Nurmi, 2017, p. 57)

5 voters	4 voters		3 voters
B	C		A
C	D	+	D
D	A		B
A	B		C

Table 11 Dodgson's rule fails on homogeneity: starting profile (Brandt, 2009, p. 460)

2 voters	2 voters	2 voters	2 voters	2 voters	1 voter	1 voter
D	B	C	D	A	A	D
C	C	A	B	B	D	A
A	A	B	C	C	B	B
B	D	D	A	D	C	C

Perhaps it is worth pointing out a peculiar property of Dodgson's rule, a property not shared by many other commonly known rules, viz. non-homogeneity. A rule is homogeneous if for any two profiles R and R' such that sizes of voter groups in R are copied n times to form R', the outcomes resulting from the application of the rule are identical in R and R'. To cite Smith (1973, p. 1029):

> Homogeneity seems an extremely natural requirement; if each voter suddenly splits into m voters each of whom has the same preferences as the original, it would be hard to imagine how "the collective preference" would change.

In the case of Dodgson's rule, such a change is, however, possible. Brandt (2009) gives an example (Table 11) involving 12 voters and four candidates.

Since there is no Condorcet winner, one looks for the minimum number of binary preference switches required to make various candidates the Condorcet winner. It turns out that one needs only three preference switches lifting A ahead of C among the eight left-most voters to make A the Condorcet winner, whereas making any other candidate the Condorcet winner would call for strictly more changes. Thus, A is the Dodgson winner in Table 11. Consider now a profile where the size of each voter group in Table 11 is multiplied by three. In the ensuing 36-voter profile candidate D requires the smallest number of preference switches between adjacent candidates to become the Condorcet winner. Thus, D is the Dodgson winner. This demonstrates the non-homogeneity of Dodgson's rule.

Non-homogeneity is a bizarre property of a voting rule.[7] Therefore, it is nice to know that not many voting rules are non-homogeneous. In the next subsection, we shall, however, discuss another non-homogenous rule that aims at bridging the gap between the Condorcet and positional intuitions.

3.4 Rules Based on Supercovering Relation

Tournament solutions have played a significant role in voting theory, e.g., in defining solution concepts for voting games (Miller, 1980; Moulin, 1986). In relatively recent papers, Pérez-Fernández and De Baets define new solution concepts based on supercovering and superdomination relation defined over preference profiles (Pérez-Fernández and De Baets, 2018a). These authors aim at finding solution concepts combining–to the extent possible–the insights of Borda and Condorcet.[8] We start with a couple of definitions for more precision and elaboration, see Pèrez-Fernàndez and De Baets (2018a).

Let A be the set of alternatives and $n(a, b)$ the number of voters preferring a to b. Let us assume that all individual preferences are strict (no ties). The first definition introduces the crucial concept of supercovering.

Definition 1 a supercovers b if (i) $n(a, b) > n(b, a)$ and (ii) $n(a, x) \geq n(b, x)$, $\forall x \in A$.

In other words, a candidate supercovers another if it beats the latter and, moreover, has at least an equal support as the latter against any other candidate. This should be distinguished from the more traditional notion of covering which requires the covering candidate to defeat the covered one and, moreover, to defeat all candidates defeated by the latter. The supercovering relation invokes 'Bordaesque' aspects—the size of support—instead of just the dichotomy: defeating or not defeating. The next definition introduces the corresponding concept of winning.

Definition 2 a is the pairwise winner if a supercovers all other candidates.

This seems to be a strengthening of the Condorcet winner. Indeed, the pairwise winner is a special case of the Condorcet winner or a sufficient(albeit not a necessary)

---

[7] That the use of a non-homogeneous voting rule makes the outcomes depend not only on the distribution of voters over preference rankings but also on the number of voters, obviously suggests that various participants may have different interests in the size of the voting body. This aspect relates the study of power to the institution design as pursued by the honoree of this volume and his associates.

[8] In a companion article, the same authors define and analyze solutions based on superdomination relation (Pérez-Fernández and De Baets, 2018b). Our focus in this paper is on the supercovering relation and the associated solutions.

condition for the existence of the latter. The Condorcet winner may or may not satisfy (ii) above and, consequently, may or may not supercover one or more other candidates. The Borda winner, in turn, may satisfy neither (i) nor (ii) and thus exist without there being a pairwise winner (Pérez-Fernández and De Baets, 2018a, Thm. 2). Conversely, the pairwise winner is necessarily the Borda winner. So, the pairwise winner is a special case of both Borda and Condorcet winners. Since it is unique, it follows that when it exists, it coincides with both Borda and Condorcet winner. As a solution concept, the pairwise winner would, thus, perfectly combine Borda's and Condorcet's intuitions of winning candidates. The drawback, obviously, is that the pairwise winner is relatively unlikely to exist. After all, it is less common than the Condorcet winner and much less common than the Borda winner. Hence, a more applicable solution would come in handy. The following definition suggests one such concept.

Definition 3 The unsupercovered set consists of candidates that are supercovered by no other candidate.

This set has the advantage of being always nonempty, but is often quite large and thus not helpful in singling out winners. Pérez-Fernández and De Baets therefore come up with another subset of candidates: the P-optimal ones.

Definition 4 P-score of a candidate measures its distance from being a pairwise winner. $P(a)$ tallies the number of pairwise switches on a's row in the pairwise comparison matrix that are required to make (i) $n(a, x)$ greater than half of the voters for all $x \in A$, and (ii) to make $n(a, y) = \max_x n(x, y)$. P-optimal candidates are those with minimum P-scores.

Some properties of the P-optimal candidates:

1. P-optimal candidates are always unsupercovered, i.e., no candidate supercovers them
2. if a pairwise winner exists, it is the unique unsupercovered and P-optimal candidate
3. P-optimal candidate may not coincide with the Condorcet winner (not even with a strong one) (Pérez-Fernández and De Baets, 2018a, p. 338))
4. P-optimal candidate may not coincide with a (strong) Condorcet winner that is simultaneously the Borda one (Table 12)

Table 12 The Borda and strong Condorcet winner is not P-optimal

9 voters	4 voters		a	b	c	B-score	P-score
a	b	a	–	9	9	18	P(a) = 4 (= 0 + 4)
b	c	b	4	–	13	17	**P(b) = 3**(= 3 + 0)
c	a	c	4	0	–	4	P(c) = 12 (= 3 + 9)

Table 13 Non-homogeneity of P-optimal outcomes: initial profile

3	1	2	1	1	1	1		a	b	c	d	P-score
b	b	c	a	b	c	d	a	–	4	5	9	P(a)=**3**
a	a	a	c	c	a	c	b	6	–	5	7	P(b)=**3**
c	d	b	d	a	d	b	c	5	5	–	8	P(c)= **3**
d	c	d	b	d	b	a	d	1	3	2	–	P(d)=12

Table 14 Non-homogeneity of P-optimal outcomes: expanded profile

6	2	4	2	2	2	2		a	b	c	d	P-score
b	b	c	a	b	c	d	a	–	8	10	18	P(a)=**4**
a	a	a	c	c	a	c	b	12	–	10	14	P(b)=5
c	d	b	d	a	d	b	c	10	10	–	16	P(c)= 5
d	c	d	b	d	b	a	d	2	6	4	–	P(d)=22

Property 4 demonstrates that the rule of searching for P-optimal candidates can radically part company with both Borda and Condorcet winners: it can result in a different outcome in situations where the Borda winner results in the same outcome as the strong Condorcet winner.

As Dodgson's rule, also the search for P-optimal candidates is associated with a bizarre twist: non-homogeneity. Table 13 which is slight modification of an example presented by Pèrez-Fernàndez and De Baets (2018a, p. 350) illustrates.

In this setting, a, b and c are elected. Suppose now that each unanimous voter group has twice the number of members as in Table 13 so that the profile of Table 14 ensues.

Now a becomes the unique P-optimal alternative.

On the positive side, it seems that the search for P-optimal candidates is invulnerable to the P-BOT paradox. To sketch an argument to support this, consider a profile where y is the winner and x is not the winner. Suppose that a group of unanimous voters, each ranking x last in their preferences, joins the electorate, *ceteris paribus*. The new pairwise comparison matrix changes accordingly. In fact, the only unchanged row is the one representing x's comparisons. The P-scores tally the distances from the column maxima and from the winning threshold. Thus, x's P-score cannot be smaller in the expanded electorate than in the original. In particular, it cannot be smaller than y's P-score because the latter received additional support at least in the x vs. y comparison. Hence, x cannot be a P-optimal candidate.

Table 15 Summary of some features examined

	P-optimal	Nanson	Dodgson	Black
Condorcet winner elected	No	Yes	Yes	Yes
Majority winner elected	No	Yes	Yes	Yes
Monotonicity	Yes	No	No	Yes
Homogeneity	No	Yes	No	Yes
P-BOT invulnerability	Yes	No	No	No

4 Concluding Remarks

Table 15 gives a summary of some features of the rules discussed.[9] It should be emphasized, though, that this is a very limited set of performance evaluation criteria. This set is motivated by the pros and cons related to pairwise comparison rules discussed in the second section above. The intuitively plausible property of monotonicity has not been discussed in the preceding. In the present context, it refers to the property of a rule according to which additional support within a fixed electorate, *ceteris paribus*, never renders winners into no-winners. In other words, if x wins under a given rule in profile P of preferences, it also wins if the P is changed so that the position of x is improved in some voters' preferences and no other changes are made. That is, additional support within a given electorate never harms the winner. The nonmonotonicity of Dodgson's and Nanson's rules as well as the monotonicity of Black's rule have been established in Felsenthal and Nurmi (2017, p. 96, 104), and Fishburn (1977, p. 478) the monotonicity of the search for P-optimal candidates has been shown in Pèrez-Fernàndez and De Baets (2018a, p. 349).

The best performing rule in this very restricted evaluation is Black's. It is basically a lexical application of two incompatible criteria with Condorcet's given the priority. In a way, it is no compromise at all. The same is true of Dodgson which introduces the positional aspect by defining a distance measure in profiles. Nanson's rule is an insightful utilization of a connection between the Borda scores and Condorcet winners: while the latter may not coincide with the Borda winners, their scores always exceed the average Borda scores of the candidates. Hence, both rules are present in the Nanson elimination process that always leads to the Condorcet winner if one originally exists. The bridge connecting the Borda and Condorcet intuitions is, however, not 'smooth': as Dodgson' rule Nanson's is non-monotonic and together with other Condorcet extensions here, vulnerable to the P-BOT paradox. Unlike Dodgson, it satisfies the most natural condition of homogeneity. In terms of monotonicity and invulnerability to the P-BOT paradox, the search for P-optimal candidates does very well. Its main flaw is non-homogeneity and the possibility for quite radical departure from both the Borda and Condorcet winners.

[9] The labels of criteria in the first column are phrased so as to make 'yes' a preferable value to 'no'. So, the more 'yes' values assigned to a rule, the better its performance in terms of this evaluation.

Many important aspects of the Borda-Condorcet controversy have been glossed over in the preceding. It is well-known that for showing the incompatibility of a rule with a criterion, all one needs is an example where the former leads to a choice not allowed by the latter. But how likely is one to encounter such an example? What is the minimum number of candidates and voters required for it? Many important results have already been achieved by Brandt and his associates using integer programming and Ehrhart theory (Brandt et al., 2019, 2022). Computing similar minimal examples of radical (in some precise sense) discrepancies between choices resulting from various voting rules would seem feasible and worthwhile if for no other reason than to find out the practical relevance of the study of voting rules. Perhaps the preceding could motivate another line of inquiry: are there fairness or power distribution differences between methods aimed at building bridges over the Borda–Condorcet divide? I am sure the honoree of this volume and his associates could provide useful insights on this question.

Acknowledgements The author wishes to thank Manfred J. Holler and Sascha Kurz for numerous constructive comments on an earlier draft.

References

Baldwin, J. M. (1926). The technique of the Nanson preferential majority system. *Proceedings of the Royal Society of Victoria, 39*, 42–52.

Black, D. (1958). *Theory of committees and elections*. Cambridge University Press.

de Borda, J.-C. (1781). Mémoire sur les élections au scrutin. Histoire de l'Académie Royale des Sciences annee 1781. English translation in De Grazia, A. (1953). Mathematical derivation of an election system. *Isis,* 44, 42–51 and in McLean, I. & Urken, A. B. (eds) (1995) *Classics in Social Choice*, pp. 83–89. Ann Arbor, MI: University of Michigan Press.

Brandt, F. (2009). Some remarks on Dodgson's voting rule. *Mathematical Logic Quarterly, 55*, 460–463.

Brandt, F., Hofbauer, J., & Strobel, M. (2019). Exploring the no-show paradox for Condorcet extensions using Ehrhart theory and computer simulations. In *Proceedings of the 18th international conference on autonomous agents and multiagent systems (AAMAS)*, pp. 520–528

Brandt, F., Matthäus, M., & Saile, C. (2022). Minimal voting paradoxes. *Journal of Theoretical Politics, 34,* 527–551.

Condorcet, Marquis de (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Imprimerie Royale.

Felsenthal, D. S., & Machover, M. (1992). After two centuries should Condorcet's voting procedure be implemented? *Behavioral Science, 37*, 250–273.

Felsenthal, D. S., & Nurmi, H. (2017). *Monotonicity paradoxes afflicting procedures for electing a single candidate*. Springer.

Felsenthal, D. S., & Tideman, N. (2013). Varieties of failure of monotonicity and participation under five voting methods. *Theory and Decision, 75*, 59–77.

Fishburn, P. C. (1974). Paradoxes of voting. *American Political Science Review, 68*, 537–546.

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

Fishburn, P. C., & Brams, S. J. (1983). Paradoxes of preferential voting. *Mathematics Magazine, 56*, 207–214.

Kurz, S., Mayer, A., & Napel, S. (2020). Weighted committee games. *European Journal of Operational Research, 282*, 972–979.

Kurz, S., Mayer, A., & Napel, S. (2021). Influence in weighted committees. *European Economic Review, 132*, 103634.

Mayer, A., & Napel, S. (2020). Weighted voting on the IMF Managing Director. *Economics of Governance, 21*, 237–244.

McLean, I. (1991). Forms of representation and systems of voting. In D. Held (Ed.), *Political theory today* (pp. 172–196). Polity Press.

McLean, I., & Urken, A. (1995). *Classics of social choice*. The University of Michigan Press

Meredith, J. C. (1913). *Proportional representation in Ireland*. Edward Ponsonby Ltd.

Miller, N. R. (1980). A new solution concept for tournaments and majority voting: Further graph-theoretical approaches to the theory of voting. *American Journal of Political Science, 24*, 68–96.

Moulin, H. (1986). Choosing from a tournament. *Social Choice and Welfare, 3*, 271–291.

Moulin, H. (1988). Condorcet's principle implies the no-show paradox. *Journal of Economic Theory, 45*, 53–64.

Nanson, E. J. (1883). Methods of election. *Transactions and Proceedings of the Royal Society of Victoria, 19*, 197–240.

Nurmi, H. (1999). *Voting paradoxes and how to deal with them*. Springer.

Nurmi, H. (2004). A comparison of some distance-based choice rules in ranking environments. *Theory and Decision, 57*, 5–24.

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

Pérez-Fernández, R., & De Baets, B. (2018). The supercovering relation, the pairwise winner, and more missing links between Borda and Condorcet. *Social Choice and Welfare, 50*, 329–352.

Pérez-Fernández, R., & De Baets, B. (2018). The superdominance relation, the pairwise winner, and more missing links between Borda and Condorcet. *Journal of Theoretical Politics, 31*, 46–65.

Richelson, J. T. (1978). Comparative analysis of social choice functions III. *Behavioral Science, 23*, 169–176.

Riker, W. H. (1982). *Liberalism against populism*. W. H: Freeman.

Saari, D. G. (1995). *Basic geometry of voting*. Springer.

Saari, D. G. (2001). *Chaotic elections!: A mathematician looks at voting*. American Mathematical Society.

Saari, D. G. (2003). Capturing the "will of the people". *Ethics, 113*, 333–349.

Smaoui, H., Lepelley, D., & Moyouwou, I. (2016). Borda elimination rule and monotonicity paradoxes in three-candidate elections. *Economics Bulletin, 36*, 1722–1728.

Smith, J. H. (1973). Aggregation of preferences with variable electorate. *Econometrica, 41*, 1027–1041.

Tideman, N. (1987). Independence of clones as a criterion for voting rules. *Social Choice and Welfare, 4*, 185–206.

Social Unacceptability for Simple Voting Procedures

Ahmad Awde, Mostapha Diss, Eric Kamwa, Julien Yves Rolland, and Abdelmonaim Tlidi

Abstract A candidate is said to be socially acceptable if the number of voters who rank her among the most preferred half of the candidates is at least as large as the number of voters who rank her among the least preferred half (Mahajne & Volji in Soc Choice Welfare 51:223–233, 2018). For every voting profile, there always exists at least one socially acceptable candidate. This candidate may not be elected by some well-known voting rules. In some cases, the voting rules may even lead to the election of a socially unacceptable candidate, that is a candidate such that the number of voters who rank her among the most preferred half of the candidates is strictly less than the number of voters who rank her among the least preferred half. In this paper, our contribution is twofold. First, since the existence of socially unacceptable candidates is not always guaranteed, we determine the probabilities that such candidates exist given the number of the running candidates and the size of the electorate. Second, we evaluate how often the Plurality rule, the Negative Plurality rule, the Borda rule and their two-round versions can elect a socially unacceptable candidate. We perform our simulations under both the Impartial Culture and the Impartial Anonymous Culture, two assumptions which

A. Awde
FEMTO-ST UMR CNRS 6174, University of Bourgogne Franche-Comté, Besançon, France
e-mail: ahmadawde2014@gmail.com

M. Diss (✉)
CRESE EA3190, University of Bourgogne Franche-Comté, Besançon, France
e-mail: mostapha.diss@univ-fcomte.fr; mostapha.diss@um6p.ma

University Mohamed VI Polytechnic, Africa Institute for Research in Economics and Social Sciences (AIRESS), Rabat, Morocco

E. Kamwa
LC2S UMR CNRS 8053, Université des Antilles, Schoelcher Cedex, Martinique, France
e-mail: eric.kamwa@univ-antilles.fr

J. Y. Rolland
LmB UMR CNRS 6623, University of Bourgogne Franche-Comté, Besançon, France
e-mail: julien.rolland@univ-fcomte.fr

A. Tlidi
University of Abdelmalek Essaâdi, FSJES of Tetouan, Tetouan, Morocco
e-mail: atlidi@uae.ac.ma

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_3

are widely used when studying the likelihood of voting events. Our results show that as the number of candidates increases, it becomes almost assured to have at least one socially unacceptable candidate; in some cases, the probability that half of the candidates in the running are socially unacceptable approaches or even exceeds 50%. It also turns out that the extent to which a socially unacceptable candidate is selected depends strongly on the voting rule, the underlying distribution of voters' preferences, the number of voters and the number of competing candidates.

1 Introduction

Assuming that voters have strict rankings (without indifference) on all running candidates, Mahajne & Volij (2018) introduce the concept of a socially acceptable candidate, which is a candidate whom at least half of the voters rank higher than at least half of the candidates in their rankings. The concept of social acceptability reflects a certain dichotomy in voters' preferences: the "good" or "desirable" candidates are above the line, and the "less good" or "undesirable" are below the line. Thus, being a socially unacceptable candidate means that a majority of voters may be uncomfortable with the election of such a candidate. The notion of social acceptability can be used, for instance, in the context of screening where we wish to keep the "most desirable" candidates while reassuring ourself that the chosen candidate would at least be acceptable. Mahajne & Volij (2018) show that there always exists at least one socially acceptable candidate for every preference profile; however, such a candidate may not be elected under some scoring rules, with the exception of a new scoring rule, the half accepted-half rejected (HAHR) rule. Furthermore, Mahajne & Volij (2019) show that a socially acceptable candidate may not be a $q-$Condorcet winner[1] and they identify some restricted preference domains that guarantee that any $q-$Condorcet winner is socially acceptable as a function of the threshold q. Diss & Mahajne (2020) extend the concept of social acceptability to multi-winner elections, i.e., when the goal is to select a given group of candidates, and perform the same analysis as Mahajne & Volij (2019) in that context.

Among the wide range of existing voting rules, scoring rules and scoring runoff rules are the most common in the literature and in practice. Under scoring rules, each voter ranks all of the alternatives from her most preferred to her least preferred candidate; then, points are awarded to candidates according to their position in voters' rankings. The score of a given candidate is defined by the total number of points received by that candidate taking into account all of the voters; the overall winner of the election is the candidate having the highest score. The most popular scoring rules are the Plurality rule, the Negative Plurality rule and the Borda rule. Although these rules are quite popular, they suffer from a number of limitations. For an overview of

[1] A $q-$Condorcet winner is a candidate who is preferred to each of the other candidates in pairwise majority comparisons by a fraction q of the total number of voters with $1/2 \leq q < 1$.

these limitations, the reader can refer, among others, to Felsenthal (2012), Gehrlein & Lepelley (2011, 2017) and Nurmi (1999) or to the recent book edited by Diss and Merlin (Diss & Merlin, 2021).

It is well known from the debate between de Borda (1781) and Condorcet (1785) that scoring rules may lead to the Condorcet winner paradox, i.e., they may fail to elect the Condorcet winner. When she exists, the Condorcet winner is a candidate who is preferred to each of the other competitors by a majority of voters. Scoring rules may also lead to the Condorcet loser paradox, as they may elect the Condorcet loser, a candidate defeated in pairwise comparisons by each of the other candidates. Even worse, a candidate ranked first by an absolute majority of the voters may not be elected (the absolute majority winner paradox) whereas a candidate ranked last by an absolute majority of the voters may be elected (the absolute majority loser paradox).[2] Avoiding the aforementioned paradoxes can be considered as an attempt to guarantee the election of a candidate supported by an "acceptable" majority of the electorate or to avoid the victory of a candidate who would be supported by only a given "minority". Note that the existence of a Condorcet winner is not always guaranteed, whereas this is always the case for at least one socially acceptable candidate.

As mentioned above, a socially acceptable candidate may not win under some scoring rules; these rules may in some cases instead select a socially unacceptable candidate. Note that unlike the assured existence of at least one socially acceptable candidate for any preference profile, the existence of a socially unacceptable candidate is not always guaranteed. If there is at least one socially unacceptable candidate, the fact is that a majority consensus leans in favor of the socially acceptable candidate(s) over the socially unacceptable candidate(s). It therefore appears that in such a case the election of a socially unacceptable candidate corresponds to an undesirable scenario. Since such scenarios can occur under some scoring rules, our goal in this paper is to find out if this is frequent or not. More exactly, we evaluate the probability that the Plurality rule, the Negative Plurality rule, the Borda rule, and their two-round versions select a socially unacceptable candidate. Prior to that, we determine the probabilities that a given number of socially unacceptable candidates exist. We focus on elections with a number of candidates in the set $\{3, 4, 5, 6, 10, 15\}$ for some values of the number of voters between 10 and 100,000. We perform our analysis by running simulations under both the Impartial Culture and the Impartial Anonymous Culture, which are two widely used assumptions when studying the likelihood of voting events. We will say more about these assumptions in the sequel.

Our results show that as the number of candidates increases, it becomes almost assured to have at least one socially unacceptable candidate; in some case, the probability that half of the candidates in the running are socially unacceptable approaches or even exceeds 50%. It also turns out that the extent to which a socially unacceptable candidate is selected depends strongly on the voting rule, the underlying distribution of voters' preferences, the number of voters, and the number of competing candidates.

---

[2] See for instance Diss et al. (2018).

The rest of the paper is organized as follows: Sect. 2 is devoted to definitions and to the basic notations that set the framework for our analysis; in this section, we will also present our simulation methodology. We present our probability results in Sect. 3 for each of the two preference models under consideration. Section 4 concludes.

2 Definitions

2.1 Preferences and Social Unacceptability

Let $A = \{a_1, \ldots, a_K\}$ be a set of K ($K \geq 3$) candidates and $N = \{1, \ldots, n\}$ a set of n ($n \geq 2$) voters. We denote by R the set of binary relations on A, and P the subset of complete, transitive, and antisymmetric binary relations on A. A preference profile is a mapping $\pi = (\succ_1, \ldots, \succ_n)$ of preference relations on A to the voters in N. For each voter, $i \in N$, $\succ_i$ represents i's preference relation over the candidates in A. We denote by P^n the set of preference profiles. A *voting situation* is a $K!$-tuple $\tilde{n} = (n_1, n_2, \ldots, n_{K!})$ that indicates the total number n_t of voters casting each of the $K!$ complete linear orders such that $\sum_{t=1}^{K!} n_t = n$.

For every subset of preference relations $C \subseteq P$, we denote by $\mu_\pi(C) = |\{i \in N : \succ_i \in C\}|$ the number of voters whose preferences are in C. For any preference profile $\pi \in P^n$, the rank of a candidate a in the preference relation $\succ$ is defined as follows: $rank_\succ(a) = K - |\{a' \in A : a \succ a'\}|$. Candidates whose ranks in the preference relation $\succ$ are less than $(K+1)/2$ are said to be placed *above the line*, those whose ranks in the preference relation $\succ$ are greater than $(K+1)/2$ are said to be placed *below the line*, and those whose ranks in the preference relation $\succ$ are equal to $(K+1)/2$ are said to be placed *on the line*. In the preference ranking $a_1 \succ a_2 \succ a_3 \succ a_4 \succ a_5$, for instance, candidates a_1 and a_2 are above the line while candidates a_4 and a_5 are below the line, and candidate a_3 is ranked on the line. It is obvious that there are no candidates on the line when the number of candidates is even.

We can now define the concept of social (un)acceptability as introduced by Mahajne & Volij (2018) for single-winner elections.

Definition 1 Let $\pi \in P^n$ be a profile of preference relations, and let $a \in A$ be a given candidate. We say that a is socially unacceptable with respect to π if the number of voters that place her below the line is strictly greater than the number of voters that place her above the line, otherwise a is called socially acceptable. Formally, a is socially unacceptable with respect to π if and only if

$$\mu_\pi\Big(\Big\{\succ : rank_\succ(a) < (K+1)/2\Big\}\Big) < \mu_\pi\Big(\Big\{\succ : rank_\succ(a) > (K+1)/2\Big\}\Big).$$

For illustration, let us consider the following simple example.

Example 1 Consider an election such that $A = \{a_1, a_2, a_3, a_4\}$ and $N = \{1, \ldots, 7\}$ with the following profile:

$$\pi = \begin{bmatrix} a_1 & a_1 & a_1 & a_2 & a_2 & a_3 & a_4 \\ a_2 & a_2 & a_3 & a_3 & a_4 & a_4 & a_3 \\ a_3 & a_4 & a_4 & a_1 & a_1 & a_1 & a_1 \\ a_4 & a_3 & a_2 & a_4 & a_3 & a_2 & a_2 \end{bmatrix}$$

The first column in π represents the preference relation of voter 1, the second one is the preference relation of voter 2 and so on. In this election, a_1 and a_4 are socially unacceptable candidates since they are ranked below the line by four voters while only three voters rank them above the line; a_2 and a_3 are socially acceptable candidates since they are ranked above the line by four voters while only three voters rank them below the line. Notice that if we erase $\succ_4$ in the above preference profile, we end up with a profile where there is no socially unacceptable candidate.

2.2 Voting Rules

Given a preference profile $\pi \in P^n$, a scoring voting rule is characterized by a list $S = \{S_1, S_2, \ldots, S_K\}$ meaning that each voter $i \in N$ assigns S_t points to the candidate that is ranked t-th in her preference relation with $t = 1, \ldots, K$. The scoring rule associated with the list S chooses the candidate(s) having the maximum total score. The well-known scoring rules under consideration in this paper are the following:

Plurality rule (PR): The Plurality score of a given candidate is the total number of voters who rank this candidate at the top of their rankings. In other words, the Plurality rule corresponds to the list $S = (1, 0, \ldots, 0, 0)$.

Negative Plurality rule (NPR): This rule picks the candidate with the lowest number of last places in voters' rankings; the vector of scores is given by $S = (1, 1, \ldots, 1, 0)$.[3]

Borda rule (BR): This rule gives $K - t$ points to a candidate each time she is ranked t-th; the vector of scores is given by $S = (K - 1, K - 2, \ldots, 1, 0)$.

We also consider a runoff version of each of the above scoring rules. Runoff rules are variously defined. Some will eliminate candidates one by one and others by blocks. In the case of a one-by-one elimination, the candidate with the lowest score is eliminated in each round; this is for example the case with the Baldwin rule (Baldwin, 1926) which at each round eliminates the candidate with the lowest Borda score. For block eliminations, the candidate(s) whose score does not meet a given threshold are eliminated in each round; this is the case for the Nanson rule

[3] This rule is also known as anti-plurality or veto.

(Nanson, 1883) or Kim-Roush voting rule (Kim & Roush, 1996).[4] For simplicity, we define runoff rules here as involving only two rounds: only the two candidates with the highest scores in the first round qualify for the second round. The winner will therefore be the candidate who wins the majority duel that governs the second round. The runoff rules that we consider in the paper at hand are therefore defined as follows:

Plurality with runoff (PRR): A majority duel pits the two candidates with the highest plurality scores against each other and the one who wins this duel is declared the winner.

Negative Plurality with runoff (NPRR): The two candidates who have been ranked last by the voters the fewest times in their rankings find themselves in a majority duel in the second round. The winner of this duel is declared the winner of NPRR.

Borda with runoff (BRR): The winner under this rule is the candidate who wins the majority duel between the two candidates with the highest Borda scores in the first round.

It is worth noting that whenever it is necessary to handle ties, we use the alphanumerical order. Note also that Mahajne & Volij (2018) introduced a scoring rule, called the half accepted-half rejected (HAHR), which always selects a socially acceptable candidate (and never selects a socially unacceptable candidate). HAHR assigns 1 point to the candidates placed above the line, -1 point to candidates below the line, and a score of 0 to the candidate (if there is one) on the line. Given a ranking t in a voter's preference relation, HAHR is formally defined by the scores S_t defined as follows:

$$S_t = \begin{cases} 1 & if \quad t < \frac{K+1}{2} \\ 0 & if \quad t = \frac{K+1}{2} \\ -1 & if \quad t > \frac{K+1}{2} \end{cases}$$

The HAHR winner is the candidate with the highest total score. Notice that given a scalar α and β, a $1 \times K$ matrix of ones, the scoring rules associated with the scores S_t and with the scores $\alpha S_t + \beta$ define one and the same rule. It then follows that for K even, HAHR is equivalent to the well-known $\frac{K}{2}$-approval rule; i.e., each voter casts a vote for half of the competing candidates. For the particular case of $K = 3$, HAHR is equivalent to the Borda rule.

To motivate our subject, let us reconsider the preferences of Example 1. One can check in this example that candidate a_1 who is socially unacceptable is elected under each of our three scoring rules and also under the three two-round rules. Thus, our voting rules in this example elect a socially unacceptable candidate even though there are two socially acceptable candidates. Example 1 can be used to construct

[4] At each round of the Nanson rule, all the candidates with less than the average Borda score are eliminated. The Kim-Roush voting rule eliminates those candidates having a score lower than the average Negative Plurality score.

profiles with any number of candidates such that our voting rules elect a socially unacceptable candidate.

2.3 Probability Models and Simulation Methodology

2.3.1 Probability Models

When computing theoretical probabilities of voting events (e.g., a paradoxical event such as the Condorcet's paradox), assumptions are needed on the distribution of voting situations or profiles. This is made through statistical models. In this paper, we address our subject by assuming the following two models which are among the most common in the literature: the Impartial Culture (IC) and the Impartial Anonymous Culture (IAC).

IC was first introduced in the social choice literature by Guilbaud (1952); this model assumes that every preference profile is equally likely to occur; each voter chooses her preference according to a uniform probability distribution and gives a probability of $\frac{1}{K!}$ for each ranking to be chosen independently. The likelihood of a given voting situation $\tilde{n} = (n_1, n_2, \ldots, n_{K!})$ is given by $\frac{n!}{\prod_{i=1}^{K!} n_i!} \times (K!)^{-n}$.

IAC was introduced by Kuga & Nagatani (1974) and Gehrlein & Fishburn (1976); this model assumes that all voting situations with n voters are equally likely to be observed. Under this model, the likelihood of a given voting event is equal to the ratio between the number of voting situations in which the event may occur over the total number of possible voting situations. The total number of possible voting situations with K candidates is given by the polynomial $\binom{n+K!-1}{K!-1}$. To determine the number of voting situations associated with a given voting event, several techniques, mathematical tools and algorithms have been proposed. For more details, we refer the reader to Bruns & Ichim (2021), Schürmann (2013), Cervone et al. (2005) and Wilson & Pritchard (2007).

The two models have been used over time in an impressive number of works. For a non-exhaustive overview of the use and technical developments around these two models, the reader can refer to the recent books edited by Diss and Merlin (Diss & Merlin, 2021) and Gehrlein and Lepelley (Gehrlein & Lepelley, 2011, 2017). Note that the techniques and algorithms mentioned above have a limit, namely the number of candidates. They are adapted for calculations involving at most four candidates; depending on the case, the calculation time can be quite variable (from one second to almost weeks). Since our analysis involves voting situations with more than four candidates, we have opted for simulations to get around these limitations. We present in the following the methodology that supports these simulations.

2.3.2 Simulation Methodology

Simulations under IC and IAC models have been produced using a handcrafted Python framework.[5] The simulation method has been based on the PrefLib code by Mattei & Walsh (2013), a reference library on preference data and algorithms for computational social choice. IC and IAC models have been extracted with modifications restricted to the random number generator. We then added the evaluations required for the socially unacceptable observations. A thorough selection has been performed to certify that no pre-existing preference data is used by any included algorithms. As such, the Python simulations are only based on random generation of preferences following IC or IAC models. The 128-bit implementation of the O'Neill's permutation congruential generator has been used as a pseudo-random generator (O'Neill, 2014) producing double-precision floating-point numbers with a period of 2^{128}.

The simulations proceed as follows: A number of candidates K, a number of voters n and a model (IC or IAC) is selected. A number $\mathcal{I}$ of preferences is produced randomly (the specific number can be different, as discussed below). Storing selected candidates and preferences, the set of socially unacceptable candidates selected for different values of the input parameters can be built. For each simulation, the results of the six voting rules described in Sect. 2.2 are investigated and the probabilities of observing the different cardinal values is computed. Results are discussed in Sect. 3.

Due to the formulation of the algorithms, and the intrinsic behavior of IC and IAC models, computation time increases sharply with values of K and n. As for the number $\mathcal{I}$ of generations, this is the main contributor in the computation times and its value restricts the maximum precision that can be expected for the probabilities of interest. To assess that dependency, we used the case of the IC model, $K = 10$ candidates, $n = 100$ voters and computed a test probability for increasing values of $\mathcal{I}$ preference generations. The variable tested is the probability to observe a set of socially unacceptable candidates selected with a cardinal value of 5. The results of this numerical calibration are shown in Table 1.[6]

The effect of $\mathcal{I}$ on computation time is linear, as expected for a single threaded implementation. Following those observations and due to time constraints, it has been decided to restrict some cases of the parametric study to a lower value of $\mathcal{I}$.[7] Concerning precision, if $\mathcal{I}$ =1,000,000 is taken as a reference, a reduction of generations by a factor of 10 leads to a relative difference of only 0.06% and a factor of 100 scales to a relative difference of 0.13%. The same study applied to other values of n and K shows similar results.

In regards of computational resources, memory consumption is not a factor in these CPU-bound simulations, neither are the memory bus speed or width. It is to be noted, however, that as of now no optimization aiming for a reduction of computation

[5] Code base is available at https://plmlab.math.cnrs.fr/jrolland/social-unacceptability-freezed-code.

[6] Precision is expressed as relative error E_{rel} to the most precise case, in percentages.

[7] All simulations for $K = 15$ as well as the specific case of IAC model, n = 100,000 and $K \in \{6, 10\}$ have been performed with a limited $\mathcal{I}$ =100,000 generations.

Table 1 Computation times t (in seconds) and precision of the simulations over the number $\mathcal{I}$ of preferences generated.

$\mathcal{I}$	10	100	1000	10,000	100,000	1,000,000
$\log(t)$	-0.62	0.34	1.32	2.32	3.32	4.56
E_{rel}	21.26%	28.61%	3.41%	0.13%	0.06%	–

times has been implemented in the Python framework. Computations for this study have been performed on two infrastructures: the LmB local computing facilities (Intel Xeon CPU - 3.20 GHz) and supercomputer facilities of the Mésocentre de calcul de Franche-Comté (Intel Xeon CPU-2.67 GHz).

3 Results

3.1 *The Probability that a Fixed Number of Socially Unacceptable Candidates Exist*

It is useful to note that with K candidates in the running, the number of socially unacceptable candidates can vary from 0 to $K - 1$. Since there is always at least one socially acceptable candidate, it is therefore impossible to have a voting situation in which all candidates are socially unacceptable. In this section, we investigate the proportion of voting situations in which one or more socially unacceptable candidates exist.

For a number of candidates $K \in \{3, 4, 5, 6, 10, 15\}$ and a number of voters n between 10 and 100,000, we compute the probability that a fixed number of socially unacceptable candidates exist. Let us denote by $P(j)$ the probability that exactly j socially unacceptable candidates exist. For space constraints, detailed results of numerical probabilities are provided in online supplementary material.[8] Figures 1 and 2 present in a synthetic way the results that we obtain under both IAC and IC. For better readability, the colors indicate in each case the number of socially unacceptable candidates.

First of all, we point out some general observations that emerge from our simulations. Whether under IC or IAC, the probabilities show similarities in their trend. Under both models, we notice that as n increases, $P(j)$ tends to decrease for $j < \frac{K}{2}$ and tend to increase when $j > \frac{K}{2}$. We also observe that for any given value of $K \in \{3, 4, 5, 15\}$ and n, $P(j)$ tends to increase for a j lower than $j^\star = \lfloor \frac{K}{2} \rfloor$ the value for which $P(j)$ is maximum and then decrease beyond that. We have a similar pattern for $K \in \{6, 15\}$ except that $j^\star$ is sometimes $j^\star = \frac{K}{2} - 1$ or $j^\star = \frac{K}{2}$. More

[8] This supplementary material is available at https://plmlab.math.cnrs.fr/jrolland/social-unacceptability-freezed-code/-/blob/master/Supplementary_materials/tables.pdf.

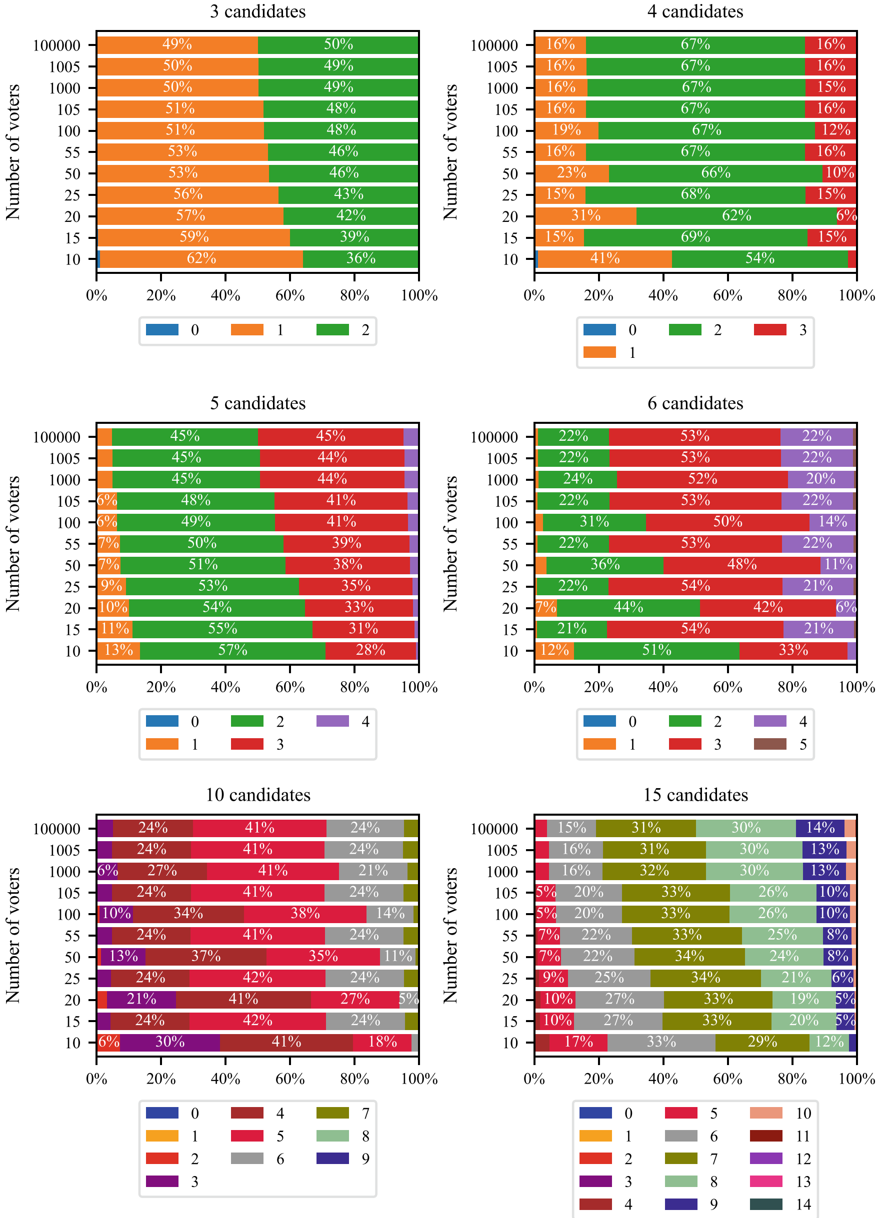


Fig. 1 Probabilities that a fixed number of socially unacceptable candidates exist under IAC model

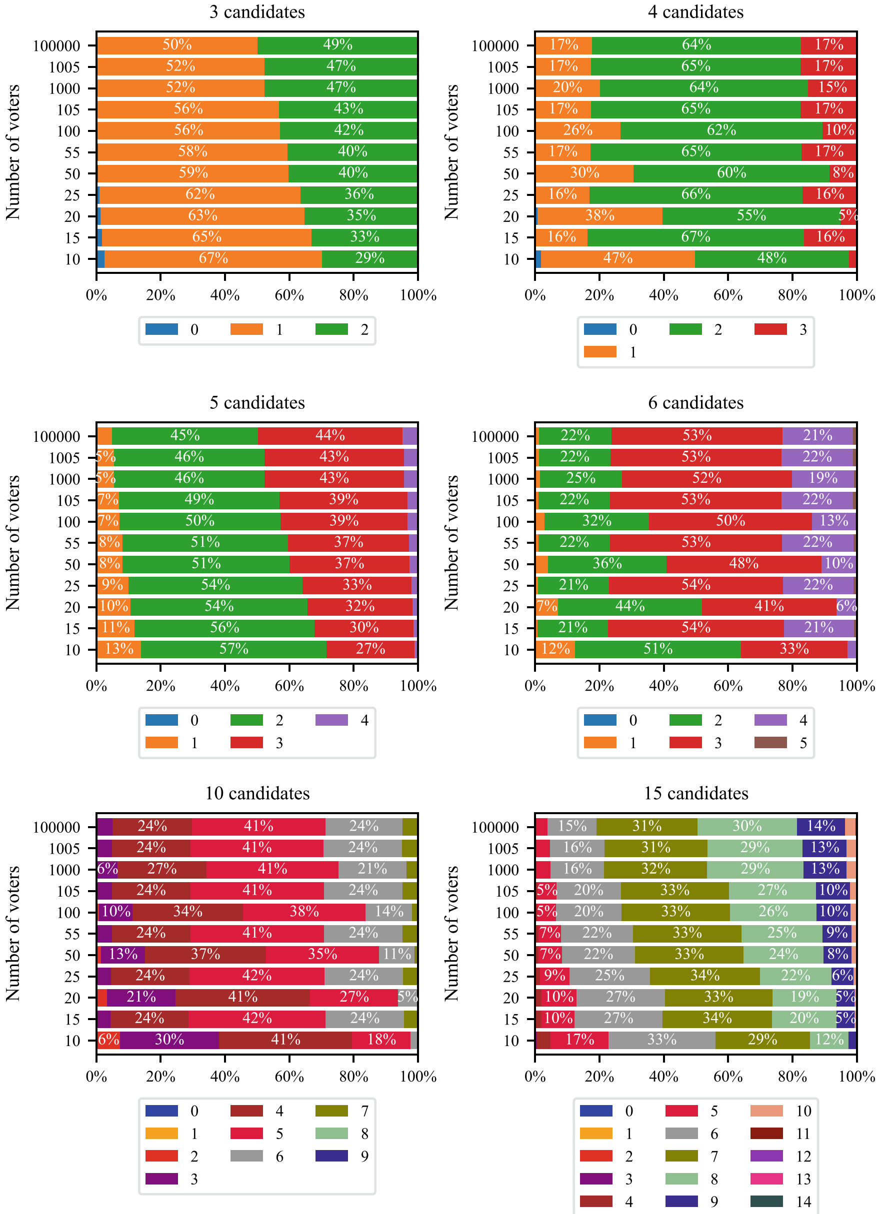


Fig. 2 Probabilities that a fixed number of socially unacceptable candidates exist under IC model

precisely, when $K = 6$, we get $j^\star = 2$ when $n \in \{10, 20\}$ while $j^\star = 3$ for the other values of n; for $K = 10$, we get $j^\star = 4$ when $n \in \{10, 20, 50\}$ and $j^\star = 5$ for the other values of n. We also notice that for an even number of candidates, $P(j)$ evolves both with the parity of the number of voters and with the parity of the number of socially unacceptable candidates.

We now turn to some probabilities that the reader will find online through the link provided above. If we look at the case of three-candidate elections ($K = 3$), we find that with ten voters there is about a 1.1% chance under IAC and about a 2.65% chance under IC that no socially unacceptable candidate exists. Thus, with such an electorate, there is a nearly 97% chance that there is at least one socially unacceptable candidate. This last conclusion is valid for all other values of n that we consider. We also note that for $K = 3$, with at least 50 voters, we reach and even exceed the 40% chance that 2 of the 3 candidates are socially unacceptable. For $K \in \{4, 6, 10\}$, there is at least a 98% chance that there is at least one socially unacceptable candidate; furthermore, for $n \in \{15, 25, 55, 105, 1005, 100{,}000\}$ the probabilities that there is no socially unacceptable candidates are very low or even zero. For $K = 15$, our results indicate that there are always at least 3 (2 for some values of n) socially unacceptable candidates; the maximum number of socially unacceptable candidates is 12 (11 for some values of n) under IAC and 13 (or even 11 or 12 for some values of n) under IC.

In sum, it seems to emerge from our simulations that for voting situations with less than ten candidates, half of the candidates in contention may (in nearly 50% of the cases or even more) be socially unacceptable; also, as the number of candidates increases, it becomes almost assured to have at least one socially unacceptable candidate.

3.2 The Probability that Some Voting Rules Select a Socially Unacceptable Candidate

The probabilities obtained in the previous section led us to the conclusion that the existence of at least one socially unacceptable candidate is almost assured for the studied voting situations. Earlier, we argued that the election of a socially unacceptable candidate is problematic because of the consensus on her status compared to that of a socially acceptable candidate. It was therefore important to verify whether the election of a socially unacceptable candidate is a rare oddity or not. As such, we computed the probabilities that a socially unacceptable candidate would be elected under PR, NPR, BR, PRR, NPRR and BRR introduced and defined above. Figures 3 and 4 show these probabilities under IAC and IC, respectively.

The first observation that emerges under both IC and IAC is that for all values of K and n, NPR appears among the voting rules to be the most likely to lead to the election of a socially unacceptable candidate; it is followed by PR, which dominates NPRR, and then PRR. BR stands out as the rule least likely to lead to the election of

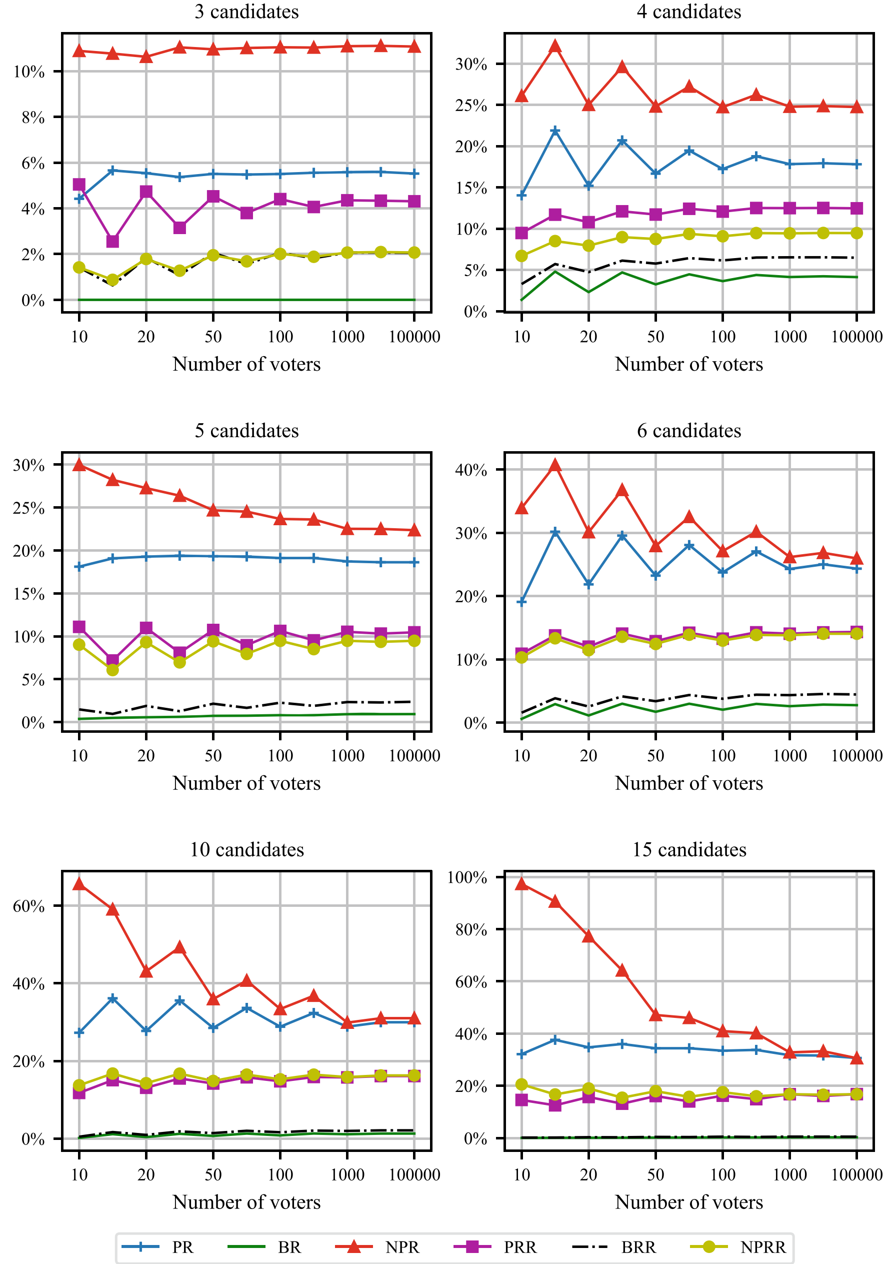


Fig. 3 Probability that some voting rules elect a socially unacceptable candidate under IAC model

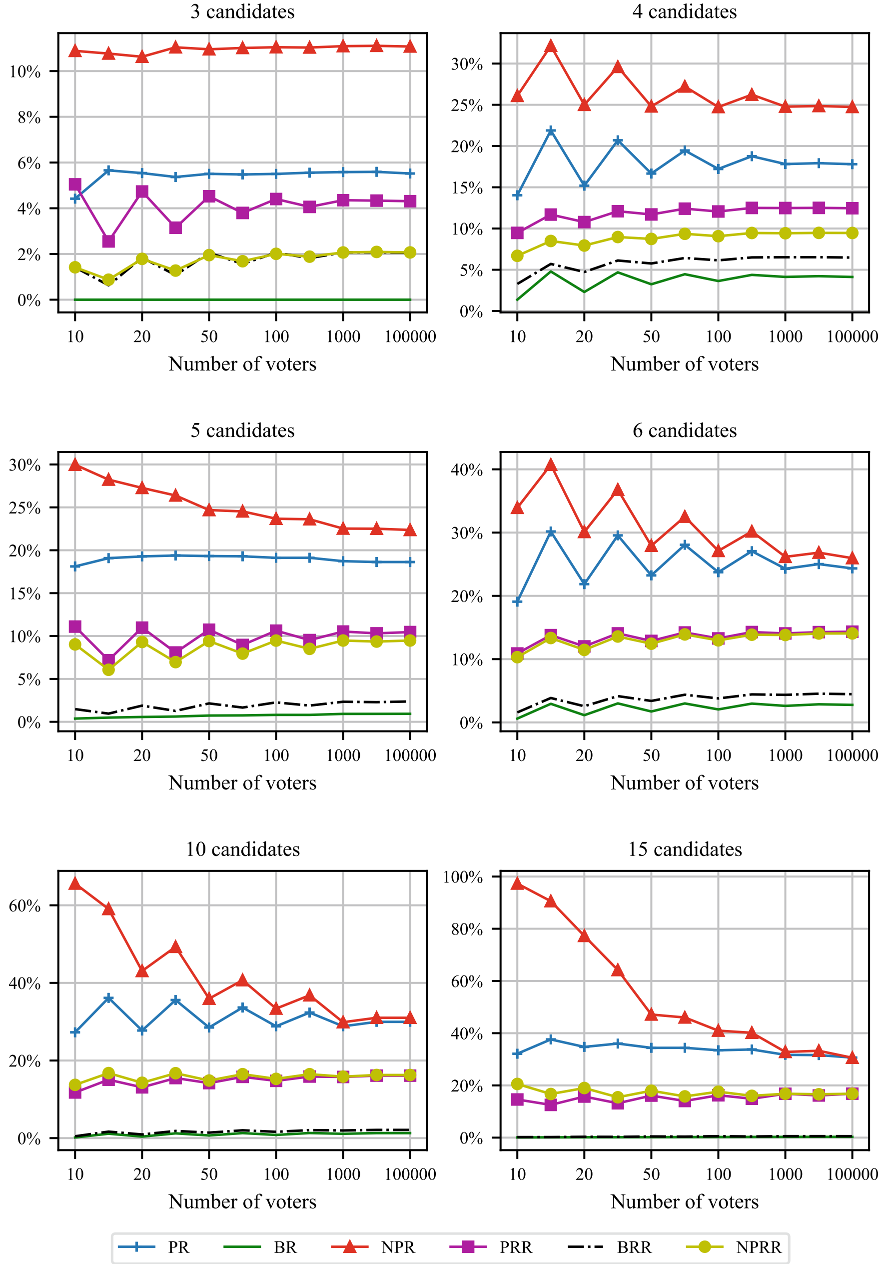


Fig. 4 Probability that some voting rules elect a socially unacceptable candidate under IC model

a socially unacceptable candidate. Let us take a closer look at the behavior of each of our rules by stressing the salient points.

With PR, for each value of K, the probabilities tend to evolve according to the parity of the number of voters. More precisely, except for $K = 3$ or $K = 5$, they tend to increase with even n and to decrease with odd n. For $K = 3$, the probabilities increase from 4.4% under IAC (5.9% under IC) with $n = 10$ to nearly 5.5% under IAC (9.4% under IC) with n = 100,000. For $K = 5$, we notice that the probabilities increase for $n \in \{10, 20, 15, 25\}$ under IAC and start decreasing for the other values of n, whereas, under IC we note a growth according to the parity of n. If we consider large values of n, the probabilities under PR tend to increase with K: we go from 5.5% under IAC (9.4% under IC) for $K = 3$ to nearly 30% (under IAC and IC) for $K = 15$. Thus, with PR, an increase in the number of candidates in the running may result into a greater possibility of electing a socially unacceptable candidate.

Under BR, the evolution of the probabilities follows an almost similar pattern under IAC and IC. The zero probabilities for $K = 3$ are well in line with the fact that BR never elects a socially unacceptable candidate since it is equivalent to HAHR. For $K = 4$, the probabilities tend to increase when n is even and to decrease when n is odd. For the other values of K, the probabilities all tend to increase according to the parity of n. We also note that it is for $K = 4$ that the probabilities are the highest under BR. We cannot clearly say, as we did with PR, how the probabilities evolve when we increase K. The fact is that when we go from $K = 4$ to $K = 5$, the probabilities tend on average to decrease significantly; they tend on average to increase significantly when we go from $K = 5$ to $K = 6$ and then we note a decrease with $K = 10$ and $K = 15$.

The pattern of probabilities obtained with NPR differs for some values of K depending on whether one is under IAC or IC. For instance, when $K = 3$, while the probabilities decrease with the parity of n under IC, they tend under IAC to decrease with n even and to increase for n odd; when $K = 4$, the probabilities decrease with the parity of n under IAC while they tend under IC to decrease with n odd and to increase for n even. When $K = 10$, the probabilities decrease with the parity of n under IAC while under IC they tend to increase with the parity of n. Even if the probabilities tend on average to increase when going from $K = 3$ to $K = 4$, and to decrease slightly when going from $K = 4$ to $K = 5$, the tendency that seems to emerge is the following: the more the number of candidates increases, the more the possibility of electing a socially unacceptable candidate increases. For all values of n considered, the probabilities seem the highest for $K = 15$: we approach 98% for $n = 10$ and 31% for n = 100,000.

In contrast to one-shot scoring rules, each of the runoff rules has, for a given K, a very similar behavior under IAC and IC: the probabilities obtained under each of these models are quite close. The probabilities tend to increase with the parity of n for each value of K: for all values of n, the probabilities are higher when $K = 4$ where they average around 6% and they are lower when $K = 15$ where they average around 0.5%. Also, given n, the probabilities tend to evolve with the parity of K. With $K = 4$, the probabilities grow from about 3.3% with 10 voters to nearly 7%

with 100,000 voters; when $K = 15$, they grow from about 0.19% with 10 voters to nearly 0.5% with 100,000 voters.

With PRR, when $K = 3$ or 6, the probabilities seem to increase for odd n and to decrease for even n. For other values of K, they simply tend to increase with the parity of n. They also tend on average to increase with K. We note that the probabilities are on average highest for $K = 15$ where they grow from almost 14.7% with 10 voters to about 16.9% with 100,000 voters. The probabilities are lowest for $K = 3$ where we reach a little over 4% with 100,000 voters.

Under IC, the probabilities with NPRR show a pattern of evolution quite close to what we have with PRR. The observation is the same under IAC except for $K = 3$ where the pattern is quite close to that of BRR. However, the probabilities are still higher than with BRR. From the similarity in the pattern of evolution of probabilities between NPRR and PRR, it is only for $K = 10$ and $K = 15$ that NPRR has higher probabilities than PRR. We notice with NPRR that except for $K = 15$ where the probabilities decrease with the parity of n, the probabilities tend to increase with the parity of n for all other values of K. Also, given n, they tend to increase with K.

To summarize, it appears that, for the rules under consideration, the probabilities of electing a socially unacceptable candidate tend to evolve both with the size of the electorate (notably its parity) and the number of candidates; BR stands out as the least likely to elect a socially unacceptable candidate. One argument that can be put forward concerning NPR's poor performance compared to the other rules is that it does not discriminate between a voter's top $K - 1$ candidates; thus, it is very likely that an candidate who is never ranked last and never ranked in the top half gets elected. We can also add that PR does not discriminate between the bottom $K - 1$ candidates, so it is likely that a candidate ranked at the top of voters' preferences wins while being ranked below the line by more than half of the voters. It seems to us that the performance of BR compared to the other rules is due to the fact that it really discriminates between the candidates by giving fewer points to a candidate below the line than to a candidate above the line; this implies "on average" that an unacceptable candidate is less likely to score better than an acceptable candidate. Nonetheless, how to explain that PRR and NPRR have quite similar behavior and perform better than their one-shot versions respectively? Why do we have the opposite behavior between BR and BRR? It seems to us that, at this stage of our analysis, it would be very daring to pronounce in a general way because things are not as simple as one might think. We also realize from the graph for $K = 15$, that for n very large, all the rules tend to have a closer behavior. Exploring situations with more candidates or exploiting real data could guide us in finer conclusions.

4 Concluding Remarks

The objective of this paper is, on the one hand, to account for the probabilities that a fixed number of socially unacceptable candidates exist. On the other hand, we wanted to determine the propensities of some popular scoring rules to elect such candidates when they exist. Using simulations under IAC and IC, we investigated voting situations with a number of candidates between 3 and 15 candidates and an electorate between 10 and 100,000 voters. Our analysis reveals under both IC and IAC that the probabilities that a fixed number of socially unacceptable candidates exist tends to decrease as the size of the electorate increases. Furthermore, there is a high probability that in some cases nearly half of the candidates will be socially unacceptable. As for the propensity of the voting rules to elect a socially unacceptable candidate, the Negative Plurality rule emerges as the most likely while the Borda rule is the least likely. In addition, the probabilities tend to evolve according to the parity of the number of voters.

This paper has then allowed us to highlight another property of the Borda rule and of its two-round version. They are the least likely to elect a socially unacceptable candidate when she exists compared to the other scoring rules and their two-round versions considered in the paper at hand. Note that a possible extension of this paper could therefore be to question the variation of our probabilities when the socially unacceptable candidate that is elected is a Condorcet winner (resp. a Condorcet loser) given that such a candidate exists. It would also be interesting to extend the notion of social acceptability to other situations than the one requiring a 50-50 splitting of individual preferences (e.g., top or bottom third, etc.).

Acknowledgements We would like to thank Sascha Kurz, Nicola Maaser and Alexander Mayer for their invitation to contribute a chapter to this volume devoted to Stefan Napel in recognition of his remarkable research career. We would also like to thank an anonymous reviewer and Alexander Mayer for their valuable comments and remarks on the first draft of this paper. We are indebted to Laboratoire de Mathématiques de Besançon (LmB) and Mésocentre de calcul de Franche-Comté for local supercomputer facilities. Mostapha Diss would like to acknowledge financial support from Région Bourgogne Franche-Comté within the program ANER 2021–2024 (project DSG). The LmB receives support from the EIPHI Graduate School (contract ANR-17-EURE-0002)

References

Baldwin, J. M. (1926). The technique of the Nanson preferential majority system of election. *Proceedings of the Royal Society of Victoria,39*, 42–52.

Bruns, W., & Ichim, B. (2021). Polytope volume by descent in the face lattice and applications in social choice. *Mathematical Programming Computation,13*, 415–442.

Cervone, D., Gehrlein, W. V., & Zwicker, W. (2005). Which scoring rule maximizes Condorcet efficiency under IAC? *Theory and Decision,58*, 145–185.

Condorcet, Marquis de (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Imprimerie Royale

de Borda, J.-C. (1781). Mémoire sur les élections au scrutin. Histoire de l'Académie Royale des Sciences

Diss, M., & Mahajne, M. (2020). Social acceptability of Condorcet committees. *Mathematical Social Sciences,105*, 14–27.

Diss, M., & Merlin, V. (2021). Evaluating voting systems with probability models, Essays by and in honor of William V. Gehrlein and Dominique Lepelley, Studies in choice and welfare. Springer.

Diss, M., Kamwa, E., & Tlidi, A. (2018). A Note on the likelihood of the absolute majority paradoxes. *Economics Bulletin,38*, 1727–1734.

Felsenthal, D. S. (2012). Review of paradoxes afflicting procedures for electing a single candidate. In D. S. Felsenthal & M. Machover (Eds.), *Electoral systems: Paradoxes, assumptions, and procedures, Studies in choice and welfare* (pp. 19–91). Springer.

Gehrlein, W. V., & Fishburn, P. C. (1976). The probability of the paradox of voting: A computable solution. *Journal of Economic Theory,13*, 14–25.

Gehrlein, W. V., & Lepelley, D. (2011). *Voting paradoxes and group coherence*. Springer.

Gehrlein, W. V., & Lepelley, D. (2017). *Elections, voting rules and paradoxical outcomes*. Springer.

Guilbaud, G.-T. (1952). Les théories de l'intérêt général et le problème logique de l'aggrégation. *Economie Appliquée,5*, 501–584.

Kim, K. H., & Roush, F. W. (1996). Statistical manipulability of social choice functions. *Group Decision and Negotiation,5*, 262–282.

Kuga, K., & Nagatani, H. (1974). Voter antagonism and the paradox of voting. *Econometrica,42*, 1045–1067.

Mahajne, M., & Volij, O. (2018). The socially acceptable scoring rule. *Social Choice and Welfare,51*, 223–233.

Mahajne, M., & Volij, O. (2019). Condorcet winners and social acceptability. *Social Choice and Welfare,53*, 641–653.

Mattei, N., & Walsh, T. (2013). PrefLib: A library of preference data http://preflib.org. In Proceedings of the 3rd international conference on algorithmic decision theory (ADT 2013), pp. 259–270

Nanson, E. J. (1883). Methods of election. *Transactions and Proceedings of the Royal Society of Victoria,19*, 197–240.

Nurmi, H. (1999). *Voting paradoxes and how to deal with them*. Springer.

O'Neill, M. E. (2014). PCG: A family of simple fast space-efficient statistically good algorithms for random number generation. Technical report, Harvey Mudd College, HMC-CS-2014-0905

Schürmann, A. (2013). Exploiting polyhedral symmetries in social choice. *Social Choice and Welfare,40*, 1097–1110.

Tideman, N., & Plassmann, F. (2012). Modeling the outcomes of vote-casting in actual elections. In D. S. Felsenthal & M. Machover (Eds.), *Electoral systems: Paradoxes, assumptions, and procedures, Studies in choice and welfare* (pp. 217–251). Springer.

Wilson, M. C., & Pritchard, G. (2007). Probability calculations under the IAC hypothesis. *Mathematical Social Sciences,54*, 244–256.

Probability of Majority Inversion with Three States and Interval Preferences

Serguei Kaniovski and Alexander Zaigraev

Abstract We examine the probability of majority inversion in a two-stage electoral process with three states. We extend May's model to state-specific general interval preferences and population weights and examine the effects of variation in population weights, and the effects of variance and bias in preferences on the inversion probability. This numerical sensitivity analysis for inversion probability is conducted using an exact formula under general interval preferences in the three states with different population weights.

1 Introduction

The main disadvantage of two-stage electoral voting is that the margin of victory in the second and final stage misrepresents the margin of victory in the first-stage based on the nationwide level of support. The most serious type of misrepresentation occurs when the outcome of the election contradicts the nationwide majority, or when the majority of voters wants a different outcome than the one obtained. This situation is known as *majority inversion*.

Figure 1 illustrates the phenomenon of majority inversion.[1] The total population of 300 voters is evenly distributed among three states. The two candidates A (red) and B (blue) differ in their popularity. The total size of the blue faction (140 voters) is smaller than that of the red faction (160 voters). Since all of a state's second-stage votes go (as a bloc) to the candidate with a majority of voters in that state, candidate B receives two votes, while candidate A receives a single vote. Majority inversion occurs because B wins the election by winning a majority of states in the

[1] This figure is borrowed from Zaigraev and Kaniovski (2020).

S. Kaniovski (✉)
Austrian Institute of Economic Research (WIFO), Vienna, Austria
e-mail: serguei.kaniovski@wifo.ac.at

A. Zaigraev
Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland
e-mail: alzaig@mat.umk.pl

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_4

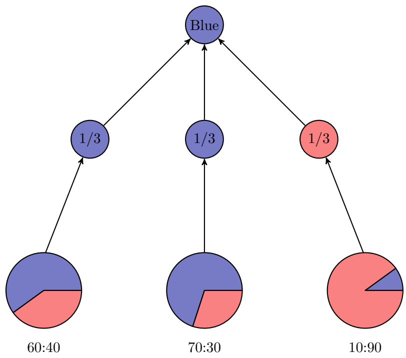


Fig. 1 Majority inversion with three states

second-stage (electoral vote), but not a nationwide majority of voters in the first-stage (popular vote). This undermines the democratic legitimacy of the election outcome, as A would beat B in a hypothetical direct election. This example assumes that the voting rule is simple majority and the second-tier votes are unweighted.

The problem of estimating the probability of majority inversion in a formal stochastic two-stage binary voting model has received considerable attention in the literature. It is fair to say that this problem is one of the oldest in the formal analysis of voting systems, dating back to the contribution by May (1948). May presented computations of the inversion probability in an idealized stochastic voting model, in which all states (constituencies) are of equal population size and the level of support for each candidates follow a standard continuous uniform distribution. This is the *standard model*, for which May computes the inversion probability for a small number of states and its limit when the number of states tends to infinity. In the recent years, several groups of researchers have been able to quantify or approximate the inversion probability under more general stochastic settings, such as the heterogeneous three-state two-stage voting system studied in this paper, or asymptotically for a large number of states (Sect. 3).

Formal analysis of stochastic voting systems encompasses a structural and a stochastic element. Both elements jointly define the set of elementary voting outcomes in the model and the probabilities of their occurrence. The structural element features the deterministic parameters of the voting system, such as the number of voters and the number of states, the weighting scheme at the second-stage of the procedure or the voting rules (majority thresholds) applied at each stage. The stochastic element of the model reflects voter behavior as an expression of voter preferences over the outcomes. By assuming that actual votes or the fraction of voters supporting a candidate or turnout are random variables, the stochastic element introduces uncertainty into the voting model and makes it necessary to consider probabilities of events. The main event under the consideration is the situation of majority inversion. The standard model assumes *homogeneity* of preference distribution across states and equality of their sizes.[2] These two assumptions allow May to compute the probability of majority inversion for a small number of states, which of course includes the case of three states (1/8), and the limit of this probability when the number of identical states tends to infinity (1/6). Section 2 discusses May's model in light of alternative stochastic voting models.

The aim of this paper is to investigate how heterogeneity in the fraction of voters supporting a candidate and population sizes affects the magnitude of the inversion probability in a general model with three states, keeping the standard model as a benchmark. Relaxing the assumption of homogeneous preferences allows us to examine how variances and biases in the preference distributions affect the inversion probability and how these effects interact with the distribution of population weights defined in the next section. To this end, we first generalize May's model by allowing distinct population weights and preference distributions, where the latter belong to rich class of uniform distributions. We call this setting the *general interval model* (Sect. 4). The analysis generalizes the results for the interval model obtained in Feix et al. (2004) for the case of three states. We then derive an exact formula for the probability of majority inversion as a function of model parameters (support intervals, population weights). This problem is stated in Sect. 5, whereas the solution can be found in the appendix. The conclusions concerning the influence of variance and bias on the inversion probability are based on numerical simulations that use the formula for the inversion probability. Section 6 discusses the simulation results and Sect. 7 summarizes the findings.

2 May's Model and Its Alternatives

Modern voting theory seems to have forgotten, rediscovered, and validated May's seminal contribution (May, 1948). To show some connections between the different contributions in the literature, let us state a more general version of May's model that allows for weighted voting. Let the number of states be n (n is odd, $n > 1$). The voters face two alternatives, A and B. For each state i, $i = 1, 2, \dots, n$,

1. the population (first-stage) weight is $w_i > 0$, such that $\sum_{i=1}^{n} w_i = 1$,
2. the voting (second-stage) weight is $v_i > 0$, such that $\sum_{i=1}^{n} v_i = 1$,
3. the share of voters who support A is a uniformly distributed random variable $\pi_i \sim U(0, 1)$, such that $\pi_1, \dots, \pi_n$ are independent.

The population weight of a state, w_i, is the number of voters in that state divided by the total number of voters in the country (the electorate). The voting weight of a state, v_i, is the number of (electoral) votes the state has in the second-stage of the voting procedure divided by the total number of such votes. The aim is to compute the inversion probability in the case of three states ($n = 3$). This is the minimal

[2] The term 'homogeneity' was introduced by Straffin (1977) in his probabilistic interpretations of classical measures of voting power (Sect. 2).

number of states required for inversions to occur. May solved this problem under the assumption that all population weights and voting weights are equal, i.e., $w_i = v_i = 1/n$, so that the states are completely identical (homogeneity). For $n = 3$, the inversion probability for varying population and voting weights has been computed in Kaniovski and Zaigraev (2018). In the variations of the above model studied in Sect. 4, we only require that *no state commands at least half of the second-stage votes* ($\max\{v_i\} < 1/2$). This implies the absence of a dictator state in the sense of the above condition.

The third assumption comes from May's idea to approach the problem of estimating the probability of majority inversion in large electorates using asymptotic analysis with respect to total population. The stochastic voting model describes the fraction of supporters of A in each state, rather than the behavior of an individual voter. Nevertheless, the choice of a uniform distribution for the fraction of supporters is not arbitrary. It has a micro foundation in the general stochastic setting that can be called the *beta-binomial* model of voting. In a state with m voters, let the probability that a voter supports candidate A be drawn from a continuous uniform distribution $U(0, 1)$. This probability is then assigned as the probability of success to a binomial random variable S representing the number of supporters of A in that state. The random variable S follows a discrete uniform distribution on $[0, 1, \ldots, m-1, m]$ and the fraction of voters who support A, or S/m, follows a discrete uniform distribution on $[0, 1/m, \ldots, (m-1)/m, 1]$. The latter becomes continuous in the limit as $m \to \infty$. This asymptotic shortcut allows to shift the focus from individual voting outcomes on the level of voters to aggregate voting outcomes on the level of states. In theoretical voting literature the above setting is known as Impartial Anonymous Culture (IAC). In statistics, this model is known as the beta-binomial model, as the uniform distribution is a special case of the Beta distribution (Beta(1,1)). In a fully general beta-binomial model, the support probability would be drawn from a general two-parameter Beta distribution (see, e.g, Casella and Berger 2002, Sect. 4.4).

The IAC or the beta-binomial model or the *homogeneity* assumption by Straffin (1977) holds a prominent place in the voting power theory.[3] The IAC is consistent with the pivot model of the Shapley-Shubik Index (SSI)—one of the two classic measures of voting power that is deeply rooted in the cooperative game theory. If the voters have a common probability of supporting candidate A and their votes are independent conditioned on the common probability, then the probability of a voter being decisive (pivotal) equals the SSI for that voter. Unconditionally, the votes are positively correlated because they have the same probability of being in favor of a candidate [see Felsenthal and Machover (1998) and De Mouzon et al. (2020)].[4]

In a recent theoretical study, Kurz and Napel (2018) show that stochastic independence of the votes can be replaced by a weaker *exchangeability* assumption. The votes

[3] For a comprehensive treatment of the theory of the measurement of voting power, see Felsenthal and Machover (1998). A survey and a discussion of the recent developments can be found in Napel (2019). For an outlook and a research agenda, see Kurz et al. (2015).

[4] Felsenthal and Machover (1998) furnish a proof of this fact, which is also discussed in Kaniovski (2008).

are exchangeable if their joint probability distribution is invariant to permutations—a certain anonymity property. This property is a necessary and sufficient condition for the voter probability of being decisive in bringing about the desired outcome, or being instrumental in the sense of Downs (1957), to be consistent with the SSI index. Exchangeability is a interesting property that also appears to be useful in the literature on Condorcet Jury Theorem (CJT), where it enters a binomial voting model based on the Bahadur parametrization (Bahadur, 1961). The parametrization provides a closed-form expression for the joint probability distribution as a function of the marginal probabilities and correlation coefficients between the votes, but the number of parameters it therefore requires is prohibitively high. This number can be significantly reduced by assuming exchangeability. This is achieved by assuming that all marginal probabilities and all correlation coefficients of the same order are equal, which is the case when all second-order or pairwise correlation coefficients are the same for each pair and all higher-order correlations equal to zero.[5] The above methodology opens a way of generalizing the binomial voting model to correlated votes. The literature on the CJT studies the probability of a correct response from a majority of expert opinions as sums of their binary votes. It can be shown that correlation between the experts can impair their ability to make correct collective decisions Kaniovski and Zaigraev (2011). Sums of exchangeable binary random variables have been studied in Zaigraev and Kaniovski (2010) and votes as exchangeable binary random variables have been studied in the context of the CJT in Zaigraev and Kaniovski (2012). Exchangeable votes have been discussed in the context of voting power in a simulation study by Kaniovski and Das (2015).

The main difference between the setting in Kurz and Napel (2018) and the models in the literature on CJT is that the latter derive exchangeability from a generalization of another probabilistic voting model known as Impartial Culture (IC). The IC model is consistent with Banzhaf's absolute measure of voting power, which is used in the literature on fair representation in two-tiered electoral systems, where the famous Penrose square root law (Penrose, 1946) has attracted considerable theoretical and even political attention in the context of the distribution of voting weights in the Council of the European Union (Kirsch, 2007).[6] Optimal weighting schemes at the second (electoral) stage of the two-stage voting procedure can be studied in the simple setting of May's model, but the number of states must be higher than three. This is because in the case of three states the assumption that no single state commands a majority of electoral votes makes the probability of majority inversion independent of the weighting scheme at the second-stage. In this situation any two states win the election. Therefore, changes in the weighting scheme at the second-stage have no

[5] For a definition of higher-order correlation coefficients, see Bahadur (1961) and Zaigraev and Kaniovski (2013). The second paper provides examples illustrating the use of the Bahadur parametrization in reliability theory and decision theory.

[6] The literature on fair representation in two-stage voting systems is comprehensive and lies outside of the present scope. The reader is referred to the classic treatment in Felsenthal and Machover (1998) and a simulation study in Maaser and Napel (2012). For a recent development, see, e.g., Kurz et al. (2017).

effect on the inversion probability in the case of three states, at least four states are required.

Distribution of the fraction of voters who support candidate A reflects preferences of the voters with respect to the candidates on the ballot. In the case of the standard uniform distribution, the model imposes minimal assumptions on the preferences. The model effectively assumes that each voter is indifferent, his or her choice is agnostic and neutral. The extension of the model studies variation in the level of support in each state, allowing for heterogeneous preferences in states that can also differ in their relative sizes (population weights). Whether preferences should be taken into account when measuring voting power is a matter of debate.[7] Felsenthal and Machover (2004) argue that a power measure should reflect *a priori* voting power. A priori power follows from the system of decision-making rules, such as voting weights and voting rules, rather than behavior of the voters. The *a priori* perspective can be based purely on cooperative game theory without a stochastic voting model consistent with it. However, if such companion stochastic model exists, as in the case of the classical measures of voting power, it is very limited (Straffin, 1977). The simple generalization of the May's model in Sect. 4 allows different states to have different probabilities of support. This departure is warranted and needed since majority inversion is a probabilistic phenomenon, whereas the same is not necessarily true for voting power.

3 Studies on Inversion Probability

May (1948) provided an exact analysis for the model with a small number of voters and states, and asymptotic approximations when at least one of those numbers becomes large. May showed that the probability of majority reversal in a two-stage electoral system with three states is equal to $1/8$, as the number of voters tends to infinity and the density of the fraction of supporters becomes continuously uniform. The probability rises to $1/6$ as the number of states tends to infinity. The recent theoretical literature tends to focus on either the exact calculation or the asymptotic analysis, as the required methodologies tend to diverge.

These early results have been confirmed in Feix et al. (2004), who obtain the same probability for the case of three and five states under IAC assumption, which is consistent with May's stochastic model. Lepelley et al. (2011) extend the asymptotic analysis and provide numerical simulations when the number of states is large. The IAC assumes that the votes are independent, conditionally on the common probability of support—this applies to any two voters in each state and any two voters from different states. Unconditionally, any two votes from the same state are positively correlated. The recent paper by De Mouzon et al. (2020) extends this setting to allow for correlation across the states. They study the probability of majority inversion in

[7] For a debate on the role of preferences in the measurement of voting power, the reader is referred to the following series of articles: Napel and Widgrén (2004a, b), Braham and Holler (2005a, b).

a model with three equally-sized states, the rate of convergence and the bounds for this probability as the number of states increases, under the assumption that any two votes (including votes from different states) in the electorate correlate. To simplify the already very complex computations, it is assumed that the states have identical population size, that simple majority rule is universally used and that the votes in the second-stage are unweighted. The IC assumption has also been studied in the existing literature. For example, Lepelley et al. (2014) show that under the IC assumption, the Penrose Square Root Law does not minimize the probability of majority inversion in the case of three states. Equalizing the ratio of the indirect voting powers in two-stage voting systems by assigning second-tier (electoral) votes proportionally to the ratio of the square-roots of population sizes does not minimize the inversion probability.

The case of three states has been analyzed in Kaniovski and Zaigraev (2018) for the general statement of May's model defined at the top of the previous section. This model features distinct voting weights in the second-stage and unequal population weights for the states. The probability of majority inversion takes the form of a quotient of two polynomials, with several distinct cases depending on the existence of a dominant voting or population weight and the weighting scheme at the second-stage of the process. The paper computes this probability and shows that inequality in the size distribution of states increases the inversion probability. The assumption of a general weighting scheme encompasses three common weighting schemes: equal voting and population weights (May's model), equal voting weights but distinct population weights (Westminster) and all voting weights equal to their respective population weights (US Electoral College). The last case is an example of a proportional weighting scheme, however, the general model allows any weighting scheme in which all voting weights and all population weights differ. The paper by Zaigraev and Kaniovski (2020) relaxes the assumption of simple majority as a voting rule by introducing arbitrary majority thresholds.

Turning to the papers on asymptotic analysis of the inversion probability, it should be noted that many asymptotic results are published together with exact calculations for small-scale problems, see, e.g., Lepelley et al. (2011). These asymptotic results are analytically obtained or uncovered by numerical simulations. The asymptotic analysis in Lepelley et al. (2011) shows that for independent and identically distributed votes the distribution of a normalized margin of victory in each state tends to a normal law. The limit value can be approximated by simulations, as in Feix et al. (2004), or analytically, as in Kikuchi (2016). The paper by Feix et al. (2021) offers a numerical analysis of different weighting schemes in the context of inversion probability that includes several stochastic voting models, while maintaining simple majority rule. The above papers are based on the classic voting models. A recent paper by Kaniovski and Zaigraev (2022) shows how to compute the full spectrum of feasible limit values of the inversion probability under a different stochastic voting model called a general binomial model. The general binomial model assumes a binomial model specific to each state, states of different population sizes and arbitrary voting quotas in both stages of the voting procedure. The asymptotic setting assumes that the number of voters tends to infinity in a way that preserved the relative sizes of the states (population weights). In this approach, population uncertainty disappears in the limit and the limit value can be obtained using combinatorial considerations and normal approximations.

4 The General Interval Model

The *standard model* by May assumes that in each state the share of voters endorsing candidate A, denoted by π, follows a standard continuous uniform distribution on a unit interval $U(0, 1)$. This distribution reflects the distribution of preferences with respect to the candidates on the ballot in each state. The standard uniform distribution is consistent with the indifference between the two candidates. The expected share of supporters $E(\pi) = 1/2$ implies $B(\pi) = |E(\pi) - 1/2| = 0$. The latter quantity can be referred to as the *absolute bias* in the preference distribution. In the standard model, there is no bias and the variance of the share of supporters is equal to $\mathrm{Var}(\pi) = 1/12$. The standard model thus reflects the absence of any prior knowledge of voter preferences, resembling the probability models underlying *a priori* reasoning behind the classical measures of voting power.

The absence of bias and the constancy of variance preclude the analysis of the effect of their variation on the inversion probability, which is the intent of the present paper. We therefore introduce a generalization of the standard model that allows for positive absolute biases and different variances in the preference distributions, while keeping the exact calculation of the inversion probability in the case of three states manageable. This allows us to introduce different preference distributions in each of the three states. The second generalization relaxes the assumption of identical population sizes. To resulting generalized model is referred to as the *interval model*. This model allows us to study how *heterogeneity* among the states in preferences and sizes affects the probability of majority inversion in a two-stage electoral procedure with three states.

The *interval model* assumes that the share of voters endorsing candidate A follows a continuous uniform distribution with a subset of the unit interval as the support. The usual definition of such distribution is based on the endpoints of the support interval of the probability density function, in the standard case 0 and 1. We parametrize the support interval using the lengths of the sub-intervals, or segments, to the left $l \in (0, 1/2]$ and to the right $r \in (0, 1/2]$ of the point $1/2$ (Fig. 2). We would thus write $\pi \sim U(1/2 - l, 1/2 + r)$, emphasizing the fact that the two parameters l and r are not the endpoints of the support interval but the distances from its endpoints to the point $1/2$.

There are several reasons for including the point $1/2$ in the support. Doing so precludes deterministic and unanimous collective decisions in the state, while still allowing for expected indifference as an important special case. The second reason is that the above parametrization allows for a more compact presentation of the inverse probability calculations, as compared to a parametrization based on the endpoints of the support interval.

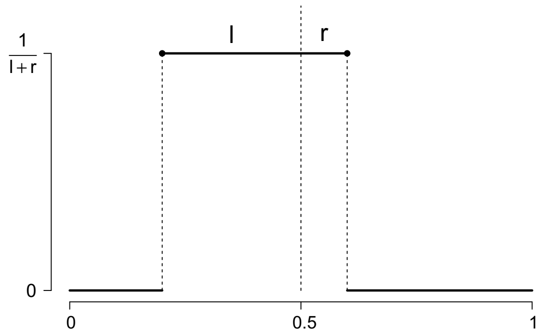


Fig. 2 Probability density of interval model

In the interval model, the share of supporters is distributed uniformly in the interval of length $l + r$, encompassing the standard density as a special case $l = r = 1/2$. This model is more flexible than the standard one, because it allows us to introduce a positive absolute bias in the preference distribution. In view of the expected value $E(\pi) = (1 - l + r)/2$, the absolute bias is equal to $B(\pi) = |l - r|/2$ and the variance is equal to $\text{Var}(\pi) = (l + r)^2/12$. The admissible values for the absolute bias belong to the interval $[0, 1/4)$, whereas the variance belongs to the interval $(0, 1/12]$.

The interval model introduces two additional features that are considered acceptable side effects given the flexibility of the distribution in modeling bias and variance. It precludes extreme levels of support, as the share of voters supporting candidate A cannot be lower than $1/2 - l$ or higher than $1/2 + r$. The second, more technical, point is that in the general interval model, the distribution of the share of voters supporting candidate A is different from the distribution of the share of voters supporting candidate B. The property of the standard distribution that $\pi \sim U(0, 1)$ implies $(1 - \pi) \sim U(0, 1)$ does not hold under the general interval model, unless $l = r$.

We relax the homogeneity assumption by allowing for different preference distributions and difference population sizes. The model is thus specified by three vectors of parameters, the vector of population weights (w_1, w_2, w_3), such that $w_i \in (0, 1)$ for $i = 1, 2, 3$ and $w_1 + w_2 + w_3 = 1$, and the two vectors of preference parameters (l_1, l_2, l_3) and (r_1, r_2, r_3), such that $l_i, r_i \in (0, 1/2]$ for $i = 1, 2, 3$. The probability of majority inversion P becomes a function of the three parameter vectors (w_1, w_2, w_3), (l_1, l_2, l_3) and (r_1, r_2, r_3).

In the special case of *homogenous preferences* given by $l = l_1 = l_2 = l_3$ and $r = r_1 = r_2 = r_3$, the following parametrization of our model: $(r - l)/2 = E$, $(r + l)/2 = D$ yields the model in Feix et al. (2004), who examine how the quantity $p = E/D = (r - l)/(r + l)$ affects the probability of majority inversion for $p > 0$ (Feix et al., 2004, Proposition 5). We relax this restriction on the sign of the fraction p and allow preferences to shift left ($r < l$) or right ($r > l$) relative to the midpoint $1/2$. In Sect. 6.3 of this paper, we compute the inversion probability under the homogeneous preferences for all cases.

5 Inversion Probability

Before proceeding with the calculation of the inversion probability, it is important to remember the assumption that no state can dictate the election outcome by commanding the majority of electoral votes. Under this assumption, a majority inversion can only occur when the outcome is backed by two electoral votes. This can happen in one of the six mutually exclusive election scenarios. In the first three scenarios, candidate A wins the election at the expense of candidate B by winning two of the three states despite failing to achieve a nationwide majority. In the remaining three scenarios, the roles are reversed and A becomes the victim of majority inversion. Calculating the inversion probability under general interval model requires evaluating the probability of these six scenarios, each of which can be defined by a system of inequalities.

Let $x_i = \pi_i - 0.5$, so that $x_i \sim U(-l_i, r_i)$ for $i = 1, 2, 3$. Then,

$$
\begin{aligned}
P_1 &= P(w_1x_1 + w_2x_2 + w_3x_3 < 0, x_1 \in (-l_1, 0), x_2 \in (0, r_2), x_3 \in (0, r_3)),\\
P_2 &= P(w_1x_1 + w_2x_2 + w_3x_3 < 0, x_1 \in (0, r_1), x_2 \in (-l_2, 0), x_3 \in (0, r_3)),\\
P_3 &= P(w_1x_1 + w_2x_2 + w_3x_3 < 0, x_1 \in (0, r_1), x_2 \in (0, r_2), x_3 \in (-l_3, 0)),\\
P_4 &= P(w_1x_1 + w_2x_2 + w_3x_3 > 0, x_1 \in (0, r_1), x_2 \in (-l_2, 0), x_3 \in (-l_3, 0)),\\
P_5 &= P(w_1x_1 + w_2x_2 + w_3x_3 > 0, x_1 \in (-l_1, 0), x_2 \in (0, r_2), x_3 \in (-l_3, 0)),\\
P_6 &= P(w_1x_1 + w_2x_2 + w_3x_3 > 0, x_1 \in (-l_1, 0), x_2 \in (-l_2, 0), x_3 \in (0, r_3)).
\end{aligned}
$$

The first inequality reflects the outcome of the popular vote, whereas the remaining inequalities reflect the voting outcomes in the three states. The above inversion scenarios represent mutually exclusive events, so the inversion probability as the probability of occurrence of one of the six scenarios is equal to the sum of their probabilities:

$$P = \sum_{i=1}^{6} P_i. \tag{1}$$

Computing each probability requires evaluating the volumes of the polytopes $A_1, A_2, \ldots, A_6$, each of which equals to the following triple integral:

$$P_i = \frac{1}{w_1w_2w_3(l_1 + r_1)(l_2 + r_2)(l_3 + r_3)} \int\int_{A_i}\int \mathrm{d}x\mathrm{d}y\mathrm{d}z, \tag{2}$$

where A_i for $i = 1, 2, \ldots, 6$ are defined as

$$
\begin{aligned}
A_1 &= \{(x, y, z) : x \in (0, w_1 l_1), y \in (0, w_2 r_2), z \in (0, w_3 r_3), x > y + z\},\\
A_2 &= \{(x, y, z) : x \in (0, w_1 r_1), y \in (0, w_2 l_2), z \in (0, w_3 r_3), y > x + z\},\\
A_3 &= \{(x, y, z) : x \in (0, w_1 r_1), y \in (0, w_2 r_2), z \in (0, w_3 l_3), z > x + y\},\\
A_4 &= \{(x, y, z) : x \in (0, w_1 r_1), y \in (0, w_2 l_2), z \in (0, w_3 l_3), x > y + z\},\\
A_5 &= \{(x, y, z) : x \in (0, w_1 l_1), y \in (0, w_2 r_2), z \in (0, w_3 l_3), y > x + z\},\\
A_6 &= \{(x, y, z) : x \in (0, w_1 l_1), y \in (0, w_2 l_2), z \in (0, w_3 r_3), z > x + y\}.
\end{aligned}
$$

Such integrals evaluate to a quotient of polynomials in the model parameters. In the appendix, we evaluate the integral in (2) by first converting it to a double integral. The evaluation of the six integrals appears tedious. However, as is shown in the appendix, once the first of the six probabilities has been computed, the remaining five probabilities can be obtained by parameter swaps.

The formula for the inverse probability as the sum of the six probabilities (2) obtained in the appendix remains valid as long as no state commands the majority of electoral votes at the second-stage of the voting procedure. In particular, the formula remains valid if more than half of the total population resides in a single state, but the largest state does not command the majority of electoral votes. If the population weight of the state exceeds $1/2$, this state dictates the outcome of the popular vote, but not the outcome of the electoral vote and the final election outcome.

6 Numerical Simulations

The exact expression for the inversion probability under the general interval model and the calculations in the appendix show that it is a complex function of the model parameters, where the parameters related to the distribution of preferences interact with the distribution of population weights in a nontrivial way. The nine parameters are collected in three vectors, one containing the population weights and the remaining two containing the parameters related to the three preference distributions.

To disentangle the effects of the nine parameters on the inversion probability, we propose a series of numerical simulations that isolate certain effects on the inversion probability by keeping the remaining effects as simple as possible. The simulations consider different voting scenarios by varying the inputs while using an exact formula for the inversion probability to produce the output. The simulations are performed using pseudo-randomly generated sets of parameters.

Since the model parameters are collected in three vectors, we do not consider the effect of individual parameters on the inversion probability. Such an analysis would require an excessive number of dimensions, making it difficult to present the results in two-dimensional figures or tables in a reasonably informative way. Instead, we summarize the model parameters along three dimensions that are conceptually relevant to our study. These dimensions are the *inequality* in the population weights, the *average variance* and the *average absolute bias* of the preference distributions.

6.1 Inequality in Population Weights

The *Gini coefficient* is a popular inequality measure that can be used to express the degree of inequality as a function of population weights.[8] For a general input vector of non-negative elements $(w_1, w_2, \ldots, w_n)$, the Gini coefficient takes values between zero and one:

$$G(w_1, w_2, \ldots, w_n) = \frac{2\sum_{i=1}^{n}(n-i+1)w_i}{n\sum_{i=1}^{n} w_i} - \frac{n+1}{n}$$

for $w_1 \geqslant w_2 \geqslant \cdots \geqslant w_{n-1} \geqslant w_n \geqslant 0$.

The value of zero indicates the equality of weights, whereas the value of $(n-1)/n$ indicates the maximal inequality of weights, or when the population is concentrated in a single state.

For three population weights, the bound on the weights simplify the expression for the Gini coefficient:

$$G(w_1, w_2, w_3) = \frac{2(w_1 - w_3)}{3}, \tag{3}$$

taking the minimal value of zero for $w_1 = w_2 = w_3 = 1/3$, and the maximal value of $2/3$ for $w_1 = 1$ and $w_2 = w_3 = 0$.

The simulations for the distribution of population weights are based on pseudo-random draws from a three-dimensional Dirichlet distribution. The Dirichlet distribution is a common choice for generating shares that need to sum to unity. The probability density of a three-point Dirichlet distribution reads:

$$f(x_1, x_2, x_3; \alpha_1, \alpha_2, \alpha_3) = \frac{x_1^{\alpha_1-1} x_2^{\alpha_2-1} x_3^{\alpha_3-1}}{\mathrm{Beta}(\alpha_1, \alpha_2, \alpha_3)},$$

where $x_i > 0$ for $i = 1, 2, 3$ and $x_1 + x_2 + x_3 = 1$. The normalizing constant contains a multivariate beta function. The chosen set of parameter $\alpha_1 = \alpha_2 = \alpha_3 = 1$ ensures that the expected value of a random vector of population weights equals to the vector $(1/3, 1/3, 1/3)$, for which the inversion probability equals $1/8$.

The top panels (a) and (b) of Fig. 3 show that the inversion probability tends to increase from $1/8$ to $1/4$ with increasing inequality in the population weights. The lower bound for the inversion probability corresponds to May's model in which the preferences are standard, $l_1 = l_2 = l_3 = r_1 = r_2 = r_3 = 1/2$, and the states have equal population weights, $w_1 = w_2 = w_3 = 1/3$. This set of parameters minimizes the probability of majority inversion, whose magnitude agrees with the value obtained by May. It is evident that the simulated inversion probabilities scatter in a relatively narrow range for a given value of the Gini coefficient. The fact that for the standard model with three (not necessarily equally-sized) states, the inversion probability

[8] The book by Yitzhaki and Schechtman (2013) gives an overview on the measurement of inequality.

increases with the inequality in the population weights has been established in Kaniovski and Zaigraev (2018), who show that the inversion probability is a Schur-convex function of population weights if there is no state commands the majority of electoral votes. Schur-convexity ensures that transferring population from a small to a large state strictly increases the degree of inequality, which is a natural property for an inequality measure.[9]

The difference between the top left and the top right panel of Fig. 3 lies in the range of admissible values for the largest population weight. In the left panel, there is no state with an absolute majority of popular votes, so that the largest population weight is smaller than $1/2$, whereas in the right panel the largest population weight $\max\{w_i\} \geqslant 1/2$. The size of the leading population weight constrains the range of attainable values for the Gini coefficient, but more importantly, it increases the lower bound on the inversion probability and the spread of the simulated probability for higher values of the Gini coefficient. If the largest population weight is not smaller than $1/2$, the minimal value of the Gini coefficient and that of the inversion probability both equal to $1/6$. In the following simulations we assume the absence of a dictator state, so that the largest state has less than one half of the total population.

The bottom panels (c) and (d) of Fig. 3 show the variation of the inversion probability with the largest population weight. In the bottom left panel, the largest population weight varies up to $1/2$. In the lower right panel, the largest weight exceeds $1/2$. The bottom panels show a similar picture to the upper panels, which is not surprising since the Gini coefficient rescales the difference between the largest and the smallest weight.

6.2 The Average Variance

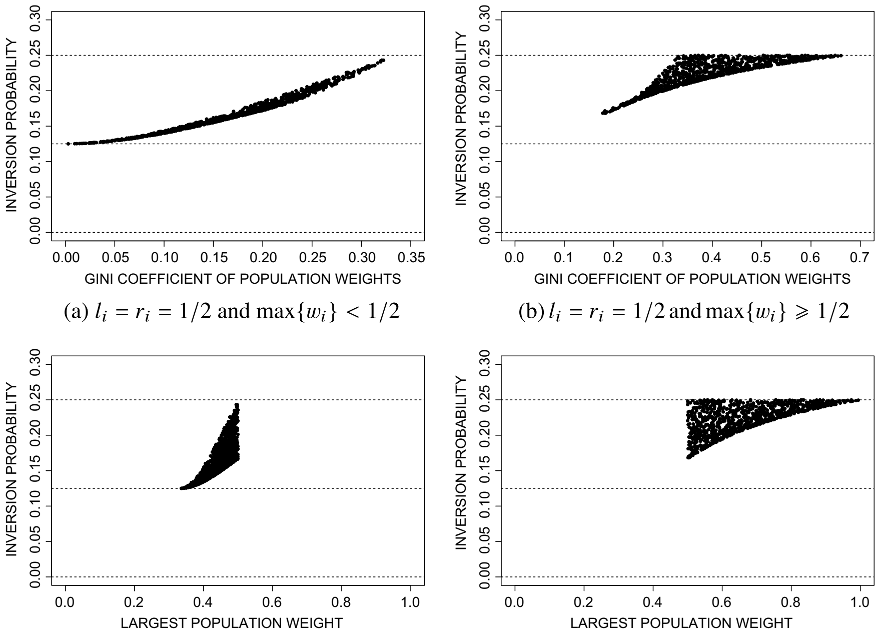
Let us now turn to the effect of the parameters of the preference distributions on the inversion probability, focusing on the variance averaged over the three states. The *average variance* is defined as follows:

$$V(l_1, l_2, l_3, r_1, r_2, r_3) = \frac{1}{3}\sum_{i=1}^{3} \mathrm{Var}(\pi_i) = \frac{1}{36}\sum_{i=1}^{3}(l_i + r_i)^2, \tag{4}$$

where the variance in state i, $i = 1, 2, 3$, is equal to $\mathrm{Var}(\pi_i) = (l_i + r_i)^2/12$, as was already noted in Sect. 4. Higher *average variance* implies more uncertainty with respect to the outcome of the popular vote.

Let $l_i = r_i$ for $i = 1, 2, 3$, so that the distributions in each state have *different variances* but are *unbiased*. The simulations with respect to the average variance in Fig. 4 show no systematic relationship between the average variance and the

[9] For details on the theory of majorization and Schur-convexity, we refer the reader to the book by Marshall et al. (2011).



(a) $l_i = r_i = 1/2$ and $\max\{w_i\} < 1/2$ (b) $l_i = r_i = 1/2$ and $\max\{w_i\} \geqslant 1/2$

(c) $l_i = r_i = 1/2$ and $\max\{w_i\} < 1/2$ (d) $l_i = r_i = 1/2$ and $\max\{w_i\} \geqslant 1/2$

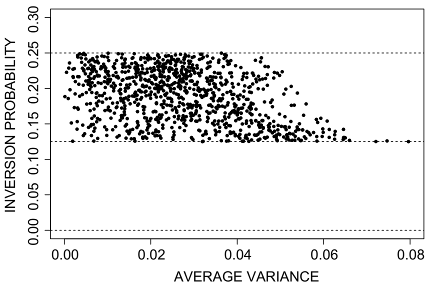
Fig. 3 Inequality in population weights

inversion probability, regardless the distribution of population weights. The scatter of the simulated inversion probabilities in the left panel with equal population weights is roughly similar to the one in the right panel with random population weights drawn from the Dirichlet distribution. However, it is evident that increasing the uncertainty of the election outcome by increasing the average variance raises the inversion probability above its value in the standard model with equally-sized states $(1/8)$. The inversion probability now lies between $1/8$ and $1/4$.

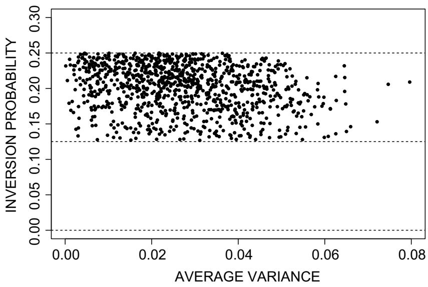
The simulation results for the inversion probability can be verified analytically using the formulas provided in the appendix. In addition to $l_i = r_i$, for $i = 1, 2, 3$, assume, without any loss of generality, that $w_1 l_1 \geqslant w_2 l_2 \geqslant w_3 l_3$. The inversion probability reads

$$P = \begin{cases} \frac{w_1 l_1 - w_2 l_2}{8 w_3 l_3} + \frac{w_1 l_1 - w_3 l_3}{8 w_2 l_2} + \frac{w_2 l_2 - w_3 l_3}{8 w_1 l_1} & \\ -\frac{w_1^2 l_1^2}{24 w_2 l_2 w_3 l_3} + \frac{w_2^2 l_2^2}{24 w_1 l_1 w_3 l_3} + \frac{w_3^2 l_3^2}{8 w_1 l_1 w_2 l_2} & \text{if } w_1 l_1 < w_2 l_2 + w_3 l_3, \\ \frac{1}{4} - \frac{w_3 l_3}{4 w_1 l_1}\left(1 - \frac{w_3 l_3}{3 w_2 l_2}\right) & \text{if } w_1 l_1 \geqslant w_2 l_2 + w_3 l_3. \end{cases}$$

The above formula is essentially the same as the formula for May's parametrization $l_i = r_i = 1/2$ in Kaniovski and Zaigraev (2018, Corollary 1), and Feix et al. (2004,



(a) $l_i = r_i$ and $w_i = 1/3$



(b) $l_i = r_i$

Fig. 4 Average variance in preferences

Proposition 5). The assumption $l_i = r_i < 1/2$ yields the same inversion probability as $l_i = r_i = 1/2$, but after rescaling the population weights. In the case of homogeneous and symmetric preferences (equal intervals, symmetric about 1/2), constraining the support of a preference distribution is equivalent to rescaling the population weights.

The sets of parameters for the variance simulation are chosen such that the preferences are unbiased, i.e. $l_i = r_i$ for all three states. This is made to exclude the effect of the bias. In the next simulations, we allow for a random bias in each state.

6.3 *The Average Absolute Bias*

Assume that $l_i + r_i = 1/2$, so that the variance is $\mathrm{Var}(\pi_i) = 1/48$ and admissible values for the absolute bias $B(\pi_i) = |l_i - r_i|/2$ defined in Sect. 4 belongs to the interval $(0, 1/4)$. Define *average absolute bias* as

$$B(l_1, l_2, l_3, r_1, r_2, r_3) = \frac{1}{3} \sum_{i=1}^{3} B(\pi_i) = \frac{1}{6} \sum_{i=1}^{3} |l_i - r_i|. \tag{5}$$

Positive *average absolute bias* implies that the voters are not indifferent between the two candidates, with the expected share of the voters supporting candidate A being lower or higher than 1/2. Let us first assume that all states have the same set of distributional parameters $l_i = l$ and $r_i = r$ for all $i = 1, 2, 3$. Assuming equal population weights makes the three states completely identical, with each state having a bias in the preference distribution given by $|l - r|/2$.

In the model with equally-sized states, increasing the average absolute bias tends to decrease the inversion probability (top left panel of Fig. 5). This relationship is not monotonic, with the inversion probability taking its maximum value 17/128 (which is larger than 1/8) at the value of the average absolute bias of $(l + r)/8 = 1/16$. The inversion probability first increases with the average absolute bias above the benchmark value of $1/8$, but then rapidly converges to zero as the average absolute bias tends to its maximum value of $1/4$. The top right panel of Fig. 5 shows the distortion in the relationship between the average absolute bias and the inversion probability introduced by varying the population weights.

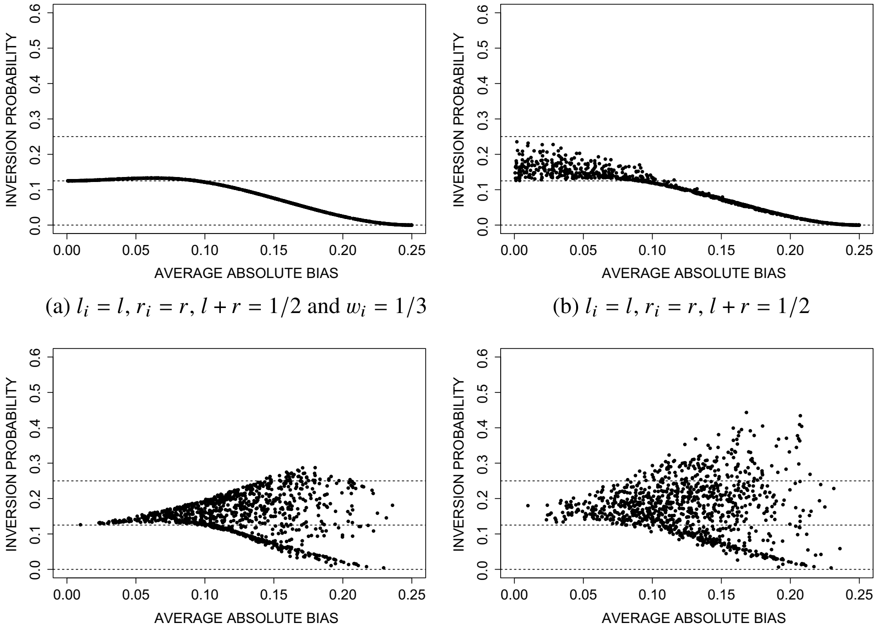
Let us now consider the most general parametrization for homogeneous states: $w_1 = w_2 = w_3,\ l = l_1 = l_2 = l_3,\ r = r_1 = r_2 = r_3$. The probability of majority inversion reads:

$$P = \begin{cases} \frac{l^2(6r-5l)}{2(l+r)^3} & \text{if } l \leqslant \frac{r}{2}, \\ \frac{l^3+r^3-2(r-l)^3}{2(l+r)^3} & \text{if } \frac{r}{2} < l < r, \\ \frac{l^3+r^3-2(l-r)^3}{2(l+r)^3} & \text{if } r \leqslant l < 2r, \\ \frac{r^2(6l-5r)}{2(l+r)^3} & \text{if } l \geqslant 2r. \end{cases}$$

The upper two equations correspond to the two equations in Feix et al. (2004, Proposition 5), which in their parametrization cover the case $p = E/D > 0$, or $r > l$ in our parametrization. The lower two equations complete the formulas for the inversion probability with the case $p = E/D < 0$, or $r < l$. The first two expressions for the inversion probability correspond to the case $p > 0$, whereas the last two expressions equal the first two when the parametrization yields $-p$ instead of p.

Note that in all the above models, it is assumed that the voting quota is equal to $1/2$. This canonical choice of the voting quota corresponds to simple majority rule in both stages of the two-stage voting procedure. However, there is an interesting relationship between the parameter values related to the distribution of preferences that maximize the inversion probability in the interval preference model and the effect of raising the voting quota from simple majority ($q = 1/2$) to qualified majority ($q > 1/2$). The maximum value of the inversion probability $17/128$ is attained for the parametrization $5l = 3r$, which implies a asymmetric preference interval around $1/2$. This maximum value was previously obtained in Zaigraev and Kaniovski (2020), who study the effect of varying the voting quota on the inversion probability in May's model with equally-sized states, i.e., for $l = r = 1/2$ and $w_1 = w_2 = w_3 = 1/3$. In the May's setting, the effect of increasing the voting quota from $1/2$ to 1 on the inversion probability is non-monotonic. The inversion probability first increases from May's calculated value of $1/8$ to $17/128$ as the quota increases from $1/2$ to $5/8$, and then decreases to 0 as the quota tends to 1. This shows that different parametrizations of the interval model can lead to the same inversion probability, either by rescaling the population weights, as was noted in Sect. 6.2, or by varying the voting quota.

The bottom panels in Fig. 5 assume heterogeneity among the states by removing the constraints $l_i = l$ and $r_i = r$, while still retaining the constraints $l_i + r_i = 1/2$ for all $i = 1, 2, 3$. They show that increasing the average absolute bias can either increase or decrease the inversion probability, but that would definitely increase its



(a) $l_i = l, r_i = r, l + r = 1/2$ and $w_i = 1/3$ (b) $l_i = l, r_i = r, l + r = 1/2$

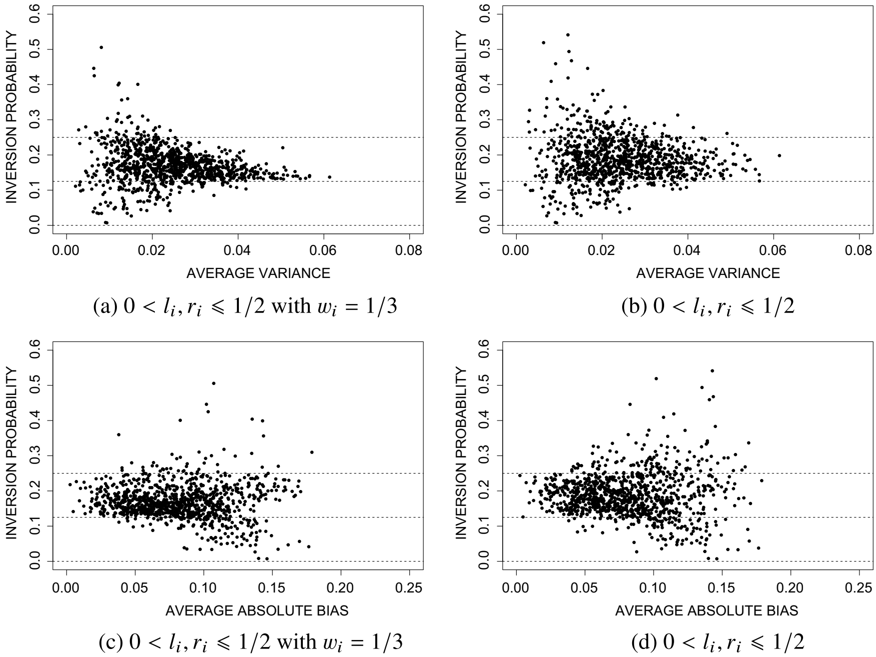
(c) $l_i + r_i = 1/2$ with $w_i = 1/3$ (d) $l_i + r_i = 1/2$

Fig. 5 Average absolute bias in preferences

spread. The spread is further increased by variation in the population weights, as is illustrated by the bottom right panel of Fig. 5.

The final round of simulations in Fig. 6 shows that the simultaneous variation in the variances and biases of the preference distributions blurs the strong individual effect of bias on the probability of majority inversion uncovered above. This underscores the benefit of considering incremental variation in model parameters coupled with precise measurement of inversion probability. We calculate the inversion probability precisely because we rely on a closed-form expression for the inversion probability, rather than on a statistical estimate of this probability obtained from the frequency of majority inversions in a random samples of voting scenarios. It is important to emphasize that the above numerical simulations are based on a suitably chosen random inputs, such pseudo-random population weights drawn from a three-point Dirichlet distribution, but the inversion probability as the output is computed in a deterministic way.

A comparison of Fig. 4 with the top panels of Fig. 6 shows that introducing bias increases the spread of inversion probabilities. Introducing variance has a similar effect, as can be seen from a comparison of the bottom panels of Fig. 5 with the bottom panels of Fig. 6, especially for small average absolute biases.



(a) $0 < l_i, r_i \leqslant 1/2$ with $w_i = 1/3$

(b) $0 < l_i, r_i \leqslant 1/2$

(c) $0 < l_i, r_i \leqslant 1/2$ with $w_i = 1/3$

(d) $0 < l_i, r_i \leqslant 1/2$

Fig. 6 Average variance and absolute bias in preferences

7 Summary

This paper shows how heterogeneity among the states in preferences and sizes affects the probability of majority inversion in a two-stage electoral procedure with three states. To this end, we first generalize May's model to state-specific interval preferences and population weights. In a second step, we examine the effects of variation in population weights and the effects of variance and bias in preferences on the inversion probability using random voting scenarios and an exact formula for the inversion probability.

Interval preferences generalize the probabilistic model of May that is based on the standard continuous uniform distribution for the proportion of supporters. The assumption of a standard continuous uniform distribution on the entire unit interval fixes the expected proportion of supporters at $1/2$ and its variance at $1/12$. By doing so, it rules out biases in the preference distribution (deviations from the indifference point $1/2$) and differences in the preference distribution among the states. The interval preference model assumes that the proportion of supporters in each state is uniformly distributed on a sub-interval of the unit interval that contains the indifference point $1/2$. This assumption allows us to stretch and shift the support of the preference distribution. This, in turn, allows us to study changes in

1. the variance of the preference distribution belongs to the interval $(0, 1/12]$,
2. the absolute bias in the preference distribution belongs to the interval $[0, 1/4)$,
3. the preference distributions for the three states.

The first generalization of May's model allows the states to have different preference distributions. The second generalization allows the states to vary in their population sizes, expressed by population weights. Together they allow for heterogeneity among the states in both preferences and sizes. The only element of May's model that we retain is the simple majority voting rule at both stages of the two-stage voting procedure.

In the reference case of May's model with three states, the inversion probability is equal to $1/8$. The numerical results for general interval model show that

1. the inequality of population weights measured by the Gini coefficient in general increases the inversion probability from $1/8$ to $1/4$,
2. for equally-sized states, lower average variance of the preference distributions (from the interval $(0, 1/12]$ versus $1/12$ in the standard model) can lead to an increase of the inversion probability up to $1/4$. The dispersion of the inversion probability versus the average variance indicates that the magnitude of this increase is not systematically related to the average variance of the preference distributions in the three states,
3. in the case of equal states, increasing the average absolute bias in the interval $[0, 1/4)$ in general decreases the inversion probability in a nonlinear manner from $1/8$ to 0. Keeping the distributional parameters equal among the states but changing their population weights introduces some variation in the inversion probability but does not change the overall conclusion. In fully heterogeneous states (preferences and sizes), increasing the average absolute bias can increase or decrease the inversion probability, but it definitely increases the spread of the inversion probability.

On a final note, we would like to emphasize that the above model has the obvious limitation of having a minimal number of states required for studying the phenomenon of majority inversion and assuming a particular distribution of preferences. Both limitations arise from the fact that we were aiming for a closed-form expression for the inversion probability as a function of the model parameters. The uniform distribution in the interval model can be replaced by other distributions, for example a triangular distribution, with a manageable increase in computational complexity associated with the need to evaluate a triple integral. However, the analysis of higher dimensional models or models with more flexible preference distributions, for example, the Beta distribution, must rely on numerical simulations. The same applies to models that assume a qualified majority as the voting rule, as may be the case with referendums where the status quo enjoys increased constitutional protection.

Acknowledgements We would like to thank the two anonymous reviewers for their comments and suggestions on the first draft.

Appendix: Computation of the Inversion Probability

The following exposition shows the essential steps in computing the inversion probability in the case of three states as a function of the population weights $\{w_i\}$ and the preference parameters $\{l_i\}$ and $\{r_i\}$, where $w_i > 0$ and $r_i, l_i \in (0, 1/2]$ for $i = 1, 2, 3$ and $w_1 + w_2 + w_3 = 1$.

Probability P_1

Consider

$$P_1 = C \iiint\limits_{A_1} \mathrm{d}x\mathrm{d}y\mathrm{d}z,$$

where $C = \frac{1}{w_1 w_2 w_3 d_1 d_2 d_3}$, $d_i = l_i + r_i$ for $i = 1, 2, 3$ and

$$A_1 = \{(x, y, z) : x \in (0, w_1 l_1), y \in (0, w_2 r_2), z \in (0, w_3 r_3), x > y + z\}.$$

Eliminating x yields

$$P_1 = C \iint\limits_{A_1'} [w_1 l_1 - z - y]\mathrm{d}y\mathrm{d}z,$$

where $A_1' = \{(y, z) : y \in (0, w_2 r_2), z \in (0, w_3 r_3), y + z < w_1 l_1\}$.

Double integrals such as the above evaluate to a quotient of polynomials. A practical difficulty in evaluating the integral lies in the number of cases that must be considered. However, once P_1 has been computed for all the cases, the remaining probabilities $P_2, P_3, \ldots, P_6$ can be computed by direct substitutions into the formula for P_1 and the corresponding inequalities. The three (mutually exclusive) cases are:

Case 1: $w_1 l_1 \geqslant w_2 r_2 + w_3 r_3$

$$P_1 = C \int_0^{w_3 r_3} \int_0^{w_2 r_2} [w_1 l_1 - z - y]\mathrm{d}y\mathrm{d}z = C \int_0^{w_3 r_3} \left[w_1 l_1 w_2 r_2 - \frac{w_2^2 r_2^2}{2} - w_2 r_2 z \right] \mathrm{d}z$$
$$= C \left[w_1 l_1 w_2 r_2 w_3 r_3 - \frac{w_2^2 r_2^2 w_3 r_3}{2} - \frac{w_2 r_2 w_3^2 r_3^2}{2} \right] = \frac{l_1 r_2 r_3}{d_1 d_2 d_3} - \frac{r_2 r_3 (w_2 r_2 + w_3 r_3)}{2 w_1 d_1 d_2 d_3}.$$

Case 2: $w_1 l_1 < w_2 r_2 + w_3 r_3$, $w_1 l_1 \leqslant w_2 r_2$

Let $m = \min\{w_3 r_3, w_1 l_1\}$. Then,

$$P_1 = C\int_0^m \int_0^{w_1l_1-z} [w_1l_1 - z - y]\mathrm{d}y\mathrm{d}z = C\int_0^m \frac{(w_1l_1-z)^2}{2}\mathrm{d}z$$
$$= \frac{C}{6}\left(w_1^3l_1^3 - (w_1l_1-m)^3\right) = \frac{Cw_1l_1m(w_1l_1-m)}{2} + \frac{Cm^3}{6},$$

or

$$P_1 = \begin{cases} \frac{l_1r_3[w_1l_1-w_3r_3]}{2w_2d_1d_2d_3} + \frac{w_3^2r_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_1l_1 > w_3r_3, \\ \frac{w_1^2l_1^3}{6w_2w_3d_1d_2d_3} & \text{if } w_1l_1 \leqslant w_3r_3. \end{cases}$$

Case 3: $w_1l_1 < w_2r_2 + w_3r_3$, $w_1l_1 > w_2r_2$
Let $m = \min\{w_3r_3, w_1l_1\}$. Then,

$$P_1 = C\left[\int_0^{w_1l_1-w_2r_2}\int_0^{w_2r_2}[w_1l_1-z-y]\mathrm{d}y\mathrm{d}z + \int_{w_1l_1-w_2r_2}^{m}\int_0^{w_1l_1-z}[w_1l_1-z-y]\mathrm{d}y\mathrm{d}z\right]$$
$$= C\left[\int_0^{w_1l_1-w_2r_2}\left[w_1l_1w_2r_2 - \frac{w_2^2r_2^2}{2} - w_2r_2z\right]\mathrm{d}z + \int_{w_1l_1-w_2r_2}^{m}\frac{(w_1l_1-z)^2}{2}\mathrm{d}z\right]$$
$$= C\left[\left(w_1l_1w_2r_2 - \frac{w_2^2r_2^2}{2}\right)(w_1l_1-w_2r_2) - \frac{w_2r_2(w_1l_1-w_2r_2)^2}{2}\right.$$
$$\left.+\frac{w_2^3r_2^3}{6} - \frac{(w_1l_1-m)^3}{6}\right]$$
$$= C\left[\frac{w_1l_1w_2r_2(w_1l_1-w_2r_2)}{2} + \frac{w_2^3r_2^3}{6} - \frac{w_1^3l_1^3}{6} + \frac{m^3}{6} + \frac{w_1l_1m(w_1l_1-m)}{2}\right],$$

or

$$P_1 = \begin{cases} \frac{l_1r_2[w_1l_1-w_2r_2]}{2w_3d_1d_2d_3} + \frac{l_1r_3[w_1l_1-w_3r_3]}{2w_2d_1d_2d_3} - \frac{w_1^2l_1^3}{6w_2w_3d_1d_2d_3} \\ +\frac{w_2^2r_2^3}{6w_1w_3d_1d_2d_3} + \frac{w_3^2r_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_1l_1 > w_3r_3, \\ \frac{l_1r_2[w_1l_1-w_2r_2]}{2w_3d_1d_2d_3} + \frac{w_2^2r_2^3}{6w_1w_3d_1d_2d_3} & \text{if } w_1l_1 \leqslant w_3r_3. \end{cases}$$

The Remaining Probabilities

For any two parameters a and b, define a parameter swap $a \leftrightarrow b$ as a mutual substitution of a for b and b for a. The remaining five probabilities can be obtained by the following parameter swaps:

$$P_2 = P_1 \text{ for } w_1 \leftrightarrow w_2, l_1 \leftrightarrow l_2, r_1 \leftrightarrow r_2,$$
$$P_3 = P_1 \text{ for } w_1 \leftrightarrow w_3, l_1 \leftrightarrow l_3, r_1 \leftrightarrow r_3,$$

$$P_4 = P_1 \text{ for } l_1 \leftrightarrow r_1, l_2 \leftrightarrow r_2, l_3 \leftrightarrow r_3,$$
$$P_5 = P_2 \text{ for } l_1 \leftrightarrow r_1, l_2 \leftrightarrow r_2, l_3 \leftrightarrow r_3,$$
$$P_6 = P_3 \text{ for } l_1 \leftrightarrow r_1, l_2 \leftrightarrow r_2, l_3 \leftrightarrow r_3.$$

Probability P_2

For $w_2 l_2 \geqslant w_1 r_1 + w_3 r_3$,

$$P_2 = \frac{r_1 l_2 r_3}{d_1 d_2 d_3} - \frac{r_1 r_3 (w_1 r_1 + w_3 r_3)}{2 w_2 d_1 d_2 d_3}.$$

For $w_2 l_2 < w_1 r_1 + w_3 r_3$, $w_2 l_2 \leqslant w_1 r_1$,

$$P_2 = \begin{cases} \frac{l_2 r_3 [w_2 l_2 - w_3 r_3]}{2 w_1 d_1 d_2 d_3} + \frac{w_3^2 r_3^3}{6 w_1 w_2 d_1 d_2 d_3} & \text{if } w_2 l_2 > w_3 r_3, \\ \frac{w_2^2 l_2^3}{6 w_1 w_3 d_1 d_2 d_3} & \text{if } w_2 l_2 \leqslant w_3 r_3. \end{cases}$$

For $w_2 l_2 < w_1 r_1 + w_3 r_3$, $w_2 l_2 > w_1 r_1$,

$$P_2 = \begin{cases} \frac{r_1 l_2 [w_2 l_2 - w_1 r_1]}{2 w_3 d_1 d_2 d_3} + \frac{l_2 r_3 [w_2 l_2 - w_3 r_3]}{2 w_1 d_1 d_2 d_3} + \frac{w_1^2 r_1^3}{6 w_2 w_3 d_1 d_2 d_3} \\ - \frac{w_2^2 l_2^3}{6 w_1 w_3 d_1 d_2 d_3} + \frac{w_3^2 r_3^3}{6 w_1 w_2 d_1 d_2 d_3} & \text{if } w_2 l_2 > w_3 r_3, \\ \frac{r_1 l_2 [w_2 l_2 - w_1 r_1]}{2 w_3 d_1 d_2 d_3} + \frac{w_1^2 r_1^3}{6 w_2 w_3 d_1 d_2 d_3} & \text{if } w_2 l_2 \leqslant w_3 r_3. \end{cases}$$

Probability P_3

For $w_3 l_3 \geqslant w_1 r_1 + w_2 r_2$,

$$P_3 = \frac{r_1 r_2 l_3}{d_1 d_2 d_3} - \frac{r_1 r_2 (w_1 r_1 + w_2 r_2)}{2 w_3 d_1 d_2 d_3}.$$

For $w_3 l_3 < w_1 r_1 + w_2 r_2$, $w_3 l_3 \leqslant w_2 r_2$,

$$P_3 = \begin{cases} \frac{r_1 l_3 [w_3 l_3 - w_1 r_1]}{2 w_2 d_1 d_2 d_3} + \frac{w_1^2 r_1^3}{6 w_2 w_3 d_1 d_2 d_3} & \text{if } w_3 l_3 > w_1 r_1, \\ \frac{w_3^2 l_3^3}{6 w_1 w_2 d_1 d_2 d_3} & \text{if } w_3 l_3 \leqslant w_1 r_1. \end{cases}$$

For $w_3l_3 < w_1r_1 + w_2r_2$, $w_3l_3 > w_2r_2$,

$$P_3 = \begin{cases} \frac{r_2l_3[w_3l_3-w_2r_2]}{2w_1d_1d_2d_3} + \frac{r_1l_3[w_3l_3-w_1r_1]}{2w_2d_1d_2d_3} + \frac{w_1^2r_1^3}{6w_2w_3d_1d_2d_3} \\ +\frac{w_2^2r_2^3}{6w_1w_3d_1d_2d_3} - \frac{w_3^2l_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_3l_3 > w_1r_1, \\ \frac{r_2l_3[w_3l_3-w_2r_2]}{2w_1d_1d_2d_3} + \frac{w_2^2r_2^3}{6w_1w_3d_1d_2d_3} & \text{if } w_3l_3 \leqslant w_1r_1. \end{cases}$$

Probability P_4

For $w_1r_1 \geqslant w_2l_2 + w_3l_3$,

$$P_4 = \frac{r_1l_2l_3}{d_1d_2d_3} - \frac{l_2l_3(w_2l_2+w_3l_3)}{2w_1d_1d_2d_3}.$$

For $w_1r_1 < w_2l_2 + w_3l_3$, $w_1r_1 \leqslant w_2l_2$,

$$P_4 = \begin{cases} \frac{r_1l_3[w_1r_1-w_3l_3]}{2w_2d_1d_2d_3} + \frac{w_3^2l_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_1r_1 > w_3l_3, \\ \frac{w_1^2r_1^3}{6w_2w_3d_1d_2d_3} & \text{if } w_1r_1 \leqslant w_3l_3. \end{cases}$$

For $w_1r_1 < w_2l_2 + w_3l_3$, $w_1r_1 > w_2l_2$,

$$P_4 = \begin{cases} \frac{r_1l_2[w_1r_1-w_2l_2]}{2w_3d_1d_2d_3} + \frac{r_1l_3[w_1r_1-w_3l_3]}{2w_2d_1d_2d_3} \\ -\frac{w_1^2r_1^3}{6w_2w_3d_1d_2d_3} + \frac{w_2^2l_2^3}{6w_1w_3d_1d_2d_3} + \frac{w_3^2l_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_1r_1 > w_3l_3, \\ \frac{r_1l_2[w_1r_1-w_2l_2]}{2w_3d_1d_2d_3} + \frac{w_2^2l_2^3}{6w_1w_3d_1d_2d_3} & \text{if } w_1r_1 \leqslant w_3l_3. \end{cases}$$

Probability P_5

For $w_2r_2 \geqslant w_1l_1 + w_3l_3$,

$$P_5 = \frac{l_1r_2l_3}{d_1d_2d_3} - \frac{l_1l_3(w_1l_1+w_3l_3)}{2w_2d_1d_2d_3}.$$

For $w_2r_2 < w_1l_1 + w_3l_3$, $w_2r_2 \leqslant w_1l_1$,

$$P_5 = \begin{cases} \frac{r_2l_3[w_2r_2-w_3l_3]}{2w_1d_1d_2d_3} + \frac{w_3^2l_3^3}{6w_1w_2d_1d_2d_3} & \text{if } w_2r_2 > w_3l_3, \\ \frac{w_2^2r_2^3}{6w_1w_3d_1d_2d_3} & \text{if } w_2r_2 \leqslant w_3l_3. \end{cases}$$

For $w_2r_2 < w_1l_1 + w_3l_3$, $w_2r_2 > w_1l_1$,

$$P_5 = \begin{cases} \frac{l_1 r_2[w_2 r_2 - w_1 l_1]}{2w_3 d_1 d_2 d_3} + \frac{r_2 l_3[w_2 r_2 - w_3 l_3]}{2w_1 d_1 d_2 d_3} + \frac{w_1^2 l_1^3}{6w_2 w_3 d_1 d_2 d_3} \\ -\frac{w_2^2 r_2^3}{6w_1 w_3 d_1 d_2 d_3} + \frac{w_3^2 l_3^3}{6w_1 w_2 d_1 d_2 d_3} & \text{if } w_2 r_2 > w_3 l_3, \\ \frac{l_1 r_2[w_2 r_2 - w_1 l_1]}{2w_3 d_1 d_2 d_3} + \frac{w_1^2 l_1^3}{6w_2 w_3 d_1 d_2 d_3} & \text{if } w_2 r_2 \leqslant w_3 l_3. \end{cases}$$

Probability P_6

For $w_3 r_3 \geqslant w_1 l_1 + w_2 l_2$,

$$P_6 = \frac{l_1 l_2 r_3}{d_1 d_2 d_3} - \frac{l_1 l_2 (w_1 l_1 + w_2 l_2)}{2w_3 d_1 d_2 d_3}.$$

For $w_3 r_3 < w_1 l_1 + w_2 l_2$, $w_3 r_3 \leqslant w_2 l_2$,

$$P_6 = \begin{cases} \frac{l_1 r_3[w_3 r_3 - w_1 l_1]}{2w_2 d_1 d_2 d_3} + \frac{w_1^2 l_1^3}{6w_2 w_3 d_1 d_2 d_3} & \text{if } w_3 r_3 > w_1 l_1, \\ \frac{w_3^2 r_3^3}{6w_1 w_2 d_1 d_2 d_3} & \text{if } w_3 r_3 \leqslant w_1 l_1. \end{cases}$$

For $w_3 r_3 < w_1 l_1 + w_2 l_2$, $w_3 r_3 > w_2 l_2$,

$$P_6 = \begin{cases} \frac{l_2 r_3[w_3 r_3 - w_2 l_2]}{2w_1 d_1 d_2 d_3} + \frac{l_1 r_3[w_3 r_3 - w_1 l_1]}{2w_2 d_1 d_2 d_3} + \frac{w_1^2 l_1^3}{6w_2 w_3 d_1 d_2 d_3} \\ +\frac{w_2^2 l_2^3}{6w_1 w_3 d_1 d_2 d_3} - \frac{w_3^2 r_3^3}{6w_1 w_2 d_1 d_2 d_3} & \text{if } w_3 r_3 > w_1 l_1, \\ \frac{l_2 r_3[w_3 r_3 - w_2 l_2]}{2w_1 d_1 d_2 d_3} + \frac{w_2^2 l_2^3}{6w_1 w_3 d_1 d_2 d_3} & \text{if } w_3 r_3 \leqslant w_1 l_1. \end{cases}$$

Probability of Majority Inversion P

Probability of majority inversion equals the sum of the six probabilities:

$$P = P_1 + P_2 + P_3 + P_4 + P_5 + P_6.$$

References

Bahadur, R. R. (1961). A representation of the joint distribution of responses to *n* dichotomous items. In H. Solomon (Ed.), *Studies in item analysis and prediction* (pp. 158–168). Stanford University Press.

Braham, M., & Holler, M. (2005a). The impossibility of a preference-based power index. *Journal of Theoretical Politics,17*, 137–157.

Braham, M., & Holler, M. (2005b). Power and preferences again. A reply to Napel and Widgrén. *Journal of Theoretical Politics,17*, 389–395.

Casella, G., & Berger, R. L. (2002). *Statistical inference*. Duxbury Press.

De Mouzon, O., Laurent, T., Le Breton, M., & Lepelley, D. (2020). The theoretical Shapley-Shubik probability of an election inversion in a toy symmetric version of the US presidential electoral system. *Social Choice and Welfare, 54,* 363–395.

Downs, A. (1957). *An economic theory of democracy*. Harper & Row.

Feix, M. R., Lepelley, D., Merlin, V. R., & Rouet, J.-L. (2004). The probability of conflicts in a U.S. presidential type election. *Economic Theory,23,* 227–257.

Feix, M. R., Lepelley, D., Merlin, V. R., Rouet, J.-L., & Vidu, L. (2021). Majority efficient representation of the citizens in a Federal Union. In M. Diss & V. R. Merlin (Eds.), *Evaluating voting systems with probability models, essays by and in honor of William V. Gehrlein and Dominique Lepelley*, *Studies in Choice and Welfare* (pp. 163–187). Springer.

Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power: Theory and practice, problems and paradoxes*. Edward Elgar.

Felsenthal, D. S., & Machover, M. (2004). A priori voting power: What is it all about? *Political Studies Review, 2,* 1–23.

Kaniovski, S. (2008). The exact bias of the Banzhaf measure of power when votes are neither equiprobable nor independent. *Social Choice and Welfare, 31,* 281–300.

Kaniovski, S., & Das, S. (2015). Measuring voting power in games with correlated votes using Bahadur's parametrisation. *Social Choice and Welfare, 44,* 349–367.

Kaniovski, S., & Zaigraev, A. (2011). Optimal jury design for homogeneous juries with correlated votes. *Theory and Decision, 71,* 439–459.

Kaniovski, S., & Zaigraev, A. (2018). The probability of majority inversion in a two-stage voting system with three states. *Theory and Decision, 84,* 525–546.

Kaniovski, S., & Zaigraev, A. (2022). Asymptotic probability of majority inversion under a general binomial voting model. *Operations Research Letters 50*, 391–398.

Kikuchi, K. (2016). *The likelihood of majority inversion in an indirect voting system*. Working Paper, Waseda University, School of Political Science and Economics.

Kirsch, W. (2007). On Penrose's square-root law and beyond. *Homo Oeconomicus, 24,* 357–380.

Kurz, S., Maaser, N., & Napel, S. (2017). On the democratic weights of nations. *Journal of Political Economy, 125,* 1599–1634.

Kurz, S., Maaser, N., Napel, S., & Weber, M. (2015). Mostly sunny: A forecast of tomorrow's power index research. *Homo Oeconomicus, 32,* 133–146.

Kurz, S., & Napel, S. (2018). The roll call interpretation of the Shapley value. *Economics Letters, 173,* 108–112.

Lepelley, D., Merlin, V. R., & Rouet, J.-L. (2011). Three ways to compute accurately the probability of the referendum paradox. *Mathematical Social Sciences, 62,* 28–33.

Lepelley, D., Merlin, V. R., Rouet, J.-L., & Vidu, L. (2014). Referendum paradox in a Federal Union with unequal populations: The three state case. *Economics Bulletin,34*, 2201–2207.

Maaser, N., & Napel, S. (2012). A note on the direct democracy deficit in two-tier voting. *Mathematical Social Sciences, 63,* 174–180.

Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: Theory of majorization and its applications*. Springer.

May, K. (1948). Probabilities of certain election results. *The American Mathematical Monthly, 55,* 203–209.

Napel, S. (2019). Voting power. In R. Congleton, B. Grofman, & S. Voigt (Eds.), *Oxford handbook of public choice* (pp. 103–126). Oxford University Press.

Napel, S., & Widgrén, M. (2004a). Power measurement as sensitivity analysis: A unified approach. *Journal of Theoretical Politics,16*, 517–538.

Napel, S., & Widgrén, M. (2004b). The possibility of a preference-based power index. *Journal of Theoretical Politics,17*, 377–387.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109,* 53–57.

Straffin, P. D. (1977). Homogeneity, independence and power indices. *Public Choice, 30,* 107–118.

Yitzhaki, S., & Schechtman, E. (2013). *The Gini methodology: A primer on a statistical methodology*. Springer.

Zaigraev, A., & Kaniovski, S. (2010). Exact bounds on the probability of at least k successes in n exchangeable Bernoulli trials as a function of correlation coefficients. *Statistics & Probability Letters, 80,* 1079–1084.

Zaigraev, A., & Kaniovski, S. (2012). Bounds on the competence of a homogeneous jury. *Theory and Decision, 72,* 89–112.

Zaigraev, A., & Kaniovski, S. (2013). A note on the probability of at least K successes in N correlated binary trials. *Operations Research Letters, 41,* 116–120.

Zaigraev, A., & Kaniovski, S. (2020). Inequality in population weights, majority threshold and the inversion probability in the case of three states. *Operations Research Letters, 48,* 732–736.

Strategic Voting and Strategic Candidacy

Markus Brill and Vincent Conitzer

Abstract Models of strategic candidacy analyze the incentives of candidates to run in an election. Most work on this topic assumes that strategizing only takes place among candidates, whereas voters vote truthfully. In this article, we extend the analysis to also include strategic behavior on the part of the voters. We also study cases where only candidates or only voters are strategic. We consider a setting in which both voters and candidates have single-peaked preferences and the voting rule is majority-consistent, and we analyze the type of strategic behavior that is required in order to guarantee desirable voting outcomes.

1 Introduction

When analyzing voting rules, the set of candidates is usually assumed to be fixed. In a pathbreaking paper, Dutta et al. (2001) have initiated the study of *strategic candidacy* by accounting for candidates' incentives to run in an election. They assumed that candidates have preferences over other candidates and defined a voting rule to be *candidate stable* if no candidate ever has an incentive not to run. In this model, it is assumed that every candidate prefers herself to all other candidates. Therefore, the winner of an election never has an incentive not to run. Non-winning candidates, on the other hand, might be able to alter the winner by leaving the election.[1] Dutta et al. (2001) showed that, under mild conditions, no non-dictatorial rule is candidate stable.

[1]Manipulating the result of an election by not participating in the voting process is reminiscent of the "no-show paradox" (Fishburn & Brams, 1983; Moulin, 1988), which is concerned with *voters'* participation incentives.

M. Brill (✉)
University of Warwick, Coventry, UK
e-mail: markus.brill@warwick.ac.uk

V. Conitzer
Carnegie Mellon University, Pittsburgh, USA
e-mail: conitzer@cs.cmu.edu

University of Oxford, Oxford, UK

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_5

This result naturally leads to the question of how voting outcomes are affected by candidates' incentives. It is straightforward to model strategic candidacy as a two-stage game. At the first stage, each candidate decides whether to run in the election or not. At the second stage, each voter casts a ballot containing a ranking of the running candidates. When analyzing this game, an important ingredient is the assumed voter behavior. That is, what assumptions are made about the votes in the second stage, conditional on the set of running candidates?

Most papers on strategic candidacy assume that voters vote truthfully, i.e., their reported ranking for any given subset of candidates corresponds to their true preferences, restricted to that subset (Dutta et al., 2001; Ehlers & Weymark, 2003; Eraslan & McLennan, 2004; Lang et al., 2013; Obraztsova et al., 2015, 2020; Polukarov et al., 2015; Rodríguez-Álvarez, 2006a, b; Samejima, 2005, 2007). However, it is well-known that this is an unrealistic assumption (Gibbard, 1973; Satterthwaite, 1975). It is therefore natural to account for strategic behavior on the part of the voters as well. Thus, in the model we consider, both candidates and voters act strategically.

The technical problem in accounting for strategic voting is that, generally speaking, too many voting equilibria exist (De Sinopoli, 2000; Myerson & Weber, 1993). If we only consider Nash equilibria, then any profile of votes for which no single voter can change the outcome is an equilibrium. In some cases, a straightforward refinement rules out many of the equilibria. For example, in a majority election between two candidates, it is natural to rule out the strange equilibria where some voters play the weakly dominated strategy of voting for their less preferred candidate. But this reasoning does not generally extend to more than two candidates. In this paper, we focus on a setting that admits natural equilibrium refinements.

Specifically, we consider single-peaked preferences (Black, 1948). It is well-known that, if the number of voters is odd, this domain restriction guarantees the existence of a Condorcet winner (namely, the most preferred candidate of the median voter) and admits a strategyproof and Condorcet-consistent voting rule (namely, the median rule) (Moulin, 1980). Dutta et al. (2001) observed that any Condorcet-consistent rule is candidate stable in any domain that guarantees the existence of a Condorcet winner. Lang et al. (2013) extended this result by showing that, in this setting, no *coalition* of candidates ever has an incentive to change their strategies as long as the Condorcet winner is running. We study the effect of strategic candidacy with single-peaked preferences when the voting rule is *not* Condorcet-consistent. Our motivation is that the voting rules that are most widely used in practice—*plurality*, *plurality with runoff*, and *single transferable vote (STV)*—may fail to select the Condorcet winner, even for single-peaked preferences. We consider the class of *majority-consistent* voting rules, which are rules that, if there is a candidate that is ranked first by more than half the voters, will select that candidate. This class includes all Condorcet-consistent rules, but also other rules such as plurality, plurality with runoff, STV, and Bucklin. For this class, we show that under some assumptions on strategic behavior, the Condorcet winner does in fact end up being elected (though for other assumptions this does not hold).

2 Related Work

Strategic candidacy was introduced by Dutta et al. (2001, 2002), who showed that every non-dictatorial voting rule might give candidates incentives not to run. Subsequently, Ehlers and Weymark (2003) and Samejima (2005) came up with alternative proofs and extensions of some of the results of Dutta et al. (2001). Mbih et al. (2009) and Ndiaye (2013) computed how frequent incentives not to run occur for specific voting rules under the impartial anonymous culture (IAC) assumption. Furthermore, models of strategic candidacy have been extended to set-valued (Eraslan & McLennan, 2004; Rodríguez-Álvarez, 2006a), probabilistic (Rodríguez-Álvarez, 2006b), and multiwinner (Obraztsova et al., 2020) voting rules.

Samejima (2007) studied strategic candidacy for single-peaked preferences and characterized the class of candidate stable voting rules for this domain. He showed that, under some mild conditions, a voting rule is candidate stable for single-peaked preferences if and only if it is a *kth leftmost peak rule* for some k. A kth leftmost peak rule fixes a single-peaked axis, identifies each voter with his most preferred candidate (his "peak"), and selects the peak of the kth leftmost voter according to the ordering given by the axis. The median rule is the special case for $k = \frac{|V|+1}{2}$, where $|V|$ denotes the number of voters, which is assumed to be odd.

Also related are two papers that precede (Dutta et al., 2001, 2002). Osborne and Slivinski (1996) and Besley and Coate (1997) study plurality equilibria in a candidacy game where all voters are potential candidates and running is costly. In both papers, preferences of voters and candidates are defined via a spatial model (which, in the one-dimensional case, yields single-peaked preferences). However, the focus of these two papers is different from ours: They are mainly interested in how the number and spatial position of candidates that run in equilibrium is affected by parameters such as entry costs, preferences, and candidates' utilities for winning. There is also a number of technical differences to our paper. For example, Osborne and Slivinski (1996) consider a continuum of voters and assume that voters vote truthfully. And Besley and Coate (1997) add a third stage to the two-stage candidacy game by letting the selected candidate choose a policy from a given policy space. None of the two papers considers strong equilibria.

Finally, a more recent line of research was initiated by Lang et al. (2013), who study for which voting rules the candidacy game admits pure equilibria under the assumption that voters vote truthfully. They also consider strong equilibria and show that, for every domain that guarantees the existence of a Condorcet winner and for every Condorcet-consistent voting rule, a set of running candidates forms a strong equilibrium if and only if the Condorcet winner is contained in the set. Their results have been extended by Polukarov et al. (2015), who study equilibrium dynamics in candidacy games, and by Obraztsova et al. (2015), who study candidacy games when running is costly.

3 Preliminaries

This section introduces the concepts and notations that are used in the remainder of the paper. For a finite set X, let $\mathcal{L}(X)$ denote the set of *rankings* (or linear orders) of X, where a ranking is a binary relation on X that is complete, transitive, and antisymmetric. For a ranking $r \in \mathcal{L}(X)$, $\text{top}(r)$ denotes the top-ranked element according to r and $r|_B$ denotes the restriction of r to a nonempty subset $B \subseteq X$.

3.1 Players and Preferences

Let C be a finite set of candidates and V a finite set of voters. Throughout this paper, we assume that $|V|$ is odd.[2] The set P of *players* is given by $P = C \cup V$. We assume that $C \cap V = \emptyset$.[3] Each player $p \in P$ has preferences over the set of candidates, given by a ranking $R_p \in \mathcal{L}(C)$. For all candidates $c \in C$, we assume that the top-ranked candidate in R_c is c itself.[4] For a player $p \in P$ and two candidates $a, b \in C$, we write $a \succeq_p b$ if $(a, b) \in R_p$ and $a \succ_p b$ if $a \succeq_p b$ and $a \neq b$.

A *preference profile* $R = (R_p)_{p \in P} \in \mathcal{L}(C)^P$ contains preferences for all players. For a preference profile R and a candidate c, let $V_R(c)$ denote the set of voters that have c as their top-ranked candidate, i.e., $V_R(c) = \{v \in V : \text{top}(R_v) = c\}$. Moreover, for a candidate $d \neq c$, let $V_R(c, d)$ denote the set of voters that prefer c to d, i.e., $V_R(c, d) = \{v \in V : c \succ_v d\}$. Candidate c is a *majority winner in* R if $|V_R(c)| > |V|/2$, and c is a *Condorcet winner in* R if $|V_R(c, d)| > |V|/2$ for all $d \in C \setminus \{c\}$. Note that both concepts ignore the preferences of candidates. Every preference profile can have at most one majority winner and at most one Condorcet winner. If a candidate is a majority winner in R, then this candidate is also a Condorcet winner in R. Finally, for a nonempty subset $B \subseteq C$ of candidates, we let $R|_B = (R_p|_B)_{p \in P} \in \mathcal{L}(B)^P$ denote the restriction of profile R to B.

---

[2] When the number of voters is even, a Condorcet winner is not guaranteed to exist even if preferences are single-peaked. However, in this case there will always be at least one *weak* Condorcet winner. The results in Sect. 5 extend to the setting with an even number of voters, with the role of the Condorcet winner taken over by one of the weak Condorcet winners, namely the one in whose favor the tie is broken.

[3] For results without this assumption, see, e.g., Dutta et al. (2001, 2002).

[4] This assumption is common in the literature on strategic candidacy, where it is often referred to as *self-preference* (Dutta et al., 2001, 2002) or *self-supporting candidate preferences* (Lang et al., 2013; Obraztsova et al., 2015, 2020; Polukarov et al., 2015). Without it, scenarios can arise where no candidate has an incentive to run.

3.2 Single-Peakedness

A well-studied structural restriction on preferences is single-peakedness (Black, 1958). Intuitively, preferences are single-peaked if the candidates can be ordered on a one-dimensional spectrum in such a way that every voter has an ideal (most preferred) point on this spectrum, and preference is declining when moving away from this ideal point. Settings in which the assumption of single-peakedness seems reasonable include elections in which candidates correspond to numerical values (e.g., voting over a tax rate) or elections in which the candidates can be assigned positions on a one-dimensional political spectrum (e.g., ranging from left-wing to right-wing political views). Our definition of a single-peaked preference profile requires not only the preferences of voters, but also the preferences of candidates to be single-peaked. The assumption that $\mathrm{top}(R_c) = c$ for every candidate c implies each candidate's ideal point coincides with their position on the spectrum.

Formally, let $\lhd \in C \times C$ be a strict ordering of the candidates. A preference profile $R = (R_p)_{p \in P}$ is *single-peaked with respect to* $\lhd$ if the following condition holds for all $a, b \in C$ and $p \in P$: if $a \lhd b \lhd \mathrm{top}(R_p)$ or $\mathrm{top}(R_p) \lhd b \lhd a$, then $b\ R_p\ a$. For a preference profile R that is single-peaked with respect to $\lhd$, the *median* of R is defined as the unique candidate c for which both $\sum_{a \in C : a \lhd c} |V_R(a)| < |V|/2$ and $\sum_{a \in C : c \lhd a} |V_R(a)| < |V|/2$. It is well known that the median is a Condorcet winner in R.

Let $c_1 \lhd c_2 \lhd \ldots \lhd c_m$ and let R be a preference profile that is single-peaked with respect to $\lhd$. The *peak distribution* of R with respect to $\lhd$ is the vector of length m whose jth entry is the number $|V_R(c_j)|$ of voters that rank c_j highest.

3.3 Voting Rules

A *voting rule* f maps a nonempty subset $B \subseteq C$ of candidates and a profile of votes $r = (r_v)_{v \in V} \in \mathcal{L}(B)^V$ to a candidate $f(B, r) \in B$. A voting rule f is *majority-consistent* if $f(B, (R_v|_B)_{v \in V}) = c$ whenever c is a majority winner in $R|_B$, and f is *Condorcet-consistent* if $f(B, (R_v|_B)_{v \in V}) = c$ whenever c is a Condorcet winner in $R|_B$. Because majority winners are always Condorcet winners, (perhaps confusingly) Condorcet-consistency implies majority-consistency.

A *scoring rule* is a voting rule that is defined by a sequence $(s^n)_{n \geq 1}$, where for each $n \in \mathbb{N}$, $s^n = (s^n_1, \ldots, s^n_n) \in \mathbb{R}^n$ is a *score vector* of length n. For a preference profile R on k candidates, the score vector s^k is used to allocate points to candidates: each candidate receives a score of s^k_j for each time it is ranked in position j by a voter. (Again, preferences of candidates are ignored.) The scoring rule then selects the candidate with maximal total score. In the case of a tie, a fixed tiebreaking ordering is used. Prominent examples of scoring rules are *plurality* $[s^n = (1, 0, \ldots, 0)]$, *Borda's rule* $[s^n = (n-1, n-2, \ldots, 0)]$, and *veto* $[s^n = (0, \ldots, 0, -1)]$.

The plurality winner is a candidate maximizing $|V_R(\cdot)|$. Plurality is majority-consistent, but not Condorcet-consistent. Borda's rule and veto are not majority-consistent and (hence) not Condorcet-consistent.

4 Game-Theoretic Model

We consider the following two-stage game. At the first stage, each candidate decides whether to run in the election or not. At the second stage, each voter casts a ballot containing a ranking of the running candidates. Throughout, we consider *complete-information* games: the preferences of the candidates and voters are common knowledge among the candidates and voters. Hence, we do not need to model games as (pre-)Bayesian and strategies do not have to condition on the player's type.

4.1 Strategies and Outcomes

Let S_p denote the set of (pure) strategies of player p. Then for each candidate $c \in C$, the set S_c is given by $\{0, 1\}$, with the convention that 1 corresponds to "running" and 0 corresponds to "not running." For each voter $v \in V$, the set S_v consists of all functions

$$s_v : 2^C \to \bigcup_{B \subseteq C} \mathcal{L}(B)$$

that map a subset $B \subseteq C$ of candidates to a ranking $s_v(B) \in \mathcal{L}(B)$. The interpretation is that $s_v(B)$ is the vote of voter v when the set of running candidates is B. In particular, each S_v contains a strategy that corresponds to *truthful voting* for voter v: this strategy maps every set B to the ranking $R_v|_B$. In general, however, a voter can rank two candidates differently depending on which other candidates run.

We are now ready to define the outcomes of the game. A *strategy profile* $s = (s_p)_{p \in P}$ contains a strategy for every player. Given a strategy profile s and a voting rule f, let $C(s) = \{c \in C : s_c = 1\}$ denote the set of running candidates[5] and let $r(s) = (s_v(C(s)))_{v \in V} \in \mathcal{L}(C(s))^V$ denote the votes cast for this set of running candidates. The outcome $o_f(s)$ of s under f is then given by $o_f(s) = f(C(s), r(s))$.

[5] If $C(s) = \emptyset$, define $o_f(s) = \top$. We assume that each candidate prefers herself to the outcome $\top$. This assumption ensures that at least one candidate will run whenever candidates act strategically.

4.2 Equilibrium Concepts

Let $s = (s_p)_{p\in P}$ be a strategy profile. For a subset $\tilde{P} \subseteq P$ and a profile of strategies $s'_{\tilde{P}} = (s'_p)_{p\in\tilde{P}}$ for players in $\tilde{P}$, let $(s'_{\tilde{P}}, s_{-\tilde{P}})$ denote the strategy profile where each player $p \in \tilde{P}$ plays strategy s'_p and all remaining players play the same strategy as in s. Fix a voting rule f and a preference profile R. For a strategy profile s and a subset $\tilde{P} \subseteq P$ of players, say that s is (R, f)*-deviation-proof w.r.t.* $\tilde{P}$ if for all $s'_{\tilde{P}}$, there exists $p \in \tilde{P}$ such that

$$o_f(s) \succeq_p o_f(s'_{\tilde{P}}, s_{-\tilde{P}}).$$

For a strategy profile $s = (s_p)_{p\in P}$, we sometimes write $s = (s_C, s_V)$, where $s_C = (s_c)_{c\in C}$ is the profile of candidate strategies and $s_V = (s_v)_{v\in V}$ is the profile of voter strategies. We can now define equilibrium behavior for both candidates and voters.

Definition 1 Let R be a preference profile and let f be a voting rule. A strategy profile $s = (s_C, s_V)$ is

- a C*-equilibrium for* R *under* f if s is (R, f)-deviation-proof w.r.t. $\{c\}$ for all $c \in C$;
- a *strong* C*-equilibrium for* R *under* f if s is (R, f)-deviation-proof w.r.t. C' for all $C' \subseteq C$;
- a V*-equilibrium for* R *under* f if for every $s'_C \in \{0, 1\}^C$, (s'_C, s_V) is (R, f)-deviation-proof w.r.t. $\{v\}$ for all $v \in V$;
- a *strong* V*-equilibrium for* R *under* f if for every $s'_C \in \{0, 1\}^C$, (s'_C, s_V) is (R, f)-deviation-proof w.r.t. V' for all $V' \subseteq V$.

We omit the reference to R and f if the preference profile or the voting rule is known from the context. In a C-equilibrium, no candidate can achieve a more preferred outcome by unilaterally changing their strategy. In a strong C-equilibrium, no coalition of candidates can change the outcome in such a way that every player in the coalition prefers the new outcome to the original one. Thus, (strong) C-equilibria correspond to (strong) Nash equilibria when strategies of voters are assumed to be fixed. For voters, the equilibrium notions are more demanding: In order to be considered a (strong) V-equilibrium, the strategies of voters are required to form a (strong) Nash equilibrium for every subset $B \subseteq C$ of running candidates.

It is instructive to relate these definitions to established game-theoretic solution concepts for extensive-form games, such as subgame-perfect equilibrium and subgame-perfect strong equilibrium. A strategy profile s is a *subgame-perfect equilibrium* of a game G if for any subgame G' of G, the restriction of s to G' is a Nash equilibrium of G', and it is a *subgame-perfect strong equilibrium* if for any subgame G' of G, the restriction of s to G' is a strong Nash equilibrium of G'. In the candidacy game, every subgame (other than the game itself) corresponds to a voting game that takes place after the candidates have decided whether or not to run. Thus, a proper subgame can be identified with the set of candidates that run in this subgame.

For candidates, playing a subgame-perfect equilibrium is not a stronger requirement than playing a Nash equilibrium, because the only subgame in which they play is the entire game itself. For voters, on the other hand, playing a subgame-perfect equilibrium entails playing a Nash equilibrium for every possible set of running candidates. Therefore, we have the following.

Fact A strategy profile is a subgame-perfect equilibrium of the candidacy game if and only if it is both a C-equilibrium and a V-equilibrium. □

For subgame-perfect *strong* equilibria, one implication is straightforward.

Fact Every subgame-perfect strong equilibrium of the candidacy game is both a strong C-equilibrium and a strong V-equilibrium. □

However, the other direction does not hold in general, because even if coalitions of either one type of players cannot successfully deviate, it is possible that a mixed coalition including players of both types can.

Splitting up the equilibrium definitions into separate requirements for C and V allows us to capture scenarios in which only players of one type (candidates or voters) act according to the corresponding equilibrium notion. In Sect. 5 we will analyze which combinations of equilibrium notions yield desirable outcomes. We will present both positive results, stating that a desirable outcome will be selected whenever a strategy profile meets a certain combination of equilibrium conditions, and negative results, stating that undesirable outcomes may be selected even if certain equilibrium conditions hold.

In sufficiently general settings, the existence of solutions is not guaranteed for any of the equilibrium concepts in Definition 1.[6] However, for all our positive results, we also show that every preference profile admits a strategy profile that meets the corresponding equilibrium conditions.

5 Results

We are going to assume that preference profiles are single-peaked with respect to a given order $\lhd$.[7] Note that our definition of single-peakedness in Sect. 3.1 also requires the preferences of *candidates* to be single-peaked with respect to $\lhd$. Given that the preferences of voters are single-peaked with respect to $\lhd$, this does not appear to be an unreasonable assumption.

We are interested in the following question: which requirements on the strategies of players are sufficient for the Condorcet winner (which is guaranteed to exist)

[6] Subgame-perfect equilibria are guaranteed to exist if one allows for *mixed strategies* and extends the preferences of players to the set of all probability distributions over $C \cup \{\top\}$ in an appropriate way.

[7] If the order is not part of the input, it can be computed in polynomial time (Bartholdi & Trick, 1986; Escoffier et al., 2008).

to be the outcome? For Condorcet-consistent rules, the answer to this question is relatively straightforward (Lang et al., 2013). However, as we have argued in the introduction, most rules that are typically used in practice are majority-consistent, but not Condorcet-consistent. The simplest and most important such rule is plurality.

It is easy to construct a plurality election in which some candidates have an incentive not to run (assuming truthful voting).

Example 1 Consider a single-peaked preference profile with candidates $a \lhd b \lhd sc$ and peak distribution[8] $(3, 2, 4)$. Under truthful voting, the plurality winner is c. However, if candidate a does not run, the three voters in $V_R(a)$ rank candidate b first, making b the plurality winner. By single-peakedness, candidate a prefers b to c.

This example also shows that plurality can fail to select the Condorcet winner when all candidates run and all voters vote truthfully. The next example shows that requiring both candidates and voters to play subgame-perfect equilibrium strategies is still not sufficient for the Condorcet winner to be chosen.

Example 2 Consider a single-peaked preference profile with candidates $a \lhd b \lhd c \lhd d \lhd e$ and peak distribution $(11, 3, 3, 3, 3)$. The Condorcet winner is b. Let s be the strategy profile in which all candidates run and all voters vote truthfully. Then $o_{\text{plurality}}(s) = a$ and no candidate other than a can change that outcome by unilaterally deviating. Therefore, s is a C-equilibrium. To see that s is also a V-equilibrium, we need to check that "truthful voting" is deviation-proof for every subset of running candidates. Deviation-proofness clearly holds whenever at most two candidates run. If at least three candidates run, single-peakedness implies that the leftmost running candidate has a plurality score of at least 11, whereas each other running candidate has a score of at most 9. Thus, no voter can change the outcome by unilaterally deviating.

We go on to show that the Condorcet winner *will* be chosen if we require stronger equilibrium notions. We first analyze strong V-equilibria. Note that this result does not require single-peaked preferences.[9]

Theorem 1 *Let R be a preference profile with Condorcet winner c^* and let f be a majority-consistent voting rule.*

(i) If $R|_B$ has a Condorcet winner for every nonempty subset $B \subseteq C$, then there exists a subgame-perfect strong equilibrium (and hence a strategy profile that is both a strong C-equilibrium and a strong V-equilibrium) for R under f in which all candidates run.

(ii) If s is a strong V-equilibrium for R under f with $s_{c^*} = 1$, then $o_f(s) = c^*$.

---

[8] We often simplify examples with single-peaked preference profiles by specifying the peak distribution only. This piece of information is clearly sufficient to identify both the Condorcet winner and, in the absence of ties, the plurality winner.

[9] In particular, note that Theorem 1 does not make any assumptions on the preferences of candidates (other than self-supportedness).

Proof For (i), denote by $c_B \in B$ the Condorcet winner in $R|_B$. Let s be a strategy profile where all candidates run and all voters rank c_B first whenever the set of running candidates is given by B. Hence, $o_f(s) = c^*$. We claim that s is a subgame-perfect strong equilibrium for R under f. In order to prove this claim, we need to show that for every subgame, there is no beneficial deviation for any coalition.

First, consider the subgame that is given by the entire game itself. Suppose, for the sake of contradiction, that there is a coalition $\tilde{P} = C' \cup V'$ of candidates and voters that can change the outcome to some $a \neq c^*$ and that all players in $\tilde{P}$ prefer a to c^*. Let $s' = (s'_{\tilde{P}}, s_{-\tilde{P}})$ denote the strategy profile that results from this deviation. Observe that $c^* \notin \tilde{P}$, because $c^* \succ_{c^*} a$. Therefore, c^* is still running under s' and all non-deviating voters $V \setminus V'$ still rank c^* highest under s'. That means that the number $|V'|$ of deviating voters has to be greater than $|V|/2$, as otherwise majority-consistency of f would yield $o_f(s') = c^*$. But then V' is a majority of voters, each preferring a over c^*. This contradicts the assumption that c^* is a Condorcet winner.

Second, consider a subgame that arises after the candidates have chosen whether or not to run. Let $B \subseteq C$ be the set of candidates that run in this subgame. If $B = \emptyset$, the outcome is $\top$ and no coalition of voters can change the outcome. If $B \neq \emptyset$, all voters rank c_B first by the definition of s. By an argument analogous to the one above, the existence of a successfully deviating coalition of voters would violate the assumption that c_B is the Condorcet winner in $R|_B$. Therefore, s is a subgame-perfect strong equilibrium.

For (ii), let s be a strong V-equilibrium for R under f with $s_{c^*} = 1$. Assume for the sake of contradiction that $o_f(s) = a \neq c^*$. We will show that s is not a strong V-equilibrium, by means of the following deviation. Let $\tilde{P} = V_R(c^*, a)$ be the set of voters that prefer c^* over a and let $s' = (s'_{\tilde{P}}, s_{-\tilde{P}})$ be the strategy profile in which all voters in $\tilde{P}$ rank c^* first whenever c^* runs. Since $s_{c^*} = 1$ and $|\tilde{P}| = |V_R(c^*, a)| > |V|/2$, majority-consistency of f implies $o_f(s') = c^*$. Moreover, $c^* \succ_p a$ for all $p \in \tilde{P}$ by definition of $\tilde{P}$. Therefore, s is not (R, f)-deviation-proof w.r.t. $\tilde{P}$, contradicting the assumption that s is a strong V-equilibrium. $\square$

The following example illustrates the proof of part (ii).

Example 3 Let R be a single-peaked preference profile with candidates $a \lhd b \lhd c \lhd d$ and peak distribution $(2, 1, 2, 4)$. The Condorcet winner is c. Consider the strategy profile s in which all candidates run and all voters vote truthfully. Then $o_{\text{plurality}}(s) = d$. If all voters in $V_R(c, d) = V_R(a) \cup V_R(b) \cup V_R(c)$ deviate and rank c first, the outcome changes to c.

We remark that part (ii) of Theorem 1 can be generalized[10] by observing that it is sufficient for f to satisfy the following condition, which is considerably weaker than majority-consistency:

---

[10] Sertel and Sanver (2004) prove a similar result in the (standard) setting where all candidates are assumed to run. A further strengthening of part (ii) of Theorem 1 was pointed out to us by François Durand: instead of requiring that voters play a strong V-equilibrium for every subset of running candidates, it is sufficient to require voters to play a strong V-equilibrium *only in those subgames that actually allow strong V-equilibria* (and to not make any assumptions on voter behavior otherwise).

> Whenever a set $V' \subseteq V$ of voters forms a majority (i.e., $|V'| > |V|/2$), then for every candidate $a \in C$ that is running and every profile of votes for voters in $V \setminus V'$, the voters in V' can vote in such a way that candidate a is chosen.

It can be shown that all unanimous C2 functions (Fishburn, 1977) satisfy this property.

The following corollary summarizes the consequences of Theorem 1 for single-peaked preference profiles.

Corollary 1 *Let R be a single-peaked preference profile with Condorcet winner c^* and let f be a majority-consistent voting rule.*

(i) There exists a subgame-perfect strong equilibrium (and hence a strategy profile that is both a strong V-equilibrium and a strong C-equilibrium) for R under f.
(ii) If s is a strong V-equilibrium and a C-equilibrium (strong or not) for R under f, then $o_f(s) = c^*$.

We provide two examples to show that the statements of Corollary 1 do not hold for rules that are not majority-consistent.

Example 4 Let R be a single-peaked preference profile with candidates $a \lhd b \lhd c$ and peak distribution $(5, 0, 4)$. If f is Borda's rule, there does not exist a strong V-equilibrium (and hence no subgame-perfect strong equilibrium). To see this, consider the case where all candidates run. Observe that in any strong V-equilibrium, the outcome would have to be a. (Suppose the outcome is not a. Then, the five voters in $V_R(a)$ can jointly deviate and change the outcome to a. They can do this by having one voter voting $a \succ b \succ c$, and the remaining four voters voting exactly the opposite rankings of the voters in $V_R(c)$.) However, there is no strong V-equilibrium that yields outcome a. This is because the voters in $V_R(c)$ prefer both other alternatives to a, and—no matter how the voters in $V_R(a)$ vote—the voters in $V_R(c)$ can jointly deviate and achieve an outcome other than a. (One of b and c will obtain a score of at least 3 from the voters in $V_R(a)$. Without loss of generality, suppose it is b. Then the voters in $V_R(c)$ can all vote $b \succ c \succ a$, making b win.)

Example 5 Let R be a single-peaked preference profile with candidates $a \lhd b \lhd c$ and five voters: three voters have preferences $a \succ b \succ c$ and two voters have preferences $b \succ c \succ a$. The Condorcet winner is a. Let f be the voting rule veto[11] and let s be the strategy profile where all candidates run and all voters vote truthfully. Then, $o_f(s) = b$. Moreover, s is a strong C-equilibrium and a strong V-equilibrium. The former holds because any deviation involving a does not change the outcome (provided b still runs), and c can only change the outcome to the less preferred alternative a. For the latter, the only interesting case is when all three candidates run. In this case, the two voters in $V_R(b)$ have no incentive to deviate from truthful voting (their favorite candidate is winning) and there is no way for the three voters in $V_R(a)$ to jointly deviate and achieve outcome a. (They can change the outcome to c by voting $a \succ c \succ b$, but they prefer b to c.) It can furthermore be shown that, when all candidates run, *every* strong V-equilibrium yields outcome b.

[11] Veto does not only violate majority-consistency, but also the weaker property defined after Theorem 1.

Theorem 2 *Let R be a single-peaked preference profile with Condorcet winner c^* and let f be a majority-consistent voting rule.*

(i) There exists a strong C-equilibrium for R under f where all voters vote truthfully.
(ii) If s is a strong C-equilibrium for R under f where all voters vote truthfully, then $o_f(s) = c^*$.

Proof For (i), let s be the strategy profile in which only c^* runs and all voters vote truthfully. We show that this is a strong C-equilibrium for R under f. Suppose, for the sake of contradiction, that $\tilde{C} \subseteq C$ is a coalition of candidates that can, by changing its strategies, make alternative $a \neq c^*$ win, and moreover that all candidates in $\tilde{C}$ prefer a to c^*. Define $C^- = \{c \in C : c \lhd c^*\}$ and $C^+ = \{c \in C : c^* \lhd c\}$, and without loss of generality suppose that $a \in C^-$. Because candidates' preferences are single-peaked and they rank themselves first, it follows that $\tilde{C} \subseteq C^-$. But this implies that still, no candidate in C^+ runs. Hence, all voters with $\text{top}(R_v) \in C^+ \cup \{c^*\}$ still rank c^* first (since they vote truthfully). Since c^* is the median, $|C^+ \cup \{c^*\}| > |V|/2$, and majority-consistency of f implies that c^* wins. This gives us the desired contradiction.

For (ii), let s be a strong C-equilibrium for R under f where all voters vote truthfully. Consider the set $C(s)$ of candidates that are running under s. Define $C_s^- = \{c \in C(s) : c \lhd c^*\}$ and $C_s^+ = \{c \in C(s) : c^* \lhd c\}$. Assume for the sake of contradiction that $o_f(s) = a \neq c^*$. Without loss of generality, suppose that $a \in C_s^-$. Consider the set $\tilde{C} = C_s^+ \cup \{c^*\}$. Define $s'_{\tilde{C}} = (s'_c)_{c \in \tilde{C}}$ by

$$s'_c = \begin{cases} 1 & \text{if } c = c^* \\ 0 & \text{if } c \in C_s^+ \end{cases}$$

and observe that $o_f(s'_{\tilde{C}}, s_{-\tilde{C}}) = c^*$. The reason for the latter is that (1) the set of voters v with $\text{top}(R_v) = c^*$ or $c^* \lhd \text{top}(R_v)$ forms a majority, (2) all of these voters satisfy $\text{top}(R_v|_{C(s'_{\tilde{C}}, s_{-\tilde{C}})}) = c^*$, and (3) all voters vote truthfully by assumption. Moreover, single-peakedness implies that all candidates in $\tilde{C}$ prefer c^* to a. Therefore, s is not (R, f)-deviation-proof w.r.t. $\tilde{C}$, contradicting the assumption that s is a strong C-equilibrium. □

Example 6 Consider again the preference profile R and the strategy profile s from Example 3. If both a and b deviate to "not running," the outcome (under plurality) changes from d to c. Therefore, s is not a strong C-equilibrium.

Similar to the case of Theorem 1, we now provide examples that show that Theorem 2 cannot be generalized in certain ways. Example 7 shows that Theorem 2 does not hold for Borda's rule (which is not majority-consistent), and Example 8 shows that Theorem 2 does not hold if the preferences of candidates are not single-peaked.

Example 7 Consider a single-peaked preference profile with candidates $a \lhd b \lhd c$ and five voters: three voters have preferences $a \succ b \succ c$ and two voters have preferences $b \succ c \succ a$. The Condorcet winner is a. Let s be the strategy profile where all candidates run and all voters vote truthfully. It is easily verified that s is a strong C-equilibrium and $o_{\text{Borda}}(s) = b$. In fact, it can be checked that the Condorcet winner is not chosen in *any* strong C-equilibrium with truthful voting. (The only other strong C-equilibrium under truthful voting has candidates b and c running and also yields outcome b.)

Example 8 Consider the following preference profile with candidates a, b, c and 14 voters: four voters have preferences $a \succ b \succ c$, four voters have preferences $b \succ a \succ c$, and six voters have preferences $c \succ b \succ a$. The preferences of the candidates are such that a prefers c over b and b prefers c over a. Whereas the preferences of the voters are single-peaked with respect to the ordering $a \lhd b \lhd c$, this is not true for the preferences of the candidates. (Therefore, this profile is not single-peaked according to the definition in Sect. 3.1.) The Condorcet winner is b and the Condorcet loser is c. Let s be the strategy profile where all candidates run and all voters vote truthfully. It is easily verified that s is a strong C-equilibrium and $o_{\text{plurality}}(s) = c$. In fact, "everybody running" is the only strong C-equilibrium under truthful voting.

Since Theorem 1 already covers the case where *both* voters and candidates play a strong (subgame-perfect) equilibrium, only one case is left to consider: candidates playing a strong C-equilibrium, and voters merely playing a V-equilibrium. The following example shows that these requirements are *not* sufficient for the Condorcet winner to be chosen.

Example 9 Consider a single-peaked preference profile with candidates $a \lhd b \lhd c$ and peak distribution $(1, 1, 1)$. The Condorcet winner is b. Let s be a strategy profile with $s_c = 1$ and voter strategies s_v that satisfy

$$\text{top}(s_v(B)) = \begin{cases} c & \text{if } c \in B \\ \text{top}(R_v|_B) & \text{otherwise} \end{cases}$$

for each $B \subseteq C$. That is, all three voters rank c first whenever c runs, and vote truthfully otherwise.[12] Obviously, $o_{\text{plurality}}(s) = c$. We claim that s is both a V-equilibrium and a strong C-equilibrium. For the former, we distinguish two cases: If c runs, then all voters rank c first and no voter can change the outcome by unilaterally deviating. If c does not run, then at most two candidates run and no voter can benefit by voting for their less preferred candidate. For the latter, no coalition of candidates can change the outcome in such a way that all members of the coalition prefer the new outcome to c. (Such a coalition would need to include candidate c, who has no incentive to deviate.)

[12] Note that the voter v with $\text{top}(R_v) = a$ plays a weakly dominated strategy, because c is her least preferred alternative. This can be avoided by introducing a fourth candidate d with $c \lhd d$ and $V_R(d) = \emptyset$.

Table 1 Overview of results

	Strong V-equilibrium	V-equilibrium	Truthful voting
Strong C-equilibrium	Yes (Corollary 1)	No (Example 9)	Yes (Theorem 2)
C-equilibrium	Yes (Corollary 1)	No (Example 2)	No (Example 2)
Naive candidacy ($s_c = 1 \ \forall c$)	Yes (Theorem 1)	No (Example 2)	No (Examples 1 and 2)

A table entry is "yes" if every strategy profile that satisfies the corresponding row and column conditions yields the Condorcet winner under every majority-consistent voting rule. Moreover, for every "yes" entry, a strategy profile satisfying the conditions is guaranteed to exist

The phenomenon illustrated in this example is perhaps somewhat surprising: Assuming that candidates play a strong C-equilibrium, both truthful voting and strong V-equilibrium voting yields the desirable outcome; however, V-equilibrium voting—a notion of sophistication that might appear to be "in between" the other two notions—does not. Table 1 summarizes our results.

6 Conclusion

We have analyzed the combination of strategic candidacy and strategic voting in a single-peaked voting setting. It would be worthwhile to study whether (some of) our positive results extend to settings where preferences are single-peaked on a tree (Demange, 1982), single-peaked on a circle (Peters & Lackner, 2020), or single-peaked in higher dimensions (Sui et al., 2013). It would also be interesting to check whether similar results can be obtained for related domain restrictions such as single-crossing (Elkind et al., 2017; Roberts, 1977) or value-restricted (Sen, 1966) preferences.

Our positive results rely on finding the right level of equilibrium refinement (strong V-equilibrium, or strong C-equilibrium with truthful voting). If we move away from restricted domains, an interesting question is whether there are other types of equilibrium refinement (Dutta & Laslier, 2010; Obraztsova et al., 2014; Thomson et al., 2013) that allow us to arrive at meaningful equilibria by ruling out "unnatural" ones.

Equilibrium *dynamics* (Meir et al., 2010; Polukarov et al., 2015) is another topic for future research. For example, in the setting with single-peaked preferences and a majority-consistent rule, are there natural dynamics that are guaranteed to lead us to an equilibrium choosing the Condorcet winner?

On a higher level, one might wonder to what extent the phenomena exhibited in strategic candidacy games can be related to other problems that involve altering the set of candidates, such as control problems (Lang et al., 2013), cloning (Tideman, 1987), and nomination of alternatives (Dutta, 1981; Dutta & Pattanaik, 1978).

Acknowledgements A previous version of this article has appeared in the Proceedings of the 29th AAAI Conference on Artificial Intelligence (Brill & Conitzer, 2015). We would like to thank John Duggan, François Durand, Alexander Mayer, and the anonymous reviewers for helpful comments. This work was supported by NSF and ARO under grants CCF-1101659, IIS-0953756, CCF-1337215, W911NF-12-1-0550, and W911NF-11-1-0332, by the Deutsche Forschungsgemeinschaft under grant BR 4744/2-1, and by a Feodor Lynen research fellowship of the Alexander von Humboldt Foundation.

References

Bartholdi, J., III., & Trick, M. (1986). Stable matching with preferences derived from a psychological model. *Operations Research Letters, 5,* 165–169.

Besley, T., & Coate, S. (1997). An economic model of representative democracy. *Quarterly Journal of Economics, 112,* 85–114.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy, 56,* 23–34.

Black, D. (1958). *The theory of committees and elections*. Cambridge University Press.

Brill, M., & Conitzer, V. (2015). Strategic voting and strategic candidacy. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)* (pp. 819–826).

De Sinopoli, F. (2000). Sophisticated voting and equilibrium refinements under plurality rule. *Social Choice and Welfare, 17,* 655–672.

Demange, G. (1982). Single-peaked orders on a tree. *Mathematical Social Sciences, 3,* 389–396.

Dutta, B. (1981). Individual strategy and manipulation of issues. *Mathematical Social Sciences,1,* 169–176.

Dutta, B., Jackson, M. O., & Le Breton, M. (2001). Strategic candidacy and voting procedures. *Econometrica, 69,* 1013–1037.

Dutta, B., Jackson, M. O., & Le Breton, M. (2002). Voting by successive elimination and strategic candidacy. *Journal of Economic Theory, 103,* 190–218.

Dutta, B., & Laslier, J.-F. (2010). Costless honesty in voting. Presented at the 10th International Meeting of the Society for Social Choice and Welfare.

Dutta, B., & Pattanaik, P.-K. (1978). Strategic manipulation of issues in group decision making. In P.-K. Pattanaik (Ed.), *Strategy and group choice* (pp. 168–183). North-Holland.

Ehlers, L., & Weymark, J. A. (2003). Candidate stability and nonbinary social choice. *Economic Theory, 22,* 233–243.

Elkind, E., Lackner, M., & Peters, D. (2017). Structured preferences. In U. Endriss (Ed.), *Trends in computational social choice* (pp. 187–207). AI Access.

Eraslan, H., & McLennan, A. (2004). Strategic candidacy for multivalued voting procedures. *Journal of Economic Theory, 117,* 29–54.

Escoffier, B., Lang, B., & Öztürk, M. (2008). Single-peaked consistency and its complexity. In *Proceedings of the 18th European Conference on Artificial Intelligence (ECAI)* (pp. 366–370).

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

Fishburn, P. C., & Brams, S. J. (1983). Paradoxes of preferential voting. *Mathematics Magazine, 56,* 207–214.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica, 41,* 587–601.

Lang, J., Maudet, N., & Polukarov, M. (2013). New results on equilibria in strategic candidacy. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT)* (pp. 13–25).

Mbih, B., Moyouwou, I., & Ndiaye, A. A. (2009). Parliamentary voting rules and strategic candidacy. *Economics Bulletin, 29,* 1371–1379.

Meir, R., Polukarov, M., Rosenschein, J. S., & Jennings, N. R. (2010). Convergence to equilibria in plurality voting. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI)* (pp. 823–828).

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice, 35,* 437–455.
Moulin, H. (1988). Condorcet's principle implies the no show paradox. *Journal of Economic Theory, 45,* 53–64.
Myerson, R. B., & Weber, R. J. (1993). A theory of voting equilibria. *American Political Science Review, 87,* 102–114.
Ndiaye, A. A. (2013). Forte stabilité et vainqueur de Condorcet. *Revue d'Economie Politique, 123,* 793–808.
Obraztsova, S., Elkind, E., Polukarov, M., & Rabinovich, Z. (2015). Strategic candidacy games with lazy candidates. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 610–616).
Obraztsova, S., Markakis, E., & Thomson, D. R. M. (2014). Plurality voting with truth-biased agents. In *Proceedings of the 6th International Symposium on Algorithmic Game Theory (SAGT)* (pp. 26–37).
Obraztsova, S., Polukarov, M., Elkind, E., & Grzesiuk, M. (2020). Multiwinner candidacy games. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 957–965).
Osborne, M. J., & Slivinski, A. (1996). A model of political competition with citizen-candidates. *Quarterly Journal of Economics, 111,* 65–96.
Peters, D., & Lackner, M. (2020). Preferences single-peaked on a circle. *Journal of Artificial Intelligence Research, 68,* 463–502.
Polukarov, M., Obraztsova, S., Rabinovich, Z., Kruglyi, A., & Jennings, N. R. (2015). Convergence to equilibria in strategic candidacy. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 624–630).
Roberts, K. W. S. (1977). Voting over income tax schedules. *Journal of Public Economics, 8,* 329–342.
Rodríguez-Álvarez, C. (2006a). Candidate stability and voting correspondences. *Social Choice and Welfare, 27*, 545–570.
Rodríguez-Álvarez, C. (2006b). Candidate stability and probabilistic voting procedures. *Economic Theory, 27*, 657–677.
Samejima, Y. (2005). Strategic candidacy, monotonicity, and strategy-proofness. *Economic Letters, 88,* 190–195.
Samejima, Y. (2007). Strategic candidacy and single-peakedness. *Japanese Economic Review, 58,* 423–442.
Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory, 10,* 187–217.
Sen, A. K. (1966). A possibility theorem on majority decisions. *Econometrica, 34,* 491–499.
Sertel, M. R., & Sanver, M. R. (2004). Strong equilibrium outcomes of voting games are the generalized Condorcet winners. *Social Choice and Welfare, 22,* 331–347.
Sui, X., Francois-Nienaber, A., & Boutilier, C. (2013). Multi-dimensional single-peaked consistency and its approximations. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 375–382).
Thomson, D. R. M., Lev, O., Leyton-Brown, K., & Rosenschein, J. (2013). Empirical analysis of plurality election equilibria. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 391–398).
Tideman, T. N. (1987). Independence of clones as a criterion for voting rules. *Social Choice and Welfare, 4,* 185–206.

Meta-agreement and Rational Single-Peaked Preferences

Olivier Roy and Maher Jakob Abou Zeid

Abstract We revisit the claim that rationality requires participants in deliberation to form single-peaked preferences once they have reached meta-agreements. We provide two different arguments that cast doubts on this claim. The first points out the rationality of having non-single-peaked preferences in cases where consuming two goods together is less valuable than consuming each of them individually. The second argument fleshes out the notion of meta-agreements in terms of reasons supporting a particular structuring dimension. These arguments show that to the extent that deliberation fosters the formation of meta-agreements and the formation of single-peaked preferences, the bridge between these two notions might not be solely a matter of rational preference formation.

1 Introduction

This paper is a philosophical contribution to the question of whether deliberation helps avoid Condorcet cycles, and, more generally, incoherent social preferences. According to what has been called the *Received View* (Rafiee Rad & Roy, 2021), deliberation has this positive effect. The view goes back at least to Miller (1992), but its most standard formulation rests on the so-called *meta-agreement hypothesis*, as articulated by List (2002) and Dryzek and List (2003). In a nutshell, the hypothesis states that deliberation fosters the formation or the discovery of underlying meta-agreements, and that once such meta-agreements are in place rationality requires the participants to form single-peaked preferences.

This paper raises doubts regarding one important aspect of the meta-agreement hypothesis, namely that, in the presence of meta-agreement, rationality requires the agents to form single-peaked preferences. We do so in two ways. We first point

O. Roy (✉)
Department of Philosophy, University of Bayreuth, Bayreuth, Germany
e-mail: olivier.roy@uni-bayreuth.de

M. J. Abou Zeid
European Commission, Brussels, Belgium
e-mail: maher-jakob.abou-zeid@ec.europa.eu

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_6

out the rationality of having non-single-peaked preferences in cases where consuming two goods together is less valuable than consuming each of them individually. We then flesh out the notion of meta-agreements in terms of reasons supporting a particular structuring dimension, and argue that there are cases where non-single-peaked preferences are permissible given a particular set of reasons. In Sect. 2 we provide the required theoretical background to our argument, introducing the meta-agreement hypothesis and single-peaked preference, and develop the counterexamples in Sect. 3.

2 The Meta-agreement Hypothesis

Black (1948) and later Arrow (1963) remarked that *single-peaked preferences* are sufficient to ensure coherent group preferences. More precisely, Black showed that pairwise majority voting always delivers a Condorcet winner when the input preference profile is single-peaked, and Arrow noticed that this aggregation method satisfies all of his other axioms once universal domain is narrowed to single-peaked preferences. See Gaertner (2001) and Puppe (2018) for comprehensive overviews of such domain conditions.

Single-peakedness is defined relative to a given ranking of the alternatives, often called a *dimension*. Informally, a particular strict ranking over a finite set of alternatives is single-peaked with respect to a given dimension whenever, as one moves away from the most preferred alternative to the left and the right along the given dimension, one always moves to strictly less preferred alternatives. Visually, one always goes "down" and never "up" again in the preference ranking. A profile of rankings is said to be single-peaked when there is a dimension along which all rankings in that profile are single-peaked.

When can the preferences of the voters be expected to be single-peaked? The Received View answers this question by bringing in the crucial notion of *meta-agreements*. A meta-agreement is an agreement regarding the structuring dimension of a particular decision problem. This is the dimension along which the participants conceptualize the problem. Later on we make this idea more precise in terms of the possible reasons that bear on the decision. For now, it suffices to point out that many such dimensions are possible. The classic example of a structuring dimension is the left-right alignments of political parties in an election. That alignment is, of course, not the only one possible. Parties might as well be aligned according to their authoritarian *vs* libertarian values. A group has formed a meta-agreement on a question when its members agree on the most relevant dimension to the problem at hand. In our example, the voters are said to have reached a meta-agreement when they agree that, for instance, the election's central issue is the choice between authoritarian and libertarian values.

Meta-agreements leave room for substantial forms of what might be called substantive disagreements. Even if the voters agree that a given election should be primarily framed as a choice between libertarian and authoritarian values, they might disagree on the extent to which specific parties embody these values. And even if they agree on that, they might have opposing views regarding the best way to strike the compromise, if at all, between authoritarian and libertarian tendencies.

Single-peakedness, even if it implies the existence of a structuring dimension, is not sufficient for meta-agreements. The problem is, as List observes, that this dimension "is only a *formal* structural condition on individual preferences" (List, 2002 [our emphasis]). It might not be meaningful for the participants (Aldred, 2004) and, if it is, they might not agree that they should frame the problem at hand in its terms. In other words, the fact that the voters' preferences happen to be single-peaked does not necessarily entail a joint conceptualization of the decision at hand along a given dimension, which is central for meta-agreements. Returning to our example, two voters might completely disagree on the main issue of a specific election. One regards it as a choice between more left-wing or right-wing policies, while the other views it as a choice between authoritarian or libertarian values. If these two dimensions happen to be sufficiently correlated among the parties, these two voters might nonetheless have single-peaked preferences, or even reach consensus, without having reached a meta-agreement.

However, the question remains whether the existence of meta-agreements is conducive to the participants forming single-peaked preferences, and this is where the meta-agreement hypothesis comes in. The hypothesis is, in effect, a proposal regarding the mechanism through which meta-agreements foster the creation of single-peaked preferences. In the version outlined by List (2002) and Dryzek and List (2003), that mechanism works in three steps. The first step is the actual formation of the meta-agreement. Deliberation, they claim, helps the participants to identify or unveil the set of norms and values that constitute the relevant dimension(s) of the problem at hand, c.f. also Dryzek and Niemeyer (2006). This could be the trade-off between authoritarian and libertarian values in our running example. In a second step, deliberation helps the participants to agree on the factual question of how the alternatives compare on that dimension. That is, it helps them to rank the alternatives along that dimension. Again, in our example, deliberation is claimed to help the participants situate each party on the libertarian/authoritarian dimension. Note that the resulting ordering of the parties is neither an individual nor a social preference ranking. It only reflects how the alternatives compare to one another with respect to the structuring dimension. This leaves open which point on that dimension is optimal or best.

The third and final step is our main focus in this paper. Each participant should individually determine what point in the agreed dimension they find best and order the other alternatives relative to it. This is where deliberation should translate meta-agreement into single-peaked preferences (Dryzek & List, 2003; List, 2002). They suggest that rationality requires the participant to form single-peaked preferences with respect to the structuring dimension that they agree on. In other words, the claim is that if a participant has agreed that the dimension identified in the first step is indeed the *only* relevant one, and she also agreed on how each alternative fares on that dimension, then she must, on pain of incoherence, be able to identify the best alternative(s) along that dimension and order the others according to their distance from them. In our example, this means that once each voter has agreed that the trade-off between libertarian and authoritarian values is the only relevant issue, and has determined how each party makes that trade-off, all that remains is to decide which trade-off is best, and order all strategies in terms of distance from that ideal point.

We should emphasize that this third step is done individually by the participants, but the result of group deliberation guides it. Taken in contrapositive, the claim is indeed that, if a participant forms non-single-peaked preferences at the third step, this must be because some other issues are important to her, after all, or because she disagrees on how the alternatives should be ranked on the unique dimension, contradicting the conclusions collectively reached at the first and second steps.

Our goal in this paper is to revisit and ultimately express doubts regarding this third step, but before presenting that, it is useful to review other existing criticisms. Ottonelli and Porello (2013) have argued that the first two steps of the mechanism require different forms of substantive agreements, and that it is debatable whether such agreements are easier to reach than fully consensual preferences. The mechanism indeed requires agreement regarding the relevant dimension and regarding the position of the alternatives on it. Both aspects are likely to involve a number of intricate or even thick, value-laden concepts. For those, it seems unlikely that the participants will reach a consensus on their meaning or their concrete realization in each of the alternatives. To start with the second aspect, even assuming in our running example that the voters agree on what are libertarian and authoritarian values, it is not implausible that deep disagreements will persist after deliberation regarding the extent to which each party embodies them. Regarding the first aspect, recall that it requires the participants to agree on the problem's relevant normative or evaluative dimension. This dimension will typically reflect a thick concept, intertwining factual with normative and evaluative questions, for instance, health, well-being, sustainability, freedom, or autonomy, to name a few. It seems rather unlikely, Ottonelli and Porello (2013) argue, that deliberation will lead the participants to agree on the meaning of such contested notions. Of course, deliberative democrats have long observed that public deliberation puts rational pressure on the participants to argue in terms of the common good (Miller, 1992), which might be conducive to an agreement on a shared dimension. But when it comes to such thick concepts, this agreement might be only superficial, involving political catchwords and thus leaving the participants using their own, possibly mutually incompatible, understandings of them. All of this does not exclude the fact that deliberation might make it more likely, in comparison with other democratic procedures, to generate single-peaked preferences from meta-agreements. The point is rather that starting from the latter puts the bar very high, especially if there appear to be other ways of reaching single-peaked preferences or of avoiding incoherent group rankings altogether.

3 Single-Peakedness Through Rationality?

The meta-agreement hypothesis, and the received view more generally, have received some empirical support, primarily reported in List et al. (2012) and Farrar et al.

(2010). They used prominence in the public sphere as a proxy for the existence of meta-agreements before deliberation. They observe stronger increases in proximity to single-peakedness in cases where the issue at hand was less prominently discussed publicly before deliberation, and interpret this data as showing that the formation or the discovery of meta-agreements through deliberation goes together with an increase in proximity to single-peakedness, has suggested by the meta-agreement hypothesis.

This evidence for the meta-agreement hypothesis is, however, indirect. While it suggests a correlation between the formation or the discovery of meta-agreements and an increase in proximity to single-peakedness, it does not directly test whether the specific mechanism postulated in the hypothesis is causally responsible for the increase. The data is silent, in particular, on whether the participant felt in any way compelled, presumably by rational pressure, to form single-peaked preferences once they have identified a structuring dimension and positioned the alternatives on it.

Rafiee Rad and Roy (2021) have, in fact, provided evidence from computational simulations that suggests that increases in proximity to single-peakedness might rather result from willingness to reach consensus than from a rational response to meta-agreements. On the one hand, they observe that rational preference change alone is insufficient to ensure an increase in proximity to single-peakedness. For decisions on three alternatives (Abou Zeid, 2021), rational deliberation sometimes insufficiently increases proximity to single-peakedness and even tends to create incoherent group preferences. Rather, the computational model suggests an alternative mechanism: the degree to which participants in deliberation are willing to reach a consensus with others might be the main driver for the increase in proximity to single-peakedness observed in deliberation—c.f. again Abou Zeid (2021) and also Rafiee Rad (2022) for additional remarks to that effect.

In this section, we want to formulate two conceptual arguments that support the search for alternative explanations of the observed correlation between meta-agreement and increases in proximity to single-peakedness. Both arguments question the claim that rationality requires the participants to form single-peaked preferences once they have reached meta-agreement. The first argument intuitively appeals to cases where the interaction of two goods naturally leads to non-convex preferences. We question why such cases should be seen as irrational once meta-agreements are reached. The second argument fleshes out this intuition by using the theory of reason-based rational choice, developed by Dietrich and List (2013). Interpreting meta-agreements as constraining the admissible reasons that can ground preference relations, we argue that it might be rational to form non-single-peaked preferences even in the presence of meta-agreements.

3.1 The Case of Non-convex Preferences

Consider a simple example of some friends deciding where to go for lunch in Munich. The options might be "Japanese", "Bavarian", and "Japanese-Bavarian", and everyone agrees that the question is how exotic the food may be, "Bavarian" being the least exotic, "Japanese" the most, "Japanese-Bavarian" in between.[1] Suppose that the friends have a fixed budget to be spent on two goods that are available in three combinations. Call the degree to which the food is exotic x, the degree to which it is conventional c. Let us furthermore assume that the goods have the same price $p_x = p_c$.

Assuming that these preferences can be represented by a *convex* utility function along the exotic/conventional dimension boils down to assuming that these preferences are single-peaked. The convexity assumption is often grounded on the fact that economics usually looks at goods that are voluntarily consumed together. However, whether this is the case is an empirical assumption and does not provide an argument to the effect that our agents are rationally compelled to have convex utility functions, i.e., single-peaked preferences.

Indeed, double-peaked/single-dipped (Barberà et al., 2012) preferences are still rational in the classical sense of being transitive and complete, and can naturally arise when consuming goods together is less valuable than consuming them individually. In our example, the participants are forced to consume x and c together, and the double-peaked preference can be modeled by the utility function $u(x, c) = (1 - x)^2(1 - c)^2$ where x and c are, again, the degrees to which the food is exotic and conventional. The indifference-curves that are implied by this are given by $I(x) = 1 - \frac{\sqrt{u}}{1-x}$ and look exactly opposite to "usual" indifference-curves. An agent with such preferences consuming these goods in isolation rather than in combination.

Intuitively, it is not clear why it is rational for some of the friends in our examples to have convex preferences over x and c, but *not* otherwise. It is commonplace to have goods that are less preferred when consumed as bundles. A good example is pickles and jam. Someone might like either on their toast, but liking the combination of both on one toast is certainly less common. In other words, many have non-convex preferences when it comes to combinations of pickles and jam. When a group of friends argues about what kind of sandwiches they should prepare for the lunch pack, it seems that the meta-agreement that the relevant issue at hand is whether a given sandwich is sweet or sour will not necessarily result in single-peaked preferences.

So even in simple, two goods models of preferences over a unique dimension, it is not clear why rationality requires the agents to have convex preferences over their bundles. If that intuition is correct, then the claim that rationality requires to form single-peaked preferences once a meta-agreement is reached must at least be qualified to the "standard" case where the agents' preferences are convex.[2]

[1] We assume that the "Japanese-Bavarian" venue serves dishes from both traditions, not necessarily that they combine them in one dish.

[2] As pointed out by an anonymous reviewer to this paper, non-convexity is not the only plausible example of rational, non-single-peaked preferences that are compatible with meta-agreements. So-called group-separable preferences (Inada, 1964) seem to provide another plausible class of examples. We are very grateful to the anonymous reviewer for this pointer.

3.2 *Meta-agreements as Constraints on Reason-Based Preferences*

A natural explanation of our intuition regarding the rationality of having non-convex preferences over certain bundles of goods is that there is a different set of reasons grounding the agents' preferences. In the case of our friends choosing a restaurant, the person preferring both more conventional and more exotic venues to anything "in-between" might do so because she dislikes unusual combinations. So for her, the choice is between venues that are more "on the beaten paths" and those that are less so. Similarly, having non-convex preferences over bundles of pickles and jam as might be explained as a choice between conventional and less conventional bread spreads, instead of between sweet and sour ones. In both cases, the explanation boils down to denying that the agents have, in effect, reached a meta-agreement. The dimension structuring their choices is not what they had purportedly agreed on.

This suggests a natural way to flesh out the idea that rationality requires to form single-peaked preferences once meta-agreements are reached, in terms of the reasons that should ground the participants' preferences. The idea, already alluded to by Ottonelli and Porello (2013), would be that what the structuring dimension singled out by the meta-agreement does is to pinpoint a set of reasons that are viewed as most prominently bearing on the discussion at hand. Reaching a meta-agreement would then mean that these are the reasons that all participants agree each should take into account while individually forming their preferences. Crucially, the agents need not agree on a weighing of these reasons. The structuring dimension might induce such a weighing (see below), but as we have seen earlier this dimension only expresses how different alternatives embody possibly incompatible values/properties. It leaves open how the agents should weigh these combinations.

To go back to our examples, the dimension structuring the choice of restaurant would then be seen as singling out the property of being conventional or exotic as the relevant reason to consider. The alternative explanation for having non-convex preferences over bundles of exotic/conventional goods would then point to a different set of reasons for the agents' choices, namely how frequent these culinary offers are. Similarly, in the case of the pickles and jam spread, the first structuring dimensions would be seen as singling out sweetness and sourness as the relevant reasons to consider, and the alternative explanation points to conventionality instead.

This idea can be made precise in the framework of reason-based rational choice developed by Dietrich and List (2013). Here we illustrate it through our simple example. Our friends need to choose one of three alternatives, Japanese (j), Bavarian (b), and Japanese-Bavarian (jb). The structuring dimension is the ranking $j < jb < b$, from less more conventional. Dietrich and List (2013) propose to view such rankings as grounded in a set $\mathcal{M}$ of possible combinations of *motivating reasons*, together with a weighing relation $\succeq$ on them. A motivating reason is, in this framework, represented by a property that can be instantiated or not by each alternative. In this case the structuring dimension $j < jb < b$ can be naturally be seen as singling out the combination of two motivating reasons, being exotic (E) or being conventional

(C), with the extension of each following our informal description: $E = \{j, jb\}$, $C = \{jb, b\}$. The dimension then stems from the following weighing of reasons: $\{C\} \succ \{E, C\} \succ \{E\} \succ \emptyset$, in the case where both E and C are motivating.

We thus propose to view meta-agreement as constraining the set of possible combinations of motivating reasons, but not necessarily how one should weigh them. So in our simple example, the structuring dimension $j < jb < b$ expresses the fact that the agents agree that two reasons are relevant to the problem at hand, whether a venue is exotic or conventional, and that any possible combination of those reasons, but *not of other reasons*, may be taken as motivating for the agents. As we have seen, the structuring dimension $j < jb < b$ is not supposed to express a value judgment regarding the alternatives. It is rather an (empirical) ranking capturing the idea that exoticism and conventionality might be at least partially incompatible. From that point of view, it seems natural to suppose the meta-agreement constraints the agents to consider only those possible combinations of reasons, but that it does not constraint them to weigh these reasons in any particular way.

If that proposal is correct, then meta-agreements need not translate into single-peaked preferences along the structuring dimension. In our case an agent could weigh the possible combinations of reasons as follow, $\{C\} \succ \{E\} \succ \{E, C\} \succ \emptyset$, resulting in the preference $b > j > jb$, which is of course not single-peaked relative to the structuring dimension. The meta-agreement rules out alternative considerations like the one we envisioned at the beginning of this section, i.e., that the agent might instead conceive the decision as one between frequent and less frequent culinary offers.

4 Conclusion

We have presented two arguments putting into question the idea that rationality requires one to form single-peaked preferences once they have reached a meta-agreement. Our first argument points out a counter-intuitive consequence of that claim, namely that agents with non-convex preferences might be seen as irrational. Our second argument fleshed out the role of meta-agreement as pinpointing the set of possible reasons motivating a decision, but leaving open how one should weigh these. By formalizing our simple running example of a choice of restaurant, we showed that this understanding of meta-agreement leaves room for rationally holding non-single-peaked preferences.

We should emphasize that these arguments do not question the Received View as a whole, nor the first two steps of the meta-agreement hypothesis. One can still take the empirical evidence to show that deliberation helps form or discover underlying meta-agreements, that this comes together with increases in proximity to single-peakedness, and that this is achieved by way of situating the alternatives along a given structuring dimension. What our argument puts into question is that rational pressure is *the* cause of the resulting single-peaked preferences. As we mentioned, other possible explanations have been put forward recently, e.g., willingness to reach consensus

(Rafiee Rad & Roy, 2021). Our argument stresses the importance of investigating further, both theoretically and empirically, how deliberation affects preference formation and preference change once a meta-agreement is in place to adjudicate better between these different possible mechanisms.

References

Abou Zeid, M. J. (2021). *Collective rationality and deliberation over five and more alternatives*. Master's thesis, University of Bayreuth.

Aldred, J. (2004). Social choice theory and deliberative democracy: A comment. *British Journal of Political Science, 34*, 747–752.

Arrow, K. J. (1963). *Social choice and individual values*. Yale University Press.

Barberà, S., Berga, D., & Moreno, B. (2012). Domains, ranges and strategy-proofness: The case of single-dipped preferences. *Social Choice and Welfare, 39*, 335–352.

Black, D. (1948). On the rationale of group decision-making. *Journal of Political Economy, 56*, 23–34.

Dietrich, F., & List, C. (2013). A reason-based theory of rational choice. *Nous, 47*, 104–134.

Dryzek, J. S., & List, C. (2003). Social choice theory and deliberative democracy: A reconciliation. *British Journal of Political Science, 33*, 1–28.

Dryzek, J. S., & Niemeyer, S. (2006). Reconciling pluralism and consensus as political ideals. *American Journal of Political Science, 50*, 634–649.

Farrar, C., Fishkin, J. S., Green, D. P., List, C., Luskin, R. C., & Paluck, E. L. (2010). Disaggregating deliberation's effects: An experiment within a deliberative poll. *British Journal of Political Science, 40*, 333–347.

Gaertner, W. (2001). *Domain conditions in social choice theory*. Cambridge University Press.

Inada, K.-I. (1964). A note on the simple majority decision rule. *Econometrica, 32*, 525–531.

List, C. (2002). Two concepts of agreement. *The Good Society, 11*, 72–79.

List, C., Luskin, R. C., Fishkin, J. S., & McLean, I. (2012). Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. *The Journal of Politics, 75*, 80–95.

Miller, D. (1992). Deliberative democracy and social choice. *Political Studies, 40*, 54–67.

Ottonelli, V., & Porello, D. (2013). On the elusive notion of meta-agreement. *Politics, Philosophy & Economics, 12*, 68–92.

Puppe, C. (2018). The single-peaked domain revisited: A simple global characterization. *Journal of Economic Theory, 176*, 55–80.

Rafiee Rad, S., Braun, S. T., & Roy, O. (2022). *Anchoring and deliberation over preference rankings*. Working Paper, University of Bayreuth.

Rafiee Rad, S., & Roy, O. (2021). Deliberation, single-peakedness, and coherent aggregation. *American Political Science Review, 115*, 629–648.

On the Individual and Coalitional Manipulability of q-Paretian Social Choice Rules

Fuad Aleskerov, Alexander Ivanov, Daniel Karabekyan, and Vyacheslav Yakuba

Abstract We study the degree of individual and coalitional manipulability of q-Paretian social choice rules under Impartial Culture. Manipulability is defined as a situation, when an agent or a coalition, which consists of some agents, misrepresents her/their preferences to obtain a better outcome of the social choice rule. We study a class of q-Paretian social choice rules, which consists of four rules: Strong q-Paretian simple majority rule, Strong q-Paretian plurality rule, Strongest q-Paretian simple majority rule, and Condorcet practical rule. For the cases of 3, 4, and 5 alternatives and for the cases from 3 to 100 agents, we use computer modelling to calculate a number of manipulability indices. We provide the analysis of the results for different cases.

1 Introduction

Manipulability is a situation when an agent or a group of agents misrepresent her/their preferences to get a better result of a social choice rule.

Let us consider an example. A profile in Table 1 consists of 6 group of agents $(p_1, \dots, p_6)$, 20 agents overall, and 3 alternatives (a, b, c).

If the plurality rule is used, the winner will be $\{a\}$, because it receives 8 votes. But it is the worst possible result for the fourth group of agents (p_4). If these 3 agents

F. Aleskerov (✉) · A. Ivanov · D. Karabekyan · V. Yakuba
HSE University, Moscow, Russia
e-mail: alesk@hse.ru

A. Ivanov
e-mail: aaivanov@hse.ru

D. Karabekyan
e-mail: dkarabekyan@hse.ru

V. Yakuba
e-mail: yakuba@ipu.ru

F. Aleskerov · V. Yakuba
Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_7

put in the ballot not their sincere preference ($b > c > a$), but insincere preference $c > b > a$, the outcome of the social choice will be $\{c\}$ with 9 votes. Thus, the coalition of 3 agents may successfully manipulate getting a better social choice ($\{c\}$ instead of $\{a\}$), and this profile is manipulable.

It was shown in Gibbard (1973) and Satterthwaite (1975) that every non-dictatorial social choice rule is manipulable, i.e. there exists at least one case when a successful manipulation is possible. Years later Duggan and Schwartz (2000), draw the same conclusion for the case of multi-valued choice, i.e. when there can be ties between alternatives in the social choice and the social choice may consist of a set of alternatives. Thus, there is a question: which social choice rules are the least manipulable?

Since then, a number of papers have been published. First, individual manipulability was studied, i.e. when one agent misrepresents her preferences to get a better social choice. For example, in Chamberlin (1985) four and in Aleskerov and Kurbanov (1999) 26 social choice rules were studied. Then, coalitional manipulation was studied, i.e. when a group of voters misrepresents their preferences (Xia, 2010). In coalitional manipulation, there are usually some assumptions on which agents may form a coalition.

In this study, we investigate a group of q-Paretian social choice rules. For fixed q, the axiomatic characterization of these rules was given in Aleskerov (1985, 1992, 1999). It was shown that they satisfy Maskin's monotonicity; hence, they are Nash-implementable (Maskin, 1979). However, with fixed q the choice might be empty. Thus, we study the rules in which q is varied, and this type of rules is direct generalization of Condorcet practical rule (de Condorcet, 1785; see also Young, 1988). They were studied only in few papers, because most papers study manipulability of popular rules, generally, scoring rules. However, q-Paretian rules for some cases show very low manipulability.

Our work is related to the framework of weighted committee games by Kurz et al. (2021). As it was noted in Kurz et al. (2020, 2021) the main difference between power indices and manipulability indices is the incentive part: manipulation should be profitable for a voter. In this paper, we use group of indices (defined below) that estimate not only the share of profitable changes, but also stability of outcome and chances of getting worse result.

The manipulability of social choice rules is studied using analytical and computational ways of estimating the degree of manipulability of social choice rules.

The analytical approach implies obtaining formulae for manipulation indices for a given social choice rule. Such an approach allows to get the exact values of indices.

Table 1 Example of a profile where manipulation is possible

Group	p_1	p_2	p_3	p_4	p_5	p_6
First best	a	a	b	b	c	c
Second best	b	c	a	c	a	b
Third best	c	b	c	a	b	a
# of agents	4	4	3	3	3	3

However, this problem can be solved only for some rules, usually for scoring rules, because for some rules (e.g. q-Paretian rules) it might be too complicated.

The computer simulation allows to get the approximate values of manipulability indices. Usually, a number of random profiles are generated, and the indices are calculated for each rule for the set of profiles. If the number of generated profiles is large, it allows finding out the least manipulable social choice rules.

We use a computational approach to estimate the degree of manipulability of four q-Paretian rules. We investigate manipulability in the classic version of the model (Gibbard, 1973; Satterthwaite, 1975), when all voters possess complete information about others' preferences, as well as coalitional manipulability when two or more coalitions cannot manipulate simultaneously. Alternative assumptions are also studied in literature for some social choice rules: for example, Veselova (2020) studies the case of incomplete information and Aleskerov et al. (2021) studies the cases of counter-threats, when another coalition might make a counter-manipulation in response to a manipulation attempt by a certain coalition.

We study individual and coalitional manipulability of q-Paretian social choice rules for the case of Impartial Culture (IC). We use several manipulability indices which allow us to estimate the share of manipulable profiles, the freedom and efficiency of manipulation, as well as resoluteness of each rule.

2 q-Paretian Rules: Definitions

A social choice rule is a mapping from the set of voters with their preferences to the set of the alternatives. If there are n agents and m alternatives, then there are $m!$ different preferences (linear orders) over the set of alternatives.

We consider four q-Paretian voting rules. We use their definitions from Aleskerov (1992, 1999).

2.1 Strong q-Paretian Simple Majority Rule

Let $f(\vec{P}; i, q) = \{x \in A \mid \#D_i(x) \leq q\}$, where $D_i(x) = \{y \in A \mid y P_i x\}$ is the upper contour set of an alternative x for an agent i and $\mathfrak{F} = \left\{I \subset N \mid \#I = \left\lceil \frac{n}{2} \right\rceil\right\}$ is the family of simple majority coalitions.

Define a function $C(A) = \bigcup_{I \in \mathfrak{F}} \bigcap_{i \in I} f(\vec{P}; i, q)$.

The social choice consists of top alternatives for each voter in at least one simple majority coalition. If there is no such an alternative, then the social choice is defined by increasing q by 1, until it is not empty.

2.2 Strong q-Paretian Plurality Rule

The rule is almost the same as the Strong q-Paretian simple majority rule with only one difference. If several alternatives are in the choice set, then for each alternative the number of coalitions which choose this alternative is counted. Then the alternative with maximal value of this number is chosen.

2.3 Strongest q-Paretian Simple Majority Rule

Define a function $C(X) = \bigcap_{I\in\mathfrak{F}} f(\vec{P}; I, q)$, where

$$f(\vec{P}; I, q) = \left\{ x \in A \mid \# \bigcap_{i\in I} D_i(x) \le q \right\}.$$

The social choice consists of the alternatives which are Pareto optimal in each simple majority coalition with $q = 0$. If there are no such alternatives, then q is increased by 1 until the choice is not empty.

The manipulability of q-Paretian social choice rules was studied in Karabekyan (2011), but only for the case of individual manipulation. In this study, we consider not only the case of individual manipulability, but also two models of coalitional manipulation.

2.4 Condorcet Practical Rule

Let $f(\vec{P}; i, q) = \{x \in A \mid \#D_i(x) \le q\}$. Define a social choice as a function $C(A) = \bigcap_{i\in N} f(\vec{P}; i, q)$ with $q = 0$. If there are no such alternatives, then $q = 1, q = 2$, etc., is considered until the social choice is not empty.

3 Extended Preferences and Multi-valued Choice

What if there is a tie between two or more alternatives in a voting procedure? One approach is to use a tie-breaking rule to determine the winner. The most popular tie-breaking rule is alphabetical one. In this case, the result of the voting procedure always consists of a single alternative.

The other approach is to allow ties. If two or more alternatives have the same score in a voting procedure, all such alternatives will be included into the social choice. In order to compare all possible sets of alternatives, i.e. all possible social choices, additional assumptions about the preferences of the voters are needed. These assumptions define so-called extended preferences. If a preference of an agent is a linear order over the set of alternatives, an extended preference is a linear order over the set of all possible social choices, i.e. over all subsets of the set of the alternatives. A detailed overview of preference extension axioms can be found in Barberà et al. (2004).

It should be noted that the voters do not know the probability of each alternative in the tie to be chosen as the final outcome—so we deal with the case of ambiguity aversion. In this paper, we introduce extended preferences as the definite way to compare sets of alternatives. In the literature, there is experimental evidence of collective choice under ambiguity (see Levati et al., 2017).

For the case of 3 alternatives, there are 4 known ways of constructing extended preferences (EP). In all cases, we assume that the voter's preference is $\{a\} > \{b\} > \{c\}$.

1. Leximin: social choices are compared alphabetically, starting from the worst alternative in the set. EP under Leximin: $\{a\} \succ \{a, b\} \succ \underline{\{b\} \succ \{a, c\} \succ \{a, b, c\}} \succ \{b, c\} \succ \{c\}$
2. Leximax: social choices are compared alphabetically, starting from the best alternative in the set. EP under Leximax: $\{a\} \succ \{a, b\} \succ \underline{\{a, b, c\} \succ \{a, c\} \succ \{b\}} \succ \{b, c\} \succ \{c\}$
3. Risk-lover: social choices are compared by the probability of the best alternative. EP under Risk-lover: $\{a\} \succ \{a, b\} \succ \underline{\{b\} \succ \{a, c\} \succ \{a, b, c\}} \succ \{b, c\} \succ \{c\}$
4. Risk-averse: social choices are compared by the probability of the worst alternative. EP under Risk-averse: $\{a\} \succ \{a, b\} \succ \underline{\{b\} \succ \{a, b, c\} \succ \{a, c\}} \succ \{b, c\} \succ \{c\}$.

For the case of 4 and 5 alternatives, there are 10 and 12 ways of constructing extended preferences, respectively. They are based on one or more axioms and/or their combinations: lexicographical (Leximin, Leximax), probabilistic (risk-averse, risk-lover) or ranking (average rank). For example, under Leximax, the EP for the case of 4 alternatives is as follows

$$\{a\} \succ \{a, b\} \succ \{a, b, c\} \succ \{a, b, c, d\} \succ \{a, b, d\} \succ \{a, c\} \succ \{a, c, d\} \\ \succ \{a, d\} \succ \{b\} \succ \{b, c\} \succ \{b, c, d\} \succ \{b, d\} \succ \{c\} \succ \{c, d\} \succ \{d\}.$$

Based on the Leximin extension, the EP for the case of 4 alternatives is represented as

$$\{a\} \succ \{a, b\} \succ \{b\} \succ \{a, c\} \succ \{a, b, c\} \succ \{b, c\} \succ \{c\} \succ \{a, d\} \succ \{a, b, d\} \\ \succ \{b, d\} \succ \{a, c, d\} \succ \{a, b, c, d\} \succ \{b, c, d\} \succ \{c, d\} \succ \{d\}.$$

4 Manipulation Models and Manipulability Indices

We study three models of manipulability of q-Paretian procedures: one case of individual manipulability and two cases of coalitional manipulability.

1. Individual manipulability. In the case of individual manipulability, it is assumed that one agent misrepresents his/her preferences to get a better result of the voting procedure.
2. Coalitional manipulability with the same preferences in the coalition. In this case, we assume that a coalition may manipulate, i.e. not only one agent, but two or more agents as well. A coalition is defined as a group of agents with exactly the same preferences.
3. Coalitional manipulability with coalitions promoting the same alternative. In this variation of coalitional manipulation, it is assumed that a coalition may consist of not only agents with the same preferences, but may consist of agents with different preferences, if they would like to promote the same alternative.

Next comes the evaluation of the degree of manipulability. The most widespread manipulability index is Nitzan–Kelly (NK) index (Kelly, 1993; Nitzan, 1985). The NK-index shows the share of manipulable profiles

$$\text{NK} = \frac{\text{number of manipulable profiles}}{\text{total number of profiles}}.$$

Later, in Aleskerov and Kurbanov (1999) additional indices were introduced. First, freedom of manipulation: I_1^+, I_1^0, and I_1^- indices. They show the shares of all possible manipulation attempts which lead to making the outcome of the social choice better, equal or worse, respectively.

Assuming that there are p possible attempts to manipulate, we can classify them

1. Successful attempts (p^+): the result after such an attempt is better for the manipulating voter or coalition.
2. Unsuccessful attempts which lead to the same choice (p^0): the result after such an attempt is the same as without this manipulation attempt.
3. Unsuccessful attempts which worsen the choice (p^-): the result after such an attempt is worse than without this manipulation attempt.

By definition, we have $p = p^+ + p^0 + p^-$. Then, we define I_1 indices as the share of attempts which make the outcome better, same or worse, i.e. $I_1^+ = \frac{p^+}{p}$, $I_1^0 = \frac{p^0}{p}$, and $I_1^- = \frac{p^-}{p}$. It can be noticed that $I^+ + I^0 + I^- = 1$.

Next comes an efficiency of manipulation. It is measured by two indices: I_2 and I_3.

Index I_2 measures average gain of manipulability. For all successful manipulation attempts, we calculate the benefit of manipulation in terms of the number of places in the preferences. For example, if we use Leximin ways of constructing extended preferences, i.e. EP is $\{a\} \succ \{a, b\} \succ \{b\} \succ \{a, c\} \succ \{a, b, c\} \succ \{b, c\} \succ \{c\}$, and if before manipulation, the choice is $\{c\}$, and after manipulation, it becomes $\{a, c\}$, then the gain is equal to 3 places. Thus, I_2 is equal to the average gain in successful manipulation attempts.

Index I_3 measures maximum gain of manipulability. For a given profile, we calculate the largest gain which can be obtained by an agent. I_3 will be equal to average of maximum gains for every profile.

Another useful measure is resoluteness (decisiveness). We use its definition from Aleskerov et al. (2009). The resoluteness index (D) is equal to the average number of alternatives in the social choice. To calculate it, one should calculate the cardinality of the social choice for each profile and then take the average cardinality. D-index may vary from 1 to m. Values close to 1 show that the rule is good in terms of resoluteness and more often returns a single-valued choice. If D is close to m, then the rule has poor resoluteness and more often returns all alternatives as a multi-valued choice. The idea is that if a certain rule often gives full set of alternatives as the social choice (e.g. $\{a, b, c\}$), it will have low manipulability, but will be almost useless from practical point of view.

We consider manipulability under Impartial Culture (IC). It means that all profiles and preferences are equally likely. For the case of m alternatives, there are $m!$ different preferences. Thus, for n agents there are $(m!)^n$ different profiles under Impartial Culture.

In this study, we evaluate NK, I_1, I_2, I_3, and D indices using computer simulation approach for the case of IC. The most accurate approach would be to generate all possible profiles. However, it can be noticed that the total number of profiles grows exponentially, and it is possible to generate all possible profiles only for small numbers of voters. Such an approach for small n was used in Aleskerov and Kurbanov (1999).

That is why we use another approach. It was shown in Karabekyan (2012) that if one generates 1 million random profiles and calculates indices, the obtained accuracy will be above than 0.001. For this reason, we generate 1 million profiles for each model and estimate the values of all considered indices.

The simulation process consists of several steps:

1. 1,000,000 random profiles are generated.
2. Each rule is considered separately.
3. For each profile, the social choice for each rule is calculated. Its cardinality will be used to calculate D-index.
4. Each of the 3 manipulation models is considered separately. For each profile for each model, all possible manipulation attempts are generated, i.e. for the case of individual manipulation all attempts from each agent are generated; for the case of coalitional manipulation, all attempts from all possible coalitions are generated.
5. Each way of constructing extended preferences is considered separately (4 ways for $m = 3$ alternatives, 10 ways for $m = 4$ alternatives, 12 ways for $m = 5$ alternatives).

6. For each manipulation attempt, the result for the profile with insincere preference is calculated.
7. If the result after a manipulation attempt is better than the sincere result, p^+ is increased. If the same, then p^0 is increased. If worse, then p^- is increased. At the end, I_1 indices are calculated.
8. If there is at least one successful manipulation attempt in a profile, it is marked as manipulable for that model for that rule. Otherwise, it is marked as non-manipulable. At the end, NK-index is calculated as the share of manipulable profiles.
9. For all successful manipulation attempts, we take average gain to calculate I_2 index.
10. For all successful manipulation attempts, we take maximum gain to calculate I_3 index.

The overall computational complexity is approximately equal to

$$
\begin{aligned}
& \text{number of profiles} \cdot \text{number of rules} \cdot \text{rule complexity} \\
& \cdot \text{number of extended preference} \cdot 100 \text{ cases of numbers of agents} \\
& \cdot 3 \text{ cases of alternatives (3, 4, and 5)} \cdot \text{number of possible coalitions} \\
= {} & 1{,}000{,}000 \cdot 4 \cdot 12 \cdot 100 \cdot 3 \cdot \text{number of coalitions} \\
= {} & 144 \times 10^8 \cdot \text{rule complexity} \cdot \text{number of coalitions}
\end{aligned}
$$

As a result, the computations had been performed for several months on 5 PCs because of the large computational complexity.

5 Results

We have obtained results for indices NK, I_1^+, I_1^0, I_1^-, I_2, I_3, and D for four q-Paretian rules for $m = 3, 4, 5$ alternatives for $n = 3, \ldots, 100$ for 4 ($m = 3$), 10 ($m = 4$) and 12 ($m = 5$) ways of constructing extended preferences. In this section, we discuss the results.

Our first point of interest is the comparison of the four rules by the share of manipulable profiles, i.e. by NK-index. Figure 1 is the chart for the case of 3 alternatives, individual manipulation, Leximax EP. The case of 4 alternatives is depicted in Fig. 2. The case of 5 alternatives is depicted in Fig. 3.

The first observation is the behaviour of Condorcet practical rule. Starting from a certain number of agents, it starts approaching NK close to 0 rather fast.

Why it is the least manipulable for almost all cases? The answer lays in its resoluteness. Let us have a look at the charts with D-index for the case of 3 alternatives, see Fig. 4, the case of 4 alternatives, see Fig. 5, and the case of 5 alternatives, see Fig. 6, respectively.

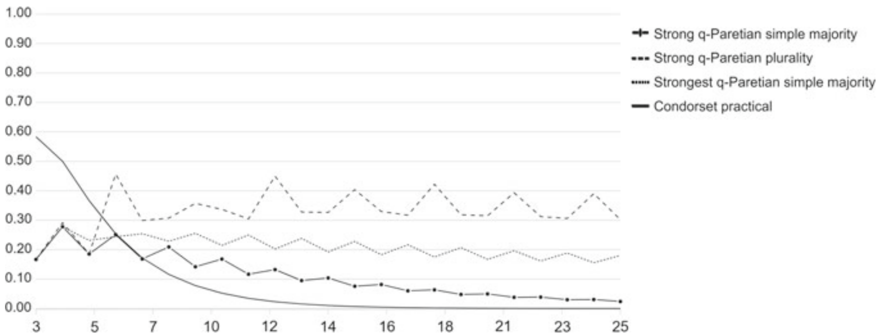


Fig. 1 *NK*-index, Leximax, 3 alternatives, individual manipulation

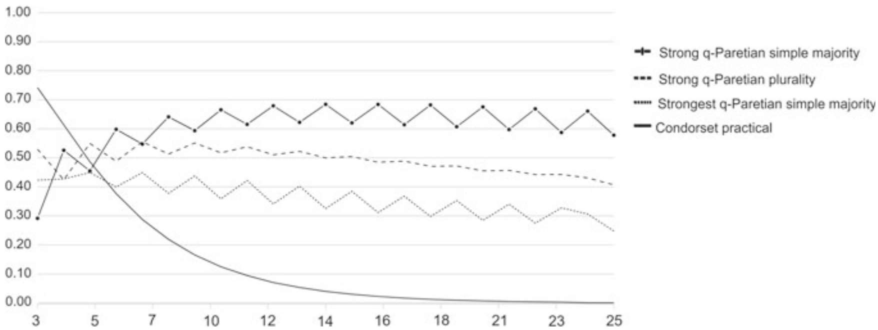


Fig. 2 *NK*-index, Leximax, 4 alternatives, individual manipulation

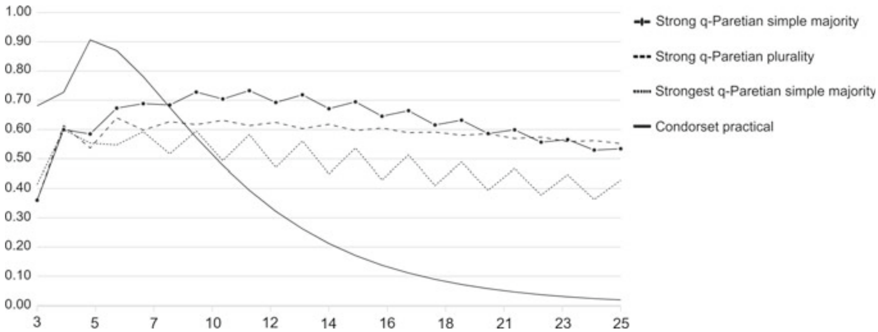


Fig. 3 *NK*-index, Leximax, 5 alternatives, individual manipulation

The values of D-index for Condorcet practical rule for all three cases approach $D = m$, i.e. the number of alternatives. It means that the rule has poor resoluteness, and it gives a complete multiple choice (e.g. $\{a, b, c\}$ for the case of 3 alternatives) as the result of the procedure in most cases.

Thus, Condorcet practical rule is the least manipulable rule due to the lack of resoluteness. We do not discuss this rule, concentrating on the other three rules.

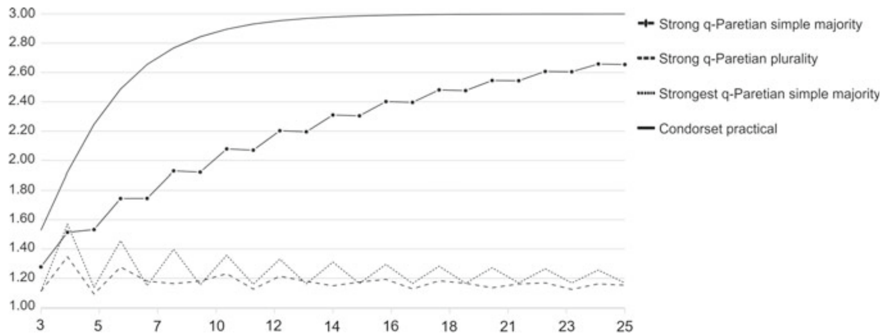


Fig. 4 *D*-index, 3 alternatives

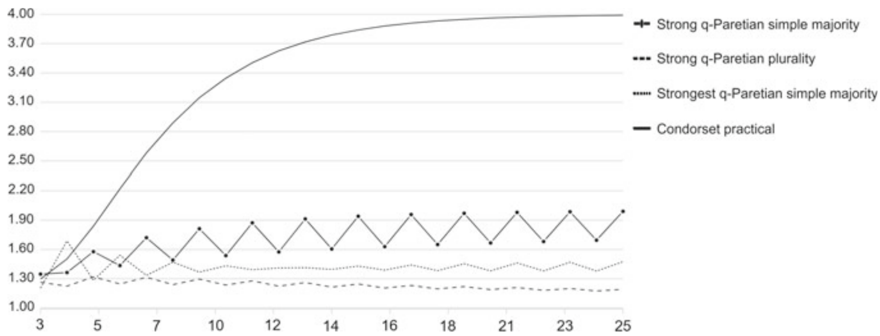


Fig. 5 *D*-index, 4 alternatives

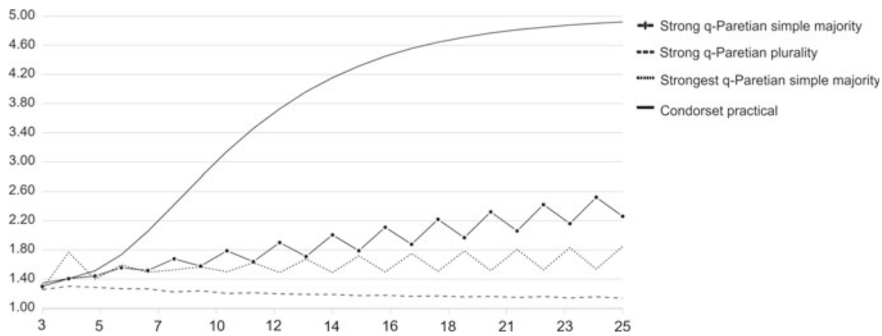


Fig. 6 *D*-index, 5 alternatives

Let us order the other three rules for the cases of $m = 3$ alternatives from the least manipulable to the most manipulable. For $n \leq 5$ voters the order is unclear, but for $n > 5$ the order is clear:

1. Strong q-Paretian simple majority;
2. Strongest q-Paretian simple majority;

3. Strong q-Paretian plurality.

However, the cases of $m = 4$ and $m = 5$ give a different order:

1. Strongest q-Paretian simple majority;
2. Strong q-Paretian plurality;
3. Strong q-Paretian simple majority.

It can be noticed that for all cases ($m = 3, 4, 5$) Strongest q-Paretian simple majority rule is less manipulable than Strong q-Paretian plurality rule. And the only difference between $m = 3$ and $m = 4, 5$ is that Strong q-Paretian simple majority is the least manipulable for the case of $m = 3$ and the worst for the cases of $m = 4$ and $m = 5$. It can be also explained by resoluteness: for $m = 3$ D-index for Strong q-Paretian simple majority rises very fast, for example, $D = 2.66$ for $n = 25$ agents.

From the chart with the cases of up to 25 agents, it is unclear what happens to the resoluteness of Strong q-Paretian simple majority rule with growing number of agents for the cases of $m = 4$ and $m = 5$. Below we provide the charts of NK-index for the cases of $n = 3, \ldots, 100$ agents, but for the cases of $n > 25$ we have made calculations not with step = 1, but with step = 10 (e.g. 29, 30, 39, 40, etc.). The case of $m = 4$ alternatives is depicted in Fig. 7, and the case of $m = 5$ alternatives is depicted in Fig. 8.

It can be seen that for $m = 4$ resoluteness of Strong q-Paretian simple majority rule is worse than for the other two rules, but does not grow: e.g. $D = 2$ for both $n = 31, 41, \ldots, 99$. For the cases of $m = 3$ and $m = 5$, the resoluteness becomes extremely bad with growing number of alternatives.

A possible hypothesis which might deserve a further theoretical study is that Strong q-Paretian simple majority rule has bad resoluteness for the cases of odd numbers of alternatives.

Now we turn to the rest of the rules: Strong q-Paretian plurality and Strongest q-Paretian simple majority. These rules behave in an opposite way: for all cases ($m = 3, 4, 5$) Strongest q-Paretian simple majority rule is less manipulable, but Strong q-Paretian plurality rule has better resoluteness.

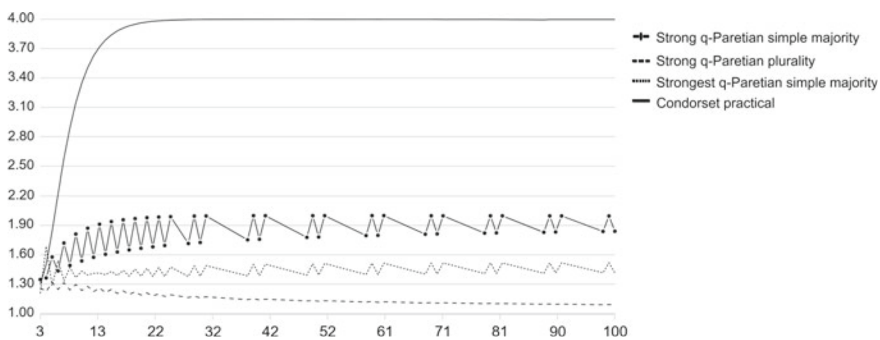


Fig. 7 D-index up to 100 agents, 4 alternatives

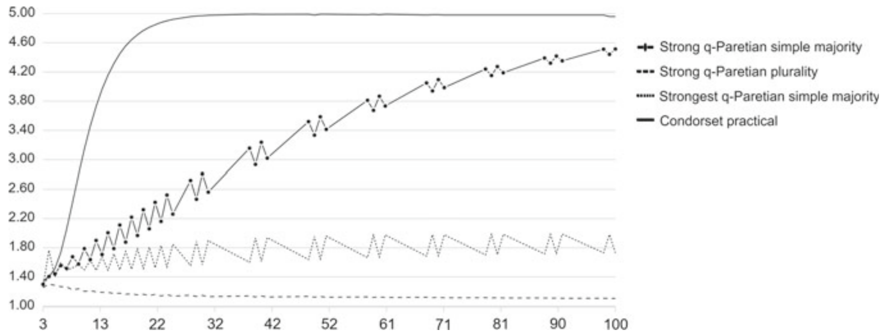


Fig. 8 *D*-index up to 100 agents, 5 alternatives

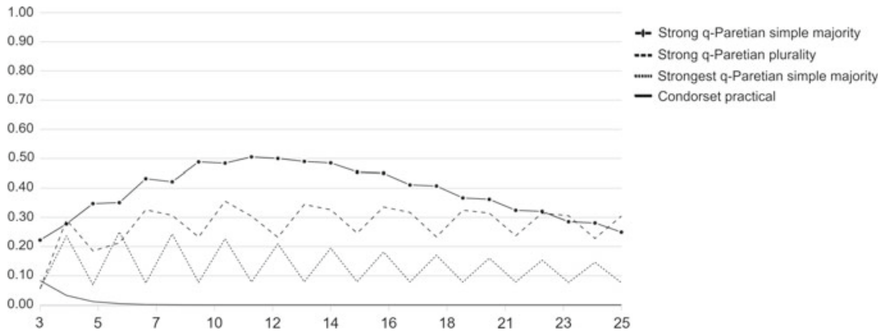


Fig. 9 *NK*-index, Leximin, 3 alternatives, individual manipulation

The next question arises: Are results for *NK*-index the same for other extended preferences? The chart of *D*-index is the same because extended preferences do not impact resoluteness, they impact how a voter compares two social choices while trying to manipulate, i.e. the form of the extended preferences strongly influences the values of the *NK*-index.

In Fig. 9, we depict the chart for the case of Leximin EP, case of 3 alternatives.

We are mostly interested in comparing Strong q-Paretian plurality rule and Strongest q-Paretian simple majority, because Condorcet practical rule starting from 5 agents has poor resoluteness, and Strong q-Paretian simple majority rule has high values of *NK*-index, which will be replaced by smaller values but with poor resoluteness with n growing.

The overall pictures for both Leximax and Leximin are the same: under Leximin EP, Strongest q-Paretian simple majority rule is less manipulable than Strong q-Paretian plurality rule. The cases of $m = 4$ and $m = 5$ are similar (we do not provide charts for these cases to save space). And it is the same for the other two ways of constructing extended preferences: risk-averse and risk-lover.

Next we discuss the case of coalitional manipulability starting with the model when agents only with the exact same preferences may manipulate. The case of $m = 3$, Leximin EP, k (coalition size constraint) $= 2$ is shown in Fig. 10.

The same pattern takes place: Condorcet practical rule shows very low values of NK-index due to poor resoluteness. Strong q-Paretian simple majority rule is highly manipulable, while it has good resoluteness and has low manipulability only when it has poor resoluteness. Strongest q-Paretian simple majority rule is less manipulable than Strong q-Paretian plurality rule.

Does the result differ for the case of larger k, i.e. when more agents may form a coalition? The chart for $k = 5$ is shown in Fig. 11.

The values of NK-index are higher, but the shape of the charts is similar. The same shapes are observed for other values of k (the coalition size constraint) as well as for the cases of $m = 4$ and 5 alternatives.

Now we turn to the model of coalitional manipulation when agents promote the same alternative. For this model, we have obtained the results only for $m = 3$ alternatives due to very high computational complexity. The chart for $m = 3$, Leximin EP looks as shown in Fig. 12.

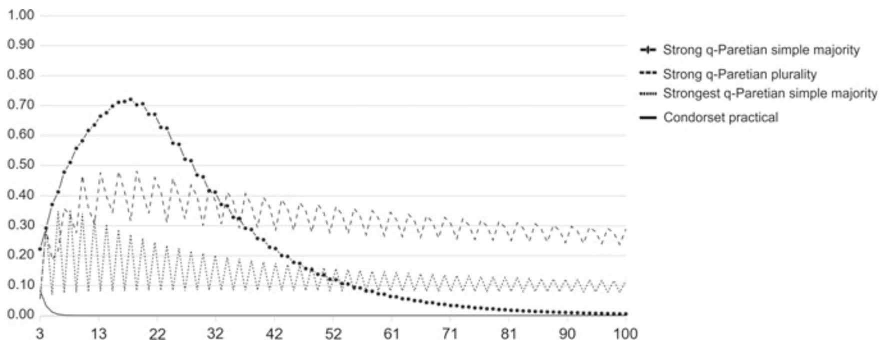


Fig. 10 NK-index, Leximin, 3 alternatives, coalitional manipulation, $k = 2$

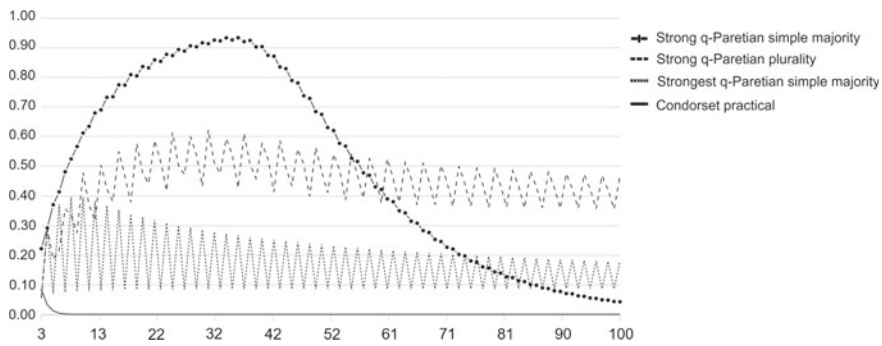


Fig. 11 NK-index, Leximin, 3 alternatives, coalitional manipulation, $k = 5$

In this model, we do not put any constraints to the size of a coalition. That is why the values of NK-index for Strong q-Paretian simple majority rule are growing fast and might decline only with larger values of n. With three other rules, the same patterns hold which we have noticed for the cases of individual manipulability and the first model of coalitional manipulability. The same patterns take place for $m = 4$ and $m = 5$.

Now let us provide the charts for the freedom of manipulation. The case of 3 alternatives individual manipulability, freedom of manipulation, and the Strong q-Paretian plurality rule is considered; see Fig. 13.

Freedom of manipulability might be interpreted as the chances of success of a random manipulation attempt. As we can see, the share of attempts to put insincere preferences which lead to a better choice (grey, upper part of the chart) is significantly smaller than the chances of getting a worse result (dark grey, lower part of the chart) or similar result (light grey, middle part of the chart).

The same situation happens for other rules, e.g. Strongest q-Paretian simple majority rule; see Fig. 14. Here the parity of the number of agents matters: for even numbers of agents, the chances of getting worse result by putting insincere preferences are higher.

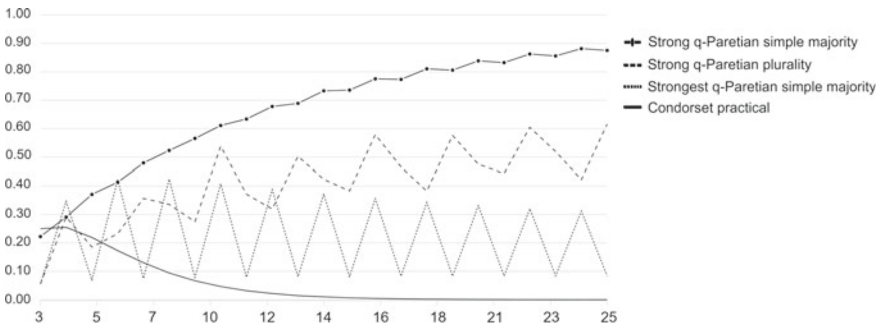


Fig. 12 NK-index, Leximin, 3 alternatives, coalitional manipulation (with coalitions promoting the same alternative)

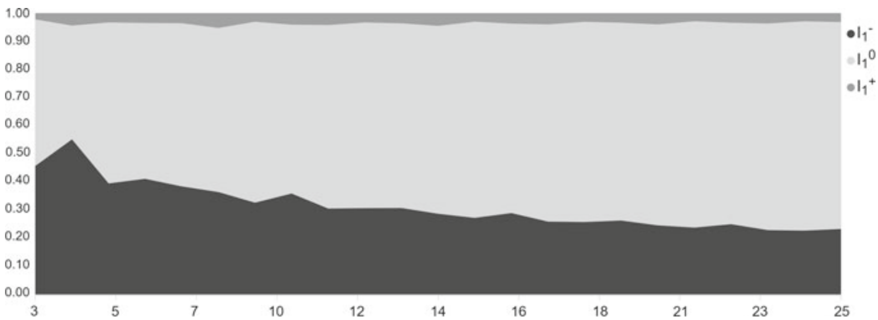


Fig. 13 Freedom of manipulation, 3 alternatives, strong q-Paretian plurality rule

Finally, we would like to study the efficiency of manipulation. Here is the chart for the average gain of manipulation (I_2) for the case of individual manipulation, 3 alternatives under Leximin EP, see Fig. 15.

The chart of the maximum gain in manipulation (I_3) for the case of 3 alternatives is presented in Fig. 16.

The average and maximum gains show close values, steadily declining with the growing number of voters. The possible explanation is that with higher number of voters, it becomes more difficult to drastically improve the outcome of the social choice rule.

The charts of I_2 and I_3 indices give similar picture to the outcomes of the charts with NK-index. Condorcet practical rule gives the lowest values of indices; however, since it has very poor resoluteness, we should take into account Strongest q-Paretian simple majority rule, because it has decent resoluteness and is the second best in terms of indices NK, I_2, and I_3.

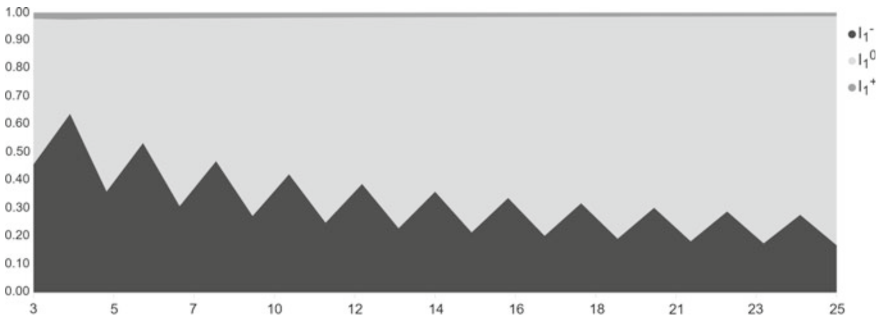


Fig. 14 Freedom of manipulation, 3 alternatives, strongest q-Paretian simple majority rule

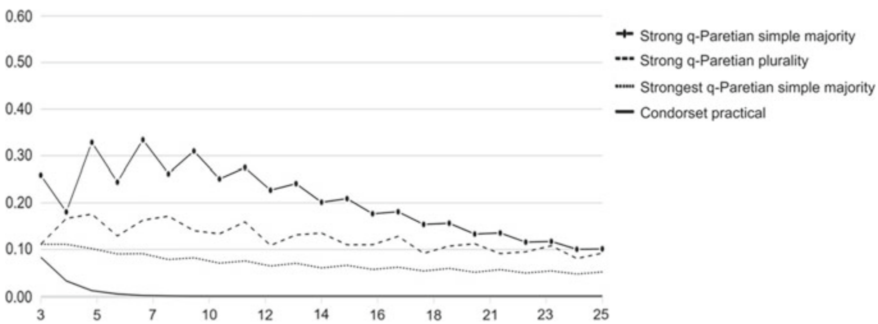


Fig. 15 Efficiency of manipulation (I_2), Leximin, 3 alternatives, individual manipulation

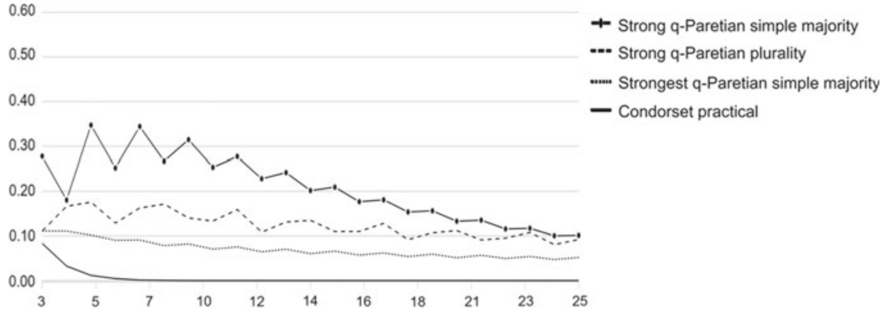


Fig. 16 Efficiency of manipulation (I_3), Leximin, 3 alternatives

6 Conclusion

We have studied manipulability of four q-Paretian social choice rules. We provided the computational scheme of estimating their degree of manipulability. We considered a list of manipulability indices: NK, I_1^+, I_1^0, I_1^-, I_2, I_3, and D.

The main conclusions are as follows:

1. Condorcet practical rule is the least manipulable for almost all cases. However, its low manipulability can be explained by poor resoluteness. It very often returns the full choice (e.g. $\{a, b, c\}$ for the case of 3 alternatives), thus, making manipulation impossible, but the practical value of such rule is also very low.
2. Strong q-Paretian simple majority rule for even number of alternatives ($m = 4$) shows the largest values of *NK*-index, i.e. the worst in terms of manipulability. For odd numbers of alternatives ($m = 3$ and $m = 5$), this rule shows high values of *NK*-index for small numbers of agents. Then, starting from a certain n, the resoluteness becomes significantly worse, while manipulability becomes low.
3. Strongest q-Paretian simple majority rule and Strong q-Paretian plurality rule show opposite results. Strongest q-Paretian simple majority rule is less manipulable, but Strong q-Paretian plurality rule has better resoluteness.

Acknowledgements This work is an output of a research project implemented as part of the Basic Research Program at the HSE University.

References

Aleskerov, F. (1985). Procedures of multicriterial choice. In *Preprints of the IFAC/IFORS Conference on Control Science and Technology for Development* (pp. 858–869).

Aleskerov, F. (1992). *Relational-functional voting operators*. California Institute of Technology, Social Science Working Paper 818.

Aleskerov, F. (1999). *Arrovian aggregation models*. Kluwer Academic Publishers.

Aleskerov, F., Karabekyan, D., Ivanov, A., & Yakuba, V. (2021). Further results on the manipulability of social choice rules—A comparison of standard and Favardin–Lepelley types of individual manipulation. In M. Diss & V. R. Merlin (Eds.), *Evaluating voting systems with probability models, essays by and in Honor of William V. Gehrlein and Dominique Lepelley, Studies in Choice and Welfare* (pp. 231–249). Springer.

Aleskerov, F., Karabekyan, D., Sanver, R., & Yakuba, V. (2009). Evaluating the degree of manipulability of certain aggregation procedures under multiple choices. *Journal of the New Economic Association, 1*, 37–61.

Aleskerov, F., & Kurbanov, E. (1999). Degree of manipulability of social choice procedures. In A. Alkan, C. D. Aliprantis, & N. C. annelis (Eds.), *Current trends in economics, Studies in Economic Theory* (Vol. 8, pp. 13–27). Springer, Berlin

Barberà, S., Bossert, W., & Pattanaik, P. K. (2004). Ranking sets of objects. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory* (pp. 893–977). Springer.

Chamberlin, J. R. (1985). An investigation into the relative manipulability of four voting systems. *Behavioral Science, 30*, 195–203.

de Condorcet, M. (1785). Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. *Imprimerie Royale*.

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

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica, 41*, 587–601.

Karabekyan, D. (2011). On the manipulability of q-Paretian social choice rules. In *Proceeding of the Seminar "Mathematical Modeling of Political Systems and Processes"* (pp. 142–156) (in Russian).

Karabekyan, D. (2012). *The problem of manipulability in the social choice theory*. Ph.D. thesis, HSE University (in Russian).

Kelly, J. S. (1993). Almost all social choice rules are highly manipulable, but a few aren't. *Social Choice and Welfare, 10*, 161–175.

Kurz, S., Mayer, A., & Napel, S. (2020). Weighted committee games. *European Journal of Operational Research, 282*, 972–979.

Kurz, S., Mayer, A., & Napel, S. (2021). Influence in weighted committees. *European Economic Review, 132*, 103634.

Levati, M. V., Napel, S., & Soraperra, I. (2017). Collective choices under ambiguity. *Group Decision and Negotiation, 26*, 133–149.

Maskin, E. (1979). Implementation and strong Nash equilibrium. In J.-J. Laffont (Ed.), *Aggregation and revelation of preferences* (pp. 433–440). North-Holland.

Nitzan, S. (1985). The vulnerability of point-voting schemes to preference variation and strategic manipulation. *Public Choice, 47*, 349–370.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory, 10*, 187–217.

Veselova, Y. A. (2020). Does incomplete information reduce manipulability? *Group Decision and Negotiation, 29*, 523–548.

Xia, L., Conitzer, V., & Procaccia, A. D. (2010). A scheduling approach to coalitional manipulation. In *Proceedings of the 11th ACM Conference on Electronic Commerce* (pp. 275–284).

Young, H. P. (1988). Condorcet's theory of voting. *American Political Science Review, 82*, 1231–1244.

Weighted Voting

Effectiveness, Decisiveness, and Success in Weighted Voting Systems: Collective Behavior and Voting Measures

Werner Kirsch

Abstract Efficiency, decisiveness, and success in a voting system depend not only on the voting rules but also on the collective behavior of the voters. The voting behavior is modeled by a voting measure which describes the interdependence (or independence) of the voters. In this paper, we define and investigate a large class of voting measures. This class can be characterized as those voting measures which are invariant under permuting the voters and which allow a natural extension to an arbitrary number of voters. The class includes the Penrose–Banzhaf measure (independent, impartial behavior), the Shapley–Shubik measure (impartial anonymous behavior). We analyze the efficiency and the success for these voting measures in weighted voting systems.

1 Introduction

The notions effectiveness, decisiveness and success are basic to the analysis of voting systems. Yet, they do not only depend on the voting rule but also on the underlying voting measure, in particular on the correlation structure of the voting behavior between the voters of the system. In the Penrose–Banzhaf case, the voting measure gives equal probability to all coalitions, thus reflecting the situation when each voter's decision is completely independent of the other voters (with probability $\frac{1}{2}$ both for 'yeah' and 'nay'). The corresponding power index (in terms of decisiveness) is the well-known total Penrose–Banzhaf power. Under *this* voting measure, there is a simple formula connecting the power index of a voter with the probability of success of this voter (see (27)).

As was emphasized in (Laruelle and Valenciano, 2015), this intimate connection between decisiveness and success is a peculiarity of the Penrose–Banzhaf measure. In particular, there is no analog for the Shapley–Shubik power index. The Shapley–Shubik index is based on decisiveness under a voting measure we call the Shapley–Shubik measure. Under this measure, a coalitions of size k has probability $\frac{1}{N+1}\frac{1}{\binom{N}{k}}$

W. Kirsch (✉)
Fakultät für Mathematik und Informatik, FernUniversität Hagen, Hagen, Germany
e-mail: werner.kirsch@fernuni-hagen.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_8

where N is the number of voters. Thus, the set of all coalitions of size k is given a probability *independent* of k, namely $\frac{1}{N+1}$.

Among others, we consider voting systems with 'simple voting rule', that is with voting weight 1 for all voters, but with arbitrary relative quota r. For such systems, we compute the probability of decisiveness and the rate of success under the Penrose–Banzhaf measure and the Shapley–Shubik measure. One of the results which we found surprising is that the rate of success under the Shapley–Shubik measure is (approximately) $\frac{3}{4}$ in the case of simple majority (i.e., $r = \frac{1}{2}$). Thus, it is remarkably bigger than the rate of success under the Penrose–Banzhaf measure. On the other hand, the rate of decisiveness in the same situation is bigger under the Penrose–Banzhaf measure, which is $\frac{1}{2}$.

We also identify the Penrose–Banzhaf measure and the Shapley–Shubik measure as special cases for Polya urn models and investigate 'Common Belief Models' (CBM) which extend the class of urn models to the general form of measures invariant under permutations of the voters.

We extend the mentioned result to general r, to weighted voting systems and to more general voting measures.

2 Some Basics

In this section, we introduce some of the concepts basic for the rest of this paper. For a thorough introduction, we recommend (Felsenthal, 1998) and (Napel, 2019), for an overview (Taylor, 2008) or (Kirsch, 2016). In the following, we restrict ourselves to 'Yes–No' voting systems.

Definition 1 A *voting system* $(V, \mathcal{V})$ consists of a (finite) set V of voters and a subset $\mathcal{V}$ of $\mathcal{P}(V)$, the system of all subsets of V, with the following properties

1. $V \in \mathcal{V}$.
2. $\emptyset \notin \mathcal{V}$.
3. If $A \in \mathcal{V}$ and $A \subset B$ then $B \in \mathcal{V}$.

Subsets of V are called *coalitions*. The coalitions in $\mathcal{V}$ are called *winning*, those not in $\mathcal{V}$ *losing*.

Definition 2 A voting system $(V, \mathcal{V})$ is called *weighted* if there is a function $w : V \to [0, \infty)$ and a $q \in (0, \infty)$ such that $A \in \mathcal{V}$ if and only if

$$\sum_{v \in A} w(v) \geq q \, . \tag{1}$$

The number $w(v)$ is called the *weight* of the voter v, q is called the *quota*. The number

$$r = \frac{q}{\sum\limits_{v \in V} w(v)} \tag{2}$$

is called the *relative quota*.
We call a weighted voting system *simple*, if $w(v) = 1$ for all v.
A simple voting system $(V, \mathcal{V})$ is called a *simple majority system* if the relative quota r is given by $r = \frac{1}{2} + \frac{1}{2N}$, where $N = |V|$ is the number of voters. In other words, winning coalitions are precisely those which contain more than half of the voters.

There are various methods to quantify the notion 'voting power' in voting systems. One of the best-known concepts goes back to Penrose (Penrose, 1946) and Banzhaf (Banzhaf, 1965). It is based on the notion of 'decisiveness' and the treatment of all coalitions as 'equally likely'.

Definition 3 Suppose $(V, \mathcal{V})$ is a voting system.

1. We call a voter $v \in V$ *winning decisive* for a coalition $A \subset V$ if $v \notin A$, $A \notin \mathcal{V}$ and $A \cup \{v\} \in \mathcal{V}$.
   We set
$$\mathcal{D}^+(v) := \{A \subset V \mid v \notin A,\ A \notin \mathcal{V},\ A \cup \{v\} \in \mathcal{V}\}\,. \tag{3}$$
2. We call a voter $v \in V$ *losing decisive* for a coalition $A \subset V$ if $v \in A$, $A \in \mathcal{V}$, $A \setminus \{v\} \notin \mathcal{V}$.
   We set
$$\mathcal{D}^-(v) := \{A \subset V \mid v \in A,\ A \in \mathcal{V},\ A \setminus \{v\} \notin \mathcal{V}\}\,. \tag{4}$$
3. We call v *decisive* for A if v is winning decisive or losing decisive for A and set
$$\mathcal{D}(v) := \mathcal{D}^+(v) \cup \mathcal{D}^-(v)\,. \tag{5}$$

Definition 4 The *Penrose–Banzhaf power* $PB(v)$ of a voter v is defined as

$$D_B(v) \; = \; \frac{|\mathcal{D}(v)|}{2^N}\,, \tag{6}$$

where $|A|$ denotes the number of elements of the set A and $N = |V|$.

$D_B(v)$ is the proportion of all coalition for which v is decisive.

Remark 1 It is well known that $|\mathcal{D}^+(v)| = |\mathcal{D}^-(v)|$, in fact $A \mapsto A \setminus \{v\}$ is a bijective map between $\mathcal{D}^+(v)$ and $\mathcal{D}^-(v)$. Consequently,

$$D_B(v) = \frac{|\mathcal{D}^+(v)|}{2^{N-1}} = \frac{|\mathcal{D}^-(v)|}{2^{N-1}}\,. \tag{7}$$

The Penrose–Banzhaf power admits a probabilistic interpretation. If we regard all coalitions in $\mathcal{P}(V)$ as equally likely ('Laplace probability') and denote the corresponding measure on $\mathcal{P}(V)$ by

$$\mathbb{P}_B(\{A\}) := \frac{1}{2^N}\,, \tag{8}$$

for each coalition A, then

$$D_B(v) = \mathbb{P}_B\Big(\mathcal{D}(v)\Big)\,. \tag{9}$$

We call $\mathbb{P}_B$ the *Penrose–Banzhaf* measure.

In the following, we shall denote the voters by integers, so that $V = \{1, \cdots, N\}$. Instead of considering $\mathbb{P}_B$ as a measure on $\mathcal{P}(V)$, we may consider $\mathbb{P}_B$ as a measure on $\{0, 1\}^N (= \{0, 1\}^V)$ by

$$\mathbb{P}_B\Big(\{(x_1, \cdots, x_N)\}\Big) := \mathbb{P}_B\Big(\Big\{\, \{i \mid x_i = 1\}\,\Big\}\Big)\,. \tag{10}$$

In the following, we will switch freely between these versions of $\mathbb{P}_B$. Moreover, to simplify notation we will write $\mathbb{P}_B\big(x_1, \cdots, x_N\big)$ instead of $\mathbb{P}_B\big(\{(x_1, \cdots, x_N)\}\big)$ and $\mathbb{P}_B(A)$ instead of $\mathbb{P}_B(\{A\})$ for any $A \in \mathcal{P}(V)$.
In this paper, we will introduce and discuss various other measures on $\mathcal{P}(V)$, respectively, $\{0, 1\}^N$ which lead to different notions of voting power, for example to the Shapley–Shubik index (Shapley and Shubik, 1954).
We now introduce this concept in an abstract setting.

Definition 5 A probability measure $\mathbb{P}$ on $\mathcal{P}(V)$ (resp. $\{0, 1\}^N$) is called a *voting measure* if

$$\mathbb{P}(A) = \mathbb{P}(V \setminus A) \text{ for all } A \subset V\,, \tag{11}$$

resp. $\mathbb{P}\Big(x_1, \cdots, x_N\Big) = \mathbb{P}\Big(1 - x_1, \cdots, 1 - x_N\Big)$.

Equation (11) means that voters are 'impartial' or a priori neutral with respect to proposals. The main idea behind voting measures is *not* that voters toss a coin to decide about their voting behavior, but rather that the proposal put before them are 'really random'; in particular, any proposal and its complete reverse are equally likely. The papers (Kirsch, 2007, 2016) contain a more detailed discussion about this issue.

As in the case of the Penrose–Banzhaf power, we may define a voting power in terms of decisiveness by

$$D^+_{\mathbb{P}}(v) := \mathbb{P}(\mathcal{D}^+(v))\,, \tag{12}$$

$$D^-_{\mathbb{P}}(v) := \mathbb{P}(\mathcal{D}^-(v))\,, \tag{13}$$

$$\text{and } D_{\mathbb{P}}(v) := \mathbb{P}(\mathcal{D}(v)) = D^+_{\mathbb{P}}(v) + D^-_{\mathbb{P}}(v)\,. \tag{14}$$

Note that we distinguish here between $D^+_{\mathbb{P}}(v)$ (the probability to make a losing coalition winning) and $D^-_{\mathbb{P}}(v)$ (the probability to make a winning coalition losing). In contrast to the case of the Penrose–Banzhaf measure, we can *not* conclude $\mathbb{P}\big(\mathcal{D}^+(v)\big) = \mathbb{P}\big(\mathcal{D}^-(v)\big)$.

Example 1 We give some examples for voting measures where we assume $|V| = N$ and $A \subset \mathcal{P}(V)$:

1. The Penrose–Banzhaf measure.

$$\mathbb{P}(A) \;=\; \frac{1}{2^N}\,. \tag{15}$$

2. The Shapley–Shubik measure.

$$\mathbb{P}_S(A) = \frac{1}{N+1}\,\frac{1}{\binom{N}{|A|}}\,, \tag{16}$$

(see (Straffin, 1977)). This measure makes coalitions of the same cardinality equally likely and satisfies

$$\mathbb{P}_S(\{A \mid |A| = k\}) = \frac{1}{N+1}\,. \tag{17}$$

By D_S, D^+_S, D^-_S we denote the quantities $D_{\mathbb{P}_S}$, $D^+_{\mathbb{P}_S}$, $D^-_{\mathbb{P}_S}$.
$\mathbb{P}_S$ can be written as

$$\mathbb{P}_S\big((x_1, \cdots, x_N)\big) = \int\limits_0^1 p^{\sum x_i}(1-p)^{N-\sum x_i}\, dp\,. \tag{18}$$

This representation of $\mathbb{P}_S$ looks overly complicated, but it is very useful for computations with large N.

3. The unanimity measure

$$\mathbb{P}_u(A) = \begin{cases} \frac{1}{2} & \text{if } A = \emptyset \text{ or } A = V \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

In this case, all voters always agree.

4. The common belief measure generalizes all three previous examples. Suppose μ is a probability measure on [0, 1] (and the Borel σ-algebra) such that for any $a, b \in [-\frac{1}{2}, \frac{1}{2}]$, $a < b$

$$\mu([\frac{1}{2}+a, \frac{1}{2}+b]) = \mu([\frac{1}{2}-b, \frac{1}{2}-a])\,, \tag{20}$$

which says that μ is reflection symmetric with respect to the midpoint $\frac{1}{2}$ of the interval [0, 1]. Then the measure

$$\mathbb{P}_\mu\big((x_1, \cdots, x_N)\big) \;=\; \int_0^1 p^{\sum x_i}(1-p)^{N-\sum x_i}\, d\mu(p) \tag{21}$$

is a voting measure because of (20). We call it the *Common Belief* voting measure or CB-measure with mixing measure μ.

Since we have more to say about the common belief measure, we introduce another way to write it which will be convenient in later sections.

We denote by P_p^1 the probability measure on $\{0, 1\}$ defined by $P_p^1(1) = p$ and $P_p^1(0) = 1 - p$ with $0 \le p \le 1$ and by P_p^N the n-fold product of P_p^1 on $\{0, 1\}^N$. Thus

$$P_p^N\big((x_1, x_2, \ldots, x_N)\big) \;=\; p^{\sum x_i}\,\big(1-p\big)^{N-\sum x_i}\,. \tag{22}$$

Whenever N is clear from the context, we write P_p instead of P_p^N.

With this notation (21) reads

$$\mathbb{P}_\mu\big(A\big) \;=\; \int_0^1 P_p^N\big(A\big)\, d\mu(p) \;=\; \int_0^1 P_p\big(A\big)\, d\mu(p) \tag{23}$$

for all $A \subset \{0, 1\}^N$.

The Penrose–Banzhaf measure corresponds to the choice $\mu = \delta_{\frac{1}{2}}$, the unanimity measure to $\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$ and the Shapley–Shubik measure to the uniform distribution (= Lebesgue measure) on [0, 1], see (18). Here, δ_a denotes the measure concentrated in the point a, in particular $\int f(x)\, d\delta_a(x) = f(a)$.

Technically speaking, the CB-measure $\mathbb{P}_\mu$ is a mixture of probability measures P_p *each* of which makes the random variables $X_1, \ldots, X_N$ *independent*, but with different expected values. Under the *mixture*, however, the random variables are *not* independent (except for the trivial case $\mu = \delta_a$)!

The CB-measure models a situation where the population of voters is influenced by some a priori opinion on the proposition at hand - represented by a number $p \in [0, 1]$. This might be a common believe inside the society or a strong influential group of opinion makers. This a priori judgment itself is random (as the proposal is random), its probability distribution is the measure μ. The voters' decision is still random but with a bias (respectively strong bias) toward acceptance if $p > \frac{1}{2}$ (respectively p close to 1) and a tendency to rejection if $p < \frac{1}{2}$.

For a detailed discussion of the common belief model and its application on two-tier voting systems, we refer to (Kirsch, 2007, 2016, 2021).

In Sects. 3 and 4, we will return to a discussion of voting measures.

Instead of looking at the decisiveness of a voter, one might define the influence of a voter by considering the probability that the outcome of the voting coincides with the voter's opinion.

Definition 6 Suppose $(V, \mathcal{V})$ is a voting system and $\mathbb{P}$ a voting measure on V.
We call the probability

$$S^{+}_{\mathbb{P}}(v) = \mathbb{P}(\{A \in \mathcal{V} \mid v \in A\}) \tag{24}$$

the *rate of affirmative success* of the voter v (with respect to $\mathbb{P}$).
Similarly,

$$S^{-}_{\mathbb{P}}(v) = \mathbb{P}(\{A \notin \mathcal{V} \mid v \notin A\}) \tag{25}$$

is called the *rate of blocking success*.
The quantity

$$S_{\mathbb{P}}(v) = S^{+}_{\mathbb{P}}(v) + S^{-}_{\mathbb{P}}(v) \tag{26}$$

is called the (total) *rate of success* of v.
For the Penrose–Banzhaf measure the rate of success does not give new information because

$$S_{\mathbb{P}_B}(v) = \frac{1}{2} + \frac{1}{2} D_B(v)\,. \tag{27}$$

This equation goes back to (Dubey and Shapley, 1979).
Equation (27) is not true for other voting measures, in fact it is *only* true for the Penrose–Banzhaf measure (Laruelle and Valenciano, 2015).
We introduce a final quantity for this section, namely the 'efficiency' of a voting system, also called the 'power of a collectivity to act'. It goes back to (Coleman, 1971) who introduced it in connection with the Penrose–Banzhaf measure.

Definition 7 If $(V, \mathcal{V})$ is a voting system and $\mathbb{P}$ a voting measure on V then

$$E_{\mathbb{P}} := \mathbb{P}(\mathcal{V}) \tag{28}$$

is called the *efficiency* of the voting system.

3 Urn Models

In this section, we discuss urn models to generate voting measures. This view on voting measures was introduced by (Berg, 1985) and further developed by Kurz–

Mayer–Napel in (Kurz et al, 2021). While those authors consider the case of an arbitrary number of alternatives we concentrate on the case of 0–1-voting.

We start with an urn containing R red balls (carrying the symbol '1' for 'yeah') and B black balls (carrying a '0' for 'nay'). Suppose the first voter draws a ball, replaces that ball and puts c additional ball of the color of the drawn ball into the urn. In other words, the voter draws a ball and (re-)places $1 + c$ balls of the same color. The next voter draws a ball from the urn, which now contains $R + B + c$ balls, puts the drawn ball together with additional c balls of the same color in the urn, and so on. Here, $c \geq 0$ is an integer. We may also interpret $c = -1$ as drawing balls without replacement. In the latter case, the process necessarily ends after $R + B$ drawings, and in all other cases, the process can be extended indefinitely.

The above-defined process is known as Polya's urn scheme. In voting theory, it serves as a model for a roll call (see (Kurz and Napel, 2018)). The voters cast their votes publicly in a particular order. A voter is influenced by those who voted before him or her, in the sense that he or she tends to vote in line with the majority, as far as it is known, when the voter is called to vote. This tendency is absent if $c = 0$; in this case, the voters cast their vote independently of the others. For fixed R and B, the tendency to align with each other is bigger when c is increased. Thus, c is a measure for cooperation between voters (as long as R and B are fixed).

Definition 8 Suppose $X_1, \ldots, X_N$ be the above described Polya urn process of drawing balls with $1 + c$ replacements with R red ('1') and B black balls ('0') in the starting urn. We call the probability distribution on $\{0, 1\}^N$ of the $X_1, \ldots, X_N$ by $\mathbb{P}_{R,B,c}$, i. e.

$$\mathbb{P}_{R,B,c}(x_1, \ldots, x_N) = \text{Prob}\,(X_1 = x_1, \ldots, X_N = x_N)\,. \tag{29}$$

For $Q, c \in \mathbb{R}$ and $k \in \mathbb{N}$ we set

$$Q^{(c,k)} \;:=\; \prod_{j=0}^{k-1} (Q + c\,j)\ . \tag{30}$$

Proposition 1 *Suppose $c \geq 0$ and $(x_1, \ldots, x_N) \in \{0, 1\}^N$*

1. *For any positive integer M*

$$\mathbb{P}_{R,B,c}(x_1, \ldots, x_N) \;=\; \mathbb{P}_{MR,MB,Mc}(x_1, \ldots, x_N)\,. \tag{31}$$

2. *If π is a permutation of $\{1, \ldots, N\}$ then*

$$\mathbb{P}_{R,B,c}(x_1, \ldots, x_N) \;=\; \mathbb{P}_{R,B,c}(x_{\pi(1)}, \ldots, x_{\pi(N)})\,. \tag{32}$$

3. *$\mathbb{P}_{R,B,c}$ is a voting measure, i. e.*

$$\mathbb{P}_{R,B,c}\,(x_1, \ldots, x_N) \;=\; \mathbb{P}_{R,B,c}\,(1 - x_1, \ldots, 1 - x_N) \tag{33}$$

if and only if $R = B$.

4. *If* $\sum_{j=1}^{N} x_i = k$ *we have*

$$\mathbb{P}_{R,B,c}(x_1, \ldots, x_N) \;=\; \frac{R^{(c,k)}\; B^{(c,N-k)}}{(R+B)^{(c,N)}} \;=\; \frac{r^{\,(1,k)}\, b^{\,(1,N-k)}}{(r+b)^{(1,N)}}\,, \tag{34}$$

where we set $r = \frac{R}{c}$, $b = \frac{B}{c}$.

5. *We have*

$$\mathbb{P}_{R,B,c}\left(\sum_{i=1}^{N} x_i \;=\; k\right) \;=\; \binom{N}{k} \frac{R^{(c,k)}\; B^{(c,N-k)}}{(R+B)^{(c,N)}}\,. \tag{35}$$

We sketch a proof of Proposition 1 in Appendix A.2. For more information on urn models, we refer to (Mahmoud, 2009).

It is clear that the random variables X_i of a Polya urn process are independent, if $c = 0$. In fact, if $c = 0$ and $R = B$ the measure $\mathbb{P}_{R,R,0}$ is just the Banzhaf measure defined in (8).

If $c \neq 0$, the X_i are *not* independent. To see this, we compute the covariance between X_1 and X_2:

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) \;&=\; \mathbb{P}_{R,B,c}(X_1 = 1, X_2 = 1) - \mathbb{P}_{R,B,c}(X_1 = 1)\, \mathbb{P}_{R,B,c}(X_2 = 1)\\ &=\; \frac{RB}{(R+B)^2\,(R+B+c)}\, c\,,\end{aligned} \tag{36}$$

so the covariance is positive if $c > 0$ and negative if $c < 0$.

Observe, that

$$\mathbb{E}_{R,B,c}\left(X_1 \cdot X_2 \cdot \ldots \cdot X_k\right) \;=\; \mathbb{P}_{R,B,c}(X_1 = 1, X_2 = 1, \ldots, X_k = 1)\,, \tag{37}$$

where $\mathbb{E}_{R,B,c}$ denotes expectation with respect to $\mathbb{P}_{R,B,c}$.

A particularly interesting case occurs for $c = 1$ and $R = B = 1$. Observing that $1^{(1,Q)} = Q!$ and $2^{(1,Q)} = (Q+1)!$, we obtain for $\sum x_j = k$

$$\mathbb{P}_{1,1,1}\left(x_1, \ldots, x_N\right) \;=\; \frac{1}{N+1}\,\frac{1}{\binom{N}{k}}\,. \tag{38}$$

So the measure $\mathbb{P}_{1,1,1}$ agrees with the Shapley–Shubik measure defined in (16).

Now, we are interested in the number of affirmative votes for large number N of voters, i.e., in the quantity

$$M_N \;:=\; \frac{1}{N} \sum_{j=1}^{N} X_j\,. \tag{39}$$

We call M_N the *voting result*. It turns out that M_N converges for large N in the sense that $\mathbb{P}_{R,B,c}(M_N \in [u, v])$ has a limit as $N \to \infty$. To formulate the limit result, we define:

Definition 9 Suppose $a, b > 0$. The *Beta function* $B(a, b)$ is defined by

$$B(a,b) \;=\; \int_0^1 x^{\,a-1}\;(1-x)^{b-1}\;dx \tag{40}$$

The *Beta distribution* is the probability measure $\beta_{a,b}$ on $[0, 1]$ with density

$$\widetilde{\beta}_{a,b}(x) \;=\; \begin{cases} \dfrac{1}{B(a,b)}\,x^{\,a-1}\;(1-x)^{b-1}\,, & \text{for } x \in (0,1); \\ 0, & \text{otherwise.} \end{cases} \tag{41}$$

For the following theorem, we suppose that $R, B, c > 0$ and set as above $r := \frac{R}{c}$ and $b := \frac{B}{c}$.

Theorem 1 *For all $u, v \in \mathbb{R}$, $u < v$*

$$\mathbb{P}_{R,B,c}\left(M_N \in [u,v]\right) \underset{N\to\infty}{\longrightarrow} \beta_{r,b}\left([u,v]\right)\,. \tag{42}$$

We sketch a proof of Theorem 1 in A.2.

Let us have a closer look at the case $R = B$ when $\mathbb{P}_{R,B,c}$ is a voting measure. Then for $a = \frac{R}{c} = \frac{B}{c} < 1$ small, the distribution on the voting result M_N is very much concentrated near unanimous rejection (0) and unanimous acceptance (1). In fact, the density function $\widetilde{\beta}_{a,a}$ goes to infinity near 0 and 1 for $a < 1$. This indicates that the voters act cooperatively. In the limit $a \searrow 0$, the distribution $\beta_{a,a}$ tends to a Dirac measure concentrated in $\{0, 1\}$ which means that voters vote always unanimously, with probability one half 'yeah' and with probability one half 'nay'. Thus, as $a \searrow 0$ the measure $\beta_{a,a}$ tends to the unanimity measure introduced in (19).

If in the contrary $a > 1$ is big, then the voting result concentrates more and more near $\frac{1}{2}$, indicating close elections with voters acting almost independently. In fact, in the limit $a \to \infty$, we recover the case of independent voting, i.e. the Banzhaf measure. For $a = 1$, we obtain the Shapley–Shubik measure which corresponds to the uniform distribution on $[0, 1]$.

4 Permutation Invariant Voting Systems

In this section, we classify voting systems and voting measures which are invariant under permutations of the voters.

Definition 10 A *permutation* on the set V is a bijective map $\pi : V \to V$, i.e. a reordering of V. For a set $A \subset V$, we set $\pi(A) := \{\pi(v) \mid v \in A\}$.

We call a voting system $(V, \mathcal{V})$ *permutation invariant* (or *invariant* for short) if for any permutation π, $A \in \mathcal{V}$ implies $\pi(A) \in \mathcal{V}$.

Invariant voting systems are easy to characterize: They obey the rule "One person, one vote!".

Proposition 2 *Every permutation invariant voting system* $(V, \mathcal{V})$ *is a weighted voting system. The weights can be chosen to be equal to* 1 *for all voters in* V.

Proof If coalitions A and B in V contain the same number of voters, then there is a permutation on V that maps A bijective onto B. It follows that $A \in \mathcal{V}$ if and only if $B \in \mathcal{V}$. In other words, whether A is winning depends only on the cardinality $|A|$ of A.

Denote by q the smallest number such that $|A| = q$ implies $A \in \mathcal{V}$. Then, by monotonicity of $\mathcal{V}$, $|B| \geq q$ implies $B \in \mathcal{V}$. Since q is the smallest such number $|B| < q$ implies $B \notin \mathcal{V}$.

Thus $(V, \mathcal{V})$ is a weighted voting system with weights $w(v) \equiv 1$ and quota q. □

Definition 11 Suppose V is a finite set. A measure $\mathbb{P}$ on V is called *permutation invariant* or *exchangeable* if $\mathbb{P}(A) = \mathbb{P}(\pi(A))$ for each $A \subset V$ and each permutation π on V.

All voting measures introduced in Example 1 as well as the urn models discussed in Sect. 3 are exchangeable.

Since we are interested in the behavior of quantities like power indices and success rates for *large* voting system, we concentrate on voting measures which can be extended to arbitrary large sets in a natural way, i.e. such that the extension is still exchangeable.

If the set V of voters has N elements, we may set $V = \{1, 2, \ldots, N\}$ without loss of generality and consider a voting measure as a measure on $\{0, 1\}^N$ as in (10).

Definition 12 We call an exchangeable measure $\mathbb{P}$ on $\{0, 1\}^N$ *extendable* if for every $N' > N$ there is an exchangeable measure $\mathbb{P}'$ on $\{0, 1\}^{N'}$ such that $\mathbb{P}$ is the restriction of $\mathbb{P}'$ on $\{0, 1\}^N$.

The voting measures of Example 1 are extendable. The urn models are extendable for $c \geq 0$, but not for $c = -1$. The following theorem says that a voting measure which is exchangeable and extendable is actually a Common Believe Model.

Theorem 2 *Suppose* $\mathbb{P}$ *is an exchangeable and extendable voting measure on* $V = \{1, 2, \ldots, N\}$ *then* $\mathbb{P}$ *is a* common belief measure *(see Example 1.4), i.e. there is a probability measure* μ *on* $[0, 1]$ *such that*

$$\mathbb{P}_\mu\big(A\big) \;=\; \int_0^1 P_p\big(A\big)\, d\mu(p)\,. \tag{43}$$

If $\mathbb{P}$ is a voting measure, then (20) holds.

Theorem 2 is a version of the celebrated theorem of de Finetti (de Finetti, 1931). De Finetti's theorem can be found at various places and in various formulations, see, e.g., (Aldous, 1985) or (Klenke, 2014). For an introduction and an elementary proof, see (Kirsch, 2019).

We mentioned already that the Banzhaf measure admits a representation as a CB-model with $\mu = \delta_{\frac{1}{2}}$. Moreover, the Shapley–Shubik measure is the CB-model with μ the uniform distribution on [0, 1]. From Theorem 2, we know that also the urn models with $c \geq 0$ can be written in the form (43). Theorem 1 identifies the corresponding measures μ for the urn models.

Theorem 3 *Suppose $\mathbb{P}_{A,B,c}$ is the probability measure (42) with parameters $A, B, c > 0$ and set $a = \frac{A}{c}$ and $b = \frac{B}{c}$. Then for $(x_1, \dots, x_N) \in \{0, 1\}^N$*

$$\mathbb{P}_{A,B,c}\,(x_1, \dots, x_N) \;=\; \int_0^1 P_p\,(x_1, \dots, x_N)\;\widetilde{\beta}_{a,b}(p)\, dp\,. \tag{44}$$

Theorem 3 follows from de Finetti's theorem and Theorem 1. For details, see, e.g., (Kirsch, 2019).

Theorem 3 identifies the measure μ abstractly given in Theorem 2 for the special case of an urn model. The representation (44) allows a computation of the left-hand side even for large N.

5 Penrose–Banzhaf Versus Shapley–Shubik

In this section, we consider the behavior of efficiency, decisiveness and rate of success in simple voting systems under the Penrose–Banzhaf and the Shapley–Shubik measure. Our first result is

Proposition 3 *Let $(V, \mathcal{V})$ be a simple majority voting system (all weights equal to 1, relative quota $r = \frac{1}{2}$) with N voters then*

1.
$$D_B(v) \approx \frac{2}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{N}} \text{ as N} \to \infty \qquad \text{and} \tag{45}$$

$$E_B \to \frac{1}{2} \textit{ as } N \to \infty\,. \tag{46}$$

2.
$$D_S(v) = \frac{1}{N} \text{ for all N} \qquad \text{and} \tag{47}$$
$$E_S \to \frac{1}{2} \text{ as N} \to \infty\,. \tag{48}$$

Definition 13 By $a_N \approx b_N$ we mean $\frac{a_N}{b_N} \to 1$ as $N \to \infty$.

Proof The proof of 1. is quite standard, see, for example, (Felsenthal, 1998).
2. follows from the fact that $\sum_{v\in V} D_S(v) = 1$ and $D_S(v) = D_S(v^{'})$ for all $v, v^{'}$. □

Remark 2 A calculation shows that in simple majority systems for odd N

$$D_S^-(v) = D_S^+(v) = \frac{1}{2}\frac{1}{N}\,, \tag{49}$$

and for even N

$$D_S^-(v) = \frac{1}{2}\frac{1}{N} - \frac{1}{2\,N(N+1)}\,, \tag{50}$$
$$D_S^+(v) = \frac{1}{2}\frac{1}{N} + \frac{1}{2\,N(N+1)}\,. \tag{51}$$

Proposition 3 has an immediate consequence for the success rate of voters. From (27), we infer that

$$S_B(v) \approx \frac{1}{2} + \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{N}} \tag{52}$$

for simple majority voting systems.
As one might expect the Penrose–Banzhaf power goes to zero as N increases and the success rate goes to $\frac{1}{2}$, the success rate (in the Banzhaf setting) of a dummy player. The Shapley–Shubik power goes to zero as $N \to \infty$ as well, in fact, even faster than the Penrose–Banzhaf power (see Proposition 3).

It may be somewhat surprising that the Shapley–Shubik success rate does *not* go to $\frac{1}{2}$ for large N, but rather stays at about $\frac{3}{4}$ independent of the size of V. We will prove this fact in greater generality below.
Now, we turn to simple voting systems with a qualified majority; i.e., we consider weighted voting systems with weights $w(v) = 1$ and arbitrary relative quota r. First, we look at the behavior of the efficiency for fixed r and large N.

Theorem 4 *Suppose* $(V, \mathcal{V})$ *is a weighted system with* N *voters, with weights* $w(v) = 1$ *for all* $v \in V$ *and relative quota* r. *Then*

1. $E_B \to \begin{cases} 1 & \text{for r} < \frac{1}{2} \\ \frac{1}{2} & \text{for r} = \frac{1}{2} \\ 0 & \text{for r} > \frac{1}{2} \end{cases} \quad \text{as N} \to \infty,$
2. *For* $r > \frac{1}{2}$

$$E_B \leq e^{-2(r-\frac{1}{2})^2 N} . \tag{53}$$

This theorem tells us that the (Banzhaf) efficiency of a voting system goes extremely fast to zero if the voting body is enlarged and the relative quota is kept fixed at $r > \frac{1}{2}$. In Sect. 6, we will extend this result to *weighted* voting system and discuss its implications for voting in the Council of the European Union in Sect. 7.

Proof Part 1. follows from the strong law of large numbers (see e.g., (Klenke, 2014)) and Proposition 3.

2. is an application of Hoeffding's inequality (see Sect. A.1 in the appendix and (Pollard, 1984)). □

In contrast to the above result, the efficiency as measured by Shapley–Shubik does not go to zero as $N \to \infty$ for $r > 1/2$.

Theorem 5 *Suppose* $(V, \mathcal{V})$ *is a weighted voting system with* N *voters, weights* $w(v) = 1$ *for all* $v \in V$ *and relative quota* r.

$$E_S \to (1-r) \text{ as } N \to \infty. \tag{54}$$

Proof From (23), we infer

$$E_S = \mathbb{P}_S(\sum X_i \geq rN) \tag{55}$$

$$= \int_0^1 P_p(\frac{1}{N} \sum X_i \geq r)\, dp \tag{56}$$

The expression under the integral in (56)

$$P_p(\frac{1}{N} \sum X_i \geq r) \tag{57}$$

is the probability with respect to P_p that the arithmetic mean of the X_i is not less than r. The random variables X_i are independent under the measure P_p. Thus, we may apply the law of large numbers to show that this expression goes to 0 for $r > p$ and to 1 for $r < p$. Consequently, (56) converges to

$$\int_r^1 dp = 1 - r \tag{58}$$

We turn to an investigation of the success rate.

Before we consider the case of arbitrary r, we discuss in detail the case $r = \frac{1}{2}$. It is quite obvious that for simple majority systems

$$S_B^+(v) \to \frac{1}{4} \text{and } \mathrm{S_B^-(v)} \to \frac{1}{4} \text{ as N} \to \infty\,, \tag{59}$$

in particular:

$$S_B(v) \to \tfrac{1}{2}\,. \tag{60}$$

The following result about S_S^+ and S_S^- is perhaps not so obvious.

Theorem 6 *Let $(V, \mathcal{V})$ be a simple majority voting system with N voters, then*

$$S_S^+(v) = \begin{cases} \frac{3}{8} + \frac{1}{8}\frac{1}{N} & \text{for odd N} \\ \frac{3}{8} - \frac{1}{8}\frac{1}{N+1} & \text{for even N} \end{cases} \tag{61}$$

$$S_S^-(v) = \begin{cases} \frac{3}{8} + \frac{1}{8}\frac{1}{N} & \text{for odd N} \\ \frac{3}{8} + \frac{3}{8}\frac{1}{N+1} & \text{for even N} \end{cases} \tag{62}$$

Consequently,

$$S_S(v) = \begin{cases} \frac{3}{4} + \frac{1}{4}\frac{1}{N} & \text{for odd N} \\ \frac{3}{4} + \frac{1}{4}\frac{1}{N+1} & \text{for even N} \end{cases} \tag{63}$$

In particular,

$$S_S(v) \to \frac{3}{4} \text{ as N} \to \infty \tag{64}$$

Proof We may assume that $V = \{1, 2, \ldots, N\}$ and $v = 1$. Let us start with N odd, say $N = 2n + 1$.

Then,

$$\begin{aligned}
S_S^+(1) &= \mathbb{P}_S(x_1 = 1, \sum_{i=2}^N x_i \geq n) \\
&= \int_0^1 P_p(x_1 = 1, \sum_{i=2}^N x_i \geq n)\, dp \\
&= \int_0^1 p \cdot P_p(\sum_{i=2}^N x_i \geq n)\, dp \\
&= \sum_{k=n}^{N-1} \binom{N-1}{k} \int_0^1 p^{k+1}(1-p)^{N-(k+1)}\, dp \\
&= \frac{1}{N+1} \sum_{k=n}^{N-1} \frac{\binom{N-1}{k}}{\binom{N}{k+1}} = \frac{1}{N+1} \sum_{k=n}^{N-1} \frac{k+1}{N} \\
&= \frac{1}{N(N+1)} \sum_{k=n+1}^{N} k = \frac{1}{N(N+1)} \frac{1}{2}\big(N(N+1) - n(n+1)\big) \\
&= \frac{1}{2} - \frac{1}{2} \frac{n(n+1)}{N(N+1)} = \frac{1}{2} - \frac{1}{8} \frac{(N-1)(N+1)}{N(N+1)} \\
&= \frac{3}{8} + \frac{1}{8} \frac{1}{N}
\end{aligned} \tag{65}$$

$$\begin{aligned}
S_S^-(1) &= \mathbb{P}_S(x_1 = 0, \sum_{i=2}^N x_i \leq n) \\
&= \sum_{k=0}^{n} \binom{N-1}{k} \int_0^1 (1-p) p^k (1-p)^{N-k-1}\, dp \\
&= \sum_{k=0}^{n} \binom{N-1}{k} \int_0^1 p^k (1-p)^{N-k}\, dp \\
&= \frac{1}{N+1} \sum_{k=0}^{n} \frac{\binom{N-1}{k}}{\binom{N}{k}} \\
&= \frac{1}{N(N+1)} \sum_{k=0}^{n} (N-k) = \frac{1}{N(N+1)} \big((n+1)N - \frac{1}{2} n(n+1)\big) \\
&= \frac{3}{8} + \frac{1}{8} \frac{1}{N}
\end{aligned} \tag{66}$$

Thus, for N odd, we have

$$S_S^+(1) = S_S^-(1) = \frac{3}{8} + \frac{1}{8}\frac{1}{N} \tag{67}$$

$$\text{and } \mathrm{S_S}(1) = \frac{3}{4} + \frac{1}{4}\frac{1}{\mathrm{N}} \to \frac{3}{4} \tag{68}$$

The calculation for even N goes along the same lines. □

Theorem 7 *Suppose* $(V, \mathcal{V})$ *is a weighted voting system with* N *voters, weights* $w(v) = 1$ *for all* $v \in V$ *and relative quota* r.

1. For arbitrary r

$$S_B(v) \to \frac{1}{2} \text{ as N} \to \infty \tag{69}$$

For $r > \frac{1}{2}$

$$S_B^+(v) \le \frac{1}{2} e^{-2(r-\frac{1}{2})^2(N-1)} \tag{70}$$

2. For arbitrary r

$$S_S^+(v) \to \frac{1}{2} - \frac{1}{2} r^2 \text{ as N} \to \infty \tag{71}$$

$$S_S^-(v) \to \frac{1}{2} - \frac{1}{2}(1-r)^2 \text{ as N} \to \infty \tag{72}$$

Consequently, as $N \to \infty$

$$S_S(v) \;\to\; 1 - \frac{1}{2}(r^2 + (1-r)^2) \;=\; \tfrac{1}{2} + r(1-r) \tag{73}$$

Remark 3 $S_S(v) \approx 1 - \frac{1}{2}(r^2 + (1-r)^2)$ is biggest for $r = \frac{1}{2}$ where it equals $\frac{3}{4}$ and smallest for $r = 0$ and $r = 1$ where it is $\frac{1}{2}$.

Proof 1. We have

$$S_B^+(1) = \mathbb{P}_B(x_1 = 1, \sum_{i=2}^{N} x_i \ge rN - 1) \tag{74}$$

$$= \frac{1}{2}\mathbb{P}_B(\sum_{i=2}^{N} x_i \ge rN - 1) \tag{75}$$

since under $\mathbb{P}_B$ the x_i are independent.
Another application of Hoeffding's inequality gives (70).

Similarly

$$S_B^-(1) = \frac{1}{2}\mathbb{P}_B(\sum_{i=2}^{N} x_i < rN)$$

so

$$\begin{aligned} S_B(1) &= \frac{1}{2}(1 + \mathbb{P}_B(\sum_{i=2}^{N} x_i \in [rN-1, rN))) \\ &\leq \frac{1}{2} + C\frac{1}{\sqrt{N}} \end{aligned} \tag{76}$$

2. A computation as in the proof of Theorem 6 shows that

$$\begin{aligned} &\mathbb{P}_S(x_1 = 1, \sum_{i=2}^{N} x_i \geq M) \\ &= \frac{1}{2} - \frac{1}{2}\frac{M(M+1)}{N(N+1)} \end{aligned} \tag{77}$$

If we insert $M = \lceil rN \rceil - 1$, where $\lceil x \rceil$ is the smallest integer not less than x, we obtain

$$\begin{aligned} S_S^+(1) &= \frac{1}{2} - \frac{1}{2}\frac{(\lceil rN \rceil - 1)\lceil rN \rceil}{N(N+1)} \\ &\to \frac{1}{2} - \frac{1}{2}r^2 \text{ as N} \to \infty \end{aligned} \tag{78}$$

Similarly

$$\begin{aligned} S_S^-(1) &= \mathbb{P}_S(x_1 = 0, \sum_{i=2}^{N} x_i \leq \lceil rN \rceil - 1) \\ &= \mathbb{P}_S(x_1 = 1, \sum_{i=2}^{N} x_i \geq N - \lceil rN \rceil) \\ &= \frac{1}{2} - \frac{1}{2}\frac{(N - \lceil rN \rceil)(N - \lceil rN \rceil + 1)}{N(N+1)} \\ &\to \frac{1}{2} - \frac{1}{2}(1-r)^2 \end{aligned} \tag{79}$$

6 Weighted Voting and the Common Belief Model

In this section, we discuss efficiency in *weighted* voting systems under the rather general case of the Common Belief model. The concepts and results in this section are mathematically slightly more involved than those in the previous sections but the have direct consequences for voting systems like the Council of the European Union.

We start with an analysis of success under the Banzhaf measure. In fact, our result in this respect is quite similar to Theorem 4. There, we showed that the efficiency decays exponentially in the *number* N of voters if the relative quota r is bigger than $\frac{1}{2}$. We will argue that for weighted voting systems N has to be replaced by the 'effective number' of voters as introduced by Laakso–Taagepera in (Laakso and Taagepera, 1979).

Definition 14 Let $\mathcal{V}$ be a weighted voting system with weights $W = \{w_1, \ldots, w_N\}$. We define the Laakso–Taagepera index of the sequence W by

$$\mathrm{LT} \;=\; \mathrm{LT}(W) \;=\; \frac{\left(\sum_{n=1}^{N} w_n\right)^2}{\sum_{n=1}^{N} w_n^2} \tag{80}$$

In a voting system comprising N voters with equal voting weights the Laakso–Taagepera index equals N. If there is one dominant voter and all other voters have almost negligible weight, the LT is close to 1.

We also remark that $a \le w_i \le b$ for all i implies $\frac{a}{b}N \;\le \mathrm{LT} \le\; \frac{b}{a}N$.

Theorem 8 *Suppose $\mathcal{V}$ is a weighted voting system with weights in $W = \{w_1, \ldots, w_N\}$.*

If the relative quota r satisfies $r > \frac{1}{2}$, then the efficiency E_B of the voting system under the Banzhaf voting measure satisfies:

$$E_B \;\le\; e^{-2(r-\frac{1}{2})^2\,\mathrm{LT}} \tag{81}$$

Proof If $X_i \in \{0, 1\}$ denotes the voting of the i^{th} voter then the X_i are independent identically distributed random variables under the Banzhaf measure. Consequently, the quantities $Y_i = w_i X_i$ are also independent (but not identically distributed in general). We have $\mathbb{E}(Y_i) = \frac{1}{2} w_i$ and $Y_i \in [0, w_i]$.

A coalition is winning if

$$\sum w_i X_i \;\ge\; r \sum_{i=1}^{N} w_i\,. \tag{82}$$

We estimate the probability that (82) is true using Hoeffding's inequality (Theorem 1) with $\sigma^2 = \sum w_i^2$ and $\lambda = (r - \frac{1}{2}) \sum w_i$. □

Now, we consider an infinite sequence of nonnegative weights $\{w_i\}_{i\in\mathbb{N}}$, and weighted voting systems $\mathcal{V}_N$ with N voters, weight $w_1, \ldots, w_N$ and a fixed relative quota r.

This serves as a model for an assembly (e.g., a union of states like the EU), which is enlarged by and by as N is increased.

The original form of the following result goes back to (Langner, 2012).

Theorem 9 *If $LT_N \to \infty$ and $\mu(\{r\}) = 0$ then the efficiency E_N under the voting measure $\mathbb{P}_\mu$ of the voting systems $(V_N, \mathcal{V}_N)$ satisfies*

$$E_N \to \mu([r, 1]) \tag{83}$$

Proof

$$E_N \;=\; \int_0^1 P_p\big(\sum_{i=1}^{N} w_i X_i \geq r \sum_{i=1}^{N} w_i\big)\, d\mu(p) \tag{84}$$

By Corollary 1 the integrand converges to 0 for $r > p$ and to 1 for $r < p$, hence

$$E_N \;\to\; \int_0^1 \chi_{\{p>r\}}(p)\, d\mu(p) \;=\; \mu([r, 1])\,. \tag{85}$$

where

$$\chi_{p>r}(p) \;=\; \begin{cases} 1, & \text{if p > r;} \\ 0, & \text{otherwise.} \end{cases}$$

We estimate

$$\begin{aligned} & P_p(|\sum w_i x_i - p \sum w_i| \geq \alpha \sum w_i) \\ & \leq \frac{1}{\alpha^2} \frac{\sum_i E_p((\sum w_i (x_i - p))^2)}{(\sum w_i)^2} \\ & = \frac{1}{\alpha^2} p(1-p) \frac{\sum_i w_i^2}{(\sum w_i)^2} = \frac{p(1-p)}{\alpha^2} LT_N \end{aligned} \tag{86}$$

It follows that

$$E_N = \int_0^1 P_p(\sum w_i x_i \geq r \sum w_i)\, d\mu(p) \tag{87}$$

converges to

$$\int_0^1 \chi(p > r)\, d\mu(p) = \mu([r, 1]) \tag{88}$$

In a similar way, we can compute the rate of success in such systems.

Theorem 10 *If $LT_N \to \infty$ and $\mu(\{r\}) = 0$, then the rate of success with respect to $\mathbb{P}_\mu$ satisfies*

$$S^+_{\mathbb{P}_\mu}(v) \to \int_r^1 p\, d\mu(p) \tag{89}$$

$$S^-_{\mathbb{P}_\mu}(v) \to \int_{r-1}^1 p\, d\mu(p) \tag{90}$$

Proof Without loss, we compute the rate of success for voter $i = 1$.

$$\begin{aligned} S^+_{\mathbb{P}_\mu}(1) &= \int_0^1 P_p(x_1 = 1, \sum_{i=2}^N w_i x_i \geq r \sum w_i - w_1)\, d\mu(p) \\ &= \int_0^1 p P_p(\sum_{i=2}^N w_i x_i \geq r \sum w_i - w_1)\, d\mu(p) \\ &\to \int_r^1 p\, d\mu(p) \end{aligned} \tag{91}$$

Remark 4 1. Observe that the success probability of a voter is asymptotically *independent* of the voters weight!
2. In the case of the Shapley–Shubik measure, i. e. if μ is the uniform distribution, Theorem 10 contains Theorem 7 as a special case.
3. Theorem 3 in combination with Theorem 10 allows to evaluate the success for the urn models with parameters $a = b$ (see Sect. 3). The following table lists a few values of $S(a, r)$, the success of a voter under the voting measure $\mathbb{P}_{\beta_{a,a}}$ with relative quota r.

a	$S(a, \frac{1}{2})$	$S(a, \frac{2}{3})$
0.1	0.942	0.936
0.5	0.82	0.80
1	0.75	0.72
2	0.69	0.65
10	0.59	0.53

7 The Council of the EU: A Case Study

The voting procedure in the Council of the European Union was fixed in the Treaty of Lisbon. It is usually referred to as the ‘Double Majority’ system although strictly speaking it is a ‘triple criteria’ system.

In many fields of politics, a proposal of the Commission is adopted by the Council if the supporters represent both 65% of the total population of the EU and 55% of the states. There is a third rule—in addition to the double majority explained before. This third rule says that a proposal is also adopted if only 3 states vote against it—even if the supporters represent less than 65% of the population of the EU.

The results above cannot be applied directly to this ‘double majority’ system since it is not a weighted voting system. In fact, as noted by (Kurz and Napel, 2016) this voting system has dimension at least seven (probably more); i.e., it cannot be written as the intersection of six or less weighted voting systems. To apply the above reasoning (in particular Theorem 8), we formalize the voting rules in the Council.

Denote by $V = \{1, \ldots, N\}$ the N member states of the EU (currently 27) and by w_i the size of the population of state i and denote for $A \subset V$ by $|A|$ the number of elements (states) in the set A. We define

$$\begin{aligned} \mathcal{V}_1 \;&=\; \{A \mid \sum_{i\in A} w_i \;\geq\; 0.65 \sum_{i\in V} w_i\} \\ \mathcal{V}_2 \;&=\; \{A \mid |A| \;\geq\; 0.55\, N\} \qquad \text{and} \\ \mathcal{V}_3 \;&=\; \{A \mid |A| \;\geq\; N-3\}\,. \end{aligned}$$

The winning coalitions in the Council are the given by:

$$\mathcal{V} \;=\; \big(\mathcal{V}_1 \cap \mathcal{V}_2\big) \cup \mathcal{V}_3\,. \tag{92}$$

Let us denote by $E_N^{(1)}$, $E_N^{(2)}$ and $E_N^{(3)}$ the Banzhaf efficiency of the voting systems $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$, respectively, where N is the number of member states. Then, it is easy to see that the efficiency E_N of the voting system $\mathcal{V}$ of the Council satisfies:

$$E_N \;\leq\; \min(E_N^1, E_N^2) \;+\; E_N^3 \;\leq E_N^2 \;+\; E_N^3\,. \tag{93}$$

where the second inequality is quite rough. Applying Theorem 4, we obtain

$$E_N \;\leq\; 2\, e^{-c\,N} \tag{94}$$

where we may choose $c = 0.005$ (as long as $N \geq 7$).

This estimate shows that the Banzhaf efficiency of the voting system in the Council of the EU decays exponentially fast in the number of member states. The above estimate is so rough that it even does not come close to the true value 0.13 for the current EU; however, it shows that the efficiency of the system goes down with each accession of new members, which may be regarded as a blunder in the Lisbon Treaty.

8 Conclusions and Outlook

In this paper, we considered voting systems equipped with voting measures. The voting measure models the collective behavior of the voters. We gave a collection of example for voting measures and classified those voting measures which are invariant under permutation of the voters as 'collective bias models'.

We investigated efficiency of weighted voting systems depending on the collective behavior of the voters. Our main interest in addition to efficiency was the rate of success of voters in various situation. In particular, we computed the rate of success in the Shapley–Shubik and in the Penrose–Banzhaf setting.

Voting measures play a prominent role in the analysis and design of two-tier voting systems. Optimal weights for the upper tier were investigated in (Kirsch, 2007) for general CB-models with independent groups (e. g. states) on the lower level. Newer developments allow also correlations between voters in different groups of the two-tier voting system (Kirsch and Langner, 2014; Kirsch, 2021).

Acknowledgements The author thanks Alexander Mayer and Stefan Napel for fruitful discussions about the topic of this paper. The author is also grateful to two unnamed referees whose comments helped to improve the manuscript considerably.

A Mathematical Appendix

A.1 Hoeffding's Inequality

For the reader's convenience, in this appendix, we present a few mathematical results needed in the main text. In particular, we formulate Hoeffding's inequality.

Theorem 1 (Hoeffding's Inequality)
Suppose $X_i, i = 1, \dots, N$ *are independent random variables such that* $X_i \in [a_i, b_i]$ *almost surely.*

Set $\sigma^2 = \sum_{i=1}^{N} (b_i - a_i)^2$. *Then*

$$\mathbb{P}\Big(\sum_{i=1}^{N} X_i \;\geq\; \sum_{i=1}^{N} \mathbb{E}(X_i) + \lambda \Big) \;\leq\; e^{-2\frac{\lambda^2}{\sigma^2}} \qquad \text{and} \tag{A.1}$$

$$\mathbb{P}\Big(\sum_{i=1}^{N} X_i \;\leq\; \sum_{i=1}^{N} \mathbb{E}(X_i) - \lambda \Big) \;\leq\; e^{-2\frac{\lambda^2}{\sigma^2}} \tag{A.2}$$

For a proof of Theorem 1, see e.g., (Pollard, 1984).

An immediate consequence of Hoeffding's inequality is the following proposition. As before P_p with $0 \leq p \leq 1$ denotes the probability measure on $\{0, 1\}^N$ given by:

$$P_p\big(x_1, x_2, \ldots, x_N\big) \;=\; p^{\sum x_i}\,\big(1-p\big)^{N-\sum x_i} \tag{A.3}$$

and E_p denotes expectation with respect to P_p.

Proposition 1 *Let $X_i, i = 1, \ldots, N$ be random variables with distribution P_p and $w_1, \ldots, w_N \in [0, \infty)$, then for $\lambda \geq 0$*

$$P_p\Big(\big|\sum_{i=1}^{N} w_i X_i - p\sum_{i=1}^{N} w_i\big| \geq \alpha \sum_{i=1}^{N} w_i\Big) \leq 2\,e^{-2\alpha^2\,\frac{(\sum w_i)^2}{\sum w_i^2}} \tag{A.4}$$

Proof The random variables $Y_i = w_i X_i$ are independent (under P_p) and take values in $[0, w_i]]$. Moreover, $E_p(Y_i) = p w_i$. Thus, (A.4) follows from Theorem 1. □

Corollary 1 *Suppose the Laakso–Taagepera index* $\mathrm{LT_N} = \frac{\sum \mathrm{w_i^2}}{(\sum \mathrm{w_i})^2}$ *goes to infinity as $N \to \infty$ then*

$$\text{If}\alpha > \text{p} \quad P_p\Big(\sum_{i=1}^{N} w_i X_i \;\geq \alpha \sum_{i=1}^{N} w_i\Big) \;\to\; 0 \quad \text{as N} \to \infty \tag{A.5}$$

$$\text{If}\alpha < \text{p} \quad P_p\Big(\sum_{i=1}^{N} w_i X_i \;\leq \alpha \sum_{i=1}^{N} w_i\Big) \;\to\; 0 \quad \text{as N} \to \infty \tag{A.6}$$

A.2 Urn Models

We start by the following observation:

Proposition 2 *Suppose $c \geq 0$. For $(x_1, x_2, \ldots, x_N) \in \{0, 1\}^N$ and $1 \leq k < N$ we set $n_k = \sum_{j=1}^{k} x_j$. Then*

$$\begin{aligned} &\mathbb{P}_{A,B,c}\,(x_1, x_2, \ldots, x_N) \\ =\; &\mathbb{P}_{A,B,c}\,(x_1, x_2, \ldots, x_k)\;\mathbb{P}_{A+n_k c, B+(k-n_k)c, c}\,(x_{k+1}, x_{k+2}, \ldots, x_N) \end{aligned} \tag{A.7}$$

Proof For $k = 1$, (A.7) is just the definition of the urn process. The general case follows by iterating the first step. □

To prove 1.2 we start with a special case

Proposition 3

$$\begin{aligned} &\mathbb{P}_{A,B,c}\,(x_1, \ldots, x_{k-1}, x_k, x_{k+1}, x_{k+2}, \ldots, x_N) \\ =\; &\mathbb{P}_{A,B,c}\,(x_1, \ldots, x_{k-1}, x_{k+1}, x_k, x_{k+2}, \ldots, x_N) \end{aligned} \tag{A.8}$$

Proof If $x_k = x_{k+1}$, the assertion is trivial. So, we assume $x_k \neq x_{k+1}$ We treat the case $x_k = 0, x_{k+1} = 1$, the other one being similar. We apply Proposition 2 three times to obtain with $\ell = \sum_j^{k-1} x_j$

$$\begin{aligned}
&\mathbb{P}_{A,B,c}\,(x_1,\ldots,x_{k-1},x_k,x_{k+1},x_{k+2},\ldots,x_N)\\
=\;&\mathbb{P}_{A,B,c}\,(x_1,\ldots,x_{k-1})\;\mathbb{P}_{A+\ell c,B+(k-1-\ell)c,c}(0)\;\mathbb{P}_{A+\ell c,B+(k-\ell)c,c}(1)\,\times\\
&\times\,\mathbb{P}_{A+(\ell+1)c,B+(k-\ell),c}\,(x_{k+2},\ldots,x_N)\\
=\;&\mathbb{P}_{A,B,c}\,(x_1,\ldots,x_{k-1})\;\frac{B+(k-1-\ell)c}{A+B+(k-1)c}\;\frac{A+\ell c}{A+B+kc}\,\times\\
&\times\,\mathbb{P}_{A+(\ell+1)c,B+(k-\ell),c}\,(x_{k+2},\ldots,x_N)\\
=\;&\mathbb{P}_{A,B,c}\,(x_1,\ldots,x_{k-1})\;\frac{A+\ell c}{A+B+(k-1)c}\;\frac{B+(k-(\ell+1))\,c}{A+B+kc}\,\times\\
&\times\,\mathbb{P}_{A+(\ell+1)c,B+(k-\ell),c}\,(x_{k+2},\ldots,x_N)\\
=\;&\mathbb{P}_{A,B,c}\,(x_1,\ldots,x_{k-1},x_{k+1},x_k,x_{k+2},\ldots,x_N)
\end{aligned}\tag{A.9}$$

Iterating Propositions 2 and 3 proves Proposition 1. To prove Theorem 1, we first note:

Remark 1 For $a, b > 0$ we have

$$B(a,b) \;=\; \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,, \tag{A.10}$$

where $\Gamma(x) = \int\limits_0^\infty t^{x-1}\, e^{-t}\, dt$ is the Gamma function.

For a proof as well as for details about the Gamma function, see e.g., (Georgii, 2008).

We now prove (42) by proving that all moments of the measures on the left-hand side converge to the moments of the right-hand side.

From (Kirsch, 2019), Theorem 5, we learn that the k^{th} moment under $\mathbb{P}_{A,B,c}$ of M_N converges to

$$m_k \;:=\; \mathbb{E}_{A,B,c}\,(X_1 \cdot X_2 \cdot \ldots \cdot X_k) \tag{A.11}$$

but m_k can be computed by (34) giving

$$m_k \;=\; \frac{a^{\,(1,k)}}{(a+b)^{(1,k)}} \tag{A.12}$$

The moments of $\beta_{a,b}$ are given by $\bar{m}_k = \frac{B(a+k,b)}{B(a,b)}$. By Proposition 1

$$
\begin{aligned}
B(a+k,b) &= \frac{\Gamma(a+k)\Gamma(b)}{\Gamma(a+b+k)} \\
&= \frac{a^{(k,1)}}{(a+b)^{(k,1)}}\,\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,. \qquad \text{(A.13)}
\end{aligned}
$$

Above we used that $\Gamma(x+k) = x^{(k,1)}\Gamma(x)$ which follows from the well-known equality $\Gamma(x) = (x-1)\Gamma(x-1)$. This proves 1.

References

Aldous, D. J. (1985). Exchangeability and related topics. In P. L. Hennequin (Ed.), *École d'été de probabilités de Saint-Flour XIII – 1983, Lecture notes in mathematics* (Vol. 1117, pp. 1–198). Springer

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*, 317–343.

Berg, S. (1985). Paradox of voting under an urn model: The effect of homogeneity. *Public Choice, 47*, 377–387.

Coleman, J. S. (1971). Control of collectivities and the power of a collectivity to act. In B. Lieberman (Ed.), *Social choice* (pp. 269–300). Gordon and Breach.

Dubey, P., & Shapley, L. S. (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research, 4*, 99–131.

de Finetti, B. (1931). Funzione caratteristica di un fenomeno aleatorio. Atti della R. Academia Nazionale dei Lincei, Serie 6. *Memorie, Classe di Scienze Fisiche, Mathematice e Naturale, 4*, 251–299.

Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power: Theory and practice, problems and paradoxes*. Edward Elgar Publishing

Georgii, H.-O. (2008). *Stochastics*. De Gruyter.

Kirsch, W. (2019). An elementary proof of de Finetti's theorem. *Statistics and Probability Letters, 151*, 84–88.

Kirsch, W. (2007). On Penrose's square root law and beyond. *Homo Oeconomicus, 24*, 357–380.

Kirsch, W. (2016). A mathematical view on voting and power. In W. König (Ed.), *Mathematics and society* (pp. 251–279). European Mathematical Society.

Kirsch, W., & Langner, J. (2014). The fate of the square root law for correlated voting. In R. Fara, D. Leech, & M. Salles (Eds.), *Voting power and procedures, Essays in honor of Dan Felsenthal and Moshé Machover, Studies in choice and welfare* (pp. 147–158). Springer.

Kirsch, W., & Toth, G. (2021). Collective bias models in two-tier voting systems and the democracy deficit. Preprint retrieved from https://arxiv.org/abs/2102.12704

Klenke, A. (2014). *Probability theory*. Springer.

Kurz, S., Mayer, A., & Napel, S. (2021). Influence in weighted committees. *European Economic Review, 134*, 103634.

Kurz, S., & Napel, S. (2016). Dimension of the Lisbon voting rules in the EU Council: A challenge and new world record. *Optimization Letters, 10*, 1245–1256.

Kurz, S., & Napel, S. (2018). The roll call interpretation of the Shapley value. *Economics Letters, 173*, 108–112.

Laakso, M., & Taagepera, R. (1979). "Effective" number of parties: A measure with application to West Europe. *Comparative Political Studies, 12*, 3–27.

Langner, J. (2012). *Fairness, efficiency and democracy deficit*. Ph.D. thesis, FernUniversität Hagen

Laruelle, A., & Valenciano, F. (2015). Assessing success and decisiveness in voting situations. *Social Choice and Welfare, 24*, 171–197.

Mahmoud, H. (2009). *Pólya urn models*. Chapman & Hill

Napel, S. (2019). Voting power. In R. Congleton, B. Grofman, & S. Voigt (Eds.), *Oxford handbook of public choice* (pp. 103–126). Oxford University Press.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109*, 53–57.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*, 787–792.

Straffin, P. (1977). Homogeneity, independence, and power indices. *Public Choice, 30*, 107–118.

Taylor, A., & Pacelli, A. (2008): *Mathematics and politics. Strategy, voting, power and proof*. Springer

All Power Structures are Achievable in Basic Weighted Games

Josep Freixas and Montserrat Pons

Abstract A major problem in decision-making is designing voting systems that are as simple as possible and able to reflect a given hierarchy of power of its members. It is known that in the class of weighted games, all hierarchies are achievable except two of them. However, many weighted games are either improper or do not admit a minimum representation in integers or do not assign a minimum weight of 1 to the weakest non-null players. These factors prevent obtaining a good representation. The purpose of the paper is to prove that for each achievable hierarchy for weighted games, there is a handy weighted game fulfilling these three desirable properties. A representation of this type is ideal for the design of a weighted game with a given hierarchy. Moreover, the subclass of weighted games with these properties is considerably smaller than the class of weighted games.

1 Introduction

1.1 Two Motivating Situations

As illustrated below, the motivation for our study is twofold.

Some members provide different contributions to a private company. They want the decision-making in the company to be modeled as a weighted game as simply as possible so that the power of its members respects the order of their individual contributions to the company. The basic principle is as follows: if member A contributes more than member B, then A must have more power than B in the decision-making of the company. Specifically, they want to make decisions through an integer representation of a weighted game that verifies the following desirable properties.

J. Freixas (✉) · M. Pons
Departament de Matemàtiques i Escola Politècnica Superior d'Enginyeria de Manresa, Universitat Politècnica de Catalunya, Manresa, Spain
e-mail: josep.freixas@upc.edu

M. Pons
e-mail: montserrat.pons@upc.edu

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_9

(i) If one member pays more than another, the first will be assigned a greater weight which will grant her more power.
(ii) If a coalition formed by some members is winning (i.e., the sum of the individual weights equals or surpasses the preset quota), then the complementary coalition should be losing.
(iii) It is as simple as possible: It is the unique minimal representation in integers of the weighted game.
(iv) It assigns a weight of 0 to null players (if any) and a weight of 1 (representing the minimum monetary contribution to the company allowed to have right to vote) to the weakest non-null contributors.

Condition (i) assigns the same weight (fee to be paid in our context) to members with the same power and the more weight more power. Condition (ii), properness, avoids the existence of two disjoint winning coalitions, which makes the weighted game stable. Condition (iii), uniqueness of a minimal weighted representation in integers, provides the simplest representation, with the least weights, of the weighted game to be chosen. Condition (iv) allocates a rate of one monetary unit to less influential contributors, and any other contribution is an integer multiple of this minimum amount.

The main question we face in this paper is whether all achievable hierarchies for weighted games are still achievable in the subclass of weighted games that admit a representation fulfilling the above properties. We prove in this paper that the answer is affirmative. The third and fourth conditions are a little bit strict, but their inclusion helps to highlight that all achievable hierarchies in weighted games are also achievable in a very small subclass of them. Thus, in these weighted games, the hierarchy given by the amounts paid to the company coincides with the hierarchy of power of the members in the company.

The following is a slightly different situation. Some members with different degrees of responsibility in a society want to make decisions through a weighted game so that the power of its members respects the ranking of responsibility in the society. In other words, we wish a representation for a weighted game that assigns a weight to each member in the society and a threshold such that the weights respect the preset hierarchy given by the degree of responsibility of its members.

Again, for each achievable hierarchy for weighted games, we wish to find a weighted game that admits a representation in integers fulfilling the four above stated properties.

Both situations involve the design of handy voting structures, a subclass of weighted games, with a pre-established hierarchy. Observe that if n is the number of members of the society, the number of possible hierarchies is 2^{n-1} since the power relation between two consecutive contributors is either strict or equal. It is known, see (Bean et al., 2008), that the number of achievable hierarchies for weighted games is $2^{n-1}-2$. Only two types of hierarchies are not achievable in weighted simple games although they are in a wider class of games (Freixas and Pons, 2010).

The purpose of the paper is to show that all achievable hierarchies for weighted games can be obtained in a tiny subclass of weighted games verifying the desirable properties stated above. From now on, we will refer to this subclass as *basic* weighted games, whose precise definition is given in the next section.

1.2 Background

The number of weighted games, although small when compared with simple games, is quite large. The structure of weighted games can be quite complex, for example, some of them do not have a unique minimal representation in integers (Dubey and Shapley, 1979; Muroga et al., 1970; Freixas and Molinero, 2009, 2010; Kurz, 2012; Freixas and Kurz, 2014), or are not necessarily proper, or the minimum positive weight for a minimal representation in integers can be greater than 1.

Let us start by formally introducing the concept of hierarchy, which specifies the ranking of elements in a complete pre-ordered set. In general, a complete pre-order $\succsim$ in a set N (a reflexive and transitive binary relation for which any two elements of N are related) generates a partition of N into equivalence classes (two elements i and j of N are equivalent if $i \succsim j$ and $j \succsim i$). Then, these equivalence classes $N_1, N_2, \ldots, N_t$ can be strictly ordered in the following way: $N_p \succ N_q$ if and only if $i \succsim j$ but $j \not\succsim i$ for any $i \in N_p$ and any $j \in N_q$. We say that the *hierarchy* given by $\succsim$ is $(n_1, n_2, \ldots, n_t)$ if there are t equivalence classes, with $N_1 \succ N_2 \succ \cdots \succ N_t$, and $n_p = |N_p|$ (n_p is the cardinality of N_p) for $1 \leq p \leq t$.

Many papers in the literature deal with the question of which hierarchies are achievable in particular subclasses of simple games: (Friedman et al., 2006; Bean et al., 2008, 2010; Bean, 2012; Friedman, 2016) for weighted games, (Bishnu and Roy, 2012) for complete games, (Carreras and Freixas, 1996; Freixas and Pons, 2010; Kurz and Tautenhahn, 2013) for weakly complete games, or (Friedman and Parker, 2009) in a general context. Usually, the pre-order considered is the desirability relation (Isbell, 1956, 1958), and a hierarchy is said to be achievable in a class of simple games if there exists a game of this class such that the players have this hierarchy by the desirability relation.

(Bean et al., 2008) proved that any hierarchy in a set of n elements is achievable by the desirability relation in the class of weighted games with the only exception of two of them: $(n-2, 1, 1)$ and $(n-3, 1, 1, 1)$. This result also tells us that three of the most prominent power indices, the Shapley–Shubik index ((Shapley, 1953; Shapley and Shubik, 1954)), the Banzhaf index ((Penrose, 1946; Banzhaf, 1965; Coleman, 1971)), and the Johnston index (Johnston, 1978) produce the same hierarchies in weighted games since in (Freixas et al., 2012), it was proven that these three power indices preserve the desirability relation.

A subclass of weighted games is that of homogeneous games, see (Von Neumann and Morgenstern, 1944; Isbell, 1956). These games admit a weighted representation for which all minimal winning coalitions have equal weight, coinciding with the threshold. Thus, all coalitions with greater weight are winning but not minimal, while coalitions with lower weight are all losing. This property implies that if a player is replaced by a weaker one (by the desirability relation) in a minimal winning coalition, then the new coalition necessarily becomes losing, something which does not occur for the rest of weighted games. These games, see (Ostmann, 1987), admit a representation in integers which is the unique minimal representation verifying that all equivalent players are assigned the same weight, all null players are assigned a weight of 0 (if any), and all voters in the smallest non-null class are assigned a weight of 1. However, homogeneous games are not necessarily proper. Proper homogeneous games verify the desired four stated properties in the previous subsection and therefore are particular cases of basic weighted games, but they are not sufficient to generate all hierarchies achievable in weighted games.

The remainder of the paper is organized as follows. Section 2 contains the basic definitions and concepts on simple games, the usual model to study binary voting systems. Section 3 states the main result of the paper, as well as the three propositions necessary to prove it, and shows some consequences of this main result. Section 4 summarizes the paper and points out the relevance of the obtained results.

2 Binary Voting Systems

Definition 1 Simple game
A *simple game* is a pair (N, W), where $N = \{1, 2, \ldots, n\}$ is the set of players, subsets of N are called *coalitions* and W is a subset of 2^N which satisfies: $\emptyset \notin W$, $N \in W$, and $S \subset T$ and $S \in W$ implies $T \in W$ (monotonicity).

A coalition S is said to be *winning* if $S \in W$ and *losing* otherwise.

Definition 2 Dual game
The *dual* of a simple game (N, W) is the simple game (N, W^*) such that $S \in W^*$ if and only if $N \setminus S \notin W$.

Definition 3 Proper simple game
A simple game (N, W) is *proper* if $S \in W$ implies $N \setminus S \notin W$.

In proper simple games, it is not possible that two disjoint coalitions win. Thus, it is a natural property to be demanded.

Definition 4 A coalition S is a *minimal winning* coalition if $S \in W$ and $T \notin W$ for all coalitions $T \subset S$. The collection of minimal winning coalitions is denoted by W^m.

A player $i \in N$ is *null* in (N, W) if $i \notin S$ for all $S \in W^m$. Null players never add value to a coalition.

A well studied subclass of simple games arises from Isbell's desirability relation (Isbell, 1958).

Definition 5 Desirability relation
We write $i \succsim j$ for two players $i, j \in N$ (we say that i is at least as desirable as j) if

$$S \cup \{j\} \in W \quad \text{implies} \quad S \cup \{i\} \in W \quad \text{for all } \; S \subseteq N \setminus \{i, j\},$$

We write $i \approx j$ whenever $i \succsim j$ and $j \succsim i$ (we say that i and j are equivalent), and $i \succ j$ if $i \succsim j$ but $j \not\succsim i$.
The desirability relation is a pre-order, i.e., it verifies

(1) $i \succsim i$ for all $i \in N$ (reflexive),
(2) $i \succsim j$, $j \succsim h$ implies $i \succsim h$ for all $i, j, h \in N$ (transitivity).

The desirability relation ranks players with respect to how much influential they are. It is a well-known property that the dual game preserves the desirability relation.

Definition 6 Complete game
A simple game W is *complete* if the desirability relation $\succsim$ is a complete (or total) pre-order, i.e., either $i \succsim j$ or $j \succsim i$ for all $i, j \in N$.

Complete games have been intensively studied (see among others, (Carreras and Freixas, 1996; Taylor and Zwicker, 1999; Bishnu and Roy, 2012)) and have been applied to several different fields.

Definition 7 Weighted game
A simple game (N, W) is called a *weighted game* if there exists a positive integer q (quota) and non-negative integers $w_1, \ldots, w_n$ (weights) such that $S \in W$ if and only if $w(S) = \sum_{i \in S} w_i \geq q$. We say that $[q; w_1, \ldots, w_n]$ is a representation of the game, and we write $W \equiv [q; w_1, \ldots, w_n]$ or simply $W \equiv [q; w]$ whenever the weight vector $w = (w_1, \ldots, w_n)$ is specified. Each weighted game has an unbounded number of integer representations.

As $w_i > w_j$ implies $i \succsim j$ and $w_i = w_j$ implies $i \approx j$, for every arbitrary representation $[q; w_1, \ldots, w_n]$ of the weighted game (N, W), it is clear that every weighted game is complete. Moreover, $i \succ j$ implies $w_i > w_j$. Furthermore, every weighted game admits a representation in which equivalent players have the same weight and therefore $w_i > w_j$ if and only if $i \succ j$. From now on, we only deal with this type of representations for weighted games. All these representations satisfy Condition i) stated in the introduction. Thus, from now on, we omit any reference to this condition.

Note that $q > w(N)/2$ for a representation $[q; w]$ of a weighted game implies that the weighted game is proper. It is well-known that a simple game is weighted if and only if its dual game is weighted, and if $[q; w_1, \ldots, w_n]$ is a representation of a weighted game, then $[w(N) - q + 1; w_1, \ldots, w_n]$ is a representation of the dual game. As at least one of these two inequalities $q > w(N)/2$, $(w(N) - q + 1) > w(N)/2$ is verified, either the weighted game is proper or its dual is proper.

Thus, throughout the paper, we do not need to check if a given weighted game is proper because if such condition fails, then its dual game is also weighted, proper and preserves the desirability relation. We refer to (Freixas and Zwicker, 2003) and (Kurz et al., 2020) for extensions of weighted games on several alternatives and to (Kurz et al., 2017) on fair allocation of weights to single delegates of differently sized groups.

Definition 8 Weighted game with a unique minimal representation in integers
A weighted game (N, W) has a unique minimal representation in integers $[q; w]$ if $w \leq w'$ and $q \leq q'$ for any other weighted representation in integers $[q'; w']$ of (N, W).

There exist weighted games without a unique minimal representation in integers, (Dubey and Shapley, 1979; Muroga et al., 1970; Freixas and Molinero, 2009, 2010; Kurz, 2012). Note that if a weighted game has a unique minimal representation in integers $[q; w]$, then $[w(N) - q + 1; w]$ is the unique minimal representation in integers of the dual game. Thus, duality preserves minimal representations.

The next definition introduces a new subclass of weighted games which is the central topic of the paper.

Definition 9 Basic weighted game
A simple game (N, W) is a *basic weighted game* if it is weighted, proper and admits a unique minimal representation in integers which assigns a weight of 1 to the weakest non-null players.

Note that the weight of equivalent players is the same in a unique minimal representation in integers. A weighted game is *homogeneous* if it admits a representation in which all minimal winning coalitions have the same weight. Homogeneous games admit a unique minimal representation in integers in which the smallest non-null players have weight 1, see (Ostmann, 1987). Thus, the conjunction of proper and homogeneous games forms a subclass of basic weighted games, but the converse is not true as shown in the next example.

Example 1 The proper weighted game for $n = 5$ with representation $[9; 5, 4, 3, 2, 1]$ is a basic weighted game which is not homogeneous. In fact, the coalitions $\{1, 3, 4\}$ and $\{1, 3, 5\}$ are both minimal but $4 \succ 5$ implies $w_4 > w_5$ for any weighted representation of the game. Thus, the weight of $\{1, 3, 4\}$ surpasses the quota for any weighted representation.

From now on, we only deal with complete games, and without loss of generality, we assume that $1 \succsim 2 \succsim \cdots \succsim n$, where $\succsim$ is the desirability relation given in Definition 5.

The desirability relation generates a partition of N into equivalence classes (two elements i and j of N are equivalent if $i \succsim j$ and $j \succsim i$). Then, these equivalence classes, denoted by $N_1, N_2, \ldots, N_t$, can be strictly ordered in the following way: $N_p \succ N_q$ if and only if $i \succ j$ for any $i \in N_p$ and any $j \in N_q$. Let $n_j = |N_j|$ for $1 \leq j \leq t$.

In this context, coalitions can be categorized into different models:

Definition 10 Models of coalitions
A coalition is of *model* $\overline{m} = (m_1, \ldots, m_t)$ if it has a total of $m_1 + \cdots + m_t$ players, with m_j players being in the class N_j $(m_j \leq n_j)$. All the coalitions of a same model have the same status in the game, in the sense that they are either all winning or all losing. If the game is weighted, then it admits a representation such that all coalitions of the same model have the same weight.

Given two models $\overline{m}$ and $\overline{p}$, we will write $\overline{m} \geq \overline{p}$ when $m_j \geq p_j$ for $j = 1, \ldots, t$, and $\overline{m} > \overline{p}$ if $\overline{m} \geq \overline{p}$ and $m_j > p_j$ for some j. The inequality $\overline{m} > \overline{p}$ means that $\overline{m}$ can be obtained from $\overline{p}$ by adding representatives of one ore more equi-desirability classes.

Example 2 On the player set $N = \{1, 2, 3, 4, 5, 6\}$, let us say that a coalition S is winning if it contains at least four players and at least two of the three first players. N is partitioned into two equivalence classes $N_1 = \{1, 2, 3\}$ and $N_2 = \{4, 5, 6\}$ such that $N_1 \succ N_2$. The models of coalitions are (m_1, m_2) with $0 \leq m_i \leq 3$ for $i = 1, 2$. The models $(2, 2)$ and $(3, 1)$ correspond to 9 and 3 minimal winning coalitions, respectively. The remaining models of winning coalitions are $(3, 2)$ and $(3, 3)$. The game is weighted but not basic. Indeed, it admits a unique minimal integer representation $[10; 3, 3, 3, 2, 2, 2]$ with a weight greater than 1 assigned to the players in N_2.

Definition 11 Hierarchy of a complete game
We say that a complete game (N, W) has the *hierarchy* $(n_1, n_2, \ldots, n_t)$ if the coalition N is of model $(n_1, n_2, \ldots, n_t)$.

Definition 12 Hierarchy achievable in a subfamily of complete games
We say that a hierarchy $(n_1, n_2, \ldots, n_t)$ is *achievable* in some subfamily of complete games (weighted, or basic weighted game in our study) if there exits a simple game in this family having this hierarchy by the desirability relation.

3 Main Result

The main theorem in this paper, Theorem 1, proves that all achievable hierarchies for weighted games are also achievable for basic weighted games, a tiny subclass of weighted games with valuable properties. For instance, the set of representations of a basic weighted game forms a discrete $n + 1$-dimensional cone with a minimum vertex (the vector of the quota and the n weights of the unique minimal representation in integers).

Proposition 1 *For every $n \geq 5$, the hierarchy $(n - 4, 1, 1, 1, 1)$ is achievable in basic weighted games.*

Proof The weighted game $[9 + 5(n - 5); 5, 4, 3, 2, 1]$, which is well-defined for any $n \geq 5$, has the hierarchy $(n - 4, 1, 1, 1, 1)$ and verifies all properties demanded to basic weighted games: It is proper since $q = 5n - 16 > (5n - 10)/2 = w(N)/2$, it assigns a weight of 1 to the weakest (non-null) player, and the representation in integers is minimal because the weights used in the representation are the five smallest positive integers, and the quota $5n - 16$ cannot be diminished without changing the game.

Proposition 2 *The following types of hierarchies are achievable in basic weighted games:*

1. (n),
2. (n_1, n_2),
3. (n_1, n_2, n_3) *for* $n_2 + n_3 > 2$,
4. (n_1, n_2, n_3, n_4) *for* $n_2 + n_3 + n_4 > 3$,
5. $(n_1, n_2, n_3, n_4, n_5)$ *for* $n_2 + n_3 + n_4 + n_5 > 4$.

Proof The proof of this proposition consists in providing, for each hierarchy, a basic weighted game having it. The games chosen for every hierarchy are shown below. The election of weights and the quota guarantees all the properties demanded to basic weighted games.

1. The weighted game $[n;\ \overbrace{1, 1, \ldots, 1}^{n}]$ is basic weighted and has hierarchy (n).
2. The weighted game $[n_1;\ \overbrace{1, 1, \ldots, 1}^{n_1}, \overbrace{0, 0, \ldots, 0}^{n_2}]$ is basic weighted and has hierarchy (n_1, n_2).
3. If $n_2 > 1$, the weighted game

$$[2n_1 + n_2 - 1;\ \overbrace{2, 2, \ldots, 2}^{n_1}, \overbrace{1, 1, \ldots, 1}^{n_2}, \overbrace{0, 0, \ldots, 0}^{n_3}]$$

is basic weighted and has hierarchy (n_1, n_2, n_3).

If $n_3 > 1$ the weighted game

$$[3n_1 + 2n_2 + n_3 - 2;\ \overbrace{3, 3, \ldots, 3}^{n_1}, \overbrace{2, 2, \ldots, 2}^{n_2}, \overbrace{1, 1, \ldots, 1}^{n_3}]$$

is basic weighted and has hierarchy (n_1, n_2, n_3).

4. If $n_2 > 1$, the weighted game

$$[2(n_3n_4 + n_3 + n_4) + 1;\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{1, \ldots, 1}^{n_4}]$$

with $x_1 = 2n_3n_4 + 2n_3 + n_4$, $x_2 = n_3n_4 + n_3 + n_4$ and $x_3 = n_4 + 1$ is basic weighted and has hierarchy (n_1, n_2, n_3, n_4).

If $n_3 > 1$, the weighted game

$$[3n_1 + 2n_2 + n_3 - 2;\ \overbrace{3, 3, \ldots, 3}^{n_1}, \overbrace{2, 2, \ldots, 2}^{n_2} \overbrace{1, 1, \ldots, 1}^{n_3}, \overbrace{0, 0, \ldots, 0}^{n_4}]$$

is basic weighted and has hierarchy (n_1, n_2, n_3, n_4).
If $n_4 > 1$, the weighted game

$$[n_1(n_2n_4 + n_3n_4 + n_2);\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{1, \ldots, 1}^{n_4}]$$

with $x_1 = n_2n_4 + n_3n_4 + n_2, x_2 = n_4 + 1$ and $x_3 = n_4$ is basic weighted and has hierarchy (n_1, n_2, n_3, n_4).

5. If $n_2 > 1$, the weighted game

$$[2(n_3n_4 + n_3 + n_4) + 1;\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{1, \ldots, 1}^{n_4}, \overbrace{0, \ldots, 0}^{n_5}]$$

with $x_1 = 2n_3n_4 + 2n_3 + n_4$, $x_2 = n_3n_4 + n_3 + n_4$ and $x_3 = n_4 + 1$ is basic weighted and has hierarchy $(n_1, n_2, n_3, n_4, n_5)$.
If $n_3 > 1$, the weighted game

$$[n_1x_1 + x_2 + x_4;\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{x_4, \ldots, x_4}^{n_4}, \overbrace{1, \ldots, 1}^{n_5}]$$

with $x_1 = 2n_2n_4n_5 + n_3n_4n_5 + 2n_2n_4 + n_2n_5 + n_3n_4 + n_3n_5 - n_4n_5 - n_4 - n_5 - 1$, $x_2 = 2n_4n_5 + 2n_4 + n_5$, $x_3 = n_4n_5 + n_4 + n_5$ and $x_4 = n_5 + 1$ is basic weighted and has hierarchy $(n_1, n_2, n_3, n_4, n_5)$.
If $n_4 > 1$, the weighted game

$$[n_1(n_2n_4 + n_3n_4 + n_2);\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{1, \ldots, 1}^{n_4}, \overbrace{0, \ldots, 0}^{n_5}]$$

with $x_1 = n_2n_4 + n_3n_4 + n_2, x_2 = n_4 + 1$ and $x_3 = n_4$ is basic weighted and has hierarchy $(n_1, n_2, n_3, n_4, n_5)$.
If $n_5 > 1$, the weighted game

$$[n_1x_1 + n_2x_2;\ \overbrace{x_1, \ldots, x_1}^{n_1}, \overbrace{x_2, \ldots, x_2}^{n_2}, \overbrace{x_3, \ldots, x_3}^{n_3}, \overbrace{x_4, \ldots, x_4}^{n_4}, \overbrace{1, \ldots, 1}^{n_5}]$$

with $x_1 = n_3n_5 + n_4n_5 + n_3n_5$, $x_2 = n_3n_5 + n_4n_5 + n_3$, $x_3 = n_5 + 1$ and $x_4 = n_5$ is basic weighted and has hierarchy $(n_1, n_2, n_3, n_4, n_5)$.

□

Proposition 3 *Any hierarchy* $(n_1, n_2, \ldots, n_t)$ *with* $t \geq 6$ *is achievable in basic weighted games.*

Proof It suffices to prove the existence of a proper homogeneous game with the given hierarchy $(n_1, n_2, \ldots, n_t)$ for any $t \geq 6$. We start by defining a representation of a weighted game with $n = n_1 + n_2 + \cdots + n_t$ players, and then, we will prove that the game has the given hierarchy and is homogeneous.

The weights x_j ($x_1 > x_2 > \cdots > x_t$), defined recursively from $j = t$ till $j = 1$, and the quota q, are the following:

$$\begin{aligned}
&x_t &&= 1\\
&x_{t-1} &&= n_t + 1\\
\text{For } j \text{ from } t-2 \text{ till } 4:\ &x_j &&= n_{j+1}x_{j+1} + x_{j+2}\\
&x_3 &&= \begin{cases} n_4x_4 + x_5 & \text{if } n_2 > 1\\ 1 + (n_4 - 1)x_4 + \sum_{j=5}^{t} n_j x_j & \text{if } n_2 = 1\end{cases}\\
&x_2 &&= n_3x_3 + x_4\\
&x_1 &&= \begin{cases} (n_2 - 1)x_2 + x_3 & \text{if } n_2 > 1\\ n_3x_3 + n_4x_4 + x_5 & \text{if } n_2 = 1\end{cases}\\
&q &&= n_1x_1 + x_2
\end{aligned} \tag{1}$$

Clearly $q > w(N)/2$ which guarantees properness. Using the notation introduced in Definition 7, let $M_j = \{i \in N \mid w_i = x_j\}$ for $j = 1, \ldots, t$. From the properties of weighted games, it is clear that all players in a fixed M_j belong to the same equivalence class. Hence, either M_j and M_{j+1} belong to the same equivalence class, which occurs when $i \approx k$ for all $i \in M_j$ and all $k \in M_{j+1}$, or $M_j \succ M_{j+1}$, which occurs when $i \succ k$ for all $i \in M_j$ and all $k \in M_{j+1}$. We will prove that $M_j \succ M_{j+1}$ for any $j = 1, \ldots, t-1$, so that $M_j = N_j$ for any $j = 1, 2, \ldots, t$, and thus, the defined game has the desired hierarchy.

The procedure for defining the weights (1) is different depending on $n_2 > 1$ or $n_2 = 1$. We will prove the statement assuming $n_2 > 1$, because the proof for the other case is similar. By the recursive equations, we observe that the defined weighted game has $t-1$ models of coalitions (see Definition 10) with weight equal to q, namely $\{\overline{a}_j\}_{j=2,\ldots,t}$, whose components are described as follows for $k = 1, \ldots, t$:

$$\overline{a}_{j,k} = \begin{cases} 1, & \text{if } j = k\\ n_1, & \text{if } k = 1,\ j \text{ even}\\ n_1 - 1, & \text{if } k = 1,\ j \text{ odd}\\ n_k, & \text{if } 1 < k < j,\ \ j \text{ and } k \text{ have different parity}\\ 0, & \text{otherwise.}\end{cases} \tag{2}$$

By the term "different parity" for subindices j and k in the previous expression, we mean that either j is odd and k even or j is even and k odd.

As for all $j = 2, \ldots, t-1$ it is $\overline{a}_{j,j} = 1$ and $\overline{a}_{j,j+1} = 0$, then any coalition S with model $\overline{a}_j$ verifies that there exists $i \in M_j \cap S$ and $h \notin S$ for any $h \in M_{j+1}$. But $w(S) = q$ so that $w((S \setminus \{i\}) \cup \{h\}) < q$. Hence, $i \succ h$, and $M_j \succ M_{j+1}$.
Finally, as $\overline{a}_{4,1} = n_1$ and $\overline{a}_{4,2} = 0$, then any coalition S with model $\overline{a}_4$ verifies that there exists $i \in M_1 \cap S$ and $h \notin S$ for any $h \in M_2$. But $w(S) = q$ so that $w((S \setminus \{i\}) \cup \{h\}) < q$. Hence, $i \succ h$, and $M_1 \succ M_2$.

Thus, we have proved that $M_j = N_j$ for all $j = 1, \ldots, t$, and therefore, the weighted game recursively defined has the given hierarchy $(n_1, \ldots, n_t)$.

We will prove now that the game is homogeneous. To this end, we study all possible models of coalitions in this game. We consider four groups: (i) models of coalitions with weight equal to the quota, already described in (2); (ii) models of coalitions which can be obtained from the models in (i) by swapping players for stronger ones (here, “weaker" and “stronger" refer to the desirability relation); (iii) models of coalitions which can be obtained from the models in (i) and (ii) by adding new players; and (iv) the rest of models of coalitions. It is clear that coalitions in (i), (ii), and (iii) are winning, with coalitions in (i) being minimal winning and coalitions in (iii) being non-minimal winning. In the following steps, we prove that all models in (ii) correspond to non-minimal winning coalitions and that all models in (iv) correspond to losing coalitions.

Let $\overline{u}$ be a model obtained from some $\overline{a}_j$ by swapping one or more players for stronger ones. Notice that it is not possible to do so in $\overline{a}_2 = (n_1, 1, 0, 0, \ldots, 0)$, so that $3 \leq j \leq t$. Now, if j is even (in this case $\overline{a}_{j,1} = n_1$), then there exists at least one k even, $k < j$, such that $\overline{u}_k \geq 1$. This implies that $\overline{u} > \overline{a}_k$, being k the minimum even index such that $\overline{u}_k \geq 1$. Similarly, if j is odd (in this case $\overline{a}_{j,1} = n_1 - 1$), then either $\overline{u}_1 = n_1$, in which case $\overline{u} > \overline{a}_2$, or $\overline{u}_1 = n_1 - 1$. In this last case, there exists at least one k odd, $k < j$, such that $\overline{u}_k \geq 1$. This implies that $\overline{u} > \overline{a}_k$, being k the minimum odd index such that $\overline{u}_k \geq 1$. Thus, $\overline{u}$ is winning but not minimal.

To prove that the remaining models of coalitions are all losing, we select those models such that any other model, not included in the already considered groups, can be obtained from them by either removing players or swapping players for weaker ones. We will see that all these models are losing. There are r models in this selection, with $r = t$ if $n_t > 1$ and $r = t - 1$ if $n_t = 1$. We denote these models by $\{\overline{c}_j\}_{j=0,1,\ldots,r-1}$, and their components are described as follows for $k = 1, \ldots, t$:

$$\overline{c}_{j,k} = \begin{cases} n_1, & \text{if } k = 1,\ j \text{ odd} \\ n_1 - 1, & \text{if } k = 1,\ j \text{ even} \\ n_k - 1, & \text{if } j = k > 1 \\ n_k, & \text{if either } \begin{cases} 1 < j < k \\ 1 < k < j, & j \text{ and } k \text{ have the same parity} \\ j \leq 1 < k, & j \text{ and } k \text{ have the same parity} \end{cases} \\ 0, & \text{otherwise.} \end{cases}$$

Let us check that all these models have a weight smaller than the quota q assuming that t is even, because the proof for t odd is analogous. Notice that the model $\overline{c}_1$ is obtained from $\overline{a}_t$ by removing the voter with weight $x_t = 1$. Thus, $w(\overline{c}_1) = q - 1 < q$. The weight of $\overline{c}_0$ is also $q - 1$ because $w(\overline{c}_0) = (n_1 - 1)x_1 + \sum_{\substack{1<k\leq t \\ k\ even}} n_k x_k$ and $w(\overline{a}_{t-1}) = (n_1 - 1)x_1 + \sum_{\substack{1<k<t \\ k\ even}} n_k x_k + x_{t-1} = q$.

For j even, $2 \leq j \leq r - 1$, it is

$$w(\overline{c}_j) = (n_1 - 1)x_1 + \sum_{\substack{1<k<j \\ k\ even}} n_k x_k + (n_j - 1)x_j + \sum_{k=j+1}^{t} n_k x_k.$$

But $w(\overline{a}_{j+1}) = q$, so that $q = (n_1 - 1)x_1 + \sum_{\substack{1<k\le j\\ k\ even}} n_k x_k + x_{j+1}$. Thus,

$$w(\overline{c}_j) = q - x_{j+1} - x_j + \sum_{k=j+1}^{t} n_k x_k$$

and using the recursive equations (1), we obtain $w(\overline{c}_j) = q - 2 < q$. An analogous reasoning proves that $w(\overline{c}_j) = q - 2$ for j odd, $2 < j \le r - 1$.

In conclusion, the defined weighted game has the hierarchy $(n_1, \ldots, n_t)$, and only the minimal winning coalitions have a weight equal to the quota. Hence, the game is homogeneous and basic weighted. □

We will show some examples of games constructed with the procedure described above.

Example 3 Consider the hierarchy (2, 3, 1, 1, 4, 5), with $t = 6$. In this case, the basic weighted game with the given hierarchy is

$$[342;\ 143, 143, 56, 56, 56, 31, 25, 6, 6, 6, 6, 1, 1, 1, 1, 1].$$

Example 4 Consider the hierarchy (3, 1, 1, 2, 4, 5, 1, 1, 3), with $t = 9$. In this case, the basic weighted game with the given hierarchy is

$$[3481;\ 935, 935, 935, 676, 467, 209, 209, 50, 50, 50, 50, 9, 9, 9, 9, 9, 5, 4, 1, 1, 1].$$

Example 5 Particularly, interesting examples are obtained when each equivalence class is a singleton, i.e., when the given hierarchy is of the type $(\overbrace{1, 1, 1, \ldots, 1, 1, 1}^{t})$ with $t \ge 6$. In these cases, the weights from x_t to x_4 follow a Fibonacci sequence, and the three remaining terms are $x_3 = x_4 + x_5 - 1$, $x_2 = x_3 + x_4$ and $x_1 = x_3 + x_4 + x_5 = 2x_3 + 1$. Let us show the basic weighted games with these hierarchies, from $t = 6$ till $t = 14$:

$t = 6:$ $[16;\ 9, 7, 4, 3, 2, 1]$
$t = 7:$ $[27;\ 15, 12, 7, 5, 3, 2, 1]$
$t = 8:$ $[45;\ 25, 20, 12, 8, 5, 3, 2, 1]$
$t = 9:$ $[74;\ 41, 33, 20, 13, 8, 5, 3, 2, 1]$
$t = 10:$ $[121;\ 67, 54, 33, 21, 13, 8, 5, 3, 2, 1]$
$t = 11:$ $[121;\ 109, 88, 54, 34, 21, 13, 8, 5, 3, 2, 1]$
$t = 12:$ $[320;\ 177, 143, 88, 55, 34, 21, 13, 8, 5, 3, 2, 1]$
$t = 13:$ $[519;\ 287, 232, 143, 89, 55, 34, 21, 13, 8, 5, 3, 2, 1]$
$t = 14:$ $[841;\ 465, 376, 232, 144, 89, 55, 34, 21, 13, 8, 5, 3, 2, 1]$

Proper weighted games with weights that follow Fibonacci sequences are studied in (Fragnelli et al., 2016; Pressacco and Ziani, 2015, 2018).

Theorem 1 *All achievable hierarchy for weighted games is achievable in the subclass of basic weighted games.*

Proof In (Bean et al., 2008), it is proved that any hierarchy of n players is achievable in weighted games with the only exception of two of them: $(n-2, 1, 1)$ for $n > 2$, and $(n-3, 1, 1, 1)$ for $n > 3$. As these two hierarchies are not achievable in weighted games, they are not achievable in basic weighted games either. Proposition 1 states that the hierarchy $(n-4, 1, 1, 1, 1)$ is achievable in basic weighted games for $n > 4$. Proposition 2 states that any other hierarchy $(n_1, n_2, \ldots, n_t)$ with $t \leq 5$ is achievable in basic weighted games, and Proposition 3 states that any hierarchy with $t \geq 6$ is achievable in basic weighted games. This finishes the proof. □

Note that the procedure described in equation (1) in the proof of Proposition 3 describes how to obtain a representation of any basic weighted game for any hierarchy with $t \geq 6$. Propositions 1 and 2 describe the method for $t \leq 5$.

We complete this section with the desired consequence of Theorem 1.

Corollary 1 *Any possible hierarchy is achievable, by either the Shapley–Shubik, the Banzhaf or the Johnston indices, in basic weighted games except* $(n-2, 1, 1)$*, and* $(n-3, 1, 1, 1)$.

4 Conclusion

The main result of the present paper is to prove that, if a particular power structure has to be designed, the class of basic weighted games can always be used for this purpose. These games have very reasonable properties that make them ideal candidates to be chosen. Moreover, the subclass is relatively small when compared with all weighted games.

The number of possible hierarchies in a set of n players is 2^{n-1}. In (Friedman et al., 2006), it was proved that all hierarchies achievable by the desirability relation in complete games are also achievable in weighted games, and that the hierarchies which are not achievable in weighted games, for n players, are $(n-2, 1, 1)$ and $(n-3, 1, 1, 1)$. Thus, the number of hierarchies achievable in weighted games, for n players, is

$$\text{Number of achievable hierarchies in weighted games} = \begin{cases} 2^{n-1} - 2, & \text{if } n \geq 4 \\ 3, & \text{if } n = 3 \\ 2, & \text{if } n = 2 \end{cases}$$

which coincides with the number of achievable hierarchies in basic weighted games.

In (Diffo Lambo and Moulen, 2002), it was proved that the Shapley–Shubik and the Banzhaf power indices generate the same hierarchy than the desirability relation in weighted games, and in (Freixas et al., 2012), it was proved that this is also true in a larger class of games and that it is also verified for the Johnston power index. Thus, the number of hierarchies achievable in weighted games by the desirability relation, specified above, coincides with the number of hierarchies achievable by anyone of these three power indices.

The study of hierarchies can be extended to voting games with abstention as a third input. Parker proves in (Parker, 2012) that the influence relation introduced by Tchantcho et al. (the I-influence relation) (Tchantcho et al., 2008) orders the voters in the same way as the classical Banzhaf and Shapley–Shubik indices do when they are extended to complete voting games with abstention. Moreover, (Parker, 2012) proves that all hierarchies are achievable, by the influence relation, in this class. In a subsequent work, (Freixas et al., 2014) prove that all hierarchies are achievable, by the influence relation, in a subclass of weighted voting games with abstention, the class of zero-sum strongly weighted games. It would be now interesting to study whether the class of basic weighted games with abstention would still be enough to generate all hierarchies achievable by the I-influence relation.

Acknowledgements This publication is part of the I+D+i project / PID2019-104987GB-I00, financed by MCIN/ AEI/10.13039/501100011033/

References

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*, 317–343.

Bean, D. (2012). Proportional quota weighted voting system hierarchies II. *Social Choice and Welfare, 39*, 907–918.

Bean, D., Friedman, J., & Parker, C. (2008). Simple majority achievable hierarchies. *Theory and Decision, 65*, 285–302.

Bean, D., Friedman, J., & Parker, C. (2010). Proportional quota weighted voting system hierarchies. *Social Choice and Welfare, 34*, 397–410.

Bishnu, M., & Roy, S. (2012). Hierarchy of players in swap robust voting games. *Social Choice and Welfare, 38*, 11–22.

Carreras, F., & Freixas, J. (1996). Complete simple games. *Mathematical Social Sciences, 32*, 139–155.

Coleman, J. S. (1971). Control of collectivities and the power of a collectivity to act. In B. Lieberman (Ed.), *Social choice* (pp. 269–300). Gordon and Breach.

Diffo Lambo, L., & Moulen, J. (2002). Ordinal equivalence of power notions in voting games. *Theory and Decision, 53*, 313–325.

Dubey, P., & Shapley, L. S. (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research, 4*, 99–131.

Fragnelli, V., Gambarelli, G., Gnocchi, N., Pressacco, F., & Ziani, L. (2016). Fibonacci representations of homogeneous weighted majority games. In N. Nguyen, R. Kowalczyk, & J. Mercik (Eds.), Transactions in computational collective intelligence XXIII, Lecture Notes in Computer Science (Vol. 9760, pp. 162–171). Springer

Freixas, J., & Kurz, S. (2014). On minimum integer representations of weighted games. *Mathematical Social Sciences, 67*, 9–22.

Freixas, J., Marciniak, D., & Pons, M. (2012). On the ordinal equivalence of the Johnston, Banzhaf and Shapley power indices. *European Journal of Operational Research, 216*, 367–375.

Freixas, J., & Molinero, X. (2009). On the existence of a minimum integer representation for weighted voting systems. *Annals of Operations Research, 166*, 243–260.

Freixas, J., & Molinero, X. (2010). Weighted games without a unique minimal representation in integers. *Optimization Methods and Software, 25*, 203–215.

Freixas, J., & Pons, M. (2010). Hierarchies achievable in simple games. *Theory and Decision, 68*, 393–404.

Freixas, J., Tchantcho, B., & Tedjeugang, N. (2014). Achievable hierarchies in voting games with abstention. *European Journal of Operational Research, 36*, 254–260.

Freixas, J., & Zwicker, W. S. (2003). Weighted voting, abstention, and multiple levels of approval. *Social Choice and Welfare, 21*, 399–431.

Friedman, J. (2016). A note on hierarchies in weighted voting games and partitions. *The Ramanujan Journal, 41*, 305–310.

Friedman, J., McGrath, L., & Parker, C. (2006). Achievable hierarchies in voting games. *Theory and Decision, 61*, 305–318.

Friedman, J., & Parker, C. (2009). Can an asymmetric power structure always be achieved? In T. Y. Chow, J. A. Gallian, & D. C. Isaksen (Eds.), *Communicating mathematics* (pp. 87–98). American Mathematical Society.

Isbell, J. R. (1956). A class of majority games. *Quarterly Journal of Mathematics, 7*, 183–187.

Isbell, J. R. (1958). A class of simple games. *Duke Mathematics Journal, 25*, 423–439.

Johnston, R. J. (1978). On the measurement of power: Some reactions to Laver. *Environment and Planning A, 10*, 907–914.

Kurz, S. (2012). On minimum sum representation for weighted voting games. *Annals of Operations Research, 196*, 361–369.

Kurz, S., Maaser, N., & Napel, S. (2017). On the democratic weights of nations. *Journal of Political Economy, 125*, 1599–1634.

Kurz, S., Mayer, A., & Napel, S. (2020). Weighted committee games. *European Journal of Operational Research, 282*, 972–979.

Kurz, S., & Tautenhahn, N. (2013). On Dedekind's problem for complete simple games. *International Journal of Game Theory, 42*, 411–437.

Muroga, S., Tsuboi, T., & Baugh, R. (1970). Enumeration of threshold functions of eight variables. *IEEE Transactions on Computers*, *C-19*, 818–825

Ostmann, A. (1987). On the minimal representation of homogeneous games. *International Journal of Game Theory, 16*, 69–81.

Parker, C. (2012). The influence relation for ternary voting games. *Games and Economic Behavior, 75*, 867–881.

Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109*, 53–57.

Pressacco, F., & Ziani, L. (2015). A Fibonacci approach to weighted majority games. *Journal of Game Theory, 4*, 36–44.

Pressacco, F., & Ziani, L. (2018). Proper strong-Fibonacci games. *Decisions in Economics and Finance, 41*, 489–529.

Shapley, L. S. (1953). A value for n-person games. In A. W. Tucker & H. W. Kuhn (Eds.), *Contributions to the theory of games II* (pp. 307–317). Princeton University Press.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*, 787–792.

Taylor, A. D., & Zwicker, W. S. (1999). *Simple games: Desirability relations, trading, and pseudoweightings*. Princeton University Press.

Tchantcho, B., Diffo Lambo, L., Pongou, R., & Mbama Engoulou, B. (2008). Voters' power in voting games with abstention: Influence relation and ordinal equivalence of power theories. *Games and Economic Behavior, 64*, 335–350.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Bargaining in Legislatures: A New Donation Paradox

Maria Montero

Abstract It is well known that being the proposer or agenda setter is advantageous in many collective decision-making situations. In the canonical model of distributive bargaining (Baron and Ferejohn, 1989), proposers are certain of being part of the coalition that forms, and, conditional on being in the coalition, a player receives more as a proposer than as a coalition partner. In this paper, I show that it is possible for a party to donate part of its proposing probability to another party and be better off as a result. This appears paradoxical, even more so since the recipient never includes the donor in its proposals. The example shows that, even though actually being selected to propose is always valuable ex post, having a higher *probability* of being proposer may be harmful.

1 Introduction

A donation paradox occurs when a player transfers an apparently valuable prerogative to another player but is better off as a result (Kadane et al., 1999). The donation paradox in power indices was identified by (Felsenthal and Machover, 1995). A power index exhibits the donation paradox when it is possible for a player to increase its power (as measured by the index) by donating part of its weight to another player. If the issue at stake is the division of a fixed pie between the players, power can be measured by the players' expected shares of this pie. (Felsenthal and Machover, 1995, p. 258–259) refer to this definition of power as P-power and argue that no reasonable measure of P-power would display the donation paradox.[1,2]

[1]If the issue at stake is binary (yes/no), (Laruelle and Valenciano, 2005) show that neither success (probability of a voter getting the final outcome they want) nor decisiveness (probability of a voter being pivotal in achieving this outcome) exhibit the donation paradox.

[2]Not all researchers have such a strict stance o0n power measures that fail this or similar postulates (such as dominance); see (Holler and Napel, 2005; Kurz et al., 2015).

M. Montero (✉)
School of Economics, University of Nottingham, Nottingham, UK
e-mail: maria.montero@nottingham.ac.uk

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_10

In this paper, I identify a donation-type paradox that arises as an equilibrium phenomenon in legislative bargaining over a fixed pie. The most influential model of bargaining in legislatures is due to (Baron and Ferejohn, 1989). In this model, n players must divide a budget by majority rule. The players have opposed preferences in the sense that each player would like to have as high a share of the budget as possible. One of the players is randomly selected to make a proposal, and the remaining players accept or reject. If the proposal is rejected, a new proposer is selected, always using the same probability distribution. Being the proposer is valuable in this model: The proposer is guaranteed a positive share and, conditional on getting a positive share, a player gets more as a proposer than as a responder (Harrington, 1990).[3] Nevertheless, this paper shows that it may be possible for a player to donate some of its *proposing probability* to a recipient and be better off. This can happen even though the recipient never includes the donor in its proposals after the donation. The effect is triggered by the fact that the donor is disproportionately likely to receive proposals by third parties after the donation, and this effect predominates over the loss of proposing probability.

2 The Model

2.1 Simple Games

Let $N = \{1, ..., n\}$ be the set of players. A *simple game* is a collection W of subsets of N such that $N \in W$, $\varnothing \notin W$, and the following *monotonicity* condition is satisfied: $S \in W$ implies $T \in W$ for all S, T such that $S \subseteq T \subseteq N$. Elements of W are called *winning coalitions*. A coalition S is *minimal winning* if $S \in W$ and $T \notin W$ for all T such that $T \subsetneq S$. A player who belongs to all winning coalitions is called a *veto player*.

A simple game W is a *weighted majority game* if there exist n nonnegative numbers (weights) $w_1, ..., w_n$ and a positive number q such that $S \in W$ if and only if $\sum_{i \in S} w_i \geq q$. A weighted majority game admits a *homogeneous representation* if there exist nonnegative numbers $w_1^h, ..., w_n^h$ and a positive number q^h such that S is minimal winning if and only if $\sum_{i \in S} w_i^h = q^h$.

[3] Proposer power or agenda-setting power arises in many settings. (Romer and Rosenthal, 1978) show that being able to set the agenda is advantageous in a collective decision setting with a unidimensional policy and single-peaked preferences. Proposer power is also found in various dynamic models with an evolving status quo, see (Kalandrakis, 2010; Diermeier and Fong, 2011; Duggan and Ma, 2018).

2.2 The Bargaining Procedure

Let W be a simple game. Players bargain over the division of a pie of size 1 as follows. At every round $t = 1, 2, ...,$ Nature selects a player randomly to be the proposer according to some probability distribution $\theta = (\theta_i)_{i \in N}$, where $\theta_i \geq 0$ for all i and $\sum_{i \in N} \theta_i = 1$. We will refer to θ_i as player i's recognition probability. The selected player proposes a payoff vector $(x_i)_{i \in N}$. This payoff vector must be feasible ($\sum_{i \in N} x_i \leq 1$), and no player can receive a negative payoff ($x_i \geq 0$ for all $i \in N$). Players vote for or against the proposal sequentially (the order does not affect the results).[4] If the set of players that voted in favor is a winning coalition, the proposal is implemented and the game ends. Otherwise, the game proceeds to the next period in which Nature selects a new proposer (always with the same probability distribution). Players are risk neutral and share a discount factor $\delta \in [0, 1]$.

We will denote the noncooperative bargaining game by $G(W, \theta, \delta)$. A pure strategy for player i is a sequence $\sigma_i = (\sigma_i^t)_{t=1}^{\infty}$, where σ_i^t, the t-th round strategy of player i, prescribes

1. A *proposal* $(x_i)_{i \in N}$.
2. A *response function* assigning "yes" or "no" to all possible proposals of the other players.

Players are free to condition their actions on the history of the game up to time t; however, the analysis is restricted to equilibria in which they choose not to do so. The solution concept is *stationary subgame perfect equilibrium* (SSPE). Stationarity requires that players follow the same (possibly mixed) strategy at every round t: The probability that the proposer makes a given proposal is the same for all t regardless of history, and the response function depends only on the current proposal and not on what happened in the previous rounds.

Given an SSPE σ^*, we will denote the associated *expected payoff* for player i (computed at the beginning of the game, before Nature chooses the proposer) by $y_i(\sigma^*)$ -we will drop σ^* to simplify notation-. The expected payoff given that a proposal is rejected is called the *continuation value*. The continuation value in a stationary equilibrium σ^* is δy_i. Continuation values play a very important role in any SSPE. A responder must accept any offer with $x_i > \delta y_i$ and reject any offer with $x_i < \delta y_i$. Players may in principle accept or reject if $x_i = \delta y_i$, but there is very little loss of generality in assuming that players accept offers with $x_i = \delta y_i$ (see (Eraslan and McLennan, 2013, Appendix A)). Given that continuation values act as prices, proposers propose the cheapest winning coalition given those prices (see (Okada, 1996)).

[4] An alternative to sequential voting is to assume that voting is simultaneous but all voters vote as if they are pivotal; see (Baron and Kalai, 1993).

2.3 The Proposer Advantage

Following common practice, I will refer to the set of players that receive a positive payoff according to the proposal that is implemented as the coalition that forms. The proposer advantage is the difference between a player's payoff conditional on being the proposer and its payoff conditional on being a coalition partner (i.e., a member of the formed coalition other than the proposer). The proposer advantage was originally established by (Baron and Ferejohn, 1989) and (Harrington, 1990) for symmetric games. Because of symmetry, each player expects $\frac{1}{n}$ if the game goes to the next period. Since the proposer only needs to convince $q-1$ players to vote for the proposal, it can offer $\frac{1}{n}$ to $q-1$ players and pocket the remaining $1-\frac{q-1}{n}=\frac{1}{n}+\frac{n-q}{n}$. Thus, there is a proposer advantage as long as $q<n$. Introducing discounting leads to an even greater proposer advantage. (Okada, 1996) shows that there is a proposer advantage in general games assuming that each player is recognized with probability $\frac{1}{n}$ and $\delta<1$. This result can be easily generalized to any recognition probabilities (see (Montero, 2006, corollary 3.1)). The following lemma shows that it can also be generalized to $\delta=1$, provided that no player has veto power or a recognition probability equal to 1 (if a player has veto power, there is still a proposer advantage if $\delta<1$; if a player has a monopoly on making proposals, the advantage of being proposer is not defined).

Lemma 1 *(cf. (Okada, 1996, theorem 1)) Let W be a simple game with no veto players, and θ a vector of recognition probabilities such that $\theta_i<1$ for all i. The game $G(W,\theta,\delta)$ has a proposer advantage in the sense that a player earns strictly more as a proposer than as a coalition partner in any SSPE. The requirement of no veto players can be replaced by $\delta<1$.*

Proof Let y_i be the expected equilibrium payoff for player i. In a stationary equilibrium, a player has the same y_i in each period, and since there is 1 unit to divide, $\sum_{j\in N} y_j \leq 1$.

First, any player with $\theta_i>0$ must have $y_i>0$. This is because, as a proposer, player i can always exclude some k with $y_k>0$ and offer everybody else slightly more than their continuation value; the proposal will pass as k is not a veto player. Since $\sum_{j\in N\setminus\{i,k\}} y_j<1$, it follows that i has a positive payoff as a proposer.[5] Moreover, given that no player can be allocated a negative payoff as a responder, any player with a positive recognition probability must have a positive expected payoff overall.

Second, player i receives exactly δy_i as a coalition partner in an SSPE. As a proposer, it receives $1-\sum_{j\in S^*\setminus\{i\}}\delta y_j$, where

$$S^* \in \arg\min_{S\ni i, S\in W} \sum_{j\in S} y_j .$$

[5] If there was no k with $y_k>0$, it would be even easier for i to have a positive payoff as a proposer by offering all players slightly more than 0.

This is because i must accept any offers above its continuation value δy_i. If offers were above δy_i, the proposer could undercut the offer slightly, and it would still be accepted. Thus, i must receive exactly δy_i as a coalition partner, and as a proposer, it receives $1 - \sum_{j \in S \setminus \{i\}} \delta y_j$ for some S. If S was not the solution to the minimization problem, there would be another coalition that could be proposed with coalition partners getting slightly more than their continuation value; acceptance would be guaranteed, and the proposer would be better off.

The difference between i's payoff as a proposer and as a coalition partner is thus $1 - \delta \sum_{j \in S^*} y_j$. This is always positive except if $\delta = 1$ and $S^* = N$. We know that $S^* \neq N$, because since no player has a monopoly on making proposals, there is at least one player $k \neq i$ who can make proposals (and will therefore have $y_k > 0$), and since there are no veto players, the proposal can still pass without that player.

The requirement of there being no veto players can be replaced by $\delta < 1$. Then, there must be a proposer advantage because the sum of the continuation values of all players is strictly less than 1, so i can get a payoff above δy_i by proposing the grand coalition and offering every player j a payoff slightly above δy_j. □

3 A New Donation Paradox

The following proposition establishes the possibility that a player may gain from donating some of its proposer probability to another player.

Proposition 1 *Let W be a weighted majority game and $\theta = (\theta_1, \ldots, \theta_n)$ and $\theta' = \left(\theta_1', \ldots, \theta_n'\right)$ two vectors of recognition probabilities such that there are players i and j with $\theta_i > \theta_i'$ and $\theta_j < \theta_j'$ and $\theta_k = \theta_k'$ for all $k \neq i, j$. Let y (resp. y') be the equilibrium payoff vector in an SSPE of $G(W, \theta, \delta)$ (resp. $G(W, \theta', \delta)$), where $0 < \delta \leq 1$. It can (but need not) happen that $y_i' > y_i$.*

We will prove this proposition by means of an example for $\delta = 1$ (the case $0 < \delta < 1$ is briefly discussed at the end of this section). Suppose there are four players in the legislature, controlling 3, 2, 2, and 1 votes, respectively, and 5 votes are needed to pass a proposal. We consider two possible scenarios: Each player is recognized with a probability proportional to its number of votes ($\theta = \left(\frac{3}{8}, \frac{2}{8}, \frac{2}{8}, \frac{1}{8}\right)$), or alternatively, each player is recognized with equal probability (thus $\theta = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$). Both scenarios are plausible: In the first case, we can think of a player with three votes as a party comprised of three members, each of them with one vote, who always follow party discipline, and each member is selected with equal probability; in the second case, we can think of parties as being treated equally in terms of voice even though they have different numbers of votes.[6] Because the medium-sized players have the same recognition probability in both scenarios, we can view the move

[6] (Diermeier and Merlo, 2004) study proposer selection empirically in the context of government formation. They find that a proportional selection process modified in order to give a premium to the largest party does best in their data.

from one scenario to the other as the large player "donating" some of its recognition probability to the small player.

(Eraslan and McLennan, 2013) show that all SSPE have the same expected payoffs; therefore, if we are only interested in payoffs and not in strategies, it is enough to find one equilibrium.

Note that players of the same type must have the same payoff in equilibrium if they have the same recognition probability [(Montero, 2002), Lemma 2]. This result also follows from Eraslan and McLennan's uniqueness result (if equilibrium payoffs are unique they must be symmetric). Thus, we can set $y_2 = y_3$ and use y_2 to denote both player 2 and player 3's payoffs. We will focus on equilibrium strategies that are symmetric in the sense that the two players of the same type play the same strategy and are treated symmetrically by other players' strategies.

What coalitions do players propose in equilibrium? The answer is straightforward for the largest and the smallest player.

The large player always proposes $\{1, 2\}$ or $\{1, 3\}$ (each with probability 0.5 since we focus on symmetric strategies). The small player is of no use to the large player as a coalition partner: Adding the small player to a coalition that already contains the large player never turns a losing coalition into a winning one.[7]

Similarly, the large player is of little use to the small player as a coalition partner. The natural coalition for the small player to propose is the only minimal winning coalition to which it belongs, $\{2, 3, 4\}$. A coalition like $\{1, 2, 4\}$ could conceivably be proposed if $y_1 \leq y_2$, but this is never the case for the recognition probabilities considered.

As for a medium-sized player like player 2, it can propose coalition $\{1, 2\}$ or coalition $\{2, 3, 4\}$, depending on how y_1 compares with $y_2 + y_4$. If $y_1 = y_2 + y_4$, we have a competitive payoff in the sense that players that can replace each other in a minimal winning coalition receive the same payoff.

Lemma 2 *Consider the weighted majority game* $[5; 3, 2, 2, 1]$, *and let* $\theta = \left(\frac{3}{8}, \frac{2}{8}, \frac{2}{8}, \frac{1}{8}\right)$. *The equilibrium payoff vector in any SSPE with* $\delta = 1$ *is* $y = \left(\frac{5}{14}, \frac{4}{14}, \frac{4}{14}, \frac{1}{14}\right)$.

Proof Suppose the large player proposes to each medium-sized player with probability 0.5, the small player proposes to both medium-sized players, and each medium-sized player randomizes between proposing to the large player (with probability λ) and proposing coalition $\{2, 3, 4\}$ (with probability $1 - \lambda$). Suppose moreover that each coalition partner is offered its continuation value (so, for example, player 2 proposes $(y_1, 1 - y_1, 0, 0)$), and players accept any offer that gives them at least their continuation value. Note that a mixed strategy can only be optimal for a medium-sized player if $y_1 = y_2 + y_4$. The following system of equations determines the expected payoffs derived from these strategies and the equilibrium value of λ.

[7] Player 4 is what (Napel and Widgrén, 2001) call an inferior player. Player 4's vote only makes a difference in coalition $\{2, 3, 4\}$, while players 2 and 3 have alternative coalitions that do not involve player 4.

$$
\begin{aligned}
y_1 &= \frac{3}{8}(1 - y_2) + \frac{4}{8}\lambda y_1 \\
y_2 &= \frac{2}{8}(1 - y_1) + \frac{3}{8}\frac{1}{2}y_2 + \frac{2}{8}(1 - \lambda)\, y_2 + \frac{1}{8}y_2 \\
y_4 &= \frac{1}{8}(1 - 2y_2) + \frac{4}{8}(1 - \lambda)\, y_4 \\
y_1 &= y_2 + y_4
\end{aligned}
$$

The solution to this system is $y_1 = \frac{5}{14}$, $y_2 = \frac{4}{14}$, $y_4 = \frac{1}{14}$, and $\lambda = \frac{1}{2}$.

The strategies described above constitute an equilibrium because responders are offered their continuation values, and proposers are proposing to the cheapest possible coalition partners. □

Lemma 3 *Consider the weighted majority game* $[5; 3, 2, 2, 1]$*, and let* $\theta = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$*. The equilibrium payoff vector in any SSPE with* $\delta = 1$ *is* $y = \left(\frac{3}{8}, \frac{2}{8}, \frac{2}{8}, \frac{1}{8}\right)$.

Proof Suppose the large player proposes to each medium-sized player with probability 0.5, the medium-sized players propose to the large player, and the small player proposes to both medium-sized players. Suppose moreover that each coalition partner is offered its continuation value (so, for example, player 2 proposes $(y_1, 1 - y_1, 0, 0)$), and players accept any offer that gives them at least their continuation value. Then continuation values are found from the following system of equations

$$
\begin{aligned}
y_1 &= \frac{1}{4}(1 - y_2) + \frac{2}{4}y_1 \\
y_2 &= \frac{1}{4}(1 - y_1) + \frac{1}{4}\frac{1}{2}y_2 + \frac{1}{4}y_2 \\
y_4 &= \frac{1}{4}(1 - 2y_2)
\end{aligned}
$$

The solution to this system of equations is $y_1 = \frac{3}{8}$, $y_2 = \frac{2}{8}$ and $y_4 = \frac{1}{8}$.

The strategies described above constitute an equilibrium because responders are offered their continuation values, and proposers are proposing to the cheapest possible coalition partners. In particular, a medium-sized player would compare proposing to the large player (and paying $\frac{3}{8}$) with proposing to the other two players (and paying $\frac{2}{8} + \frac{1}{8} = \frac{3}{8}$). Because the alternative coalition is no better than the one that is being proposed, there is no profitable deviation. □

What is the effect of the donation from player 1 to player 4? The direct effect is negative: Player 4 always proposes coalition $\{2, 3, 4\}$, so if players did not change their strategies it would be the case that y_2 would go up (as the medium-sized players receive more proposals), y_4 would go up (as the small player is more likely to be recognized), and y_1 would go down (as the large player is less likely to be recognized). But then, it would no longer be optimal for players 2 and 3 to play a mixed strategy, as $y_1 < y_2 + y_4$. In the new equilibrium, the medium-sized players are more likely to propose to the large player than before. This indirect effect (the large player is more likely to receive proposals from the medium-sized players) brings the equilibrium back to a competitive situation in which $y_1 = y_2 + y_4$. Nevertheless, the individual values of y_1, y_2 and y_4 are not the same as before, and player 1 is better off in this new competitive equilibrium.

More generally, there is a range of probabilities such that player 1 can move from a competitive allocation to another competitive allocation that is more favorable by donating some probability to player 4. Fix the probability of being proposer for a medium-sized player at $\frac{1}{4}$, and let θ_1 be the probability that the large player is selected to be proposer; then, the small player is selected with probability $\frac{1}{2} - \theta_1$. For $\frac{1}{4} \leq \theta_1 \leq \frac{1}{2}$, the equilibrium is such that a medium-sized player is indifferent between proposing to the large player and proposing to the other medium-sized player and the small player, or equivalently $y_1 = y_2 + y_4$. Let λ be the probability that a medium-sized player proposes to the large player. Then, expected payoffs are found from the following equations

$$\begin{aligned}
y_1 &= \theta_1 (1 - y_2) + \frac{1}{2}\lambda y_1 \\
y_2 &= \frac{1}{4}(1 - y_1) + \frac{\theta_1}{2} y_2 + \frac{1}{4}(1 - \lambda) y_2 + \left(\frac{1}{2} - \theta_1\right) y_2 \\
y_4 &= \left(\frac{1}{2} - \theta_1\right)(1 - 2y_2) + \frac{1}{2}(1 - \lambda) y_4 \\
y_1 &= y_2 + y_4
\end{aligned}$$

The solution for λ is $2(1 - 2\theta_1)$. It starts at 1 for $\theta_1 = \frac{1}{4}$, and it approaches 0 when θ_1 approaches $\frac{1}{2}$. This is intuitive: If a player is less likely to be proposer with strategies being unchanged, it becomes cheaper and will receive more proposals. What is surprising is the overcompensation, so that the player is better off when it is less likely to be proposer. It turns out that $y_1 = \frac{2(1-\theta_1)}{5-4\theta_1}$, which is decreasing in θ_1.[8] Payoffs for the other two types are $y_4 = \frac{1-2\theta_1}{5-4\theta_1}$ (which is decreasing in θ_1 as one would intuitively expect; the direct effect of the donation is stronger than the indirect effect), and $y_2 = \frac{1}{5-4\theta_1}$ (which must be increasing in θ_1 since the other payoffs are decreasing in θ_1).

For $\delta < 1$, the range of probabilities where a mixed strategy equilibrium occurs becomes progressively smaller as δ gets closer to 0. Fixing the proposal probability for each medium-sized player at $\frac{1}{4}$, a mixed strategy equilibrium exists for $0.25 \leq \theta_1 \leq 0.50$ when $\delta = 1$ as discussed above. For $\delta = 0.5$, a mixed strategy equilibrium exists for $0.3229 \leq \theta_1 \leq 0.4182$ (values rounded to four decimal places). Similarly, for $\delta = 0.005$, the condition (again with rounded values) is $0.3746 \leq \theta_1 \leq 0.3753$. It is still the case that player 1's payoff decreases in θ_1 in this region, so the paradox still occurs. For $\delta = 0$, expected payoffs coincide with θ, and the paradox cannot be observed.

[8] It is also concave in θ_1. This implies that a donation (i.e., a reduction in θ_1) increases y_1 at a decreasing rate.

4 Generalizing the Example to a Class of Games

Suppose there are $n = m + 2$ players, where $m \geq 2$. Minimal winning coalitions are of two types: $\{1, i\}$ for $i \notin \{1, n\}$ and $N\backslash\{1\}$. Hence, there are three types of players: a large player (player 1), m medium-size players (each of which can form a winning coalition with player 1), and one small player. A homogeneous representation would be $q = 2m + 1$, $w_1 = 2(m - 1) + 1$, $w_i = 2$ for $i \notin \{1, n\}$ and $w_n = 1$. The game is close to an apex game, the only difference being that the large player and the small player together do not have a majority. Analogously to the case of apex games, there is exactly one game in this class for each $m \geq 2$. The example discussed in the previous section corresponds to $m = 2$.

Suppose player 1 is selected to propose with probability θ_1, each medium-sized player is selected with probability θ_m, and player n is selected with probability $1 - \theta_1 - m\theta_m \geq 0$. As in the previous section, we focus on the SSPE with symmetric strategies. Player 1 has no use for player n as a coalition partner and will propose to each medium-sized player with probability $1/m$. We investigate the parameter region where the SSPE is such that medium-sized players are indifferent between proposing to the large player or to the other medium-sized players and the small player, that is, $y_1 = (m - 1)y_m + y_n$. Assuming $\delta = 1$, the following conditions must be satisfied in such an equilibrium.

$$
\begin{aligned}
y_1 &= \theta_1(1 - y_m) + m\theta_m\lambda y_1 \\
y_m &= \theta_m(1 - y_1) + \frac{\theta_1}{m} y_m + (m - 1)\theta_m(1 - \lambda)y_m + (1 - \theta_1 - m\theta_m)y_m \\
y_n &= (1 - \theta_1 - \theta_m m)(1 - m y_m) + m\theta_m(1 - \lambda)y_n \\
y_1 &= (m - 1)y_m + y_n \\
\lambda &\in [0, 1] \\
y_1 &\geq y_m
\end{aligned}
$$

The solution to the system of the first four equations is

$$
\begin{aligned}
\lambda &= \frac{1 - 2\theta_1}{m\theta_m} \\
y_1 &= \frac{(1 - \theta_1)(m - 1)}{2(m - 1)(1 - \theta_1) + m\theta_m} \\
y_m &= \frac{m\theta_m}{2(m - 1)(1 - \theta_1) + m\theta_m} \\
y_n &= \frac{(1 - \theta_1 - m\theta_m)(m - 1)}{2(m - 1)(1 - \theta_1) + m\theta_m}
\end{aligned}
$$

We now add the condition that $\lambda = \frac{1-2\theta_1}{m\theta_m}$ must be between 0 and 1. This is the case when $\frac{1-m\theta_m}{2} \leq \theta_1 \leq \frac{1}{2}$. For $\theta_1 = \frac{1}{2}$, $\lambda = 0$, and player 1 receives no proposals. For $\theta_1 = \frac{1-m\theta_m}{2}$, $\lambda = 1$.

The final condition that needs to be satisfied is $y_1 \geq y_m$; if this condition fails, player n would be better off replacing one of the medium-sized players with player 1. This condition is satisfied by the solution because $y_1 = (m-1)y_m + y_n$, so $y_1 < y_m$ would imply $y_n < 0$. The value of y_n found as a solution of the system is clearly nonnegative.

So, the mixed strategy equilibrium region is nonempty for all $m \geq 2$. It is given by $\theta_m > 0$, $1 - m\theta_m - \theta_1 \geq 0$, and $\frac{1-m\theta_m}{2} \leq \theta_1 \leq \frac{1}{2}$.

For θ_1 in this region and keeping θ_m constant, player 1's payoff decreases in θ_1 (so increases if θ_1 is reduced, holding θ_m constant, which is a "donation" from the large to the small player). Note that if $\theta_1 = \frac{1-m\theta_m}{2}$, then the small player also has a proposing probability of $\frac{1-m\theta_m}{2}$; hence, the donation is only profitable up to the point where the large and the small player have the same proposing probability. This implies that the egalitarian protocol is invulnerable to this paradox in this class of games.

5 Discussion

The proposer advantage means that, if we fix the vector of recognition probabilities θ and thus the equilibrium payoffs, actually being recognized to be proposer is always good news. However, if we allow θ to change (and strategies to adjust to the new equilibrium), having a higher probability of being recognised can be bad news.

It is known that the indirect effect can offset the direct effect. (Baron and Ferejohn, 1989, p. 1194)) have an example with different recognition probabilities but equal payoffs. Similarly, (Montero, 2002) shows that, for apex games, all values $0 < \theta_1 \leq 0.5$ of the apex player's recognition probability lead to the same expected payoffs. If the apex player becomes the proposer more often, it receives proposals less often so that the competitive payoff for apex games $y_1 = (n-2)y_2$ is maintained.

An important difference between the game [5; 3, 2, 2, 1] (and other games in the class discussed in the previous section) and apex games is that the competitive payoff vector is unique for apex games because apex games have a unique homogeneous representation (up to rescaling). The indifference condition $y_1 = (n-2)y_2$ together with $y_1 + (n-1)y_2 = 1$ determines expected payoffs uniquely in apex games.

The game [5; 3, 2, 2, 1] has many competitive payoff vectors because it has many homogeneous representations, so that assuming that the outcome is competitive does not lead to a unique payoff vector. For example, [7; 4, 3, 3, 1] is a homogeneous representation of the same game obtained by increasing the weights of the medium-sized players and adjusting w_1 and q accordingly. If we normalize the weights so that they add up to 1, it is easy to compute all homogeneous representations. Clearly, players 2 and 3 must have the same weight in any homogeneous representation.

Denote the weights by w_1, w_2 and w_4, respectively. Normalization implies that

$$w_1 + 2w_2 + w_4 = 1 \tag{1}$$

Homogeneity implies that $w_1 + w_2 = 2w_2 + w_4$, or

$$w_1 = w_2 + w_4 \tag{2}$$

Note that the homogeneity condition is the same as the indifference condition that we obtained previously for a medium-sized player, but with weights instead of payoffs. Solving this system, we obtain

$$w_2 = 1 - 2w_1 \tag{3}$$
$$w_4 = 3w_1 - 1 \tag{4}$$

It turns out that w_2 is negatively related to w_1, whereas w_4 is positively related to w_1.[9]

Since $\{1, 4\}$ and $\{2, 3\}$ are losing coalitions, there are two additional constraints: $w_2 > w_4$ guarantees that $\{1, 4\}$ is losing, and $w_4 > 0$ guarantees that $\{2, 3\}$ is losing. Taking these constraints into account, we find that any value w_1 such that $\frac{1}{3} < w_1 < \frac{2}{5}$ leads to a homogeneous representation (the corresponding intervals for the other two players are $\frac{1}{3} > w_2 > \frac{1}{5}$ and $0 < w_3 < \frac{1}{5}$).

If we assume a "competitive" equilibrium in which $y_1 = y_2 + y_4$ (equivalent votes receive the same payoff), expected payoffs must satisfy Eqs. (3) and (4), and the payoffs of 1 and 4 must vary together. This goes some way toward explaining the phenomenon (if a donation from 1 to 4 affects y_1 and y_4, it must have a paradoxical effect) though it does not explain why payoffs change when 1 donates probability to 4 instead of remaining constant.

6 Concluding Remarks

Being recognized as a proposer is always a good thing *ex post*. However, having a higher recognition probability can hurt a player. The reason is that the indirect effect of this donation may outweigh the direct effect: The recipient is now less likely to receive proposals, and that effect more than compensates for the increase in the recognition probability.

In the example identified and its generalization, the donation is only profitable up to the point where the donor and the recipient have the same proposer probability. It follows that the egalitarian protocol is invulnerable to the paradox in this class of games. Whether this result extends to all games is an open question.

[9] This generalizes to the class of games in the previous section as $w_m = 1 - 2w_1$ and $w_n = (2m - 1)w_1 - m + 1$.

The paradox appears to be connected to the fact that the set of minimal winning coalitions is not rich enough, so that the homogeneous representation of the game is not unique. Identifying a class of games for which the paradox does not occur (besides apex games) would be an interesting topic for future research.[10]

References

Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review, 83*, 1181–1206.

Baron, D. P., & Kalai, E. (1993). The simplest equilibrium of a majority-rule division game. *Journal of Economic Theory, 61*, 290–301.

Diermeier, D., & Fong, P. (2011). Legislative bargaining with reconsideration. *Quarterly Journal of Economics, 126*, 947–985.

Diermeier, D., & Merlo, A. (2004). An empirical investigation of coalitional bargaining procedures. *Journal of Public Economics, 88*, 783–797.

Duggan, J., & Ma, Z. (2018). Extreme agenda setting power in dynamic bargaining games. Working paper

Eraslan, H., & McLennan, A. (2013). Uniqueness of stationary equilibrium payoffs in coalitional bargaining. *Journal of Economic Theory, 148*, 2195–2222.

Felsenthal, D. S., & Machover, M. (1995). Postulates and paradoxes of relative voting power: A critical re-appraisal. *Theory and Decision, 38*, 195–229.

Felsenthal, D. S., & Machover, M. (1998). The measurement of voting power: Theory and practice, problems and paradoxes. Edward Elgar Publishing

Harrington, J. E. (1990). The power of the proposal maker in a model of endogenous agenda formation. *Public Choice, 64*, 1–20.

Holler, M., & Napel, S. (2005). Local monotonicity of power: Axiom or just a property? *Quality and Quantity, 38*, 637–647.

Kadane, J. B., Stone, C. A., & Wallstrom, G. (1999). The donation paradox for peremptory challenges. *Theory and Decision, 47*, 139–151.

Kalandrakis, T. (2006). Proposal rights and political power. *American Journal of Political Science, 50*, 441–448.

Kalandrakis, T. (2010). Minimal winning coalitions and endogenous status quo. *International Journal of Game Theory, 39*, 617–643.

Kurz, S., Maaser, N., Napel, S., & Weber, M. (2015). Mostly sunny: A forecast of tomorrow's power index research. *Homo Oeconomicus, 32*, 133–146.

Laruelle, A., & Valenciano, F. (2005). A critical reappraisal of some voting power paradoxes. *Public Choice*, *125*, 17–41.

Montero, M. (2002). Noncooperative bargaining in apex games and the kernel. *Games and Economic Behavior, 41*, 309–321.

Montero, M. (2006). Noncooperative foundations of the nucleolus in majority games. *Games and Economic Behavior, 54*, 380–397.

Napel, S., & Widgrén, M. (2001). Inferior players in simple games. *International Journal of Game Theory, 30*, 209–220.

[10] To the best of my knowledge, there are no general results on the comparative statics of changing recognition probabilities. (Kalandrakis, 2010) shows that *any* expected payoffs can be obtained for *some* recognition probabilities but contains no claims on what probabilities lead to what payoffs.

Okada, A. (1996). A noncooperative coalitional bargaining game with random proposers. *Games and Economic Behavior, 16*, 97–108.

Romer, T., & Rosenthal, H. (1978). Political resource allocation, controlled agendas, and the status quo. *Public Choice, 33*, 27–43.

Egalitarian Collective Decisions as 'Good' Corporate Governance?

Federica Alberti, Werner Güth, Hartmut Kliemt, and Kei Tsutsui

Abstract Value-neutral (Robbinsian) economic science cannot directly address substantive normative issues. Economics can, however, provide analytical and empirical methods that make implications and consequences of normative premises (more) transparent and thereby indirectly contribute to normative opinion formation. To this effect, we translate substantive normative premises of stakeholder value approaches into operational axioms that characterize a class of collective decision mechanisms. If such implications seem less attractive to stakeholder theorists than the high-minded values from which they started in criticism of shareholder value approaches, they should come up with alternative collective decision mechanisms or a modified set of values.

1 Introduction and Overview

The conflict between so-called shareholder and stakeholder approaches to corporate governance forms one if not the defining controversy of modern business ethics. According to Milton Friedman's traditional shareholder value approach, owners (respectively their managerial agents) are entitled (in competition with other likewise entitled owners) to pursue owner – or – in case of stock-companies – shareholder

F. Alberti (✉)
University of Portsmouth, Portsmouth, UK
e-mail: federica.alberti@port.ac.uk

W. Güth
Max Planck Institute for Research on Collective Goods, Bonn, Germany
e-mail: gueth@coll.mpg.de

H. Kliemt
Justus-Liebig-Universität Giessen, Giessen, Germany
e-mail: hartmut.kliemt@wirtschaft.uni-gissen.de

K. Tsutsui
University of Bath, Bath, UK
e-mail: k.tsutsui@bath.ac.uk

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_11

value.[1] Advocates of 'socially responsible management' lean toward Edward Freeman's view that not only shareholders but *all* stakeholders of a company have to be taken into account with ('quasi-Kantian') *equal* respect as 'ends in themselves'.[2]

Taking an ethical stance on these issues, violates the Weberian *methodological norm* of value-neutrality which respects the 'fact-value' ('quid facti—quid juris') and the parallel 'descriptive-prescriptive' distinction.[3] Subscribing to such a rather orthodox view we do not deny that like all human practices, the practices of science are norm-guided. It is a constitutive norm of value-neutral scientific practice that scientists may claim the authority of science only in answering questions concerning the suitability of means but in their role as scientists must remain silent on ends.[4]

In the next Sect. 2, we recapitulate central implications of Weberian value-neutrality in Robbins' "essay on the nature and significance of economic science" (Robbins, 1932) and beyond. In Sect. 3, we show that Friedman and Freeman *both* do not live up to Robbins' methodological norms of value-neutral science. Section 4 presents in an exemplary manner what value-neutral economics can contribute to foundational controversies like that between stakeholder and shareholder conceptions by a translation of what adherents of quasi-Kantian respect preach into specific procedural proposals concerning means to such ends. Section 5 further scrutinizes the proposed procedure as a 'technology' in terms of its game-form[5] and the relation to the game that might emerge from this. Section 6 concludes that perhaps all applied economics should be conducted in the engineering spirit of exploring alternative technological means rather than discussing the ends for which the means might be used.

[1] In a normative welfare economic perspective, they would, to secure Pareto efficiency, also be normatively obliged to follow market signals. This fairly common assumption among economists—apparently including Friedman himself—is, however, not in line with a privilege of ownership that conceptualizes the owner as 'sovereign' in disposing of private property within the limits of law: the competitive market may provide extrinsic motives to choose in ways conducive to the attainment of efficiency—as defined in terms of the game *form*—but the intrinsic motive to act accordingly need not prevail and neither be seen as a morally binding precept to further the common weal.

[2] We do not claim to present an interpretation true to what Kant could have meant. We speak of Kantian ideals of equal interpersonal respect in opposition to utilitarian (aggregationist) ideals of the common weal. This is the spirit in which Rawls and also Buchanan are often seen as being part of the Kantian camp of moral philosophy and, say, Harsanyi of the utilitarian.

[3] This in turn demands acknowledging a. that prescriptive statements cannot be derived from sets of exclusively descriptive statements and b. that prescriptive statements cannot corroborate or falsify descriptive ones.

[4] For the value-neutral conception of technological research in applied science, see (Albert, 1985) and (Albert, 2010, 2022).

[5] We use the concept of the game-form as familiar from social choice theoretical debates basically for the objective rules of the game. The Game as an object of common knowledge arises once subjective aspects—subjective preferences and beliefs are factored in. The latter are like the game-form also beyond choices within plays of the Game and in this sense part of the 'rules of the Game' along with its game-form.

2 Relativism and Applied Economic Science

In his seminal essay, Lionel Robbins restricts applied economics to nomological knowledge of cause-and-effect relations and to providing technological advice on how desired ends may be achieved in the shadow of scarcity. The arguably best characterization of the underlying narrow conception of reason has been provided by Alisdair MacIntyre:

> Reason is calculative, and it can assess truths of fact and mathematical relations but nothing more. In the realm of practice, therefore, it can speak only of means. About ends it must be silent. [6] (MacIntyre 1984, p. 54)

According to this narrow conception of reason, ends (aims or values) are to be treated as exogenous to Robbinsian economics. We merely add that this conception is adequate not only for economics as an applied science but also—and perhaps even more so—for management science if it is interpreted as a technological discipline that proposes means to ends without defending or censoring the ends as such.

As Robbins (1938) himself emphasizes the silence of science about (the legitimacy of) ends must be extended to the problem of fixing trade-offs between the achievement of ends that in the shadow of scarcity cannot be realized simultaneously (at least not all to the maximal separately desired extent).[7] In consequence, scientific advice seems restricted either to the diagnosis of Pareto inferior states, and the suggestion of how to overcome them by Pareto superior moves or to the proposal of decision-making mechanisms that let the individuals themselves determine the relevant trade-offs. In the latter case, which is center stage in the present context, many economists seem to assume that unanimity in collective decisions is akin to a normative default option by which the advisees themselves are procedurally restricted to decisions akin to Pareto superior moves.[8]

If a narrow conception of reason is accepted, the preceding line of argument, familiar as it may be, is nevertheless fundamentally flawed: if economics is understood as a *technological* applied science (Albert, 1985) that singles out means to the achievement of ends that must be given exogenously to economic science, then the restriction of economic *advice* to Pareto improvements and/or to the pursuit of aims, ends, or values that could be unanimously accepted by all concerned is itself

---

[6] MacIntyre (1984) who provides this apt characterization of a narrow conception of reason is an ardent critic of it.

[7] Obviously, even in case of several Pareto improvements, not all need be maximally advantageous to each separate individual so that even then inter-individual conflict concerning which of several improvements should be realized prevails.

[8] In Buchanan's scheme of things, the diagnosis that an alternative is a potential Pareto improvement is a hypothetical that must be corroborated ('ratified') by unanimous agreement to bringing it about. Unanimous agreement of the actors themselves rather than the external diagnosis of Pareto inferiority is the legitimation.

a *substantive* value judgment that *cannot* be justified *within* economic science but is exogenous to it.[9]

At the risk of beating this to death, technological institutional proposals of value-neutral economics can point out how to reach the ends of particular individuals at the expense of other individuals. Very much like technologies of bomb-building, economic technologies can assist each of a multiplicity of parties in getting their way in competition with others independently of the relative merits of individual ends that may or may not prevail.[10] That many economists may for themselves seek to contribute to the peaceful resolution rather than the exacerbation of conflicts is their personal value but nothing that is supported by methodological norms like fact-orientation and value-neutrality of economics as a science.

Quite in line with this traditional medical ethics—as opposed to public health ethics—has been decidedly *not* (!) in line with the requirements of symmetric respect and the common weal of *all* patients—of some collectivity—but demanded of the medical doctor to be *partisan* for her particular patient. Likewise in traditional management counseling, the established *role obligation* of the counselor is that of an agent who (similar to an attorney) provides advice on what furthers the partial interests of her particular client rather than allegedly universalizable interests of all.[11]

In any event, many of the ethical disputes between shareholder and stakeholder conceptions can be side-stepped if the exogenous normative premises of arguments are treated as hypotheticals on whose legitimacy the theorist remains silent while focusing on means furthering aims, ends, or values. Before pursuing this strategy in an exemplary manner in Sects. 4 and 5 with respect to stakeholder approaches, let us briefly show that our criticisms are not directed at strawmen.

3 Friedman, Freeman as Brothers in Sin

3.1 Friedman

Shareholder value conceptions like that of Milton Friedman start from a specific order of ‘private law’ which fulfills the general “three fundamental rules of jus-

[9] Substantive restrictions to Pareto improvements on the diagnostic level and to unanimity requirements on the level of mechanisms may as a matter of fact be desired by particular persons. Yet, according to value-neutrality as a constitutive methodological norm of Robbinsian economics, the theorist has a reason to (re-)search exclusively such suggestions if those requiring her advice as a matter of fact demand to receive exclusively advice concerning Pareto superior moves.

[10] Which party will prevail is ultimately a matter of the ability to get ones way rather than a matter of argument. Of course, the fact that economic counsel concerning instruments may or may not be private information to the counseled party is relevant here, too. To avoid this, economists routinely assume that information is common knowledge—or at least subject to a process of dissemination that will approach common knowledge.

[11] As in the medical case, in case of management, partisanship rather than impartiality is the basis of the trust relationship between counselor and client.

tice, the stability of possession, its transference by consent, and the performance of promises" (Hume 1739, Bk III, part II, sec. xi). In Western, Educated, Industrialized, Rich, Democratic, WEIRD, societies on which Friedman and Freeman focus a quite extended realm is politically specified as private and citizens are legally entitled to follow their personal inclinations.[12] They are legally entitled to exert consequences of their choices even if those are regarded as negative, even resented as repugnant, by third parties (unless, of course, the externalities are deemed illegitimate by law). To the extent that the specific prevailing legal conventions of corporate law fulfill the general criteria laid down by Hume they empower 'owners' in general and shareholders in particular to do with their possessions as seems fit to them.[13]

It is worth emphasizing that Milton Friedman claims that "*the social responsibility of business is to increase its profits*" (Friedman, 1984) does not follow from an acknowledgement of private law entitlements. Quite to the contrary, assuming otherwise would contradict the privilege of private property that adherents of shareholder value conceptions seek to defend. Entrepreneurs who own a firm are entitled to choose (for whatever reasons) as seems fit to them. They cannot be under a private law obligation to follow market signals in profit increasing ways even if such behavior can be demonstrated to further wealth or some conception of the common weal.

Friedman's claim is a personal value judgment that has, as it stands neither the authority of law nor that of science on its side.[14] Atleast, single owners are clearly legally entitled *not* to live up to the social responsibility that Friedman ascribes to them. Only in case of a firm with multiple owners, a technological argument in favor of increasing shareholder value can be offered: in this special case, the management can discharge its fiduciary contractual obligations to multiple owners of shares who pursue a plurality of values best by maximizing the value of tradeable shares. For, if those who disagree with management decisions can sell at maximal share value, this enables them to pursue their own aims, ends, or values to the largest extent outside the contract nexus of the firm. They signal their implicit assent by *not* selling shares on liquid markets. As long as they do not go short, they can be held responsible for not doing so. According to the time-honored maxim 'volenti non fit iniuria', they cannot under such consensual conditions be (ab-)used as mere means by management.[15]

---

[12] So-called consumer sovereignty is just one example of this sovereignty in matters that are by constitutional politics put beyond in period politics.

[13] Böhm (1966), an influential German constitutional lawyer, uses this expression to characterize markets as a subsystem of the economic order of a free western society. Like other members of the so-called ordo liberal Freiburg School, he strongly sympathized with the values of the Mont Pelerin society among whose 'noble' members were besides its founder F. A. v. Hayek, scholars like G. Becker, J. M. Buchanan, M. Friedman, V. Smith.

[14] By assumption, the relevant specification of law is assumed to comply with 'the three fundamental rules of justice' which rule such a positive law obligation out, likewise, by assumption the value-neutral practice of science leaves no room for justifying moral responsibilities as conclusions of science.

[15] A more detailed presentation of this argument can be found in Kliemt (2022).

The preceding line of argument shows that maximizing share value is expressive of respect for the aims, ends, and values of shareholders. Yet, this operationalization of quasi-Kantian norms of interpersonal respect rests on owner privilege as characterized by “the three fundamental rules of justice”. Freeman does not accept this.

3.2 Freeman

In opposition to prevailing shareholder, value conceptions of good corporate governance which treat the firm “as a nexus of contracts” protected under the owner privilege Edward Freeman intend to “revitalize the concept of managerial capitalism”:

> My thesis is that we can revitalize the concept of managerial capitalism by replacing the notion that managers have a duty to stockholders with the concept that managers bear a fiduciary relationship to stakeholders. Stakeholders are those groups who have a stake in or claim on the firm. Specifically, I include suppliers, customers, employees, stockholders, and the local community, as well as management in its role as agent for these groups. I argue that the legal, economic, political, and moral challenges to the currently received theory of the firm, as a nexus of contracts among the owners of the factors of production and customers, require us to revise this concept. That is, each of these stakeholder groups has a right not to be treated as a means to some end and therefore must participate in determining the future direction of the firm in which they have a stake. (Freeman 1984, p. 184)

This remarkable passage has the air of a political-ethical appeal rather than that of a contribution to a theoretical debate in economics and management. Yet, it is certainly meant also as a disciplinary contribution to the discourse about ‘what managers should do’. So let us take it as such a contribution.

The last sentence of the passage connects two crucial theses about Freeman’s rather comprehensive list of stakeholder groups to be respected equally (i.e., without owner privilege) by management: first, “‘that each of these stakeholder groups has a right not to be treated as means to some end’ and second, ‘therefore must participate in determining the future direction of the firm in which they have a stake’”.

Though much more might be said about the passage, we focus in our comments on the comprehensiveness of the list of stakeholders and that it is a list of *groups* of stakeholders whose participation is demanded.

Freeman sees the manager in a fiduciary relationship to all who may be affected by management decisions: “I include suppliers, customers, employees, stockholders, and the local community”. This first person-statement may correctly represent Freeman’s ‘ethical opinions’ but it does not demonstrate that a corresponding ‘right’ does *exist* as part of positive law. The mere demand that a right exist is not that right, “want is not supply—hunger is not bread”.[16] Neither does it suffice for the existence of rights that broadly Kantian principles of respect are as a matter of fact widely

[16] Article 2 of Bentham (1843).

shared among citizens of WEIRD societies and lead to a demand for such rights. Finally, in Freeman’s account, the basis for the fiduciary obligations he assumes are not legal (contractual) facts.[17]

So, let us turn to group orientation. The preceding extended quote speaks of groups rather than of individuals that should never be treated merely as means. In view of the traditional concern that in collective decisions, individual members of minorities are permanently at risk of being used as means to the ends of majorities it seems at least prima facie surprising that Freeman demands that in order to prevent being used as a means “each of these stakeholder groups ... must participate in determining the future direction of the firm”.

Since Freeman rejects the ‘currently received theory of the firm, as a nexus of contracts’ as defining the fiduciary relationships of management his approach makes sense only if the ‘right not to be treated as mere means to some end’ is not sufficiently protected by the rules of the ‘private law society’ with its protection of individual rights by a polity-wide constitution. As a remedy, he proposes participatory procedures of collective decision-making concerning strategic decisions of companies.

What Freeman seems to have in mind is akin to established institutions of managerial co-determination familiar from German corporate law. But it has obvious roots in the intellectual environment of American liberal economics as well. After all, the Darden School at which Freeman spent most of his academic career is part of the University of Virginia, Charlottesville. In the UVA economics department, there has been an ongoing tradition of efforts to incorporate Kantian interpersonal respect norms into economics. For instance, in his 1956 UNESCO report “on the state of economics in the United States of America” Rutledge Vining, a then leading member of the UVA economics faculty characterizes, (Virginia) Political Economy as endorsing substantive interpersonal respect norms. In the course of his discussion, he states: “To require of each individual that he takes no action which impairs the freedom of any other individual is to accept the moral principle that no individual should treat another simply as a means to an end” (Vining 1956, p. 19). This is almost literally what Kant says.

James M. Buchanan, who along with Ronald Coase had joined the UVA faculty in the mid 1950s on Vining’s initiative, has always worked in the equal mutual respect framework to which he himself refers as ‘politics as exchange’. The voluntariness of exchange represents interpersonal respect, and the inclusion of all affected is represented by the assumption that “‘political exchange’ necessarily involves *all* members of the relevant community rather than the two trading partners that characterize economic exchange”. (Buchanan 1979, p. 50, emphasis in original).[18]

This supports the view that Freeman definitely has more in mind than ‘boring’ German ‘corporate democracy’. What this might be is, however, hard to say without

---

[17] Even if we would assume that there is objective knowledge of what is ethically right and/or wrong this would as such still *not* explain the existence of rights as parts of legal and/or moral *institutions*. As it stands, Freeman’s list of stakeholders is no more than a personal wish list.

[18] For more on this strand of ‘normative economics’, see Brennan and Kliemt (2019) and Kliemt (2011).

translating high-minded ethical ideals of stakeholder approaches into concrete rules. Such a translation can make it also easier to assess the relative merits of stakeholder vis a vis shareholder approaches.

Before we turn to that, let us note for the record that as far as the argument goes Freeman's appeals to ethical ideals are—like Friedman's endorsement of the responsibility to increase profits—rely personal opinions. Value-neutral arguments cannot deliver a foundation for such ethical values and norms but they can contribute to a better understanding of their potential implications and consequences. Next, we outline in an exemplary constructive manner what such a contribution might look like.

4 An Outline of a Procedure of Stakeholder Participation as Egalitarian Bidding

Taken seriously the programmatic acknowledgment that each of the many "stakeholder groups has a right not to be treated as a means to some end and therefore must participate in determining the future direction of the firm in which they have a stake" can hardly be operationalized in ways that allow real-world implementation.[19] To illustrate the argument in principle, a discussion of employee participation along with participation of management as agent of owners must and can do.

So let us *hypothetically* accept the quasi-Kantian ideal of not using others merely as means in an n-stakeholder setting. Our aim is to add transparency by translating values of equal respect into axioms that characterize a procedure of employee participation. The result is an outline of a procedure that could conceivably serve as an instrument of protecting the "right not to be treated as means to some end".

Adding further ideals that are prevalent in WEIRD market societies, we assume that the corporate governance structure should procedurally support the search for universal advantage and agreement under constraints of inter-individual respect. To this effect, basically, Buchanan's ideal of "politics as (multilateral) exchange" must be extended from political to non-political corporate actors.[20]

The mechanism we shall outline next incorporates both a common denominator (facilitating compromise and agreement to trade off) and group-based veto power (making compromise and concessions a necessity). As a common denominator that facilitates compromise and agreement outside a market also in a participatory context, the measuring rod of money will be used.

[19] That participation in interactions in regular markets can be operationalized in implementable ways is, to put it mildly, not the least advantage of this social 'technology'. But it is ruled out in the group participation framework endorsed by Freeman.

[20] See (Buchanan (1999), vol. 1), passim and for further discussion of some specific aspects (Brennan and Kliemt, 2018).

4.1 An Axiomatic Characterization of Stakeholder Participation

Assume that the management considers plans of how to (re)structure the firm. Let M denotes the set of mutually exclusive plans and refer to $m \in M$ as a typical element of that set. Assume that for each plan m, management states a 'surplus claim' $S(m)$.

Being informed about M and $S = S(m) : m \in M$ by management, which has the prerogative of an agenda setter, the stakeholders $i = 1, \cdots, n (\geq 1)$ can participate via bidding, i.e., stating bids $b_i(m) \in \mathbb{R}$ for all $m \in M$ according to the following rules:

Veto condition: If $b_1(m) + \cdots + b_n(m) < S(m)$, for some $m \in M$, then plan m is rejected.

Since stakeholder groups $i = 1, \cdots, n$—respectively, their representatives—are all 'free to choose' their monetary bids $b_i(m) \in \mathbb{R}$ (*including negative values*), they can obviously veto a plan by some appropriately low bid so that for any $m \in M$ the veto condition is fulfilled. By their veto, they incur the opportunity cost that m is not realized. Otherwise, *voicing* an 'appropriately low' figure as a bid is a purely expressive act that imposes no higher (transaction) costs on them than voicing a higher bid.

If $b_1(m) + \cdots + b_n(m) \geq S(m)$, plan $m \in M$ is not vetoed. A necessary condition for implementing $m \in M$ is met, and management is authorized by the rules to consider it an eligible option.

It may be worth noting that management can subsidize ($S(m) < 0$) certain plans from other resources of the company to win stakeholders over. Yet, there is no guarantee that any plan will meet the necessary condition for implementing it; that is, after, stakeholders have been bidding on all $m \in M$ the subset

$$M_a = m \in M : b_1(m) + \cdots + b_n(m) \geq S(m)$$

of acceptable plans $m \in M$ may be empty.

Status quo condition: If $M_a = \phi$ or if, in case of $M_a \neq \phi$, none of the plans in M_a is realized by management the status quo is maintained.

Stating all payoffs in relation to the status quo, we can assume that management and stakeholders $i = 1, \cdots, n$ all receive 0-payoffs if either $M_a = \phi$ or $M_a \neq \phi$ and management abstains from realizing any of the acceptable plans $m \in M_a$ despite $M_a \neq \phi$.

Equal split condition: If $M_a \neq \phi$ and plan $m \in M_a$ is realized by management, then management receives $S(m)$ and stakeholder groups $i = 1, \cdots, n$ earn

$$v_i(m) - b_i(m) + \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n}.$$

$b_1(m) + \cdots + b_n(m) - S(m)$ is the surplus and $v_i(m)$ is the *true* value of stakeholder group i when $m \in M_a$ is realized. Of course, $v_i(m)$ usually is i's private information while all the other factors determining stakeholder i's information are commonly known from overt bidding.

If $v_i(m)$ is i's private information, it is impossible to find solutions for the game-form that underlies the mechanism, unless the total value $v_1(m) + \cdots + v_n(m) = V(m)$ is known. However, if $V(m)$ is common knowledge, it can be shown that the strategic equilibrium requires $\sum_{i=1}^{n} b_i(m) = S(m)$, hence

$$V(m) - \sum_{i=1}^{n} b_i(m) = V(m) - S(m).$$

If the manager is a profit maximizer, the manager leaves the smallest monetary unit ϵ to each stakeholder group in the subgame-perfect equilibrium:

$$v_1(m) - b_1(m) = \cdots = v_n(m) - b_n(m) = \epsilon.$$

Therefore, the manager's surplus claim is given by

$$\begin{aligned} V(m) - \sum_{i=1}^{n} b_i(m) &= n\epsilon \\ V(m) - S(m) &= n\epsilon \\ S(m) &= V(m) - n\epsilon. \end{aligned}$$

The mechanism grants veto power with the only—possibly negative—'opportunity cost' that arises from the fact that bidders must jointly meet the requirement of $M_a \neq \phi$ if any positively valued change is to occur. Therefore, whatever their $v_i(m)$, the stakeholders usually have an incentive to underbid.[21] Still, despite the underbidding incentive, bidding of the kind we suggest is at least one *feasible* way of translating into procedural terms the ideal of Kantian equal respect for stakeholders along with an adequately privileged role of management as agenda setter. Moreover, as we endeavor to indicate next, the procedure seems normatively and empirically more reasonable than economic folk wisdom on mechanism design may initially suggest.

---

[21] Only in case of commonly known $v_1(m) + \cdots + v_n(m) = S(m)$ underbidding incentives would not exist.

4.2 Properties of Participation as Egalitarian Bidding

The egalitarian bidding mechanism has three properties:

(V) The mechanism guarantees “*voluntariness*”.

It fulfills

$$v_i(m) - b_i(m) + \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n} \geq 0$$

if $b_i(m) \leq v_i(m)$ for $i = 1, \cdots, n$.

That is, stakeholder groups can only lose, relative to the status quo, when overbidding; furthermore, by bidding low enough, possibly via $b_i(m) < 0$ or $b_i(m) < v_i(m)$, each stakeholder (group) i can veto any plan $m \in M$.[22]

As has already been mentioned, the values $v_i(m)$ for all $m \in M$ can, and as a rule, will be private information of stakeholder groups. This obviously renders procedural guarantees of equal treatment with respect to *private values* impossible. However, this does not rule out guarantees of *equal* treatment with respect to *monetary bids* $b_i(m)$. The latter and the equal treatment with respect to them are overt acts that are observable.

(E) The mechanism guarantees “*equal respect according to (overt) bids*”.

Substituting $v_i(m)$ by $b_i(m)$ in the payoff specification for stakeholders yields *equal respect according to (overt) bids*

$$\begin{aligned} & b_i(m) - b_i(m) + \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n} \\ & = \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n} \quad \text{for } i = 1, \cdots, n. \end{aligned}$$

With respect to their interpersonally observable *bids*, all stakeholder groups $i = 1, \cdots, n$ are treated equally by receiving an equal share of $b_1(m) + \cdots + b_n(m) - S(m)$. That is, the procedure not only grants equal veto power but also *equal treatment as far as overt payoff consequences relative to the status quo* are concerned (and is egalitarian in this sense).

---

[22] The so-called “hold out problem”, that this gives rise to, probably motivates many scholars to reject the unanimity requirement. It should not be neglected, though, that some stakeholders i with $v_i(m) < 0$ for some $m \in M$ may get compensated when moderately underbidding in terms of $b_i(m) < v_i(m)$. In that case, m may still be acceptable and possibly implemented by management. In any event, the interest of guaranteeing that $m \in M_a$ applies will recommend to exercise some moderation in strategic underbidding.

(O) When only considering whether a given plan m or the status quo should be realized,[23] the co-determination mechanism is *overbidding proof*.

Underbidding incentives and overbidding proofness for a given plan m, of course, can only matter when they determine the consequences of an implemented plan m: in case of not implementing m, the consequence is only the status quo due to the focus on a given plan m for which we analyze how consequences depend on under-, respectively, overbidding. It does not pay for any stakeholder group $i = 1, \ldots, n$ to overbid the value $v_i(m)$ since, relative to truthful bidding $b_i(m) = v_i(m)$, overbidding would yield a disadvantage for stakeholder group i due to

$$v_i(m) - b_i(m) + \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n} < \frac{b_1(m) + \cdots + b_n(m) - S(m)}{n} \quad \text{for } i = 1, \cdots, n.$$

This reduction of payoff would apply if management implemented a plan $m \in M$ for which even the truthful bid $b_i(m) = v_i(m)$ would not guarantee acceptability, so that $b_1(m) + \cdots + b_n(m) \geq S(m)$for plan $m \in M$ results exclusively from i's overbidding. If for $b_i(m) = v_i(m)$ plan m is unacceptable, i.e., $b_1(m) + \cdots + b_{i-1}(m) + v_i(m) + b_{i+1}(m) + \cdots + b_n(m) < S(m)$, stakeholder i would suffer a loss when overbidding – $b_i(m) > v_i(m)$ – due to

$$\begin{aligned} & v_i(m) - b_i(m) \\ & \quad + \frac{b_1(m) + \cdots + b_{i-1}(m) + v_i(m) + b_{i+1}(m) + \cdots + b_n(m) - S(m)}{n} \\ & \quad + \frac{b_i(m) - v_i(m)}{n} \\ & = \frac{n-1}{n}[v_i(m) - b_i(m)] \\ & \quad + \frac{b_1(m) + \cdots + b_{i-1}(m) + v_i(m) + b_{i+1}(m) + \cdots + b_n(m) - S(m)}{n} < 0 \end{aligned}$$

by assumption.

The co-determination mechanism of "stakeholder participation as egalitarian bidding" shares the properties **(V)**, **(E)**, and **(O)** with familiar institutions like, in particular, first-price auctions.[24] Belonging to a class of familiar institutions is certainly desirable with respect to practical uses of a procedure. Yet, this is not sufficient to vindicate a mechanism like the proposed one against other possible objections.

[23] This neglects bidding by which one bids for m also in order to prevent the implementation of another plan $\tilde{m} \neq m$. In such cases, under-, respectively, overbidding incentives can depend on the expected bids after plan $\tilde{m}$ for which property **(O)** may not be true.

[24] It can be shown that some such mechanisms can be fully characterized by requiring that they be "envy-free" and meeting requirement **(E)** with respect to overt bids; see Güth (2011).

5 Critical Assessment of Incentives

5.1 *Underbidding Incentives*

As indicated, if ideals of quasi-Kantian equal respect are procedurally expressed in terms of "stakeholder participation as egalitarian bidding" the resulting mechanism is not underbidding proof. It invites bid shading, i.e., $b_i(m) < v_i(m)$ for $m \in M$ and $i = 1, \ldots, n$. Yet, whether this forms a decisive argument against relying on the mechanism depends on the aims, ends, or values that are pursued by those who consider using it.

First, there is no procedure fulfilling properties **(V)**, **(E)**, and **(O)** that is over and underbidding proof and generally implementable.[25] A trade-off between fulfilling desirable properties cannot be avoided in procedural implementations of the ideals of stakeholder theories.

Second, we cannot imagine a mechanism that confers discretionary power to make collective decisions on collective bodies and at the same time procedurally grants veto power to each and every member of the decision-making body unless property **(V)**—or some variant of it—is fulfilled (at least for groups). Adherents of stakeholder conceptions who reject "stakeholder participation as egalitarian bidding" but like Freeman accept that they need to go beyond mere appeals to quasi-Kantian ideals should come up with constructive procedural counter proposals if they intend to uphold their values.

Third, requiring **(E)** along with **(V)** translates ideals of substantive equality into procedural specifications. To the extent that the effects of **(E)** and **(V)** are perceived as expressive of ideals of substantive equality by participating stakeholders, this may psychologically (that is, *causally*) reduce their proclivity to underbid by strengthening the *intrinsic motivation* to bid truthfully or even to overbid due to some crowding in of, say, *corporate identity or corporate social responsibility* concerns.

5.2 *Intrinsic Motivation and Extrinsically Motivating Incentives*

The practical proof of quasi-Kantian appeals expressive of the values of stakeholder participation which is in the institutional 'eating'. This is why we took much care to specify rules that conceivably allow to test the implications of approximating Kantian normative ideals of stakeholder theories in practice. We are not naively assuming that the implementation of our (or any alternative) procedural translation of the ideals of stakeholder theories will lead to attractive results independent of context. Yet, there is some empirical (experimental) evidence showing that

[25] Using, for instance, the revelation principle would require the highly restrictive common knowledge assumption of game theoretic equilibrium analysis.

under favorable circumstances intrinsic motives may exert much stronger influences on practical stakeholder behavior than typical conceptions of mechanism design and principal agent theory assume in their search for 'knave proof' institutions.[26]

Experimental research on bidding mechanisms has provided ample evidence that rather subtle aspects of implementation may matter. Moreover, some first findings of bounded under- and even some overbidding in an explorative experiment on "egalitarian bidding" suggest seeking ways of strengthening such effects.[27] In particular, explicit framing to enhance awareness of properties **(V)**, **(E)**, and **(O)** may be expected to strengthen intrinsic motivation of stakeholders to act 'fairly' in what they regard as the common interest of all stakeholders.

Of course, ultimately any co-determination mechanism for corporate governance must live up to the test of competitive market evaluation: in countries in which the market for corporate control is working reasonably well instances of testing 'the Kantian pudding in the institutional eating' could in fact arise after some companies implemented values of stakeholder theory in terms of procedurally fair bidding for at least some types of decisions.

6 Concluding Remarks

The restriction to universally advantageous policy proposals that prevails in traditional so-called 'normative economics' whether it be Paretian, Buchanan-type contractarian, or Pigovian-utilitarian is incompatible with the narrow conception of reason and value-neutrality of economics as a science. Once stakeholder approaches of good corporate governance are translated into specific procedural proposals of collective decision and choice making—in the example of this paper—they can be evaluated in terms of the likely effects of implementing these procedures. The result of value-neutral economic analyzes as outlined in the preceding is technological blueprints of mechanisms that may be used in pursuit of exogenously given values. The result of such analyzes may support either acceptance or rejection of efforts to implement the technologies. In any event, the technologies are instrumentally rele-

[26] This point has been rightly emphasized by adherents of stakeholder conceptions from the start. It can be supported by many findings of psychology and experimental economics. Starting from Hume's well-known remark "that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a knave and to have no other end, in all his actions, than private interest". (Hume 1777, Essay VI, 42). Bowles (2016) provides an excellent overview over relevant experimental results concerning the validity of the behavioral assumptions underlying mechanism design in relation to 'moral motives'.

[27] In an explorative experimental study (Alberti et al., 2022) that implemented "stakeholder participation as egalitarian bidding", bounded bid shading and even some systematic overbidding could in fact be observed. It seems that some stakeholders did not want to block a plan that might be 'good for the firm' even though affecting themselves negatively. Obviously more research concerning effects of fairness perceptions and intrinsic motivation based on procedurally fair bidding is necessary before stronger evidence-based claims can be made.

vant *only* for those who as a matter of contingent fact do share suitable values. No claim needs to be made concerning the universal validity of the values that are treated as exogenous to economic science.

The analysis is in the spirit of traditional management theory that allows for giving advice conducive to particular rather than allegedly common and/or universal interests. In this spirit, stakeholder theories should make an extended effort to demonstrate, first, that the values they propagate can be spelled out by operational rules, second, how the rules can be implemented institutionally in corporate governance, and third, that corporations that implement corresponding mechanisms as part of their corporate governance structures can as a matter of fact survive and thrive in inter-firm competition.[28]

If these three conditions could in fact be met, this would take the sting out of the somewhat overblown controversy between stakeholder and shareholder conceptions in business ethics. Even adherents of shareholder value conceptions could then reasonably suggest that shareholders should in pursuit of their own values support the implementation of collective decision rules that are expressive of the values of stakeholder conceptions of good corporate governance. Companies could be "doing 'well' for their shareholders in the narrow sense of Friedman by doing 'good' in Freeman's sense to (all) other stakeholders". Whether that in fact be the case is an empirical issue.

We are skeptical that implementing proposals like those outlined here will in fact contribute to the attainment of democratic egalitarian ideals prevalent in WEIRD societies. On the whole, the outline seems to demonstrate that serious efforts to implement the appealing values invoked by stakeholder theorists may lead to extreme consequences. But we believe that the search for procedural implementations of the ideals of stakeholder theories could revitalize discussions of both managerial capitalism a la Freeman and market capitalism a la Friedman by putting them on track toward value-neutral analyzes. Merely preaching ethical ideals and to discussing them in ideal ethical theory terms seems a dead end.

References

Albert, H. (1985). *Treatise on critical reason*. Princeton University Press.

Albert, M. (2010). Critical rationalism and scientific competition. *Analyse and Kritik, 32*, 247–266.

Albert, M. (2022). How to escape from model platonism in economics: Critical assumptions, robust conclusions, and approximate explanations. Homo Oeconomicus, forthcoming

Alberti, F., Güth, W., & Tsutsui, K. (2022). Experimental effects of institutionalizing codetermination by a procedurally fair bidding rule. *Journal of Business Ethics*, forthcoming

Alchian, A. A. (1950). Uncertainty, evolution, and economic theory. *Journal of Political Economy, 58*, 211–221.

Bentham, J. (1843). Anarchical fallacies. In J. Bowring (Ed.), The works of Jeremy Bentham (Vol. 2, pp. 489–534). William Tait

[28] In the spirit of Alchian (1950).

Böhm, F. (1966). Privatrechtsgesellschaft und Marktwirtschaft. ORDO: Jahrbuch für die Ordnung von Wirtschaft und Gesellschaft, *17*, 75–151

Bowles, S. (2016). The moral economy: Why good incentives are no substitute for good citizens. Yale University Press

Brennan, G., & Kliemt, H. (2018). The constitution of markets. In R. E. Wagner (Ed.), James M. Buchanan: A theorist of political economy and social philosophy (pp. 807–838). Palgrave Macmillan

Brennan, G., & Kliemt, H. (2019). Kantianism and political institutions. In R. D. Congleton, B. Grofman, & S. Voigt (Eds.), *Oxford handbook of public choice* (pp. 796–813). Oxford University Press.

Buchanan, J. M. (1979). Politics without romance: A sketch of positive public choice theory and its normative implications. In G. Brennan, H. Kliemt, & R. D. Tollison (Eds.), The collected works of James M. Buchanan (1999, Vol. 1, The logical foundations of constitutional liberty, pp. 45–59). Liberty Fund

Buchanan, J. M. (1999). The logical foundations of constitutional liberty. In G. Brennan, H. Kliemt, & R. D. Tollison (Eds.), The collected works of James M. Buchanan (1999). Liberty Fund

Freeman, R. E. (1984). Stakeholder theory of the modern corporation. In W. M. Hoffman, R. E. Frederick, & M. S. Schwartz (Eds.), *Business ethics: Readings and cases in corporate morality* (pp. 184–191). Wiley-Blackwell.

Friedman, M. (1984). The social responsibility of business is to increase its profits. In W. M. Hoffman, R. E. Frederick, & M. S. Schwartz (Eds.), *Business ethics: Readings and cases in corporate morality* (pp. 180–183). Wiley-Blackwell.

Güth, W. (2011). Rules (of bidding) to generate equal stated profits: an axiomatic approach. *Journal of Institutional and Theoretical Economics, 167*, 608–612.

Hume, D. (1739). A treatise of human nature (2nd edition 1978). Oxford University Press

Hume, D. (1777). Essays IV and XII. In E. F. Miller (Ed.), Essays: Moral, political, literary (1985, based on the 1777 ed., pp. 32-36 and 87–96). Liberty Fund

Kliemt, H. (2011). Bukantianism—Buchanan's philosophical economics. *Journal of Economic Behavior and Organization, 80*, 275–279.

Kliemt, H. (2022). What should moral managers do? Management Revue, forthcoming

MacIntyre, A. (1984). *After virtue*. University of Notre Dame Press.

Robbins, L. C. (1932). *The nature and significance of economic science*. Macmillan.

Robbins, L. C. (1938). Interpersonal comparisons of utility: A comment. *Economic Journal, 48*, 635–641.

Vining, D. R. (1956). Economics in the United States of America: A review and interpretation of research. UNESCO

Interpretation and Measurement of Power

Liability Situations with Successive Tortfeasors

Frank Huettner and Dominik Karos

Abstract Given a tort that involves several tortfeasors, an allocation scheme attributes to each of them that part of the damage that reflects their responsibility. We consider successive torts—i.e., torts that involve a causality chain—and show that simple and intuitive principles, which are well-known in the law of tort, uniquely define an allocation scheme. We show that this scheme incentivizes agents to exhibit a certain level of care, creating an efficient prevention of accidents. We further describe the unique rule according to which a liability situation has to be adjusted after a partial settlement such that incentives to settle early are created.

1 Introduction

A driver negligently hits a pedestrian and a physician negligently treats, thereby aggravating, the pedestrian's injury ... if the first tortfeasor had not acted tortiously, the entire injury to the plaintiff—initial plus incremental damages—might have been avoided, whereas if only the second injurer had been nonnegligent only the incremental damages could have been avoided (Landes and Posner, 1980). In such a situation, the natural question arises: how much should each tortfeasor pay the victim in order to compensate for the damage? The principles of attributing damages to several tortfeasors depend on the legal regime; here we briefly summarize the discussions in (Kornhauser and Revesz, 2000) and (Wall, 1986) that are relevant for our purposes. Under *non-joint liability*, a damage must be attributed to each defendant and each defendant has to pay a compensation for the damage attributed to him. Hence there is a direct need for an allocation scheme that allocates damages to tortfeasors. Under *joint liability*, there is no direct need for such a scheme as the plaintiff can recover the

F. Huettner (✉)
SKK Graduate School of Business, Sungkyunkwan University, Seoul, Korea
e-mail: info@frankhuettner.de

D. Karos
Center for Mathematical Economics, Bielefeld, Germany
e-mail: dominik.karos@uni-bielefeld.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_12

full damage from any defendant he prevails, subject to not being overcompensated. However, resulting claims among the defendants may call upon an allocation scheme for the original situation to refer to. Similarly, an allocation scheme will help in the aftermath of settlements between the plaintiff and some of the defendants: under the *pro tanto* set-off rule the damage attributable to the remaining defendants is reduced by the settlement amount, whereas under the *apportioned set-off* rule it is reduced by the damage attributable to the settling defendants. Hence while in the first case an allocation scheme can serve as an anchor for settlements, in the second case it is necessary to calculate the set-off rule.

How should such an allocation scheme be designed? There are normative principles—axioms—such a scheme should satisfy: (1) It should ensure that all damages that have been caused negligently, should be recovered. (2) It should be consistent with a rule of liability, i.e., tortfeasors should have to pay only if they are liable according to such a rule (Shavell, 1980). (3) It should treat equals equally. (4) In case of successive torts, the amount attributed to a tortfeasor should not depend on the damages that preceded his involvement. (5) A tortfeasor who created no incremental damage but whose action provoked the next step in the causality chain is treated in the same way as those involved in the next step. We formalize these principles, and we show that there is only one allocation scheme that satisfies them all.

While the axiomatic approach relies solely on normative principles, there is a second factor that should be taken int account when an allocation scheme is chosen, namely, its ability to prevent accidents. When deciding how much care people should take to prevent accidents, a government faces a trade-off between the (expected) cost of an accident and the cost of reducing the probability that it occurs. We shall assume that there is a unique optimal level of care, and we refer to it as the *standard of care*. Agents have to be incentivized to act in accordance with the standard of care cf. (Landes and Posner, 1980). We define a game where each agent can choose how much to invest into the prevention of accidents and show that every allocation scheme that satisfies two very simple axioms incentivizes agents to invest an efficient amount.

In a liability situation with multiple tortfeasors some of them may settle with the plaintiff. Thus, a third factor that is relevant for the design of an allocation scheme is its effect on settlements. Both under apportioned set-off and pro tanto the payments of the remaining agents have to be adjusted—ideally in a way that incentivizes all agents to settle early. For the allocation scheme we provide there is a unique key according to which the settlement amount should be distributed as to affect the claims against the remaining defendants in the desired way.

Throughout the paper we will use the following numerical example to illustrate our points. Two drivers have an accident and a pedestrian who stands nearby is injured, implying a damage of $100,000. A physician negligently treats the pedestrian causing an incremental damage of $900,000. The three tortfeasors have to compensate for an overall damage of $1,000,000. The remainder of the article is structured as follows. In Sect. 2 we focus on the normative approach and develop the unique allocation scheme with all desired properties. In Sect. 3 we provide a minimal requirement of the allocation scheme so as to achieve the efficient prevention of accidents. The proper adjustment of damages after settlements is discussed in Sect. 4. Section 5 closes the article with a brief discussion. The mathematics are postponed to the appendix.

2 Obtaining the Compensation Payments

Throughout the paper we consider liability situations where groups of agents (tortfeasors) successively cause (incremental) damages to a plaintiff, and where the damage caused by any but the first group is only possible because of the behavior of the previous groups. We say that a tortfeasor *directly* causes an incremental damage if he belongs to the group that caused that damage, and we say he *indirectly* causes a damage if the incremental damage was caused in the sequel of his group's action. The plaintiff is not a member of any of these groups, and each other agent belongs to exactly one group, i.e., we do not consider situations where an agent contributes through multiple own acts.

An *allocation scheme* is a rule that specifies, for *any* liability situation, for how much of the overall damage each tortfeasors must compensate the plaintiff. The literature on tortfeasors suggests several principles such an allocation scheme should follow. We shall precisely formulate them here as axioms and investigate what allocation scheme satisfies them all simultaneously.

In order to be compensated for any damage by a defendant, the plaintiff has to prove that the damage lies in the scope of liability of the defendant, and that there is an applicable rule of liability under which the defendant has to compensate the plaintiff (Shavell, 1980). We shall not discuss the scope of liability in this article as such a discussion would distract from the point we wish to make. The only thing that needs to be decided is whether or not there is an applicable rule of liability. For the arguments of this section it is not necessary to specify this rule; for simplicity assume that the negligence rule is in place. We will have a closer look at the "optimal" rule of liability in Sect. 3. The first two axioms an allocation scheme should satisfy are easily formulated and do not need much discussion.

Axiom 1. A defendant's compensation payment is strictly positive if and only if he negligently, either directly or indirectly, caused a strictly positive damage.

Axiom 2. The sum of the compensation payments covers exactly all damages that are caused negligently, either directly or indirectly.

To keep things simple we shall assume that there is no discrimination between different degrees of negligence; either a tortfeasor was negligent or not. In particular, there should be no discrimination between negligent tortfeasors who caused the same incremental damages.

Axiom 3. The compensation payments of two defendants are equal whenever (i) both are negligent, and (ii) they entered the scene together.

In liability situations with only one tort, the axioms we have formulated so far are sufficient to pin down a unique allocation scheme.

Theorem 1 *There is a unique allocation scheme that satisfies Axioms 1–3 on all liability situations that are caused by simultaneous torts. It distributes the full compensation payment equally among the negligent agents.*

This result is hardly surprising, and the far more challenging task is the generalization of this allocation scheme to situations in which not all tortfeasors acted simultaneously. A first step is the following "uncontroversial principle" (Landes and Posner, 1980) in the law of tort.[1]

Axiom 4. A defendants compensation payment does not depend on the damages that precede the damage he has caused.

In some situations a tortfeasor might actually not have caused a *direct* damage, but only opened the door for future damages: suppose that in the initial example the pedestrian was not injured in the accident, but during the check-up by a paramedic who appeared at the scene. In this case the car drivers would still be (partly) responsible for this injury. However, as their acts would no longer create a separate damage, they can be treated as if they had caused the injury together with the paramedic. The following axiom captures this reasoning.

Axiom 5. Any agent who did not cause a damage but made the subsequent damages possible is treated as if he belonged to the first subsequent group that actually caused a damage.

These five axioms are, in fact, sufficient to uniquely specify an allocation scheme.[2]

Theorem 2 *There exists an allocation scheme that satisfies Axioms 1–5, and this allocation scheme is unique. It proceeds as follows. All damages that are directly or indirectly caused by at least one negligent agent will be covered. Each incremental damage is equally distributed between all negligent agents that are directly or indirectly responsible for it.*

When we apply this allocation scheme to our initial example, we see that the damage of \$100,000 has to be equally divided between the two drivers, and the incremental damage of \$900,000 has to be divided between all three tortfeasors. Hence, the compensation payments will be \$350,000 for each driver and \$300,000 for the physician.

Interestingly the axioms in the foregoing theorem do not only uniquely define an allocation scheme; each of them is needed to guarantee uniqueness: whenever one of the axioms is left out, there are several allocation schemes that satisfy those remaining.

[1] Note that Axiom 4 is a implied by Young's Strong Monotonicity for the respective TU game; an axiom that follows from Shapley's Additivity axiom would be the following. Two cases can be dealt with either separately or at once—the outcome should not depend on this.

[2] A similar result for the less general case in which all players are negligent was independently discovered by (Dehez and Ferey, 2016).

The allocation scheme we present here is related to a well-known concept from game theory, namely the Shapley value (Shapley, 1953) of an appropriately defined cooperative game (Dehez and Ferey, 2013). Bearing this in mind there are several generalizations one can think of. First, the plaintiff might be (partly) responsible for (some of) the damages. In this case one can obtain a similar result if one includes him as a tortfeasor in the liability situation. His net compensation payment is then the amount of all (negligently caused) damages reduced by the payment the allocation scheme allocates to him. Second, there might be discrimination between tortfeasors depending on their degree of negligence. In this case Axiom 3 must necessarily be dropped. However, if one assumes the degree of negligence is independent of the monetary value of a damage, one obtains (together with the other axioms) a family of allocation rules that only depend on the degrees of negligence of all tortfeasors. Such an allocation scheme would correspond to the weighted Shapley value (Shapley, 1953b; Kalai and Samet, 1987).

3 Efficiency and Deterrence

Suppose that taking care, i.e., avoiding accidents, is costly to agents: taking more care is more expensive and results in a lower probability of an accident. Agents then face a trade-off between saving the cost of taking care and reducing the risk of being involved in an accident. We assume throughout that agents are rational and risk neutral, so that they minimize their *private* (expected) cost. The *social* cost is given by the aggregated private cost of care plus the *expected* cost of an accident, which is the probability of an accident multiplied by the damage the accident would cause. It shall be assumed that there is a unique level of care that minimizes the social cost: in this case the additional cost of any further damage prevention would outweigh the additional reduction of the expected cost of an accident, and the savings from taking less care would be outweighed by the increase in the expected cost of an accident. In order to minimize social cost government must incentivize agents to apply this care level; henceforth, we shall refer to it as the *standard of care*. On the other hand, the agents' objective is to choose their care level so as to minimize their expected private cost. This poses a free-rider problem as each agent may rely on the other agents' measures to prevent accidents. In particular, it is not at all clear that the social and private interests are aligned, see for instance (Shavell, 1980; Landes and Posner, 1980; Kornhauser and Revesz, 2000). One way for the government to reconcile these interests is to combine Axioms 1 and 2 with an appropriate liability rule.

Axiom 1*. An agent has to pay a positive compensation payment if and only if (i) he caused a positive damage and (ii) his care level was lower than the standard of care.

Axiom 2*. The sum of the compensation payments covers exactly all damages that have been directly or indirectly caused by at least one agent with a care level below the standard of care.

In contrast to Axioms 1 and 2, the Axioms 1* and 2* now refer to the socially optimal standard of care. In this sense, they build a specification of the former, which allowed for other interpretations of negligence. The implied liability rule, together with the full recovery principle, already has a very strong implication. To formulate it, we need the concept of *Nash equilibrium* (Nash, 1950): a Nash equilibrium is a list of choices, one choice for each agent, such that no agent has an incentive to change their choice, given the choices of the others. Hence, a Nash equilibrium is a situation in which the agent's choices are self-sufficient; whenever the society is not in a Nash equilibrium, there is at least one agent who could unilaterally improve by choosing differently (provided the others stick to their choices).

Theorem 3 *In a society where the implemented allocation scheme satisfies Axioms 1* and 2*, there is only one Nash Equilibrium. In this Nash equilibrium, each player will choose their care level equal to the standard of care.*

It is not surprising that choosing the standard of care constitutes a Nash equilibrium – perhaps more surprising is the fact that this Nash equilibrium is unique: whenever some potential tortfeasors do not behave appropriately, at least one of them would benefit by choosing the standard of care.

A direct consequence of this theorem is that the replacement of Axioms 1 and 2 in Theorems 1 and 2 by Axioms 1* and 2* leads to the characterizations of (unique) allocation schemes that create the desired incentives.

4 Settlements

In this section we shall take the allocation scheme that has been described in Sect. 2 as given. When one of the defendants settles, the liability situation is adjusted: the settling agent is removed and the remaining agents have to compensate for the original damage either net the settlement amount (pro tanto rule) or net the amount the allocation scheme would attribute to the settling agent (apportioned set-off rule). In the case of simultaneous tortfeasors this reduction is easily calculated, as the remaining damage is equally split between all agents who did not settle. If, however, there have been several subsequent torts, it is not obvious what part of the settlement amount should be used in order to reduce each incremental damage. A *settlement amendment scheme* is a list of weights, one for each incremental damage, that add up to 100%. It serves as the key according to which any settlement amount is distributed between all the damages of the causality chain. (In particular, it is independent of the settlement amount.) So, any settlement leads to a new liability situation where the settling agent is removed from his groups, and the damage of each group is adjusted by that percentage of the settlement amount that is specified by the settlement amendment scheme.

Under the apportioned set-off rule the payments of non-settling agents should not be affected by someone else's settlement; otherwise there would be at least one tortfeasor who benefits from somebody else's early settlement. Under the pro tanto rule, one would expect the new payments of all non-settling agents to be higher whenever the settlement is lower than what is originally specified by the allocation rule. If a settlement amendment scheme achieves the latter, it *promotes settlements*. The emphasis here lies on *all* non-settling agents: if the settlement amount is lower than the allocation payment, it is clear that under the pro tanto there is at least *one* agent who will have to pay more. Requiring that this holds for *all* agents, however, pins down a unique settlement amendment scheme.

Theorem 4 *There is a unique settlement amendment scheme that promotes settlements. According to this scheme each incremental damage is adjusted as follows:*

1. *if the settling agent was neither directly nor indirectly responsible for the incremental damage, it is not adjusted,*
2. *otherwise it is divided by the number of negligent agents who are directly or indirectly responsible for it (including the settling agent),*
3. *this per capita incremental damage is multiplied by the ratio between the settling agent's settlement amount and the amount he would have to pay according to the allocation scheme,*
4. *the damage is reduced by this amount.*

We shall apply this settlement amendment scheme to our initial example in the case that one of the drivers settles at an amount of \$210.000. (A general formula for the weights in this settlement amendment scheme is provided in the appendix.) The driver is (directly or indirectly) responsible for all incremental damages, so step 1 does not apply. The driver is involved in two incremental damages, namely \$100.000 (with one other tortfeasor) and \$900.000 (with two other tortfeasors.) Thus, the per capita damages in step 2 are \$50,000 and \$300,000. The driver settles at \$210.000 which is 60% of the amount specified by the allocation scheme, which was \$350.000. So, following step 3, we find $60\% \times \$50,000 = \$30,000$ and $60\% \times \$300,000 = \$180,000$. In step 4, the incremental damages are reduced by these amounts. That is, after the settlement, the first incremental damage is reduced to $\$100.00 - \$30,000 = \$70,000$ and the second one to $\$900.000 - \$180,000 = \$720,000$. Thus, the remaining agents face a new liability situation with incremental damages \$70,000 and \$720,000; the overall amount left to compensate for is \$790.000. Applying the original allocation scheme described in Sect. 2 to this new situation leads to payments of \$430,000 for the non-settling driver and \$360,000 for the physician.

Interestingly, this is also the only settlement amendment scheme that ensures that under the apportioned set-off rule the payments of non-settling agents in the original and new liability situations are equal.

5 Conclusion

We have shown that the rigid use of reasonable and commonly accepted principles in the law of tort uniquely determines a scheme according to which compensation payments should be allocated to several tortfeasors. Even under the joint and several liability rule (where such an allocation scheme is not explicitly used), these payments can be used as an anchor. In particular, together with a unique settlement amendment scheme, they promote early settlements. The allocation scheme we provided also incentivizes agents to apply the efficient standard of care, so that a socially efficient outcome can be achieved. Indeed, every allocation scheme that satisfies consistency with a rule of liability (which is defined by a standard of care) and ensures the recovery of all damages that have negligently been caused, will lead to an efficient outcome in the unique Nash equilibrium. Further solution concepts from cooperative game theory, in particular the nucleolus are studied by (Dehez and Ferey, 2016). An obvious desideratum for future research is to provide a characterization for the nucleolus. Another question would be to investigate, e.g., egalitarian Shapley values, according to which a given percentage of the damage is equally split whereas the remainder is split according to the Shapley value.[3] At first, this might appear surprising; however, it is common to have all involved parties suffer to some extend (e.g., to spend time and money on legal advice). Further, it might set incentives to all parties to abstain from hidden actions so to prevent the creation of damage.

Acknowledgements We would like to thank Lewis A. Kornhauser and Tse-Ling Teh for their useful comments and insightful suggestions.

Mathematical Appendix

Proofs of Theorems 1 and 2

A *liability situation* is a triple $(N, \mathbf{S}, \Delta)$ where N is the (finite) set of tortfeasors, $\mathbf{S} = (S_k)_{k=1}^m$ is a vector of coalitions $S_k \subseteq N$ with $S_k \cap S_l = \emptyset$ for all $k \neq l$, and $\bigcup_{k=1}^m S_k = N$, and $\Delta = (\Delta_k)_{k=1}^m \in \mathbb{R}^m$ is a vector of non-negative damages. An *allocation scheme* is a map f that maps any liability situation $(N, \mathbf{S}, \Delta)$ on a vector $(f_i\,(N, \mathbf{S}, \Delta))_{i \in N} \in \mathbb{R}^N_{\geq 0}$ of *compensation payments*. For a liability situation $(N, \mathbf{S}, \Delta)$ denote by S_k^* the members of S_k who were negligent, let $N^* = \bigcup_{k=1}^m S_k^*$, and let

$$\Delta_k^* = \begin{cases} \Delta_k & \text{if } \bigcup_{l \leq k} S_l^* \neq \emptyset, \\ 0 & \text{otherwise} \end{cases}$$

be the damages that are caused (directly or indirectly) by at least one negligent agent. The mathematical formulation of the axioms 1–5 is as follows.

[3] A characterization similar to the one presented in Theorem 2 could be developed based on (Casajus and Huettner, 2014).

Axiom 1. For any liability situation $(N, \mathbf{S}, \Delta)$ and any $i \in N$ it holds that $f_i(N, \mathbf{S}, \Delta) > 0$ if and only if there is k such that $i \in S_k^*$ and $\sum_{l \geq k} \Delta_l > 0$.
Axiom 2. For any liability situation $(N, \mathbf{S}, \Delta)$ it holds that $\sum_{i \in N} f_i (N, \mathbf{S}, \Delta) = \sum_{k=1}^{m} \Delta_k^*$.
Axiom 3. For any liability situation $(N, \mathbf{S}, \Delta)$ and any two agents $i, j \in N$ with $i, j \in S_k^*$ for some $k = 1, \dots, m$ it holds that $f_i (N, \mathbf{S}, \Delta) = f_j (N, \mathbf{S}, \Delta)$.
Axiom 4. If $(N, \mathbf{S}, \Delta)$ and $\left(N, \mathbf{S}, \Delta'\right)$ are such that there is $k \leq m$ with $\Delta_h' = \Delta_h$ for all $h \geq k$, then $f_i (N, \mathbf{S}, \Delta) = f_i \left(N, \mathbf{S}, \Delta'\right)$ for all $i \in \bigcup_{l=k}^{m} S_l$.
Axiom 5. For all liability situations $(N, \mathbf{S}, \Delta)$ with $\Delta_k = 0$ for some k it holds that

$$f (N, \mathbf{S}, \Delta) = f (N, \mathbf{S}_{-k}, \Delta_{-k}),$$

where

$$\begin{aligned} S_{-k} &= (S_1, \dots, S_{k-1}, S_k \cup S_{k+1}, \dots, S_m) \\ \Delta_{-k} &= (\Delta_1, \dots, \Delta_{k-1}, \Delta_{k+1}, \dots, \Delta_m). \end{aligned}$$

Any liability situation $(N, \mathbf{S}, \Delta)$ can be naturally associated with a characteristic function form game $\left(N, v^{N,\mathbf{S},\Delta}\right)$ by setting

$$v^{N,\mathbf{S},\Delta} (T) = \sum_{k: \bigcup_{l=1}^{k} S_l^* \subseteq T} \Delta_k^*.$$

for all $T \subseteq N$ (Dehez and Ferey, 2013). In particular, players who are not negligent are null players in this game.

Proof of Theorem 1.

The proposed allocation scheme f is given by

$$f_i (N, \mathbf{S}, \Delta) = \begin{cases} \frac{\Delta_1^*}{|N^*|} & \text{if } i \in N^*, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, f satisfies Axioms 1–3. Suppose there is another allocation scheme g that satisfies all three axioms as well. By Axiom 1 $g_i (N, \mathbf{S}, \Delta) = 0 = f_i (N, \mathbf{S}, \Delta)$ for all $i \notin N^*$. If there is at least one $i \in N^*$ then, by Axiom 2, $\sum_{i \in N^*} g_i (N, \mathbf{S}, \Delta) = \Delta_1^* = \Delta_1$. Hence, by Axiom 3, $g_i (N, \mathbf{S}, \Delta) = \frac{\Delta_1}{|N^*|} = f_i (N, \mathbf{S}, \Delta)$. Hence, f and g coincide. *Q.E.D.*

Note that the Axioms 1–3 are independent. The allocation scheme that splits the damage equally among all players satisfies Axioms 1 and 3, but not 2. The allocation scheme that assigns 50% of the damage according to our allocation scheme and leaves the remaining damage undistributed satisfies Axioms 1 and 3, but not 2. The allocation scheme that assigns the weighed Shapley value to each player satisfies Axioms 1 and 3, but not 2.

Proof of Theorem 2.

We have to prove that the allocation scheme f defined by

$$f_i(N, \mathbf{S}, \Delta) = \sum_{k: i \in \bigcup_{l=1}^{k} S_k^*} \frac{\Delta_k^*}{\left|\bigcup_{l=1}^{k} S_k^*\right|}.$$

for all $i \in N$ is the only allocation scheme that satisfies axioms 1–5. It can easily be seen that

$$f(N, \mathbf{S}, \Delta) = Sh\left(N, v^{N,\mathbf{S},\Delta}(S)\right), \tag{1}$$

where Sh is the *Shapley value* (Shapley, 1953). Hence, Axiom 1–3 follow from the null player property, efficiency, and symmetry of the Shapley value. Axiom 5 holds as the two liability situations $(N, \mathbf{S}, \Delta)$ and $(N, \mathbf{S}_{-k}, \Delta_{-k})$ are associated with the same characteristic function form game. Axiom 4 is satisfied because of the strong monotonicity of the Shapley value (Young, 1985) as $v^{N,\mathbf{S},\Delta}(S) - v^{N,\mathbf{S},\Delta}(S \setminus \{i\}) = v^{N,\mathbf{S},\Delta'}(S) - v^{N,\mathbf{S},\Delta'}(S \setminus \{i\})$ for all such i.

For the uniqueness of f suppose that there is another allocation scheme g that satisfies the axioms as well. For any liability situation $(N, \mathbf{S}, \Delta)$ let $I(\Delta) = |\{k : \Delta_k > 0\}|$. If $I(\Delta) = 0$, then there are no positive damages, and Axiom 1 implies that $g_i(N, \mathbf{S}, \Delta) = 0 = f_i(N, \mathbf{S}, \Delta)$ for all $i \in N$. Let $(N, \mathbf{S}, \Delta)$ be such that $I(\Delta) \geq 1$, and that the claim is true for all liability situations $\left(N, \mathbf{S}', \Delta'\right)$ with $I\left(\Delta'\right) < I(\Delta)$. By Axiom 5 we can assume without loss of generality that $\Delta_1 > 0$. Define $\left(N, \mathbf{S}, \Delta'\right)$ by $\Delta_1' = 0$ and $\Delta_k' = \Delta_k$ for all $k \geq 2$. Then

$$g_i(N, \mathbf{S}, \Delta) = g_i\left(N, \mathbf{S}, \Delta'\right) = f_i\left(N, \mathbf{S}, \Delta'\right) = f_i(N, \mathbf{S}, \Delta)$$

for all $i \in \bigcup_{l=2}^{m} S_l$ by Axiom 4 and the induction hypothesis. Hence, $g_i(N, \mathbf{S}, \Delta) = f_i(N, \mathbf{S}, \Delta)$ for all $i \in S_1$ by Axioms 2, and 3. *Q.E.D.*

For the independence of the axioms note that the allocation scheme $f_i^1(N, \mathbf{S}, \Delta) = \sum_{k: i \in \bigcup_{l=1}^{k} S_l} \frac{\Delta_k^*}{\left|\bigcup_{l=1}^{k} S_l\right|}$ satisfies all axioms but Axiom 1. Further, $f_i^2(N, \mathbf{S}, \Delta) = \frac{1}{2} f_i(N, \mathbf{S}, \Delta)$ satisfies all axioms but Axiom 2. If the Shapley value in Eq. (1) is replaced by a *weighted* Shapley value (Shapley, 1953b; Kalai and Samet, 1987) one obtains an allocation scheme that satisfies all axioms but Axiom 3. The allocation scheme

$$f_i^4(N, \mathbf{S}, \Delta) = \begin{cases} \frac{1}{\left|\bigcup_{k=1}^{m} S_k^*\right|} \sum_{k=1}^{m} \Delta_k^*, & \text{if } i \in S_k^* \text{ for some } k = 1, \ldots, m \\ 0, & \text{otherwise} \end{cases}$$

satisfies all axioms but Axiom 4.

The allocation scheme

$$f_i^5(N, \mathbf{S}, \Delta) = \begin{cases} \sum_{k: i \in S_k^*} \frac{\Delta_k^*}{|S_k^*|}, & \text{if } i \in S_k^* \text{ for some } k = 1, \ldots, m \\ 0, & \text{otherwise} \end{cases}$$

satisfies all axioms but Axiom 5.

Proof of Theorem 3

Let $(N, \mathbf{S}, \Delta)$ be a (fixed) liability situation. Let $x_i \in \mathbb{R}$ be agent i's level of care and denote by $C_i(x_i)$ the associated private cost of agent i, where C_i is an increasing function. By $p_k\left((x_i)_{i \in S_k}\right)$ denote the probability that group S_k causes damage Δ_k provided that each $i \in S_k$ chooses x_i as his level of care; let p_k be decreasing in all coordinates. Then the expected social costs of the liability situation are given by

$$SC(x) = \sum_{i \in N} C_i(x_i) + \sum_{k=1}^{m} p_k\left((x_i)_{i \in S_k}\right) \Delta_k. \tag{2}$$

Assume that SC has a unique minimum,[4] and denote the minimizer of SC by $x^* \in \mathbb{R}^N$. We interpret x_i^* as the standard of care that applies to agent i; in particular, different standards of care may apply to different agents in the liability situation. The adapted axioms then read as follows.

Axiom 1*. For any liability situation $(N, \mathbf{S}, \Delta)$ and any $i \in N$ it holds that $f_i(N, \mathbf{S}, \Delta) > 0$ if and only if $x_i < x_i^*$ and $\sum_{k: i \in \bigcup_{l=1}^{k} S_k} \Delta_k > 0$.

Axiom 2*. For any liability situation $(N, \mathbf{S}, \Delta)$ it holds that $\sum_{i \in N} f_i(N, \mathbf{S}, \Delta) = \sum_{k=1}^{m} \Delta_k^*$, where

$$\Delta_k^* = \begin{cases} \Delta_k & \text{if there is } i \in \bigcup_{l \leq k} S_l \text{ with } x_i < x_i^*, \\ 0 & \text{otherwise.} \end{cases}$$

Proof of Theorem 3.

We first show that x^* is a Nash equilibrium. It is obvious that no agent has an incentive to choose $x_i > x_i^*$. Suppose that all agents $j \in N \setminus \{i\}$ chose $x_j = x_j^*$ and assume that $x_i < x_i^*$. Then agent i's expected payment is

$$C_i(x_i) + \sum_{k: i \in \bigcup_{l=1}^{k} S_l} p_k\left(\left(x_j\right)_{j \in S_k}\right) \Delta_k.$$

[4] Existence and uniqueness can be guaranteed for instance if C_i is convex for all i and p_k is strictly convex.

as i has to recover the full damage alone. As x^* is the unique minimizer of the social cost function in Eq. (2) it holds that

$$\begin{aligned} SC\left(x^*\right) &= \sum_{j\in N} C_j\left(x_j^*\right) + \sum_{k=1}^{m} p_k\left(\left(x_j^*\right)_{j\in S_k}\right)\Delta_k \\ &< \sum_{j\neq i} C_j\left(x_j^*\right) + C_i\left(x_i\right) + \sum_{k:i\notin\bigcup_{l=1}^{k} S_l} p_k\left(\left(x_j^*\right)_{j\in S_k}\right)\Delta_k \\ &\quad + \sum_{k:i\in\bigcup_{l=1}^{k} S_l} p_k\left(x_i, \left(x_j^*\right)_{j\in S_k\setminus\{i\}}\right)\Delta_k \end{aligned}$$

and therefore

$$\begin{aligned} C_i\left(x_i\right) + \sum_{k:i\in\bigcup_{l=1}^{k} S_l} p_k\left(x_i, \left(x_j^*\right)_{j\in S_k\setminus\{i\}}\right)\Delta_k &> C_i\left(x_i^*\right) + \sum_{k:i\in\bigcup_{l=1}^{k} S_l} p_k\left(\left(x_j^*\right)_{j\in S_k}\right)\Delta_k \\ &\geq C_i\left(x_i^*\right). \end{aligned}$$

Hence, x_i imposes higher expected costs on i than x_i^*, so x^* is a Nash Equilibrium.

We now show that x^* is, in fact, the only Nash Equilibrium. For this purpose let x be vector of care levels, and assume that x is a Nash Equilibrium. Let A be the set of agents who choose $x_j = x_j^*$, and let $B = N \setminus A$ the set of agents who choose $x_i < x_i^*$. (Recall that no agent would choose $x_j > x_j^*$ in a Nash Equilibrium.) Then, by Axioms 1* and 2*, the total expected costs that the agents in B have to bear are

$$\sum_{i\in B} C_i\left(x_i\right) + \sum_{k:\bigcup_{l=1}^{k} S_l\cap B\neq\emptyset} p_k\left(\left(x_j^*\right)_{j\in A\cap S_k}, \left(x_i\right)_{i\in B\cap S_k}\right)\Delta_k.$$

(Recall that $\Delta_k = \Delta_k^*$ for all k with $\bigcup_{l=1}^{k} S_l \cap B \neq \emptyset$.) By the definition of x^* we have $SC\left(x^*\right) < SC\left(x\right)$ and therefore,

$$\begin{aligned} &\sum_{i\in B} C_i\left(x_i\right) + \sum_{k:\bigcup_{l=1}^{k} S_l\cap B\neq\emptyset} p_k\left(\left(x_j^*\right)_{j\in A\cap S_k}, \left(x_i\right)_{i\in B\cap S_k}\right)\Delta_k \\ &\quad > \sum_{i\in B} C_i\left(x_i^*\right) + \sum_{k:\bigcup_{l=1}^{k} S_l\cap B\neq\emptyset} p_k\left(\left(x_i^*\right)_{i\in S_k}\right)\Delta_k \\ &\quad \geq \sum_{i\in B} C_i\left(x_i^*\right). \end{aligned}$$

Since the aggregated expected costs of the agents in B are strictly greater if they choose $\left(x_i^*\right)_{i\in B}$ than if they choose $\left(x_j^*\right)_{j\in B}$, there must be at least one agent $i \in B$ whose private expected costs are strictly greater from choosing x_i than from choosing x_i^*. Hence, x cannot be a Nash Equilibrium. *Q.E.D.*

Proof of Theorem 4

A settlement amendment scheme is a vector $(r_k)_{k=1}^m$ with $r_k \geq 0$ for all $k = 1, \ldots, m$ and $\sum_{k=1}^m r_k = 1$. If agent i settles and pays an amount θ, the remaining agents face the new liability situation $\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right)$ with

$$S'_k = S_k \setminus \{i\}$$
$$\Delta'_k = \Delta_k - r_k\theta.$$

The settlement amendment scheme r promotes settlements if

$$f_j\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right) > f_j(N, \mathbf{S}, \Delta)$$

for all $j \neq i$ if and only if $\theta < f_i(N, \mathbf{S}, \Delta)$.

Proof of Theorem 4.

The proposed settlement amendment scheme is given by

$$r_k = \begin{cases} \frac{\Delta_k^*}{\left|\bigcup_{l=1}^k S_l^*\right|} \frac{1}{f_i(N, \mathbf{S}, \Delta)} & \text{if } i \in \bigcup_{l=1}^k S'_l \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that r has the desired property; in fact, this follows from the consistency of the Shapley value (Hart and Mas-Colell, 1989).

Let r promote settlements and let i be an agent with $f_i(N, \mathbf{S}, \Delta) > 0$ who settles at θ. Then $f_j\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right) \leq f_j(N, \mathbf{S}, \Delta)$ for all $j \neq i$ if and only if $\theta \geq f_i(N, \mathbf{S}, \Delta)$. In case that $\theta = f_i(N, \mathbf{S}, \Delta)$ one further obtains

$$\sum_{j \neq i} f_j\left(N, \mathbf{S}', \Delta'\right) = \sum_{k=1}^m \Delta_k^* - \theta = \sum_{k=1}^m \Delta_k^* - f_i(N, \mathbf{S}, \Delta) = \sum_{j \neq i} f_j(N, \mathbf{S}, \Delta).$$

Hence, in this case it must hold that $f_j\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right) = f_j(N, \mathbf{S}, \Delta)$ for all $j \neq i$. Using the definitions of f and $\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right)$ this leads to

$$\begin{aligned} \sum_{k: j \in \bigcup_{l=1}^k S_l} \frac{\Delta_k^*}{\left|\bigcup_{l=1}^k S_l\right|} &= f_j(N, \mathbf{S}, \Delta) = f_j\left(N \setminus \{i\}, \mathbf{S}', \Delta'\right) \\ &= \sum_{k: j \in \bigcup_{l=1}^k S_l} \frac{\Delta_k^* - r_k f_i(N, \mathbf{S}, \Delta)}{\left|\bigcup_{l=1}^k S_l \setminus \{i\}\right|} \end{aligned}$$

or equivalently

$$\sum_{k:j\in\bigcup_{l=1}^{k} S_l} \frac{r_k f_i\,(N, \mathbf{S}, \Delta)}{\left|\bigcup_{l=1}^{k} S_l \setminus \{i\}\right|} = \sum_{k:j\in\bigcup_{l=1}^{k} S_l} \left(\frac{\Delta_k^*}{\left|\bigcup_{l=1}^{k} S_l \setminus \{i\}\right|} - \frac{\Delta_k^*}{\left|\bigcup_{l=1}^{k} S_l\right|} \right) \tag{3}$$

for all $j \neq i$. If $S_1 \neq \{i\}$ these are m linear equations that are linearly independent (as the system is triangular and all diagonal entries are strictly positive), so the solution is unique. If $S_1 = \{i\}$ the requirement that $\sum_{k=1}^{m} r_k = 1$ is another constraint that is linearly independent of the $m - 1$ independent (non-trivial) equations in (3). Hence, in both cases the solution is unique. As the proposed settlement amendment scheme r solves this linear equation system, it is the only solution. *Q.E.D.*

References

Casajus, A., & Huettner, F. (2014). Weakly monotonic solutions for cooperative games. *Journal of Economic Theory, 154*, 162–172.

Dehez, P., & Ferey, S. (2013). How to share joint liability: A cooperative game approach. *Mathematical Social Sciences, 66*, 44–50.

Dehez, P., & Ferey, S. (2016). Multiple causation, apportionment and the Shapley value. *Journal of Legal Studies, 45*, 143–171.

Hart, S., & Mas-Colell, A. (1989). Potential, value and consistency. *Econometrica, 57*, 589–614.

Kalai, E., & Samet, D. (1987). On weighted Shapley values. *International Journal of Game Theory, 16*, 205–222.

Kornhauser, L., & Revesz, R. (2000). Joint tortfeasors. In B. Bouckaert & G. Geest (Eds.), *Encyclopedia of law and economics volume II: Civil law and economics* (pp. 625–643). Edward Elgar Publishing.

Landes, W., & Posner, R. (1980). Joint and multiple tortfeasors: An economic analysis. *The Journal of Legal Studies, 9*, 517–555.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*, 48–49.

Shapley, L. S. (1953). A value for n-person games. In H. Kuhn & A. Tucker (Eds.), *Contributions to the theory of games II* (pp. 307–317). Princeton University Press.

Shapley, L. S. (1953b). Additive and non-additive set functions. Ph.D. Thesis, Princeton University

Shavell, S. (1980). An analysis of causation and the scope of liability in the law of torts. *The Journal of Legal Studies, 9*, 463–516.

Wall, D. (Ed.) (1986). Contribution and claim reduction in antitrust litigation. Monograph No. 11, American Bar Association Antitrust Section

Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory, 14*, 65–72.

Solidarity and Fair Taxation in TU Games

André Casajus

Abstract We consider an analytic formulation of the class of efficient, linear, and symmetric values for TU games that, in contrast to previous approaches, which rely on the standard basis, rests on the linear representation of TU games by unanimity games. Unlike most of the other formulae for this class, our formula allows for an economic interpretation in terms of taxing the Shapley payoffs of unanimity games. We identify those parameters for which the values behave economically sound, i.e., for which the values satisfy desirability and positivity. Put differently, we indicate requirements on fair taxation in TU games by which solidarity among players is expressed.

1 Introduction

Consider the Barbie family consisting of mother Barbie, father Ken, and their little daughter Aqua,[1] which is of an age where she neither generates income nor affects the generation of income by her parents. If the family income would be spent on its members well-being according to their contributions to the former, then poor Aqua would end up very miserable. In reality, some sort of "solidarity" is expressed among the members of a family—part of the family income goes to "unproductive" members like Aqua.

Situations like those above can be modeled by TU games. The Shapley value (Shapley, 1953) probably is the most eminent point-solution concept for TU games. Its standard characterization involves four axioms: efficiency, additivity/linearity, symmetry, and the null player axiom. In a sense, it is mainly the latter property that prevents the Shapley value to allow for solidarity among the players. Irrespective of

[1] Note that Aqua also is the name of a Danish-Norwegian Eurodance music group, best known for their 1997 multi-platinum single "Barbie Girl".

A. Casajus (✉)
Leipzig Graduate School of Management, Leipzig, Germany
e-mail: mail@casajus.de

Dr. Hops Craft Beer Bar, Leipzig, Germany

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_13

the productivity within the whole society, the null player axiom requires unproductive players to obtain a zero payoff. Moreover, together with additivity, the null player property already entails marginality (Young, 1985); i.e., the players' payoffs depend only on their *own* productivities measured by marginal contributions, respectively.

So, if one wishes values to allow for solidarity considerations, one has to drop the null player axiom from the list of required properties. But then we are down to the class of values obeying efficiency, linearity, and symmetry. Obviously, this large class contains a lot of values that deviate from the Shapley value not only by economically sound solidarity considerations. For example, the equal surplus division value (Driessen & Funaki, 1991) and the consensus value (Ju et al., 2007) are members of this class, but fail positivity (Kalai & Samet, 1987). That is, these values may assign negative payoffs in monotonic games, i.e., in games where no player ever is destructive in terms of her marginal contributions. We feel that this does not fit well our intuitions on solidarity. Moreover, values in this class may not meet desirability (Maschler & Peleg, 1966); i.e., a player who is more productive than another one may end up with a lower payoff. Again, this would overstretch our sense of solidarity.

Formulae for the class of efficient, linear, and symmetric values (ELS values) have been proposed by Ruiz et al. (1998), Driessen and Radzik (2003), Chameni Nembua and Andjiga (2008), and Hernandez-Lamoneda et al. (2008). Recently, Chameni Nembua (2012) and Malawski (2013) came up with a more interpretational one. In essence, the players' marginal contributions within a coalition are taxed at a rate depending on its size, while the tax revenue is distributed evenly among the other players in the coalition under consideration.

We suggest and explore an alternative formula for this class, already indicated by Radzik and Driessen (2009, p. 5), which also is interpretable in terms of taxation. The main idea of our approach is to tax and redistribute the Shapley payoffs of unanimity games. First, the Shapley payoffs are taxed at a certain rate, which depends on the cardinality of the set of productive players in such a game. And second, the overall tax revenue is distributed equally among *all* players. Linearity extends these payoffs to general TU games.

Radzik and Driessen (2013) provide conditions on the parameters of the formula due to Driessen and Radzik (2003) such that the resulting value satisfies desirability or both positivity and desirability, respectively. Recently, (Radzik, 2021, Theorem 6) provides conditions on the parameters of this formula such that the resulting value satisfies positivity.

In this paper, we attempt analogous conditions on the parameters of our formula. First, we prepare our main results by showing the relation between the parameters of our formula and the parameters of the Driessen–Radzik formula (Propositions 2 and 3). Then, we identify those parameters that entail desirability (Theorem 1). Since there seems to be no nice way to describe the parameters that yield positivity, we first identify those parameters that entail positivity for null players both for the Driessen–Radzik formula (Theorem 2) and for our formula (Theorem 3). Only then, we describe the parameters that yield positivity (Remark 4). Combining the afore-mentioned results, we obtain requirements on the parameters that imply both desirability and positivity (Theorem 4). Finally, we identify the parameters that guarantee the acceptability properties suggested by Joosten et al. (1994) and by Radzik and Driessen (2013) (Propositions 4 and 5).

This paper is organized as follows: In the second section, we introduce basic definitions and notation. The third section surveys formulae for ELS values and introduces a new parametrization for this class. In section four, we provide conditions on the parameters of our formulae such that one or another of the desirable properties mentioned above are satisfied. The appendix contains the lengthier proofs.

2 Basic Definitions and Notation

A **(TU) game** is a pair (N, v) consisting of a non-empty and finite set of players N and a **coalition function** $v : 2^N \to \mathbb{R}$ with $v(\emptyset) = 0$. Let $\mathbb{V}(N)$ denote the set of coalition functions for N. Since we work within a fixed player set, we frequently drop the player set as an argument. In particular, we write $\mathbb{V}$ instead of $\mathbb{V}(N)$ and address $v \in \mathbb{V}$ as a game. Subsets of N are called **coalitions**; $v(S)$ is called the worth of coalition S. For $v, w \in \mathbb{V}$ and $\lambda \in \mathbb{R}$, the coalition functions $v + w \in \mathbb{V}$ and $\lambda \cdot v \in \mathbb{V}$ are given by $(v + w)(S) = v(S) + w(S)$ and $(\lambda \cdot v)(S) = \lambda \cdot v(S)$ for all $S \subseteq N$. For $T \subseteq N$, $T \neq \emptyset$, the game $u_T \in \mathbb{V}$, $u_T(S) = 1$ if $T \subseteq S$ and $u_T(S) = 0$ for $T \nsubseteq S$, is called a **unanimity game**. For $T \subseteq N$, $T \neq \emptyset$, the game $e_T \in \mathbb{V}$, $e_T(S) = 1$ if $T = S$ and $e_T(S) = 0$ for $T \neq S$, is called a **standard game**. A game v is called **monotonic** if $v(S) \geq v(T)$ for all $S, T \subseteq N$ such that $T \subseteq S$. Any $v \in \mathbb{V}$ can be uniquely represented by unanimity games,

$$v = \sum_{T \subseteq N : T \neq \emptyset} \lambda_T(v) \cdot u_T, \tag{1}$$

where the Harsanyi dividends, $\lambda_T(v)$, $T \subseteq N$, $T \neq \emptyset$ (Harsanyi, 1959) are given implicitly by

$$v(S) = \sum_{T \subseteq S : T \neq \emptyset} \lambda_T(v) \qquad \text{for all } S \subseteq N,\ S \neq \emptyset. \tag{2}$$

For $v \in \mathbb{V}$, $i \in N$, and $S \subseteq N \setminus \{i\}$, the **marginal contribution** of i to S in v is given by $MC_i^v(S) := v(S \cup \{i\}) - v(S)$. Player $i \in N$ is called a **null player** in $v \in \mathbb{V}$ if $MC_i^v(S) = 0$ for all $S \subseteq N \setminus \{i\}$; players $i, j \in N$ are called **symmetric** in $v \in \mathbb{V}$ if $MC_i^v(S) = MC_j^v(S)$ for all $S \subseteq N \setminus \{i, j\}$.

A **value** on N is an operator φ that assigns a payoff vector $\varphi(v) \in \mathbb{R}^N$ to any $v \in \mathbb{V}$. For $S \subseteq N$, we denote $\sum_{i \in S} \varphi_i(v)$ by $\varphi_S(v)$. The **Shapley value** (Shapley, 1953), Sh, given by

$$\mathrm{Sh}_i(v) := \sum_{T \subseteq N : i \in T} \frac{\lambda_T(v)}{|T|} \qquad \text{for all } i \in N,\ v \in \mathbb{V} \tag{3}$$

is the unique value on N that satisfies the axioms **E**, **A** (or **L**), **S**, and **N** below.

Efficiency, E. For all $v \in \mathbb{V}$, $\varphi_N(v) = v(N)$.

Additivity, A. For all $v, w \in \mathbb{V}$, $\varphi(v + w) = \varphi(v) + \varphi(w)$.

Null player, N. For all $v \in \mathbb{V}$ and all $i \in N$, who are null players in v, $\varphi_i(v) = 0$.

We further refer to the following standard axioms.

Linearity, L. For all $v, w \in \mathbb{V}$ and $\lambda \in \mathbb{R}$, $\varphi(v + w) = \varphi(v) + \varphi(w)$ and $\varphi(\lambda \cdot v) = \lambda \cdot \varphi(v)$.

Symmetry, S. For all $v \in \mathbb{V}$, $i \in N$, and all bijections $\pi : N \to N$, $\varphi_{\pi(i)}\left(v \circ \pi^{-1}\right) = \varphi_i(v)$.

Continuity, C. The mapping $\varphi : \mathbb{V} \to \mathbb{R}^N$ is continuous.

Moreover, we refer to the following values, which also obey **E**, **L**, and **S**. The **equal division value**, ED, is given by

$$\mathrm{ED}_i(v) := \frac{v(N)}{|N|} \qquad \text{for all } i \in N,\ v \in \mathbb{V}.$$

The **egalitarian Shapley values** (Joosten, 1996), Sh^α, $\alpha \in [0, 1]$, are given by $\mathrm{Sh}^\alpha = \alpha \cdot \mathrm{Sh} + (1 - \alpha) \cdot \mathrm{ED}$. The **equal surplus division value** (Driessen & Funaki, 1991), ES, is given by

$$\mathrm{ES}_i(v) := v(\{i\}) + \frac{v(N) - \sum_{j \in N} v(\{j\})}{|N|} \qquad \text{for all } i \in N,\ v \in \mathbb{V}.$$

The **solidarity value** (Nowak & Radzik, 1994), So, is given by

$$\mathrm{So}_i(v) := \sum_{S \subseteq N : i \in S} \frac{1}{\binom{|N|}{|S|} \cdot |S|} \sum_{j \in S} \frac{v(S) - v(S \setminus \{j\})}{|S|} \qquad \text{for all } i \in N,\ v \in \mathbb{V}.$$

The **consensus value** (Ju et al., 2007), Con, is given by $\mathrm{Con} = \frac{1}{2} \cdot \mathrm{Sh} + \frac{1}{2} \cdot \mathrm{ES}$. The **least-square pre-nucleolus** (Ruiz et al., 1996), LSPN, is given by

$$\mathrm{LSPN}_i(v) := \mathrm{Ba}_i(v) + \frac{v(N) - \sum_{j \in N} \mathrm{Ba}_j(v)}{|N|} \qquad \text{for all } i \in N,\ v \in \mathbb{V},$$

where Ba stands for the Banzhaf value (Banzhaf, 1965; Owen, 1975),

$$\mathrm{Ba}_i(v) := \sum_{T \subseteq N : i \in T} \frac{\lambda_T(v)}{2^{|T|-1}} \qquad \text{for all } i \in N,\ v \in \mathbb{V}.$$

3 Efficient, Linear, and Symmetric Values

In this section, we first provide the formulae for the class of efficient, linear, and symmetric values (henceforth, **ELS values**) mentioned in the introduction. The formulae below apply to all $v \in \mathbb{V}$ and $i \in N$.

Ruiz et al. (1998): For $\rho = \left(\rho_1, \ldots, \rho_{|N|-1}\right) \in \mathbb{R}^{|N|-1}$, the ELS value RVZ^ρ is given by

$$\mathrm{RVZ}_i^\rho(v) := \frac{v(N)}{|N|} + \sum_{S \subsetneq N : i \in S} \frac{\rho_{|S|}}{|S|} \cdot v(S) - \sum_{S \subseteq N \setminus \{i\} : S \neq \emptyset} \frac{\rho_{|S|}}{|N| - |S|} \cdot v(S).$$

Driessen and Radzik (2003)[2]: For $b = \left(b_1, \ldots, b_{|N|-1}\right) \in \mathbb{R}^{|N|-1}$, the ELS value DR^b is given by

$$\mathrm{DR}_i^b(v) := \frac{v(N)}{|N|} + \sum_{S \subsetneq N \setminus \{i\}} \frac{b_{|S|+1} \cdot v(S \cup \{i\})}{\binom{|N|}{|S|+1}(|S|+1)} - \sum_{S \subseteq N \setminus \{i\} : S \neq \emptyset} \frac{b_{|S|} \cdot v(S)}{\binom{|N|}{|S|+1}(|S|+1)}. \tag{4}$$

A major disadvantage of the above formulae is that the parameters can hardly be interpreted in economic terms. To remedy this, (Chameni Nembua, 2012) proposes another type of parametrization.[3] For $\alpha = \left(\alpha_2, \ldots, \alpha_{|N|}\right) \in \mathbb{R}^{|N|-1}$, the ELS value CN^α is given by

$$\mathrm{CN}_i^a(v) := \frac{v(\{i\})}{|N|} + \sum_{S \subseteq N : i \in S, |S| > 1} \frac{AMC_i^v(S, \alpha)}{\binom{|N|}{|S|} \cdot |S|},$$

where

$$AMC_i^v(S, \alpha) := \alpha_{|S|} \cdot [v(S) - v(S \setminus \{i\})] + \frac{1 - \alpha_{|S|}}{|S| - 1} \sum_{j \in S \setminus \{i\}} [v(S) - v(S \setminus \{j\})].$$

According to this formula, a player's payoff is some average of marginal contributions, both of her own ones and the other players' ones. Within a coalition S, the marginal contribution of player $i \in S$ is taxed at a rate of $1 - \alpha_{|S|}$, leav-

[2] Chameni Nembua and Andjiga (2008) and Malawski (2013) (and personal communication) consider essentially the same formulae, the latter under the name *inversely procedural values*. Moreover, (Hernandez-Lamoneda et al., 2008) consider similar parametrizations, which just rescale the parameters. Actually, they consider continuous values and require just additivity. Yet, it is well known that linearity entails continuity and that additivity combined with continuity implies linearity.

[3] Malawski (2013) suggests essentially the same formulae as the *procedural values*. Instead of marginal contributions to coalitions, he considers marginal contributions for orders of the player set.

ing him a share of $\alpha_{|S|} \cdot [v(S) - v(S \setminus \{i\})]$, while the tax revenue amounting to $(1 - \alpha_{|S|}) \cdot [v(S) - v(S \setminus \{i\})]$ is distributed evenly among the *other* players in S.

Despite of the structural differences of the formulae above, they are closely related. By applying these values to standard games, the parameters can be recovered in a similar fashion. In particular, for $T \subsetneq N$, $T \neq \emptyset$, and $i \in T$, we have

$$\begin{aligned} \rho_{|T|} &= |T| \cdot \mathrm{RVZ}_i^{\rho}(e_T), \\ b_{|T|} &= \binom{|N|}{|T|} \cdot |T| \cdot \mathrm{DR}_i^{b}(e_T), \\ \alpha_{|T|+1} &= \binom{|N|}{|T|} \cdot |T| \cdot \mathrm{CN}_i^{\alpha}(e_T). \end{aligned} \tag{5}$$

Hence, conditions on the parameters as for example imposed by Radzik and Driessen (2013) for the formula suggested by Driessen and Radzik (2003) can easily be translated into conditions for the parameters of the other formulae above.

We advocate another formula for the class of ELS values, already indicated by Radzik and Driessen (2009, p. 5). In contrast to the approaches above, our formula is based on unanimity games, i.e., on the Harsanyi dividends $\lambda_T(v)$ in (1). We consider the following class of values on N. For $\tau = (\tau_1, \ldots, \tau_{|N|-1}) \in \mathbb{R}^{|N|-1}$, the value ζ^{τ} on N is given by

$$\zeta_i^{\tau}(v) = \frac{\lambda_N(v)}{|N|} + \sum_{T \subsetneq N : T \neq \emptyset} \frac{\tau_{|T|}}{|N|} \cdot \lambda_T(v) + \sum_{T \subsetneq N : i \in T} \left(\frac{1 - \tau_{|T|}}{|T|} \right) \cdot \lambda_T(v), \tag{6}$$

for all $i \in N$, $v \in \mathbb{V}$.

Formula (6) can be interpreted as follows: The Harsanyi dividend $\lambda_T(v)$ represents the gains from cooperation in the game v that can be attributed to coalition T. On the right-hand side of (6), the rightmost term indicates that $\lambda_T(v)$ is first taxed at a rate of $\tau_{|T|}$ and then distributed equally among the players in T. The middle term indicates that the tax revenue is (re)distributed equally among all players in N. The leftmost term covers the gains from cooperation of the grand coalition N. Here, taxation and redistribution would lead to an equal distribution among all players in N, independent of the tax rate. Note that whereas the Shapley value distributes $\lambda_T(v)$ equally among the players in T (tax rate of 0), the equal division value distributes it equally among all players in N (tax rate of 1). Further examples of tax rates related to ELS values can be found in Table 1.

While the formulae in the previous section are closely related via (5), our formula is distinct. Instead of standard games, unanimity games are employed to recover the parameters. For $T \subsetneq N$, $T \neq \emptyset$, and $i \in N \setminus T$, we have

$$\tau_{|T|} = |N| \cdot \zeta_i^{\tau}(u_T). \tag{7}$$

The parameters $(\tau_1, \ldots, \tau_{|N|-1})$ can be interpreted as tax rates that are applied to (scaled) unanimity games. For $\lambda \cdot u_T$, $\lambda \in \mathbb{R}$, $T \subseteq N$, $T \neq \emptyset$, we obtain

Table 1 Tax rates for some ELS values

	τ_1	τ_2	$\cdots$	τ_t	$\cdots$	$\tau_{\|N\|-1}$
Sh	0	0	$\cdots$	0	$\cdots$	0
Sh^{α}	$1-\alpha$	$1-\alpha$		$1-\alpha$		$1-\alpha$
CON	0	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}$	$\cdots$	$\frac{1}{2}$
ES	0	1	$\cdots$	1	$\cdots$	1
LSPN	$1-\frac{1}{\|N\|}$	$1-\frac{1}{\|N\|}$	$\cdots$	$1-\frac{t}{2^{t-1}}\frac{1}{\|N\|}$	$\cdots$	$1-\frac{\|N\|-1}{2^{\|N\|-2}}\frac{1}{\|N\|}$
ED	1	1	$\cdots$	1	$\cdots$	1

$$\zeta_i^{\tau}\left(\lambda \cdot u_T\right) = \frac{\tau_{|T|} \cdot \mathrm{Sh}_N\left(\lambda \cdot u_T\right)}{|N|} + \left(1-\tau_{|T|}\right) \cdot \mathrm{Sh}_i\left(\lambda \cdot u_T\right) \qquad \text{for all } i \in N.$$

That is, player i's Shapley payoff is taxed at a rate of $\tau_{|T|}$, leaving him a net income of $\left(1-\tau_{|T|}\right) \cdot \mathrm{Sh}_i\left(\lambda \cdot u_T\right)$, while the resulting overall tax revenue amounting to $\tau_{|T|} \cdot \mathrm{Sh}_N\left(\lambda \cdot u_T\right)$ is distributed evenly among *all* players. Note that this kind of taxation and redistribution would not affect the payoffs for $\lambda \cdot u_N$. Hence, there is no tax rate $\tau_{|N|}$. The following proposition is immediate from (6).

Proposition 1 *A value φ on N satisfies **L**, **E**, and **S** if and only if there is some $\tau \in \mathbb{R}^{|N|-1}$ such that $\varphi = \zeta^{\tau}$, where ζ^{τ} is as in (6).*

A number of values in the literature belong to the class of ELS values. In Table 1, we provide the tax rates $\tau \in \mathbb{R}^{|N|-1}$ for some of them. Unfortunately, there seems to be no "nice" expressions for the tax rates that produce the solidarity value.

4 Solidarity and Fair Taxation

Within the class of ELS values dwells a huge number of values that do not show certain economically sound properties. In this section, we provide conditions on the parameters of our Formula (6) such that one or another of the desirable properties mentioned in the introduction is satisfied. These properties can be viewed as requirements of fair taxation.

4.1 Technical Preliminaries

Later on, we will make use of the following definitions. For $m \in \mathbb{N}$ and $x \in \mathbb{R}^m$, the forward differences $\Delta_t^k x$, $t \in \{1, \ldots, m\}$, $k \in \{0, \ldots, m-t\}$ are given recursively by

$$\Delta_t^0 x := x_t \text{ and } \Delta_t^{k+1} x := \Delta_t^k x - \Delta_{t+1}^k x \; \forall t \in \{1, \ldots, m\}, \; k \in \{0, \ldots, m-t\}. \tag{8}$$

It is well known that $\Delta_t^k x$ is given by

$$\Delta_t^k x = \sum_{\ell=0}^{k} (-1)^{\ell} \cdot \binom{k}{\ell} \cdot x_{t+\ell} \qquad \text{for all } t \in \{1, \ldots, m\}, \; k \in \{0, \ldots, m-t\}. \tag{9}$$

The following lemma easily follows from (8) by induction on $t + k$.

Lemma 1 *Let $m \in \mathbb{N}$ and $x \in \mathbb{R}^m$. Then, $\Delta_t^{m-t} x \geq 0$ for all $t \in \{1, \ldots, m\}$ implies $\Delta_t^k x \geq 0$ for all $t \in \{1, \ldots, m\}$, $k \in \{0, \ldots, m-t\}$.*

Moreover, we employ a transformation $x \mapsto \hat{x}$ for $x \in \mathbb{R}^m$, $m \in \mathbb{N}$ defined by

$$\hat{x}_t := \frac{x_t}{t} \qquad \text{for all } t \in \{1, \ldots, m\}. \tag{10}$$

Let $\mathbf{0}, \mathbf{1} \in \mathbb{R}^m$ be given by $\mathbf{0}_t = 0$ and $\mathbf{1}_t = 1$ for all $t \in \{1, \ldots, m\}$. By induction on k, one easily shows

$$\Delta_t^k \left(\rho \cdot \hat{\mathbf{1}} \right) = \frac{\rho}{(t+k) \binom{t+k-1}{k}} \tag{11}$$

for all $\rho \in \mathbb{R}$, $t \in \{1, \ldots, m\}$, and $k \in \{0, \ldots, m-t\}$.

4.2 *Relation Between Parameters*

We prepare our main results by establishing the relation between our parameters and the parameters of the other formulae for ELS values. In view of (5), we focus on the formula suggested by Driessen and Radzik (2003). First, we show how our parameters can be translated into parameters for the latter formula. The proof of the proposition is referred to the appendix.

Proposition 2 *For $b \in \mathbb{R}^{|N|-1}$ and $\tau \in \mathbb{R}^{|N|-1}$, we have $\mathrm{DR}^b = \zeta^\tau$ if and only if*

$$b_t = 1 - \frac{\Delta_t^{|N|-1-t} \hat{\tau}}{\Delta_t^{|N|-1-t} \hat{\mathbf{1}}} \qquad \textit{for all } t \in \{1, \ldots, |N|-1\}.$$

Now, we show how the parameters of the formula suggested by Driessen and Radzik (2003) can be translated into the parameters of our formula. The proof of the proposition is referred to the appendix.

Proposition 3 *For $b \in \mathbb{R}^{|N|-1}$ and $\tau \in \mathbb{R}^{|N|-1}$, we have $\mathrm{DR}^b = \zeta^\tau$ if and only if*

$$\tau_t = 1 - \frac{1}{\binom{|N|-1}{t}} \sum_{s=t}^{|N|-1} \binom{s-1}{t-1} \cdot b_s \qquad \textit{for all } t \in \{1, \ldots, |N|-1\}\,.$$

4.3 Desirability

Even if players express solidarity among themselves, the payoffs should reflect their individual productivity. At least, payoff differentials should not be opposite to their productivities. This idea is expressed by the desirability axiom.[4]

Desirability, D (Maschler & Peleg, 1966). For all $v \in \mathbb{V}$ and $i, j \in N$ such that $MC_i^v(S) \geq MC_j^v(S)$ for all $S \subseteq N \setminus \{i, j\}$, we have $\varphi_i(N, v) \geq \varphi_j(N, v)$.

The ELS value DR^b in (4) meets desirability if and only if $b_t \geq 0$ for all $t \in \{1, \ldots, |N|-1\}$ (Radzik & Driessen, 2013, Theorem 1). Combining this result with Proposition 2, we obtain the following requirements on our parameters to guarantee desirability. Casajus (2012) provides a direct proof of this result.

Theorem 1 *The ELS value ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ obeys desirability (**D**) if and only if $\Delta_t^{|N|-1-t}\hat{\mathbf{1}} \geq \Delta_t^{|N|-1-t}\hat{\tau}$ for all $t \in \{1, \ldots, |N|-1\}$.*

Remark 1 Theorem 1 together with Lemma 1 imply the following necessary requirements on $\tau \in \mathbb{R}^{|N|-1}$ for ζ^τ to satisfy desirability: $\Delta_t^k\hat{\mathbf{1}} \geq \Delta_t^k\hat{\tau}$ for all $t \in \{1, \ldots, |N|-1\}$ and $k \in \{0, \ldots, |N|-t-1\}$. In some strong sense, taxes should be smaller than 1. In particular, we have (i) $\Delta_t^0\hat{\mathbf{1}} \geq \Delta_t^0\hat{\tau}$, i.e., $1 \geq \tau_t$ for all $t \in \{1, \ldots, |N|-1\}$; i.e., the players should not be "overtaxed". Further, (ii) $\Delta_t^1\hat{\mathbf{1}} \geq \Delta_t^1\hat{\tau}$, i.e., $\tau_{t+1} \geq \tau_t - \frac{1-\tau_t}{t}$ for all $t \in \{1, \ldots, |N|-2\}$. Given $1 \geq \tau_t$, this means that tax rates should not decrease too much when t increases. In particular, if $\tau_t = 1$ for some t, then $\tau_s = 1$ for all $s \geq t$.

4.4 Positivity for Null Players

In monotonic games, no player ever is destructive; i.e., all players always have a non-negative marginal contributions. Hence, even if players show solidarity to less productive ones, nobody should end up with a sub-zero payoff. This idea is expressed by the positivity axiom.[5]

Positivity (Kalai and Samet, 1987)**, P.** For all $v \in \mathbb{V}$ that are monotonic and all $i \in N$, we have $\varphi_i(N, v) \geq 0$.

[4] Desirability is also known as *local monotonicity* (of values) (see Levinský & Silársky, 2004), or *fair treatment* (see Radzik & Driessen, 2013).

[5] Positivity is also known as *monotonicity* (see Radzik & Driessen, 2013).

Radzik and Driessen (2013) do not give exact conditions on $b \in \mathbb{R}^{|N|-1}$ such that the value DR^b meets positivity. Careful inspection of the proof of their Theorem 2 shows that one actually has a nice description of those $b \in \mathbb{R}^{|N|-1}$ that let DR^b obey a weaker property, positivity for null players. Note that the negative Shapley value, $-\mathrm{Sh}$, for example, fails positivity but not positivity for null players. Moreover, the negative equal division value, $-\mathrm{ED}$, fails both positivity for null players and positivity, for example.

Positivity for null players, PN. For all $v \in \mathbb{V}$ that are monotonic and all $i \in N$ who are null players in v, we have $\varphi_i(N, v) \geq 0$.

Theorem 2 *The value* DR^b, $b \in \mathbb{R}^{|N|-1}$ *obeys positivity for null players (**PN**) if and only if* $1 \geq b_t$ *for all* $t \in \{1, \ldots, |N|-1\}$.

Proof Sufficiency: Let $v \in \mathbb{V}$ be monotonic and $i \in N$ be a null player in v. Let further $b \in \mathbb{R}^{|N|-1}$ be such that $1 \geq b_t$ for all $t \in \{1, \ldots, |N|-1\}$. This part of the proof heavily relies on the proof of , which makes use of the additional assumption (*) $b_t \geq 0$ for all $t \in \{1, \ldots, |N|-1\}$. Since i is a null player in v, we can avoid to appeal to (*).

For notational parsimony, we just indicate how this can be done and refer the reader to Radzik and Driessen (2013, proof of Theorem 2) , which is by induction. Condition (*) is applied twice, first, in the induction basis, and second, in the final step after induction. We obtain

$$\begin{aligned}\mathrm{DR}_i^b(v) &\overset{(4)}{=} \frac{v(N)}{|N|} + \sum_{S \subsetneq N \setminus \{i\}} \frac{b_{|S|+1} \cdot v(S \cup \{i\})}{\binom{|N|}{|S|+1} \cdot (|S|+1)} - \sum_{S \subseteq N \setminus \{i\} : S \neq \emptyset} \frac{b_{|S|} \cdot v(S)}{\binom{|N|}{|S|+1} \cdot (|S|+1)} \\ &= \left(1 - b_{|N|-1}\right) \cdot \frac{v(N)}{|N|} + \sum_{S \subsetneq N \setminus \{i\}} \frac{\left(b_{|S|+1} - b_{|S|}\right) \cdot v(S \cup \{i\})}{\binom{|N|}{|S|+1} \cdot (|S|+1)},\end{aligned}$$

where the second equation follows from i being a null player. This already establishes the induction basis (second equation after Equation (23) of the proof). Since $v(\emptyset \cup \{i\}) = 0$ for the null player i, we do not need (*) in the last equation/final step of the proof.

Necessity: Let now $b \in \mathbb{R}^{|N|-1}$ be such that DR^b meets **PN**. Fix $t \in \{1, \ldots, |N|-1\}$ and $i \in N$ and let $v \in \mathbb{V}$ be given by $v(S) = 1$ if $|S| > t$ and $i \in S$, $v(S) = 1$ if $|S| \geq t$ and $i \notin S$, and $v(S) = 0$ else. One easily checks that v is monotonic and that i is a null player in v. Moreover Radzik and Driessen (2013, Proof of Theorem 2) show $\mathrm{DR}_i^b(v) = \frac{1-b_t}{|N|}$. Hence, $b_t \leq 1$ for all $t \in \{1, \ldots, |N|-1\}$. □

Combining this result with Proposition 2, we obtain the following requirements on our parameters to guarantee positivity for null players. Casajus (2012) provides a direct proof of this result.

Theorem 3 *The value* ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ *obeys positivity for null players (**PN**) if and only if* $\Delta_t^{|N|-1-t}\hat{\tau} \geq \Delta_t^{|N|-1-t}\hat{\mathbf{0}} = 0$ *for all* $t \in \{1, \ldots, |N|-1\}$.

Remark 2 Theorem 3 together with Lemma 1 implies the following necessary requirements on $\tau \in \mathbb{R}^{|N|-1}$ for ζ^τ to satisfy positivity for null players: $\Delta_t^k \hat{\tau} \geq \Delta_t^k \hat{\mathbf{0}} = 0$ for all $t \in \{1, \ldots, |N| - 1\}$ and $k \in \{0, \ldots, |N| - t - 1\}$. In some strong sense, taxes should be non-negative. In particular, we have (i) $\Delta_t^0 \hat{\tau} \geq 0$, i.e., $\tau_t \geq 0$ for all $t \in \{1, \ldots, |N| - 1\}$, i.e., the players should not be subsidized. Further, (ii) $\Delta_t^1 \hat{\tau} \geq 0$, i.e., $\frac{t+1}{t}\tau_t \geq \tau_{t+1}$ for all $t \in \{1, \ldots, |N| - 2\}$. Given $\tau_t, \tau_{t+1} \geq 0$, this means that tax rates should not increase too much when t increases. In particular, if $\tau_t = 0$ for some t, then $\tau_s = 0$ for all $s \geq t$.

Remark 3 Casajus and Huettner (2013) consider a considerable sharpening of positivity for null players, the null player in a productive environment property below. Instead of restricting attention to monotonic games, they extend the implication of **PN** to games where the grand coalition generates a non-negative worth.

Null player in a productive environment, NPE. For all $v \in \mathbb{V}$ and $i \in N$ such that i is a null player in v and $v(N) \geq 0$, we have $\varphi_i(v) \geq 0$.

Their Proposition 1 entails that the value ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ obeys the null player in a productive environment property if and only if $\tau_t = \tau_1 \geq 0$ for all $t \in \{1, \ldots, |N| - 1\}$. By (11) and Proposition 2, the values DR^b, $b \in \mathbb{R}^{|N|-1}$ satisfy this property if and only if $1 \geq b_t = b_1$ for all $t \in \{1, \ldots, |N| - 1\}$.

Remark 4 Recently, Radzik (2021, Theorem 6) shows that the value DR^b, $b \in \mathbb{R}^{|N|-1}$ obeys positivity (**P**) if and only if[6] $b_t \leq 1$ and

$$\sum_{\ell=t}^{T} b_\ell \geq -1 \qquad \text{for all } t, T \in \{1, \ldots, |N| - 1\}, \ t \leq T.$$

By Proposition 3, the value ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ obeys positivity (**P**) if and only if $\Delta_t^{|N|-1-t} \hat{\tau} \geq \Delta_t^{|N|-1-t} \hat{\mathbf{0}} = 0$ and

$$T - t + 2 \geq \sum_{\ell=t}^{T} \frac{\Delta_\ell^{|N|-1-\ell} \hat{\tau}}{\Delta_\ell^{|N|-1-\ell} \hat{\mathbf{1}}} \qquad \text{for all } t, T \in \{1, \ldots, |N| - 1\}, \ t \leq T.$$

4.5 *Desirability and Positivity*

The ELS value DR^b, $b \in \mathbb{R}^{|N|-1}$ satisfies desirability and positivity if and only if $1 \geq b_t \geq 0$ for all $t \in \{1, \ldots, |N| - 1\}$ (Radzik and Driessen, 2013, Theorem 2). Combining this result with Proposition 2, we obtain the following requirements on our parameters to guarantee the combination of desirability and positivity.

Theorem 4 *The value ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ obeys desirability (**D**) and positivity (**P**) if and only if $\Delta_t^{|N|-1-t} \hat{\mathbf{1}} \geq \Delta_t^{|N|-1-t} \hat{\tau} \geq \Delta_t^{|N|-1-t} \hat{\mathbf{0}} = 0$ for all for all $t \in \{1, \ldots, |N| - 1\}$.*

[6] Malawski (2013, Lemma 5) already shows that these conditions are necessary.

Remark 5 Theorem 4 together with Lemma 1 implies the following necessary requirements on $\tau \in \mathbb{R}^{|N|-1}$ for ζ^{τ} to satisfy desirability and positivity for null players: $\Delta_t^k \hat{\mathbf{1}} \geq \Delta_t^k \hat{\tau} \geq \Delta_t^k \hat{\mathbf{0}} = 0$ for all $t \in \{1, \ldots, |N|-1\}$ and $k \in \{0, \ldots, |N|-t-1\}$. In a strong sense, tax rates are required to fall between 0 and 1.

Remark 6 Theorems 1, 3, and 4, alternatively, Theorem 2 and Radzik and Driessen (2013, Theorems 1 and 2) imply that an ELS value that satisfies desirability and positivity for null players also satisfies positivity. Casajus (2012) provides a direct proof of this result.

We now demonstrate the power of Theorem 4 with an example. A technical lemma facilitates the application of the theorem.

Lemma 2 *Let $m \in \mathbb{N}$ and $f : [1, m] \to \mathbb{R}$ be differentiable up to order $m-1$ and such that $(-1)^k \cdot f^{(k)}(\xi) \geq 0$ for all $\xi \in [1, m]$ and $k \in \{0, \ldots, m-1\}$. For $x \in \mathbb{R}^m$ given by $x_t = f(t)$ for all $t \in \{1, \ldots, m\}$, we have $\Delta_t^k x \geq 0$ for all $t \in \{1, \ldots, m\}$ and $k \in \{0, \ldots, m-t\}$.*

Proof Let m and f be as in the lemma. For $t \in \{1, \ldots, m\}$, we have

$$\Delta_t^0 x = x_t = f(t) = f^{(0)}(t) = (-1)^0 \cdot f^{(0)}(t) \geq 0.$$

By induction on k, one easily shows

$$\Delta_t^k x = (-1)^k \int_t^{t+1} \int_{i_2}^{i_2+1} \int_{i_3}^{i_3+1} \ldots \int_{i_k}^{i_k+1} f^{(k)}(\xi)\, d\xi\, di_k \ldots di_3 di_2$$

for all $t \in \{1, \ldots, m\}$ and $k \in \{1, \ldots, m-t\}$. The claim now follows from $(-1)^k \cdot f^{(k)}(\xi) \geq 0$ for all $\xi \in [1, m]$. □

Example 1 For $\alpha \in [0, 1]$, we consider the tax system $\tau^{\alpha} \in \mathbb{R}^{|N|-1}$ such that

$$\zeta_i^{\tau^{\alpha}}(u_T) = \alpha \cdot \zeta_j^{\tau^{\alpha}}(u_T), \qquad \text{for all } T \subsetneq N,\ T \neq \emptyset,\ i \in N \setminus T,\ j \in T.$$

That is, in unanimity games, unproductive players obtain α times the payoff of productive players. By (6), we obtain

$$\tau_t^{\alpha} = \frac{\alpha \cdot |N|}{(1-\alpha) \cdot t + \alpha \cdot |N|} \qquad \text{for all } t \in \{1, \ldots, |N|-1\}.$$

The resulting value $\zeta^{\tau^{\alpha}}$ meets **D** and **P**. To see this, let $f, g : [1, |N|-1] \to \mathbb{R}$ be given by

$$f(\xi) = \frac{1}{\xi}, \qquad g(\xi) = \frac{1-\alpha}{(1-\alpha) \cdot \xi + \alpha \cdot |N|} \qquad \text{for all } \xi \in [1, |N|-1].$$

By (10), we have $[\tau^{\alpha}]_t = f(t) - g(t)$ and $([\mathbf{1}] - [\tau^{\alpha}])_t = g(t)$ for $t \in \{1, \ldots, |N| - 1\}$. Moreover, one obtains $f^{(0)}(\xi) \geq g^{(0)}(\xi) \geq 0$, and

$$(-1)^k \cdot f^{(k)}(\xi) = \frac{(-1)^{2k} \cdot k!}{\xi^{k+1}} \geq \frac{(-1)^{2k} \cdot k! \cdot (1-\alpha)^{k+1}}{((1-\alpha) \cdot \xi + \alpha \cdot |N|)^{k+1}} = (-1)^k \cdot g^{(k)}(\xi) \geq 0$$

for all $\xi \in [1, |N| - 1]$, $k \in \{0, \ldots, |N| - 2\}$. By Lemma 2, we have $\Delta_t^{|N|-1-t}\hat{\mathbf{1}} \geq \Delta_t^{|N|-1-t}\hat{\tau}^{\alpha} \geq 0$ for all $t \in \{1, \ldots, |N| - 1\}$. Finally, the claim follows from Theorem 4.

4.6 Social Acceptability

Joosten et al. (1994) consider the social acceptability axiom.

Social acceptability, SA. For all $T \subseteq N$, $i \in T$, and $j \in N \setminus T$, we have $\varphi_i(u_T) \geq \varphi_j(u_T) \geq 0$.

Social acceptability imposes rather weak fairness requirements. Since unanimity games are monotonic, the requirement $\varphi_i(u_T) \geq 0$ and $\varphi_j(u_T) \geq 0$ above is equivalent to positivity restricted to unanimity games. In u_T, the players in T are more productive than those in $N \setminus T$. Hence for ELS values, demanding $\varphi_i(u_T) \geq \varphi_j(u_T)$ for $i \in T$ and $j \in N \setminus T$ is equivalent to desirability restricted to unanimity games.

Since the values ζ^{τ} are closely related to the linear representation of games by unanimity games, we state the following obvious proposition with some diffidence and mainly for completeness' sake.

Proposition 4 *The value ζ^{τ}, $\tau \in \mathbb{R}^{|N|-1}$ obeys social acceptability (**SA**) if and only if $1 \geq \tau_t \geq 0$ for all $t \in \{1, \ldots, |N| - 1\}$.*

Proof Fix $T \subseteq N, |T| = t < |N|$. Let $i \in T,\ j \in N \setminus T$. By (6), we have $\zeta_i^{\tau}(u_T) - \zeta_j^{\tau}(u_T) = \frac{1-\tau_{|T|}}{|T|} \geq 0$ if and only if $\tau_t \leq 1$ and $\zeta_j^{\tau}(u_T) = \frac{\tau_t}{|N|} \geq 0$ iff $\tau_t \geq 0$. Further, $\zeta_i^{\tau}(u_N) = |N|^{-1} > 0$ for all $i \in N$. □

Remark 7 Compare the results of the proposition with analogous findings for the formulae based on standard games. The ELS value DR^b, $b \in \mathbb{R}^{|N|-1}$ satisfies social acceptability if and only if

$$0 \leq \frac{|N| \cdot t}{|N| - t} \cdot \binom{|N|}{t}^{-1} \cdot \sum_{s=t}^{|N|-1} \binom{s}{t} \cdot \frac{b_s}{s} \leq 1$$

for all $t \in \{1, \ldots, |N| - 1\}$ (Radzik and Driessen, 2013, Theorem 3).

4.7 General Acceptability

Radzik and Driessen (2013) consider another notion of acceptability, general acceptability.

General acceptability, GA. For all $S, T \subseteq N$ and $i \in N$ such that $S \subseteq T$ and $i \in S$, we have $\varphi_i(u_S) \geq \varphi_i(u_T)$.

Within the class of ELS values, general acceptability coincides with strong monotonicity for unanimity games. Note that on the domain of all TU games, there is a unique ELS value that meets strong monotonicity, the Shapley value (Young, 1985, Theorem 2).

Strong monotonicity, Mo$^+$ (Young, 1985). For all $v, w \in \mathbb{V}$ and $i \in N$ such that $v(K \cup \{i\}) - v(K) \geq w(K \cup \{i\}) - w(K)$ for all $K \subseteq N \setminus \{i\}$, we have $\varphi_i(v) \geq \varphi_i(w)$.

Proposition 5 *The value ζ^τ, $\tau \in \mathbb{R}^{|N|-1}$ obeys general acceptability (**GA**) if and only if (i) $\tau_t \leq 1$ for all $t \in \{1, \ldots, |N|-1\}$ and (ii)*

$$\tau_{t+1} - \tau_t \geq \frac{\tau_t - 1}{t} \cdot \frac{|N|}{|N| - t - 1}$$

for all $t \in \{1, \ldots, |N|-2\}$.

Proof Let $t \in \{1, \ldots, |N|-1\}$ and $T \subseteq N$, $|T| = t$. By (6), $\zeta_i^\tau(u_T) \geq \zeta_i^\tau(u_N)$ iff $\tau_t \leq 1$. Let $s \in \{1, \ldots, |N|-2\}$ and $S, T \subseteq N$, $S \subseteq T$, $|S| = s$, $|T| = s+1$. By (6), $\zeta_i^\tau(u_S) \geq \zeta_i^\tau(u_T)$ if and only if

$$\tau_{s+1} \geq \tau_s - \frac{1 - \tau_s}{s} \cdot \frac{|N|}{|N| - s - 1},$$

which entails the second part of the requirement. □

Remark 8 Proposition 5 first requires that there is no overtaxing, $\tau_t \leq 1$. Given this, the second requirement says that tax rates should not decrease too much when t increases. In particular, if $\tau_t = 1$ for some t, then $\tau_s = 1$ for all $s \geq t$. Recall some necessary requirements for desirability due to Theorem 1, (i) $\tau_s \leq 1$ for all $t \in \{1, \ldots, |N|-1\}$ and (ii) $\tau_{t+1} \geq \tau_t - \frac{1-\tau_t}{t}$ for all $t \in \{1, \ldots, |N|-2\}$. Since $\tau_t - 1 \leq 0$ and $\frac{|N|}{|N|-t-1} > 1$, desirability implies general acceptability for ELS values.

Remark 9 Compare the results of the proposition with analogous findings for the formulae based on standard games. The ELS value DR^b, $b \in \mathbb{R}^{|N|-1}$ satisfies general acceptability if and only if

$$0 \leq \sum_{s=t}^{|N|-1} \frac{|N| - s}{s} \cdot \binom{s}{t} \cdot b_s$$

for all $t \in \{1, \ldots, |N|-1\}$ (Radzik and Driessen, 2013, Theorem 4).

5 Concluding Remarks

There essentially are two types of parametrized formulas for a class of values that has attracted attention in the literature, the ELS values. In a sense, theses formulas are based on two types of "basic" games, standard games or unanimity games. For the former, conditions for the parameters already have been established such that the resulting values satisfy certain "desirable" properties. In this paper, we establish such conditions for the latter. This helps to identify or to construct ELS values that satisfy one or another of these properties.

Acknowledgements We are grateful to Frank Huettner for helpful discussions on this paper.

Appendix: Omitted Proofs

Proof of Proposition 2

Necessity: Let $S \subsetneq N$, $S \neq \emptyset$, $i \in S$ and $j^* \in N \setminus S$. First, we determine the payoff $\zeta_i^\tau(e_S)$. It is well-known that $\lambda_T(e_S) = \begin{cases} 0, & S \not\subseteq T, \\ (-1)^{|T|-|S|}, & S \subseteq T \end{cases}$ for all $T \subseteq N$.

We this ($\star$), we obtain

$$
\begin{aligned}
&\zeta_i^\tau(e_S) \\
&\overset{(6)}{=} \frac{\lambda_N(e_S)}{|N|} + \sum_{T \subsetneq N: T \neq \emptyset} \frac{\tau_{|T|}}{|N|} \cdot \lambda_T(e_S) + \sum_{T \subsetneq N: i \in T} \left(\frac{1-\tau_{|T|}}{|T|} \right) \cdot \lambda_T(e_S) \\
&= \frac{\lambda_N(e_S)}{|N|} + \frac{1}{|N|} \sum_{j \in N} \sum_{j \in T \subsetneq N} \frac{\tau_{|T|}}{|T|} \cdot \lambda_T(e_S) + \sum_{T \subsetneq N: i \in T} \frac{1-\tau_{|T|}}{|T|} \cdot \lambda_T(e_S) \\
&\overset{i \in S, \star}{=} \frac{(-1)^{|N|-|S|}}{|N|} + \frac{1}{|N|} \sum_{j \in N} \sum_{S \subseteq T \subsetneq N: j \in T} \frac{\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} \\
&+ \sum_{S \subseteq T \subsetneq N} \frac{1-\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} \\
&= \sum_{S \subseteq T \subseteq N} \frac{(-1)^{|N|-|S|}}{|N|} + \frac{1}{|N|} \sum_{j \in S} \sum_{S \subseteq T \subsetneq N} \frac{\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} \\
&+ \frac{1}{|N|} \sum_{j \in S} \sum_{S \subseteq T \subsetneq N: j \in T} \frac{\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} - \sum_{S \subseteq T \subsetneq N} \frac{\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} \\
&\overset{j^* \in N \setminus S}{=} \sum_{S \subseteq T \subseteq N} \frac{(-1)^{|N|-|S|}}{|N|} - \frac{|N|-|S|}{|N|} \sum_{S \subseteq T \subsetneq N} \frac{\tau_{|T|}}{|T|} \cdot (-1)^{|T|-|S|} \\
&+ \frac{|N|-|S|}{|N|} \sum_{S \subseteq T \subsetneq N \setminus \{j^*\}} \frac{\tau_{|T \cup \{j^*\}|}}{|T \cup \{j^*\}|} \cdot (-1)^{|T \cup \{j^*\}|-|S|}
\end{aligned}
$$

$$= \sum_{t=|S|}^{|N|} \binom{|N|-|S|}{t-|S|} \cdot \frac{(-1)^{t-|S|}}{t} - \frac{|N|-|S|}{|N|} \sum_{t=|S|}^{|N|-1} \binom{|N|-|S|}{t-|S|} \cdot \frac{\tau_t}{t} \cdot (-1)^{t-|S|}$$

$$+ \frac{|N|-|S|}{|N|} \sum_{t=|S|}^{|N|-2} \binom{|N|-1-|S|}{t-|S|} \cdot \frac{\tau_{t+1}}{t+1} \cdot (-1)^{t+1-|S|}$$

$$= \sum_{t=0}^{|N|-|S|} \binom{|N|-|S|}{t} \cdot \frac{(-1)^t}{t+|S|} - \frac{|N|-|S|}{|N|} \sum_{t=0}^{|N|-1-|S|} \binom{|N|-|S|}{t} \cdot \frac{\tau_{t+|S|}}{t+|S|} \cdot (-1)^t$$

$$+ \frac{|N|-|S|}{|N|} \sum_{t=0}^{|N|-2-|S|} \binom{|N|-1-|S|}{t} \cdot \frac{\tau_{t+1+|S|}}{t+1+|S|} \cdot (-1)^{t+1}. \tag{12}$$

Moreover, we have

$$\sum_{t=0}^{|N|-1-|S|} \binom{|N|-|S|}{t} \cdot \frac{\tau_{t+|S|}}{t+|S|} \cdot (-1)^t$$

$$= \binom{|N|-|S|}{0} \frac{\tau_{0+|S|}}{0+|S|} \cdot (-1)^0 + \sum_{t=1}^{|N|-1-|S|} \binom{|N|-1-|S|}{t} \cdot \frac{\tau_{t+|S|}}{t+|S|}$$

$$+ \sum_{t=1}^{|N|-1-|S|} \binom{|N|-1-|S|}{t-1} \cdot \frac{\tau_{t+|S|}}{t+|S|}$$

$$= \sum_{t=0}^{|N|-1-|S|} \binom{|N|-1-|S|}{t} \cdot \frac{\tau_{t+|S|}}{t+|S|}$$

$$+ \sum_{t=0}^{|N|-2-|S|} \binom{|N|-1-|S|}{t} \cdot \frac{\tau_{t+1+|S|}}{t+1+|S|}. \tag{13}$$

Combining (12) and (13) gives

$$\zeta_i^\tau(e_S) = \sum_{t=0}^{|N|-|S|} \binom{|N|-|S|}{t} \frac{(-1)^t}{t+|S|} - \frac{|N|-|S|}{|N|} \sum_{t=0}^{|N|-1-|S|} \binom{|N|-1-|S|}{t} \frac{\tau_{t+|S|}}{t+|S|}$$

$$\overset{(9)}{=} \frac{1}{|S| \cdot \binom{|N|}{|S|}} - \frac{|N|-|S|}{|N|} \Delta_{|S|}^{|N|-1-|S|}[\tau]$$

$$\overset{(11)}{=} \frac{1}{|S| \cdot \binom{|N|}{|S|}} \left(1 - \frac{\Delta_{|S|}^{|N|-1-|S|}[\tau]}{\Delta_{|S|}^{|N|-1-|S|}[\mathbf{1}]} \right).$$

In view of (5), the claim now is immediate.

Sufficiency follows from the fact that both formulae cover the class of ELS values. □

Proof of Proposition 3

Necessity: Let $T \subsetneq N$, $T \neq \emptyset$, $i \in N \setminus T$. First, we determine the payoff $\mathrm{DR}_i^b(u_T)$. One obtains

$$\begin{aligned}
\mathrm{DR}_i^b(u_T) &\overset{(4)}{=} \frac{u_T(N)}{|N|} + \sum_{S \subsetneq N\setminus\{i\}} \frac{b_{|S|+1} \cdot u_T(S \cup \{i\})}{\binom{|N|}{|S|+1}(|S|+1)} \\
&- \sum_{S \subseteq N\setminus\{i\}: S \neq \emptyset} \frac{b_{|S|} \cdot u_T(S)}{\binom{|N|}{|S|+1}(|S|+1)} \\
&\overset{i \in N\setminus T}{=} \frac{1}{|N|} + \sum_{T \subseteq S \subsetneq N\setminus\{i\}} \frac{b_{|S|+1}}{\binom{|N|}{|S|+1}(|S|+1)} \\
&- \sum_{T \subseteq S \subseteq N\setminus\{i\}: S \neq \emptyset} \frac{b_{|S|}}{\binom{|N|}{|S|+1}(|S|+1)} \\
&= \frac{1}{|N|} + \sum_{s=|T|}^{|N|-2} \binom{|N|-1-|T|}{s-|T|} \frac{b_{s+1}}{\binom{|N|}{s+1}(s+1)} \\
&- \sum_{s=|T|}^{|N|-1} \binom{|N|-1-|T|}{s-|T|} \frac{b_s}{\binom{|N|}{s+1}(s+1)} \\
&= \frac{1}{|N|} + \frac{1}{(|N|-|T|)\binom{|N|}{|T|}} \left(\sum_{s=|T|}^{|N|-2} \binom{s}{|T|} \cdot b_{s+1} - \sum_{s=|T|}^{|N|-1} \binom{s}{|T|} \cdot b_s \right) \\
&= \frac{1}{|N|} - \frac{b_{|T|}}{(|N|-|T|)\binom{|N|}{|T|}} + \frac{1}{(|N|-|T|)\binom{|N|}{|T|}} \sum_{s=|T|+1}^{|N|-1} \left(\binom{s-1}{|T|} - \binom{s}{|T|} \right) \cdot b_s \\
&= \frac{1}{|N|} - \frac{b_{|T|}}{(|N|-|T|)\binom{|N|}{|T|}} - \frac{1}{(|N|-|T|)\binom{|N|}{|T|}} \sum_{s=|T|+1}^{|N|-1} \binom{s-1}{|T|-1} \cdot b_s \\
&= \frac{1}{|N|} - \frac{1}{(|N|-|T|)\binom{|N|}{|T|}} \sum_{s=|T|}^{|N|-1} \binom{s-1}{|T|-1} \cdot b_s .
\end{aligned}$$

In view of (7), the claim is immediate.

Sufficiency follows from the fact that both formulae cover the class of ELS values. □

References

Banzhaf, J. F. (1965). Weighted voting does not work: A mathematical analysis. *Rutgers Law Review, 19*, 317–343.

Casajus, A. (2012). Solidarity and fair taxation in TU games. Working paper 111, Faculty of Economics and Management Science, University of Leipzig, Germany. URL https://ideas.repec.org/p/zbw/leiwps/111.html

Casajus, A., & Huettner, F. (2013). Null players, solidarity, and the egalitarian Shapley values. *Journal of Mathematical Economics, 49*, 58–61.

Chameni Nembua, C. (2012). Linear efficient and symmetric values for TU-games: Sharing the joint gain of cooperation. *Games and Economic Behavior, 74*, 431–433.

Chameni Nembua, C., & Andjiga, N. G. (2008). Linear, efficient and symmetric values for TU-games. *Economics Bulletin, 3*(71), 1–10.

Driessen, T., & Funaki, Y. (1991). Coincidence of and collinearity between game theoretic solutions. *OR Spectrum, 13*(1), 15–30.

Driessen, T., & Radzik, T. (2003). Extensions of Hart and Mas-Colell's consistency to efficient, linear, and symmetric values for TU-games. In L. A. Petrosjan & D. W. K. Yeung (Eds.), *ICM millennium lectures on games* (pp. 147–166). Springer.

Harsanyi, J. C. (1959). A bargaining model for cooperative *n*-person games. In A. W. Tucker, & R. D. Luce (Eds.), Contributions to the theory of games IV (Vol. 2, pp. 325–355). Princeton University Press

Hernandez-Lamoneda, L., Juarez, R., & Sanchez-Sanchez, F. (2008). Solutions without dummy axiom for TU cooperative games. *Economics Bulletin, 3*(1), 1–9.

Joosten, R. (1996). Dynamics, equilibria and values. Ph.D. thesis, Maastricht University, The Netherlands

Joosten, R., Peters, H. J. M., & Thuijsman, F. (1994). Socially acceptable values for transferable utility games. Report M94-03, Department of Mathematics, University of Maastricht, The Netherlands

Ju, Y., Borm, P., & Ruys, P. (2007). The consensus value: A new solution concept for cooperative games. *Social Choice and Welfare, 28*, 685–703.

Kalai, E., & Samet, D. (1987). On weighted Shapley values. *International Journal of Game Theory, 16*(3), 205–222.

Levinský, R., & Silársky, P. (2004). Global monotonicity of values of cooperative games: An argument explaining the explanatory power of Shapley's approach. *Homo Oeconomicus, 20*, 473–492.

Malawski, M. (2013). "Procedural" values for cooperative games. *International Journal of Game Theory, 42*(1), 305–324.

Maschler, M., & Peleg, B. (1966). A characterization, existence proof and dimension bounds for the kernel of a game. *Pacific Journal of Mathematics, 18*(2), 289–328.

Nowak, A. S., & Radzik, T. (1994). A solidarity value for *n*-person transferable utility games. *International Journal of Game Theory, 23*, 43–48.

Owen, G. (1975). Multilinear extensions and the Banzhaf value. *Naval Research Logistics Quarterly, 22*, 741–750.

Radzik, T. (2021). On a new method of analyzing properties of efficient, symmetric and linear values of TU-games. *Applicationes Mathematicae, 48*, 173–208.

Radzik, T., & Driessen, T. (2009). Socially acceptable values for cooperative TU-games. Working paper, Institute of Mathematics and Computer Science, Wroclaw University of Technology, Poland

Radzik, T., & Driessen, T. (2013). On a family of values for TU-games generalizing the Shapley value. *Mathematical Social Sciences, 65*(2), 105–111.

Ruiz, L. M., Valenciano, F., & Zarzuelo, J. M. (1996). The least square prenucleolus and the least square nucleolus. Two values for TU games based on the excess vector. *International Journal of Game Theory, 25*, 113–134.

Ruiz, L. M., Valenciano, F., & Zarzuelo, J. M. (1998). The family of least square values for transferable utility games. *Games and Economic Behavior, 24*, 109–130.
Shapley, L. S. (1953). A value for *n*-person games. In H. Kuhn & A. Tucker (Eds.), *Contributions to the theory of games II* (pp. 307–317). Princeton University Press.
Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory, 14*, 65–72.

Analyzing the Zerkani Network with the Owen Value

Encarnación Algaba, Andrea Prieto, Alejandro Saavedra-Nieves, and Herbert Hamers

Abstract This paper introduces a new centrality measure based on the Owen value to rank members in covert networks. In particular, we consider the Zerkani network responsible for the Paris attack of November 2015 and the Brussels attack of March 2016. We follow the line of research introduced in Hamers et al. [Handbook of the Shapley value. Taylor and Francis Group: CRC Press, pp 463–481 (2019)]. First, we consider two different appropriate cooperative games defined on the Zerkani network. Both games take into account the strengths of the links between its members and the individual contribution of its members. Second, for each game the Owen value is calculated, that provides a ranking of the members in the Zerkani network. For this calculation, we need to create a suitable partition of the members in the network, and, subsequently, we will use the approximation method introduced in Saavedra-Nieves et al. [The mathematics of the uncertain: A tribute to Pedro Gil. Springer, pp 347–356 (2018)]. Moreover, we can provide specific error bounds for the approximation of the Owen value. Finally, the obtained rankings are compared to the rankings established in Hamers et al. [Handbook of the Shapley value. Taylor and Francis Group: CRC Press, pp 463–481 (2019)].

E. Algaba (✉)
Matemática Aplicada II and Instituto de Matemáticas de la Universidad de Sevilla, Sevilla, Spain
e-mail: ealgaba@us.es

A. Prieto
Matemática Aplicada II, Escuela Superior de Ingenieros, Camino de los Descubrimientos, Sevilla, Spain
e-mail: aprietogarcia@us.es

A. Saavedra-Nieves
Departamento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Santiago de Compostela, Spain
e-mail: alejandro.saavedra.nieves@usc.es

H. Hamers
CentER, Department of Econometrics and Operations Research and TIAS Business School, Tilburg University, Tilburg, The Netherlands
e-mail: H.J.M.Hamers@uvt.nl

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_14

1 Introduction

The last decades covered organizations have executed many massive attacks. We mention the 9/11 attack in 2001 by Al-Qaeda [cf. (Krebs, 2002)], the Bali bombing in 2002 by Jemaah Islamiyah [cf. (Wise, 2005)], and the mass shooting and suicide attack in Paris and Brussels in 2015 and 2016, respectively, by the Zerkani network [cf. (Basra & Neumann, 2016)]. The main goal of governments and their intelligence services is to neutralize potential attacks of possible covert networks using information from such a network, to prevent damage to society in terms, primarily, of human lives. For this reason, it is essential to identify the role of each member within such a network with the aim to expose their leaders. It is clear that information management is necessary but not always available. Knowledge of this information enables law enforcement agencies to advance their investigations and to proceed in an unhurried manner, maximising their effectiveness and minimizing the network's capacity to act.

A covert network can be represented by a graph, where its members become the nodes and the edges represent the interaction between each pair of individuals. Subsequently, standard graph theoretical centrality measures can be used to create a ranking of its members [cf. (Koschade, 2006; Sparrow, 1991; Klerks, 2001), or (Farley 2003)] despite not being originally intended for this purpose. The main limitation of these standard centrality measures is the fact that they only take into account the structure of the network. On the one hand, these measures do not take into account the importance of the interaction between members (e.g., kind of communication, as cell phone, Internet or exchange devices, among others) and, on the other hand, the individual importance (e.g., financial means, skills to make an improvised explosive device, known in acronyms as an IED). These aspects were taking into account in Lindelauf et al. (2013), Husslage et al. (2015), van Campen et al. (2018) and Hamers et al. (2019) by using cooperative game theory and applied to several covert networks. In fact, in these papers, a ranking of the members of the network according to its relevance was determined from the Shapley value (Shapley, 1953), a value used in a wide range of applications [cf. (Algaba et al., 2019)].

In this paper, we introduce a centrality measure based on the Owen value (Owen, 1977), an extension of the Shapley value. The Owen value is defined for cooperative games with a priori unions. This also enables to take into account possible cooperative restrictions among the members in a network, unlike the Shapley value, in addition to the strength of the links and the individual influence. The Owen value is applied in several real-life settings in which an exact expression of the Owen value can be derived. We mention airport games (Vázquez-Brage et al., 1997), maintenance of railways infrastructures (Costa, 2016), and methodologies for computing solutions for weighted multiple majority games [cf. (Algaba et al., 2003, 2007; Alonso-Meijide & Bowles, 2005)], as well as the development of heuristic and exact solutions to find voting systems that generate a power distribution, called the inverse power index problem (Kurz & Napel, 2014). Unfortunately, similar to the Shapley value, in general, the computational complexity of the Owen value increases exponentially with the number of players. Sampling techniques are often introduced to create approximations. Following the ideas in Castro et al. (2009) for the Shapley value approximation, Saavedra-Nieves et al. (2018) adapt them for the estimation of the Owen value and provide both theoretical and experimental results for such procedure.

As an application, we will focus on the Zerkani network. On November 13, 2015, the Zerkani network carried out simultaneous attacks in Paris (France), killing more than a hundred people and wounding hundreds. Few months later, on March 22, 2016, part of those who were involved behind of the Paris attacks managed to launch another massive attack in Brussels, detonating suicide bombs at Zaventem International Airport and in Maelbeek subway station, killing thirty-two people and injuring many more. The main figures responsible for the tactical operations of the Paris and Brussels' attacks were Abdelhamid Abaaoud and jihadist recruiter Khalid Zerkani. The Zerkani network provided personnel, training, planning, attack, escape and evasion. The Zerkani network is analyzed in Hamers et al. (2019), providing a ranking based on an approximation of the Shapley value. We focus on ranking the members of the Zerkani network considering the existence of different degrees of relationships among them. In this sense, these affinities are modeled by several cooperative games with a priori unions when establishing an a priori coalitional structure. Specifically, we will use the Owen value as a mechanism of ranking.

This paper is organized as follows. Section 2 discusses the Owen value and the approximation method of Saavedra-Nieves et al. (2018). The network analysis which uses two cooperative games is explained in Sect. 3. In the subsequent section, the analysis of the Zerkani network is presented and the final section concludes.

2 The Owen Value and an Approximation Algorithm

This section recalls the definition of the Shapley value, the Owen value and the sampling method for its approximation, as well as their error bounds introduced in Saavedra-Nieves et al. (2018).

A *transferable utility game*, or TU-game, is a pair (N, v), where $N = \{1, 2, ..., n\}$ is the set of players (the grand coalition) and v is a map that assigns a value $v(S)$ to each coalition $S \subseteq N$ such that $v(\emptyset) = 0$. The set of all cooperative games with player set N is denoted by $\mathcal{G}^N$. We say that a game (N, v) is non-negative if $v(S) \geq 0$, for all $S \subseteq N$. A game (N, v) is said to be monotonic if for all $S \subseteq T \subseteq N$ then $v(S) \leq v(T)$. A game is called superadditive if for all $A, B \subseteq N$ with $A \cap B = \emptyset$, then $v(A \cup B) \geq v(A) + v(B)$.

One of the most important solutions concepts for cooperative games is the Shapley value (Shapley, 1953), see Algaba et al. (2019) for a wide range of theoretical and applied results about this value. The *Shapley value* assigns to each $(N, v) \in \mathcal{G}^N$, and each $i \in N$, the real number defined by

$$Sh_i(N, v) = \frac{1}{|\Pi(N)|} \sum_{\pi \in \Pi(N)} m_v^\pi(i), \tag{1}$$

where $\Pi(N)$ is the set of all possible permutations of N, and $m_v^\pi(i)$ denotes the marginal contribution of player i in a given permutation $\pi \in \Pi(N)$, which is defined as

$$m_v^\pi(i) = v\left(P_i^\pi \cup \{i\}\right) - v\left(P_i^\pi\right),$$

P_i^π being the set of predecessors of i in π, i.e., $P_i^\pi = \{j \in N | \pi(j) < \pi(i)\}$.

A cooperative game with a priori unions is given by a triplet (N, v, P) where $(N, v) \in \mathcal{G}^N$ and $P = \{P_1, ..., P_m\}$ is a partition of N. In this case, we assume that P is interpreted as the coalition structure that restricts the cooperation among the players in N. The set of all cooperative games with a priori unions with set of players N is denoted by $\mathcal{U}^N$.

The *Owen value* (Owen, 1977) assigns to each $(N, v, P) \in \mathcal{U}^N$ and each $i \in N$, the real number defined by

$$O_i(N, v, P) = \frac{1}{|\Pi_P(N)|} \sum_{\pi \in \Pi_P(N)} m_v^\pi(i), \tag{2}$$

where $\Pi_P(N)$ denotes the set of all permutations of N which are compatible with a coalition structure P, meaning that the elements of each union of P are not separated by π, i.e., $\pi \in \Pi_P(N)$ if and only if, for all $i, j, k \in N$, $P_h^\pi \in P$, it holds that

$$\text{if } i, j \in P_h^\pi \text{ and } \pi(i) < \pi(k) < \pi(j), \text{ then } k \in P_h^\pi.$$

The idea of distributing the worth of cooperation of N is made in two stages. First, the profit is allocated among the unions using the Shapley value, and then, the amount received per union is distributed among the members belonging to it, by using again the Shapley value. Observe, if the coalition structure is formed by unitary unions or only by the grand coalition, the Owen value and the Shapley value prescribe the same allocation.

Although the notion of marginal contribution of a player is intuitively clear, computing the Owen value becomes a hard task when the amount of players involved in the cooperative game substantially increases.

In what follows, we formally describe an algorithm introduced in Saavedra-Nieves et al. (2018) for estimating the Owen value for a cooperative game with a priori unions (N, v, P), with $P = \{P_1, ..., P_m\}$, based on simple random sampling with replacement on the set of compatible permutations.[1]

[1] This methodology ensures that each $\pi \in \Pi_P(N)$ is equally likely. Alternatively, other sampling techniques, as stratified sampling, can be considered, that implies that not all permutations are taken with equal probability, see for instance Castro et al. (2017).

The steps of the sampling procedure are the ones depicted below:

1. The sampling population $\mathcal{S}$ is the set of all permutations of N compatible with P, i.e., $\Pi_P(N)$.
2. The vector of parameters to be estimated is $O = (O_i)_{i\in N}$, where O_i denotes $O_i(N, v, P)$, for all $i \in N$.
3. The characteristics under study for each sampling unit $\pi \in \Pi_P(N)$ are the marginal contributions of the players according to π, i.e., the vector

$$(m_v^\pi(i))_{i\in N} = (v\left(P_i^\pi \cup \{i\}\right) - v\left(P_i^\pi\right))_{i\in N}.$$

4. Each permutation $\pi \in \Pi_P(N)$ is taken with the same probability: a random permutation of the elements of each P_k, with $k \in \{1, ..., m\}$, is chosen and then, it is selected a random permutation of the elements of $\{1, ..., m\}$, i.e., of elements of P. Joining this collection of $m + 1$ permutations, we obtain $\pi \in \Pi_P(N)$.
5. The mean of the marginal contribution vectors over the sample $\mathcal{S}$ corresponds to the Owen value estimation, i.e., $\hat{O} = (\hat{O}_i)_{i\in N}$, such that

$$\hat{O}_i = \frac{1}{\ell}\sum_{\pi\in\mathcal{S}} m_v^\pi(i),$$

for all $i \in N$, where $\ell = |\mathcal{S}|$ equals the sample size.

A fundamental issue in the problem focuses on bounding the error of the estimation, that is often not possible to be measured in practice. For this reason, the following probabilistic bound is theoretically provided in Saavedra-Nieves et al. (2018) instead:

$$\mathbb{P}(|\hat{O}_i - O_i| \geq \varepsilon) \leq \alpha, \text{ with } \varepsilon > 0 \text{ and } \alpha \in (0, 1].$$

Theorem 1 *Saavedra-Nieves et al. (2018) Let $\varepsilon > 0$, $\alpha \in (0, 1)$, (N, v, P) be a cooperative game with a priori unions, and $r_i = \max_{\pi,\pi'\in\Pi_P(N)} (m_v^\pi(i) - m_v^{\pi'}(i))$.*

$$\text{If } \ell \geq \min\left\{\frac{1}{4\alpha\varepsilon^2}, \frac{\ln(2/\alpha)}{2\varepsilon^2}\right\} r_i^2 \text{ then, } \mathbb{P}(|\hat{O}_i - O_i| \geq \varepsilon) \leq \alpha \text{ for every } i \in N. \quad (3)$$

As illustration, take $r_i = 300$ for all $i \in N$. Then, Table 1 shows the theoretical errors with respect to (3).

Thus, the estimated Owen value usually becomes a good approximation of the real one when sampling sizes sufficiently enlarge.

Table 1 Theoretical errors (ε) for $\ell = 10^3$, $\ell = 10^5$ and $\ell = 10^6$

	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
$\ell = 10^3$	11.61068	12.88408	15.44099
$\ell = 10^5$	1.16107	1.28841	1.54410
$\ell = 10^6$	0.36716	0.40743	0.48828

3 A New Game Theoretic Centrality Measure

In this section, we follow the approach in Hamers et al. (2019) in which two classes of cooperative games are considered that take into account the structure of the network, the relational and the individual strength of the members in the network. For both classes, unlike Hamers et al. (2019), we will determine a ranking of the Zerkani network using the Owen value in Sect. 4.

A covert network can be represented by an undirected graph $G = (N, E)$, where the node set N represents the set of members of the network and the set of links E denotes all relationships between these members. A relationship between member i and j is denoted by ij, with $ij \in E$. For a coalition $S \subseteq N$, the subnetwork G_S consists of the members of S and its links in E, i.e., $G_S = (S, E_S)$ where $E_S = \{ij \in E : i, j \in S\}$. A coalition $S \subseteq N$ is said to be a connected coalition, if the network G_S is connected; otherwise, S is called disconnected.

The influence of individuals in $G = (N, E)$ is represented by a set of weights on player set N, i.e., $\mathcal{I} = \left\{w_j\right\}_{j \in N}$ with $w_j \geq 0$, and the relational strength between members of the network is given by a set of weights on the edges E, i.e., $\mathcal{R} = \{k_{lh}\}_{lh \in E}$ with $k_{lh} \geq 0$.

The *weighted connectivity game (wconn)* (N, v^{wconn}) with respect to $G = (N, E)$ based on $\mathcal{I}$ and $\mathcal{R}$ is defined as in Hamers et al. (2019). Thus, for any connected coalition S, (N, v^{wconn}) is given by

$$v^{\text{wconn}}(S) = f(S, \mathcal{I}, \mathcal{R}), \tag{4}$$

f being a context specific and tailor-made non-negative function depending on S, $\mathcal{I}$ and $\mathcal{R}$. Function f measures the effectiveness of coalitions in the network to best reflect the situation and information at hand.

For a disconnected coalition S, the characteristic function is defined by

$$v^{\text{wconn}}(S) = \max_{T \in \Sigma_S} v^{wconn}(T), \tag{5}$$

with Σ_S the set of components (i.e., maximal connected coalitions) in G_S. The value of each disconnected coalition is based on the most effective component of this coalition.

An *additive weighted connectivity game (awconn)* (N, v^{awconn}) with respect to network $G = (N, E)$ based on $\mathcal{I}$ and $\mathcal{R}$ is, for a connected coalition S, defined by (4), identically to the corresponding weighted connectivity cooperative game (N, v^{wconn}), i.e., $v^{\text{awconn}}(S) = v^{\text{wconn}}(S)$, for all connected coalition $S \subseteq N$.

The difference between these two games is the definition of the disconnected coalitions. For a disconnected coalition S, the worth of S in an additive weighted connectivity game is given by

$$v^{\text{awconn}}(S) = \max_{A \in P(S)} \sum_{T \in A} v^{\text{wconn}}(T), \tag{6}$$

with $P(S)$ the set of all possible partitions of coalition S in connected coalitions. In general, this game is different from the one considered in Hamers et al. (2019),[2] where the value of disconnected coalitions S equals $v^{\text{awconn}}(S) = \sum_{T \in \Sigma_S} v^{\text{wconn}}(T)$, i.e., the sum of the values on the set of components of coalition S, denoted by Σ_S. Hence, notice that in the additive weighted connectivity game, we look for the most effective partition in connected coalitions, for each given coalition.

4 The Owen Value Approximation in the Zerkani Network

In this section, we introduce an explicit expression for f in the Zerkani network. Recall that f depends on the coalition at hand S, the importance of the individuals ($\mathcal{I}$) and the strength of the relationship between the links ($\mathcal{R}$). Subsequently, we will approximate the Owen value for the weighted connectivity game and the additive weighted connectivity game, respectively, and we will provide an error bound for these approximations. Finally, we will analyze the rankings obtained from both approximations.

4.1 The Zerkani Network Analysis

This section focuses on ranking the members of the Zerkani network by approximating the Owen value for the *wconn* and *awconn* cooperative games, respectively. Before determining these approximations, we need to define f to obtain explicit expressions for the TU-games in (5) and (6). Here, we follow Hamers et al. (2019), where $f(S, \mathcal{I}, \mathcal{R})$ is defined by

$$f(S, \mathcal{I}, \mathcal{R}) = \begin{cases} \left(\sum_{j \in S} w_j\right) \cdot \max_{lh \in E_S} k_{lh}, & \text{if } |S| > 1, \\ w_S, & \text{if } |S| = 1. \end{cases} \tag{7}$$

[2] Conditions to assure the equality between both games can be found in Algaba et al. (2001).

From this definition and considering the weights on the links (edges) and on the members (nodes) of the Zerkani network given in Table 2, the following result follows, immediately, for both games: If $f(S, \mathcal{I}, \mathcal{R})$ is defined by (7), then weighted connectivity games and additive weighted connectivity games are non-negative and monotonic games.

Following Algaba et al. (2001), since the weighted connectivity game, defined for the Zerkani network, is superadditive on the set of connected coalitions and, the fact that, in this setting, the set of components of a coalition S form a partition of this coalition, we have

$$v^{\text{awconn}}(S) = \max_{A \in P(S)} \sum_{T \in A} v^{\text{wconn}}(T) = \sum_{T \in \Sigma_S} v^{\text{wconn}}(T), \tag{8}$$

where $P(S)$ is the set of all possible partitions of coalition S in connected coalitions and Σ_S the set of components (i.e., maximal connected coalitions) of coalition S. This last expression coincides with the game introduced in Myerson (1977) to analyze the cooperation, when there are restrictions in communication defined by an undirected graph, which has been also used in more general settings of restricted cooperation as van den Nouweland et al. (1992) or Algaba et al. (2000), among others.

The Zerkani network consists of 47 players obtained from Gartenstein-Ross et al. (2016). Since the considered games does not lead to a clean expression for the Owen value which enables its computation in polynomial time, we will approximate the Owen value using the method described in Saavedra-Nieves et al. (2018) (see Sect. 2). We will focus on the top 10 of the rankings, although we completely rank all 47 individuals. The associated graph of the Zerkani network is displayed in Fig. 1.

We consider the weights used in Hamers et al. (2019) on the links (edges) and on the members (nodes) of the network. Initially, it was assigned a weight equal to one to each link/member. Later, these weights were increased following Table 2.

After its usage, the following members have a weight larger than one: Abdelhamid Abaaoud, Fabien Clain (weights equal to 4, respectively), Khalid Zerkani (5), Miloud F. (2) and Mohamed Belkaid (3).

We conclude this subsection by using the explicit expression of f in the setting of the Zerkani network in Theorem 1. Consider the weighted connectivity game (N, v^{wconn}). Then,

$$\begin{aligned} v^{\text{wconn}}(S \cup \{i\}) - v^{\text{wconn}}(S) &\leq v^{\text{wconn}}(S \cup \{i\}) \\ &\leq v^{\text{wconn}}(N), \end{aligned} \tag{9}$$

where the first inequality holds because $v^{\text{wconn}}(S) \geq 0$, for all $S \subseteq N$ and the second by monotonicity of the game v^{wconn}.

Since the marginal contributions satisfy $0 \leq m^{\pi}_{v^{\text{wconn}}}(i) \leq v^{\text{wconn}}(N)$, for all $i \in N$, it follows immediately that for all $i \in N$, we substitute r_i, the parameter in Theorem 1, by $v^{\text{wconn}}(N)$, i.e., $r_i = v^{\text{wconn}}(N)$. This implies that, by using Theorem 1 on (N, v^{wconn}), the following theorem directly results.

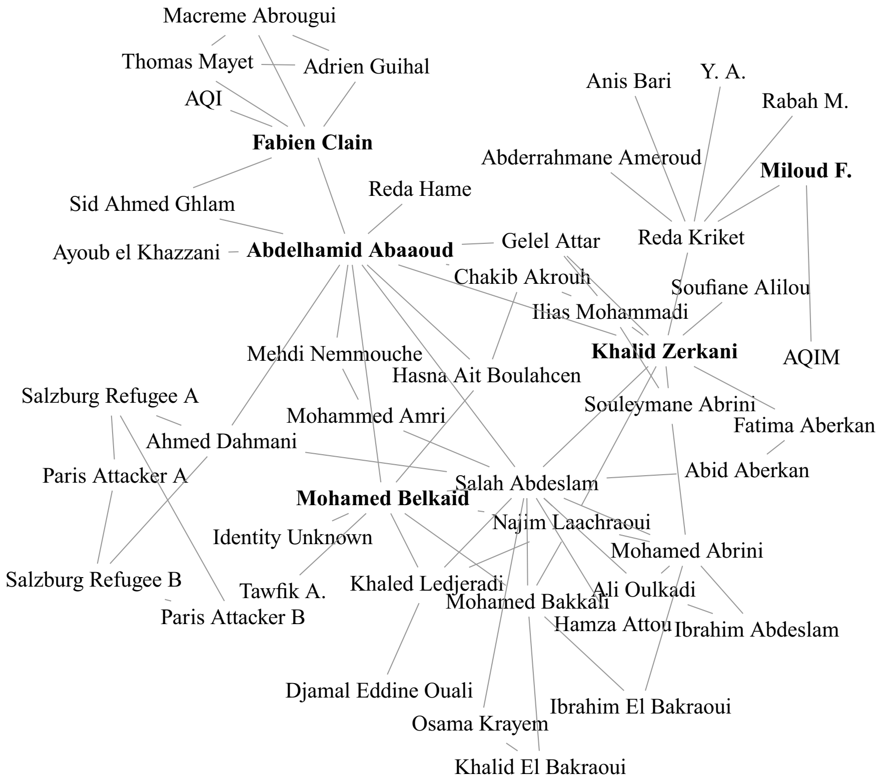


Fig. 1 Graph of the Zerkani network

Table 2 List of relationships, weights for links and weights for starting nodes

Relationships	Weights on links	Extra weight for starting nodes
“Associate of”	2	0
“Brother of”	1	0
“Commander of”	2	2
“Family relationship”	1	0
“Funded”	1	2
“Lived with”	2	0
“Nephew of”	1	0
“Recruiter of”	1	1
“Supporter of”	1	1
“Traveled to Syria with”	2	0
“Traveled with”	2	0
“Associate and traveled with”	4	0
“Traveled and lived with”	4	0

Theorem 2 *Let $\varepsilon > 0$, $\alpha \in (0, 1)$, (N, v^{wconn}, P) be a weighted connectivity game with a priori unions P, $r_i = v^{wconn}(N)$, for all $i \in N$.*

$$\text{If } \ell \geq \min\left\{\frac{1}{4\alpha\varepsilon^2}, \frac{\ln(2/\alpha)}{2\varepsilon^2}\right\}(v^{wconn}(N))^2, \tag{10}$$

then $\mathbb{P}(|\hat{O}_i - O_i| \geq \varepsilon) \leq \alpha$, for every $i \in N$.

The same result, using the same arguments, holds for additive weighted connectivity games with a priori unions $(N, v^{\text{awconn}}, P)$. Observe that taking into account (10), the resulting bound, for the sample size ℓ, does not depend on the partition P considered in the game (N, v^{wconn}, P).

4.2 On the Partition P Considered

In this subsection, we provide a partition of the Zerkani network that integrates affinities or certain features of the terrorists to achieve enhanced outcomes. The partition in our case has ten unions, i.e., $P = \{P_1, P_2, \ldots, P_{10}\}$. Union P_1 groups those high ranked members of the Zerkani network, that is, those majors and those devoted to recruit terrorist. They are Abdelhamid Abaaoud, Fabien Clain and Khalid Zerkani with an associated weight larger than one. Union P_2 corresponds to the associated upper-level charges of the Zerkani network. This group is formed by Chakib Akrouh, Gelel Attar, Hasna Ait Boulahcen, Fatima Aberkan, Osama Krayem, Souleymane Abrini, Ayoub el Khazzani, Mehdi Nemmouche, Thomas Mayet, Macreme Abrougui, Ahmed Dahmani and Adrien Guihal. Union P_3 groups those individuals that are considered in an inferior rank (those recruited or under the authority of a major). Their names (or alias) are Sid Ahmed Ghlam, Reda Hame, AQI, Ilias Mohammadi, Soufiane Alilou, Najim Laachraoui and Khalid El Bakraoui. Unions P_4, P_5 and P_6 involve those terrorists that created strong relationships or they may have hidden intentions when traveling. Those terrorists under alias Paris Attacker A, Paris Attacker B, Salzburg Refugee A and Salzburg Refugee B form union P_4; Mohammed Amri, Hamza Attou, Mohamed Abrini and Abid Aberkan form P_5, and Mohamed Belkaid, Salah Abdeslam, Mohamed Bakkali and Ibrahim El Bakraoui belong to P_6. The Kriket network is captured in union P_7. They are Reda Kriket, Rabah M., Y. A., Abderrahmane Ameroud, Miloud F., Anis Bari and AQIM. The individuals associated with $\mathit{Forging\ Ring}$ form union P_8 and corresponds Khaled Ledjeradi and Djamal Eddine Ouali. Union P_9 is given by a person with unknown identity and Tawfik A and union P_{10} is formed by Ibrahim Abdeslam and Ali Oulkadi. Each of them is given by the two individuals who were arrested simultaneously in Forest.

4.3 Numerical Results

As mentioned, the main purpose is to rank the members of the Zerkani network according to the decreasing order using the outcomes corresponding to the approximations of the Owen value.

Thus, we apply the approximation method described in Sect. 2 to the TU-games (N, v^{wconn}) and (N, v^{awconn}), respectively. The sample size in the approximation method is chosen equal to 1000, i.e., $\ell = 1000$. Subsequently, we have simulated 1000 times the Owen value. In this way, we obtain 1000 approximations of the Owen value for both games. We also consider the quality of the approximation using Theorem 2. Finally, we look for both games the Owen value obtained by averaging all of these 1000 approximations. The final result is equivalent to obtain one simulation with $\ell = 10^6$.

Using the input of Table 2, the network structure, its relations and the definition of the (additive) weighted connectivity game, we can deduce that $v(N) = 300$. Hence, in Theorem 2 we choose $r_i = 300$, for all $i \in N$. Therefore, Table 1, displayed in Sect. 2, provides the theoretical errors for estimating the Owen value in the Zerkani network by using the cooperative games (N, v^{wconn}) and (N, v^{awconn}).

Table 3 depicts the top-10 list of terrorists belonging to the Zerkani network and the corresponding results for the games (N, v^{wconn}) and (N, v^{awconn}). The overall list of the Owen value estimations for the members of the Zerkani network can be also supplied on request. Khalid Zerkani, the most influential terrorist, under the weighted connectivity game, goes to the second position in the additive case. Abdelhamid Abaaoud moves to the fourth position under the second point of view. However, Mohamed Belkaid moves up from the third position in the weighted connectivity game to the first position under additivity. Analogous comments can be extracted from the remainder of the list of members in the top 10, who most of them are in the same positions (in fact, at most, they differ in one position).

Once the Owen value is estimated, we check how the sampling proposal for approximating the Owen value performs in this example, in terms of the variability of the results, through a small simulation study. By construction, we have obtained the results in Table 3 by averaging 1000 estimations of the Owen value for the Zerkani network by using sample sizes equal to $\ell = 10^3$. We separately focus our attention on those 10 terrorists belonging to the top of the considered rankings.

Table 4 summarizes, from a purely statistical point of view, the 1000 obtained results for the 10 most relevant terrorists in the Zerkani network by using the estimated Owen value with $\ell = 10^3$ for the weighted connectivity game. Notice that the order established for the top 10 in Table 3 can be also maintained when using as criteria the main statistical measures.

Analogous conclusions can be obtained from the case of the additive approach, in view of the statistical summary for the 1000 estimations of the Owen value for the 10 terrorists in the top 10 in the ranking of Table 3. The numerical results are included in Table 5.

Table 3 Top 10 of the ranking of terrorists in the Zerkani network, according to the estimated Owen value for games (N, v^{wconn}) and (N, v^{awconn}) with $\ell = 10^6$

	Ranking R_{wconn}		Ranking R_{awconn}	
Pos.	Terrorist	$\hat{O}$	Terrorist	$\hat{O}$
1	Khalid Zerkani	39.357363	Mohamed Belkaid	28.373644
2	Abdelhamid Abaaoud	36.255303	Khalid Zerkani	27.728460
3	Mohamed Belkaid	29.162472	Mohamed Bakkali	27.133642
4	Mohamed Bakkali	27.802943	Abdelhamid Abaaoud	26.225820
5	Salah Abdeslam	27.070990	Salah Abdeslam	22.451192
6	Fabien Clain	13.569693	Fabien Clain	16.427384
7	Reda Kriket	11.621222	Reda Kriket	11.395351
8	Ahmed Dahmani	10.808112	Ahmed Dahmani	6.167458
9	Khaled Ledjeradi	4.685000	Miloud F.	6.089226
10	Miloud F.	4.203003	Khaled Ledjeradi	5.268594

Table 4 Statistical summary of the 1000 estimations of the Owen value for the *wconn* game

	Terrorist	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	Khalid Zerkani	36.833	38.781	39.338	39.357	39.902	42.167
2	Abdelhamid Abaaoud	33.959	35.820	36.271	36.255	36.709	39.586
3	Mohamed Belkaid	26.581	28.470	29.202	29.163	29.768	32.169
4	Mohamed Bakkali	24.484	26.979	27.801	27.803	28.672	32.832
5	Salah Abdeslam	24.354	26.398	27.050	27.071	27.743	30.355
6	Fabien Clain	12.251	13.273	13.547	13.570	13.882	15.274
7	Reda Kriket	10.614	11.375	11.612	11.621	11.876	12.768
8	Ahmed Dahmani	9.853	10.575	10.805	10.808	11.045	12.078
9	Khaled Ledjeradi	4.270	4.602	4.689	4.685	4.766	5.035
10	Miloud F.	3.714	4.091	4.201	4.203	4.315	4.739

From these results, we also can check the variability of rankings. For instance, by taking the *awconn* game, we check that Khalid Zerkani moves up to the first position in the ranking when considering the minimum values. On the other hand, Mohamed Bakkali also moves up to the top of the ranking when the maximum estimations of the Owen value are considered.

We complete this analysis measuring the computational effort required in obtaining these 1000 estimations. Table 6 includes a summary of the processing times (in seconds) for each of these 1000 repetitions. Note that more than 75% of the estimations have been obtained in less than 2 h of real computing time.

Note that obtaining these estimations requires significant computational effort that can be substantially minimized by parallelising their implementation.

Table 5 Statistical summary of the 1000 estimations of the Owen value for the *awconn* game

	Terrorist	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	Mohamed Belkaid	25.912	27.736	28.385	28.374	28.973	31.464
2	Khalid Zerkani	26.328	27.420	27.739	27.729	28.029	29.069
3	Mohamed Bakkali	24.042	26.369	27.128	27.134	27.978	31.908
4	Abdelhamid Abaaoud	24.892	25.970	26.239	26.226	26.494	27.753
5	Salah Abdeslam	19.924	21.828	22.423	22.451	23.060	25.564
6	Fabien Clain	15.522	16.253	16.435	16.427	16.601	17.177
7	Reda Kriket	10.766	11.235	11.389	11.395	11.561	12.268
8	Ahmed Dahmani	5.730	6.056	6.166	6.168	6.276	6.732
9	Miloud F.	5.742	6.013	6.087	6.089	6.171	6.488
10	Khaled Ledjeradi	5.010	5.212	5.270	5.269	5.326	5.488

Table 6 Statistical summary of the 1000 processing times (in seconds) for the estimations of the Owen value

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Proc. time	2365.370	4965.364	5552.082	5637.076	6265.814	13,287.390

4.4 *A Brief Comparison with the Ranking Based on the Shapley Value*

In this section, we make a brief discussion on the rankings obtained for the estimations of the Owen value and the Shapley value, respectively. To this purpose, we use the sampling procedure considered in Castro et al. (2009), based on simple random sampling with replacement. By simplicity, and for the sole purpose of comparing the scenarios, we estimate 100 times the Shapley value by taking $\ell = 1000$ permutations of N. Thus, we average these estimations and we compare such result with the average of 100 of the estimations of the Owen value considered in Sect. 4.3. In both cases, the theoretical bounds of the error in their estimation, for all $i \in N$, are shown in the second row corresponding to Table 1.

Table 7 illustrates the top 10 of the rankings obtained. The complete list of terrorists based on these results can be supplied on request.

Below, we briefly make some comments on the resulting rankings of the top-10 members of the Zerkani network. We emphasize the fact that the most of the positions change when using the approach considered in this paper. Probably, this may be due to the organizational and logistical role played by such terrorists, which are now incorporated in the cooperative games through an a priori coalitional structure, and the fact that their weights will change with the approach under consideration. Undoubtedly, the possibility of integrating information on the cooperation and affinities of members of the network with the Owen value, makes that the Owen value can

Table 7 Top 10 of terrorists in the Zerkani network, according to the average of 100 estimations of the Owen value and the Shapley value with $\ell = 1000$ for games (N, v^{wconn}) and (N, v^{awconn})

Ranking R_{wconn}			
Terrorist	$\hat{Sh}$	Terrorist	$\hat{O}$
Ab. Abaaoud	17.108	Khalid Zerkani	39.242
Khalid Zerkani	15.026	Ab. Abaaoud	36.129
Salah Abdeslam	14.741	Mohamed Belkaid	29.236
Mohamed Belkaid	14.249	Mohamed Bakkali	27.845
Najim Laachraoui	7.918	Salah Abdeslam	27.042
Mohamed Bakkali	7.356	Fabien Clain	13.661
Fabien Clain	5.884	Reda Kriket	11.642
Reda Kriket	3.696	Ahmed Dahmani	10.754
Ahmed Dahmani	3.369	Khaled Ledjeradi	4.702
Mohamed Abrini	2.917	Miloud F.	4.209
Ranking R_{awconn}			
Terrorist	$\hat{Sh}$	Terrorist	$\hat{O}$
Mohamed Belkaid	13.987	Mohamed Belkaid	28.460
Khalid Zerkani	12.332	Khalid Zerkani	27.677
Ab. Abaaoud	11.850	Mohamed Bakkali	27.168
Salah Abdeslam	11.453	Ab. Abaaoud	26.157
Fabien Clain	8.295	Salah Abdeslam	22.439
Mohamed Bakkali	7.625	Fabien Clain	16.404
Najim Laachraoui	7.549	Reda Kriket	11.395
Reda Kriket	4.923	Ahmed Dahmani	6.142
Mohamed Abrini	2.996	Miloud F.	6.093
Miloud F.	2.827	Khaled Ledjeradi	5.271

be considered more representative than the Shapley value, in reference to the reality of the Zerkani network.

Notice Khalid Zerkani is leading the ranking when a priori unions system exist under the weighted connectivity game, with respect to the one associated to the case of the Shapley value, keeping his position for the additive approach. Recall that this man directed a recruitment network in the Brussels area. He was not present coordinating the attacks of Paris and Brussels, but he had a high influence on all those who related to him. He is currently imprisoned on terrorism-related charges. In the additive scenario, the person who always occupies the first position in both rankings is Mohamed Belkaid. On the other hand, Najim Laachraoui and Mohamed Abrini do not belong to the top 10 when using the Owen value in ranking with both games. Their positions are completed with Khaled Ledjeradi and Miloud F., when the Owen value is considered for the weighted connectivity game. However, under the additive approach, Najim Laachraoui and Mohamed Abrini are changed by Ahmed Dahmani and Khaled Ledjeradi in the lists.

Table 8 Spearman's correlation matrix for the rankings of the Zerkani network

		R_{wconn}		R_{awconn}	
		$\hat{Sh}$	$\hat{O}$	$\hat{Sh}$	$\hat{O}$
R_{wconn}	$\hat{Sh}$	1.000	0.927	**0.855**	0.830
	$\hat{O}$	0.927	1.000	0.889	**0.925**
R_{awconn}	$\hat{Sh}$	0.855	0.889	1.000	0.944
	$\hat{O}$	0.830	0.925	0.944	1.000

In general, these results are in line with the reality of the Zerkani network. Along with Khalid Zerkani and Abdelhamid Abaaoud, the role of Mohamed Bakkali is also important since he is the alleged intellectual author of the attack of Paris. It is believed that he selected those who were going to be in the war zones or in Europe. He died in a police raid, when Chakib Akrouh detonated his explosives belt. Another individual in the ranking is Salah Abdeslam. He was the most wanted man in Europe after the Paris attack. Fabien Clain was one planner of Paris attack and explored the different places where to perform the blows. Then, Reda Kriket was a recruiter for the network and provided money to it. Meanwhile, Khaled Ledjeradi was someone very required in the network, since he was the leader of an organization that created fake documents for the members of the network, allowing them to travel. Finally, about Miloud F. not much information is available, but he was arrested in Turkey in 2005, and this allowed for arresting Reda Kriket later. About Salzburg Refugee A and Salzburg Refugee B refugees, barely there is information about them, but they are the points of union of the attackers Paris A and B to the network. So, the supervision of these last two subjects mentioned may have been key to the cessation of the attack. The decision not to increase police surveillance on them, even if it was not a good one, can be justified from our results since only under the additive perspective, these individuals rank high in the ranking. Notice that those that carry out the suicide and therefore the action (Salzburg Refugee A or B, among others) usually appear from eleventh position onwards.

The degree of similarity of all rankings considered is now studied through the computation of Spearman's correlation coefficient on their positions (see Table 8). In view of such results, when comparing the Owen value and the Shapley value for both approaches, the level of association under the additive approach is slightly higher than under that one given by the weighted connectivity game. In addition, we check that the Owen values, obtained for both games, have a stronger association than the respective Shapley values (both correlations, next, in bold).

5 Conclusions

The Shapley value (Shapley, 1953) and the Banzhaf value (Banzhaf, 1965) are two of the most well known and studied values in the literature of cooperative games. The idea of this chapter is how to determine the key members in a terrorist network defined by a graph with tools of cooperative game theory. With this objective, the Owen value is presented as a new centrality measure for networks generalizing the studies made with the Shapley value in Hamers et al. (2019). The advantage of this new measure is the possibility of considering relevant information of the members of the network through a partition that realistically describes their affinities. Based on this same idea, in this framework, the Banzhaf–Owen value (Owen, 1982) has been studied in Algaba et al. (2022) extending likewise the Banzhaf value and presenting these values to analyze these networks. Therefore, when choosing the approach based on a priori unions, a previous and deep analysis of the network is needed to capture the most important features and information of it. This way, we can integrate this information in the results and obtain more precise outcomes in line with the income data.

Acknowledgements The authors wish to thank Sascha Kurz for his invitation to contribute with a chapter to this volume devoted to Stefan Napel in recognition of his remarkable research career. They also thank two anonymous referees for their valuable comments and suggestions.
Financial support from R&D&I Project Grant PGC2018-097965-B-I00, funded by MCIN/ AEI/ 10.13039/501100011033/ and by "ERDF A way of making Europe"/EU is gratefully acknowledged. A. Saavedra-Nieves acknowledges the financial support of FEDER/Ministerio de Ciencia, Innovación y Universidades - Agencia Estatal de Investigación under grant MTM2017-87197-C3-3-P, and of the Xunta de Galicia through the European Regional Development Fund (Grupos de Referencia Competitiva ED431C 2021/24). Authors also thank the computational resources of the Centro de Supercomputación de Galicia (CESGA).

References

Algaba, E., Fragnelli, V., & Sánchez-Soriano, J. (Eds.). (2019). *Handbook of the Shapley value*. Taylor and Francis Group: CRC Press.

Algaba, E., Prieto, A., & Saavedra-Nieves, A. (2022). Rankings in the Zerkani network by a game theoretical approach. Working paper 1-28,

Algaba, E., Bilbao, J. M., Borm, P., & López, J. J. (2000). The position value for union stable systems. *Mathematical Methods of Operations Research, 52,* 221–236.

Algaba, E., Bilbao, J. M., & López, J. J. (2001). A unified approach to restricted games. *Theory and Decision, 50,* 333–345.

Algaba, E., Bilbao, J. M., Fernández García, J. R., & López, J. J. (2003). Computing power indices in weighted multiple majority games. *Mathematical Social Sciences, 46*(1), 63–80.

Algaba, E., Bilbao, J. M., & Fernández, J. R. (2007). The distribution of power in the European Constitution. *European Journal of Operational Research, 176*(3), 1752–1766.

Alonso-Meijide, J. M., & Bowles, C. (2005). Generating functions for coalitional power indices: An application to the IMF. *Annals of Operations Research, 137*(1), 21–44.

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19,* 317–343.

Basra, R., & Neumann, P. (2016). Criminal pasts, terrorist futures: European jihadists and the new crime-terror nexus. *Perspectives on Terrorism, 10*(6), 25–40.
Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research, 36,* 1726–1730.
Castro, J., Gómez, D., Molina, E., & Tejada, J. (2017). Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation. *Computers & Operations Research, 82,* 180–188.
Costa, J. (2016). A polynomial expression for the Owen value in the maintenance cost game. *Optimization, 65*(4), 797–809.
Deng, X., & Papadimitriou, C. H. (1994). On the complexity of cooperative solution concepts. *Mathematics of Operations Research, 19*(2), 257–266.
Farley, J. D. (2003). Breaking AlQaeda cells: A mathematical analysis of counterterrorism operations. *Studies in conflict and terrorism, 26*(6), 299–314.
Fragnelli, V., García-Jurado, I., Norde, H., Patrone, F., & Tijs, S. (2000). How to share railways infrastructure costs? In *Game practice: Contributions from applied game theory* (pp. 91–101). Springer.
Gartenstein-Ross, D., & Barr, N. (2016, April 27). *Recent attacks illuminates the Islamic State's Europe attack network*. The Jamestown Foundation. https://jamestown.org/program/hot-issue-recent-attacks-illuminate-the-islamic-states-europe-attack-network/
Hamers, H., Husslage, B., & Lindelauf, R. (2019). Analysing ISIS Zerkani network using the Shapley value. In E. Algaba, V. Fragnelli, & J. Sánchez-Soriano (Eds.), *Handbook of the Shapley value* (pp. 463–481). CRC Press, Taylor and Francis Group.
Husslage, B., Borm, P., Burg, T., Hamers, H., & Lindelauf, R. (2015). Ranking terrorists in networks: A sensitivity analysis of Al Qaeda's 9/11 attack. *Social Networks, 42,* 1–7.
Klerks, P. (2001). The network paradigm applied to criminal organizations. *Connections, 24,* 53–65.
Koschade, S. (2006). A social network analysis of Jemaah Islamiyah: The applications to counterterrorism and intelligence. *Studies in Conflict and Terrorism, 29*(6), 559–575.
Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections, 24,* 43–52.
Kurz, S., & Napel, S. (2014). Heuristic and exact solutions to the inverse power index problem for small voting bodies. *Annals of Operations Research, 42,* 137–163.
Lindelauf, R., Hamers, H., & Husslage, B. (2013). Cooperative game theoretic centrality analysis of terrorist networks: The cases of Jemaah Islamiyah and Al Qaeda. *European Journal of Operational Research, 229,* 230–238.
Littlechild, S. C., & Owen, G. (1973). A simple expression for the Shapley value in a special case. *Management Science, 20*(3), 370–372.
Myerson, R. (1977). Graphs and cooperation in games. *Mathematics of Operations Research, 2,* 225–229.
Owen, G. (1977). Values of games with a priori unions. In R. Henn & O. Moeschlin (Eds.), *Mathematical economics and game theory* (pp. 76–88). Springer.
Owen, G. (1982). Modification of the Banzhaf-Coleman index for games with a priori unions. In M. J. Holler (Ed.), *Power, voting, and voting power* (pp. 232–238). Springer.
Saavedra-Nieves, A., García-Jurado, I., & Fiestras-Janeiro, M. G. (2018). Estimation of the Owen value based on sampling. In E. Gil, E. Gil, J. G. Gil, & M. Á. Gil (Eds.), *The mathematics of the uncertain: A tribute to Pedro Gil* (pp. 347–356). Springer.
Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games, 2*(28), 307–317.
Sparrow, M. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks, 13*(3), 251–274.
van Campen, T., Hamers, H., Husslage, B., & Lindelauf, R. (2018). A new approximation method for the Shapley value applied to the WTC 9/11 terrorist attack. *Social Networks Analysis and Mining, 8*(1), 1–12.

van den Nouweland, A., Borm, P., & Tijs, S. H. (1992). Allocation rules for hypergraph communication situations. *International Journal of Game Theory, 20,* 255–268.
Vázquez-Brage, M., van den Nouweland, A., & García-Jurado, I. (1997). Owen's coalitional value and aircraft landing fees. *Mathematical Social Sciences, 34*(3), 273–286.
Wise, W. M. (2005). *Indonesia's war on terror*. United States-Indonesia Society.

The Power of Closeness in a Network

Manfred J. Holler and Florian Rupp

Abstract This paper considers the question of whether it is profitable for a weaker player to be closely linked to a strong/powerful player—our conjecture is 'Yes'—and whether it is more beneficial to a strong/powerful player to be closely linked with a weak player than being linked with a strong player. Our understanding of power is based on the Public Good Index. We will demonstrate that in this sense power is a non-local concept indicating that strong players form a 'hot-region' about the strongest player. To obtain this result, we present an easy to perform algorithm for the computer-based determination of the Public Good Index on networks that equips us with instruments for studying the voting power in small networks.

1 Introduction

This paper puts a spotlight on the, presumably, ancient question of how close one should be to powerful players. The setting we are discussing is that of a democratic network where players form winning coalitions to execute power. The measure of power that is used throughout the paper is that of the Public Good Index (PGI) subject to networks, see Holler and Rupp (2019, 2020a, 2020b). This instrument helps us to reframe our questions more precisely of whether it is profitable for a weaker player to be closely linked to a strong player—our conjecture is 'Yes'—and whether it is more beneficial for a strong player to be closely linked to a weak player than being linked to a strong player. More generally, we will ask when closeness is favorable or unfavorable, where closeness between two nodes is measured using the shortest path between these nodes and the number of other nodes on that path between these nodes. If no path connects them, then as per definition the value of closeness is

M. J. Holler (✉)
University of Hamburg, Germany and CCR-Munich, Hamburg, Germany
e-mail: manfred.holler@uni-hamburg.de

F. Rupp
Kutaisi International University, Georgia and CCR-Munich, Kutaisi, Georgia
e-mail: florian.rupp@kiu.edu.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_15

zero (and the graph decomposes into more than one connected component). We will demonstrate that the answers to these questions depend on the properties of the network.

The first indication about powerless players in a voting network is given in Bozzo and Franceschet (2016). There, power is measured in terms of the power equation, and the least powerful nodes have been detected as end nodes of a voting network that are just connected to a single other player. Here, we will show that in terms of power measured by the PGI it indeed seems to have advantages to connect to the most powerful player, and connecting to the least powerful player may hand the red lantern, assigning the lowest PGI, value to the connecting new node.

2 Algorithmic and Computational Aspects

Our analysis focuses on a weighted voting game of the type $v = (d; w)$ where d is the decision rule (quota or quorum) and $w = (w_1, \ldots, w_i, \ldots, w_n)$ is the vector of voting weights. $N = \{1, \ldots, n\}$ is the set of decision makers (agents or players) and i is an element of N. A subset S of N is a coalition. The value of a coalition S is $v(S)$. In general, voting games are modeled in the form of simple games, i.e., $v(S) = 1$ if S is a winning coalition if $\sum_{i \in S} w_i \geq d$, and $v(S) = 0$ if S is a losing coalition if $\sum_{i \in S} w_i < d$.

If $v(S) - v(S - i) = 1$, then i is a critical player, i.e., a swing player in coalition S that has the 'power' to turn a winning coalition into a losing coalition and a losing coalition into a winning one. This characterizes the notion of 'power' underlying our analysis—voting power. S is a minimum winning coalition (MWC) if all elements of S are critical players. The set of MWCs is a subset of the set of winning coalitions W. If W is nonempty then, obviously, a MWC exists.[1]

If c_i represents the number of MWCs that have i as a member then the Public Good Index (PGI) is defined by

$$h_i(v) = \frac{c_i}{\sum_{i \in N} c_i}. \tag{1}$$

This measure has been introduced in Holler (1982b) and axiomatized in Holler and Packel (1983) and, "axiomatization completed," in Napel (1999, 2001).[2] Note that the PGI considers the value of a coalition to be a public good $v(S)$ such that the standard properties of non-rivalry in consumption and non-excludability apply. The PGI value $h_i(v)$ expresses the 'influence' that i has on the creation of $v(S)$. If in the following, we nevertheless speak of shares of power, then this is a shorthand for

[1] Though, it can be that no winning coalitions exist. For instance, assume a set of players who are not connected with each other. If none has a weight larger or equal to the quorum then no winning coalition can be formed.

[2] For alternative axiomatizations, see Alonso-Meijide et al. (2008) and Safokem et al. (2021).

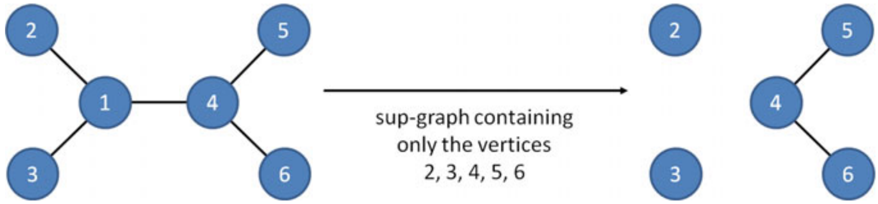


Fig. 1 Illustration of algorithm's key step 2

expressing the degree of influence. The discussion of the essence of power is not our focus in this paper[3]; we are interested in the effects of networks and networking.

Based on the efficient implementation of power indices for simple voting games in programming and software environments, like the program R, it is straightforward to extend these implementations to (large) network structures that test the feasibility of coalitions in simple voting games in the network. This feasibility testing considers whether a given coalition leads to a connected sub-graph in the network or not.

Thus, in terms of the PGI of a voting game $v = (d; w, \Gamma)$, where all players have the same weight, on a network Γ the algorithmic procedure for its computation can read as

Algorithm: PGI on Networks

1. Compute the Minimum Winning Coalitions (MWCs) of the simple voting game $v = (d; w)$, where all players have the same weight.
2. For each of these MWC, check if it is feasible within the network structure, i.e., whether the MWC induces a sub-graph on Γ that has exactly one connected component.
3. Compute the PGI based on the remaining feasible MWCs.

The benefit of starting with the MWCs of the simple voting game and then reducing their number in accordance with the concrete graph structure is that we will not forget any possible MWCs. The key step is to check the connectedness of the sub-graph induced by the MWC. Figure 1 gives an illustration of a case in which this condition is violated. The simple voting game $v = (5; 1, 1, 1, 1, 1, 1)$ consists of six players of equal weights and with a quorum rule $d = 5$ such that the Players 2, 3, 4, 5 and 6 form a MWC in this (unrestricted) simple voting game. Though, if the network restriction Γ is imposed, then the set of Players 2, 3, 4, 5 and 6 is no longer connected, as the Player 1 is not included in this set that connects, e.g., Players 2 and 4. Therefore, the set of Players 2, 3, 4, 5 and 6 is not a MWC in the voting game $v = (d; w, \Gamma)$ on the given network Γ.

[3] We refer to the corresponding literature on power measures like Holler (1982b), Holler and Owen (2001), Holler and Nurmi (2013), and Napel (2019). Kurz et al. (2015) offer "A Forecast of Tomorrow's Power Index Research" and possible applications.

For a voting system on a network in which every player has the same voting weight, the highest PGI value is attained at that node, which is part of the most connected sub-graphs with cardinality q. An implementation of this algorithm in the software environment R is provided in the appendix.

We will see later in the context of star networks that for voting games $v = (d; w, \Gamma)$ on a network Γ with non-equal weights of the players there can be MWC on the network that may not be MWC of the simple voting game. Here, a player is required who assures the connectivity of MWC on the network and who is not critical in the simple voting game. Thus, a more general algorithm must take the MWCs on the network into account. Though, this requires more advanced operations for networks.

Note that due to the collection of all possible MWCs in step 1 the performance of this algorithm follows the computational performance of determining the PGI in simple voting games. This performance is limited by combinatorial effects that essentially grow with the number of decision makers. In particular, Sperner's theorem (Sperner, 1928) bounds this number as follows

$$1 \;\leq\; \text{number of MWCs in a simple voting game} \;\leq\; \binom{N}{\lfloor \frac{N}{2} \rfloor}.$$

Due to the network structure, this number is dramatically reduced for the voting game defined on the network. A speed-up of the algorithm will be gained by directly determining the MWCs on the network via consistent paths between the decision makers.

3 Power in Small Unweighted Networks

The following networks are prototypical examples for the comparison of graph theoretic centrality measures in the sense that they display rather small graphs such that the typical centrality measures like betweenness centrality, closeness centrality, degree centrality, and eigenvector centrality can be analyzed.

Betweenness centrality quantifies the number of times a vertex acts as a bridge along the shortest path between two other vertices. It was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network. In his conception, vertices that have a high probability to occur on a randomly chosen shortest path between two randomly chosen vertices have a high betweenness. The **degree of a node** is the number of other nodes it is connected to, i.e., the number of edges that are associated with this node. The **degree central** node has the highest degree (there can be more than one such node). **Closeness centrality** of a node is the average length of the shortest path between the node and all other nodes in the graph; i.e., it is the reciprocal of the farness. Thus, the more central a node is, the closer it is to all other nodes. Finally, **eigenvector centrality** assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. Google's PageRank and the Katz centrality are variants of the eigenvector centrality. Let A be the adjacency matrix of the network, then the mentioned relative scores are provided by the entries of the eigenvector x that has positive entries only such that $\lambda x = Ax$. The unique existence of a pair (λ, x), where x has positive entries only, is guaranteed by the theorem of Perron–Frobenius.

For a more in-depth discussion of these measures, we refer to the literature, e.g., Bonacich (1987), Newman (2010), Todeschini and Consonni (2009), Jackson (2019) and Holler and Rupp (2019). For these simple networks, we refer to Brandes and Hildebrand (2014) as a reference for the application of the mentioned centrality measures.

In Krackhardt's kite (Fig. 2, top left), node 8 has the highest betweenness and the highest eigenvector score, nodes 6 and 7 have the highest closeness score, and node 4 has the highest degree score. Our algorithm tells us that node 8 (betweenness and eigenvector winner) also has the highest PGI value.

Everett's network (Fig. 2, top right) is an example where betweenness, closeness, degree and eigenvector centrality are located at different nodes. We call such networks BCDE networks. Node 2 has the highest betweenness score, node 3 has the highest closeness degree, node 4 has the highest degree score, and node 5 has the highest eigenvalue score. Here, node 3 (closeness winner) has the highest PGI value.

Pott's network (Fig. 2, bottom left) is also of BCDE type. Node 8 has the highest betweenness score, node 6 has the highest closeness score, node 3 has the highest degree score, and node 5 has the highest eigenvector score. Moreover, node 8(betweenness winner) has the highest PGI value.

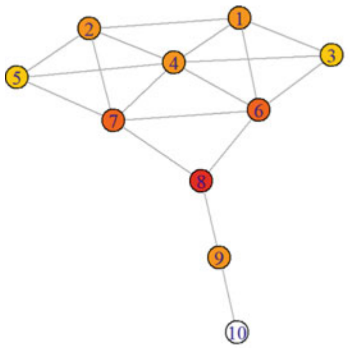
Finally, in the Medici network (Fig. 2, bottom right), that we discussed in detail in Holler and Rupp (2021, 2022), node 12, representing the fifteenth-century Medici family, has the highest betweenness score, the highest closeness score, the highest degree score, the highest eigenvalue score as well as the highest PGI value.

Figure 3 shows further examples for networks and the distribution of Public Good Power within them. The network in Fig. 3, left, is a real network given in a study involving students in a distance learning group using a computer-supported collaborative learning (CSCL) environment; i.e., it is a network of the usage of the CSCL's instant messenger (IM) system by the students (see Sundararajan, 2008). The two-star network in Fig. 3, right, is an example given by Massimo Franceschet.[4] Both examples illustrate that the nodes with the highest PGI values seem to be quite often linked directly. However, the Medici network shows that also bridges between such nodes are possible as well.

The examples illustrate two essential results: First, power measured by the PGI is different from a power assignment based on the traditional graph theoretic measures of betweenness, closeness, degree and eigenvector centrality. Second, power measured by the PGI in networks is a non-local concept: We see the most powerful node to be surrounded by the next powerful nodes giving rise to some 'power hot-region' rather than that the powerful node is supported by less powerful ones.

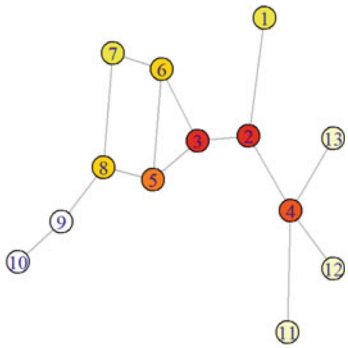
[4] https://users.dimi.uniud.it/~massimo.franceschet/networks/nexus/properties.html.

Krackhardt's kite network:



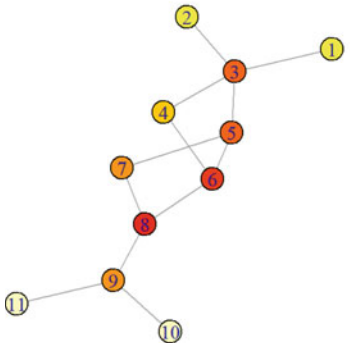
d = 6. PGI-vector: $h_K = (0.0952, 0.0952, 0.0857, 0.1071, 0.0857, 0.1190, 0.1190, 0.1500, 0.1023, 0.0404)$

Everett's BCDE-network:



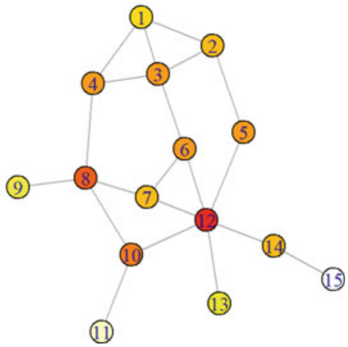
d = 7. PGI-vector: $h_E = (0.0599, 0.1405, 0.1428, 0.1198, 0.1013, 0.0875, 0.0645, 0.0852, 0.0391, 0.0138, 0.0483, 0.0483, 0.0483)$

Potts's BCDE-network:



d = 6. PGI-vector: $h_P = (0.0530, 0.0530, 0.1250, 0.0946, 0.1212, 0.1287, 0.0984, 0.1477, 0.1022, 0.0378, 0.0378)$

The Medici network:

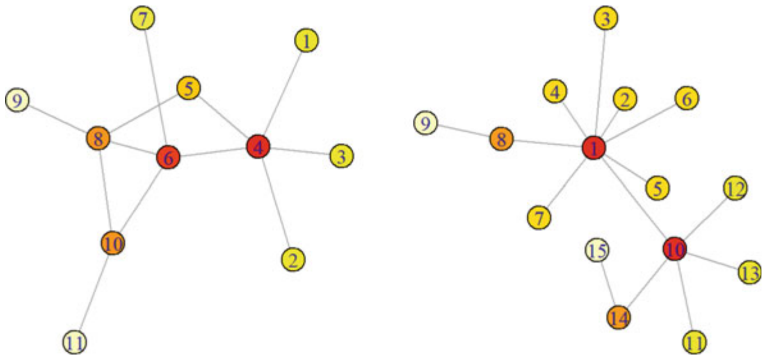


d = 8. PGI-vector: $h_M = (0.05606061, 0.06469697, 0.07303030, 0.07348485, 0.07212121, 0.07272727, 0.06560606, 0.09136364, 0.03893939, 0.08787879, 0.03742424, 0.11787879, 0.05196970, 0.06969697, 0.02712121)$

Fig. 2 Display of network structures and PGI-vectors—h_K, h_E, h_P and h_M—for Krackhardt's kite, Everett's BCDE-network, Potts' BCDE-network and the Medici network. (The darker the fill color, the higher the PGI value of the corresponding node.)

The networks in Fig. 4 are designed to address the question of whether it is profitable for a weaker player to be closely linked to a strong player. They have a kite-like structure with a tail of the length of the quorum. Based on our insights, this design is chosen to guarantee that the end of the tail can be present in one and only one minimal winning coalition, whereas the node with the highest PGI value is the con-

Collaborative learning environment: **Two-star network:**



d = 6. PGI-vector: h = (0.07341270, 0.07341270, 0.07341270, 0.15873016, 0.09325397, 0.13888889, 0.06547619, 0.11706349, 0.05158730, 0.10912698, 0.04563492)

d = 8. PGI-vector: h = (0.12500000, 0.05864865, 0.05864865, 0.05864865, 0.05864865, 0.05864865, 0.05864865, 0.07959459, 0.03405405, 0.12405405, 0.05783784, 0.05783784, 0.05783784, 0.07864865, 0.03324324)

Fig. 3 Collaborative learning environment and two-star network

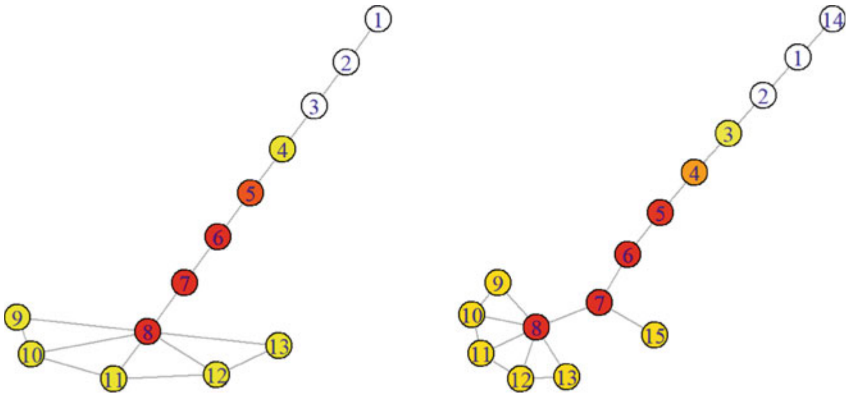
nector between the tail and the remaining part and thus obtains options for a critical participation in minimal winning coalitions form both of these parts. The initial kite (Fig. 4, left) and the modified kite with two additional nodes (Fig. 4, right) both have node 7 as that with the highest PGI value. The additional nodes 14 and 15 in the modified kite are attached to the node with the initially highest (node 7) and that with the initially lowest PGI value (node 1). We see that when adding a new node 14 to the network the red lantern is handed to that new node: assigning the lowest PGI value to the new node. Nodea 14 and node 1 show a PGI value of 0.001893939 and 0.005681818, respectively.

A coalition with the least powerful node alone does not lead to a gain of power, which seems to be intuitively clear in the context of politics. On the other hand, node 15 is, by purpose, connected to the most powerful node 7. We see that this link to the strongest player assigns, in this example, even more power to the newcomer node 15 than to nodes 9–13 which were initially at the periphery.

4 Power in Weighted Networks

So far, we discussed networks in which each node had the same voting weight. In this section, we drop this assumption and allocate different weights. These weights can be understood as representing resources, e.g., vote shares. Let us continue with several prototypical small network structures that commonly occur as sub-graphs in

Kite: **Modified kite (new players** 14 **and** 15**):**



d = 7. PGI-vector: h = (0.004329004, 0.008658009, 0.030303030, 0.073593074, 0.116883117, 0.138528139, 0.142857143, 0.138528139, 0.069264069, 0.069264069, 0.069264069, 0.069264069, 0.069264069)

d = 8. PGI-vector: h = (0.005681818, 0.017045455, 0.045454545, 0.083333333, 0.111742424, 0.123106061, 0.125000000, 0.121212121, 0.060606061, 0.060606061, 0.060606061, 0.060606061, 0.060606061, 0.001893939, 0.062500000)

Fig. 4 Kite networks designed to address the question of whether it is profitable for a weaker player to be closely linked to a strong player (The darker the fill color the higher the PGI value of the corresponding node.)

larger networks such that they can be considered as prototypical building blocks that additionally display voting networks with nodes of different weights. Figure 5 provides an overview of the network structures, we are going to discuss: linear networks, circle networks and star networks each with five players. Particularly, in weighted voting games, the connectivity of the players has a huge effect on the distribution of power. In this section, we propose the voting game $v^\circ = (51; 35, 20, 15, 15, 15)$ and apply the described various linear, circle and star network structures to it. In what follows we will call Players 1 and 2 'strong players' and 3, 4 and 5 'weak players' because of their voting weights. In some cases, Player 1 is addressed by 'the strong player'. But we will also speak of strong and weak players concerning the values of the power index.

The voting game $v^\circ = (51; 35, 20, 15, 15, 15)$ is the notorious example used to demonstrate that the PGI violates local monotonicity. If no network structure is considered, the corresponding PGI is $h(v^\circ) = \left(\frac{4}{15}, \frac{2}{15}, \frac{3}{15}, \frac{3}{15}, \frac{3}{15}\right)$. Note that Player 2 has more votes than Players 3, 4 or 5, but a smaller PGI value, i.e., local monotonicity, is violated.[5] Thus, a player that is strong in (voting) weights can be a weak player when

[5] We do not want to discuss this issue here—for a discussion, see Holler (2018, 2019)—, but point out that the more popular power indices by Shapley–Shubik and Penrose–Banzhaf also lose their local monotonicity property if engrafted by a network structure.

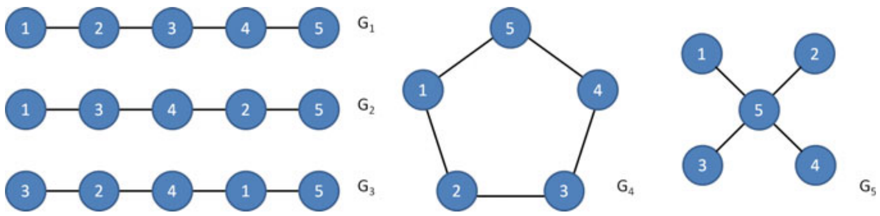


Fig. 5 Some connectivity network structures for voting games with five players: linear networks (left), a circle network (middle) and a star network (right)

ranked by its PGI value—even when there is no network introducing restrictions on coalition formation. In the following, the PGI values $h(v^{\circ})$ serve as a base of reference for calling a player strong or weak.

4.1 The Linear Network Case

To start with, we assume the structure G_1:

$$G_1 : \quad (1) - (2) - (3) - (4) - (5)\,.$$

The set of MWCs is $M(G_1) = \{\{1,2\},\{2,3,4,5\}\}$. The corresponding PGI is $h_1 = \left(\frac{1}{6}, \frac{2}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$. The introduction of a network structure had a significant impact on the power structure. Player 2, victim to the violation of monotonicity in the unstructured game, is the strongest player if network structure G_1 holds—its power value is larger than the power value of Player 1 who controls more votes than Player 2. It seems that Player 2 profits from being more central than Player 1. But if we compare the power values of Player 2 and Player 4—the two players have the same values of centrality however measured—then we may conclude that Player 2 profits to be directly linked to the 'strong' Player 1 as measured by 1's voting weight. Does it pay to be close to a strong player?

To come closer to the answer to this question, we rearrange the network positions of the five players along the linear dimension such that network structure G_2 results.

$$G_2 : \quad (1) - (3) - (4) - (2) - (5)\,.$$

The set of MWCs is $M(G_2) = \{\{1,3,4\},\{2,3,4,5\}\}$; therefore, we get $h_2 = \left(\frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{2}{7}, \frac{1}{7}\right)$. Again, closeness to a 'strong player' seems to be profitable. But is it profitable for the strong one? More specifically, is it more profitable for a strong player to be next to a stronger player than to a weaker one?

Let us further check the effects of closeness by assuming network structure G_3:

$$G_3 : \quad (3) - (2) - (4) - (1) - (5)\,.$$

Here, we gain the set of MWCs as $M(G_3) = \{\{2, 4, 1\}, \{4, 1, 5\}\}$, and thus $h_3 = \left(\frac{2}{6}, \frac{1}{6}, \frac{0}{6}, \frac{2}{6}, \frac{1}{6}\right)$. It seems that centrality is 'profitable' (compare Players 3 and 4!), but neighboring a 'strong' player seems 'profitable' as well. Note that Players 3, 4 and 5 have equal vote shares. We may conclude that Player 4 benefits of centrality and closeness to the strong Player 1.

Under the assumption that all players are connected and thereby potential members of some coalitions, there are $n! = 5! = 120$ possible orderings representing alternative linear network structures for five layers. If we check these alternatives, can we conclude that it is beneficial to be a neighbor of a strong player?

Due to Players 3, 4 and 5 having the same weight, the number of $5! = 120$ possible orderings actually reduces to the 10 essential ones listed in Table 1. For instance, the PGI values $h_1 = \left(\frac{1}{6}, \frac{2}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$, which derives for network G_1, also describe the PGI values of the following five networks:

$$(1) - (2) - (3) - (5) - (4), \qquad (1) - (2) - (4) - (5) - (3),$$
$$(1) - (2) - (5) - (3) - (4),$$
$$(1) - (2) - (4) - (3) - (5), \qquad (1) - (2) - (5) - (4) - (3).$$

Note, these networks, together with G_1, contain the six possible orderings of Players 3, 4 and 5, connected with the given ordering of Players 1 and 2 as the starting places, read from left to right. Additionally, the same number of orderings occurs for the ending ordering of the Players 2 and 1, read from the right to the left. Thus, the $h_1 = \left(\frac{1}{6}, \frac{2}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right)$ represents the essence of the twelve orderings (Table 2).

4.2 *The Circle Network Case*

The linear structure has a long tradition in economics and theoretical and empirical public choice. For example, it is used to model a political left-right dimension in the political arena.[6] However, to separate the closeness effect from centrality effects due to positions toward the beginning and end of the linear space, we chose the circular structure G_4 as given in Fig. 2 (middle).

Given the voting game $v^\circ = (51; 35, 20, 15, 15, 15)$, we have $M(G_4) = \{\{1, 2\}, \{4, 5, 1\}, \{2, 3, 4, 5\}\}$ and thus $h_4 = \left(\frac{2}{9}, \frac{2}{9}, \frac{1}{9}, \frac{2}{9}, \frac{2}{9}\right)$. Here, Players 4 and 5 are neighbors of the strong player 1, and coalitions with the strong Player 1 include either the direct neighbor 5 or Player 4 (who is still close to Player 1), or both of them.

---

[6] Hotelling (1929) applied such a model to discuss competition in a one-dimensional spatial duoply market. However, he also applied his model to discuss spatial competition between the Democrats and the Republicans in the USA. The Public Choice literature by and large ignored Hotelling's work and results. See the pioneering work by Anthony Downs (1957).

Table 1 The 10 essential linear networks $v^\circ = (51; 35, 20, 15, 15, 15)$ together with the number of their occurrences which has to add up to 120 in total as well as the corresponding PGI values. (Note that due to having the same weights, the nodes 3, 4 and 5 are interchangeable without loss of generality.)

The 10 essential linear networks	No of occurrences	PGI of Player 1	PGI of Player 2	PGI of Player 3	PGI of Player 4	PGI of Player 5
G_1: (1) – (2) – (3) – (4) – (5)	12	0.1666	0.3333	0.1666	0.1666	0.1666
(2) – (1) – (3) – (4) – (5)	12	0.4000	0.2000	0.2000	0.2000	0.2000
(1) – (3) – (2) – (4) – (5)	12	0.0000	0.2500	0.2500	0.2500	0.2500
(1) – (3) – (4) – (5) – (2)	12	0.1428	0.1428	0.2857	0.2857	0.1428
G_2: (1) – (3) – (4) – (2) – (5)	12	0.1428	0.1428	0.2857	0.2857	0.1428
(2) – (3) – (4) – (1) – (5)	12	0.3333	0.0000	0.1666	0.3333	0.1666
(3) – (1) – (2) – (4) – (5)	12	0.5000	0.5000	0.0000	0.0000	0.0000
(3) – (2) – (1) – (4) – (5)	12	0.4000	0.2000	0.2000	0.2000	0.2000
G_3: (3) – (2) – (4) – (1) – (5)	12	0.3333	0.1666	0.0000	0.3333	0.1666

Table 2 Essential circle networks $v^\circ = (51; 35, 20, 15, 15, 15)$

Circle Network	PGI of Player 1	PGI of Player 2	PGI of Player 3	PGI of Player 4	PGI of Player 5
G_4: (1) – (2) – (3) – (4) – (5)	0.2222	0.2222	0.1111	0.2222	0.2222
(1) – (3) – (2) – (4) – (5)	0.2000	0.1000	0.2000	0.2000	0.3000

Note that the initial and the terminal player are connected

4.3 *The Star Network*

Finally, we discuss the star network G_5 (see Fig. 2, right), where the Players 1, 2, 3 and 4 are all connected to Player 5 and have no further direct links between each other:

$$G_5 : \quad (1) - (5)\,, \;\; (2) - (5)\,, \;\; (3) - (5)\,, \;\; (4) - (5)\,.$$

With the MWCs $M(G_5) = \{\{5, 1, 4\}, \{5, 1, 3\}, \{5, 2, 3, 4\}\}$, the Public Good Index of this star network is $h_5 = \left(\frac{2}{10}, \frac{1}{10}, \frac{2}{10}, \frac{2}{10}, \frac{3}{10}\right)$. Thus, the hub position of Player 5 is 'very profitable'. Since player 5 is pivotal to any coalition, neither the difference in weight between Player 2 and Players 3 and 4 matters nor the maximum voting

Table 3 Comparison of the PGI in possible star networks formed by 5 players with different hub players and voting rule $v^{\circ} = (51; 35, 20, 15, 15, 15)$

Star network with hub position	PGI of Player 1	PGI of Player 2	PGI of Player 3	PGI of Player 4	PGI of Player 5
1	0.36363636	0.09090909	0.18181818	0.18181818	0.18181818
2	0.1666667	0.3333333	0.1666667	0.1666667	0.1666667
3	0.2	0.1	0.3	0.2	0.2
4	0.2	0.1	0.2	0.3	0.2
5 (i.e., G_5)	0.2	0.1	0.2	0.2	0.3

weight of Player 1 assures a maximum of the voting power. In terms of Freeman, Player 5 satisfies point centrality: "the point at the center of a star or the hub of a wheel [...] is the most central possible position. A person located in the center of a star is universally assumed to be structurally more central than any other person in any other position in any other network of similar size" (Freeman 1978/79: 218). But centrality does not equal power. When it comes to collective action, a central player needs the support of other players which gives power to the latter.

Table 3 summarizes the PGI values for the star network for the cases that Players 1, 2, 3 4 or 5 are in the hub position, alternatively. The latter case is shown in Fig. 2 (right); it represents the star network G_5.

Finally, let us consider the star network, where Player 1 with weight $w_1 = 33$, 2 with weight $w_2 = 33$ and 3 with weight $w_3 = 33$ are connected with player 4 with weight $w_4 = 1$:

$$S : \quad (1) - (4)\,, \;\; (2) - (4)\,, \;\; (3) - (4)\,.$$

The voting game, we consider is $v^{\circ} = (51; 33, 33, 33, 1)$. Here, the MWCs on the star network are $M(S) = \{\{1, 2, 4\}, \{1, 3, 4\}, \{2, 3, 4\}\}$. Here, player 4 is critical to the MWCs because if this player leaves a coalition the remaining partners are no longer connected. On the other hand, in the simple voting game, the MWCs are $\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$, and here, player 4 is not critical for any MWC (and has no power).

Note, in unweighted voting games $v_{\Gamma} = (d; w, \Gamma)$ on a network the MWCs on the network are always a subset of the MWCs of the corresponding simple voting game $v = (d; w)$.

5 Discussion

Based on our derived observations for power measured by the PGI, we can justify the following three statements: First, power measured by the Public Good Index (PGI) is different from a power assignment based on the traditional graph theoretic measures of betweenness, closeness, degree and eigenvector centrality. Second, power measured by the PGI in networks is a non-local concept. In the networks in which each node had the same voting weight, our examples seem to indicate straightforward results concerning closeness: We see the most powerful node to be surrounded by the next powerful nodes giving rise to some ‘power hot-region’ (see the Medici network for a bridge between the two most powerful nodes) rather than that the powerful vertex being a singularly isolated node is supported by lesser powerful ones. Third, our examples illustrate that the nodes with the highest PGI values seem to be quite often linked directly.

As yet we cannot propose any general conditions for powerful nodes being linked or connected by bridges. General conclusions are even less likely when resources are distributed unevenly. Then, both resources and the player’s position in the social context determine the player’s potential to be a valuable element of a winning coalition. This invites us to distinguish between players which are strong measured by their resources and players who are strong because of their network positions. The corresponding PGI values are $h = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{2}{6}, \frac{1}{6}\right)$. The PGI values summarize the two components. In the extreme, a player can exert power without resources—due to the player’s position in the network. Let us assume the network structure G_3: $(3) - (2) - (4) - (1) - (5)$ and a weighted voting game $v = (51; 35, 25, 25, 0, 15)$. Above ‘strong’ refers to resources; however, the example demonstrates that this label is not always adequate.

Because of the nonmonotonicity of the PGI, it can happen that the power value of a player with larger resources is smaller than the value of a player with smaller resources even when there are no constraints on forming alliances with other players. Such effects can be amplified or counterbalanced if the social context takes the form of a particular network. Closeness to a strong player seems to be profitable. But is it profitable for the strong one as well? More specifically, is it more profitable for a strong player to be next to a stronger player than to a weaker one? Does the PGI value of a strong player i increase if we substitute a connected player j with a weaker player k? Our conjecture is: it depends?

So far, we assumed that the networks were exogenously given. But, of course, it would interesting, and perhaps even beneficial, to see whether we can apply our apparatus to explain the formation of particular networks endogenously. Think about a dinner party where people of different power arrive at random, which bar tables would they join?

Acknowledgements The Authors would like to thank Luciano Andreozzi, Nicola Friederike Maaser and Aurelien Mekuko for their very helpful comments.

Appendix: Essential Source Code for the Computation of the PGI in Voting Games on Networks

As outlined, the programming environment of our choice is R, and for the following computation of power indices, the commands require the packages `igraph` and `CoopGame` to be installed and loaded. The first step is to define the underlying voting game:

```
n <- 10          # number of vertices in the graph/ network
d <- 6           # decision rule
w <- rep(1, n)   # weight vector in case that all vertices have the
                # same weight 1
# network structure, together with the dummy element 0
g <- make_graph(~ 0, 1-2, 1-3, 1-4, 1-6, 2-4, 2-5, 2-7, 3-4, 3-6,
+                  4-5, 4-6, 4-7, 5-7, 6-7, 6-8, 7-8, 8-9, 9-10)
```

Next, we pave the ground by discussing the simple voting game to gain information about its MWCs:

```
v <- weightedVotingGameVector(n, w, d)
M <- getMinimumWinningCoalitions(v)
m <- nrow(M)                               # total number of MWCs
```

As a preparation for the PGI on the network structure, we generate an auxiliary Matrix M_2, such that each row of M_2 contains the numbers of the elements that constitute the MWC described in the row, together with the dummy element 0. Note, we assume that the grand coalition (containing all elements) is not a MWC, i.e., each row of M_2 contains at least one dummy element.

```
M2 <- matrix(0, m, n)     # mxn-matrix filled with zeros
M3 <- matrix(0, m, n)     # image of M with better handling properties
for( rcount in 1:m ){
   for( ccount in 1:n ){
      if( M[rcount, ccount] == 1 ){
         M2[rcount, ccount] <- ccount
          M3[rcount, ccount] <- 1      }}}
```

Now, we turn to the network structure, and check whether a MWC of the simple voting game is feasible within the network structure:

```
for( rcount in 1:m ){
   cc <- sort( c( M2[rcount,]) )
  g1 <- induced_subgraph( g, as.character(cc) )
  # if the MWC is not feasible, then the corresponding row in M is
   # equaled to a zero-vector
   if( count_components(g1) >= 3 ){
      M3[rcount,] <- rep(0, n)    }}
```

Finally, we compute the total number of MWC a vertex belongs to and thus the PGI on the network structure:

```
mwcg <- rep(0,n)
for( i in 1:n ){ mwcg[i] <- sum( M3[,i] ) }
pgi <- mwcg/sum(mwcg)    # Public Good Index on the network structure
```

References

Alonso-Meijide, J. M., Casas-Mendez, B., Holler, M. J., & Lorenzo-Freire, S. (2008). Computing power indices: Multilinear extensions and new characterizations. *European Journal of Operational Research, 188*, 540–554.

Barabasi, A.-L. (2016). *Network science*. Cambridge University Press.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology, 92*, 1170–1182.

Bozzo, E., & Franceschet, M. (2016). A theory on power in networks. *Communications of the ACM, 59*, 75–83.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology, 25*, 163–177.

Brandes, U., & Hildenbrand, J. (2014). Smallest graphs with distinct singleton centers. *Network Science, 2*, 416–418.

Downs, A. (1957). *An economic theory of democracy*. Harper & Brothers.

Freeman, L. C. (1978/79). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215–239.

Holler, M. J. (1982). Forming coalitions and measuring voting power. *Political Studies, 30*, 262–271.

Holler, M. J. (Ed.). (1982b). *Power, voting, and voting power*. Physica.

Holler, M. J. (2018). The Public Good Index: A Brief Introduction. In T. Airaksinen, M. J. Holler, & H. Nurmi (Eds.), *Finnish-German Yearbook of Political Economy 1* (pp. 31–39). Accedo.

Holler, M. J. (2019). The story of the poor public good index. In N. T. Nguyen, R. Kowalzyk, J. Mercik, & A. Motylska-Kruzma (Eds.), *Transactions on computational collective intelligence 34* (pp. 171–179). Springer.

Holler, M. J., & Nurmi, H. (Eds.). (2013). *Power, voting, and voting power: 30 years after*. Springer.

Holler, M. J., & Owen, G. (Eds.). (2001). *Power indices and coalition formation*. Kluwer Academic Publishers.

Holler, M. J., & Packel, E. W. (1983). Power, luck, and the right index. *Journal of Economics (Zeitschrift für Nationalökonomie), 43*, 21–29.

Holler, M. J., & Rupp, F. (2019). Power in networks: A PGI analysis of Krackhardt's kite network. In N. T. Nguyen, R. Kowalzyk, J. Mercik, & A. Motylska-Kruzma (Eds.), *Transactions on computational collective intelligence 34* (pp. 21–34). Springer.

Holler, M. J., & Rupp, F. (2020). Some strategic decision problems in networks. In N. T. Nguyen, R. Kowalzyk, J. Mercik, & A. Motylska-Kruzma (Eds.), *Transactions on computational collective intelligence 35* (pp. 133–147). Springer.

Holler, M. J., & Rupp, F. (2020). Power in networks and the urban space. In E. Macri, V. Morea, & M. Trimarchi (Eds.), *Cultural commons and urban dynamics: A multidisciplinary perspective* (pp. 37–52). Springer.

Holler, M. J., & Rupp, F. (2021). Power in networks: The Medici. *Homo Oeconomicus, 38*, 59–75.

Holler, M. J., & Rupp, F. (2022). Power in networks, the urban space, and the Medici. *Munich Social Science Review*, New Series, 5, (forthcoming).

Hotelling, H. (1929). Stability in competition. *Economic Journal, 39*, 41–57.

Jackson, M. O. (2019). *The human network: How your social position determines your power, beliefs, and behaviors*. Pantheon Books.

Krackhardt, D. (1990). Assessing the political landscape: Structure, cognition, and power in organizations. *Administrative Science Quarterly, 35*, 342–369.

Kurz, S., Maaser, N., Napel, S., & Weber, M. (2015). Mostly sunny: A forecast of tomorrow's power index research. *Homo Oeconomicus, 32*(1), 133–146.

Napel, S. (1999). The Holler-Packel axiomatization of the public good index completed. *Homo Oeconomicus, 15*(4), 513–520.

Napel, S. (2001). Note on the Holler-Packel axiomatization of the public good index (PGI). In M. J. Holler & G. Owen (Eds.), *Power indices and coalition formation* (pp. 143–151). Kluwer Publishing.

Napel, S. (2019). Voting power. In R. Congleton, B. Grofman, & S. Voigt (Eds.), *Oxford handbook of public choice* (Vol. 1, Ch. 6, pp. 103–126). Oxford University Press.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.

Safokem, A., Moyouwou, I., & Mekuko, A. (2021). Axiomatizing the public good index via merging and new arrival properties. *Discrete Applied Mathematics, 305*, 86–102.

Sperner, E. (1928). Ein Satz über Untermengen einer endlichen Menge. *Mathematische Zeitschrift, 27*, 544–548.

Sundararajan, B. (2008). Determining an actor's network capacity and network utilization: A Markov model of human agency. *Connections, 28*(2), 28–42.

Todeschini, R., & Consonni, V. (2009). *Molecular descriptors for chemoinformatics*. (Vols. I & II). Wiley-VCH.

Political Power on a Line Graph

René van den Brink, Gerard van der Laan, Marina Uzunova, and Valeri Vasil'ev

Abstract We consider situations of majority voting, where the players are ordered linearly. This order may be based on, for example, ideology or political preferences over economic policy, ethical principles, environmental issues, and so on. Winning and losing coalitions are given by a majority voting game, while restrictions on cooperation are determined by a line graph, where only connected coalitions are feasible and can form a (winning) coalition. Various solutions for line-graph games can then be viewed as power indices measuring the ability of political parties to turn losing coalitions into winning ones, taking into account the cooperation restrictions among the parties. Here, we start by observing that a number of existing power indices either are not core stable, or do not reward intermediate veto players. Then, we take a closer look at the average hierarchical outcome, called hierarchical index in the context of this paper, and the τ-index. These indices are core stable and, moreover, reward all veto players. Specifically, the τ-index rewards all veto players equally, while the hierarchical index always assigns higher power to the two extreme veto players than to intermediate veto players. We axiomatically characterize the (i) hierarchical index by core stability and a weaker version of component fairness and (ii) the τ-index by core stability and a weaker version of Myerson's [Math Oper Res 2(3), 225–229 (1977)] fairness property.

R. van den Brink (✉) · G. van der Laan
School of Business and Economics and Tinbergen Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: j.r.vanden.brink@vu.nl

M. Uzunova
John Stuart Mill College, Vrije Universiteit, Amsterdam, The Netherlands
e-mail: m.g.uzunova@vu.nl

V. Vasil'ev
Sobolev Institute of Mathematics, Novosibirsk, Russia
e-mail: vasilev@math.nsc.ru

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_16

1 Introduction

In this paper, we will be concerned with measuring voting power in specific types of *majority voting situations*. Among the many existing power indices in the literature, some of the most famous ones are the *Shapley–Shubik index* (Shapley & Shubik, 1954), which is equivalent to the *Shapley value* (Shapley, 1953) applied to the associated voting game, and the *Banzhaf–Penrose index* (Banzhaf, 1965; Penrose, 1946). By the voting power of a given political party, we mean its ability to effect a change in the outcome: turning losing coalitions into winning ones or vice versa. To determine these abilities, it is sufficient to know the winning and losing coalitions in a voting situation. These coalitions can be summarized by an associated *majority voting game*: a cooperative or coalitional game where the worth of any coalition is either one (when it has the required qualified majority and is thus winning) or zero (when it does not have the required qualified majority and is thus losing).

The most famous set-valued solution for cooperative games is the *core* (Gillies, 1953), which, for any game, is the set of all efficient payoff vectors such that every coalition earns at least its own worth. In the context of voting games, we will mostly refer to payoff vectors as *power* vectors. For a majority voting game, this means that, in every core payoff vector, the payoffs or powers of the parties in a winning coalition must sum up to one. Specifically, the powers of *all* parties must sum up to one since the grand coalition—the set of all players in the game—is winning. Then, it immediately follows that the core of a voting game is non-empty if and only if the game has veto players: players who belong to every winning coalition. In that case, the core consists of all allocations where the full power of one is allocated over the veto players and where non-veto players are assigned a power of zero.

In a cooperative game, any coalition of players can form and obtain its worth. In voting games, this means that any coalition of parties can cooperate and try to form a majority or winning coalition. However, in real-life politics, not every combination of parties can form a coalition. In some cases, parties exclude each other from forming coalitions. What's more, even when parties do not exclude each other, it might be that two parties can only belong to the same coalition if another (ideologically intermediate) party belongs to that coalition as well. One way to model such cooperation restrictions is by using Myerson's (1977) (communication) graph game model. In this model, the players in a cooperative game are also the nodes in an undirected graph such that two players are linked if and only if they can cooperate together without any other player. The feasible coalitions in that case are the connected coalitions in the graph. Various existing solutions take account of these cooperation restrictions in allocating the payoffs over the players. One of the first such solutions was introduced by Myerson (1977) and is obtained by applying the Shapley value to the so-called restricted game. This restricted game is obtained by assigning to every coalition the sum of the worths of its maximally connected subsets (components). This solution was later called the *Myerson value*. Myerson (1977) characterized this solution as the unique solution for communication-graph games that satisfies *component efficiency* and *fairness*. Component efficiency requires that the sum of the payoffs of all players in a maximally connected subset (i.e. component) is equal to the worth of that component. Fairness requires that deleting an edge between two players has the same effect on their payoffs.

For cycle-free graph games, (Demange, 2004) introduced the concept of *hierarchical outcomes*, while (Herings et al., 2008) considered the average of all hierarchical outcomes and characterized this solution by component efficiency and *component fairness*. Component fairness requires that deleting an edge between two players in a cycle-free graph game has the same effect on the per capita payoffs in the two newly created components (each containing one of the two players whose link is broken). Béal et al. (2010) considered weighted combinations of hierarchical outcomes. Hierarchical outcomes are special marginal vectors of the game where the order in which the players enter the grand coalition is restricted by the graph. An interesting property of the hierarchical outcomes (and their convex combinations) is that they always belong to the core of the restricted game if the game is superadditive and the graph is cycle free. Since majority voting games are superadditive and line graphs are cycle free, this immediately implies that the hierarchical outcomes and their convex combinations assign core power vectors when we restrict a majority voting game to a line graph. In van den Brink et al. (2007), the special class of line-graph games is considered and particular attention is paid to two hierarchical outcomes (the so-called upper equivalent solution and the lower equivalent solution) and their average. These outcomes also regularly appear in the applied economics and operations research literature. Specifically, the *upper equivalent solution*, which assigns to every line-graph game the *marginal vector* where the players enter from left to right, yields the downstream incremental solution for river games in Ambec and Sprumont (2002) or the drop out monotonic solution for sequencing games in Fernández et al. (2005). The *lower equivalent solution*, which assigns to every line-graph game the marginal vector where the players enter from right to left, coincides with the upstream incremental solution for river games in Ambec and Ehlers (2008). The average of the upper and lower equivalent solutions yields the equal gain split rule introduced for one-machine sequencing games by Curiel et al. (1993, 1994).

In this paper, we consider majority voting games where the political parties are ordered on a line according to their political preferences over, for example, economic policy, ethical issues, environmental problems, and so on. We apply various solutions to the associated line-graph games as a way of measuring the parties' power, taking into account their positions on the line. The upper equivalent solution mentioned above assigns full power to the pivotal party in a *majority voting line-graph game* when the parties enter from left to right. The lower equivalent solution assigns full power to the pivotal party in a majority voting line-graph game when the parties enter from right to left. These two parties are also the most right-wing, respectively left-wing, veto players. Applying the average of the upper and lower equivalent solutions assigns equal (half) power to the left- and right-wing pivotal party.

Since all hierarchical outcomes and their convex combinations belong to the core of a majority voting line-graph game, it must hold that these two hierarchical outcomes, and their average, fully allocate power over veto players and give zero power to non-veto players. This is an important difference with the Shapley value (Shapley–

Shubik index (Shapley & Shubik, 1954) for voting games) or the Banzhaf value (Banzhaf, 1965) which, when applied to the restricted game of a majority voting line-graph game, ascribe positive power to non-veto players, and thus do not belong to the core of the restricted game.

The goal of this paper is to find core solutions that also reward non-extreme veto players. Whereas the average of the upper and lower equivalent solutions (the equal gains split rule) allocates the full power of one equally over the two extreme (left and right-wing) veto players. Other combinations of hierarchical outcomes allocate the full power over *all* veto players, assigning 'intermediate' or non-extreme veto players some positive power. Typically, however, in the average of all hierarchical outcomes, the two extreme veto players get the highest power. We refer to the power index that assigns to every majority voting line-graph game the average hierarchical outcome as the *hierarchical index*. We then axiomatically characterize this power index by core stability and a weaker version of Herings, van der Laan, and Talman's (Herings et al., 2008) component fairness property, where we only consider deleting edges between veto players.

An interesting power index in this context is the one that allocates the full power *equally* over all veto players. In this paper, we will show that this rule is obtained by applying the τ*-value* (Tijs, 1981) to the restricted game of a majority line-graph voting game. We call the resulting power index the τ*-index*. Moreover, we axiomatically characterize this power index by core stability and a weaker version of Myerson's (1977) fairness property, where again we only consider deleting edges between veto players.

This paper is organized as follows. In Sect. 2, we discuss preliminaries on majority voting games, line-graph games, and solutions on line-graph games. In Sect. 3, we apply known results on line-graph games to the measurement of political power on majority voting line-graph games. In Sect. 4, we consider the hierarchical outcomes and the τ-value as measures of political power on line-graph games and provide an axiomatization. Section 5 contains concluding remarks.

2 Preliminaries

2.1 Cooperative Games

A situation in which a finite set of players can obtain certain payoffs through cooperation can be described by a *cooperative game with transferable utility* or, simply, a cooperative game. A cooperative game is a pair (N, v), where $N = \{1, \ldots, n\}$ is a finite set of n players and $v \colon 2^N \to \mathbb{R}$ is a *characteristic function* on N such that $v(\emptyset) = 0$. For any coalition $S \subseteq N$, the real number $v(S)$ is the *worth* of coalition S; that is, the members of coalition S can obtain a total payoff of $v(S)$ by agreeing to cooperate.

In this paper, we assume that N is fixed. This allows us to refer to a cooperative game (N, v) simply by its characteristic function v. We denote the collection of all cooperative games on N (represented by their characteristic function) by $\mathcal{G}^N$. We first recall some properties of cooperative games. A cooperative game v is *superadditive* if $v(S \cup T) \geq v(S) + v(T)$ for any pair of subsets $S, T \subseteq N$ such that $S \cap T = \varnothing$. Further, a cooperative game v is *convex* if $v(S \cup T) + v(S \cap T) \geq v(S) + v(T)$ for all $S, T \subseteq N$. Every convex game is superadditive. A special class of convex games are unanimity games. For each non-empty $T \subseteq N$, the *unanimity* game u_T is given by $u_T(S) = 1$ if $T \subseteq S$, and $u_T(S) = 0$ otherwise. It is well known that unanimity games form a basis for $\mathcal{G}^N$. Specifically, every game v can be expressed as a unique linear combination of unanimity games,

$$v = \sum_{S \subseteq N,\, S \neq \varnothing} \Delta_S(v) u_S,$$

where $\Delta_S(v)$ are *the Harsanyi dividends* [see Harsanyi (1959)], given by

$$\Delta_S(v) = \sum_{T \subseteq S} (-1)^{|S|-|T|} v(T), \quad S \subseteq N, S \neq \varnothing. \tag{1}$$

Equivalently, by applying the Möbius transformation, we have

$$v(S) = \sum_{T \subseteq S} \Delta_T(v), \quad S \subseteq N, S \neq \varnothing. \tag{2}$$

So, the worth of coalition S is equal to the sum of the dividends of all subcoalitions of S. This also gives a recursive definition of the Harsanyi dividends. The dividend of every one-player coalition is equal to its worth, while, recursively, the dividend of every coalition with at least two players is equal to its worth minus the sum of the dividends of all of its proper subcoalitions. In this sense, the dividend of a coalition S can be interpreted as the extra benefit from cooperation among the players in S that they cannot realize through cooperation in smaller coalitions.

Solution concepts.

A *payoff vector* of a cooperative game (N, v) is an n-dimensional vector that assigns a payoff to any player $i \in N$. A *point-valued solution* is a function f that assigns a single payoff vector $f(v) \in \mathbb{R}^N$ to any game (N, v). A point-valued solution f is *efficient* if, for any game (N, v), it distributes precisely the worth of the grand coalition: $\sum_{i \in N} f_i(v) = v(N)$ for all $v \in \mathcal{G}^N$. An example of an efficient point-valued solution is the famous Shapley value (Shapley, 1953): the average of the so-called marginal contribution vectors.[1]

For a permutation $\pi : N \to N$ that assigns rank number $\pi(i) \in N$ to any player $i \in N$, we define $\pi^i = \{j \in N \mid \pi(j) \leq \pi(i)\}$, that is, π^i is the set of all players with rank number at most equal to the rank number of i, including i itself. Then the *marginal contribution vector* $m^\pi(v) \in \mathbb{R}^N$ of game v and permutation π is given by

[1] For a recent survey on the Shapley value, see Algaba et al. (2019).

$$m_i^\pi(v) = v(\pi^i) - v(\pi^i \setminus \{i\}), \quad \text{for all } i \in N.$$

The vector $m_i^\pi(v)$ thus assigns to player i its marginal contribution to the worth of the coalition consisting of all of its predecessors in π. The *Shapley value*, Sh, assigns to every game the average of the marginal contribution vectors over all permutations and is thus defined by

$$Sh_i(v) = \frac{1}{|N|!} \sum_{\pi \in \Pi(N)} m_i^\pi(v), \quad \text{for all } i \in N,$$

where $\Pi(N)$ is the collection of all permutations on N.[2]

Writing

$$m_i^S(v) = v(S \cup \{i\}) - v(S),$$

as the marginal contribution of player $i \in N$ to coalition $S \subseteq N \setminus \{i\}$, the *Banzhaf value* [see Owen (1975) and Dubey and Shapley (1979) as an extension of Banzhaf (1965)], Ba, is defined by

$$Ba_i(v) = \frac{1}{2^{|N|-1}} \sum_{S \subseteq N \setminus \{i\}} m_i^S(v), \quad \text{for all } i \in N.$$

Thus, $Ba_i(v)$ is the average marginal contribution of player i to every coalition that does not contain i assuming that every coalition has equal probability of occurring. The Banzhaf value is not efficient.[3]

The τ-value is defined in Tijs (1981) as an efficient solution for the class of quasi-balanced games. In order to define the τ-value and the class of quasi-balanced games, we first need to define two types of payoff bounds. Specifically, let the upper payoff bound of game v be given by the so-called *utopia payoff vector* $M(v)$ defined as

$$M_i(v) = v(N) - v(N \setminus \{i\}), \quad \text{for all } i \in N.$$

The vector $M(v)$ thus assigns to every player their marginal contribution to the grand coalition N. Further, let the lower payoff bound of game v be given by the so-called *minimal right vector* defined as

[2] Alternatively, for any game, the Shapley value distributes equally the Harsanyi dividend of coalition S over the players in S.

[3] Efficient normalizations of the Banzhaf value are the multiplicative and additive normalizations. The multiplicative normalization allocates $v(N)$ proportionally to the Banzhaf values of the players [see van den Brink and van der Laan (1998)]. The additive normalization is the least square value [see Ruiz et al. (1998)] obtained by adding or subtracting the same amount from the Banzhaf value payoffs of the players so that an efficient payoff vector results.

$$m_i(v) = \max_{S \subseteq N,\, i \in S} \left(v(S) - \sum_{j \in S \setminus \{i\}} M_j(v) \right), \quad \text{for all } i \in N.$$

The class of quasi-balanced games on N, denoted by QB^N, is the class of games for which $M(v)$ and $m(v)$ constitute genuine upper and lower bounds in the following sense: (1) each player's utopia payoff is at least as large as that player's minimal right, (2) all utopia payoffs sum up at least to the worth $v(N)$ of the grand coalition, and (3) all minimal rights sum up at most to the worth $v(N)$ of the grand coalition:

$$QB^N = \left\{ v \in \mathcal{G}^N \;\middle|\; m(v) \leq M(v) \text{ and } \sum_{i \in N} m_i(v) \leq v(N) \leq \sum_{i \in N} M_i(v) \right\}.$$

The τ-value is defined on the class of quasi-balanced games and, for every $v \in QB^N$, is given by

$$\tau(v) = m(v) + \alpha(M(v) - m(v)),$$

where $\alpha \in \mathbb{R}$ is such that the τ-value is efficient: $\sum_{i \in N} \tau_i(v) = v(N)$. The τ-value assigns to each player, in an efficient way, their minimal right plus a (uniform) share of the margin by which their utopia payoff exceeds their minimal right.

A *set-valued solution* for cooperative games is a mapping F that assigns to every game (N, v) a set of payoff vectors $F(v) \subset \mathbb{R}^N$. The most famous set-valued solution, the *core*, introduced by Gillies (1953), is the set of all efficient payoff vectors that cannot be improved upon by any coalition; that is, any payoff vector in the core is efficient and each coalition gets at least its own worth:

$$core(v) = \left\{ x \in \mathbb{R}^N \;\middle|\; \sum_{i=1}^{n} x_i = v(N), \text{ and } \sum_{i \in S} x_i \geq v(S), \text{ for all } S \subseteq N \right\}.$$

The core of a game can be empty. It is well known that $core(v)$ is non-empty if and only if v is balanced [which was shown independently by Bondareva (1963) and Shapley (1967)].

2.2 Line-Graph Games

Line-graph games are a special class of games with communication (graph) structure studied in Myerson (1977). We may assume, without loss of generality, that a line graph reflects the natural ordering from 1 to n. The structure on the set of players then is given by a line graph (N, L), where N is the set of players and $L \subseteq \overline{L} = \{\{i, i+1\} \mid i = 1, \dots, n-1\}$ is the set of (undirected) *edges*. Notice that $\overline{L}$ is a linear order on N. However, we allow for any subset L of $\overline{L}$ to be a line graph; hence, a line graph can consist of disconnected parts. Let $\mathcal{L}^N$ be the power set of $\overline{L}$, that is, the set of all line graphs on N given the natural ordering from 1 to n. We further denote the collection of all line-graph games on N by $\mathcal{G}^N \times \mathcal{L}^N$. For short, we denote the game (N, v) with line graph (N, L) as the *line-graph game* (v, L).

Following Myerson (1977) and Greenberg and Weber (1986), in a line-graph game $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, players can only cooperate when they are able to communicate with each other. This means that a coalition $S \subseteq N$ can only realize its worth $v(S)$ when S is *connected* in the line graph (N, L). Clearly, for the 'full' line graph $(N, \overline{L})$, the set $\mathcal{I}$ of (non-empty) connected coalitions is given by[4]

$$\mathcal{I} = \{S \subseteq N \mid S = [i, j],\ 1 \leq i \leq j \leq n\},$$

where $[i, j]$ denotes the set of consecutive players $\{i, i+1, \ldots, j-1, j\} \subseteq N$. For any line graph (N, L), the set of connected coalitions is a subset of $\mathcal{I}$ and consists of those coalitions $[i, j]$ where i and j belong to the same component. Connected coalition $T = [l, m] \in \mathcal{I}$ is a *component* in $L \subseteq \overline{L}$ if $\{i, i+1\} \in L$ for all $i \in [l, m-1]$ and $\{\{l-1, l\}, \{m, m+1\}\} \cap L = \varnothing$. For any coalition $S \subseteq N$, we denote the collection of components in line graph $L(S) = \{\{i, j\} \in L \mid i, j \in S\}$ by $C_L(S)$. When there is no ambiguity about the line graph L, we simply refer to this as the collection of components of S. Observe that the collection $C_L(S)$ of components of S forms a partition of S.

The set of connected coalitions in line graph $L \subseteq \overline{L}$ is denoted by

$$\mathcal{I}(L) = \{[i, j] \in \mathcal{I} \mid \text{there exists a } T \in C_L(N) \text{ such that } i, j \in T\}.$$

In the restricted game introduced by Myerson (1977) (for arbitrary graph games), a connected coalition earns its worth, but when coalition S is not connected, the players in S can only realize the sum of the worths of its components. So, for a given line-graph game (v, L), the restricted game $v^L \in \mathcal{G}^N$ induced by line graph (N, L) is given by[5]

$$v^L(S) = \begin{cases} v(S), & \text{if } S \in \mathcal{I}(L), \\ \sum_{T \in C_L(S)} v(T), & \text{if } S \notin \mathcal{I}(L). \end{cases} \tag{3}$$

We defined a line graph L to be any subset of the (complete) linear order $\overline{L}$. But, notice that restricting the restricted game v^L further on the linear order $\overline{L}$ does not have an impact: $v^L = (v^L)^{\overline{L}}$.

[4] In the more general model of Myerson (1977), the players belong to a communication structure that is represented by a *graph* (N, A), where the player set N is the set of nodes and where $A \subseteq \{\{i, j\} \mid i, j \in N,\ i \neq j\}$, a collection of unordered pairs, is the set of edges reflecting the communication possibilities among the players. A coalition $S \subseteq N$ can realize its worth $v(S)$ when S is *connected* in graph (N, A), that is, when for any two players i and j in S, there is a subset $\{\{i_k, i_{k+1}\} \mid k = 1, \ldots, t\} \subseteq A$ of edges such that $i_1 = i$, $i_{t+1} = j$, and $\{i_2, \ldots, i_t\} \subseteq S$.

[5] For definitions on arbitrary graph games, we refer to Myerson (1977).

Applying a formula stated in Owen (1986) for cycle-free graph games[6] gives the following useful expression for the Harsanyi dividends in line-graph games.

Theorem 1 *[Owen (1986), Bilbao (1998), van den Brink et al. (2007)]*[7] *Consider line-graph game* (v, L). *Then, the dividends of the restricted game* v^L *are given by* $\Delta_S(v^L) =$

$$\begin{cases} 0, & \text{if } S \notin \mathcal{I}(L), \\ v(\{i\}), & \text{if } S = \{i\}, \\ v[i,j] - v[i+1,j] - v[i,j-1] + v[i+1,j-1], & \text{if } S = [i,j] \in \mathcal{I}(L),\ j > i. \end{cases} \tag{4}$$

This implies that the dividend of any coalition S is fully determined by the worths of at most four coalitions, irrespective of the size of S. In contrast, for general cooperative games, the Harsanyi dividend of coalition S depends on the worths of all $2^{|S|}$ subsets of the coalition. Expression (4) turns out to be very useful for applications, as we will see in Sect. 3 where we discuss majority voting line-graph games.

It is well known that the restricted game of a superadditive game on the complete line graph $\overline{L}$ is balanced [see, for example Le Breton et al. (1992), Demange (1994), and Potters and Reijnierse (1995)]. This result follows immediately from Granot and Huberman (1982), who showed that a so-called permutationally convex game is balanced. More precisely, let u and ℓ be the two permutations on N defined by

$$u(i) = i,\ i = 1, \ldots, n,$$

respectively,

$$\ell(i) = n + 1 - i,\ i = 1, \ldots, n.$$

When v is superadditive, the restricted game $v^{\overline{L}}$ satisfies the permutational convexity condition of Granot and Huberman (1982) for the two permutations u and ℓ. Further, it then follows that the two marginal vectors $m^u(v^{\overline{L}})$ and $m^\ell(v^{\overline{L}})$ are in the core of $v^{\overline{L}}$ [see also Demange (2004)]. Since v^L is superadditive for any superadditive game v and any line graph $L \subseteq \overline{L}$,[8] these results also hold for any superadditive game v restricted to a line graph $L \subseteq \overline{L}$.

---

[6] See also Bilbao (1998) for the more general class of cycle-complete graphs.

[7] In van den Brink et al. (2007), the last line of (4) is shown to hold for $\overline{L}$. Applying it to $(v^L)^{\overline{L}}$ yields, for $[i, j] \in \mathcal{I}(L)$ and $j > i$, $v^L[i, j] - v^L[i+1, j] - v^L[i, j-1] + v^L[i+1, j-1]$. This expression reduces to (4) since $v^L(S) = v(S)$ and $\Delta_S(v^L) = \Delta_S(v^{\overline{L}})$ for any connected coalition $S \in \mathcal{I}(L)$.

[8] This follows since, if v is a superadditive game, then for any $S, T \subseteq N$ with $S \cap T = \varnothing$, it holds that $v^L(S \cup T) = \sum_{H \in C_L(S \cup T)} v(H) \geq \sum_{H \in C_L(S)} v(H) + \sum_{H \in C_L(T)} v(H) = v^L(S) + v^L(T)$.

Properties.

We recall some properties of solutions for line-graph games. First, component efficiency requires that the sum of the payoffs of the players in any component equals the worth of that component.

- A point-valued solution f on $\mathcal{G}^N \times \mathcal{L}^N$ satisfies **component efficiency** if $\sum_{i \in T} f_i(v, L) = v(T)$ for all $T \in C_L(N)$.

In van den Brink et al. (2007), four solutions are axiomatized with the use of component efficiency and one of the following four axioms, all of which concern the removal of edges.[9]

For $i = 1, \dots, n-1$, let $(N, L(i))$ be the graph on N, where $L(i) = L \setminus \{\{i, i+1\}\}$ is the set of edges obtained by deleting the edge $\{i, i+1\}$ from L. Notice that $(v, L(i))$ is also a line-graph game for every $i \in \{1, \dots, n-1\}$.

- A point-valued solution f on $\mathcal{G}^N \times \mathcal{L}^N$ is called **fair** if, for any $i = 1, \dots, n-1$ and any $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, it holds that $f_i(v^L) - f_i(v^{L(i)}) = f_{i+1}(v^L) - f_{i+1}(v^{L(i)})$.
- A point-valued solution f on $\mathcal{G}^N \times \mathcal{L}^N$ is called **upper equivalent** if, for any $i = 1, \dots, n-1$ and any $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, it holds that $f_j(v^L) = f_j(v^{L(i)})$, $j = 1, \dots, i$.
- A point-valued solution f on $\mathcal{G}^N \times \mathcal{L}^N$ is called **lower equivalent** if, for any $i = 1, \dots, n-1$ and any $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, it holds that $f_j(v^L) = f_j(v^{L(i)})$, $j = i+1, \dots, n$.
- A point-valued solution f on $\mathcal{G}^N \times \mathcal{L}^N$ is said to have the **equal loss property** if, for any $i = 1, \dots, n-1$ and any $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, it holds that $\sum_{j=1}^{i} (f_j(v^L) - f_j(v^{L(i)})) = \sum_{j=i+1}^{n} (f_j(v^L) - f_j(v^{L(i)}))$.

The first property is the famous fairness property introduced by Myerson (1977) for arbitrary graph games. It states that deleting the edge between i and $i+1$ hurts (or benefits) both players, i and $i+1$, equally. The equal loss property can also be conceived as referring to a type of fairness, but instead of the individual payoffs of the players on the edge that is deleted, it concerns the total payoff of *all* players at both sides of the deleted edge, requiring that these total payoffs change by the same amount. Upper equivalence requires that the payoff of a player does not depend on the presence of 'downward' edges, while lower equivalence requires that the payoff of a player does not depend on the presence of 'upward' edges. Which property is most appropriate depends on the respective application, something we will discuss after the next theorem and in the following sections.

Let f^u, f^ℓ, f^e, and f^s be the point-valued solutions on $\mathcal{G}^N \times \mathcal{L}^N$ defined by $f^u(v, L) = m^u(v^L)$, $f^\ell(v, L) = m^\ell(v^L)$, $f^e(v, L) = \frac{1}{2}(m^u(v^L) + m^\ell(v^L))$, and

[9] In van den Brink et al. (2007), these four solutions are axiomatized as so-called Harsanyi solutions of the restricted game. These latter solutions allocate the Harsanyi dividends of coalitions over the corresponding players according to a fixed weight system per coalition, which implies component efficiency.

$f^s(v, L) = Sh(v^L)$ for all $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$.[10] The solution f^s is known as the Myerson value and was introduced and axiomatized by Myerson (1977) for arbitrary graph games.

Theorem 2 *[van den Brink et al. (2007)]*
Let $f : \mathcal{G}^N \times \mathcal{L}^N \to \mathbb{R}^N$ be a component-efficient solution on the class $\mathcal{G}^N \times \mathcal{L}^N$ of line-graph games. Then,

(i) f is fair if and only if $f = f^s$.
(ii) f is upper equivalent if and only if $f = f^u$.
(iii) f is lower equivalent if and only if $f = f^\ell$.
(iv) f satisfies the equal loss property if and only if $f = f^e$.

Myerson (1977) already showed that, on the class of communication-graph games, the Shapley value f^s is characterized by component efficiency and fairness.[11] So, the 'if' part of item (i) in Theorem 2 also follows immediately from Myerson (1977). The 'only if' part shows that uniqueness also holds on the (smaller) subclass of line-graph games.

As mentioned before, for any superadditive line-graph game (v, L), both the lower equivalent solution f^ℓ and the upper equivalent solution f^u are in the core of the game; hence, all convex combinations, including the equal loss solution f^e, are also in the core. In the rest of this paper, we focus on a superadditive class of line-graph games, called majority voting line-graph games, where the Shapley value need not belong to the core. We note, however, that when v is convex, the Shapley value *does* belong to the core of the restricted game v^L. This follows (1) from the fact that the Shapley value of any convex cooperative game belongs to its core, and (2) from van den Nouweland and Borm (1991) [see also Algaba et al. (2001)], who show that when v is convex, the restricted game v^L is also convex.[12] What's more, when v is convex, as mentioned in van den Brink et al. (2007), the restricted game v^L is *almost positive*, that is, $\Delta_S(v^L) \geq 0$ whenever $|S| \geq 2$ [see also Vasil'ev (1978, 2006)].[13] For such games, not only does the core contain the Shapley value, but it also coincides with the so-called *selectope* or *Harsanyi set* [(Vasil'ev, 1978, 2006)], denoted by $H(v)$, which is, for any game, the set of all allocations that distribute the Harsanyi dividend of any coalition S over the players in S.[14] The following result thus follows from

[10] For sequencing games, Curiel et al. (1993, 1994) introduced the function f^e as the β-rule; see also the next section.

[11] See van den Brink (2001) for a related result on the class of cooperative games. A non-cooperative implementation of the Shapley value can be found in Pérez-Castrillo and Wettstein (2001). Slikker (2007) provides a strategic implementation of the Myerson value and other graph game solutions.

[12] In van den Nouweland and Borm (1991), this is shown for all so-called *cycle-complete* graphs, being those graphs such that if there is a cycle, then the subgraph on that cycle is complete. This class obviously contains all cycle-free graphs and thus all line graphs.

[13] In van den Brink et al. (2007), a line-graph game with an almost positive v^L is called *linear-convex*.

[14] See Vasil'ev (2006) and Derks et al. (2000) for these results. Since the Shapley value is one way of allocating the Harsanyi dividends (namely equally), it always belongs to the Harsanyi set.

van den Nouweland and Borm (1991) together with, for example, Vasil'ev (2006) or Derks et al. (2000).[15]

Corollary 1 *For any line-graph game $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$, if v is convex, then the restricted game v^L is almost positive and hence convex, which implies that $Sh(v^L) \in core(v^L) = H(v^L)$.*

Examples of a convex line-graph game are Ambec and Sprumont's (2002) river game and Curiel et al.'s (1989) sequencing game. We now turn to a class of line-graph games, majority voting line-graph games, that are superadditive but not convex.

3 Political Power in Majority Voting Line-Graph Games

In this section, we consider majority voting games between political parties in a parliament. We consider majority voting games where the political parties are ordered on a line according to their political preferences over, for example, economic policy, ethical issues, environmental problems, and so on.

A *majority voting situation* consists of (i) a set of political parties, (ii) a number of seats for each party, and (iii) a quota expressing how many of the seats are necessary to pass a bill. Formally, a majority voting situation is a triple (N, s, q), where

1. $N = \{1, \dots, n\}$ is the set of players representing the parties in a parliament,
2. $s = (s_i)_{i \in N}$ is the seat distribution with s_i the number of seats (votes) of party (player) $i \in N$, and
3. q such that $\frac{1}{2}\sum_{i \in N} s_i < q \leq \sum_{i \in N} s_i$ is the quota, being the minimum number of seats necessary to pass a ballot.

We denote the total number of seats by $w = \sum_{i \in N} s_i$. Voting power refers to the ability of the political parties to turn losing coalitions into winning ones, or vice versa. To determine these abilities, it is sufficient to know what the winning and losing coalitions are. This can be summarized by an associated *majority voting game*: a cooperative game where the worth of any coalition is either one (when it has a qualified majority) or zero (when it does not have a qualified majority). In other words, the majority voting game associated with voting situation (N, s, q) is the game (N, v) given by

$$v(S) = \begin{cases} 1 & \text{if } \sum\limits_{i \in S} s_i \geq q, \\ 0 & \text{if } \sum\limits_{i \in S} s_i < q. \end{cases}$$

A coalition S with $v(S) = 1$ is called a *winning* coalition, and a coalition S with $v(S) = 0$ is called a *losing* coalition. A winning coalition S is called a *minimal*

[15] Note that, given expression (4), for Corollary 1 to hold, full convexity of v is sufficient but not necessary. Instead, we may require a weaker property: $v(A \cup B) + v(A \cap B) \geq v(A) + v(B)$ for all $A, B \subseteq N$ such that $|A \setminus B| = |B \setminus A| = 1$.

winning coalition (MWC) if $v(S \setminus \{i\}) = 0$ for all $i \in S$. Two special types of players are veto players and null players. A player i is a *veto player* if $v(S) = 1$ implies that $i \in S$. A player i is a *null player* if $v(S \setminus \{i\}) - v(S) = 0$ for any S containing i. Notice that, since $v(N) = 1$, every null player in a majority voting game is a non-veto player, but there can be non-veto players who are not null players.

A majority voting game is a special type of a *simple game*: a game v such that $v(S) \in \{0, 1\}$, $v(\varnothing) = 0$, and $v(N) = 1$.[16] It is well known that a simple game has a non-empty core if and only if there is at least one veto player. Furthermore, the core distributes the worth $v(N) = 1$ among the veto players and assigns a zero power to all non-veto players. This follows straightforwardly from the observation that the sum of the non-negative powers of all parties in a winning coalition, including the grand coalition N, must be equal to one.

Two well-known power indices that measure the voting power of political parties in majority voting situations are the Shapley–Shubik index (Shapley & Shubik, 1954) and the Banzhaf index (Banzhaf, 1965). These can be obtained by applying the Shapley value (Shapley, 1953), respectively, the Banzhaf value (Owen, 1975), to the associated majority voting game. Since both values assign positive power to non-null players in a majority voting game, they assign positive power to non-veto players, even when there are veto players, and thus the associated power vector is not in the core of the majority game.

We now consider a situation where the parties can be ordered linearly according to their political preferences. Without loss of generality, suppose that the parties can be indexed successively from player 1 (the most left-wing party) to player n (the most right-wing party). We refer to a pair $(v, L) \in \mathcal{G}^N \times \mathcal{L}^N$ with v a majority voting game, as a *majority voting line-graph game* (or, for short, *majority line-graph game*). In such a political structure, it is reasonable to suppose that only connected coalitions will form; that is, this situation can be modelled by the line-graph game $(v, \overline{L})$ where $\mathcal{I} = \{S \subseteq N \mid S = [i, j]$ for some $i <= (smallerorequal) j\}$ is the collection of feasible coalitions. If, for some reason, two idealogical neighbours refuse to cooperate, then the cooperation restrictions can be modelled by a line-graph game (v, L) with $L \subset \overline{L}$, $L \neq \overline{L}$.[17] As mentioned in Sect. 2, since v is superadditive, it follows from Granot and Huberman (1982), Le Breton et al. (1992), Demange (1994), and Potters and Reijnierse (1995) that the core of the restricted game v^L is non-empty and, specifically, that $f^u(v, L)$, $f^\ell(v, L)$, and $f^e(v, L)$ belong to the core of the restricted game v^L. Since the core is non-empty, v^L has at least one veto player. Indeed, if $v^L(N) = 1$, since only coalitions of successive parties can form, there is at least one player who necessarily belongs to both the most left-wing MWC

[16] $v(\varnothing) = 0$ requires that unanimous opposition implies rejection, and $v(N) = 1$ requires that unanimous support implies acceptance. In the literature, sometimes $v(N) = 1$ is not required for a game to be a simple game.

[17] Note that even though we require that the grand coalition be winning in the majority game v (that is, $v(N) = 1$), we do not require that N be winning in the restricted game v^L. That is, $v(N) = 1$ does *not* imply $v^L(N) = 1$. For example, when $L = \varnothing$ and $s_i < q$ for all $i \in N$, $v(N) = 1$ but $v^L(N) = 0$.

and the most right-wing MWC and thus to any majority coalition.[18] Moreover, the set of veto players is a connected coalition in the line graph L. If $v^L(N) = 0$, then all players are veto players and the unique core power vector assigns zero power to all players.

Proposition 1 *Consider the majority line-graph game* (v, L).

(i) *For* $h \leq k$, *let* $[1, k]$ *be the most left-wing MWC (i.e.* $v([1, k]) = 1$ *and* $v([1, k-1]) = 0$*) in* $(v, \overline{L})$, *and* $[h, n]$ *be the most right-wing MWC (i.e.* $v([h, n]) = 1$ *and* $v([h+1, n]) = 0$*) in* $(v, \overline{L})$. *Then* $[h, k]$ *is the set of veto players in* $v^{\overline{L}}$.

(ii) *For every* $L \subseteq \overline{L}$, *there is at most one component* $T = [p, q] \in C_L(N)$ *such that* $v(T) = 1$. *Let* $[p, k]$ *be the most left-wing MWC (i.e.* $v([p, k]) = 1$ *and* $v([p, k-1]) = 0$*) in* (v, L), *and* $[h, q]$ *be the most right-wing MWC (i.e.* $v([h, q]) = 1$ *and* $v([h+1, q]) = 0$*) in* (v, L). *Then* $[h, k]$ *is the set of veto players in* v^L.[19]

Proof Since (i) follows as a corollary from (ii) (for $T = N$, $p = 1$, and $q = n$), we only prove (ii). We show that, for all $i \in [h, k]$, if $i \notin S \subset N$, then $v^L(S) = 0$. First, note that $v(S) = 0$ if $\{h, k\} \not\subset S$ since h and k are veto players. It follows that, for all $i \in [h, k]$ and $S \subseteq N \setminus \{i\}$, $v^L(S) = v^L(S \cap [1, i-1]) + v^L(S \cap [i+1, n]) = 0$ since $k \notin [1, i-1]$ and $h \notin [i+1, n]$. □

Example 1 Take $N = \{1, 2, 3, 4, 5, 6, 7\}$, $w = 85$ with $s_1 = s_2 = s_6 = s_7 = 10$, $s_3 = s_4 = s_5 = 15$, and $q = 60$. In the restricted game $v^{\overline{L}}$, $[1, 5]$ is the most left-wing MWC, $[3, 7]$ the most right-wing MWC, and $[3, 5] = \{3, 4, 5\}$ the set of veto players. □

In the remainder of this section, we focus on the full line graph $\overline{L}$, but the results can be straightforwardly generalized to any $L \subseteq \overline{L}$ by considering the unique winning component $T \in C_L(N)$, if any, rather than N. When $h = k$, it follows that $f^u(v, \overline{L}) = f^\ell(v, \overline{L}) = f^e(v, \overline{L}) = e(h)$, where $e(i) \in \mathbb{R}^N$, given by $e_i(i) = 1$ and $e_j(i) = 0$ for all $j \neq i$, is the unique core element. When $h < k$, then $f^u(v, \overline{L}) = e(k)$, $f^\ell(v, \overline{L}) = e(h)$, and $f^e(v, \overline{L}) = \frac{1}{2}(e(k) + e(h))$. So, $f^u(v, \overline{L})$ assigns full power to the most right-wing veto player, $f^\ell(v, \overline{L})$ assigns full power to the most left-wing veto player, and $f^e(v, \overline{L})$ divides the power equally between the two extreme veto players. Observe that, according to these solutions, no power is assigned to any other player, including intermediate veto players between the two extreme veto players, h and k. This seems reasonable when the two extreme veto players are considered to be critical. When the most left-wing coalition $[1, k]$ is formed, the most right-wing veto player k has the highest incentive (or lowest objection) to break away and form another MWC. So, if this player is willing to cooperate in $[1, k]$, then it can be expected that any other player in $[1, k]$ is willing to cooperate in $[1, k]$, including any other veto player. Similarly, this holds for h in the MWC $[h, n]$. The equal loss solution f^e, giving both players a power of $\frac{1}{2}$,

[18] Notice that, since $q > w/2$, the restricted game v^L is a *proper* simple game; that is, the complement of any winning coalition, including a MWC, is a losing coalition.

[19] Note that a MWC in the line-graph game (v, L) need *not* be a MWC in the majority game v.

seems to be an appropriate power index for a political situation before it is known whether a left-wing or right-wing majority coalition will be formed. Note that both the Shapley value and the (normalized) Banzhaf power index assign positive power to every player and, thus, they are not in the core.

The next proposition states that, in the restricted game $v^{\overline{L}}$, the dividend of each MWC in $\mathcal{I}$ is equal to 1. The dividend of any other coalition in $\mathcal{I}$ is equal to 0 or -1.[20]

Proposition 2 *Let $v^{\overline{L}}$ be the restricted game of a majority line-graph game $(v, \overline{L})$. Then, for $S \in \mathcal{I}$, $\Delta_S(v^{\overline{L}}) = 1$ if S is a MWC and $\Delta_S(v^{\overline{L}}) \in \{-1, 0\}$ otherwise.*

Proof First, observe that $\Delta_T(v^{\overline{L}}) = 0$ for any $T = [i, j]$ with $v[i, j] = 0$. Next, let $S = [i, j]$ be a MWC; that is, $v[i, j] = 1$ and $v(T) = 0$ for all $T \subset S$, $T \neq S$. From Theorem 1, it follows that $\Delta_{[i,j]}(v^{\overline{L}}) = v[i, j] - v[i, j-1] - v[i+1, j] + v[i+1, j-1] = 1 - 0 - 0 + 0 = 1$. Next, given a MWC $[i, j]$, consider any coalition $[i, k]$ with $k > j$. Then, $k - 1 \geq j$ and thus $v[i, k] = v[i, k-1] = 1$. Further, if $v[i+1, k-1] = 1$, then $v[i+1, k] = 1$, and thus $\Delta_{[i,k]}(v^{\overline{L}}) = 0$ in that case. Otherwise, if $v[i+1, k-1] = 0$, since $v[i+1, k] \in \{0, 1\}$, $\Delta_{[i,k]}(v^{\overline{L}}) \in \{-1, 0\}$. Similarly, given a MWC $[i, j]$, this holds for any coalition $[h, j]$ with $h < i$. Finally, given a MWC $[i, j]$, when $S = [h, k]$ is such that $h < i < j < k$, then $v[h, k] = v[h+1, k] = v[h, k-1] = v[h+1, k-1] = 1$ and thus $\Delta_{[h,k]}(v^{\overline{L}}) = 0$. □

Since the worth $v^{\overline{L}}(N) = 1$ is equal to the sum of all dividends in the restricted game, this proposition implies that the number of coalitions with dividend equal to -1 is one fewer than the number of MWCs.

Notice that, in a standard majority game, it follows from Formula (1) that $\Delta_S(v) = 1$ if S is a MWC, since $v(S) = 1$ and $v(T) = 0$ for any $T \subset S$, $T \neq S$. However, as the next example shows, in a standard majority game v, other (winning) coalitions may also have a positive dividend and even a dividend larger than one.

Example 2 Take $N = \{1, 2, 3, 4, 5\}$, $s_i = 1$ for all $i \in N$ and $q = 3$, so that $v(S) = 1$ if and only if $|S| \geq 3$. Hence, any coalition of precisely three players is a MWC and has a dividend of one. Further, any coalition of four players contains four subcoalitions of three players. By applying Formula (2), it follows that $\Delta_S(v) = -3$ when $|S| = 4$. Finally, the grand coalition N contains ten subcoalitions of three players, each with dividend 1, and five subcoalitions of four players, each with dividend -3. Hence, the dividends of N's subcoalitions sum up to $10 + 5(-3) = -5$, which means that $\Delta_N(v) = 1 - (-5) = 6$. □

[20] This shows that the restricted game of a majority line-graph game is not almost positive.

4 Rewarding Intermediate Veto Players: Hierarchical Outcomes and the τ-Index

A power index is a mapping f that assigns a power vector $f(v, L)$ to every majority line-graph game (v, L). In this paper, we are mainly interested in power indices that assign a power vector in the core of the restricted game v^L.[21]

Axiom *A power index f for majority line-graph games is called* **core stable** *if, for every majority line-graph game (v, L), it holds that $f(v, L) \in core(v^L)$.*

From the indices considered in the previous section, f^u, f^ℓ, and f^e reward either one or both of the extreme veto players, but assign zero power to the intermediate veto players. As we saw, applying the Shapley value to the restricted line-graph game does reward the intermediate veto players, but it also rewards non-veto players and thus assigns a power vector that does not belong to the core of the restricted game. (The same holds for the Banzhaf value.)

Next, we consider two types of power indices that do reward intermediate veto players without also rewarding non-veto players. Thus, unlike the Shapley and Banzhaf values, these indices are core stable.

4.1 Hierarchical Outcomes

The upper and lower equivalent solutions for line-graph games are examples of hierarchical outcomes. Hierarchical outcomes are defined by Demange (2004) for connected cycle-free communication-graph games as specific core-stable payoff or power vectors of the restricted game. In this paper, we consider line graphs (being a special type of cycle-free graphs), but do not require that the graph be connected. We directly define the hierarchical outcomes for this type of graph games, and refer to Demange (2004) for definitions on connected cycle-free graph games.

First, consider the line-graph game (v, L). For each player, there is a corresponding hierarchical outcome. For $i \in N$, let $C_L^i(N) = [h_i, k_i] \in C_L(N)$ be such that $i \in C_L^i(N)$, that is, $C_L^i(N)$ is the component in L that contains player i.

The *hierarchical outcome* corresponding to player i is the payoff vector given by[22]

[21] We remark that a different core concept for games with restricted cooperation assigns to every graph game the set of component-efficient payoff vectors such that the sum of the payoffs of all players in any connected coalition is at least equal to the worth of this coalition. For line-graph games, this gives the solution $\overline{core}(v, L) = \{x \in \mathbb{R}^N \mid \sum_{i \in C} x_i = v(C)$ for all $C \in C_L(N)$, and $\sum_{i \in S} x_i \geq v(S)$ for all $S \in \mathcal{I}(L)\}$. It is obvious that $\overline{core}(v, L) = core(v^L)$ if the game v is monotonic (i.e. $v(S) \leq v(T)$ if $S \subseteq T \subseteq N$) and superadditive. Since majority voting games are monotonic and superadditive, for the games considered in this paper, the two core concepts boil down to the same.

[22] With some abuse of notation, we define $[h, k]$ to be the empty set $\varnothing$ if $k < h$.

$$h_j^i(v, L) = \begin{cases} v([h_i, j]) - v([h_i, j-1]) & \text{if } h_i \leq j < i \\ v([j, k_i]) - v([j+1, k_i]) & \text{if } i < j \leq k_i \\ v(C_L^i(N)) - v([h_i, j-1]) - v([j+1, k_i]) & \text{if } j = i \\ 0 & \text{if } j \in N \setminus C_L^i(N). \end{cases} \quad (5)$$

Thus, the hierarchical outcome $h^i(v, L)$ allocates to player j to the left (respectively, to the right) of i the contribution of player j to all players to its left (respectively, right) in its component, while player i gets the surplus of its component that is left after all other players are assigned their power.

Example 3 Consider the majority voting situation of Example 2, that is, $s_i = 1$ for all $i \in N = \{1, 2, 3, 4, 5\}$, and $q = 3$. Further, let $L = \{\{1, 2\}, \{3, 4\}, \{4, 5\}\}$. Then, $h_1^3(v, L) = h_2^3(v, L) = h_5^3(v, L) = 0$, $h_4^3(v, L) = v(\{4, 5\}) - v(\{5\}) = 0 - 0 = 0$, and $h_3^3(v, L) = v(\{3, 4, 5\}) - v(\{4, 5\}) = 1 - 0 = 1$, yielding $h^3(v, L) = (0, 0, 1, 0, 0)$. □

For majority line-graph games, we will refer to the power index that assigns to every majority line-graph game (v, L) the hierarchical outcome $h^i(v, L)$ as the *i-hierarchical index*. Notice that, for $L = \overline{L}$, the n-hierarchical index (respectively, the 1-hierarchical index) is equivalent to the upper equivalent solution f^u (respectively, the lower equivalent solution f^ℓ) applied to this class of majority line-graph games. Further, f^e is obtained as the average of these two hierarchical indices.

As Demange (2004) has showed, for superadditive games (N, v), and connected cycle-free (tree) graphs (N, L), each hierarchical outcome of the associated tree game (v, L), $h^i(v, L)$, $i \in N$, is an extreme core allocation of the restricted game v^L. Since majority games are superadditive, this also holds for majority line-graph games with $L = \overline{L}$.

Herings et al. (2008) characterized the *average tree solution* which assigns to every cycle-free graph game the (component-wise) average hierarchical outcome given by

$$\overline{h}_j(v, L) = \frac{1}{|C_L^j(N)|} \sum_{i \in C_L^j(N)} h_j^i(v, L) \text{ for all } j \in N. \quad (6)$$

We refer to the power index that assigns to every majority line-graph game (v, L), the average hierarchical outcome $\overline{h}(v, L)$ as the *hierarchical index*. By convexity of the core of a game and the fact that all hierarchical outcomes belong to the core of the restricted game, the hierarchical index is core stable. From now on, we denote by $Veto(v, L)$ the set of veto players in the restricted game v^L.

For each $i \in N$, the hierarchical outcome associated with i can be expressed as follows. If $v^L(N) = 0$, then all players are veto players and every hierarchical outcome assigns zero power to all players, which is also the unique core power vector. If $v^L(N) = 1$, then the set of veto players, $Veto(v, L)$, is a subset of a component, $Veto(v, L) \subseteq S$ for some $S \in C_L(N)$. It follows then that the hierarchical outcome associated with any player outside S is the zero vector. Next, consider the hierarchical outcomes associated with players in S. As S is connected, each hierarchical outcome

$h^i(v, L)$, for $i \in S$, is an extreme point in the core and hence allocates full power to a veto player. Let $Veto(v, L) = [h, k]$. For $i \in Veto(v, L)$, $h^i(v, L) = e(i)$. For $i \in S \setminus Veto(v, L)$, if $i < h$, then $h^i(v, L) = e(h)$, and if $i > k$, then $h^i(v, L) = e(k)$. In other words, the hierarchical outcome associated with a non-veto player $i \in S$ is the extreme core-stable power vector that assigns full power to the veto player that is closest to i. This is summarized in the following theorem.

Theorem 4 *Let v^L be the restricted game of a majority line-graph game (v, L). If $v^L(N) = 0$, then $h(v, L) = \mathbf{0}$. If $v^L(N) = 1$, then there is exactly one $S \in C_L(N)$ with $v^L(S) = 1$ and, denoting $Veto(v, L) = [h, k] \subseteq S$,*

$$h^i_j(v, L) = \begin{cases} 1 & \text{if } (j = i \in Veto(v, L)), \\ & \text{or } (j = h \text{ and } i \in S \text{ with } i < h), \\ & \text{or } (j = k \text{ and } i \in S \text{ with } i > k), \\ 0 & \text{otherwise.} \end{cases}$$

Using Theorem 4, it is now easy to express the hierarchical index $\bar{h}(v, L)$ given by (6) for majority line-graph games as follows.

Theorem 5 *Let (v, L) be a majority line-graph game with restricted game v^L such that $v^L(N) = 1$. Further, let $[l, m]$ denote the component containing the veto players in (v, L), that is, $v^L([l, m]) = 1$. We distinguish two cases.*

(1) Suppose that $|Veto(v, L)| = 1$. Then,

$$\bar{h}_i(v, L) = \begin{cases} 1 & \text{if } \{i\} = Veto(v, L), \\ 0 & \text{otherwise.} \end{cases}$$

(2) Suppose that $|Veto(v, L)| > 1$. Let $Veto(v, L) = [h, k] \subseteq [l, m] \in C_L(N)$. Then,

$$\bar{h}_i(v, L) = \begin{cases} \frac{h-l+1}{m-l+1} & \text{if } i = h, \\ \frac{m-k+1}{m-l+1} & \text{if } i = k, \\ \frac{1}{m-l+1} & \text{if } |Veto(v, L)| > 2 \text{ and } i \in [h+1, k-1], \\ 0 & \text{otherwise.} \end{cases}$$

In majority line-graph games, the hierarchical index rewards all (and only) veto players. It thus belongs to the core of the restricted game.

Corollary 2 *The hierarchical index $\bar{h}$ is core stable, that is, if v^L is the restricted game of a majority line-graph game (v, L), then $\bar{h}(v, L) \in core(v^L)$.*

It also follows from Theorem 5 that the hierarchical index does not reward all veto players equally. More precisely, it rewards the two extreme veto players h and k more than it does the intermediate veto players in $[h + 1, k - 1]$. Further, it rewards the left (respectively, right) more than it does the right (respectively, left) extreme veto player, depending on the number of other (non-veto) players positioned to the left (respectively, to the right) of the respective veto player in the component.

Herings et al. (2008) axiomatized[23] the average tree solution by component efficiency and a component fairness property, which says that breaking an edge in a cycle-free graph game has the same per capita effect on the payoffs of the players in the two newly created components. In this paper, we focus on core-stable power indices. Notice that core stability implies component efficiency. It turns out that, to characterize the hierarchical index using core-stability, we can do with a weaker component fairness axiom, where we only consider the breaking of edges between two veto players. Recall that $L(i) = L \setminus \{\{i, i+1\}\}$ is the set of edges obtained by deleting the edge $\{i, i+1\}$ from L.

Axiom *A power index f for majority line-graph games is called* **component veto fair** *if, for every majority line graph (v, L) with $Veto(v, L) = [h, k]$, and for all $i \in [h, k-1]$, it holds that*

$$\frac{1}{|C^i_{L(i)}(N)|} \sum_{j \in C^i_{L(i)}(N)} \big(f_j(v, L) - f_j(v, L(i))\big)$$
$$= \frac{1}{|C^{i+1}_{L(i)}(N)|} \sum_{j \in C^{i+1}_{L(i)}(N)} \big(f_j(v, L) - f_j(v, L(i))\big)$$

□

The hierarchical index $\bar{h}$ is the only power index that satisfies core stability and component veto fairness.

Theorem 7 *Let f be a power index for majority line-graph games. Then, f is core stable and component veto fair if and only if $f = \bar{h}$.*

Proof The 'if' part follows from the more general result in Herings et al. (2008) and Corollary 2. Uniqueness could be showed in a way similar to that in Herings et al. (2008). But since it is not a corollary of their result, and we use the stronger core stability but weaker component veto fairness property, we give the uniqueness proof for completeness.[24]

Let f be a power index for majority line-graph games that satisfies core stability and component veto fairness. If $v^L(N) = 0$, then core stability implies that $f(v, L) = \mathbf{0}$.

Now, suppose that $v^L(N) = 1$. By core stability, we have

$$f_i(v, L) = 0 \text{ for all } i \in N \setminus Veto(v, L). \tag{7}$$

[23] A strategic implementation of the hierarchical outcomes and their average can be found in van den Brink et al. (2013).

[24] We remark that the proof follows the one in Herings et al. (2008), except that we have fewer equations from component veto fairness, but more equations from core stability.

Let $Veto(v, L) = [h, k] \subseteq [l, m] \in C_L(N)$. We distinguish the following two cases.

When $|Veto(v, L)| = 1$, then $f_j(v, L) = 1$ for $\{j\} = Veto(v, L)$ follows from core stability (particularly, from efficiency). This, together with the $n - 1$ equations in (7), determines $f(v, L)$.

Suppose that $|Veto(v, L)| > 1$. We proceed by induction on the number of edges $|L|$. When $L = \emptyset$, core stability implies that the core consists of a unique vector, the zero vector. Hence, $f_i(v, L) = 0$ for all $i \in N$. Assume that $f(v, L')$ is determined for all L' with $|L'| < |L|$.

By component veto fairness, we have, for all $i \in [h, k - 1]$,

$$\begin{aligned}
&\frac{1}{|C^i_{L(i)}(N)|} \sum_{j \in C^i_{L(i)}(N)} \big(f_j(v, L) - f_j(v, L(i))\big) \\
&= \frac{1}{|C^{i+1}_{L(i)}(N)|} \sum_{j \in C^{i+1}_{L(i)}(N)} \big(f_j(v, L) - f_j(v, L(i))\big). \qquad (8)
\end{aligned}$$

Next, by core stability and (7), we have

$$\sum_{i \in [h,k]} f_i(v, L) = 1. \qquad (9)$$

Since $f(v, L(i))$ in (8) is determined by the induction hypothesis, there are $n - (k - h + 1) = n - k + h - 1$ equations of type (7), and $k - h$ equations of type (8). Hence, together with the equation in (9), there are n independent equations that determine the n unknowns $f_i(v, L)$, $i \in N$. Since the hierarchical index satisfies the two axioms, f must be equal to $\bar{h}$. □

4.2 The τ-Index

An interesting power index for political line-graph games is the solution that assigns *equal* power to all veto players and zero power to all non-veto players. It turns out that this power index is obtained by applying the τ-value (Tijs, 1981) to the associated restricted game.

First, observe that the restricted game of a majority line-graph game is quasi-balanced and applying the τ-value gives the following power vectors.

Theorem 8 *Let v^L be the restricted game of a majority line-graph game (v, L). Then, $v^L \in QB^N$. If $v^L(N) = 0$, then $\tau(v^L) = \mathbf{0}$. If $v^L(N) = 1$, then*

$$\tau_i(v^L) = \begin{cases} \frac{1}{|Veto(v,L)|} & \text{if } i \in Veto(v, L), \\ 0 & \text{otherwise.} \end{cases}$$

Proof Let v^L be the restricted game of a majority line-graph game (v, L). If $v^L(N) = 0$, then $M(v^L) = m(v^L) = \mathbf{0}$, and the result follows. Now, suppose that $v^L(N) = 1$. Then,

$$M_i(v^L) = \begin{cases} 1 & \text{if } i \in Veto(v, L), \\ 0 & \text{otherwise.} \end{cases}$$

Next, consider a player's minimal right and distinguish the following two cases:

(1) Suppose that $|Veto(v, L)| > 1$. We then have

$$m_i(v^L) = 0 \text{ for all } i \in N.$$

(2) Suppose that $|Veto(v, L)| = 1$. We then have

$$m_i(v^L) = \begin{cases} 1 & \text{if } i \in Veto(v, L), \\ 0 & \text{otherwise.} \end{cases}$$

In both cases, v^L is quasi-balanced and $\tau(v^L)$ is as stated in the theorem. □

Since the τ-value allocates the whole power in the game over the veto players, and none to non-veto players, we immediately have the following corollary.

Corollary 3 *Let v^L be the restricted game of a majority line-graph game (v, L). Then, $\tau(v^L) \in core(v^L)$.*

Next, we consider the power index that assigns to every majority line-graph game (v, L), the τ-value of the restricted game v^L: $\tau(v, L) = \tau(v^L)$. We refer to this power index as the *τ-index*. We axiomatize this power index with the following two axioms. The first is core stability as defined in the previous subsection, which requires that the index always assigns a power vector that belongs to the core of the restricted game. By Corollary 3, this is obviously satisfied for the τ-index.

It is straightforward to show that Theorem 2.(i) also holds for the specific class of majority line-graph games; that is, the Shapley value is the unique solution for majority line-graph games that satisfies component efficiency and fairness. Since every core-stable solution satisfies component efficiency and the Shapley value is not core stable, this implies that there is no core-stable solution that satisfies fairness.[25] It turns out that the τ-index satisfies a weaker fairness property where equal gains or losses are required only when one breaks an edge between veto players.

Axiom *A power index f for majority line-graph games is called* **veto fair** *if, for every majority line graph (v, L) with $Veto(v, L) = [h, k]$, it holds that*

$$f_i(v, L) - f_i(v, L(i)) = f_{i+1}(v, L) - f_{i+1}(v, L(i)) \text{ for all } i \in [h, k-1].$$

[25] In van den Brink et al. (2021), a similar observation is made about assignment games. They characterize the τ-value in such games, where it coincides with the fair division point (Thompson, 1981), see Núñez and Rafels (2002).

The τ-index is the only power index that satisfies core stability and veto fairness.

Theorem 10 *Let f be a power index for majority line-graph games. Then f is core stable and veto fair if and only if $f = \tau$.*

Proof From Corollary 3 it follows that the τ-index is core stable. Veto fairness follows since (i) if $v^L(N) = 0$ then $v^{L'}(N) = 0$ for all $L' \subset L$ and $\tau(v, L) = \mathbf{0}$, and (ii) if $v^L(N) = 1$ then, for every $i, i+1 \in Veto(v, L)$,

$$\tau_i(v, L) - \tau_i(v, L(i)) = \frac{1}{|Veto(v, L)|} - 0 = \tau_{i+1}(v, L) - \tau_{i+1}(v, L(i)).$$

The proof of uniqueness follows a well-known structure. To prove uniqueness, suppose that f is a power index for majority line-graph games that satisfies core stability and veto fairness. Core stability implies that $f(v, L) = \mathbf{0}$ if $v^L(N) = 0$.

Next, suppose that $v^L(N) = 1$. Core stability implies that $f_i(v, L) = 0$ for all $i \in N \setminus Veto(v, L)$.

Recall that the set of veto players is a set of connected players $[h, k]$; see Proposition 1.

If $|Veto(v, L)| = 1$, then by component efficiency $f_i(v, L) = 1$, with $\{i\} = Veto(v, L)$, is determined by core stability (which implies component efficiency).

If $|Veto(v, L)| > 1$, then we prove uniqueness by induction on $|L|$. If $|L| = 0$, that is, $L = \varnothing$, then, since there are at least two veto players, $core(v^L)$ is a singleton with the zero vector as its only element, and thus $f_i(v, L) = 0$ for all $i \in N$.[26] Proceeding by induction, assume that $f(v, L')$ is uniquely determined for all L' with $|L'| < |L|$. Let $Veto(v, L) = [h, k]$.

Veto fairness implies that

$$f_i(v, L) - f_i(v, L(i)) = f_{i+1}(v, L) - f_{i+1}(v, L(i)) \text{ for all } i \in [h, k-1]. \quad (10)$$

Core stability implies that

$$f_i(v, L) = 0 \text{ for all } i \in N \setminus [h, k], \quad (11)$$

and thus

$$\sum_{i \in [h,k]} f_i(v, L) = 1. \quad (12)$$

Since $f(v, L(i))$ is determined by the induction hypothesis, (10)–(12) give $(k - 1 - h + 1) + (n - k + h - 1) + 1 = n$ independent equations in the n unknown powers $f_i(v, L)$, $i \in N$, implying that $f(v, L)$ is uniquely determined. Since the τ-index satisfies the axioms, f must be equal to τ. □

Notice that in the axiomatization of the τ-index in Theorem 10, we used a similar weakening of fairness as we did with component fairness for the hierarchical index $\bar{h}$

[26] Recall also the observation in Footnote 16.

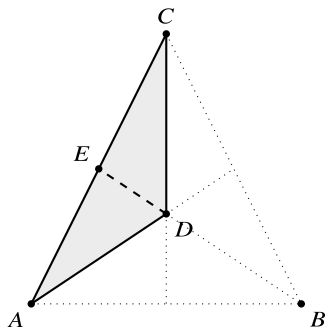


Fig. 1 Core of a majority line-graph game (v, L) with three veto players, where $A = (1, 0, 0)$, $B = (0, 1, 0)$, $C = (0, 0, 1)$, $D = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, and $E = (\frac{1}{2}, 0, \frac{1}{2})$

in Theorem 7. Instead of deleting any edge between two neighbours, we only consider deleting edges between veto players. But also notice the different impact of these results. The standard component fairness axiom is compatible with core stability, but requiring stronger core stability, instead of component efficiency, allows us to use the weaker component veto fairness. The standard fairness axiom is incompatible with core stability, but the weaker veto fairness is compatible with core stability and the two axioms together characterize the τ-index.

4.3 Illustration

We conclude this section by visually comparing the power indices discussed in this paper using a majority line-graph game (v, L) with three veto players, $|Veto(v, L)| = 3$, and any number of other non-veto players. Suppose that the veto set is part of a connected component with $v^L(N) = 1$, that is, $Veto(v, L) \subseteq [l, m] \in C_L(N)$. As the power indices we are concerned with are core stable, all of them assign zero power to players outside $Veto(v, L)$, and, hence, we can focus on the three veto players.

The simplex in Fig. 1 illustrates the core of the game, and thus, any point in ABC is a core allocation (and any point outside it is not). Points A and C are where the left and right, respectively, extreme veto players are assigned full power. Similarly, at B, the intermediate veto player gets full power. In point D, all veto players get equal power. Our interest, in this paper, has been in the power allocations in the shaded core area, ADC. As we know from Theorem 8, the τ-index allocation depends only on the number of veto players, since it allocates the full power of one equally among all veto players. Thus, regardless of the size of $N \setminus Veto(v, L)$ or the edges outside $[l, m]$, $\tau(v, L) = D$, where each veto player is assigned equal power. Similarly, regardless of $N \setminus Veto(v, L)$ or the edges outside $[l, m]$, $f^u(v, L) = C$, $f^l(v, L) = A$, and $f^e(v, L) = E$, since f^u (respectively, f^l) assigns full power to the right (respectively, left) extreme veto player, while f^e equally splits the power between them.

Now, suppose that $Veto(v, L) = [j, j+1, j+2]$ and consider the hierarchical indices. Any i-hierarchical index where $l \leq i \leq j$ allocates full power to the left-most extreme veto player j and hence in that case $h^i(v, L) = e(j) = A$. Similarly, for $j+2 \leq i \leq m$, the i-hierarchical index allocates full power to the right-most veto player $j+2$, and thus in that case $h^i(v, L) = e(j+2) = C$. When $i = j+1$, $h^i(v, L) = e(j+1) = B$.

Next, consider the hierarchical index $\bar{h}$. This index may assign any allocation in triangle ADC, *excluding* the sides AC, AD, and DC, but including point D. The location of $\bar{h}$ depends on the size, and distribution along the line, of $[l, m] \setminus Veto(v, L)$. If the veto component coincides with the veto set, $[l, m] = Veto(v, L)$, then $\bar{h}(v, L) = D$. Now, as one starts adding an increasing *equal* number of non-veto players to the left of j and to the right of $j+2$, the hierarchical index moves north-west along DE, tending towards E as more players are added on both sides (but never reaching E). When the number of players to the left of the left-most extreme veto player j is higher than the number of players to the right of the right-most extreme veto player $j+2$, then $\bar{h}$ lies in the interior of ADE. When the reverse is the case, $\bar{h}$ lies in the interior of CDE. This means that the $\bar{h}$ index of, particularly, the extreme veto players depends crucially on the respective left/right-side non-veto players. Under $\bar{h}$, we might think of the set of non-veto players on the respective left/right side as giving leverage for the respective left/right-most extreme veto player.

5 Concluding Remarks

Our main focus in this paper has been on majority games where players can be ordered linearly according to their political preferences. The core of such majority line-graph games is non-empty, and it consists of allocations that reward only the veto players in the game. Two of these allocations, which reward only the two extreme veto players, are those recommended by the upper and lower equivalent solutions. Instead, here, we focused on two core-stable point-valued solutions, the hierarchical index $\bar{h}$ and the τ-index, that reward positively *all* veto players, both extreme and intermediate.[27] More precisely, according to the τ-index, all veto players are equally powerful, while, according to the hierarchical index $\bar{h}$, the extreme (left and right-wing) veto players are no less (and are often more) powerful than the intermediate veto players, depending on the number of players on each side of these extreme veto players. As we saw in Sect. 4, for majority line-graph games, the two indices can be characterized by core stability with an additional weak (veto) fairness axiom. For the hierarchical index $\bar{h}$, this is the component veto fairness axiom—a weaker version of Herings et al.'s (2008) component fairness property—while, for the τ-index, this is the veto fairness axiom—a weaker version of Myerson's (1977) fairness property.

[27] Another solution, specifically defined for line-graph games, is the spectrum value introduced in Álvarez-Mozos et al. (2013). However, as the Shapley value and the Banzhaf value, this solution is also not core stable for majority voting line-graph games.

Interestingly, while—in contrast to the component fairness property—the standard fairness axiom is not compatible with core stability, its weaker veto version is; indeed, these two axioms characterize the τ-index on majority line-graph games.

Some years ago in a private meeting, Stefan Napel remarked that solutions that assign nonzero power to null players are particularly interesting. We note here that majority line-graph games make the hierarchical index $\bar{h}$ and the τ-index interesting in this sense. Notice that null players in the majority game v need *not* be null players in the restricted game v^L. In fact, players who are null players in v might be *veto* players in v^L.[28] Such players may thus be assigned positive power by a core-stable power index and are always assigned positive power by the core-stable hierarchical index $\bar{h}$ and τ-index. This elevated status of null players in v is a consequence of their position in the (linear) ordering on the players and the cooperation possibilities between them captured by the line graph (N, L). Indeed, such a linear ordering, reflecting alliance possibilities among the member states, might explain the Luxembourg 'gaffe' in the 1958–1972 iteration of the EU Council of Ministers.[29]

Majority line-graph games are also interesting because, in such games, a number of popular core-stable solutions agree in their recommendations. More precisely, as Potters and Reijnierse (1995) have showed, in superadditive *tree* games, (1) the *bargaining set* (Aumann & Maschler, 1964) coincides with the core, and (2) the *kernel* (Davis & Maschler, 1965), which is a subset of the bargaining set, is a singleton that contains only the *nucleolus* (Schmeidler, 1969). Since the restricted game v^L is superadditive and line graphs are trees, these results also hold for majority line-graph games. In fact, it is easy to see that, in these games, the nucleolus coincides with the τ-index.[30]

We close by referring back to an observation about the hierarchical index $\bar{h}$ noted at the end of the preceding section. Recall that, according to the hierarchical index, the power of the left and right-wing extreme veto players depends on the number of non-veto players occupying positions on either side. Put succinctly, the higher the number of non-veto players in the respective 'flank' of an extreme veto player—that is, the non-veto players to the left (right) of the left (right) extreme veto player—the more powerful that veto player is. Thus, while non-veto players have no power according to the hierarchical index $\bar{h}$, they do affect the power of their closest extreme veto player. We might then say that, while powerless according to $\bar{h}$, the non-veto players in the winning component of a line-graph game have *leverage* over the $\bar{h}$-

[28] Note, however, that null players in v can only be *intermediate*, and never extreme, veto players in v^L.

[29] Famously, Luxembourg was a null player in the Council at that time, which consisted of Germany, Italy, France, The Netherlands, Belgium, and Luxembourg with weights of 4, 4, 4, 2, 2, and 1, respectively, and a quota of 12 (Felsenthal & Machover, 1997). It is easy to verify that in all orders where Luxembourg 'separates' the big four-weight countries and the small two-weight countries (e.g. take the order Belgium, Italy, Luxembourg, France, The Netherlands, and Germany), Luxembourg is an intermediate veto player in the respective majority line-graph game.

[30] This also follows from the fact that the kernel of cooperative games satisfies symmetry (it rewards symmetric players equally; see Maschler (1992, p. 621) and that veto players in line-graph games are symmetric.

power of the two extreme veto players. The same does not hold for the τ-index, which rewards all veto players equally and according to which the power of a veto player is not sensitive to the number of non-veto players in the winning component. These observations suggest an interesting line for future research: the formulation of leverage measures that can capture the effect players have over the power of other players, given an underlying power index.[31] In the case of core-stable power indices, power*less* (i.e. non-veto) players may nevertheless hold leverage over power*ful* (i.e. veto) players. In such cases then, we would need to keep the ideas of 'power' and 'leverage' conceptually and formally distinct. Developing these ideas in more detail, however, is left for another occasion.

References

Algaba, E., Bilbao, J. M., & López, J. J. (2001). A unified approach to restricted games. *Theory and Decision, 50*, 333–345.

Algaba, E., Fragnelli, V., & Sánchez-Soriano, J. (Eds.). (2019). *Handbook of the Shapley value*. Taylor and Francis Group: CRC Press.

Álvarez-Mozos, M., Hellman, Z., & Winter, E. (2013). Spectrum value for coalitional games. *Games and Economic Behavior, 82*, 132–142.

Ambec, S., & Ehlers, L. (2008). Sharing a river among satiable agents. *Games and Economic Behavior, 64*(1), 35–50.

Ambec, S., & Sprumont, Y. (2002). Sharing a river. *Journal of Economic Theory, 107*(2), 453–462.

Aumann, R. J., & Maschler, M. B. (1964). The bargaining set for cooperative games. In M. Dresher, L. S. Shapley, & A. W. Tucker (Eds.), Advances in game theory (Vol. 52, pp. 443–476). Princeton University Press.

Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review, 19*(2), 317–343.

Béal, S., Rémila, E., & Solal, P. (2010). Rooted-tree solutions for tree games. *European Journal of Operational Research, 203*(2), 404–408.

Bilbao, J.-M. (1998). Values and potential of games with cooperation structure. *International Journal of Game Theory, 27*(1), 131–145.

Bondareva, O. (1963). Some applications of linear programming methods to the theory of cooperative games. *Problemy Kibernetiki, 10*, 119–139.

Casajus, A. (2021). Second-order productivity, second-order payoffs, and the Shapley value. *Discrete Applied Mathematics, 304*, 212–219.

Curiel, I., Pederzoli, G., & Tijs, S. (1989). Sequencing games. *European Journal of Operational Research, 40*(3), 344–351.

Curiel, I., Potters, J., Prasad, R., Tijs, S., & Veltman, B. (1993). Cooperation in one machine scheduling. *Zeitschrift für Operations Research, 38*(2), 113–129.

Curiel, I., Potters, J., Prasad, R., Tijs, S., & Veltman, B. (1994). Sequencing and cooperation. *Operations Research, 42*(3), 390–575.

Davis, M., & Maschler, M. B. (1965). The kernel of a cooperative game. *Naval Research Logistics, 12*(3), 223–259.

Demange, G. (1994). Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics, 23*(1), 45–58.

[31] See Casajus (2021) for such a recent contribution with respect to the Shapley value for cooperative games, where what we call a player's 'leverage' is called a player's 'second-order productivity'.

Demange, G. (2004). On group stability in hierarchies and networks. *Journal of Political Economy, 112*(4), 754–778.
Derks, J., Haller, H., & Peters, H. (2000). The selectope for cooperative games. *International Journal of Game Theory, 29*(1), 23–38.
Dubey, P. K., & Shapley, L. S. (1979). Mathematical properties of the Banzhaf power index. *Mathematics of Operations Research, 4*(2), 99–131.
Felsenthal, D. S., & Machover, M. (1997). The weighted voting rule in the EU's council of ministers, 1958–1995: Intentions and outcomes. *Electoral Studies, 16*(1), 33–47.
Fernández, C., Borm, P., Hendrickx, R., & Tijs, S. (2005). Drop out monotonic rules for sequencing situations. *Mathematical Methods of Operations Research, 61*(3), 501–504.
Gillies, D. B. (1953). Some theorems on n-person games. Ph.D. dissertation. Princeton University.
Granot, D., & Huberman, G. (1982). The relationship between convex games and minimum cost spanning tree games: A case for permutationally convex games. *SIAM Journal on Matrix Analysis and Applications, 3*(3), 288–292.
Greenberg, J., & Weber, S. (1986). Strong Tiebout equilibrium under restricted preferences domain. *Journal of Economic Theory, 38*(1), 101–117.
Harsanyi, J. (1959). A bargaining model for cooperative n-person games. In A. W. Tucke, & R. D. Luce (Eds.), Contributions to the theory of games (Vol. IV, pp. 325–355). Princeton University Press.
Herings, P.J.-J., van der Laan, G., & Talman, D. (2008). The average tree solution for cycle-free graph games. *Games and Economic Behavior, 62*(1), 77–92.
Le Breton, M., Owen, G., & Weber, S. (1992). Strongly balanced cooperative games. *International Journal of Game Theory, 20*(4), 419–427.
Maschler, M. B. (1992). The bargaining set, kernel and nucleolus. In R. J. Aumann & S. Hart (Eds.), Handbook of game theory with economic application (Vol. I, pp. 591–667). Elsevier.
Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of Operations Research, 2*(3), 225–229.
Núñez, M., & Rafels, C. (2002). The assignment game: The τ-value. *International Journal of Game Theory, 31*(3), 411–422.
Owen, G. (1975). Multilinear extensions and the Banzhaf value. *Naval Research Logistics Quarterly, 22*(4), 741–750.
Owen, G. (1986). Values of graph-restricted games. *SIAM Journal on Algebraic Discrete Methods, 7*(2), 210–220.
Penrose, L. S. (1946). The elementary statistics of majority voting. *Journal of the Royal Statistical Society, 109*(1), 53–57.
Pérez-Castrillo, D., & Wettstein, D. (2001). Bidding for the surplus: A non-cooperative approach to the Shapley value. *Journal of Economic Theory, 100*(2), 274–294.
Potters, J., & Reijnierse, H. (1995). γ-component additive games. *International Journal of Game Theory, 24*(1), 49–56.
Ruiz, L. M., Valenciano, F., & Zarzuelo, J. M. (1998). The family of least square values for transferable utility games. *Games and Economic Behavior, 24*(1–2), 109–130.
Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics, 17*(6), 1163–1170.
Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), Contributions to the theory of games (Vol. II, pp. 307–317). Princeton University Press.
Shapley, L. S. (1967). On balanced sets and cores. *Naval Research Logistics Quarterly, 14*(4), 453–460.
Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *The American Political Science Review, 48*(3), 787–792.
Slikker, M. (2007). Bidding for surplus in network allocation problems. *Journal of Economic Theory, 137*(1), 493–511.
Thompson, G. L. (1981). Auctions and market games. In R. J. Aumann (Ed.), *Essays in game theory and mathematical economics in Honor of Oskar Morgenstern.* Bibliographisches Institut.

Tijs, S. (1981). Bounds for the core and the τ-value. In O. Moeschlin, & D. Pallaschke (Eds.), *Game Theory and Mathematical Economics: Proceedings of the Seminar on Game Theory and Mathematical Economics*, Bonn/Hagen, 7–10 October 1980 (pp. 123–132). North-Holland Publishing Company.
van den Brink, R. (2001). An axiomatization of the Shapley value using a fairness property. *International Journal of Game Theory, 30*(3), 309–319.
van den Brink, R., & van der Laan, G. (1998). Axiomatizations of the normalized Banzhaf value and the Shapley value. *Social Choice and Welfare, 15*(4), 567–582.
van den Brink, R., van der Laan, G., & Vasil'ev, V. A. (2007). Component efficient solutions in line-graph games with applications. *Economic Theory, 33*(2), 349–364.
van den Brink, R., van der Laan, G., & Moes, N. (2013). A strategic implementation of the average tree solution for cycle-free graph games. *Journal of Economic Theory, 148*(6), 2737–2748.
van den Brink, R., Núñez, M., & Robles, F. (2021). Valuation monotonicity, fairness and stability in assignment problems. *Journal of Economic Theory, 195*, 105277.
van den Nouweland, A., & Borm, P. (1991). On the convexity of communication games. *International Journal of Game Theory, 19*(4), 421–430.
Vasil'ev, V. A. (1978). Support function of the core of a convex game. *Optimizacija, 21*, 30–35.
Vasil'ev, V. A. (2006). Core and generalized NM-solutions for some classes of cooperative games. In T. S. Driessen, G. van der Laan, V. A. Vasil'ev, & E. B. Yanovskaya (Eds.), *Russian contributions to game theory and equilibrium theory* (pp. 91–149). Springer.

EU

Double Proportionality for the European Parliament: The Tandem System

Jo Leinen and Friedrich Pukelsheim

Abstract The tandem system proposes a double proportional electoral system for the European Parliament. It offers a forum for europarties to contest an election with power, visibility and influence. The tandem system proceeds in three steps. The first step apportions all parliamentary seats among europarties by aggregating the electorate's votes at Union level. Thus, with regard to the division of the Union's citizens by political persuasion, the tandem system obeys the One Person–One Vote principle. The second step, disaggregation of the unionwide apportionment, allots the seats by Member State and Europarty in a way safeguarding the seat contingents of the Member States. Thus, with regard to the Union's layout by Member State, the tandem system respects the principle of degressive representation. The third step assigns the seats of a party in a Member State to domestic candidates by means of the same provisions that Member States have been employing in the past, thus complying with the Union's principle of subsidiarity.

1 Introduction

The elections to the ninth European Parliament (EP) took place during 23–26 May 2019. The EP constitutes a single political body, yet it is customary to use the plural "elections" when talking about electing a EP. As a matter of fact, the event decomposes into a patchwork of twenty-seven separate elections, one per Member State. Lack of uniformity is a hallmark of EP elections. The diffuse appearance of the electoral event has been lamented before and after previous EP elections and is again moaned in assessments of the 2019 elections (Hrbek, 2019; Kaeding et al., 2019; Oelbermann et al., 2019).

The current status has its roots in the past. The Electoral Act was conceived in 1976, amended in 2002 and 2018, and is again on the agenda of the incumbent

J. Leinen
European Parliament, Strasbourg, France

F. Pukelsheim (✉)
Institute for Mathematics, University of Augsburg, Augsburg, Germany
e-mail: pukelsheim@math.uni-augsburg.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_17

parliament.[1] As soon as the 1976 Electoral Act had to pass practical tests its deficiencies came to light. Quite a few proposals for amendment were tabled during past legislative periods, see Anastassopoulos (2002), Duff (2011), pp. 32–51 and Costa and Jouvenat (2016).

The 2002 amendment achieved some progress. It decreed that in each Member State members of the EP shall be elected on the basis of proportional representation. The term "proportionality" addresses a specific group of stakeholders, political parties. Parties are institutions mediating between the many voters and the few representatives. The term "proportional representation" stipulates that the number of seats allotted to a party ought to be proportional to the number of votes cast for this party. Back in 2002, EP elections were conducted as an ensemble of separate elections per Member States. The parties relevant in those days were the domestic parties of the Member States.

The involvement of domestic parties naturally inspired visions to launch corresponding political institutions at Union level. An initial regulation, on "political parties at European level" in 2003 was superseded by a subsequent regulation on "European political parties" in 2014. The topic is again on the agenda of the incumbent EP.[2]

Originally a political party at European level was taken to be an association of like-minded domestic parties from the Member States, as indicated by the alternate designation as a "European party family". Hopes were raised that eventually a Union polity would evolve as soon as European party families would mutate into 'true' europarties. A 'true' europarty would set a proper political agenda at Union level, reconnect with the Union's citizens, and contest EP elections by shaping the electoral campaign (Bardi, 2005; Leinen & Pescher, 2014; Hecke, 2018).

It is rather sensible for the AFCO committee to review the Electoral Act and the Regulation on European political parties in parallel. The true functioning of europarties is a supposition underlying all proposals for enhanced uniformity when electing the EP (Farrell & Scully, 2005; Hix & Hagemann, 2009; Oelbermann & Pukelsheim, 2011).

Here we boldly assume that europarties are properly operating, strive for political power, and aim to play a vital role at European elections. Our focus is on the intricacy of design of the electoral procedure. The tandem system, a double proportional system, takes into account two dimensions each of which reflects the representation of the Union's citizens in the EP. One dimension is the electorate's political division by partisan vote, the other, the electorate's geographical division by Member State, see Duff et al. (2015), Pukelsheim (2017), Sect. 14 and Costa & Jouvenat (2021).

---

[1] Official Journal of the European Union (OJ) L 278 (8.10.1976) 1–11; OJ L 283 (21.10.2002) 1–4; OJ L 178 (16.7.2018) 1–3; Dossier AFCO 2020/2220(INL), rapporteur Domènec Ruiz Devesa (ES-S&D). – A consolidated version of the 2002 Act is in Duff (2011), pp. 9–14—The 2018 Act is still pending; see Cicchi (2021).

[2] OJ L 297 (15.11.2003) 1–4; OJ L 317 (4.11.2014) 1–27; Dossier AFCO 2021/2018(INI), co-rapporteur Charles Goerens (LU-Renew) and Rainer Wieland (DE-PPE).

As for the representation by Member State, Article 14(2) TEU[3] demands that "representation of citizens shall be degressively proportional". That is, representation of citizenries may deviate from strict proportionality in the direction of degressivity. In view of this specification the term "double proportionality" sounds inappropriate. We opt for a specific label, "tandem system".

Our paper is organized as follows. Section 2 reviews the success of double proportional electoral systems in Swiss Cantons. Section 3 describes the prospective use of double proportionality for the EP in form of the tandem system. The system is illustrated using the data of the 2019 elections in Sect. 4. Section 5 concludes the paper with some general considerations.

2 Double Proportionality in Swiss Cantons

Elections for the EP share a typical characteristic with elections for Swiss canton parliaments in that the electoral region is subdivided into several electoral districts and that this subdivision is considered constitutive. The European Union is subdivided into Member States. For a canton, the subdivision is into communities such as townships, counties or villages.

Cantonal communities differ by population figures. Theoretically, a community with a population too small to form a district may merge with its neighbors in order to assemble a district of reasonable size. People gain little, though, when communities are located in valleys disassociated from each other by mountain massifs of thousands of meters in altitude as in Valais or Grisons. More generally, there may exist historical, federalistic, cultural, linguistic, or religious reasons calling for preservation of communities when subdividing a canton.

When a canton is subdivided into electoral districts, the districts' seat contingents are allocated well ahead of polling day so that people know how many representatives they will elect in their district. The allocation is determined in proportion to population figures. A small community may command no more than one or two seats.

Traditionally, the election is evaluated in each district separately. A separate evaluation may cause severe legal problems when a cantonal constitution decrees that the election must follow the principles of proportional representation. Proportionality is hardly possible when there is no more than two seats to fill. Parties finishing third or yet less successful will not gain a seat. The votes of their supporters turn ineffective because the two seats are dealt out between the two major parties. Such situations violate the electoral principle of equality.

[3] OJ C 326 (26.10.2012) 13–45.—For the determination of the Member States' seat contingents see Pukelsheim and Grimmett (2018).—Note also that the Qualified Majority Voting system in the Council, while technically disjoint from the apportionment of seats in the EP, constitutes another representational issue that is highly sensitive on the political level.

This is where a double proportional electoral system comes to the rescue. Double proportionality aggregates the votes of the entire electorate at canton level. Then it apportions all seats of the parliament to parties in proportion to canton wide vote sums. It becomes irrelevant whether votes are cast in a small, medium, or large community. All votes are treated equally, in accord with the *One Person–One Vote* principle.

The new element added by double proportionality is the allotment of seats by community and party. This new step allots the parties' cantonwide seats to districts in such a way that every district ends up with its preordained seat contingent. In this way double proportionality warrants equality of votes across the whole canton, while at the same time it verifies the subdivision of the canton into several districts of different size.

The world premiere of double proportionality took place in 2006 in the canton of Zurich. Since then, more cantons adopted a double proportional system: Schaffhausen 2008, Aargau 2009, Zug 2014, Nidwalden 2014, Schwyz 2016, Valais 2017, Uri 2020, Grisons 2021. In some cantons the amendment of the electoral law had to be approved by a popular referendum. Acceptance was overwhelming, despite of blustering polemics of sullen politicians who interpreted the quest for electoral equality to be an attack on cantonal sovereignty, see Pukelsheim and Grimmett (2011), Senti (2013), and Pukelsheim (2017), Sect. 14.5.

The exigencies of electoral equality are settled by the Bundesgericht (Swiss Federal Court) in Lausanne, based on the Swiss constitution together with the canton constitution. The Court repeatedly pointed out that seat contingents when too small would become unacceptable in cantons whose constitution demands proportionality. An infringement of constitutionally warranted equality would be even less acceptable since double proportionality provides a solution which does justice to the constitutional demands without ifs and buts, see Bundesgericht (2010)—The Court refers to double proportionality with the tag "Doppelter Pukelsheim".

3 Double Proportionality for the EP

In order to apply double proportionality to EP elections there needs to be a sensible way of aggregating all votes at Union level. To this end we introduce three categories of political entities. A first category are the European political parties registered with the Authority for European Political Parties and European Political Foundations.[4] Since conditions for registration are ambitious, it seems appropriate to allow for a second category of party-like entities not (yet) registered with the Authority, euromovements. A group of domestic parties from two or more Member States qualifies as a euromovement, as does a European political movement such as VOLT. We

[4] European political parties should not be confounded with political groups in the EP. Political parties cater to the citizenry of the Union, while political groups are institutional units to organize parliamentary business.

use the label "europarties" as a generic term spanning both categories, (registered) European political parties as well as (non-registered) euromovements.

Moreover, domestic parties may choose not to associate with any europarty but to remain solitary. This gives rise to a third category, "stand-alone parties", comprising domestic parties who contest the EP election just in their home state.

The European political parties assumed relevant at the 2019 elections are the ones listed on the webpage of the Authority for European Political Parties and European Political Foundations:

ALDE	Alliance of Liberals and Democrats for Europe Party
EPP	European People's Party
PES	Party of European Socialists
EDP	European Democratic Party
EFA	European Free Alliance
EGP	European Green Party
PEL	Party of the European Left
ECR	European Conservatives and Reformists Party
ECPM	European Christian Political Movement
ID	Identité et Démocratie Parti

Domestic parties who cooperate with a European political party usually may choose between joining as a full member, an associate member, or an observer. For our 2019 illustration we restricted attention to full members. Since we failed to retrieve reliable membership rosters of any of the europarties listed, we compiled them ourselves from their webpages and the information in Wikipedia. Most likely, our compilations contain errors or outdated information.

As an example of a non-registered europarty we include into our illustration the European movement VOLT. At the 2019 elections, its German section was the sole section to win a seat. Other VOLT sections failed the domestic electoral threshold, or garnered too few votes to validate a seat, or contested the election with an independent candidate who was not successful.

Votes included into the 2019 example are those cast for domestic parties who pass the pertinent domestic threshold and who obtain at least one seat. The tandem system re-evaluation of the 2019 elections disregards all votes that were cast for dwarf parties, whether they are members of European political parties or not. These limitations are imposed solely for enabling us to use the 2019 data as an example; in actual applications the limitations should be relieved. Vote counts are taken from the study (Oelbermann et al., 2019), disregarding all vote counts which in the study are labeled "Others".

4 The Tandem System

Our illustration of the tandem system uses the data of the 2019 elections, disregarding the results from the United Kingdom. Even though the seat assignments resulting from the tandem system turn out to be close to those actually implemented they cannot be taken to indicate a systematic trend of any political significance. Due to the instructive character of the example some hypothetical adjustments are unavoidable.

The tandem system proceeds in three steps.

4.1 Apportionment of Seats at Union Level

The aggregation of votes at Union level provides the base to apportion the 705 EP seats among europarties and stand-alone parties. The apportionment calculations use the divisor method with standard rounding (Sainte-Laguë method). This first step realizes the *One Person–One Vote* principle and secures electoral equality for all voters in the Union.

Table 1 displays a total of 163,374,809 votes that enter into the process of apportioning the 705 EP seats.[5] Every 231,400 votes justify roughly one seat, i.e., dividing the Union divisor 231,400 into "Votes" yield "Quotients" that are rounded in the standard fashion to obtain the desired "Seats". The electoral key 231,400 is determined so that the sum of all "Seats" is equal to the number of seats available, 705.

The upper block of Table 1 exhibits the aggregated results for the eleven europarties. They are apportioned a total of 624 seats. These seats need to be disaggregated by Member State and europarty, disaggregation is carried out in the next step.

The lower block of Table 1 features thirty-four stand-alone parties, i.e., domestic parties who are not a member of any europarty. They are labeled by the two-letter code[6] of the Member State where they are active, together with their party acronym. Altogether the stand-alone parties are apportioned a total of 81 seats. This apportionment is definitive, there is no need to subject these seats to any further disaggregation mechanism.

4.2 Allotment of Seats by Member State and Europarty

The synchronizing potential of the tandem system comes to light in the allotment of seats by Member State and europarty. Since the 81 seats of the stand-alone parties are final, they are subtracted from the states' seat contingents. The reduced contingents provide a total of 624 seats to be allotted to europarties.

The task then is to merge two dimensions that are interacting: the layout by Member State, and the division by europarties:

[5] "Votes" are divided by the Union divisor 231,400 to obtain "Quotients", then "Quotients" are rounded to yield "Seats". The divisor is determined so that the sum of all "Seats" is equal to the number of seats available, 705.

[6] Interinstitutional Style Guide (February 2022), Sect. 7.1.1.

Table 1 Apportionment of 705 seats at Union level

EP2019-Aggregation	Votes	Quotient	Seats
Eleven europarties, totalling 624 seats			
EPP	39,338,118	170.0	170
PES	32,347,309	139.8	140
ALDE	18,656,812	80.6	81
ID	16,182,413	69.9	70
EGP	14,835,208	64.1	64
ECR	11,329,360	49.0	49
PEL	6,261,560	27.1	27
EFA	2,195,733	9.49	9
EDP	2,023,884	8.7	9
ECPM	741,034	3.2	3
VOLT	416,171	1.8	2
Thirty-four stand-alone parties, totalling 81 seats			
IT-M5S	4,569,089	19.7	20
DE-AfD	4,104,453	17.7	18
FR-LFI	1,428,548	6.2	6
ES-JUNTS	1,018,435	4.4	4
DE-DIE PARTEI	899,079	3.9	4
PL-WIOSNA	826,975	3.6	4
HU-DK	557,081	2.4	2
DE-TIERSCHUTZ	542,226	2.3	2
DE-ÖDP	369,869	1.6	2
BE-2PTB	355,883	1.54	2
CZ-PIRATI	330,844	1.4	1
EL-KKE	302,603	1.3	1
DK-DF	296,978	1.3	1
SE-V	282,300	1.2	1
EL-XA	275,734	1.2	1
FI-PS	253,176	1.1	1
DE-PIRATEN	243,302	1.1	1
EL-EL	236,347	1.0	1
NL-PvdD	220,938	1.0	1
HU-JOBBIK	220,184	1.0	1
NL-50+	215,199	0.9	1
IE-SF	196,001	0.8	1
NL-PVV	194,178	0.8	1
CZ-KSCM	164,624	0.7	1
LT-LVZS	158,190	0.7	1
IE-I4C	124,085	0.54	1
SK-KLSNS	118,995	0.51	1
LT-DP	113,243	0.49	0
IE-2indep	85,034	0.4	0
HR-MK	84,765	0.4	0
LT-AMT	82,005	0.4	0
CY-AKEL	77,241	0.3	0
HR-ZZ	60,847	0.3	0
CY-DIKO	38,756	0.2	0
Sum (Union divisor)	163,374,809	(231,400)	705

Table 2 Allotment of seats by Member State and europarty—part 1

EP2019	624	EPP	170	PES	140	ALDE	81	ID	70	EGP	64
AT	19	1,305,956	6	903,151	5	319,024	1	650,114	4	532,193	3
BE	19	849,976	2	1,085,159	3	1,148,705	3	811,169	3	1,011,563	4
BG	17	725,678	8	474,160	5	323,510	3				
CY	6	81,539	4	29,715	2						
CZ	19	447,943	5	502,343	6	216,718	4				
DE	69	10,794,042	21	5,916,882	13	2,028,594	4	7,677,071	21		
DK	13	170,544	1	592,645	3	926,132	5	364,895	3		
EE	7	34,188	1	77,375	2	134,959	3	42,265	1		
EL	18	1,873,137	8	436,726	2						
ES	55	4,510,193	11	7,359,617	20	2,726,642	7				
FI	13	380,460	3	267,603	3	363,439	3	292,892	3		
FR	73	1,920,407	7	1,403,170	6	5,079,015	17	5,286,939	28	3,055,023	15
HR	12	244,076	5	200,976	5	55,829	1				
HU	18	1,824,220	14	229,551	2	344,512	2				
IE	11	496,459	5	52,753	1	277,705	3	190,755	2		
IT	56	2,493,858	6	6,107,545	16	9,175,208	30				
LT	10	248,736	4	200,105	4	83,083	1				
LU	6	264,665	2	152,900	1	268,910	1	237,215	2		
LV	8	124,193	2	82,604	2	58,763	1				
MT	6	58,699	2	124,441	4						
NL	26	669,555	4	1,045,274	6	1,194,792	6	599,283	4		
PL	48	4,009,958	17	1,239,977	6						
PT	21	930,191	6	1,104,694	8	396,060	4				
RO	33	3,447,949	13	2,040,765	9	2,028,236	7				
SE	20	1,056,626	5	974,589	6	619,060	3	478,258	3		
SI	8	180,155	4	89,936	2	74,431	2				
SK	13	194,715	4	154,996	4	99,128	2				
Party div.		1.098		1		1.165		0.77		0.818	

- Within a Member State, the sum of the seats must meet the state's reduced seat contingent.
- Within a europarty, the sum of the seats must exhaust their due seats at the Union level.

Tables 2 and 3 resolve the task by using the double proportional variant of the divisor method with standard rounding.[7] Double proportionality employs two sets of electoral keys, state divisors and party divisors. Once these are published, the vote count which has been recorded in state S for party P is divided by the state divisor for state S and by the party divisor for party P. The resulting quotient is rounded to the nearest whole number to yield the seat number sought, i.e., the number of seats allotted to europarty P in state S.

[7] The votes are divided by two divisors, the associated "State divisor" and the associated "Party divisor", and then rounded to "Seats". Row-sums match the states' seat contingents, and column-sums meet the parties' apportionments at Union level.

Table 3 Allotment of seats by Member State and europarty—part 2

(cont.)	ECR	49	PEL	27	EFA	9	EDP	9	ECPM	3	VOLT	2	State div.
AT													200,000
BE					954,048	4					20,385	0	330,000
BG	143,830	1									3,500	0	88,000
CY													18,000
CZ	344,885	4											76,000
DE			2,056,049	6			806,703	3	273,828	0	249,098	1	457,500
DK			151,903	1									170,000
EE													40,000
EL			1,343,595	8									210,000
ES	1,388,681	3	2,258,857	8	1,212,139	4	633,265	2			32,432	0	360,000
FI			126,063	1									106,000
FR													249,400
HR									91,546	1			44,000
HU													120,000
IE													94,000
IT	1,726,189	4											392,000
LT	69,347	1											54,000
LU											4,606	0	160,000
LV	77,591	2			29,546	1							46,000
MT													30,000
NL	602,507	3							375,660	2	106,004	1	170,000
PL	6,192,780	25											221,000
PT			325,093	3									134,000
RO							583,916	4					235,000
SE	636,877	3									146	0	176,000
SI													40,000
SK	146,673	3											40,000
Party div.	1.1031		0.8		0.8		0.705		1.44		1		

Small scale illustrations can be found in Balinski (2004), Chap. 7 or in Pukelsheim (2017), Chap. 14. Calculation of state divisors and party divisors is cumbersome and needs a computer program, see Pukelsheim (2017), Chap. 15. On the positive side, once the divisors are obtained and published, everybody can verify the seat numbers via simple divisions and a rounding operation.

As an example, the Austrian contingent of nineteen seats is allotted as follows. EPP garners 1,305,956 votes. The Austrian divisor is 200,000, the EPP divisor is 1.098. This leads to the quotient $1,305,956/(200,000 \times 1.098) = 5.9$, justifying six seats for the Austrian EPP-member ÖVP. The other successful europarties are allotted five, one, four, and three seats, which are handed over to their respective domestic parties.

In this way the allotment by Member State and europarty guarantees that every Member State receives its due number of seats and so does every europarty.

4.3 Assignment of Seats to Candidates

The tandem system concludes with the assignment of seats to candidates. Simply, domestic provisions of a Member State are applied as in the past. Thus the tandem system perpetuates the kind of accountability that Union citizens and representatives are accustomed to. Since domestic provisions differ and since the tandem system respects these differences, every Member State must be reviewed on its own. The twenty-seven reviews decompose into three classes, see Leinen and Pukelsheim (2021).

The first class embraces thirteen Member States where every europarty is in a one-to-one correspondence with a unique domestic member party. The seats allotted to europarties are handed over to the corresponding domestic parties without further ado.

The second class consists of eleven Member States where one of the europarties is in a one-to-many correspondence with its domestic member parties. For every europarty with several member parties, its seats are parceled out proportionally to the votes its members tallied.

The third class assembles three Member States which are special because of establishing multiple constituencies (Belgium and Ireland), or because of using single transferable vote schemes (Ireland and Malta). Slight adjustments accommodate these special cases.

5 Conclusion

There remains the crucial task of raising citizens' awareness that what is at stake is their representation at Union level. Expedient operational procedures, such as the tandem system, are necessary but not sufficient to reach this aim. The mediators for conveying this message are political parties, domestic parties as well as europarties. They ought to be offered incentives to act in concert and to spread the logic of cooperative synergies, see Leinen and Pukelsheim (2022).

The tandem system aligns citizens and Member States in a synchronized (i.e., *tandem*) way. Conceptually, it amends the current Electoral Act in various directions:

- The tandem system achieves electoral equality among all citizens of the Union by aggregating votes at Union level rather than performing separate evaluations per Member State.
- The unionwide alignments are arranged in a manner safeguarding the composition of the EP, i.e., the allocation of the seats of the EP between the Member States.
- Member States retain many domestic provisions, such as ballot structure, vote pattern, and rules to assign the seats of a domestic party to this party's candidates.
- The tandem system promotes a unionwide view of EP elections by involving europarties through political power, public visibility, and coordinating influence.

- The tandem system offers a forum for europarties to promote their spitzenkandidaten and their lead personnel for staffing political offices in the new legislative period.

The tandem system summarizes an EP election across the entire European Union in exhibits such as Tables 1, 2, and 3. The complexity of the tables mirrors the complexity of the Union. The synoptic view of the tandem system furnishes a more informative and less disorienting electoral portrait of the Union than the patchwork of segmented elections as in the past. Of course other options to achieve more uniformity in the European Electoral Act should also be considered, such as Müller (2022).

Finally we note that the tandem system resolves a long-standing friction of primary law. It ends the controversy whether degressive representation of the Member States is at odds with electoral equality of the Union's citizens. The tandem system aligns the two goals without any conflict. It safeguards degressivity, yet it also implements the *One Person–One Vote* principle for all voters in the Union irrespective of their Member State provenance.

References

Anastassopoulos, G. (2002). The debate on the system of electing the members of the European parliament. From a uniform procedure to the common principles of the Treaties. A contribution to the problem of enhancing the democratic and representative nature of the European Parliament. Foreword by Professor Dimitris Tsatsos. MEP, Athens.

Balinski, M. (2004) Le suffrage universel inachevé. Belin.

Bardi, L. (2005). Parties and party system in the European Union. In K. R. Luther & F. Müller-Rommel (Eds.), *Political parties in the New Europe: Political and analytical challenges* (pp. 293–322). Oxford University Press.

Bundesgericht. (2010). Decision of 7 July 2010 (1C_541/2009) re canton Nidwalden; Decision of 20 December 2010 (1C_127/2010) re canton Zug; Decision of 19 March 2012 (1C_407/2011) re canton Schwyz; Decision of 12 February 2014 (1C_495/2012) re canton Valais; Decision of 12 October 2016 (1C_511/2015) re canton Uri, 2010.

Cicchi, L. (2021). Europeanising the elections of the European Parliament—Outlook on the implementation of Council decision 2018/994 and harmonisation of national rules on European elections. *Study, European Parliament, Directorate-General for Internal Policies,* PE 694.199.

Costa, O. (2016). The history of European electoral reform and the electoral act 1976. Issues of democratisation and political legitimacy. *Study, European Parliament, European Parliament History Series*, PE 563.516.

Costa, O., & Jouvenat, P. (2021). *Towards a European political space. The challenges of European electoral law. College of Europe Policy Brief CoFoE 2.21.*

Duff, A. (2011). *Report on a proposal for a modification of the act concerning the election of the members of the European Parliament by direct universal suffrage of 20 September 1976*. European Parliament, Committee on Constitutional Affairs, PE 440.210v04-00

Duff, A., Pukelsheim, F., & Oelbermann, K.-F. (2015). *The electoral reform of the European Parliament. Composition, procedure and legitimacy*. In-depth Analysis for the AFCO-Committee, European Parliament, Directorate-General for Internal Policies, PE 510.002

Farrell, D. M., & Scully, R. (2005). Electing the European Parliament. How uniform are 'uniform' electoral systems? *Journal of Common Market Studies, 34*, 969–984.
Hix, S., & Hagemann, S. (2009). Could changing the electoral rules fix European Parliament elections? *Politique Européenne, 28*, 37–52.
Hrbek, R. (2019). Europawahl 2019. Neue politische Konstellationen für die Wahlperiode 2019–2024. *Integration, 42*, 167–186.
Kaeding, M., Müller, M., & Schmälter, J. (Eds.). (2020). *Die Europawahl 2019*. Springer Nature.
Leinen, J., & Pescher, F. (2014). Von Parteibündnissen zu 'echten Parteien' auf europäischer Ebene? Hintergrund, Gegenstand und Folgen der neuen Regeln für Europäische Parteien. *Integration, 37*, 228–246.
Leinen, J., & Pukelsheim, F. (2021). The tandem system. A new electoral frame for the European Parliament. *MIP—Zeitschrift für Parteienwissenschaften, 27*, 115–124.
Leinen, J., & Pukelsheim, F. (2022). Europäisierung der Wahlen zum Europäischen Parlament. *Das Tandemsystem. Integration, 45*, 60–73.
Müller, M. (2022). Wahlgleichheit und degressive Proportionalität versöhnen: Ein europäischer Verhältnisausgleich durch transnationale Listen. *Integration, 45*, 139–146.
Oelbermann, K.-F., & Pukelsheim, F. (2011). Future European Parliament elections. Ten steps towards uniform procedures. *Zeitschrift für Staats- und Europawissenschaften, 9,* 9–28.
Oelbermann, K.-F., Pukelsheim, F., & Lehmann, W. (2020). *The European elections of May 2019. Electoral systems and outcomes*. Study, European Parliament, European Parliamentary Research Service, PE 652.037.
Pukelsheim, F. (2017). *Proportional representation. Apportionment methods and their applications* (2nd ed.). Springer Cham.
Pukelsheim, F., & Grimmett, G. R. (2018). Degressive representation of member states in the European Parliament 2019–2024. *Representation, 54*, 147–158.
Pukelsheim, F., & Schuhmacher, C. (2011). Doppelproporz bei Parlamentswahlen - ein Rück- und Ausblick. *Aktuelle Juristische Praxis - Pratique Juridique Actuelle, 20*, 1581–1599.
Senti, M. (2013, March 16). Aufstand am Fusse der Mythen. *Neue Zürcher Zeitung*.
van Hecke, S. (Lead author) (2018). *Reconnecting European political parties with European Union citizens*. International IDEA Discussion Paper 6/2018, International Institute for Democracy and Electoral Assistance.

Explaining Contestation: Votes in the Council of the European Union

Arash Pourebrahimi, Madeleine O. Hosli, and Peter van Roozendaal

Abstract In this chapter, we study voting behavior in the Council of the European Union (EU) for the time span of 2010 to 2021. We use Council voting data, examining the impact of different independent variables on member states' voting behavior: net contributions to the EU budget, voting power, left–right policy positions, and finally, the distance of a member state's ideological position from the position of the winning coalition under the qualified majority voting (QMV) rule. We investigate more than 1229 legislative decisions taken in the Council, based on over 30,000 votes. Controlling for public attitudes toward the EU and whether a member state held the Council presidency, we use a random effects binomial logit model in which we divide votes into two categories: support and objection. Moreover, we also apply an ordered logit model in which voting decisions are ordered based on the level of support for a vote. Our results show that net contributors to the EU budget are more likely to contest a vote in the Council of the EU. Similarly, the further the ideological position of a member state from the one of a winning coalition, the higher the chance it contests the vote. We find no evidence, however, for a clear relation between voting power and the probability of contestation.

1 Introduction

Decision-making in the Council of the European Union (EU) is often a long and enigmatic process, involving many actors both within EU institutions and at the level of individual Member States. Formal voting in the Council of Ministers may take place at the end of this extensive process. Many policy proposals are never formally

A. Pourebrahimi · M. O. Hosli (✉)
Leiden University, Leiden, The Netherlands
e-mail: m.o.hosli@fgga.leidenuniv.nl

A. Pourebrahimi
e-mail: a.pourebrahimi.andouhjerdi@fgga.leidenuniv.nl

P. van Roozendaal
Netherlands Court of Audit, Leiden, The Netherlands
e-mail: p.vanroozendaal@rekenkamer.nl

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_18

voted upon, as participants in Council meetings—most notably the Presidency—are keenly aware of whether or not a required majority of EU states support a proposal in respective discussions. Depending on the type of policy proposal at hand, a simple majority, a qualified majority, a reinforced qualified majority, or unanimity is required for formal Council decision-making. Since the Treaty of Lisbon, most issues are dealt with based on the "ordinary legislative procedure" and require qualified majority voting (QMV) in the Council of the EU. This procedure starts with a proposal by the European Commission, continues with a reading by both the European Parliament (EP) and the Council of the EU, and has a potential second reading; in case there is no agreement on the proposal yet and if after this second round, the EP and Council still disagree, a conciliation committee is installed. From the total of 991 legislative decisions with unique inter-institutional codes taken in the Council between 2010 and 2021, however, only six decisions went through the third reading.

In this chapter, we investigate a total of 1229 QMV decisions taken in the Council between 2010 and 2021 to shed light on the factors that affect the likelihood of a contesting vote cast in the Council. Our dataset[1] contains a total of 33,910 votes, for which we examine why an EU member state might have decided not to join the winning coalition (i.e., the set of member states that passed the decision) and instead contested a decision by casting a no vote or abstained. Applying a random effects binomial logistic regression analysis, we examine the impact of variables such as a Member State's net contribution to the EU, its relative power, government composition, and its ideological distance from the winning coalition on voting decisions. We control whether an EU state held the (rotating) presidency of the Council of the EU and domestic public opinion toward the EU. In addition to a binomial logistic regression, we apply ordered logit models to distinguish between abstentions and negative votes cast.

Our results show that net contribution to the EU budget as a share of GDP is positively correlated with the probability of contestation in the Council. Similarly, when a Member State's left–right policy position is further away from the winning coalition's policy position, the likelihood of contestation is higher. Our results also show that holding the presidency of the Council lowers the contestation probability. We include power (Banzhaf index) in our dataset and track changes in Member States' power levels over time. Our random effects model does not show any correlation between power and voting behavior, while a binomial logit model without being defined as a panel would show that relationship.

Our research contributes to the literature on investigating voting behavior in the Council with roll call data. Further research on the same topic can zoom in further to consider the ideological positions of the ministers in Council configurations instead of the government as well. Our chapter is structured as follows. The next section discusses potential determinants of member states' voting behavior in the Council, based on earlier literature on this subject and own additional reflections. Section 3

[1] This dataset has been compiled by Arash Pourebrahimi. It is the basis for his forthcoming dissertation on decision-making in the European Union.

presents our dependent and independent variables and discusses the sources used to operationalize them. Section 4 of our paper presents the results of our analysis, while Sect. 5 summarizes and concludes our paper.

2 Potential Determinants of Voting Behavior in the Council

In terms of distributional divisions within the EU, some EU member states are net contributors to the EU budget while others are net beneficiaries. Similarly, the share of net contributions or benefits compared to the size of a member state's economy varies among member states. Zimmer et al. (2005), for example, argue that the division between net contributors and net beneficiaries in the EU is the most important source of conflict in the Council. Empirical studies on the impact of net contributions to the EU budget on voting behavior, however, have shown mixed results. Hosli et al. (2011) find that this effect differs between new and old member states, considering the extensive 2004 enlargement: While among old member states, net contributors were more likely to oppose, net beneficiaries among the new member states were more likely to oppose decisions supported by the majority (Hosli et al., 2011). In their study, the coefficient of budget balance for all member states revealed that higher net contributors were more likely to contest decisions. Mattila (2004) has shown that net contributors had a higher probability to contest decisions. However, this relationship was significant only when "no"-votes and abstentions together were considered to constitute oppositional votes. As discussed by Mattila (2004), net contributors might feel more entitled to oppose the majority in the Council given that they were contributing more to the Union. More recently, Franchino et al. (2022) demonstrated that net receipts from the EU budget are correlated with governments' responsiveness to their publics' left–right positions. However, the authors do not examine the impact of net receipts on voting behavior. Bailer et al. (2015) demonstrated that net contributors to the EU budget were more likely to oppose decisions in the Council than were net beneficiaries. The line of argument was that member states contributing more demonstrated their discontent with EU decisions in the sense of sending a signal to their domestic voters. While it is notoriously difficult to assess who the "net contributors" or "net beneficiaries" to the EU budget are (as there are different streams of income and expenditures within the EU, spread over different policy areas), we will use this basic cleavage as one of the possible explanations for vote choices in the Council. Accordingly, our first hypothesis is as follows:

H1: *The higher the net contribution of an EU member state, the more likely it is to contest a decision in the Council of the EU.*

Another potential explanatory variable could be the relative power a member state has in the Council. Due to differences in terms of their (population) size, the distribution of power among member states in the Council is not evenly distributed among them. It can hence be expected that different levels of power are likely to affect a member state's voting behavior in the Council. In recent literature on the subject, the focus has mainly been on the impact of voting power in the framework of bargaining success in the Council. Bailer (2004), for example, found that voting power as an exogenous source of power improved a member state's bargaining success only in selected policy areas. Exploring the impact of voting power on voting behavior in the Council, Mattila (2004) demonstrated that EU member states possessing a higher number of votes opposed decisions in the Council more often than smaller member states did. Similarly, Hosli et al. (2011) when analyzing all member states (after the 2004 enlargement) found that higher voting power was correlated with a higher number of opposing votes (i.e., no or abstain). However, this result was not found for the category of new member states separately. More recently, Perarnaud and Arregui (2022) examined the impact of voting power on the speed with which national positions are shaped and on bargaining success in the Council. With a focus on economic and monetary integration in the EU, Lundgren et al. (2019) showed that a member state's power resources, including voting power, did not contribute to bargaining success in negotiations on the reform of the Eurozone conducted between 2010 and 2015. These negotiations, however, were largely conducted in the context of intergovernmental negotiations at the level of the European Council. Following the literature above, we present our second hypothesis:

H2: *More powerful member states in the EU are more likely to contest decisions in the Council.*

Another potential independent variable affecting vote choice in the Council is the ideological position of a member state on the left–right policy scale. For example, Bressanelli et al. (2020) discuss how domestic politics can create bottom-up pressures on member states' behavior at the EU level. Most of the studies on governments' ideological positions have focused on left–right policy positions (e.g., (Hix, 1999; Hosli et al., 2011; Mattila, 2004; Pircher & Farjam, 2021)). Investigating the impact of left–right positions of member states on their voting behavior can be based either on their left–right positions or the distance from the average left–right position (Hosli et al., 2011). Another potential operationalization of this variable is the left–right position of a government and that of its political opposition (Pircher & Farjam, 2021). Party ideologies can also affect the formation of cooperation networks in the Council (Huhe et al., 2022). Accordingly, our third and fourth hypotheses are as follows:

H3: *An EU member state's left–right position is correlated with its voting behavior in the Council.*

H4: *Member states that are further away from the left–right policy position of a winning coalition in the Council of the EU are more likely to cast an opposing vote in this institution.*

To explore the potential effects of the independent variables above, based on a new data collection covering the period 2010 to 2021, control variables will be added. First, our analysis controls for whether an EU member state at the time of the vote held the rotating presidency of the Council of the EU. Analyses of this issue

Table 1 Distribution of decisions in the Council by voting rule, 2010 to 2021

Voting rule	Number	Percentage
Qualified majority voting	1229	91.92
Unanimity	107	8.00
Reinforced QM	1	0.08

Source Data collection based on the Consilium Website, for the period 2010 to 2021

have demonstrated that when a member state holds the presidency, this can impact its voting behavior (Tallberg, 2003; van Gruisen et al., 2019). Similarly, attitudes toward the EU in public opinion are likely to affect a member state's voting behavior in the Council (Franchino et al., 2022; Hagemann et al., 2017; Pircher & Farjam, 2021; Wratil, 2018). It is also likely to affect bargaining success in the Council of the EU (Mariano & Schneider, 2022). Accordingly, we will also control for attitudes toward the EU in public opinion, in each member state, and for each year.

3 Variable Operationalization and Data

In this section, we discuss our dependent variables, independent variables of interest, and controls in more detail and discuss how we operationalize them.

3.1 Votes in the Council

We examine decisions made by the Council in the period between 2010 and 2021. The data have been collected from the Council's Consilium Website.[2] To get a query from the Consilium's database, we have used the EURLEX R package (Ovádek, 2021).[3] The dataset on which our analysis will be based includes 1337 decisions made by the Council between 2010 and 2021. For most of these decisions, this concerns the phase in which the Treaty of Lisbon was applicable, and QMV in the Council was the relevant voting rule. For the entire dataset, Table 1 depicts the distribution of decisions based on the voting rule that was applicable.

In this chapter, our analysis will focus on the QMV cases, as with 91.92 of all decisions taken by the Council, this constituted the clear majority of cases. EU member states when it comes to a vote in the Council can either vote in favor, against, or abstain. Moreover, they can decide not to participate in the voting process. Out of 1229 QMV decisions taken by the Council between 2010 and 2021, in 572 cases, at least one member state did not vote in favor to the proposal (casting a vote

[2] https://www.consilium.europa.eu/en/general-secretariat/corporate-policies/transparency/open-data/voting-results.

[3] This extraction has been conducted by Arash Pourebrahimi and constitutes part of his PhD dissertation (Leiden University).

Table 2 Contested QMV decisions in the Council of the EU, 2010–2021

QMV proposals	Number	Share in total number of votes (%)
Not supported by at least one Member State (against, abstention, or non-participating)	571	46.37
Contested by at least one Member State (voting against or abstaining)	490	39.86

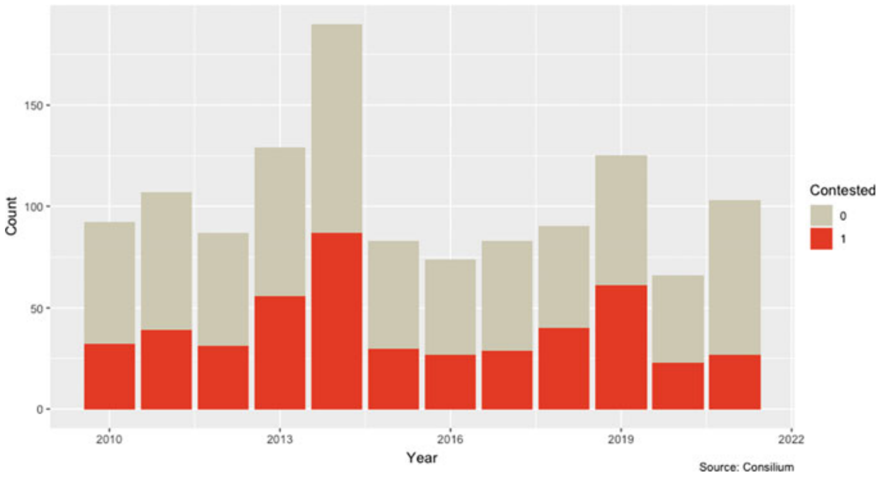


Fig. 1 Contested decisions in the Council, 2010–2021

against the proposal, abstaining, or not participating in the vote). In 490 cases, at least one member state voted against a proposal or abstained. Please note that this is a remarkable increase compared to earlier periods of voting in the Council of the EU (e.g., Mattila (2004)). Table 2 provides an overview of the percentage of proposals that were not supported or contested by at least one EU member state in the Council.

We assume that a proposal is contested if there is at least one opposing vote or an abstention vote on the proposal because these have the same impact on the results in terms of forming a winning coalition when the Council decides by QMV. By comparison, we do not consider non-participation to constitute contestation. Though not participating in a vote can be considered to demonstrate a lack of support for a proposal, in some cases, non-participation may simply reflect that a proposal is irrelevant to a member state. Hence, we do not consider this to reflect contestation. Figure 1 shows the number of QMV decisions in the Council between 2010 and 2021 that were contested. A fascinating side-result of the analysis is that in the year 2014—when the double-majority rule became effective that was incorporated into the Treaty of Lisbon—a spike of decisions was taken in the Council.

Figure 2 shows the percentage of QMV decisions that were contested in the Council of the EU for each year between 2010 and 2021. Next to the spike of legislative

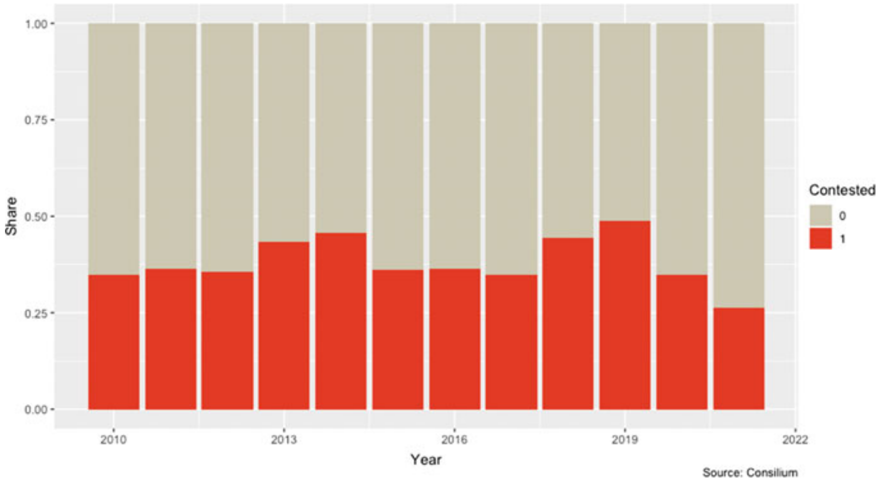


Fig. 2 Share of contested decisions in the Council, 2010–2021

acts voted on in 2014, the occurrence of votes was lower in 2020 and 2021, and potentially, the COVID-19 pandemic. This might be related to Brexit, as the United Kingdom left the Union in early 2020.

Around 40% of proposals in the Council, for the period analyzed here, were contested. However, the contestation rate varied among different policy areas: more than half of the proposals voted on were contested in some policy areas, including Social Policy, Industry, Employment, and Environment. Figure 3 shows the total and the number of contested proposals voted on in the Council, by policy areas. Figure 4 depicts the contestation rate for each policy area.

Each legislative decision in the Council is decided in a specific Council configuration. Ministers and permanent representatives of each member state take decisions in sessions held in each Council configuration. Figures 5 and 6 show the distribution of decisions in Council configurations and the share of contested decisions in each configuration, respectively. The highest contestation rates in terms of Council configurations are in the areas of Justice and Home Affairs, Education, Youth, and Culture. This finding needs more investigation, since Education, Youth, and Culture as separate policy areas do not show such a high rate of contestation.

The 1337 QMV decisions included in our dataset contain 33910 decisions taken by EU member states. Note that this includes the years before Croatia joined the EU (in 2013), and after Brexit. Thus, the number of decisions taken by member states is not simply the number of decisions times a constant number of member states. Less than four percent of these 33910 decisions were contested votes. Table 3 summarizes the frequency of member state vote choices.

The contestation rate shown in Table 2 is much higher than the contestation rate in Table 3. The reason for this remarkable difference is that contesting coalitions in the Council are usually consisting of a small number of member states. For instance,

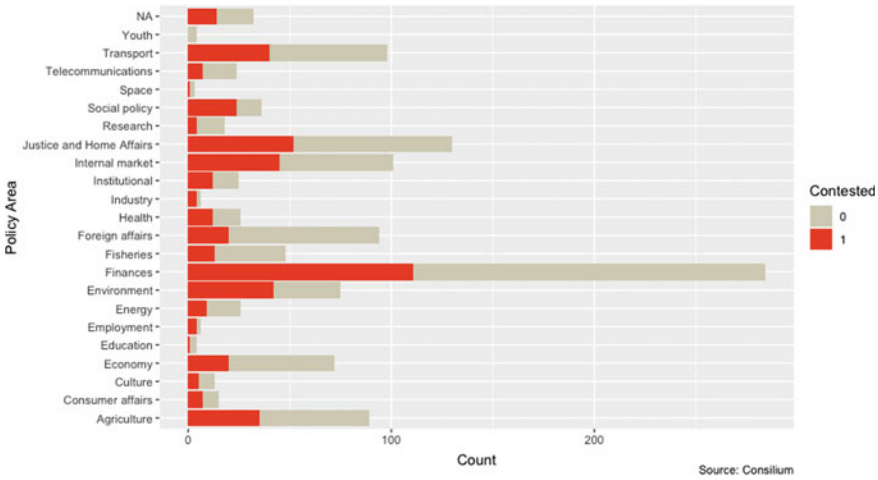


Fig. 3 Contested decisions in the Council by policy area, 2010–2021

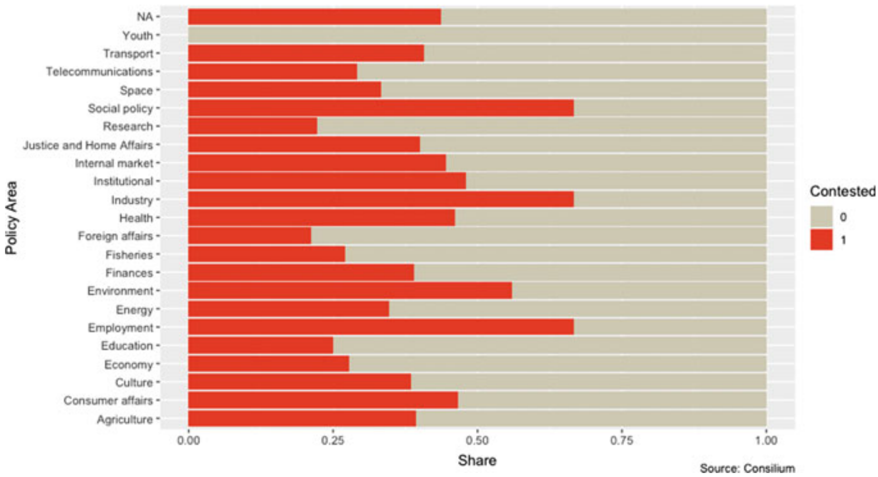


Fig. 4 Share of contested decisions in the Council by policy area, 2010–2021

if for a proposal 26 member states vote in favor and only one opposes, the proposal is considered to be a contested proposal, while only 3.7% of votes cast constitute opposing votes.

Table 4 summarizes different voting behavior patterns of member states in the Council of the EU. The last column of Table 4 reports the contestation rate of member states. The United Kingdom, with 18.95%, had the highest rate of contestation, followed by Austria (5.69%), Hungary (5.53%), and the Netherlands (5.13%).

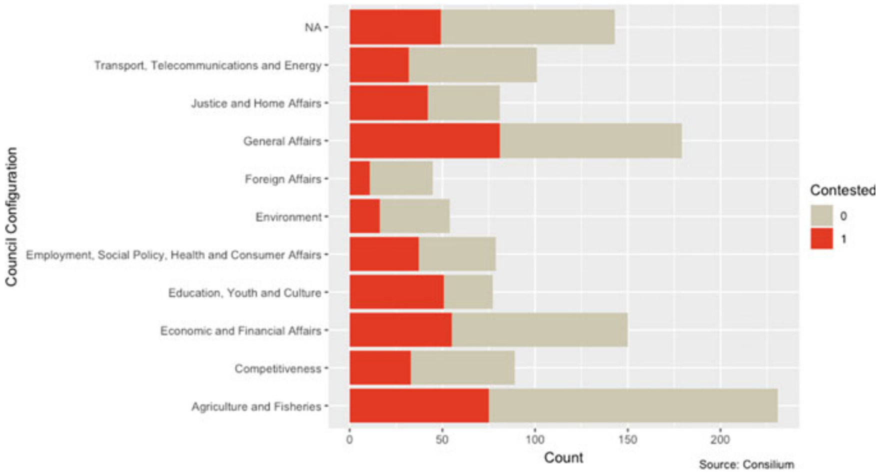


Fig. 5 Decisions in the Council by configuration, 2010–2021

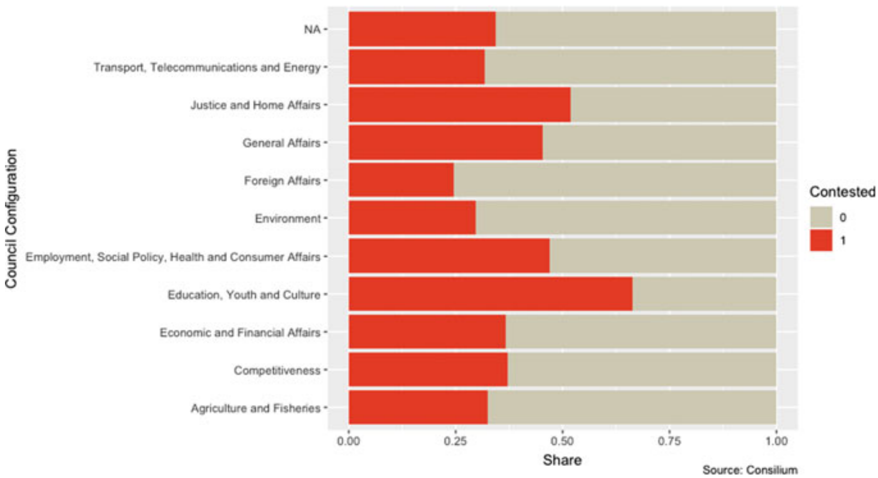


Fig. 6 Share of contested decisions in the Council by configuration, 2010–2021

Table 3 Distribution of vote choices in the Council of the EU, 2010–2021

Decision	Number	Percentage of total
In favor	32,654	96.3
Not participating	250	0.74
Abstention	526	1.55
Against	480	1.42

Table 4 Contestation Rates by EU Member States, 2010–2021

EU Member State	Yes (%)	Non-participation (%)	Abstention (%)	No (%)	Contestation rate (no or abstention) (%)
Austria	94.3	0	3.25	2.44	5.69
Belgium	97.56	0	1.87	0.57	2.44
Bulgaria	96.83	0.08	1.55	1.55	3.1
Croatia	98.44	0.11	1.22	0.22	1.44
Cyprus	98.62	0	0.08	1.3	1.38
Czech Republic	96.26	0.08	2.28	1.38	3.66
Denmark	88.75	8.39	0.81	2.04	2.85
Estonia	98.13	0	1.14	0.73	1.87
Finland	98.86	0	0.41	0.73	1.14
France	99.76	0	0.16	0.08	0.24
Germany	95.2	0	2.52	2.28	4.8
Greece	99.27	0	0.24	0.49	0.73
Hungary	94.39	0.08	2.52	3.01	5.53
Ireland	92.91	5.79	0.98	0.33	1.31
Italy	98.7	0.08	0.41	0.81	1.22
Latvia	98.45	0	0.9	0.65	1.55
Lithuania	98.7	0.08	0.49	0.73	1.22
Luxembourg	97.88	0	1.22	0.9	2.12
Malta	97.97	0	0.73	1.3	2.03
Netherlands	94.87	0	1.55	3.58	5.13
Poland	95.2	0.08	2.36	2.36	4.72
Portugal	98.45	0	1.06	0.49	1.55
Romania	98.45	0.08	0.65	0.81	1.46
Slovakia	97.88	0	1.14	0.98	2.12
Slovenia	98.7	0	0.98	0.33	1.31
Spain	98.21	0.08	0.81	0.9	1.71
Sweden	96.34	0.08	0.57	3.01	3.58
UK	74.84	6.22	13.1	5.94	18.95

3.2 Government Composition

To measure the ideological position of each government in each EU member state, we use the ParlGov database (Döring et al., 2022). In this database, each party's left–right policy position is measured on a spectrum ranging from 0 to 10, where 0 stands for extreme left and 10 for the extreme right. To measure the government

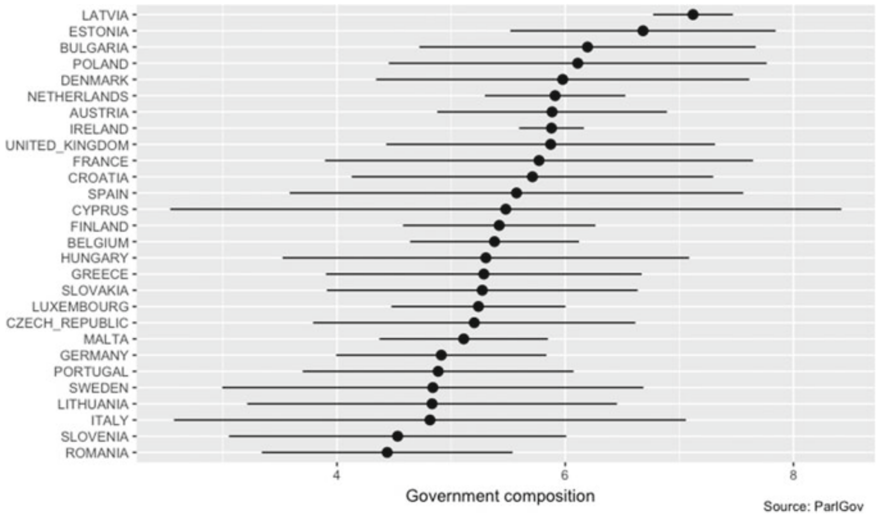


Fig. 7 Member States' average government composition, 2000–2021

composition of each government, we calculate the weighted average of the parties included in each government. The weights are based on the number of members each party has in the parliament; the weight of each party is the number of members of parliament of that party, divided by the total number of members of parliament of all parties that form the government.

Another possibility to measure the weight of each party in the government is to use the number of ministers of that party over the total number of ministers in the government. However, for this method, we need to assume that the weights (importance) of all ministries are equal, which may not be a realistic assumption. Moreover, it is not clear when applying this approach how to deal with the fact that the head of the government will be from a given political party: The office of the head of the government could be included like another ministerial position, with equal weight, or alternatively, be attributed a higher weight in the calculations. But if the latter solution is chosen, how much weight should that be? Hence, using the share of each party in the parliament may be a less arbitrary and straightforward approach to reflect the weight of each party in government.

Figure 7 shows the average government compositions in all EU member states between 2000 and 2021, with two standard deviations on both sides. Higher variance means that the government composition has fluctuated more during that period. A more stable government composition (reflected by lower variance), however, does not necessarily mean that parties forming an EU member state's government did not change much. It just implies that the weighted average of the ideological positions of the parties in government did not vary much during the period analyzed.

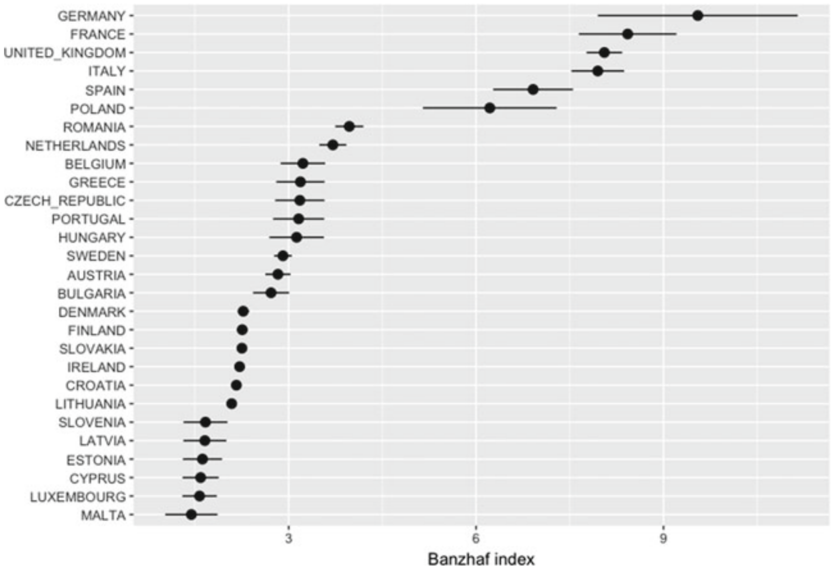


Fig. 8 Member States' power distribution, 2010–2022

3.3 Power

To measure the power of each EU member state for the period of this analysis, we use the Banzhaf power index. The reason for applying the Banzhaf instead of the Shapley Shubik index is that assumptions underlying the Banzhaf power index, in our view, align better with the Council's voting system. Most notably, the Shapley Shubik index assumes that players vote in sequence, while in the framework of the Banzhaf index, players cast their votes simultaneously. To calculate the Banzhaf index for each EU member state, we apply the same method as used by Gábor (2020). To measure the Banzhaf index, we use the computer program created by Leech and Leech (2004). This platform allows us to also account for the double-majority system. We calculate the (non-normalized) Banzhaf index of each member state based on the triple majority system before 2014 and the double-majority system since November 2014. We use EUROSTAT[4] as a source for the population size of each member state for each year analyzed (as of January 1st).

The variation in each member state's Banzhaf power index is due to Croatia's accession to the EU in July 2013, the change of the voting system in the Council of the EU (November 2014), population size changes since 2014, and finally, Brexit (February 2020). These changes affected the power indices of the member states. For Croatia and the UK, the power indices are calculated only for the period they were members of the EU (Fig. 8).

[4] https://ec.europa.eu/eurostat/web/population-demography/demography-population-stock-balance/database.

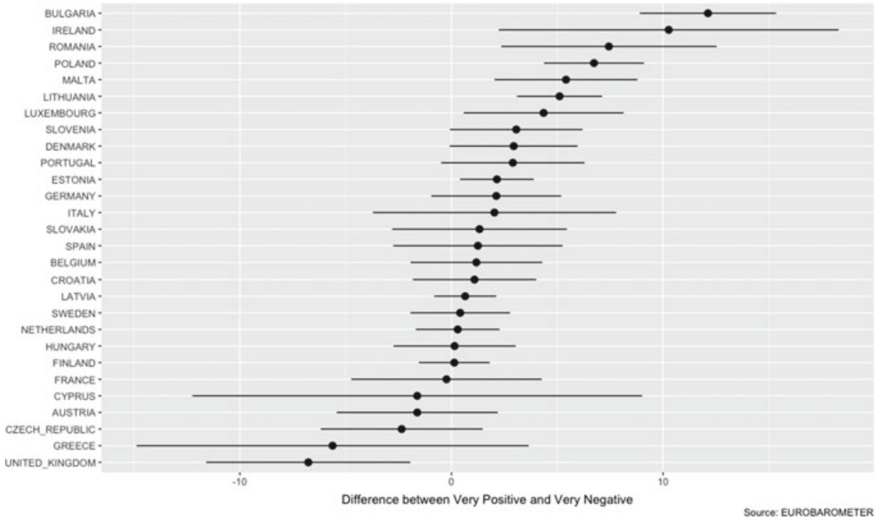


Fig. 9 Member States' public opinion on the EU, 2010–2021

3.4 Public Attitudes Towards the EU

To gauge public opinion in EU member states, we will use public attitudes toward the EU, based on data published by the Eurobarometer survey. These are conducted twice a year among European citizens and include a high number of questions, on a variety of topics. For our analysis, we will use the question "In general, does the EU conjure up for you a very positive, fairly positive, neutral, fairly negative, or very negative image?" We use the difference between the percentage of very positive and very negative as a measure of public sentiment: the higher this number, the more positive the attitude toward the EU is in an EU member state. The assessments we use for each year constitute the average of the answers to these items in two rounds of the Eurobarometer survey in that year. Figure 9 shows the average public opinion towards the EU for each member state for the timespan 2010 to 2021. To make the figure clear, we multiply the respective numbers by 100 (but do not use this extension when applying the models).

3.5 Net Contributions

To calculate the net contribution of each member state to the EU budget, we use EU spending and revenue data published by the European Commission. Net contributions of each member state in each year are measured as total national contributions plus total own resources minus total expenditures. Negative contributions imply that

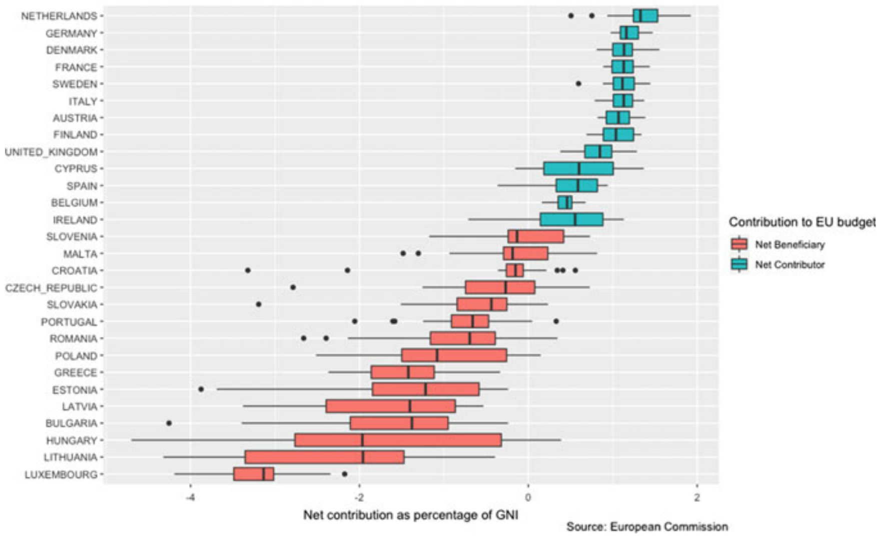


Fig. 10 Member States' net contribution to the EU budget, 2010–2020

the member state is a net beneficiary. In our model, we use net contributions as a percentage of Gross National Income (GNI). GNI numbers are from the same source, published by the European Commission. However, this dataset does not cover the year 2021; hence, we limit this analysis to 2010–2020. Figure 10 shows the average annual net contributions over GNI for each EU member state between 2010 and 2020.

3.6 Presidency

For each voting decision by each EU member state, we record if it was made while the member state held the (rotating) presidency of the Council of the EU. The presidency rotates every six months. Given this rotating scheme, most vote decisions by a member state will be taken when it does not hold the presidency. Table 5 shows votes while an EU member state was holding the rotating presidency, and for time spans it did not: Member states holding the presidency contested a vote (i.e., voted no or abstained) only in 0.9% of all cases, while the contestation ratio for those not holding the presidency was more than three percent on average.

Table 5 Contesting votes for member states holding the rotating presidency compared to regular member states

Vote choice	Yes		Non-participation		Abstention		No		Total
	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage	
Holding presidency	1209	98.37	9	0.73	4	0.33	7	0.57	1229
Not holding presidency	31,445	96.22	241	0.74	522	1.6	473	1.45	32,681

4 Explaining Voting Outcomes in the Council of the European Union

What drives vote choice in the Council of the EU? Based on the independent variables presented above as well as the control variable, we will now proceed to estimate vote choice based on such explanatory factors. We will first apply binomial logistic regression analysis, followed by an ordered logistic regression. For our binomial logit model, we capture vote choice by member states as a binary variable, with each decision by a member state (i.e., its minister or delegation in the Council) constituting one observation. Accordingly, the (binary) dependent variable is support for (coded as 0) or contestation of (coded as 1) a proposal voted on in the Council. A member state supports a legislative proposal by voting in favor (yes). Opposing votes (no) and abstentions are classified as contesting votes. We do not take non-participation in the vote into account, as even though not participating implies not supporting a decision, this does not constitute an action considered to be contestation: A member state might not participate in a decision simply because that decision is not relevant from its perspective.

The independent variables of our logistic model, as discussed above, are a member state's net contribution to the EU budget, its power index, government composition, and the distance of its government (assessed on a left–right policy scale) from the one of the winning coalition. We control for whether a member state held the presidency at the time of the vote and for domestic attitudes toward the EU, based on Eurobarometer data. To assess the net contributions, as discussed above, we use contributions to the EU budget plus traditional own resources minus EU expenditures, divided by the GNI. The presidency is captured as a binary variable indicating whether the member state held the Council presidency (coded as 1) or not (coded as 0). The power index is the (normalized) Banzhaf power index.

For the analysis, we use the weighted average of the left–right government positions and the absolute distance of each government from the average of government compositions as contained in the winning coalition in the Council of the EU for each proposal voted on. We use this variable to examine whether being at a larger distance from the winning coalition in terms of the left–right ideological position is correlated with the probability that a member state contests a vote in the Council of the EU. Public opinion is captured by citizens in a member state stating they are very positive

Table 6 Binomial logit model (random effects)

	Dependent variable
	Decision (contest = 1)
Net contribution divided by GNI	0.139**
	(0.059)
Presidency	−0.901***
	(0.308)
Power	0.036
	(0.050)
Government composition left–right	0.046
	(0.030)
Government distance left–right from winning coalition	0.347***
	(0.071)
Domestic public opinion	0.999
	(1.000)
Observations	29,388

Note *p < 0.1; **p < 0.05; ***p < 0.01

about the EU minus those that are very negative. The results of the binomial logit model (with random effects) are displayed in Table 6.

Our results show that the higher the net contribution of a member state to the EU budget, the higher the probability that it contests a vote in the Council of the EU. This probably, as expected, decreases when a member state holds the presidency of the Council of the EU. Our binomial logit model shows that the further the government is on the left–right ideological spectrum from the winning coalition in the Council of the EU, the higher the probability it contests a vote.

In addition to a binomial logit model, we run an ordered logit regression. For this model, we order the decisions as yes (coded as 1), non-participation (coded as 2), abstention (coded as 3), and no votes (coded as 4). We use this order because yes votes reflect support, non-participation is likely to reflect a neutral position, abstention reflects contestation in the sense that it diminishes the fraction of member states supporting the proposal, and no demonstrates a higher level of contestation, even though practically, abstention and no votes have the same effect. Table 7 shows the results of our ordered logit model.

The insights generated by the ordered logit models align with the ones based on the logistic models. However, in this model, the government composition on the left–right policy scale is significant at the 10% level, with right-wing governments being more likely to contest decisions in the Council.

Based on the results of the estimates generated by the models above, there is not sufficient empirical evidence to reject hypothesis one. Accordingly, it is likely that member states contributing more to the EU's budget feel more confident in terms of contesting decisions in the Council of the EU than member states that are net beneficiaries. This might be because these member states allocate a higher share

Table 7 Ordered logit model (random effects)

	Dependent variable
	Decision
Net contribution divided by GNI	0.148**
	(0.059)
Presidency	−0.518**
	(0.232)
Power	0.024
	(0.051)
Government composition left–right	0.046*
	(0.027)
Government distance left–right from winning coalition	0.413***
	(0.068)
Domestic public opinion	0.381
	(0.084)
Observations	29,618

Note $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

of their total income to the EU and feel more confident demonstrating contestation when they do not fully support a new piece of legislation, even though they cannot change the outcome. Sending a signal to domestic voters might be another reason for these net contributors to contest a vote. However, member states are less likely to demonstrate opposition to a vote when they hold the Council presidency: In this role, the mandate is to support the collective agenda and strive for consensus, which is still a guiding principle in (intergovernmental) EU negotiations. It would be interesting to explore whether member states holding the presidency also aim to lower the rate of contestation in the Council when they hold the presidency (and to see which member states succeed in doing so).

Our results do not show that member states' power indices, assessed by the Banzhaf power index, are correlated with their voting behavior. Thus, our second hypothesis can get rejected. However, when we run a logistic model, not as a panel model, the power index as an explanatory variable generates statistically significant (and robust) results. A reason why the power index does not generate statistically significant results as a predictor could be because this variable does not vary much during the period assessed. Larger member states have more voting power. Thus, the investigation of the impact of power on the probability of contestation is related to the differences between the voting behavior of large member states compared to others in the Council.

The results generated by both models imply that the probability of contestation is lower when a member state's ideological (left–right) policy position as reflected in its government constellation is closer to the one of the winning coalition. Thus, we cannot reject hypothesis 4. The reason for this might be that coalitions in the Council are partly formed based on ideologies (as captured by left–right policy positions). This aspect may need more thorough investigation, notably as the left–right ideological positioning of ministers who attend Council sessions might affect the voting behavior of a member state in a specific Council configuration. However, the results in terms of the impact of the ideology of a government on voting behavior (see hypothesis 3) are mixed: The ordered logit model demonstrated that right-wing governments in EU member states are more likely to contest decisions in the Council. However, the effect can differ between policy areas. For instance, right-wing governments might not support the EU's agenda on topics such as foreign and security policy (van Kersbergen & de Vries, 2007) or immigration, where their perspectives may be nationalistic, while this is not necessarily the case in other policy domains.

5 Conclusion

Using all voting data for the Council of the EU between 2010 and 2021, this paper has examined the determinants of member states' contesting votes in the Council of the EU. Our dataset includes more than 29,000 votes based on the QMV rule. The results of our random effects binomial logit model and an ordered logit model show that member states allocating a higher net share of their GNI to the EU are more likely to oppose votes in the Council of the EU. Member states with an ideological left–right policy position at a higher distance from the one of the winning coalitions in the Council are more likely to contest votes. However, our results in terms of the impact of left–right positions on voting behavior are mixed. Moreover, our analysis does not provide evidence showing that a member state's voting power is linked to its voting behavior in the Council.

Our analysis has the same limitations that Council roll call data studies face: We only have information on the decisions that made it through the decision-making process in the EU. Decisions that did not have a chance to get approved in the Council of the EU would not show up in the formal voting procedures. Similarly, given the culture of consensus that exists in the Council of the EU, the number of contested votes is low overall. Despite these limitations, however, our study can shed light on some aspects of voting behavior in the Council of the EU. Given the fact that the EU is continuously evolving and more decisions are taken, more research can further improve our understanding of the functioning of the EU. In addition, more information on ideological differences among EU member states, in particular at the ministerial level, can shed further light on the dynamics of legislative decision-making in the Council of the EU.

References

Bailer, S. (2004). Bargaining success in the European Union: The impact of exogenous and endogenous power resources. *European Union Politics, 5*, 99–123.

Bailer, S., Mattila, M., & Schneider, G. (2015). Money makes the EU go round: The objective foundations of conflict in the Council of Ministers. *Journal of Common Market Studies, 53*, 437–456.

Bressanelli, E., Koop, C., & Reh, C. (2020). EU Actors under pressure: Politicisation and depoliticisation as strategic responses. *Journal of European Public Policy, 27*, 329–341.

Döring, H., Constantin, H., & Manow, P. (2022). *Parliaments and governments database (ParlGov): Information on parties, elections and cabinets in established democracies*. Development version

Franchino, F., Kayser, M. A., & Wratil, C. (2022). Electoral competitiveness and responsiveness: Rational anticipation in the EU Council. *Journal of European Public Policy, 29*, 42–60.

Gábor, J. (2020). Impact of Brexit on voting power in Council of the European Union. *Open Political Science, 1*, 192–197.

Hagemann, S., Hobolt, S. B., & Wratil, C. (2017). Government responsiveness in the European Union: Evidence from Council voting. *Comparative Political Studies, 50*, 850–876.

Hix, S. (1999). Dimensions and alignments in European Union politics: Cognitive constraints and partisan responses. *European Journal of Political Research, 35*, 69–106.

Hosli, M. O., Mattila, M., & Uriot, M. (2011). Voting in the Council of the European Union after the 2004 enlargement: A comparison of old and new member states. *Journal of Common Market Studies, 49*, 1249–1270.

Huhe, N., Thomson, R., Arregui, J., & Naurin, D. (2022). Intergovernmental cooperation networks, national policy positions and partisan ideologies: Longitudinal evidence from the Council of the European Union. *Journal of European Public Policy, 29*, 78–96.

Leech, D., & Leech, R. (2004). *A computer program for the analysis of qualified majority voting power under the treaty of nice*. http://www.warwick.ac.uk/ecaae/

Lundgren, M., Bailer, S., Dellmuth, L. M., Tallberg, J., & Târlea, S. (2019). Bargaining success in the reform of the Eurozone. *European Union Politics, 20*, 65–88.

Mariano, N., & Schneider, C. J. (2022). Euroscepticism and bargaining success in the European Union. *Journal of European Public Policy, 29*, 61–77.

Mattila, M. (2004). Contested decisions: Empirical analysis of voting in the European Union Council of Ministers. *European Journal of Political Research, 43*, 29–50.

Ovádek, M. (2021). Facilitating access to data on European Union laws. *Political Research Exchange, 3*, 1.

Perarnaud, C., & Arregui, J. (2022). Do member states' permanent representations matter for their bargaining success? Evidence from the EU Council of Ministers. *Journal of European Public Policy, 29*, 97–116.

Pircher, B., & Farjam, M. (2021). Oppositional voting in the Council of the EU between 2010 and 2019: Evidence for differentiated politicisation. *European Union Politics, 22,* 472–494.

Tallberg, J. (2003). The agenda-shaping powers of the EU Council Presidency. *Journal of European Public Policy, 10*, 1–19.

van Gruisen, P., Vangerven, P., & Crombez, C. (2019). Voting behavior in the Council of the European Union: The effect of the trio presidency. *Political Science Research and Methods, 7*, 489–504.

van Kersbergen, C. J., & de Vries, C. E. (2007). Interests, identity and political allegiance in the European Union. *Acta Politica, 42*, 307–328.

Wratil, C. (2018). Modes of government responsiveness in the European Union: Evidence from Council negotiation positions. *European Union Politics, 19*, 52–74.

Zimmer, C., Schneider, G., & Dobbins, M. (2005). The contested Council: Conflict dimensions of an intergovernmental EU institution. *Political Studies, 53*, 403–422.

Codecision in Context Revisited: The Implications of Brexit

Nicola Maaser and Alexander Mayer

Abstract The paper analyzes the implications of the UK's leave from the European Union for the distribution of power between the Council of the European Union and the European Parliament and within the Council under the EU's codecision procedure. Unlike previous studies on this issue, we do not treat the Council in isolation but follow the richer framework of Maaser & Mayer (2016). That is, we model the codecision procedure as a bargaining game between the European Parliament and the Council under various a priori preference assumptions. We find that the withdrawal of the UK has no significant effect on the power distribution between the European Parliament and the Council and that it is mainly the large member states that benefit from the UK's leave.

1 Introduction

On December 31, 2020, the transition period for the United Kingdom (UK) to withdraw from the European Union (EU) officially came to an end. This so-called Brexit—short for "British exit"—marked the end of a year-long process that started with a referendum on June 23, 2016, when 52% of voters voted in favor of leaving and handed a surprise victory to the "leave" campaign. It then took more than four years to negotiate both a *Withdrawal Agreement* and a *Trade and Cooperation Agreement* governing the future economic relationship between the EU and the UK. Since Jan-

N. Maaser
Department of Economics and Business Economics, Aarhus University, Aarhus, Denmark
e-mail: nmaaser@econ.au.dk

A. Mayer (✉)
Department of Economics, University of Bayreuth, Bayreuth, Germany
e-mail: alexander.mayer@uni-bayreuth.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_19

uary 1, 2021, and after nearly 50 years of membership, the UK is no longer a member of the EU.[1]

While there is already a large body of work on the economic impact of Brexit (see, e.g., Born et al. 2019; Breinlich et al. 2022; Sampson 2017), only few papers investigate the consequences for EU decision-making. We will contribute to the latter by exploring how Brexit affects the distribution of power (i) between the EU's two main decision-making bodies, the European Parliament (EP) and the Council of the European Union (CEU) and (ii) within the Council. Kóczy (2021) and Grech (2021) have already analyzed the power consequences of Brexit in a framework that treats the Council in isolation assuming that it is the only decision-making body of the EU. Both find that it is mainly the large members that benefit from the UK's exit.

We take a broader perspective and consider that the EP and the Council in the codecision procedure—today the most important decision-making procedure in the EU—are commonly regarded by EU observers as equal co-legislators. Also according to the EP's own description (European Parliament, 2012, p. 5), the "ordinary legislative procedure is based on the principle of parity between the [...] European Parliament, representing the people of the Union, and the Council, representing the governments of Member States."

Applied models on negotiations between the EP and the Council arrive at remarkably different conclusions regarding the distribution of power in the EU. The findings—theoretical as well as empirical—vary from a genuine, balanced two-chamber system (Crombez, 2000; Garrett & Tsebelis, 2000; Moser, 1997) to a more or less pronounced asymmetry in favor of the Council (Costello & Thomson, 2013; Maaser & Mayer, 2016; Napel & Widgrén, 2006).

We take the model of Maaser and Mayer (2016) and try to identify which of the two institutions, if any, benefits (more) from Brexit. Furthermore, we check whether the conclusion drawn by Kóczy (2021) and Grech (2021) that it is primarily the larger countries that benefit from Brexit still holds in a more realistic setting of EU codecision characterized by the following main features. First, the model takes into account the fact that negotiations between the EP and the Council are characterized by mutual gains and concessions. Such a principle of juste retour results in a cooperative rather than conflictual mode of negotiation. This observation can be captured by modeling bargaining between the EP and the Council in the final stage of the codecision procedure using the *Kalai–Smorodinsky bargaining solution* (Kalai & Smorodinsky, 1975). Second, Maaser and Mayer (2016) take into account heterogeneity across member states. That is, they assume that there is a closer political connection within countries than between them. In particular, they assume that both citizens' and delegates' ideal points are a priori identically distributed, but not independent. The final assumption of the model of Maaser and Mayer (2016) concerns the fact that Council delegates are supposed to represent national governments, while Members of the European Parliament (MEPs) are supposed to directly represent citizens. If the

[1] Formally, the UK has left the EU already on January 31, 2020. However, there was a transition period for the remainder of 2021. See, e.g., The Economist (2021) for an extensive discussion of the Brexit process.

Council's delegates are faithful agents of their national governments, then the preferences of a member state's Council delegate—usually a disciplined bureaucrat—are congruent with the respective national median voter. In contrast, a country's MEPs are modeled such that their ideal points are drawn from a symmetric triangular distribution over the respective policy interval. This is consistent with the observation that MEPs are ideologically very diverse and cover the entire political space.

The remainder of the paper is organized as follows. Section 2 presents the codecison procedure in more detail. Section 3.1 derives the EP's and the Council's respective bargaining positions, Sect. 3.2 the predicted policy outcome of the codecision game and both institutions' influence on the outcome. Section 3.3 introduces various assumptions on preferences. Section 4 discusses the results from the quantitative power analysis. Section 5 concludes.

2 The Codecision Procedure

The *ordinary legislative procedure* as laid down in Article 294 of the *Treaty on the Functioning of the European Union* (TFEU) provides that the EP and the Council must reach consensus on the basis of a Commission proposal by means of alternate amendments. If no agreement is achieved during the first two readings, a compromise is sought in the third and final stage, the *Conciliation Committee*. It is composed of one delegate per Council member (i.e., 28 members before and 27 members after Brexit) and an equal number of EP delegates.[2] In the event of a successful conciliation, the final joint draft is voted upon under closed rule, i.e., neither institution can amend the proposal. Adoption requires a simple majority of votes cast in the EP and a qualified majority in the Council; otherwise (or if no joint text has been produced), the proposal fails and the legal status quo applies.

We portray the EU's codecision procedure as a non-cooperative extensive form game of perfect information between the Council and the EP (see Fig. 1) and assume one-dimensional spatial preferences for members of the EP and the Council delegates.[3] Imposing standard rationality assumptions on the players, we can derive the codecision outcome by backward induction. It is determined by the anticipated outcome of the last stage, i.e., the Conciliation Committee (cf. Sect. 3.2).

Following the approach developed by Napel and Widgrén (2004; 2006), we do not consider it appropriate to force the extensive form game in Fig. 1 into the form of sim-

[2] Still, the two delegations are potentially asymmetric, as each Council member is involved in the Conciliation Committee, but the EP delegates in the Committee are merely agents whose interests need not be fully aligned with those of their principals (see Franchino and Mariotto 2013).

[3] Since the Commission has no formal say in the negotiations (i.e., EP and Council can implement any policy they jointly agree on), but only fulfills a mediating and facilitating role, it is not treated as a relevant player in the codecision game. However, this contradicts some qualitative analyses as well as the prevailing view in the media that the Commission has a major influence on legislation. A first approach to incorporate a strategic forward-looking Commission in models of EU codecision is offered by van Gruisen and Crombez (2021).

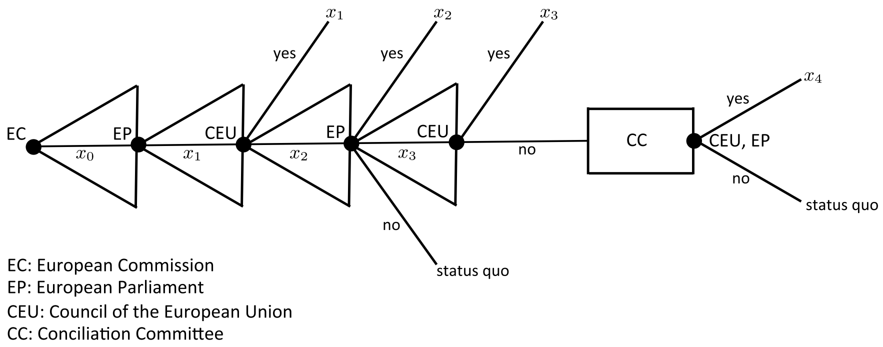


Fig. 1 Stylized codecision game tree

ple voting games and calculate standard binary indices such as the Shapley–Shubik index (Shapley & Shubik, 1954). Instead, in the next section, we will introduce a specific policy space, preferences for players, and some more structure for negotiations in the Conciliation Committee.

3 The Model

Drawing on the model of Maaser and Mayer (2016), we consider a convex policy space $X \subseteq \mathbb{R}^n$ equipped with metric ρ. For most of the analysis, we assume $n = 1$ and ρ to be the Euclidean distance. Let $q \in X$ denote the status quo regarding the issue in question. Assume that agents (i.e., MEPs, Council members and the representatives in the Conciliation Committee) have single-peaked preferences such that agent i with ideal point $\lambda_i \in X$ has utility $u_i(x) = -|\lambda_i - x|$ from policy $x \in X$. That is, utility falls linearly in distance between λ_i and x.

Given these preferences, we can predict the outcome of conciliation, and thus codecision, by specifying, first, how EP's and Council's internal decision rules translate individual preferences of MEPs and Council members into the common ideal points π for the EP and μ for the Council (*intra-institutional bargaining*), and second, how these collective preferences jointly determine a bargaining agreement (*inter-institutional bargaining*).

3.1 Intra-institutional Bargaining

Before the conciliation procedure begins, the respective bargaining positions of the EP and the Council must be agreed in accordance with the respective internal decision-making rules of each institution. MEPs decide on any conciliation compromise by simple majority rule, which—in our one-dimensional Euclidean setting—makes the median MEP pivotal.

Denoting the ordered ideal points of the MEPs by $\pi_{(1)}, \ldots, \pi_{(l)}$, where $l = 705$ for EU27 and $l = 751$ for EU28, the median MEP's ideal point is $\pi^{EU28} = \pi_{(376)}$ for EU28 and $\pi^{EU27} = \pi_{(353)}$ for EU27, respectively.

Agreement on the Council's common ideal point μ is subject to the voting rules laid down in the Treaty of Lisbon and is therefore more involved.[4] Denoting the ordered ideal points of the individual Council members by $\mu_{(1)}, \ldots, \mu_{(m)}$, $m = 27, 28$, we distinguish two cases. First, if the Council wants to replace the status quo q by a policy $x > q$, members with positions $\mu_{(m)}, \mu_{(m-1)}$, etc., will be the most enthusiastic about this change. The pivotal member of the Council is then the country that first brings about the required qualified majority as less and less enthusiastic members in favor of policy x are added. We will refer to this member as the Council's *left* pivot L^{EU28} and L^{EU27}, respectively, with respective ideal point μ_L^{EU28} and μ_L^{EU27}. Similarly, the critical member for a policy change to the left of q is referred as the Council's right pivot R^{EU28} (R^{EU27}) with associated ideal point μ_R^{EU28} (μ_R^{EU27}).

Denoting EU28's population by P^{EU28} (and EU27's population by P^{EU27}), the voting weight of the member with ideal point $\mu_{(i)}$ by $w(\mu_{(i)})$ and the population she represents by $p(\mu_{(i)})$, we have

$$R^{\text{EU28}} = \min\left\{ \min\left\{r \in \{16, ..., 28\}\colon \sum_{i=1}^{r} p(\mu_{(i)}) \geq 0.65 P^{\text{EU28}}\right\}, 25\right\}$$

and

$$L^{\text{EU28}} = \max\left\{ \max\left\{l \in \{1, ..., 13\}\colon \sum_{i=l}^{28} p(\mu_{(i)}) \geq 0.65 P^{\text{EU28}}\right\}, 4\right\},$$

for EU28 and

$$R^{\text{EU27}} = \min\left\{ \min\left\{r \in \{15, ..., 27\}\colon \sum_{i=1}^{r} p(\mu_{(i)}) \geq 0.65 P^{\text{EU28}}\right\}, 24\right\}$$

and

$$L^{\text{EU27}} = \max\left\{ \max\left\{l \in \{1, ..., 13\}\colon \sum_{i=l}^{27} p(\mu_{(i)}) \geq 0.65 P^{\text{EU28}}\right\}, 4\right\},$$

for EU27.

[4] Under the Treaty of Lisbon, a qualified majority must consist of at least 55% of member states and must represent at least 65% of total EU population. Additionally, a blocking minority must include at least four Council members.

3.2 Inter-institutional Bargaining

One way—namely that of Napel & Widgrén (2006)—to arrive at predictions about the outcome of negotiations in the Conciliation Committee is offered by bargaining solutions such as the *Nash bargaining solution* (Nash, 1950). To incorporate that actual codecision negotiations are characterized by (i) a cooperative mode of behavior and (ii) an informal principle of juste retour that recognizes that each member should gain something from the negotiations, we on the contrary follow Maaser and Mayer (2016) and consider the *Kalai–Smorodinsky solution* (Kalai & Smorodinsky, 1975) instead of the Nash solution. The Kalai–Smorodinsky solution has several attractive features that make it, at least in our view, a better fit for bargaining situations characterized by mutual concessions.[5]

To describe the Kalai–Smorodinsky solution, we first need to define the so-called *utopian point* u^* as the —typically unrealisable—point at which *every* player i receives her *aspiration level* a_i, which is the highest possible surplus for player i, provided that all other players receive at least their disagreement payoff. The Kalai–Smorodinsky solution assumes that all players cut back proportionally with respect to the utopian point in such a way that the ratio of their aspirations is preserved. Formally, it is defined by

$$\xi^{KS}(\mathcal{U}, d) = d + \bar{\lambda}(u^* - d),$$

where $\bar{\lambda} = \max\left\{\lambda \in \mathbb{R} : d + \lambda(u^* - d) \in \mathcal{U}\right\}$. Geometrically, the Kalai–Smorodinsky solution $\xi^{KS}(\mathcal{U}, d)$ is just the intersection of $\mathcal{U}$'s Pareto frontier and the straight line connecting the disagreement point d and the utopian point u^*.

In our setting, we will—without loss of generality—assume that $\text{sign}(q - \pi) = \text{sign}(q - \mu)$, i.e., gains from trade exist, and $|\pi - q| \leq |\mu - q|$, i.e., the EP's ideal point π is closer to q than the Council's ideal point μ. This implies that $u^*_{\text{EP}} = 0$ and

$$u^*_{\text{CEU}} = \begin{cases} 0 & \text{if } |\pi - q| \geq |\pi - \mu| \\ -|\pi - \mu| + |\pi - q| & \text{otherwise.} \end{cases}$$

Figure 2 geometrically illustrates the location of the Kalai–Smorodinsky solution for these two cases (assuming $q = 0$).

The precise location of the policy x^{KS} that corresponds to ξ^{KS} can be determined as follows:

Proposition 1 *Maaser and Mayer (2016, p. 223). Whenever there are gains from trade (i.e.,* $\text{sign}(q - \pi) = \text{sign}(q - \mu)$*), the Kalai-Smorodinsky solution to the bar-*

[5] See Maaser and Mayer (2016, pp. 220f) for a detailed discussion of why the Kalai–Smorodinsky solution seems to reflect actual codecision negotiations better than the Nash solution. Maaser and Mayer (2016, p. 223) also showed that the Kalai–Smorodinsky solution has empirical support in the context of EU codecision: using real policy issues reported in the DEUII dataset (Thomson et al., 2012, 2006), they find that the Kalai–Smorodinsky solution does significantly better in predicting the actual codecision outcome than the Nash solution.

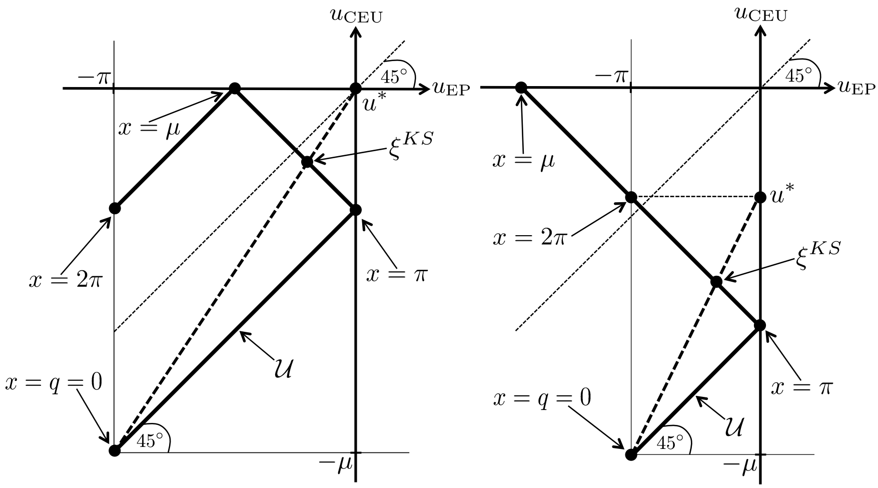


Fig. 2 Kalai–Smorodinsky bargaining solution with $u^* = (0, 0)$ in the left panel and $u^* = \big(0, -|\pi - \mu| + |\pi - q|\big)$ in the right panel

gaining problem $(\mathcal{U}, d)$ *corresponds to agreement on a policy* x^{KS} *which is located on the Pareto frontier but nearer to the ideal point which is closer to the status quo. More specifically,*

$$x^{\text{KS}}(\pi, \mu, q) = \begin{cases} \pi + \frac{\mu-\pi}{1+(\mu-q)/(\pi-q)} & \textit{if } |\pi - q| \le |\mu - q| \textit{ and } |\pi - q| \ge |\pi - \mu|, \\ \mu + \frac{\pi-\mu}{1+(\pi-q)/(\mu-q)} & \textit{if } |\pi - q| > |\mu - q| \textit{ and } |\mu - q| > |\pi - \mu|, \\ \pi + \frac{\pi-q}{3} & \textit{if } |\pi - q| \le |\mu - q| \textit{ and } |\pi - q| < |\pi - \mu|, \\ \mu + \frac{\mu-q}{3} & \textit{if } |\pi - q| > |\mu - q| \textit{ and } |\mu - q| \le |\pi - \mu|. \end{cases}$$

Proof See Maaser and Mayer (2016, pp. 234f). The qualitative finding of Proposition 1 remains true for multidimensional policy spaces $X \in \mathbb{R}^n$ with $n \ge 2$. A proof is available from the authors upon request.

The expected influence of the EP, the Council and of individual Council members on EU codecisions can be quantified using the *power as outcome sensitivity* approach introduced by Napel and Widgrén (2004). This approach conceives of a posteriori power as the sensitivity of the equilibrium outcome with respect to small changes in a player's behavior or preferences. The *strategic measure of power* (SMP) then evaluates a priori power as expected a posteriori power, using a probability measure with a priori credentials.

Taking the partial derivatives of the predicted outcome, the a posteriori power of the EP for a *given* realization of the status quo q and ideal points of MEPs and Council members then is

$$\frac{\partial x^{\mathrm{KS}}(\pi,\mu,q)}{\partial \pi} =$$
$$\begin{cases} \frac{(q-2\mu)(2q-\pi-\mu)+(\pi+\mu)q-2\pi\mu}{(2q-\pi-\mu)^2} & \text{if } (q<\pi<\mu \text{ or } \mu<\pi<q) \text{ and } |\pi-q|>|\pi-\mu|,\\ \frac{(q-2\mu)(2q-\pi-\mu)+(\pi+\mu)q-2\pi\mu}{(2q-\pi-\mu)^2} & \text{if } (q<\mu<\pi \text{ or } \pi<\mu<q) \text{ and } |\mu-q|>|\pi-\mu|,\\ \frac{4}{3} & \text{if } (q<\pi<\mu \text{ or } \mu<\pi<q) \text{ and } |\pi-q|<|\pi-\mu|,\\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for an individual member k of the Council, we obtain

$$\frac{\partial x^{\mathrm{KS}}(\pi,\mu(\mu_1,\ldots,\mu_m),q)}{\partial \mu_k} =$$
$$\begin{cases} \frac{(q-2\pi)(2q-\pi-\mu)+(\pi+\mu)q-2\pi\mu}{(2q-\pi-\mu)^2} & \text{if } (q<\pi<\mu \text{ or } \mu<\pi<q) \text{ and } |\pi-q|>|\pi-\mu|,\\ \frac{(q-2\pi)(2q-\pi-\mu)+(\pi+\mu)q-2\pi\mu}{(2q-\pi-\mu)^2} & \text{if } (q<\mu<\pi \text{ or } \pi<\mu<q) \text{ and } |\mu-q|>|\pi-\mu|,\\ \frac{4}{3} & \text{if } (q<\mu<\pi \text{ or } \pi<\mu<q) \text{ and } |\mu-q|<|\pi-\mu|,\\ 0 & \text{otherwise,} \end{cases}$$

where $\mu = \mu_k$, i.e., member k is the Council's pivotal member.

3.3 Assumptions on Preferences

To account for the fact that (i) the EP represents citizens while the Council represents states (or rather governments) and (ii) there are reasons that are likely to lead to closer political ties between voters within countries than between them, we follow Maaser and Mayer (2016) and include the following assumptions about the ideal points of each decision-maker in our analysis.[6]

(SPA) All individual voters have spatial preferences, characterized by ideal point ν^i in policy space X.

We are aware that a constitutional analysis should ignore knowledge about specific preferences. From a normative perspective of constitutional design, it can therefore be appealing to assume that all individual ideal points are independently and identically distributed (i. i. d.). However, while the normative perspective certainly implies that all ideal points ν^i should be drawn from the same marginal distribution F, in our view positive correlation within countries should be allowed. We therefore consider the partition $\mathfrak{C} = \{\mathcal{C}_1,\ldots,\mathcal{C}_m\}$, $m = 27, 28$ of the EU voter population into m *constituencies* with $n_j = |\mathcal{C}_j| > 0$ members each.

[6] See Maaser and Mayer (2016, pp. 225–228) for a detailed discussion and implications of these facts.

In particular, we determine individual ideal points ν^i by a two-step random experiment: First, we draw a constituency-specific shock θ_j independently for each $j = 1, \ldots, m$ from a distribution G with standard deviation σ_{ext}. This parameter captures the degree of *external heterogeneity* between $\mathcal{C}_1, \ldots, \mathcal{C}_m$ for the policy issue at hand. The parameter θ_j is supposed to reflect the expected ideal point of citizens from $\mathcal{C}_j$. Each citizen $i \in \mathcal{C}_j$ is then assigned an individual ideal point ν^i from a distribution F_j that has mean θ_j and is otherwise just a shifted version of the same distribution F for each constituency $j = 1, \ldots, m$.[7] F's standard deviation σ_{int} is a measure of the *internal heterogeneity* in any constituency. It intuitively reflects the differences in opinion within any given $\mathcal{C}_j$. In summary, we account for heterogeneity among countries by assuming that

(HET) The ideal points of all citizens are identically distributed with convoluted a priori distribution $G * F$ but not independent: Citizens in constituency $\mathcal{C}_j$ experience a shock θ_j, which is independent of θ_k for any $k \neq j$.

The introduction of heterogeneity has an additional benefit: It is worth including another institutional fact, namely the *degressive proportionality* in the national composition of the EP (see Table 1 for the respective number of MEPs).[8]

Our two final assumptions reflect the idea that ideal points of Council members and MEPs should not be identically distributed:

(MED) The preferences of country j's representative in the Council are congruent with the country's median voter.[9] More formally, representative j has ideal point

$$\mu_j = \text{median}\{\nu^i : i \in \mathcal{C}_j\}.$$

(CRD) MEPs who are elected in country j are a clustered random draw from that country's electorate. More formally, let s_j denote the number of seats allocated to country j. If μ_j is the median voter position in $\mathcal{C}_j$, then the ideal points $\pi_1^j, \ldots, \pi_{s_j}^j$ of j's MEPs are distributed according to the symmetric triangular distribution $F(a_j, \mu_j, b_j)$ on the interval $[a_j, b_j]$ with peak location μ_j, where a_j and b_j, respectively, are the lower and upper bound of country j's policy space.

[7] Specifically, we draw θ_j from a uniform distribution $\mathbf{U}(-a, a)$ with variance σ_{ext}^2 and then obtain $\nu^i = \theta_j + \varepsilon$ with $\varepsilon \sim \mathbf{U}(0, 1)$.

[8] Note that 27 of UK's 73 seats in the EP were distributed between the remaining members, with France and Spain getting the most additional seats (five each).

[9] Many EU member states have coalition governments consisting of only a subset of parliamentary parties; the more realistic assumption would hence be to consider the median voter within the government coalition instead of the entire population. This, however, would require information about each member state's type of government coalition. Since many coalition governments consist of parties from the center of the political spectrum, we believe it is a good approximation to consider the median of the entire population.

4 Results

Since we are not interested in a player's influence on a single issue but rather in *expected* influence, we quantify a priori power by computing the average of a posteriori power (Sect. 3.2) over a large number of uniformly distributed issues.

In a first step, we draw a country specific shock θ_j that reflects the expected ideal point of citizens from country C_j. Second, the ideal points of voters from different countries are drawn from different distributions with mean θ_j, which are shifted versions of some distribution F. Specifically, the ideal points of MEPs are drawn from shifted triangular distributions $F_1, \ldots, F_m$ with the median of the respective country as the peak (i.e., for EU27, 96 ideal points are drawn from the distribution of Germany, 79 from the distribution of France, etc.). The ideal points of the Council members are drawn from the respective shifted beta distributions F_1, ..., F_m with parameters $((n_j+1)/2, (n_j+1)/2)$.[10] The status quo q is drawn from a uniform distribution over an interval that captures all possible preferences in our heterogenous EU. Third, we sort the realized ideal points and determine the respective pivot positions of EP and Council according to their respective internal decision rules (cf. Sect. 3.1). Fourth, we identify the policy outcome x^{KS} and derive the a posteriori power of EP and Council for the given realizations of q, π and μ. By repeating this procedure up to 10^9 times, we obtain numerical estimates of the SMP numbers of the EP, the CEU and the individual Council members.[11]

We report the simulation results for the EU28 with UK and the EU27 without UK in Table 1. The first noteworthy result is that Brexit has basically no impact on the respective ex ante power relations between the EP and the Council. The SMP of the EP slightly increases by about 1.5 %, while the Council's SMP decreases by a negligible 0.5 %. This suggests that some formerly pivotal positions of the Council which were held by the UK now belong to the EP or result in deadlock due to the lack of gains of trade. Second, our results regarding the intra-institutional distribution of power show—with very few exceptions like Spain and France—a clear pattern: The larger a country in terms of its population size, the more it benefits from Brexit (in relative terms). While Germany, the largest member of the EU sees its SMP increase by about 20 %, and the EU's smallest member, Malta, loses about 4 % of its ex ante power. The relationship between the influence effect of Brexit and population size is also illustrated in Figs. 3 and 4.

Let us remark that the change in a country's SMP results from two opposing effects. The first effect is positive: In the configurations in which the UK was doubly

---

[10] This follows from assuming that voter ideal points within a *given* country are uniformly distributed (see Arnold et al., 1992, pp. 13f).

[11] To reduce the variance of our SMP estimator ($\widehat{\mathrm{SMP}}$), we use the difference between the exactly calculated Shapley–Shubik index (SSI) and the estimated Shapley–Shubik index ($\widehat{\mathrm{SSI}}$, which is a by-product of our simulation) as a control variate for our SMP estimator. It can be shown that $\widehat{\mathrm{SMP}}_i + c(\widehat{\mathrm{SSI}}_i - \mathrm{SSI}_i)$ is an unbiased estimator of the true SMP of country i. The variance of this estimator gets minimized for $c = -\mathrm{cov}(\widehat{\mathrm{SMP}}_i, \widehat{\mathrm{SSI}}_i)/\mathrm{var}(\widehat{\mathrm{SSI}}_i)$. With this approach, the variance can be reduced by up to 41 %.

Table 1 2020 Eurostat population, EP seats, and power for EU28 (w/ UK) and EU27 (w/o UK)

Member state	Population	EP seats		SMP ($\times 10^{-2}$)		Δ SMP %
		EU28	EU27	EU28	EU27	
Germany	83,166,711	96	96	9.35	11.26	+20.34
France	67,320,216	74	79	7.25	8.54	+17.34
United Kingdom	67,025,542	73	–	7.21	–	–
Italy	59,641,488	73	76	6.31	7.46	+18.17
Spain	47,332,614	54	59	4.91	5.89	+19.95
Poland	37,958,138	51	52	4.06	4.47	+10.12
Romania	19,328,838	32	33	2.38	2.55	+6.89
Netherlands	17,407,585	26	29	2.19	2.34	+6.84
Belgium	11,522,440	21	21	1.62	1.73	+6.66
Greece	10,718,565	21	21	1.54	1.64	+6.47
Czech Republic	10,693,939	21	21	1.54	1.64	+6.51
Sweden	10,327,589	20	21	1.51	1.60	+6.23
Portugal	10,295,909	21	21	1.50	1.60	+6.52
Hungary	9,769,526	21	21	1.45	1.55	+6.28
Austria	8,901,064	18	19	1.38	1.46	+6.07
Bulgaria	6,951,482	17	17	1.19	1.25	+5.31
Denmark	5,822,763	13	14	1.09	1.13	+4.78
Finland	5,525,292	13	14	1.06	1.11	+4.41
Slovakia	5,457,873	13	14	1.05	1.10	+4.62
Ireland	4,964,440	11	13	1.01	1.05	+3.90
Croatia	4,058,165	11	12	0.92	0.95	+3.10
Lithuania	2,794,090	11	11	0.81	0.82	+0.86
Slovenia	2,095,861	8	8	0.75	0.74	−0.59
Latvia	1,907,675	8	8	0.73	0.72	−0.85
Estonia	1,328,976	6	7	0.68	0.66	−2.27
Cyprus	888,005	6	6	0.64	0.62	−3.49
Luxembourg	626,108	6	6	0.61	0.59	−3.87
Malta	514,564	6	6	0.60	0.58	−4.41
CEU aggregate				65.36	65.06	−0.46
EP				26.29	26.69	+1.53

pivotal (i.e., both in the Council and in the Conciliation Committee), another country is now the pivotal Council member. Some of these altered pivotal positions translate into doubly pivotal positions for the remaining countries; others lead to pivotality of the EP or in deadlock. While this effect is positive for all countries, the larger countries intutitively stand to benefit more as they are closer substitutes to the UK. The second effect is ambiguous: It stems from configurations in which the UK was not pivotal, but provided the necessary weight to make some other country pivotal. It

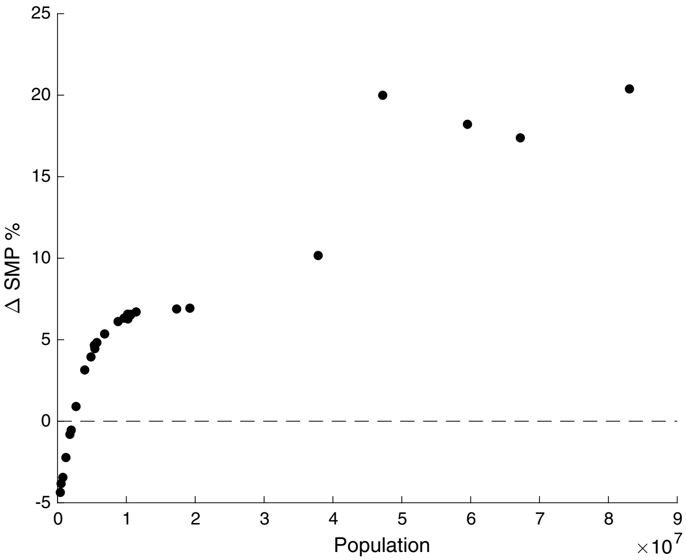


Fig. 3 Relationship between population size and percentage change in SMP

is intuitively clear that the exit of the third-largest country has a particularly negative impact on the opportunities of smaller countries in this respect. The effect on larger countries is, however, mixed: They lose some formerly pivotal positions since the UK does no longer contribute the necessary weight, but they also gain new pivotal positions formerly captured by small countries. As a result, the overall effect depends on population size, with large countries benefiting most from the first effect, while for small countries the second effect outweighs the first.

5 Concluding Remarks

This paper provides, at least to our knowledge, one of the most comprehensive analyses to date of the Brexit implications on the distribution of power within the European Union. Unlike existing studies, we do not treat the Council of the European Union in isolation, but consider many previously neglected features of the broader EU institutional framework. Nevertheless, the qualitative result of the binary power analysis that it is mainly the large, more populous countries that benefit from the UK's exit remains robust. Also, small countries continue to be overrepresented, e.g., Germany's SMP now is about 20 times that of Malta, although Germany has about 160 times as many inhabitants; before Brexit, the balance of power was even more skewed at 15:1.

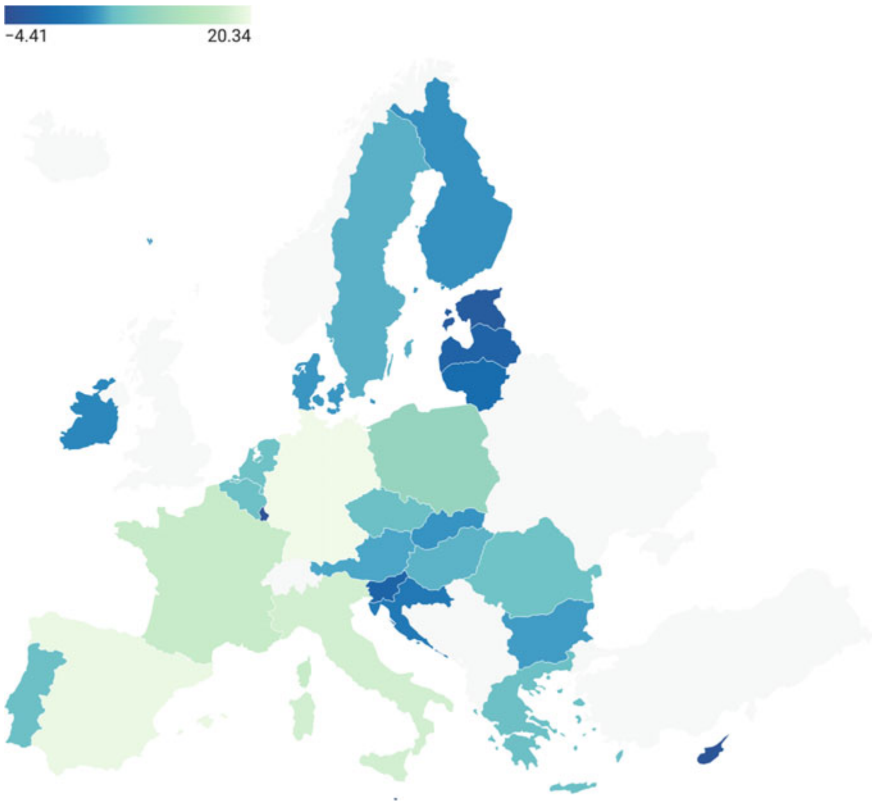


Fig. 4 Map of winners and losers from Brexit

An important aspect that remains hidden in this a priori analysis is the fact that countries have different degrees of connection with the UK and are therefore affected differently by its exit. Brexit has already prompted some realignments of the remaining member states in policy areas where the UK used to play a significant role within the EU such as budgetary policy or trade. At this stage, it is too early to say whether new groupings such as the Frugal Four (Austria, Denmark, the Netherlands, Sweden) or the New Hanseatic League (the Baltic States, Denmark, Finland, Ireland, the Netherlands and Sweden) will develop into deep partnerships that coordinate their activities in a similar way to the Benelux Union. If so, a future analysis of the impact of Brexit on EU decision-making could model the new EU27 as a composite game in which these alliances act as a single player.[12] Such an approach would however leave the a priori framework and rather hold a middle ground between a priori and a posteriori power analysis.

[12] Mayer (2018) provides a power analysis for the composite game that treats the Benelux union as one bloc.

Acknowledgements Parts of this article were published in *Social Choice and Welfare 46*, 213–237, 2016. We are grateful to Hannu Nurmi for valuable feedback on an earlier draft.

References

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A First Course in Order Statistics*. Wiley.

Born, B., Müller, G. J., Schularick, M., & Sedláček, P. (2019). The costs of economic nationalism: Evidence from the Brexit experiment. *Economic Journal, 129*, 2722–2744.

Breinlich, H., Leromain, E., Novy, D., & Sampson, T. (2022). The Brexit vote, inflation and U.K. living standards. *International Economic Review, 63*, 63–93.

Costello, R., & Thomson, R. (2013). The distribution of power among EU institutions: Who wins under codecision and why? *Journal of European Public Policy, 20*, 1025–1039.

Crombez, C. (2000). Institutional reform and co-decision in the European Union. *Constitutional Political Economy, 11*, 41–57.

European Parliament. (2012). Guide to codecision and conciliation under the Treaty of Lisbon. DV 898248. Conciliations and Codecision Unit, European Parliament.

Franchino, F. ,& Mariotto, C. (2013). Explaining negotiations in the Conciliation Committee. *European Union Politics*, *14*, 345–365.

Garrett, G., & Tsebelis, G. (2000). Legislative politics in the European Union. *European Union Politics, 1*, 9–36.

Grech, P. D. (2021). Power in the Council of the EU: Organizing theory, a new index, and Brexit. *Social Choice and Welfare, 56*, 223–258.

Kalai, E., & Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica, 43*, 513–518.

Kóczy, L. Á. (2021). Brexit and power in the Council of the European Union. *Games, 12*, 51.

Maaser, N., & Mayer, A. (2016). Codecision in context: Implications for the balance of power in the EU. *Social Choice and Welfare, 46*, 213–237.

Mayer, A. (2018). Luxembourg in the early days of the EEC: Null player or not? *Games, 9*, 29.

Moser, P. (1997). A theory of the the conditional influence of the European Parliament in the cooperation procedure. *Public Choice, 91*, 333–350.

Napel, S., & Widgrén, M. (2004). Power measurement as a sensitivity analysis: A unified approach. *Journal of Theoretical Politics, 16*, 517–538.

Napel, S., & Widgrén, M. (2006). The inter-institutional distribution of power in EU codecision. *Social Choice and Welfare, 27*, 129–154.

Nash, J. F. (1950). The bargaining problem. *Econometrica, 18*, 155–162.

Sampson, T. (2017). Brexit: the economics of international disintegration. *Journal of Economic Perspectives, 31*, 163–184.

Shapley, L. S., & Shubik, M. (1954). A method for evaluating the distribution of power in a committee system. *American Political Science Review, 48*, 787–792.

The Economist (2021). *How Brexit happened*. https://www.economist.com/britain/2021/01/02/how-brexit-happened

Thomson, R., Arregui, J., Leuffen, D., Costello, R., Cross, J., Hertz, R., & Jensen, T. (2012). A new dataset on decision-making in the European Union before and after the 2004 enlargements (DEUII). *Journal of European Public Policy, 19*, 604–622.

Thomson, R., Stokman, F. N., Achen, C. H., & König, T. (2006). *The European Union Decides*. Cambridge University Press.

van Gruisen, P., & Crombez, C. (2021). The Commission and the Council Presidency in the European Union: Strategic interactions and legislative powers. *European Journal of Political Economy, 70*, 102040.

Field Experiments and Quasi-experiments

Proximity-Based Preferences and Their Implications Based on Data from the Styrian Parliamentary Elections in 2019

Christian Klamler

Abstract Single-peaked preferences in political models are usually determined on a left-right political scale via the distance between a voter's optimal location along that scale and the candidates' positions. In this paper, we want to use a sort of proximity approach based on data from an exit poll undertaken right after the Styrian Parliamentary Elections in 2019. First, we show that a single-peakedness model does not perfectly fit the data, because there is no unanimous agreement on the order of the parties along the political left-right scale. Second, declared preferences and proximity-based preferences do differ significantly indicating that other factors do play a role in determining the voters' preferences. Third, the actual impact of those other factors on the election results will be specified by comparing the (hypothetical) election results, using different well-known voting rules, for two preference profiles, one based on the stated preferences and one determined by proximity-based preferences.

1 Introduction

The theory of social choice investigates the aggregation of individual preferences over a set of parties, candidates or alternatives into a group choice or ranking. Originally, the focus was mostly on theoretical issues, with the most famous results being those from Arrow (1963), Sen (1979), Gibbard (1973) and Satterthwaite (1975).

However, with the rise of behavioral social choice (see, e.g., Regenwetter and Tsetlin (2004) and Regenwetter et al. (2006)) and in the field of political science (see, e.g., Cox (1997) and Blais and Degan (2019)), empirical issues received increased attention in the scientific literature and provided insight into the consequences of the rather negative theoretical results for the practical use of voting rules.

Consequently, more intensive focus has been put on the analysis of real-world elections illuminating various aspects, from the performance of voting rules to strategic behavior of voters (see, e.g., Baujard et al. (2014, 2018), Roescu (2014), Alòs-

C. Klamler (✉)
Institute of Public Economics, University of Graz, Graz, Austria
e-mail: christian.klamler@uni-graz.at

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_20

Ferrer and Grani (2012), Darmann et al. (2017, 2019), Kawai and Watanabe (2013), Stephenson et al. (2018)).

One widely studied topic within social choice is domain restrictions which guarantee a winner and/or a transitive social ranking in voting situations. Probably, the most famous of those domain restrictions is single-peakedness (Black, 1958), which is based on analyzing voting behavior in one dimension (for a more thorough discussion consult, e.g., surveys by Gaertner (2001), Monjardet (2009) or Grofman (2019a)). Whenever the candidates can be aligned on a single dimension and all voters' preferences based on the order of candidates are single peaked, there will be a Condorcet winner, i.e., a candidate that is preferred by a majority to every other candidate.[1] That is, preferences are based on distances from a voter's ideal position in this single dimension to the voter's perceived positions of the candidates. Obviously, single-peakedness would require that all voters consider the positions of the candidates to be the same. This is a rather unrealistic assumption in many political models. However, individual proximity-based preferences might be a good approximation to a voter's actual preferences.

In this paper, we want to investigate proximity-based preferences, i.e., preferences determined by the distance between a voter's own political position on a left-right scale and her perception of the political parties, using data from the 2019 parliamentary elections in the Austrian region of Styria. To be able to do so, during election day voters were asked, in front of the polling stations, to respond to various questions concerning their preferences and their evaluation of themselves and the parties along a usual political left-right scale. This data is used in the following way:

First, to analyze the consistency of actually stated preferences in the sense of comparing them with proximity-based preferences and second, how the results for various well-known voting rules would change based on those different preference profiles.

The paper is structured as follows: Sect. 2 introduces the experimental design. Section 3 discusses the one-dimensional approach using distances from a voter's ideal point on a political left-right scale to her perceived positions of the running parties. Whether the voters' stated preferences are consistent with their hypothetical proximity-based preferences will be analyzed in Sect. 4. The created proximity-based preferences will then be used in Sect. 5 to determine the election outcomes for various well-known voting rules and are compared to the outcomes received for the actually stated preferences. Finally, Sect. 6 concludes the paper.

2 Experimental Design and Data

The data for this experimental study was collected during election day for the 2019 Styrian Parliamentary Elections. We developed a design for the experiments and undertook an exit poll. In front of nine randomly chosen voting stations in the city of

[1] Strictly speaking, single-peakedness does not necessarily lead to a strong Condorcet winner if there is an even number of voters and/or there are ties in peaks.

Graz, voters were invited to provide information about their preferences. A total of 937 voters participated in the exit poll. The survey contained questions about a full ranking of the six parties running for election[2], their approval preferences, evaluative voting, their personal left-right orientation as well as their evaluation of the parties along that dimension, how they actually voted in the real election, and various other (statistical) questions.[3] In those nine polling stations, we reached about 5.5% of the total voters. Given that the voters participated voluntarily in the experiment, the raw data obviously has a certain participation bias. This can be seen in comparing the actual voting results with the declared votes by the participants. We aimed at correcting that bias by the use of weights; the weights were determined by dividing the shares for each party in the official result in those nine polling stations by the share the same party received in the exit poll. For example, the SPÖ had a share of 18.5% of the votes in the official result and received a support of 13.76% in the exit poll. Therefore, participants supporting the SPÖ were under-represented in the exit poll leading to a weight of $1.34 = \frac{18.50}{13.76}$. If we compare the calculated weights, we observe that, in addition to the SPÖ, also the ÖVP and the FPÖ have weights larger than one and are therefore under-represented. The other parties are over-represented in the survey leading to weights smaller than one. Because for different questions in the survey the response-rate was different, weights have been adapted accordingly based on the number of participants responding to that question using the same procedure as explained before.[4] Table 1 provides an overview over official and declared voting results.

3 Left-Right Dimension

Given the huge interest in spatial voting in the literature (e.g., Enelow & Hinich (1984), Saari (1995) or, more recently, Enelow and Hinich (2008), Grofman (2019a; 2019b), Schofield (2019)), in the experiment the participants were explicitly asked to communicate their evaluations of the parties on the usual political left-right dimension. In particular, they were able to assign numbers from 1 (politically far left)

---

[2] Throughout the paper, the parties are abbreviated as follows: SPÖ—Sozialdemokratische Partei Österreichs (Social Democratic Party of Austria); ÖVP—Österreichische Volkspartei (Austrian People's Party); FPÖ—Freiheitliche Partei Österreichs (Freedom Party of Austria); GREENS—Die Grünen (The Green Party); KPÖ—Kommunistische Partei Österreichs (Communist Party of Austria); NEOS - Das Neue Österreich und Liberales Forum (The New Austria and Liberal Forum). On a left-right ideological scale, in general, the KPÖ, GREENS and SPÖ are considered to be rather left, NEOS in the center, ÖVP central to center-right, and FPÖ right-wing. This perception is also confirmed by the voters who participated in this exit poll and were asked to position the parties on a left-right scale.

[3] A translation of the used survey can be found in the appendix.

[4] For example, because not all participants provided both, complete preferences and evaluations along the left-right scale, for the latter the weights have been corrected by just taking those voters into account who actually provided that information.

Table 1 Official results, declared votes and weights

Parties	Official results (%)	Declared official votes (%)	Weights
SPÖ	18.50	13.76	1.34
ÖVP	23.94	16.76	1.43
FPÖ	14.62	5.90	2.48
GREENS	21.67	36.07	0.60
KPÖ	13.24	16.42	0.81
NEOS	8.04	11.10	0.72

Table 2 Mean, median and mode positions

	SPÖ	ÖVP	FPÖ	GREENS	KPÖ	NEOS	Self
Mean	4.32	7.82	10.23	3.07	2.74	5.84	4.64
Median	4	8	11	3	2	6	4
Mode	4	9	11	3	1	6	6

to 11 (politically far right). Additionally, they were asked about their own political orientation along that dimension. Roughly 75% of the participants answered those questions. According to those responses, one is able to provide mean, median and mode positions for the parties stated in Table 2.

In addition, one interesting question is how the parties have been perceived in general, i.e., what is the distribution of their perceived political orientation by the voters on the left-right scale. Those distributions are presented in Fig. 1.

There seems to be a clear trend that the KPÖ is considered to be the most left party followed by the Greens. The SPÖ and the NEOS are more seen as central parties, whereas ÖVP and FPÖ are considered right and far-right parties, respectively. Compared with the voters' own political orientation, it seems quite obvious that for the current dataset, SPÖ, Greens and NEOS are—on average—closer to the voters' ideal points, and therefore, any consideration of distances of party evaluations to individual ideal points should be to the benefit of those parties. However, it is also clearly observable that there is no real consensus about the positions of the parties along the left-right scale. Hence, an approach based on single-peakedness can, in principle, not be applied because the underlying order of the parties is not unanimously accepted (see Feld and Grofman (1986) for consequences of partial single-peakedness).

Besides the actual positions of the parties, one could also check for the general pairwise evaluation of the parties along the left-right scale. As can be seen in Table 3, where the numbers indicate the percentage of voters that consider the row party to be politically to the left of the column party, there is a strong agreement on the relative positions of pairs of parties for at least some pairs.

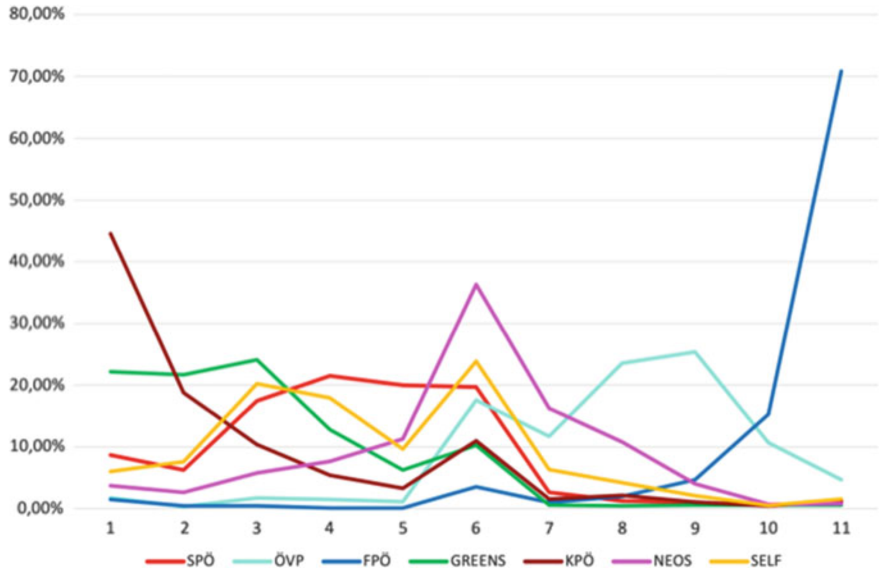


Fig. 1 Left-right distribution

Table 3 Percentage of voters considering row party to the left of column party

	SPÖ (%)	ÖVP (%)	FPÖ (%)	GREENS (%)	KPÖ (%)	NEOS (%)
SPÖ		89.31	96.31	16.46	12.65	69.16
ÖVP	2.95		90.29	1.97	4.55	7.49
FPÖ	0.49	1.84		0.37	0.49	0.86
GREENS	66.09	91.52	96.44		25.31	85.50
KPÖ	71.99	87.59	95.95	54.30		78.38
NEOS	14.37	76.04	94.84	3.56	11.79	

Although the evaluation is clear for certain pairs of parties (e.g., the Greens and the FPÖ), for other parties it is rather mixed (in particular between the Greens and the KPÖ). Again, this indicates that there is no unanimous agreement on an order of the parties along the left-right scale.

4 Proximity-Based Preferences and Consistency

If political models in one dimension are to be considered a good approximation of actual voting behavior, it might be of interest to check to what extent proximity-based preferences match the voters' actual preferences. First let us illustrate how proximity-based preferences are determined using an example.

Table 4 Example proximity-based preferences

	Party A	Party B	Party C	Party D	Self
Position	8	3	9	7	4
Distance	4	1	5	3	
Rank	3	1	4	2	

Consider four parties $\{A, B, C, D\}$ evaluated by a voter on a left-right scale starting at 1 (extreme left) and ending at 11 (extreme right). Table 4 states the voter's evaluation (positions on the left-right scale) of the parties and her self-evaluation, i.e., her optimal political position.

The first row indicates the personal evaluation of the parties by the voter, who considers her optimal position to be at 4. Taking the absolute value of the difference between the voter's self-evaluation and the evaluation of a party determines the distance, shown in row two. Because parties closer to the voter's optimal position are considered more preferred, row three states the actual ranks of the proximity-based preference which would be (from best to worst) $B \succ D \succ A \succ C$ with B top ranked and C bottom ranked.

Given our data, independent of the general evaluation of the parties, for the 713 voters that responded to the question about political positions, it is possible to analyze how consistent their communicated rankings of the parties are w.r.t. the distance of their own ideal position compared to their perceived positions of the parties. Hence, we can simply apply a proximity-based approach in one dimension. Obviously, such a one-dimensional approach might be too scarce to really identify the main reasons for political preferences, because other aspects might influence a voter's preference (Grofman, 2019a). Those could be the personality of the main candidate, certain activities or scandals in the recent past, strategic or coalitional considerations, or plenty of other reasons. Moreover, even if the voters think in one dimension, the left-right political spectrum might not be the relevant dimension on which their voting decisions are based.

Nonetheless, the goal of this paper is to analyze to what extent voters' preferences are consistent with preferences based on distances from a voter's optimal position on the left-right scale to the position of the parties, i.e., proximity-based preferences.[5] In contrast to the stated ranking in the survey, proximity-based preferences can of course contain indifferences. In our analysis, we will keep those indifferences, although other approaches would be possible. Now, if one builds preferences in such a way, we are able to determine the consistency between the stated preferences, i.e., supposedly the voters' true rankings, and the proximity-based preferences based on the stated evaluations along the left-right scale. Of course, different ways of comparing consistency are possible. A simple option is based on the positions of the

[5] As discussed before, this alludes to single-peaked preferences. However, because there is no real consensus about the parties positions on the left-right scale, we can only determine individual proximity-based single-peaked preferences.

Table 5 Overlap between proximity-based rankings and actual rankings

	Less than 33% (%)	Between 33 and 66% (%)	More than 66% (%)
Positional overlap	52.09	38.38	9.52
Pairwise overlap	12.27	31.75	55.99

Table 6 Overlap of more than 66% by the respective party supporters

	SPÖ (%)	ÖVP (%)	FPÖ (%)	GREENS (%)	KPÖ (%)	NEOS (%)
Positional overlap	8.99	3.23	10.34	9.85	19.42	2.56
Pairwise overlap	52.81	37.63	44.83	65.53	65.05	48.72

parties in the two rankings. In Table 5, the first row states the positional overlap, i.e., the percentages of voters that have a certain number of parties ranked in exactly the same position in both, their stated and proximity-based preferences.

As can be seen in Table 5, less than 10% of the participants had an overlap of at least five positions when comparing their stated preferences with their proximity-based preferences.[6] The vast majority did substantially deviate from those proximity-based preferences. One problem in looking at exact positions obviously comes from indifferences occurring in proximity-based preferences, another from the fact that a low overlap in ranks can occur even if most of the pairwise comparisons would be proximity consistent. For example, assume a set of four candidates $\{A, B, C, D\}$ and the two rankings (from best to worst) $A \succ_1 B \succ_1 C \succ_1 D$ and $B \succ_2 C \succ_2 D \succ_2 A$. In this situation, there is zero overlap in terms of positions; however, in a pairwise comparison three of the six pairwise comparisons are consistent. The second row in Table 5 shows that, if consistency is measured in terms of pairwise overlaps, the percentages of voters ranking at least two-thirds of the pairs of parties exactly the same increase massively to 56%.

However, the numbers still indicate that one has to be careful when using a strictly one-dimensional approach (or considering the left-right scale being the one dimension). As discussed before, many other factors can influence voters' preferences.

Interestingly, the consistency of preferences w.r.t. distances on the left-right scale differs considerably among supporters of different parties. For example, the percentages of supporters that are consistent in more than two-thirds of the positions in their declared rankings vary between 2.5% (for the NEOS) and 19% (for the KPÖ). With respect to pairwise overlap, two-thirds consistency ranges from 37% (for the ÖVP) to 65% (for the Greens). The detailed numbers can be seen in Table 6.

This section provided some insight into the consistency between stated and proximity-based preferences. In the following section, we want to investigate to

[6] Actually, less than 3% ranked the parties exactly according to the parties' distances to their optimal position on the left-right scale.

what extent a purely proximity-based behavior by the voters might change voting outcomes. Therefore, we provide a comparison of the hypothetical election results for various voting rules based on the declared preferences and the results based on proximity-based preferences. Although we have previously criticized the use of proximity-based preferences in terms of leaving out various relevant factors in a voter's formation of preferences, we still consider such an exercise to provide some interesting insight. In particular, it shows the actual impact of other considerations, such as a party's main candidate or preferences over potential coalitions, on election outcomes. Because of lack of specific data, a more distinguished analysis of the impact of the individual factors can, however, not be provided.

5 Comparing (Hypothetical) Election Outcomes

We intentionally decided in the exit poll to not communicate particular voting rules to the participants before they stated their preferences. Hence, we assume all preference information and the information about the political orientation to be sincere. Obviously, because the voting rules were unknown to the participants, there was also no incentive to behave strategically in stating the preferences.[7] Of course, as has been shown in many empirical papers (see Stephenson et al. (2018) for a detailed overview) and can be seen in our data when comparing the voters' stated actual vote with their declared preferences, strategic behavior does indeed occur in real elections. Specific voting rules have not been communicated to the participants for essentially two reasons: First, most alternative voting rules would probably be too complex to be understood on the spot. Second, given that there was no real election based on such an alternative voting rules, the participants would most likely not have provided more than a sincere vote anyway, because any deeper strategic considerations would not have had any real-world consequences. Hence, when comparing the impact of different voting rules on the outcome, we will use exclusively the weighted (and sincere) preference and political orientation data from the survey. Moreover, the current election was a parliamentary election. Therefore, because it is not clear how to distribute seats based on any election outcome using more information than just top-ranked alternatives, we will stick to "social rankings" of the parties only. In addition, the major goal of the parties was not purely to receive votes but to ensure that seats in the parliament are gained, something which was of course also known to the voters.[8]

As explained in the previous section, we eventually had two different preference profiles. One based on the voters' announced preference rankings and one determined

[7] One could of course imagine that voters stated their preferences with the actually used voting rule in the election in mind. However, because the election used plurality rule, where only one (top-ranked) party can be announced, it seems rather unlikely that this had a significant impact on how to state a full preference ranking of the six parties.

[8] Given the current election rules in the region, which are based on getting at least a base mandate in one of the sub-regions, this had obvious consequences on the parties' campaigning strategies.

by the distances between a voter's self-evaluation along the left-right scale and her evaluation of the parties. Using those two preference profiles, we will, in the rest of this section, compare, for various well-known voting rules, the hypothetical election outcomes given as rankings of the parties.

In general, the voting rules can, to a certain extent, be classified in terms of the preference information used. Probably, the simplest rules are those that only use information about the top (or bottom)-ranked parties, in certain cases in a sequential manner. As a second class, we identify rules which require information about the full preference ranking, and the third category uses preference information beyond ordinal rankings, something which could be more or less difficult to provide.

5.1 Plurality Rule

Plurality rule is a widely applied voting rule for political elections in many countries. Each voter can vote for exactly one party (or candidate). The overall ranking of the parties is then determined by the number of received votes. If we assume sincere voting for the voters' top-ranked party based on their stated preference rankings on the one hand, and voting for the closest party to their self-evaluation on the other hand, the weighted vote shares are given in Table 7.

Although the votes have been weighted accordingly, the results based on both, the sincere declared preferences and the proximity-based preferences, differ considerably from the actual votes obtained by the parties at the corresponding polling stations (compare to Table 1). In terms of social rankings, the actual winner (ÖVP) is overtaken by the Greens based on declared preferences, whereas the SPÖ is winning under proximity-based preferences with the ÖVP moving down to fourth position. This means that in the real election, a significant fraction of the voters did not vote for their most-preferred party and/or their ideologically closest party. Hence, they either voted strategically, based on other criteria, or both.

That the SPÖ is the winner in a proximity-based model is not really surprising given the valuations in Table 2. The average position of the SPÖ along the left-right scale is very similar to the average position of the voters themselves. Hence, it is very likely that, on average, the SPÖ is the closest party to the voters' self-evaluations.

Table 7 Results—plurality rule

	Declared preferences (%)	Proximity-based preferences (%)
GREENS	27.60	20.42
ÖVP	25.84	16.43
SPÖ	16.00	21.19
FPÖ	11.50	7.67
KPÖ	10.65	15.69
NEOS	8.41	18.77

Table 8 Results—anti-plurality rule

	Declared preferences (%)	Proximity-based preferences (%)
ÖVP	2.20	0.93
SPÖ	4.66	2.59
NEOS	4.70	1.24
GREENS	9.79	10.19
KPÖ	11.33	14.31
FPÖ	67.32	70.74

5.2 *Anti-plurality Rule*

Whereas plurality rule allows to vote for one party, the **anti-plurality rule** takes into account the objection of voters against parties, by giving them the possibility to vote against the party they dislike most. The social ranking of the parties is then determined by the number of received (negative) votes in ascending order, i.e., the winner is the party with the lowest number of votes against. For proximity-based preferences, the negative vote is given to the party which is furthest away from a voter's ideal position on the left-right scale. The weighted results for both situations are given in Table 8.

Anti-plurality rule makes the ÖVP the winner in both scenarios. There is a rather intuitive reason for this result. Although the ÖVP was considered rather far to the right on the political left-right spectrum by many of the participants, more than 90% (see Table 3) perceived the FPÖ to be even further to the right. Hence, there were only very few voters that either considered them the lowest ranked or furthest from their ideal point on the left-right scale. Interestingly, there was quite some strong opposition against the Greens by a considerable share of the voters (around 10%).

5.3 *Pairwise Majority Rule*

Pairwise majority rule[9] is based on pairwise comparisons of the parties. That is, when considering preferences, for each pair of parties we determine which party is more preferred than the other party by a majority of the voters. Using proximity-based preferences, we consider which party is closer to a voter's ideal position on the left-right scale. Hence, it is, in principle, based on a full ranking of the parties. The weighted pairwise tallies for both scenarios can be found in Tables 9 and 10, where the numbers indicate the percentages of (weighted) voters who prefer the row party over the column party (and bold numbers indicate majorities).

[9] Pairwise majority rule is also called Condorcet rule, due to the Marquis de Condorcet being the first to promote this rule at the end of the eighteenth century.

Table 9 Pairwise tallies—declared preferences

	SPÖ (%)	ÖVP (%)	FPÖ (%)	GREENS (%)	KPÖ (%)	NEOS (%)
SPÖ		**51.26**	**76.50**	37.63	**56.53**	49.50
ÖVP	48.74		**85.59**	43.18	**52.46**	48.71
FPÖ	23.50	14.41		22.22	26.36	21.93
GREENS	**62.37**	**56.82**	**77.78**		**68.52**	**64.61**
KPÖ	43.47	47.54	**73.64**	31.48		48.50
NEOS	**50.50**	**51.29**	**78.07**	35.39	**51.50**	

Table 10 Pairwise tallies—proximity-based preferences

	SPÖ (%)	ÖVP (%)	FPÖ (%)	GREENS (%)	KPÖ (%)	NEOS (%)
SPÖ		**62.58**	**77.65**	**57.18**	**66.07**	**54.28**
ÖVP	37.42		**88.37**	43.17	45.88	30.42
FPÖ	22.35	11.63		27.10	28.45	15.47
GREENS	42.82	**56.83**	**72.90**		**64.50**	48.86
KPÖ	33.93	**54.12**	**71.55**	35.50		44.64
NEOS	45.72	**69.58**	**84.53**	**51.14**	**55.36**	

Pairwise majority rule is, in general, considered to be a very attractive voting rule respecting major democratic principles. It has a major drawback though, namely that there could occur so called Condorcet cycles, i.e., cycling pairwise majorities, in both situations.[10] However, as observed in Tables 9 and 10, no such Condorcet cycles do exist in this election, although in the case of the declared preferences, the margins between SPÖ, ÖVP and NEOS are very small (between one and three percent of the total weighted votes). Hence, this preference profile is very "close" to containing such a cycle.

The rankings based on pairwise majority rule are given in Table 11.

Again, the results do change considerably whenever proximity-based preferences are assumed. Whereas the Greens are winning based on declared preferences, it is the SPÖ which beats every other party based on distances from the voters' self-evaluation. In addition, the NEOS, who are ranked last under the plurality rule, do benefit from such pairwise comparisons and turn out to be second ranked in both scenarios. However, as displayed in Table 9, the winning margins in three of their contests (against SPÖ, ÖVP and KPÖ) based on declared votes are very small and

[10] Of course single-peaked preferences are a sufficient condition for transitive social preferences. However, because there is no general agreement among the voters concerning the order of the parties along the left-right scale, this does not apply in our case and theoretically Condorcet cycles could occur with individually proximity-based preferences.

Table 11 Results—pairwise majority rule

Declared preferences	Proximity-based preferences
GREENS	SPÖ
NEOS	NEOS
SPÖ	GREENS
ÖVP	KPÖ
KPÖ	ÖVP
FPÖ	FPÖ

Table 12 Results—Borda rule—declared preferences

	points share (%)
GREENS	21.98
ÖVP	18.57
SPÖ	18.08
NEOS	17.77
KPÖ	16.31
FPÖ	7.28

below five percent of the total weighted votes. The results in the proximity-based preference profile are in some sense more robust with the closest margin being between Greens and NEOS.

5.4 Borda Rule

The **Borda Rule** is a widely discussed aggregation method whenever full ranking information is available. It works by assigning pre-determined points to the different ranking positions, i.e., in case of k parties, $k-1$ points for every top-rank, $k-2$ points for every second-rank, down to 0 points for being bottom ranked.[11] Because the points are pre-determined, the rule is still based purely on ordinal information. Therefore, it does not take into account the perceived intensity between different parties by the voters. Based on our data, the weighted share of the total points for the parties is given in Table 12.

As shown in Darmann et al. (2017), using the full ranking as input is especially harmful to parties, which are considered *polarizing*. Those are parties which receive strong positive support from a considerable fraction of the voters but, at the same time, are strongly disapproved by a large fraction of the voters. In the Austrian political

[11] The Borda rule is a special case among the huge class of *scoring rules*. Those are rules which are based on a specific scoring vector. The scoring vector for the Borda rule, for example, is $s_B = (k-1, k-2, ..., 1, 0)$, whereas the plurality rule scoring vector is $s_P = (1, 0, 0, ..., 0)$.

Table 13 Results—Borda rule—proximity-based preferences

	Points share (%)
SPÖ	20.22
NEOS	19.65
GREENS	18.77
KPÖ	16.56
ÖVP	16.45
FPÖ	8.33

situation, it is, in particular, the FPÖ which is such a polarizing party. Compared to its result under plurality rule, it is significantly worse off under the Borda rule. On the other hand, more *medium* parties, i.e., those that do not create strong feelings in a positive or negative way, benefit from such a rule, e.g., the NEOS. The Greens would again win in a Borda election; however, the winning margin would be much smaller. But this is of no surprise, because using Borda scores implies an upper limit for the difference in the vote shares of the winning party and its runner-up. This is also one reason for the rather close contest between the ÖVP, SPÖ, KPÖ and the NEOS.

We can also apply the Borda rule on our proximity-based preference profile. From Table 2, it is obvious that the parties whose perceived mean positions are closest to the voters' own positions should benefit from the Borda rule. This is indeed the case as can be seen in Table 13 which provides the shares of the total Borda scores for the different parties in the proximity-based preference profile.

Again, besides for the bottom-ranked FPÖ, the rankings of all the other parties do change considerably w.r.t. the actual election outcome and the outcome based on declared preferences.

5.5 *Approval Voting*

An interesting extension of plurality rule is **Approval Voting** Brams & Fishburn (1983), which allows voters to not only vote for exactly one party, but to approve of as many parties as they like. Each of them will receive one point, and the ranking of the parties is determined by summing up the points over all voters. Obviously, approval preferences cannot be determined purely from ordinal preference information, but preferences have, in principle, to be dichotomous.[12] According to the declared approval preferences in the exit poll, the participants, on average, approved of 2.12 different parties, the median number of approvals (as well as the mode) being 2. Interestingly, only about 25% of the participants approved of just one party. To some extent, this indicates one of the problems of plurality rule, namely that many

[12] In some sense, this means both more and less information than ordinal rankings.

Table 14 Results—approval voting

	Approvals (%)
GREENS	55.37
NEOS	35.01
KPÖ	34.26
ÖVP	33.36
SPÖ	29.82
FPÖ	17.87

voters consider more than one party acceptable. The weighted results under approval voting, given the announced approvals, are presented in Table 14.

Again, the Greens get the largest number of approvals, whereas the FPÖ receives by far the lowest approval and can therefore, from a dichotomous point of view, be seen as disapproved by most of the voters. The other four parties have very similar numbers of approvals. One interesting fact is that the NEOS and the KPÖ, both being considered to be rather medium parties, do receive substantial approval although voters do not vote for them if only one vote is possible. Hence, those parties benefit substantially in case more information (in the form of approvals) can be communicated by the voters.

Approval preferences can also be determined using a proximity-based approach. In principle, if voters strictly operate according to the one-dimensional political spectrum, the assumption could be that they approve of all parties which lie within a certain distance around their ideal points. Of course, how large this distance should be is not so obvious. If one considers the average number of approvals (2.12) as some kind of key value, then this number would be achieved by using a distance somewhere between two and three. Table 15 shows the parties' percentages of approvals for various distances, ranging from one to six. For example, "dis (3)" indicates the parties' percentages, whenever all parties are approved by the voters that lie within a distance of three from the voters' ideal points. Interestingly, the results, compared to the declared approvals, change significantly for the SPÖ. It moves from fifth position to become the clear winner for all considered distances. Again, this seems to be a

Table 15 Results—approval voting—proximity-based approvals

	Dis (1) (%)	Dis (2) (%)	Dis (3) (%)	Dis (4) (%)	Dis (5) (%)	Dis (6) (%)
SPÖ	27.51	46.89	60.81	72.81	85.88	87.94
GREENS	22.28	40.63	52.93	59.98	74.35	81.55
NEOS	17.67	37.90	54.17	65.95	77.27	80.43
KPÖ	15.27	33.12	50.12	58.28	73.19	78.53
ÖVP	12.28	26.18	38.33	49.43	63.26	74.37
FPÖ	2.38	9.43	16.71	26.82	44.30	54.65

clear indication that the actual preferences (especially concerning the SPÖ) are highly influenced by other factors (and those factors have significant impact). The rest of the social order remains unchanged for most of the proximity-based approval cases.

6 Conclusion

In this paper, based on data from an exit poll during the 2019 Styrian Parliamentary Elections, we compare actual preferences with proximity-based preferences determined from voters' perceptions of the parties and themselves in the political left-right spectrum. First it is shown that voters do not fully agree on an underlying order of the parties along the left-right scale. In addition, voters' preferences can not be seen as purely proximity-based, i.e., actual preferences do substantially deviate from preferences based on distances from a voter's ideal point to the perceived locations of the parties in a one-dimensional political left-right spectrum. This could have various reasons: Perhaps voters consider other political dimensions, on which they build their actual preferences, more important. Or their proximity-based preferences are just not single peaked. Also strategic considerations or specific views about possible coalitions could play an important role. To determine the consequences of such other factors which influence the determination of voters' preferences, we then compare hypothetical election outcomes from the two different preference profiles for various well-known voting rules, namely the plurality rule, the anti-plurality rule, the pairwise majority rule, the Borda rule and approval voting. Significant differences in the outcomes do occur. The proximity-based approach does, in particular, favor parties which, on average, are considered closer to the average self-evaluation than other parties. A summary of the hypothetical election results is given in Table 16.

To conclude, the results show that one-dimensional approaches, be it single-peaked models or just proximity-based models, have to be taken with some care. First,

Table 16 Outcomes declared preferences versus proximity-based (PB) preferences

Voting rule	1st	2nd	3rd	4th	5th	6th
Official result	ÖVP	GREENS	SPÖ	FPÖ	KPÖ	NEOS
Plurality rule	GREENS	ÖVP	SPÖ	FPÖ	KPÖ	NEOS
PB plurality rule	SPÖ	GREENS	NEOS	ÖVP	KPÖ	FPÖ
Anti-plurality rule	ÖVP	SPÖ	NEOS	GREENS	KPÖ	FPÖ
PB anti-plurality rule	ÖVP	NEOS	SPÖ	GREENS	KPÖ	FPÖ
Pairwise majority	GREENS	NEOS	SPÖ	ÖVP	KPÖ	FPÖ
PB pairwise majority	SPÖ	NEOS	GREENS	KPÖ	ÖVP	FPÖ
Borda rule	GREENS	ÖVP	SPÖ	NEOS	KPÖ	FPÖ
PB Borda rule	SPÖ	NEOS	GREENS	KPÖ	ÖVP	FPÖ
Approval voting	GREENS	NEOS	KPÖ	ÖVP	SPÖ	FPÖ
PB approval voting (dis 2)	SPÖ	GREENS	NEOS	KPÖ	ÖVP	FPÖ

in voting situations, a voter's behavior can probably be approximated by proximity-based preferences only to some extent. Second, voting rules are obviously sensitive to differences in preference profiles whenever individual preferences do significantly change in a proximity-based approach. Of course, it would be interesting to investigate in much more detail what are the key factors that lead to such differences, but this would require different data and is therefore left for future research.

Acknowledgements I am grateful to three anonymous reviewers for their helpful comments. This research has been supported by the University of Graz research project "Wahl-Auswahl-Entscheidung".

Appendix: Supplementary Material

UNIVERSITÄT GRAZ
UNIVERSITY OF GRAZ
Sozial- und Wirtschaftswissenschaftliche Fakultät

UNI GRAZ

Welcome to the

Survey for the Styrian Parliamentary Election 2019 on 24 Nov 2019

By participating and conscientiously answering the questions below, you are contributing to a project on alternative electoral procedures and agree to the evaluation of your information. Your information will be treated confidentially and analysed ANONYMOUSLY.

Question 1: *Suppose you had been able to use your vote in the election not only to choose one party, but to cast as many votes as you like. Which parties would have received your votes? (Please tick)*

☐Greens ☐ÖVP ☐FPÖ

☐KPÖ ☐NEOS ☐SPÖ

Question 2: *Suppose you had to rank ALL parties, starting with the best party for you and ending with the worst party. What would this ranking look like from position 1 to position 6) (Note: 1 is the best party for you and 6 is the worst party for you).*

SPÖ		KPÖ		Greens	
NEOS		ÖVP		FPÖ	

Question 3: *Is it easy for you to rank ALL six parties as you did in question 2?*

☐easy ☐rather easy ☐rather difficult ☐difficult ☐no answer

Question 4: *In political questions, people usually talk about left and right. How would you classify the parties and yourself (=self-assessment)? (Please tick the appropriate box).*

	LEFT					MIDDLE					RIGHT
NEOS	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
FPÖ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Greens	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
KPÖ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
SPÖ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
ÖVP	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
self-assessment	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Question 5: *Suppose instead of one vote you could give points to the parties in the election, from -20 (very bad) to +20 (very good). What would your score be? (Note: You can also assign the same number to several parties).*

FPÖ			NEOS			ÖVP	
SPÖ			Greens			KPÖ	

Question 6: *Suppose you had to rate the parties according to your acceptance. Please tick your rating for each party, where "+" means that the party is acceptable to you, "-" means that the party is unacceptable to you and "o" means that you consider the party neutral.*

	+	o	-		+	o	-		+	o	-
ÖVP	☐	☐	☐	FPÖ	☐	☐	☐	KPÖ	☐	☐	☐
Greens	☐	☐	☐	SPÖ	☐	☐	☐	NEOS	☐	☐	☐

Question 7: *Would you prefer an electoral system that uses information like in questions 1,2,5 or 6 instead of the current electoral system?*

☐Yes ☐No ☐no answer

STATISTICS

Question a: *Which party did you vote for in today's election*

☐KPÖ ☐SPÖ ☐NEOS

☐ÖVP ☐FPÖ ☐Greens

Question b: *Please indicate your gender:*

☐male ☐female ☐other ☐no answer

Question c: *Please state your age:*

☐16 to 19 ☐20 to 29 ☐30 to 39 ☐40 to 49

☐50 to 59 ☐60 to 69 ☐70 years or older ☐no answer

Question d: *What is your highest educational level?*

☐compulsory school ☐compulsory school with apprenticeship

☐middle school (BMS/technical school) ☐AHS/BHS without Matura (A-level)

☐Matura (A-level) ☐university/university of applied sciences

☐no answer

Thank you very much for your time!

References

Alòs-Ferrer, C., & Granič, D. (2012). Two field experiments on approval voting in Germany. *Social Choice and Welfare*, *39*(1), 171–205

Arrow, K. J. (1963). *Social choice and individual values*. Wiley.

Baujard, A., Gavrel, F., Igersheim, H., Laslier, J. F., & Lebon, I. (2014). Who's favored by evaluative voting? An experiment conducted during the 2012 French presidential election. *Electoral Studies, 34*, 131–145.

Baujard, A., Gavrel, F., Igersheim, H., Laslier, J. F., & Lebon, I. (2018). How voters use grade scales in evaluative voting. *European Journal of Political Economy, 55*, 14–28.

Black, D. (1958). *The theory of committees and elections*. Cambridge University Press.

Blais, A., & Degan, A. (2019). The study of strategic voting. In R.D. Congleton, B. Grofman, & S. Voigt (Eds.), *The Oxford handbook of public choice* (vol. 1, ch. 14, pp. 292–309). Oxford University Press.

Brams, S. J., & Fishburn, P. C. (1983). *Approval voting*. Birkhauser.

Cox, G. W. (1997). *Making votes count: Strategic coordination in the world's electoral systems*. Cambridge University Press.

Darmann, A., Grundner, J., & Klamler, C. (2017). Election outcomes under different ways to announce preferences: an analysis of the 2015 parliament election in the Austrian federal state of Styria. *Public Choice, 173*(1), 201–216.

Darmann, A., Grundner, J., & Klamler, C. (2019). Evaluative voting or classical voting rules: Does it make a difference? empirical evidence for consensus among voting rules. *European Journal of Political Economy, 59*, 345–353.

Enelow, J. M., & Hinich, M. J. (1984). *The spatial theory of voting: An introduction*. Cambridge University Press.

Enelow, J. M., & Hinich, M. J. (Eds.) (2008). *Advances in the spatial theory of voting*. Cambridge University Press

Feld, S. L., & Grofman, B. (1986). Partial single-peakedness: An extension and clarification. *Public Choice, 51*, 71–80.

Gaertner, W. (2001). *Domain conditions in social choice theory*. Cambridge University Press.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica, 41*(4), 587–601.

Grofman, B. (2019a). Political choices in one dimension: Theory. In R. D. Congleton, B. Grofman, & S. Voigt (Eds.), *The Oxford handbook of public choice* (Ch. 8, pp. 147–166). Oxford University Press.

Grofman, B. (2019b). Political choices in one dimension: Applications. In R. D. Congleton, B. Grofman, & S. Voigt (Eds.), *The Oxford handbook of public choice* (Ch. 9, pp. 167–186). Oxford University Press.

Kawai, K., & Watanabe, Y. (2013). Inferring strategic voting. *American Economic Review, 103*, 624–662.

Monjardet, B. (2009). Acyclic domains of linear orders: A survey. In S. J. Brams, W. V. Gehrlein, & F. S. Roberts (Eds.), *The Mathematics of preference, choice and order. Studies in choice and welfare* (pp. 139–160). Springer.

Regenwetter, M., & Tsetlin, I. (2004). Approval voting and positional voting methods: Inference, relationship, examples. *Social Choice and Welfare, 22*(3), 539–566.

Regenwetter, M., Grofman, B., Marley, A., & Tsetlin, I. (2006). *Behavioral social choice*. Cambridge University Press.

Roescu, A. (2014). Preferences, voting rules, behaviour and outcomes: A field experiment on the local elections in Romania. *The Romanian Journal of Society and Politics, 9*(1), 7–29.

Saari, D. G. (1995). *Basic geometry of voting*. Springer.

Satterthwaite, M. (1975). 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.

Schofield, N. (2019). Spatial social choice. In R. D. Congleton, B. Grofman, & S. Voigt (Eds.), *The Oxford handbook of public choice* (Ch. 11, pp. 208–244). Oxford University Press

Sen, A. (1979). The impossibility of a Paretian liberal. *Journal of Political Economy, 78*(1), 152–157.

Stephenson, L. B., Aldrich, J. H., & Blais, A. (Eds.) (2018). *The many faces of strategic voting*. University of Michigan Press.

Participation in Voting Over Budget Allocations: A Field Experiment

Clemens Puppe and Jana Rollmann

Abstract We study the effect on the participation rate of employing different voting rules in the context of the problem to allocate a fixed monetary budget to two different public projects. Specifically, we compare the mean rule according to which the average of the individually proposed allocations is implemented with the median rule which chooses the allocation proposed by the median voter as the social outcome. We report the results of a field experiment in which subjects (students of KIT) could allocate money to fund two different public projects, the student's bike shop and a campus garden project. The treatment variable was the collective decision rule employed. While the mean and median rules have very different properties in theory, we found no significant treatment effect on the participation rate. Our results nevertheless shed important light on the use of different voting rules in the context of budget allocation in practice.

1 Introduction

The problem of participation in elections is mostly either discussed in the context of large democratic elections (Downs, 1957; Riker & Ordeshook, 1968; Tullock, 1967), and/or assuming the election procedure to be simple majority voting among two alternatives, e.g. political candidates or parties (Ledyard, 1984; Palfrey & Rosenthal, 1983, 1985). By contrast, the present paper focuses on voters' participation decisions in (small) committees and under further assumptions about preferences. The main

C. Puppe (✉) · J. Rollmann
Department of Economics and Management, Karlsruhe Institute of Technology, Karlsruhe, Germany
e-mail: clemens.puppe@kit.edu; clemens.puppe@balliol.ox.ac.uk

J. Rollmann
e-mail: jana.rollmann@kit.edu

C. Puppe
Higher School of Economics (HSE), Perm Campus, Moscow, Russian Federation

Balliol College, Oxford University, Oxford, UK

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_21

question addressed in the present paper is if the choice of the specific voting rule applied has an influence on the participation rates.

Specifically, we consider the collective decision on the level of a one-dimensional variable (spending on a public project) under the assumption of single-peaked preferences. This problem has been theoretically studied in Müller and Puppe (2020), Müller (2022), Osborne et al. (2000). These contributions show that the general game-theoretic analysis of the corresponding participation/voting games is complex. In particular, even under the (strong) assumption of complete information, the existence of pure Nash equilibria is not guaranteed, and in many cases in which existence can be proved the equilibria are not unique. The situation is even more intricate under incomplete information in which case there is little hope to successfully apply standard Bayesian analysis in order to arrive at behavioural predictions in real-life situations. Indeed, in contrast to the case of two alternatives/candidates in which voters' beliefs can be parametrized by a single number (e.g. the proportion of the supporters of the first candidate), the relevant beliefs are high-dimensional when there is a large number of alternatives as in our context.

We therefore take a 'bounded rationality' approach in the present paper. Concretely, we hypothesize that a voter's participation decision is to a large part influenced by the *potential impact* that her or his vote can have on the outcome resulting from the collective decision procedure. In order to test this hypothesis, we conducted a field experiment in which subjects could decide on the allocation of a fixed budget using two different aggregation rules. The first is the simple *mean rule* according to which the collective outcome is the average of the individual votes (here, a 'vote' is simply the amount of money allocated to one public project) (Renault & Trannoy, 2005). The second is the *median rule* that picks the median vote as the social outcome. Under the mean rule, the potential impact of a vote, i.e. the extent to which one's vote can change the social outcome, only depends on the number of other voters and not on the distribution of their votes. On the other hand, voting truthfully under the mean rule is clearly not optimal in general (because a rational voter would want to 'exaggerate' in expressing her or his preference whenever the true most preferred does not coincide with the social outcome). By contrast, the great advantage of the median rule is that sincere voting is a weakly dominant strategy (if the number of individuals is odd). Therefore, it has been widely employed and studied in the literature following the seminal work by Black (1948), Downs (1957), Moulin (1980). On the other hand, the potential impact under the median rule is uncertain or, more precisely, ambiguous. First, only one subject can change the outcome at all in the case of an odd number of participants (and maximally two subjects in the case of an even number of participants). Moreover, the extent to which pivotal median voter(s) can potentially change the outcome depends on the neighbouring votes: if these are close, the potential impact is small; see Kurz et al. (2017, p. 1607f) for a lucid and detailed further discussion. Summarizing then, the impact of one's vote is easy to understand and to assess under the mean rule, but it is uncertain and depends on the precise distribution of the votes of the other participants under the median rule.

Our main hypothesis was therefore that we would find a higher participation rate in our field experiment under the mean rule as compared to the median rule.

While we do indeed observe effects in this direction, a closer examination reveals that the difference between the two rules is statistically not significant. We tested a number of other hypotheses as well and again found no significant treatment effects (with respect to the aggregation rule as main treatment variable). While our main hypotheses are not supported by our data, the results from our field experiment nevertheless have interesting and important implications. First, real subjects seem to be motivated by a number of factors different from pivotality in the voting process. Secondly, strategic behaviour is much less prevalent in the field than one would expect by looking at corresponding data from laboratory experiments (Block, 2014; Marchese & Montefiori, 2011; Louis et al., 2022; Puppe & Rollmann, 2021).

The plan of the paper is as follows. After giving more background and an overview of the literature, we present a theoretical model of the role of the potential vote impact in Sect. 2. The field experiment and the hypotheses are described in Sect. 3 and the results in Sect. 4. Section 5 concludes. An appendix contains further information, including the instructions, screenshots, and the questionnaire that we used during the experiment.

1.1 Background and Overview of the Literature

With their work on rational choice theory and the calculus of voting, Downs (1957), Tullock (1967) as well as Riker and Ordeshook (1968) provide a decision-theoretic model of participation in elections. The question that these models face is why a rational individual would vote if the return from voting is often outweighed by the costs that emerge in the voting process. Even if the cost to participate in an election is small, the probability that a single vote affects the outcome is almost zero in large electorates. As an example, Gelman et al. (1998) estimate the ex post probability of a single vote being decisive in the 1992 US presidential election to be 1 in 10 million. Therefore, the rational choice model predicts turnout levels that are far below the actual participation rates in elections. This discrepancy is often referred to as the *paradox of voting*.

The calculus of voting has been tested empirically in a variety of studies in the 70s and 80s. Aldrich (1993) have pointed out that using aggregate data like Barzel and Silberberg (1973), Settle and Abrams (1976), as well as (Silberman and Durden, 1975) do, yields a correlation between pivotality and turnout, while survey data as in the studies by Ferejohn and Fiorina (1975) or Foster (1984) do not. Enos and Fowler (2014) review some 70 articles on voter turnout and the relation between pivotality and participation rates. They find that in a majority of studies pivotality is an important driving force for turnout and that most models on turnout in fact focus mainly on pivotality. In their own study, the authors describe a rare circumstance of an exact tie in electing a candidate for the Massachusetts State House in 2010, which led to a re-election and thus gave a unique opportunity for a field experiment to measure the effect of pivotality on turnout. They informed subjects about the closeness of the election but found a significant increase in turnout only for a subgroup of frequent voters. Thus, Enos and Fowler (2014) concluded that pivotality is not as relevant for turnout decisions as the 'calculus of voting' would predict.

Ferejohn and Fiorina (1974, 1975) argue that subjects often do not act as expected utility maximizers since the relevant probabilities of the model are unknown. Instead of basing their participation decisions upon pivotality, subjects are assumed to follow a strategy based on minimax regret. These authors argue that the minimax regret approach leads to a higher participation rates than the pivotality approach.

Palfrey and Rosenthal (1983, 1985) and Ledyard (1984) formulate the pivotal voter model using a game-theoretic approach. In the Palfrey-Rosenthal participation game, two groups of subjects prefer either one or another candidate. Each subject may vote for his or her preferred candidate (voting for the opponent is strictly dominated in the two-candidate case) or abstain. Participation is costly, while abstention is free. The candidate that gets the majority of votes wins. In their equilibrium analysis, Palfrey and Rosenthal (1983) show that there not only exist equilibria with low turnout levels but also equilibria with substantial turnout if participants face identical costs and complete information on the distribution of preferences. Ledyard (1984) endogenizes pivotality and highlights that the participation decision of all subjects is made simultaneously. His model implements uncertainty about preferences as well as costs, and turnout levels lie between zero and full participation in equilibrium. Building on Ledyard (1984), Palfrey and Rosenthal (1985) implement uncertainty about the individual voting costs and show that this lack of information causes individuals to abstain even when participation would be optimal under full information. Hence, in large electorates the unique Bayesian equilibrium displays low turnout under incomplete information.

Blais (2000) surveys numerous empirical studies and provides a review on rational choice models. He concludes that the rational choice model has limited explanatory power in order to explain empirical turnout rates. Dhillon and Peralta (2002) provide a complementary survey on the existing models and theoretical literature on participation.

The Palfrey-Rosenthal participation game is used widely in the literature that tests the pivotal voter model in experimental studies. A laboratory experiment conducted by Levine and Palfrey (2007) tests the voter turnout predictions of the Palfrey-Rosenthal model with asymmetric information, in which participation costs are private information. The authors find a 'size effect', meaning that in large elections, participation rates are lower as compared to small electorates. The data also reveal a 'competition effect', i.e. elections that are expected to be close are associated with a higher voter turnout. Another finding is the so-called underdog effect: groups that support a less popular alternative have higher turnout rates as compared to the supporters of the popular alternative.

Duffy and Tavits (2008) perform a laboratory experiment of the complete information pivotal voter model and additionally elicit the subjects' beliefs about the probability of a close election. Therefore, the authors are able to directly test the pivotal voter model and focus on the correlation of beliefs about being pivotal and the participation decision. The study finds that a higher belief about the probability of being pivotal increases the likelihood of participation and that subjects tend to overestimate this probability of being pivotal. In another laboratory experiment, Agranov et al. (2017) test the effect of pre-election information on voter turnout. The study finds that pre-election polls influence participation, and the effects depend on the expectation on the closeness of the election. When a poll reveals that the election is expected to be close bandwagon effects appear (higher voter turnout among the majority), whereas when landslide victories are predicted, the authors find underdog effects. The authors also find that landslide elections occur more often in treatments with more information and that voter turnout is higher the more likely subjects expect the preferred alternative to win.

Grillo (2017) presents a game-theoretic model that takes risk aversion into account and by doing so is able to explain bandwagon effects in cases in which the standard pivotal voter model would otherwise predict an underdog effect. Blais et al. (2014) study the rational choice model in the laboratory in the context of participation in elections and compare different voting rules. Remarkably, they find that a large share of subjects (62%) make the 'wrong' decision, i.e. they vote when they should have abstained and vice versa. Even when controlling for beliefs of the opponents' behaviour, the rational choice model fails to explain the decision of voting and abstaining, as subjects do not appear to maximize their payoff.

Börgers (2004) develops a costly voting model assuming that costs are private information. He compares compulsory to voluntary voting and finds that compulsory voting is Pareto-dominated by voluntary voting. In a related model, Krasa and Polborn (2009) find that paying subsidies to participants can prevent 'wrong' electoral decisions and increase social welfare by increasing the electorate. Another extension of the costly voting model is provided by Arzumanyan and Polborn (2017) who consider plurality rule among more than two candidates. The interesting new aspect is that strategic voting becomes possible, i.e. voting but not for the own top candidate (something that is never optimal in the two-candidate setting). However, the authors find that for three candidates all equilibria exhibit only sincere voting, a finding that hinges crucially on the fact that voting is costly.

Voting over the level of a one-dimensional variable without participation costs has been investigated in Block (2014), Louis (2022), Marchese and Montefiori (2011), and Rollmann (2020); the latter contribution theoretically and experimentally studies general 'trimmed means', of which the mean and median rules are both special cases.

2 Theoretical Framework

Consider a set of individuals $I = \{1, \dots, n\}$ that have to collectively decide on the allocation of a fixed budget $Q \geq 0$ on m public projects $J = \{1, \dots, m\}$. We assume that the entire budget has to be spent and that no project can receive negative funding. The set of feasible allocations is thus given by

$$\mathcal{B} := \{x \in \mathbb{R}^m_{\geq 0} | \sum_{j \in J} x^j = Q\},$$

where x^j is the amount of money allocated to project j.

2.1 Aggregation Rules

Individuals decide whether or not to participate in the voting process; formally, each individual i faces a participation decision $\vartheta_i \in \{0, 1\}$ that takes the value 1 for participation and 0 in case of abstention. An individual that decides to participate, i.e. $\vartheta_i = 1$, submits a vote that is taken into account in the calculation of the social outcome (such an individual is referred to as 'participant' in the following). The set of participants is denoted by $I^{-*} := \{i \in I | \vartheta_i = 1\}$ and the number of participants by $k = |I^{-*}|$. Throughout, we assume $k > 0$. Each participant $i \in I^{-*}$ submits a vote $q_i = (q_i^1, \ldots, q_i^m) \in \mathcal{B}$. We refer to the vote of individual i also as voter i's *proposal*. The vector of all proposals/votes is given by $q = (q_1, \ldots, q_k) = (q_i)_{i \in I^{-*}}$ and will be accounted for in the aggregation process to determine the social outcome. If an individual abstains, i.e. $\vartheta_i = 0$, no vote is submitted and we define $q_i = *$. We define the set of abstainers by $A := \{i \in I | q_i = *\}$.

In our setting, the social outcome $x(q) = \big(x^1(q), \ldots, x^m(q)\big) \in \mathcal{B}$ is calculated either by the mean or by the median rule. Under the *mean rule*, all votes are added separately for each project and divided by the number of votes:

$$\text{Mean}(q) = \frac{1}{k} \sum_{i \in I^{-*}} q_i. \tag{1}$$

Note that the social outcome under the mean rule always satisfies the budget constraint, i.e. $\text{Mean}(q) \in \mathcal{B}$.

The median rule selects, for every project, the middle proposal if the number of participants is odd or the average of the two middle votes if it is even. Specifically, if we denote for each project j, by $q^j_{[1]}, \ldots, q^j_{[k]}$ the individual votes in ascending order, the *median rule* $\text{Med}(q)$ is defined by the m coordinate-by-coordinate median values:

$$\text{Med}^j(q) = \begin{cases} q^j_{[\frac{k+1}{2}]}, \text{ if } k \text{ is odd} \\ \frac{1}{2} \cdot (q^j_{[\frac{k}{2}]} + q^j_{[\frac{k}{2}+1]}), \text{ if } k \text{ is even.} \end{cases} \tag{2}$$

A problem of the median rule is that the coordinate-by-coordinate median values do not satisfy the total budget in multidimensional allocation problems in general even when the individual proposals do, i.e. it is well possible that $\sum_{j=1}^m \text{Med}^j(q) \neq Q$ if $m > 2$. There are several ways to respond to this problem in general; see Lindner (2011). For our purposes, however, this does not pose a difficulty because we will assume throughout that $m = 2$, i.e. that there are only two different public projects. As is easily verified, the median rule always satisfies the budget constraint in this case.

In fact, with two public projects, due to the budget restriction, it is sufficient to indicate both the vote or the social outcome only for one project—the value for the other project then follows directly from the budget equality: $x^1(q) + x^2(q) = Q$.

We will therefore in the following omit superscripts referring to projects with the convention that the given value refers to the amount allocated to project $m = 1$, the value for project $m = 2$ is then given by the difference to the total budget Q.

2.2 Preferences

We assume that participation in the voting process is costly. Each participant faces a cost $c_i > 0$ if $\vartheta_i = 1$. By contrast, abstention ($\vartheta_i = 0$) is costless. Individuals are assumed to have single-peaked preferences over outcomes. More specifically, we assume *symmetric* single-peaked preferences (sometimes referred to as 'Euclidean' preferences): there is a unique most preferred outcome p_i (voter i's *peak*), and an outcome x is preferred to y if and only if x is closer than y to p_i in the standard Euclidean distance $d(\cdot, \cdot)$. Summarizing, voter i's preference can be represented by the following utility function:

$$u_i(x(q)) = \begin{cases} -d(p_i, x(q)) - c_i, \text{ if } \vartheta_i = 1 \text{ and } k > 0, \\ -d(p_i, x(q_{-i})), \text{ if } \vartheta_i = 0 \text{ and } k > 0, \\ -\infty, \text{ if } k = 0. \end{cases} \tag{3}$$

Observe that we set individual utility to $-\infty$ if no one participates in the voting process. Alternatively, we could have set that value to a large negative number. The idea is that if the group does not reach a decision via the voting process, some external institution determines the social outcome, and that all involved individuals would prefer *any* collectively determined outcome to that exogenous outcome.

According to the pivotal voter model, an individual will participate in voting process if and only if the utility from doing so exceeds the utility of abstention, i.e. if and only if

$$-d(p_i, x(q_{-i}, q_i^*)) - c_i \geq -d(p_i, x(q_{-i})), \tag{4}$$

where q_i^* is the optimal vote of individual i given the votes q_{-i} of all other participants. In fact, the inequality (4) describes the participation decision in Nash equilibrium. Alas, the Nash equilibria of the corresponding voting game are very complex. Even under the (strong) assumption of complete information (i.e. each voter's preferences are common knowledge) and the restrictive assumption of symmetrically single-peaked ('Euclidean') preferences, there are multiple Nash equilibria for many parameter constellations; for other parameter constellations, there do not exist pure Nash equilibria at all; see Müller (2022). Under the (in many applications more realistic) assumption of incomplete information, there is no hope of deriving general results on the structure and existence of Bayesian Nash equilibria even under restrictive assumptions (e.g. equal participation costs across individuals).

Thus, instead of concentrating on solutions that assume perfectly rational individuals (with a common prior and Bayesian updating under incomplete information), we explore in the following *boundedly* rational behaviour and test various hypotheses by means of a field experiment. Concretely, we hypothesize that an important determinant of the participation decision is the potential *impact* that an individual's vote can have under the different voting rules.

2.3 Impact of a Vote

The *option set* of individual i, given the decisions of all other individuals, is defined by[1]

$$\mathcal{OS}_i(q_{-i}) = \{\beta \in \mathcal{B} \mid \exists\, q_i \in \mathcal{B}, x(q_i, q_{-i}) = \beta\}\,.$$

The option set describes the set of all possible social outcomes that an individual i can induce given the decisions q_{-i} of all others. Since both the mean and median rule are monotonic and continuous in q_i given any fixed vector q_{-i}, we obtain

$$\mathcal{OS}_i = \big[x(0, q_{-i})\,,\; x(Q, q_{-i})\big] \subseteq [0, Q] \tag{5}$$

if $x(q)$ is determined either by the mean or by the median rule.

Observe that under the mean rule, $x(0, q_{-i}) = \sum_{j \neq i} q_j / k$ where k is the number of participants including voter i, and $x(Q, q_{-i}) = (Q + \sum_{j \neq i} q_j)/k$; in particular, the outcome from individual i abstaining, i.e. $\sum_{j \neq i} q_j/(k-1)$, can also be induced by individual i also voting for the $q_i = \sum_{j \neq i} q_j/(k-1)$. (Of course, with positive participation costs, individual i would never want to cast a vote that does not change the outcome as compared to abstention.)

Similarly, under the median rule we have $x(0, q_{-i}) = q_{[\frac{k-1}{2}]}$ and $x(Q, q_{-i}) = q_{[\frac{k+1}{2}]}$ if $k-1$ is even (and the $q_{[j]}$, $j = 1, \ldots, k-1$ or the other individuals' votes in ascending order); for $k-1$ odd, we obtain

$$x(0, q_{-i}) = \frac{1}{2} \cdot \Big(q_{[\frac{k-2}{2}]} + q_{[\frac{k}{2}]}\Big) \;\text{ and }\; x(Q, q_{-i}) = \frac{1}{2} \cdot \Big(q_{[\frac{k}{2}]} + q_{[\frac{k+2}{2}]}\Big)\,.$$

We define the *potential impact* of individual i given the distribution q_{-i} of the other voters as the length of the option set and denote it by $\mathrm{imp}_i(q_{-i})$; thus,

$$\mathrm{imp}_i(q_{-i}) = d\Big(x(0, q_{-i})\,,\; x(Q, q_{-i})\Big). \tag{6}$$

When no confusion can arise, we will omit the argument and simply write imp_i.

Clearly, the potential impact of a vote of individual i in general depends both on the rule and on the distribution of the other votes. But for the mean rule, it in fact only depends on the total number of voters k. Since every vote under the mean rule

[1] We neglect the possibility that no individual participates in the election process. This is justified by the fact that the outcome from universal abstention is strictly worse than any other outcome for all individuals.

has the same weight of $\frac{1}{k}$, the higher k the smaller is a voter's option set, and hence the smaller the potential impact. Indeed, it follows at once from the calculations of $\mathrm{Mean}(0, q_{-i})$ and $\mathrm{Mean}(Q, q_{-i})$ above that under the mean rule, for all q_{-i},

$$\mathrm{imp}_i(q_{-i}) = \frac{Q}{k}. \tag{7}$$

As a simple example, take the mean rule and let $Q = 100$, $\mathrm{Mean}(q_{-i}) = 20$ and $k = 4$. Individual i's option set is $\mathcal{OS}_i = [15, 40]$, resulting in a potential impact of $\mathrm{imp}_i = 25$. With an additional voter, the option set is $\mathcal{OS}_i = [16, 36]$, which reduces the potential impact to $\mathrm{imp}_i = 20$. For $k = 100$, individual i's potential impact decreases to $\mathrm{imp}_i = 1$.

If the median rule is used, the precise location of votes matters, more precisely the positions of the votes that are ranked next to the median as described above. Denote by $q_{\mathrm{Med}-}$ the vote that is ranked one position left of the median and by $q_{\mathrm{Med}+}$ the vote ranked one position right of the median. Consider the following example: $Q = 100$, $\mathrm{Med}(q_{-i}) = 44$, $q_{\mathrm{Med}-} = 18$, $q_{\mathrm{Med}+} = 70$ and $(k-1)$ is even. Including individual i, there is an odd number of voters k, and thus, the median outcome is the vote that is ranked at the middle position $[\frac{k+1}{2}]$. For any $q_i \leq 18$, the social outcome is $\mathrm{Med}(q_i) = 18$, and for any $q_i \geq 70$, the social outcome is $\mathrm{Med}(q_i) = 70$. Therefore, $\mathrm{imp}_i = 52$. Take the same example but now let $(k-1)$ be odd, which means that there exists a median voter who votes for $q_{\mathrm{Med}} = 44$. For any $q_i \leq 18$, the social outcome is $\mathrm{Med}(q_i) = \frac{1}{2} \cdot \big(q_{\mathrm{Med}-} + \mathrm{Med}(q_{-i})\big) = 31$. For any $q_i \geq 70$, $\mathrm{Med}(q_i) = \frac{1}{2} \cdot \big(\mathrm{Med}(q_{-i}) + q_{\mathrm{Med}+}\big) = 57$, resulting in $\mathrm{imp}_i = 26$ or half the size of the potential impact in the even case.

Importantly, we hypothesize that the potential impact that an individual enjoys has a positive effect on this individual's participation decision. According to this hypothesis, a higher potential impact would thus induce a greater participation probability. However, as we have just seen, the potential impact generally depends on the distribution of the other participants' votes. This implies that, under incomplete information, the participation probability is affected by an individual's beliefs about the vote distribution of the other participants. How should we model this uncertainty? We will now briefly discuss two standard approaches to this problem: the 'complete ignorance' view and the 'Bayesian view'.

2.3.1 Minimal and Maximal Impact of Participation

According to the complete ignorance view, two important reference points are the *minimal* possible impact and the *maximal* possible impact (where the min and the max are taken over all distributions of the other participants' votes). Indeed, the pessimistic 'maxmin' principle would focus exclusively on the minimal impact, the optimistic 'maxmax' principle would focus on the maximal impact, and the well-known 'Hurwicz criterion' would consider a convex combination of both; see Luce and Raiffa (1957).

As observed above, the potential impact under the mean rule only depends on the number of voters and is given by $\text{imp}^{\text{Mean}} = \frac{Q}{k}$. Thus, for any given number of participants the minimal and maximal possible impacts coincide. Observe in particular that the potential impact is always positive under the mean rule and constant for any given number of participants. Summarizing, we have for all i,

$$\min_{q_{-i}} \text{imp}_i^{\text{Mean}}(q_{-i}) \;=\; \max_{q_{-i}} \text{imp}_i^{\text{Mean}}(q_{-i}) \;=\; \frac{Q}{k} \;>\; 0. \tag{8}$$

By contrast, the minimal impact under the median rule can well be zero; for instance, if sufficiently many of the other participants cast the same vote, no single vote can change the location of the median. In other words,

$$\min_{q_{-i}} \text{imp}_i^{\text{Med}}(q_{-i}) \;=\; 0.$$

By contrast, the maximal impact under the median rule can be large. It depends on whether k is even or odd; specifically, we have

$$\max_{q_{-i}} \text{imp}_i^{\text{Med}}(q_{-i}) \;=\; \begin{cases} Q, \text{ if } k \text{ is odd} \\ \\ \frac{Q}{2}, \text{ if } k \text{ is even.} \end{cases} \tag{9}$$

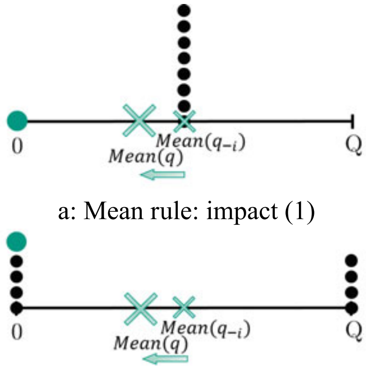
To prove (9), consider for odd k a situation in which half of the other $k-1$ participants vote for 0, and the other half for Q; then evidently $\mathcal{OS}_i = [0, Q]$ resulting in a maximal impact of Q. Similarly, if k is even, $k/2$ of the other voters are at 0 and $k/2 - 1$ are at Q, we have $\mathcal{OS}_i = [0, Q/2]$ resulting in a maximal impact of $Q/2$; one also easily verifies that the impact for even k cannot be larger than $Q/2$.

Comparing the mean and the median rule, we thus find that the minimal impact is always larger under mean rule; the maximal impact is equal under both rules for $k = 2$ and strictly larger under the median rule for $k \geq 3$. While large impacts are possible under the median rule, they require quite special ‘polarized’ constellations of the votes of the other participants.

2.3.2 Expected Impact of Participation

We turn to the *expected (potential) impact* of an individuals’ vote on the social outcome. Under the median rule, the expected impact depends on the belief about the distribution of the other votes. In fact, it turns out that it can be equal, larger, or smaller than the impact under the mean rule.

Let us consider different symmetric distributions with mean $\frac{Q}{2}$. Under a uniform distribution of $(k-1)$ votes, the expected impact under the median rule is identical to that under the mean rule. Indeed,



a: Mean rule: impact (1)

b: Mean rule: impact (2)

Fig. 1 Impact of vote q_i under the mean rule. Impact of a vote under the mean rule does not depend on the distribution of votes

$$E_{\text{unif}}(\text{imp}_i) = \frac{Q}{k}. \tag{10}$$

The reason is that under a uniform distribution of $(k-1)$ votes, the expected distance of two adjacent votes is exactly $\frac{Q}{k}$, and as noted above, it is this distance which is relevant for the potential impact under the median rule. To illustrate, consider the following example: $Q = 100$, $(k-1) = 3$, and a uniform distribution of votes at $q_1 = 25$, $q_2 = 50$, and $q_3 = 75$. Under either rule, we obtain the outcome $x(q_{-i}) = 50$. The expected impact of an additional vote q_i under the mean rule is identical to the one under the median rule and equal to $Q/k = 25$ as stated in (10); indeed, this is the length of the option set $\mathcal{OS}_i = [37.5, 62.5]$ under both rules.[2]

If the other participants' votes are not uniformly distributed, the expected impact under the median rule may be smaller or larger than under the mean rule. For instance, if the other participants' votes are normally distributed around the mean $\frac{Q}{2}$, the expected impact under the median rule is smaller than the expected impact under the mean rule which remains at $\frac{Q}{k}$. The reason is that the *median interval*, i.e. the distance between the vote to the left and the vote to the right of the median, is smaller than $\frac{Q}{k}$ in expectation under a normal distribution. By contrast, an analogous argument shows that the expected impact under the median rule is larger than $\frac{Q}{k}$ for a (symmetric) bi-modal distribution.

For further illustration, consider Figs. 1 and 2 which show examples with $k = 9$ voters. Specifically, we have a set of eight voters, represented by the black circles that either all vote for the allocation of $\frac{Q}{2}$ (Figs. 1a and 2a), or they split equally between 0 and Q (Figs. 1b and 2b). These examples can be understood as extreme cases of a normal distribution (with zero variance) and of a bi-modal distribution.

[2] Observe that by using an appropriate notion of 'uniform distribution' in the discrete case, we do not need to distinguish between even and odd numbers of voters. For instance, the 'uniform distribution' of four votes would correspond to $q_1 = 20$, $q_2 = 40$, $q_3 = 60$ and $q_4 = 80$. The resulting expected potential impact of an additional vote is easily calculated to be equal to $Q/k = 100/5 = 20$ for both rules.

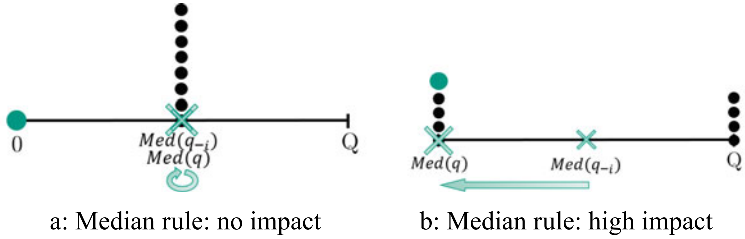


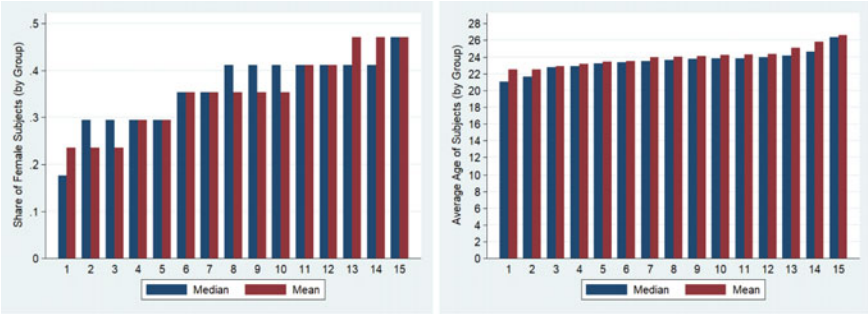
Fig. 2 Impact of vote q_i under the mean rule. Impact of a vote under the median rule depends on the distribution of votes. The impact of a vote under the median rule depends on the distribution of votes

Under both rules, the outcome is identical and denoted by Mean(q_{-i}) and Med(q_{-i}), respectively. The green circle displays the vote of participant i, which is $q_i = 0$ in all cases. (A vote $q_i = Q$ would yield symmetric results.) The mean outcome including voter i is denoted by Mean(q), and according to our observations above the impact of i's vote under the mean rule does not depend on the distribution of the q_{-i}; see Fig. 1a, b. By contrast, under the median rule the impact is zero in Fig. 2a and equal to Q in Fig. 2b.

Summarizing, the impact under the mean rule is small for large k but certain, while the (expected) impact under the median rule can be smaller, equal to, or larger than that of the mean rule. In any case, the variance of the (expected) impact under the median rule is larger than zero. We therefore hypothesized that risk averse, resp. ambiguity averse, individuals would tend to participate less under the median rule as compared to the mean rule. In order to test this and related hypotheses, we conducted the field experiment described in the remainder of this paper.

3 The Field Experiment

The field experiment was conducted at the Karlsruhe Institute of Technology. We set up a vote over the allocation of a donation on two campus projects using either the mean or the median rule. Our main focus are the participation rate and the role of the impact of a vote under both voting rules. Subsequent to the vote, we implemented a survey in order to elicit beliefs about the allocation result, about the participation rate and about the impact on the social outcome. Additionally, we asked for strategic voting behaviour and elicited risk preferences.



a: Share of female subjects b: Average age of subjects

Fig. 3 Gender and age of subjects, by group (increasing). Subject pool is randomized in terms of gender and age

3.1 General Set-up and Design

The field experiment went on for one week in July 2017 at the Karlsruhe Institute of Technology. We invited 510 subjects to participate in a vote over the allocation of a donation on two campus projects: the *bike workshop* (an assisted workshop which provides tools, help, and space for students that would like to repair their bike) and the *campus garden* (the possibility to grow plants, cultivate, and harvest fruits and vegetables). Both projects are on campus and accessible free of charge for all students. We assessed wide interest among the students for these projects and made sure that the implementation could be realized by the General Students' Committee (AStA) for every possible outcome. Pictures of the projects can be found in "The Projects" of Appendix.

We randomized subjects using the HROOT subject pool (Bock et al. 2014) of the KD2Lab at Karlsruhe Institute of Technology. Our randomly selected pool consists of 510 subjects which were divided into 30 groups of 17 members each. The subjects received an invitation to participate in the voting process via e-mail on 11 July 2017. Each group voted over the allocation of 100 Euros on the two projects. We use a between-subjects design with the treatment variable voting rule: 255 subjects were randomly assigned to the mean rule and 255 to the median rule. Our pool consisted of 90 female subjects under the mean rule and 92 under the median rule. The average age was 24.05 under the mean rule and 23.52 under the median rule. Figure 3 shows the average age and share of female subjects by group in increasing order for both treatments. We conclude that our subject pool is balanced in terms of gender and age.

In the e-mail, subjects were informed that they belonged to a group with 16 other persons who were also invited to vote over the allocation of 100 Euros on the bike workshop and the campus garden. We explained the voting rule (mean or median) and the requirements for a valid vote. Subjects were also informed that other groups would vote in a similar fashion over the allocation of another 100 Euros and that for an implementation of the group outcome at least one vote within the group is necessary. Together with the remark that no individual payments to the voters will be made followed the link to submit one's vote.

The vote for an allocation on the two projects was accompanied with a short questionnaire. The questionnaire asked for information about the individual beliefs regarding the allocation result, the number of participants in each group and the degree of the impact of their vote. We further asked if the voting rule was well understood and if the participants revealed their true preferred allocation. Besides, we asked for demographic data and elicited risk preferences. In "Voting Over Donation Projects at KIT" of Appendix, we provide the e-mail that subjects received under both treatments. Screenshots of the voting process and the questionnaire can be found in "The Voting Process" and "The Questionnaire" in Appendix.

3.2 Eliciting Impact Beliefs and Risk Preferences

When comparing the mean and median rule for relatively small groups, we hypothesize that differences in participation rates are driven by different beliefs about how participation affects the social outcome. We capture these beliefs in two different ways.

First, we asked the participants directly to assess the impact of their vote. For this, we asked them to choose one out of six qualitatively described impact categories, ranging from 'my vote does not have any impact' to 'my vote is decisive for the outcome'. While these categories are probably subject to individual interpretation, this is arguably the most suitable way to elicit impact beliefs directly.

As a more indirect way to assess the beliefs about the impact of one's own vote, we asked participants about their belief on the number of participants in their respective group. For the mean rule, the impact strictly decreases with the number of voters, and also under the median rule, the possibilities for affecting the outcome are higher if the number of participating voters is low.

Since we expected that differences in participation rates can be traced back to risk attitudes via impact beliefs, we also elicited risk preferences. Charness et al. (2013) present and evaluate several methods for eliciting risk preferences in experiments. The authors argue that simple methods are especially suitable for capturing treatment effects, which is the aim of our study. We elicit risk preferences using the Eckel and Grossman (2002) method, where subjects have to choose one out of a series of lotteries.[3] We adapt the values of the original gambles as suggested by Dave et al. (2010). Dave et al. (2010) let participants choose between six lotteries, each involving a high and a low payoff with equal probability of 50%. Table 1 presents the lottery choices. Lottery 1 represents a secure option as subjects receive a payoff of 28$ for sure.[4] Expected payoffs increase together with the risk level from lottery 1 to lottery 5. Lottery 5 represents risk neutrality as it comes with the highest expected

[3] Note that the choice was not incentivized.

[4] We adapt the currency from $ to Euros in our experiment but otherwise stick to the same values.

Table 1 Lottery choices from Dave et al (2010)

	Low payoff	High payoff	Exp. return	S.D.
Lottery 1	28	28	28	0
Lottery 2	24	36	30	6
Lottery 3	20	44	32	12
Lottery 4	16	52	34	18
Lottery 5	12	60	36	24
Lottery 6	2	70	36	34

Lotteries 1–4 indicate risk aversion, lottery 5 risk neutrality and lottery 6 risk-seeking behaviour

return combined with a lower standard deviation as compared to lottery 6, which implies risk-seeking behaviour.

Reynaud and Couture (2012) compare the Eckel and Grossman method to the well-known elicitation method of Holt and Laury (2002); they perform a non-incentivized field experiment and find that while there exist differences among elicitation methods, the risk attitudes are significantly correlated across the different lottery tasks. The main advantage of the Eckel and Grossman method for our purposes is that by letting individuals choose only one lottery, we exclude inconsistent decisions like subjects switching lotteries in the Holt and Laury method. Moreover, the Eckel and Grossman task is simpler and the explanation can be done faster. One should keep in mind the subjects participated in the vote on a voluntary basis and we wanted to keep the questionnaire as short as possible.

3.3 Hypotheses

While we have shown above that the (expected) potential impact on the social outcome can be larger under the median rule than under the mean rule, its quantification is much more involved under the median rule. By contrast, the impact under the mean rule is easy to understand; moreover, it only depends on the number of participants and not on the distribution of the other participants' votes. Therefore, we expected that subjects overestimate the impact their vote would have under the mean rule as compared to the median rule.

Hypothesis H1. ***The belief about the real impact is higher under the mean rule as compared to the median rule.***

Hypothesis H1 is further backed up by the following consideration. Ex post, every voter's vote is pivotal for the outcome under the mean rule; by contrast, under the median rule no participant is pivotal except one (the median voter if the number of participants is odd), or possibly two (the two middle voters if the number of participants is even). Thus, also from the perspective of *how many* participants are pivotal (ex post), the mean rule fares better than the median rule.

In the light of the greater variance of the expected potential impact under the median rule, we also expected that there is a *selection effect* in the set of participants under the two rules. This is based on the following two presumptions. First, we assume that (in line with many empirical studies) most subjects are risk averse (independently of treatment); second, for risk-averse individuals the probability of participation under the median rule is lower precisely because the impact of one's vote is 'more uncertain' under the median rule than under the mean rule. We therefore hypothesized:

Hypothesis H2. ***Actual participants are more risk averse in mean voting as compared to median voting.***

Based on H1 and H2, together with standard assumptions about individual preferences, in particular about the widespread trait of risk aversion, see e.g. Holt and Laury (2002), we expect that the voter turnout under the mean rule is higher as compared to the median rule.

Hypothesis H3. ***The actual number of participants is higher under the mean rule as compared to the median rule.***

Our last hypothesis does not address voter turnout but the structure of the actual votes. It is well known that under the mean rule (in a complete information setting) the unique Nash equilibrium of the voting game with a fixed number of participants prescribes all but possibly one participant to vote for one of the extreme options, i.e. 0 or Q (Block, 2014; Renault & Trannoy, 2005). By contrast, as is also well known, truth-telling is a weakly dominant strategy under the median rule. The 'game' that our subjects played is of course different in two important respects: First, it took place under conditions of incomplete information, and secondly, it involved not only a voting but also a participation decision. Nevertheless, the qualitative difference of the equilibria of the corresponding voting games (under complete information and a fixed number of participants) made us hypothesize:

Hypothesis H4. ***The variance of the participants' votes is higher under the mean rule as compared to the median rule.***

The field experiment allows us to test all four hypotheses. The data for H3 and H4 are the direct observations of the voter turnout and the votes of all participants. The data used for H1 and H2 are collected via the subsequent survey filled out by participants.

4 Results

Our main focus in the following is whether there is a difference in the participation rates under the two voting rules, and if so, whether our data on impact beliefs and risk preferences are able to explain it. We also study if subjects voted optimally given their beliefs.

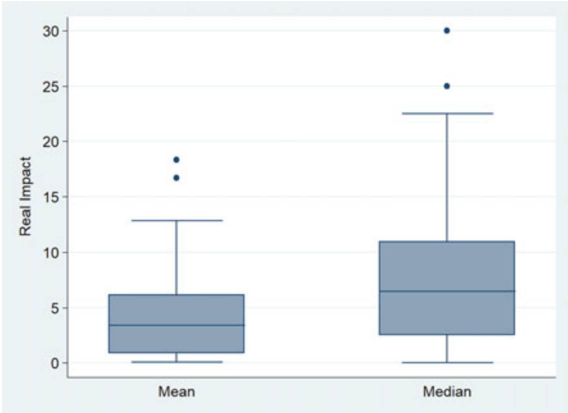


Fig. 4 Real imapct. Real impact of votes is higher under the median rule

4.1 Real Impact Versus Assessed Impact

For the analysis of the experimental results, we define the *real impact* of a participant's vote as the difference of the actual social outcome and the hypothetical outcome without this voter's participation given the actual distribution of votes for each group, i.e. the absolute difference between $x(q_{-i})$ and $x(q)$. This can be always be calculated because in all groups there were at least two participants. The real impact ranges across individuals, i.e. the interval from the lowest to the highest real impact, is [0, 30] under the median rule, and the range under the mean rule is [0.08, 18.33]; see the boxplot in Fig. 4. Our observations align well with the theoretical considerations above: the real impact under the mean rule is small but always strictly positive; by contrast, it can be zero under the median rule and has higher variance. The average individual real impact is 4.48 Euros under the mean rule, and the real impact values are significantly lower as compared to the median rule with an average individual real impact of 7.90 Euros (Mann–Whitney U test, $p = 0.018$).

Is this significant difference in the real impact, as one could have expected from the theoretical analysis, also reflected in the beliefs of the subjects? We elicited the belief about the impact by letting the participants evaluate their impact in six qualitative categories. The categories range from 'my vote has no impact on the social outcome' to 'my vote is decisive'. The share of participants that chose the respective categories is displayed in Fig. 5.

Figure 5 indicates a slightly higher belief about the impact for mean rule participants, but the difference is in fact not statistically significant; in particular, our data do not support Hypothesis H1 (Mann-Whitney U test, $p = 0.159$).

Figure 6 plots the belief about the impact against the real impact for each participant. Each participant is represented by a bubble, and larger bubbles represent several subjects. We classify the real impact values into six categories, indicated by the blue frames in Fig. 6: 'No impact' corresponds to an impact of 0 Euros, 'very low' is classified as an alteration of the outcome by more than 0 and up to 5 Euros and so on;

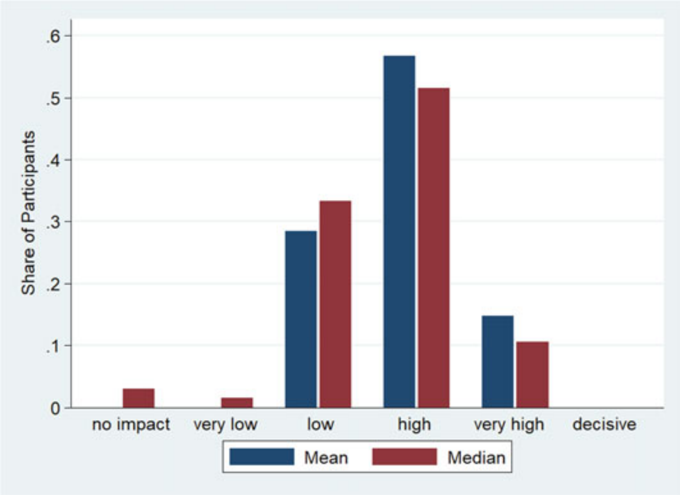


Fig. 5 Beliefs about impact. No treatment effect regarding participants' beliefs about the impact of their vote

'decisive' is classified as an alteration by more than 20 Euros (recall that the highest value of the real impact was 30). Participants whose belief matched their real impact (given our classification) are situated within the blue areas and make for 12.16% of mean and 18.18% of median rule participants. Very few subjects underestimated their real impact under the mean rule (6.76%), a somewhat higher share did so under the median rule (18.18%) (these are represented by the bubbles to the top-left of the blue areas). An overestimation occurs for 81.08% of the participants under the mean and for 63.64% under median rule (the bubbles below the blue areas). We find that the differences in proportions of over- and underestimation are significant (two-sample test of proportions, $p = 0.019$ for underestimation, $p = 0.010$ for overestimation); but clearly, these figures depend on the chosen classification.

We elicited beliefs about the impact also indirectly by asking the participants about their belief about the allocation result. Figure 7 plots this against the impact beliefs. The horizontal axis shows the classification of the impact belief, while the vertical axis shows the distance between the actual vote and the belief about the result of the respective participant (again represented by a bubble). We additionally draw the regression line and find that there is no correlation. It is remarkable that some subjects believe to have had a very high impact when asked directly and at the same time indicate that they believe the group result will differ from their own vote by more than 30 Euros.

We also examine gender differences in impact beliefs by performing a Mann–Whitney U test on the impact belief depending on gender.[5] We are able to reject the H_0 hypothesis that male and female participants have an equal belief about their

[5] Two participants did not state their sex; therefore, we exclude these observations.

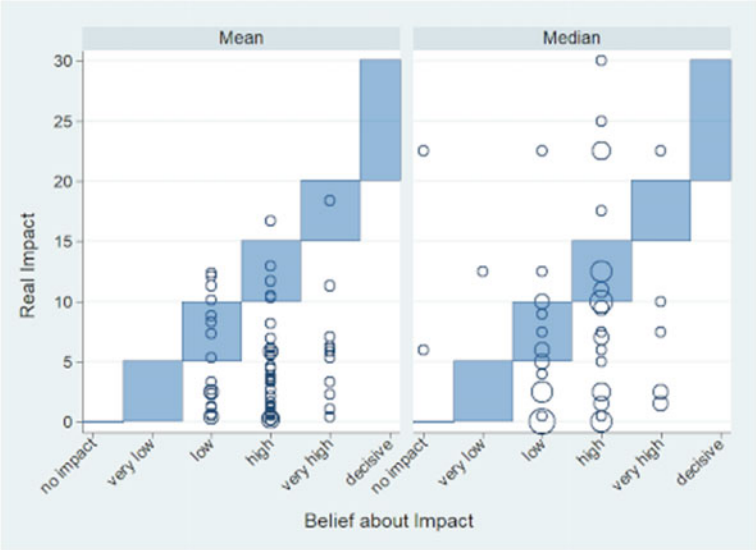


Fig. 6 Belief about impact versus real impact. We categorized the belief about the impact (horizontal axis) and compared it to the real impact (vertical axis). Very few subjects (represented by bubbles) underestimated their real impact under the mean rule; a slightly higher, yet still low share did so under the median rule. A large number of participants overestimated their impact under both treatments with a yet higher share under the mean rule

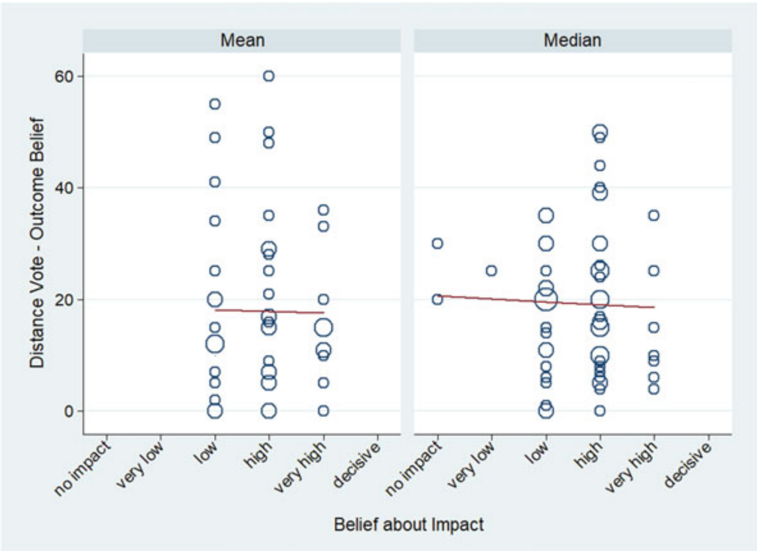


Fig. 7 Measures of impact belief. Different measures of impact belief: direct question (horizontal axis) versus distance between actual vote and belief about the result (vertical axis). No correlation found

impact with $p = 0.034$ in support of the H_1 that the impact belief of men is on average higher as compared to women. Since female participants have on average a lower belief about their impact, we tested in a second step if women also overestimate their impact less frequently. Out of the 102 participants that overestimated their impact, 63 were male and 37 were female. Therefore, the share of male participants that overestimated their impact is 68.48% (63 out of 92), while the share for women was 80.43% (37 out of 46). However, we do not find a significant correlation between overestimation and gender (chi-squared test, $p = 0.138$).

When analysing the votes separately for male and female participants, we find some interesting differences. While 24 male participants, i.e. 26.09% of all men, voted for an extreme allocation of either 0, 1, 99 or 100 Euros, the share of female participants that voted extreme is only 8.70% or four women. Building on these differences in the voting behaviour, we performed another Mann–Whitney U test to compare the real impact dependent on gender and find indeed that the real impact of male participants is on average significantly higher as compared to the real impact of women ($p = 0.026$).

Considering the high shares of overestimation of the real impact, the question arises if these high beliefs are driven by an underestimation of the number of participants (which is negatively correlated with impact for both rules). In both treatments, subjects believed on average that the number of participants per group is 8.8 (no difference, two-sample t-test, $p = 0.966$). This means that the share of participants that overestimated the number of participants as compared to the real number of participants is 84.85% in the median (with a true group average participation rate of 4.4) and 74.32% in the mean groups (with a true average of 4.9 participants).

If one measures the belief about impact indirectly by the belief about the number of participants, most subjects underestimate this impact, since the belief about the number of participants is higher than the actual number of participants. Therefore, participants overestimate their real impact directly (as inferred from Fig. 6) and at the same time overestimate the number of participants. Figure 8 sheds light on the indicated beliefs about the impact in combination with the beliefs about the number of other participants per group. Remarkably, the red regression line has a positive slope in the mean treatment; the correlation between the belief about the number of participants and the belief about one's own vote impact is 0.188 (Spearman correlation, $p = 0.109$). Since the actual impact of participation under the mean rule is negatively correlated with the number of participants, subjects display inconsistent beliefs under the mean rule. Under the median rule, the correlation is slightly negative (-0.072) so that we do not find the same inconsistency in beliefs for the median rule (Spearman correlation, $p = 0.566$).

4.2 Risk Preferences

Since we only have the survey data for the subjects who completed the questionnaire, the following analysis refers to these subjects only. We elicited risk preferences using

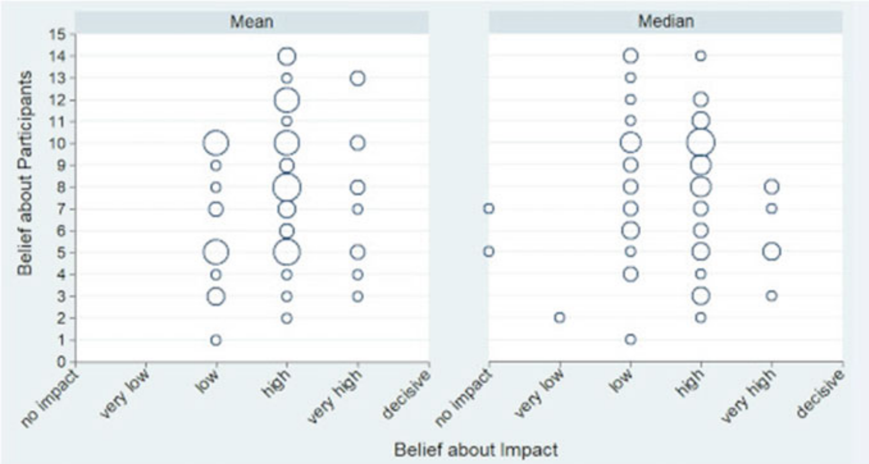
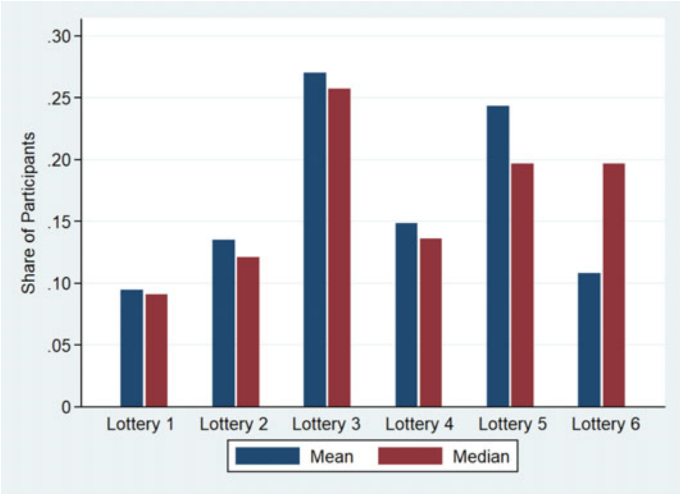


Fig. 8 Belief about impact versus belief about participants. Different measures of impact belief: direct question (horizontal axis) versus beliefs about the number of participants per group (vertical axis). Since the number of participants is negatively correlated with the impact of a vote, the positive slope of the regression line under the mean treatment indicates inconsistent beliefs

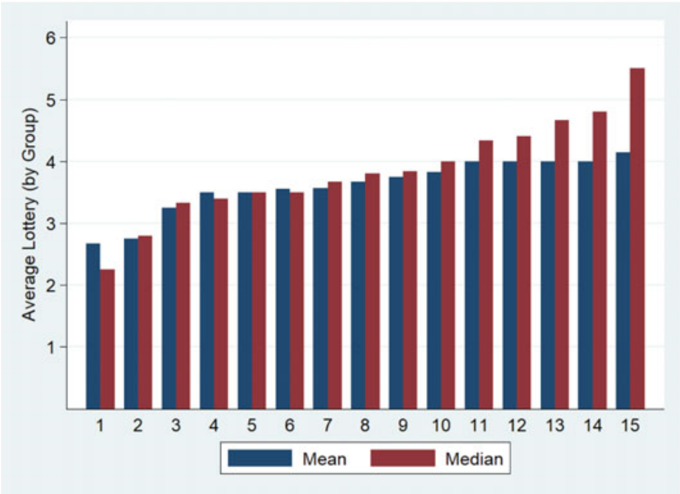
the Eckel and Grossman (2002) method with the values given in Table 1 above. Figure 9a displays for each of the six lotteries the share of participants by voting rule. The average lottery number per individual is 3.64 under the mean rule and slightly higher under the median rule, where the average lottery is 3.82. Both values indicate risk aversion at the individual level, since the average is below lottery 4, which among the risk-averse lotteries is the one with the highest expected return and the highest standard deviation. At the individual level, we however do not find a significant difference between the two rules (two-sample t-test, $p = 0.485$). Thus, our individual data do not support Hypothesis H2.

We label the six lotteries by corresponding values from 1 to 6 and run a regression of the chosen lottery on a set of independent variables. The results are displayed in Table 2, and again we see that the voting rule (the relevant coefficient is 'rulemean', which takes the value 1 for the mean rule and 0 for the median rule) is negative but not significantly so. Interestingly, we find a highly significant and positive coefficient for the dummy variable 'gendermale', which takes the value 1 if the participant is male and 0 for female participants.

In order to further analyse gender differences, we perform a two-sample t-test for the individual risk attitudes of male vs. female participants. We are able to reject the H_0 hypothesis that there is no difference in the risk preference between men and women (independent of the voting rule) at a p-value below 1% and find support for H_1: the average risk preference level is lower for female as compared to male participants (on average 2.98 vs. 4.09). Since we do find a gender difference regarding the risk preferences, our next question is if the share of female participants under the two voting rules is different. As argued above, risk-averse subjects could be assumed



a: Shares by lottery



b: Average lottery, by group (increasing)

Fig. 9 Risk preferences. Lotteries indicating average risk-averse preferences. No treatment effect found

to prefer the sure impact under the mean rule, and the higher degree of risk aversion by women would therefore predict a higher female participation rate in mean voting as compared to median voting. The overall share of female participants was 33.33% and splits up into 35.14% under the mean rule and 31.25% under the median rule. Our overall subject pool (participants and non-participants) consisted of 91 women under the mean and 92 under the median rule. The adjusted share of female participants

Table 2 Regression results risk preferences

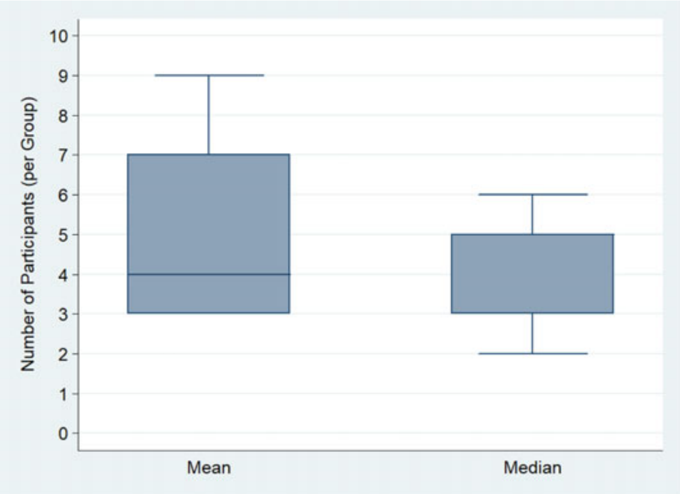
Variables	Lottery
Rulemean	−0.0985 (0.256)
Extremevote	0.0138 (0.336)
Partbelief	0.0135 (0.0403)
Man	0.574 (0.494)
Impactbelief	−0.127 (0.176)
Gendermale	1.113*** (0.275)
Constant	3.230*** (0.580)
Observations	138
R-squared	0.130

Standard errors in parentheses
*** p <0.01, ** p <0.05, * p <0.1
No treatment effect regarding risk preferences. However, male participants chose significantly higher lottery, i.e. they tend to be more risk seeking

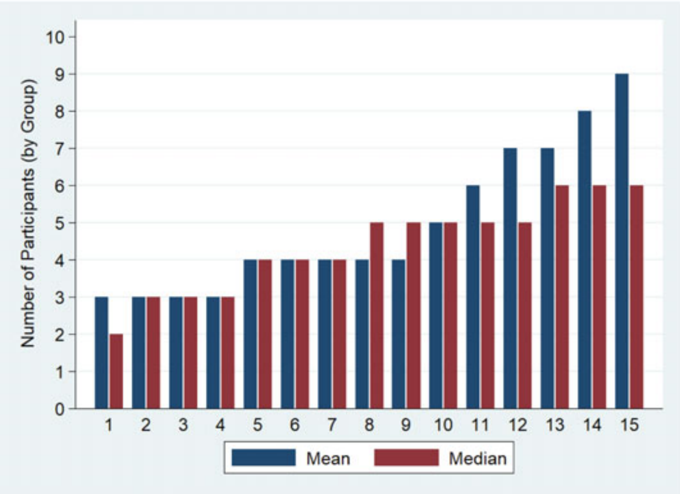
out of all female subjects is therefore 28.57% for the mean rule and 21.74% for the median rule, which makes an even higher difference. Nevertheless, we do not find that the difference in participation rates by gender is statistically significant (chi-squared test, $p = 0.381$).

4.3 Voter Turnout

Out of the $n = 510$ subjects that were invited, $k = 140$ participated and completed the entire voting process. Further 21 subjects visited the survey platform but did not complete the form. Of the 140 participants, 74 subjects participated under the mean rule and 66 under the median rule. The overall participation rate was quite high with 29.0% and 25.9%, yielding 4.9 participants on average per group under the mean and 4.4 under the median rule. There were at least two participants in each group and a maximum number of nine participants, which occurred under the mean rule. Figure 10a presents a boxplot containing the number of participants per group. The median number of participants per group is four under the mean and five under the median rule. The detailed number of participants by voting rule and groups is displayed in Fig. 10b, ordered by increasing number of participants. As one can see in the two figures, the spread of participation rates is higher under the mean rule. In six



a: Boxplot of participants



b: Participants by group (increasing)

Fig. 10 Average number of participants. No treatment effect regarding the participation rate

of the ordered groups, the number of participants under the mean rule exceeds the one under the median rule; the opposite is true for only two groups. We do find support for the hypothesis that the variance of mean rule participation is higher as compared to participation under the median rule groups (variance ratio test, $p = 0.046$).

In order to test our Hypothesis H3, we run a regression of the dichotomous variable 'participation' on a dummy variable for the voting rule and gender. These are the only independent variables that we have for participants and abstainers, as the subject pool contains information on gender and the voting rule was assigned by us. The coefficient for 'rulemean', which takes the value 1 for the mean rule and 0 for the median rule, is $\beta_1 = 0.0370$; the positive value indicates that the voter turnout is higher for the mean rule. However, we cannot reject the null hypothesis that the deviation is significantly different from zero. The same is true for the coefficient of the gender dummy variable, indicating that male subjects have a higher participation rate, yet the difference is not significant ($\beta_2 = 0.0314$).

We additionally consider the aggregated group values for participation and perform a Mann–Whitney U test. The rank sum of the average group participation rate is higher under the mean rule, but the difference is also not statistically significant ($p = 0.735$). Our data thus do not support our Hypothesis H3; indeed, we do not find evidence for different numbers of participants among the two rules. This result has to be considered with some caution, however. The number of groups that we compare is limited to fifteen, and the average participation rates per group differ in only eight cases.

4.4 Distribution of Votes

The actual distribution of votes is displayed in Fig. 11a. The horizontal axis shows the votes for the bike workshop in Euros. The vertical axis ('share of votes') represents the percentage of participants that voted for the respective amount. The overall distribution shows the highest percentage at 100 Euros, indicating the preference to allocate all the money on the bike workshop. Further remarkable values are those close to a (70, 30) split as well as the symmetric (30, 70), indicating the existence of reference points besides the extreme allocations or the equal split, which only two participants under the median rule and three under the mean rule voted for. We do not find a significant difference across the rules in the votes themselves (Mann–Whitney U test, $p = 0.680$), in the distribution of votes (two-sample Kolmogorov–Smirnov test, $p = 0.830$), nor in the variance of votes (variance ratio test, $p = 0.553$). A boxplot of the votes for the bike workshop by rules is shown in Fig. 11b.

These results imply that Hypothesis H4 is not supported either; in fact, we do not find that the variance of votes is higher under the mean rule. This is an indication that participants did not vote strategically under the mean rule, for which extreme voting (i.e. $q_i = 0$ or $q_i = Q$) is almost always optimal.[6] In the light of the laboratory experiments of Puppe and Rollmann (2021), Block (2014), and Rollmann (2020), this is a particularly remarkable finding; indeed, in the laboratory situation we consistently find strategic behaviour under the mean rule even in situations of incomplete information. Our present results thus suggest that in a voting context the

[6] Observe that the optimality of extreme voting for almost all voters under the mean rule carries over to an incomplete information setting for all Bayesian rational players. This is because an interior vote $\tilde{q}_i \in (0, Q)$ can only be optimal under the mean rule if the social outcome is exactly at $\tilde{q}_i$.

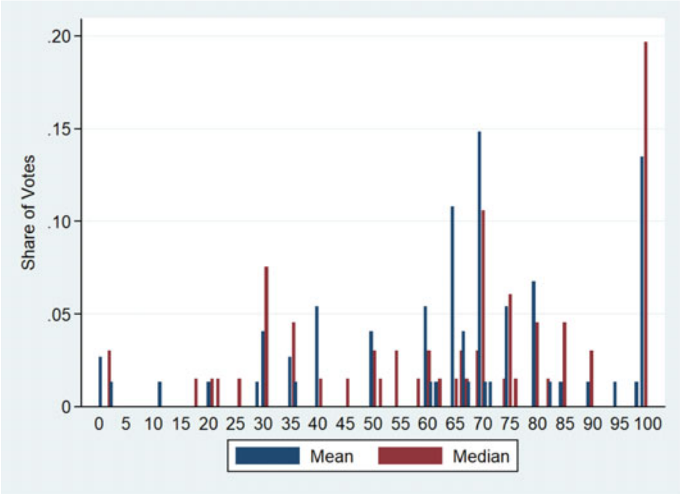
natural environment of field experiments plays an important role and is crucial for the behaviour of subjects. To identify the confounding effects in the field as compared to the more tightly controlled laboratory situation seems an important task for future research.

Also observe that since the distribution of votes did not differ significantly among the voting rules, we can also conclude that the difference that we found in the real impact is not driven by differences in the distribution of votes.

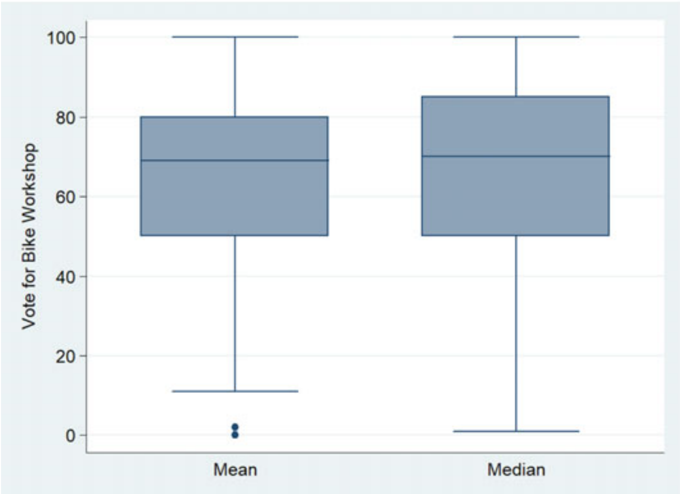
4.5 Allocation Outcomes

We also do not find a significant difference in the group allocation results across the two voting rules (two-sample t test, $p = 0.227$). The allocation result ($bike,\ garden$) is on average (64.97, 35.03) under the mean rule and (70.47, 29.53) under the median rule. Figure 12a shows a boxplot of the allocation results by rule. Though the group results are not significantly different, we do observe a greater spread in the group results under the median rule, where one group result was to donate the total budget of 100 Euros to the bike workshop project. In this group, three subjects participated and the vector of votes were $q = (85, 100, 100)$. Also, only under the median rule the social outcome was once below 50 Euros for the bike workshop. The greater variance in the group allocation results under median voting is statistically significant (variance ratio test, $p = 0.021$). The total donation for the bike workshop adds up to 2,031.60 Euros and for the campus garden to 968.40 Euros.

We also asked for the belief about the allocation result, which is slightly more balanced as compared to the actual result: (62.10, 37.90) is the average belief under the mean and (62.87, 37.13) under the median rule. We do not find a significant difference in the belief about the allocation result across voting rules (two-sample t-test, $p = 0.786$). Interestingly, we do find a significant difference in the correct belief about the result (two-sample t-test, $p = 0.019$). Specifically, we calculated the difference between the belief about the result and the real result and find that the average difference under the mean rule is -2.08, as compared to -7.08 under the median rule. The negative sign indicates that participants under both rules believed that the result for the bike workshop is lower than it really was, i.e. on average they underestimated the share for the bike workshop, or in other words, they believed that the result was more balanced among the projects. We find that the average value for correct estimation is significantly higher under the mean rule, which means that the average of the participants' beliefs is only 2.08 Euros lower as the real result. Under the median rule, the average deviation from the real result is 7.08 Euros. Figure 13 plots the belief about the result versus the real result. The correlation is positive for both rules (0.061 for the mean rule and 0.467 for the median rule); however, the coefficient is significant only for the median rule (Spearman correlation, $p = 0.608$ and $p < 0.001$). We find that the difference between the rules is significant (two-sample Fisher's z-test, $p = 0.037$).

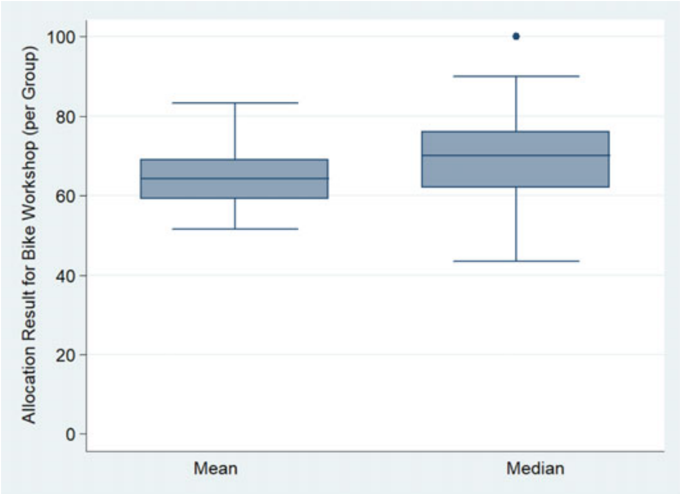


a: Distribution of votes

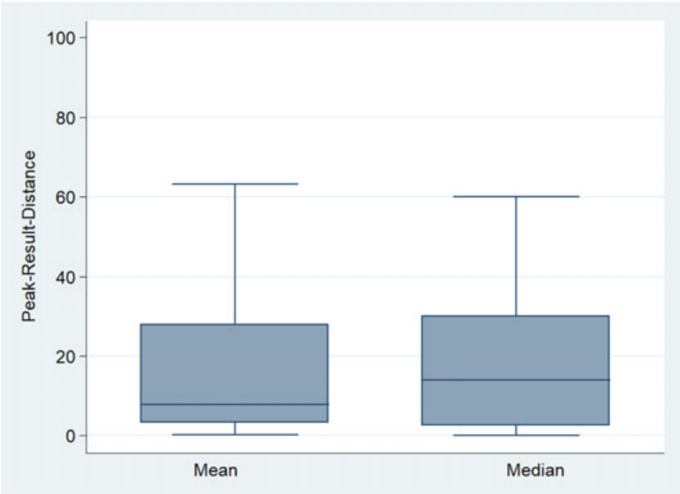


b: Boxplot of votes

Fig. 11 Votes for bike workshop. No treatment effect regarding the distribution of votes. No treatment effect regarding the distribution of votes



a: Allocation results for bike workshop



b: Distance between peak and result

Fig. 12 Allocation results and distance to peaks. No treatment effect regarding the allocation results nor the distance between the peak and the allocation result

In order to (roughly) assess welfare under both voting rules, we assume that the votes correspond to the true preference peaks of the participants (which seems well justified in the absence of strategic voting) and calculate for each subject the distance between the peak (vote) and the aggregated group results. Implicitly, we thus also assume that the true underlying preferences are symmetrically single peaked, and that the costs of participation are identical across subjects. Figure 12b displays the

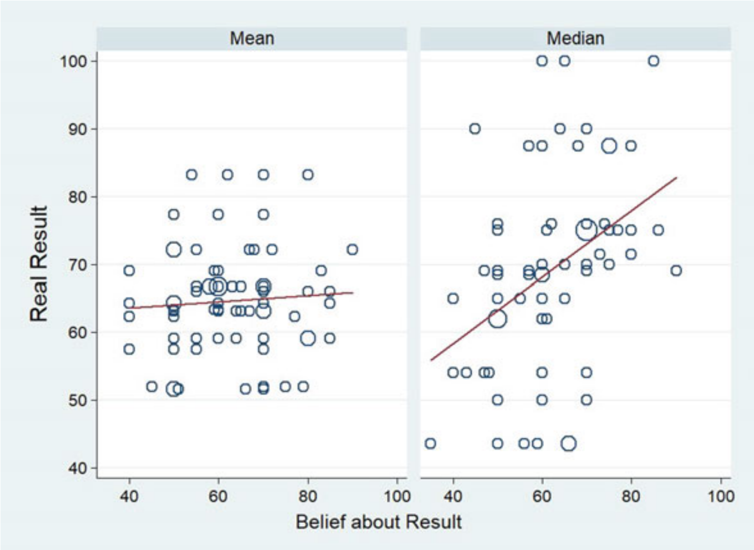


Fig. 13 Belief about result versus real result. Positive correlation between the belief about the result and the real result; the coefficient is significant for the median rule

boxplots for the distance between peak and result for both voting rules. We do not find a significant difference in the average distance (two-sample t-test, $p = 0.534$) nor in the variance between both rules (variance ratio test, $p = 0.837$).

4.6 Non-truthful and Strategic Voting

Clearly, in our field experiment we cannot know the underlying preferences and hence cannot infer if subjects voted truthfully or strategically purely from the observation of their votes. Nevertheless, one can ask them about the motives for their behaviour and that is what we did. Specifically, we asked them if they casted their vote truthfully and if not about their true peak p_i.

As already noted, under the mean rule the belief about the social outcome is sufficient for deciding where to place one's vote. For instance, if the belief about the social outcome is $\mathrm{Mean}(q_{-i}) = 70$ and the most preferred outcome of i $p_i < 70$, submitting a vote $q_i < p_i$ is optimal but of course non-truthful. (Recall that we only list the expenditure for the first project, which is the bike workshop in the present context.) Table 3 lists all participants under the mean rule that stated to have voted non-truthfully. The share of all votes that are non-truthful (as per the statements of the participants) is 6.76% and rather low (five subjects). Based on the reported true peaks p_i, the beliefs about the social outcome $b_i^{x(q)}$, and about the number of participants

Table 3 Non-truthful voting, mean rule

p_i	$b_i^{x(q)}$	$b_i^{(k-1)}$	$b_i^{x(q_{-i})}$	$q_i(b_i^{x(q)}, b_i^{(k-1)})^{(*)}$	$q_i(b_i^{x(q)}, b_i^{(k-1)})^*$	q_i
85	51	10	46.10	435.10	100	100
80	64	5	56.80	172.80	100	100
60	60	10	58.00	78.00	78	80
50	50	8	56.20	6.25	6	0
50	70	8	68.75	−81.25	0	80

Five individuals stated that they voted non-truthfully under the mean rule. The first two individuals submitted a vote that corresponds to the best response. Individuals in lines 3 and 4 submitted a vote very close to the best response. The vote of the individual in the last line appears to be 'non-strategic'

$b_i^{(k-1)}$, we calculate a theoretical best response $q_i(b_i^{x(q)}, b_i^{(k-1)})^{(*)}$ for each of the five individuals. After correcting for allocations that are feasible (only natural numbers between zero and 100), we are able to compare the theoretical belief-based best responses $q_i(b_i^{x(q)}, b_i^{(k-1)})^*$ to the actual votes q_i. Two individuals submitted a vote that corresponds to the best response ($q_i = q_i(b_i^{x(q)}, b_i^{(k-1)})^*$), which implies that these votes are not only strategic but also optimal given the beliefs. Two further votes were very close to the best response and therefore strategic, as they deviate from the true peak and decrease the distance to the social outcome belief. The one individual that places a 'non-strategic' vote according to the beliefs argued that he could not find a reason why his peak should be different from the equal split ('I don't find an argument why one project should be better for the community than the other'). Nevertheless, he voted for more budget on the project that seemed more useful personally ('I will probably never use the campus garden').[7] This argumentation puts the vote into perspective and makes it in some way strategic as well, as it seems that the indicated peak was based on the benefit for the community.

Table 4 summarizes the corresponding data for the six participants that stated to have cast a non-truthful vote under the median rule. The share of non-truthful votes is 9.10% and, surprisingly, higher than under the mean rule. Beneficial strategic voting under the median rule (and single-peaked preferences) is possible only for an even number of participants and only if the corresponding peak is one of the two in the middle. With the beliefs about the social outcome and the true peak, we get a hint on whether participants vote strategically according to their beliefs. As strategic voting is not possible for an odd number of voters, the best response given that $b_i^{(k-1)}$ is even is a vote that is at least as high as the belief about the median outcome $b_i^{x(q)}$ if $p_i > b_i^{x(q)}$, or at most as high as $b_i^{x(q)}$ if $p_i < b_i^{x(q)}$. The three individuals that believe $(k-1)$ to be even play a best response given their beliefs, but the belief-based distance between their peak and the social outcome is not reduced.

[7] The original statements are in German.

Table 4 Non-truthful voting, median rule

p_i	$b_i^{x(q)}$	$b_i^{(k-1)}$	$q_i(b_i^{x(q)}, b_i^{(k-1)})^*$	q_i
85	75	10	≥ 75	100
78	61	11	?	62
70	60	8	≥ 60	85
50	50	3	?	35
30	40	14	≤ 40	1
26	66	9	?	17

Under the median rule, six individuals stated that they voted non-truthfully. The three participants that believe $(k-1)$ to be even play a best response given their beliefs

5 Conclusion

Instead of providing a comprehensive summary, let us conclude by highlighting the following two findings. First, it appears that voting over a one-dimensional variable is governed by very different principles in the field compared to the laboratory. Secondly, we find much less strategic behaviour (under the mean rule) than one could have expected from the results of corresponding laboratory experiments.

We did not explicitly ask subjects about their motivation regarding participation or abstention. However, our results suggest that neither the voting rule nor the belief about the impact of one's vote was a main driving force behind the participation decision. We conclude that the motivation of participation was not so much personal benefit but rather the chance to contribute to a donation for the campus. This is also backed by the observation that only few participants voted non-truthfully resp. strategically.

We are only at the beginning of understanding the motivations underlying participation and voting decisions in our context, and there is certainly more work to be done. We believe that (controlled) field experiments can contribute significantly to our understanding of voting and participation in elections, and we hope that the present paper might stimulate further work in this direction.

Acknowledgements This work is based on Chapters 2 and 3 of the second author's Ph.D. dissertation; it has been presented at the ESA European Meeting 2017 in Vienna (Austria), at the the 13th HeiKaMaX Workshop 2018 in Mannheim (Germany), at the 14th Meeting of the Society for Social Choice and Welfare 2018 in Seoul (South Korea), at the Conference 'Improving Public Debate and Political-Decision Making' 2019 at the New College of the Humanities in London (UK), at the International Summer School on 'Pluridisciplinary Analysis of Collective Decision Making' 2019 at Université de Caen Normandie (France), and in Seminars at the University of Paderborn (2018), Université de Paris Dauphine (2019) and Higher School of Economics, Perm Campus (2021). We are grateful to the participants for helpful comments and suggestions. Special thanks to Tobias Dittrich and Michael Müller for continuous discussions on the topic of the present paper and related topics throughout the years and to Tim Cleffmann for technical support in the experiment. Needless to say that all errors are our own

Fig. 14 Bike workshop and campus garden (https://www.asta-kit.de)

Appendix: Supplementary Material

The Projects

See Fig. 14.

Voting Over Donation Projects at KIT

Mean Treatment

Hello,

The Chair of Economic Theory at KIT would like to invite you to participate in an election. Besides you, **16 other people** have been invited to participate in this election. Your group votes on the allocation of **100 Euro** on two campus projects. For this purpose, each group member can make an individual proposal for the allocation. The result of the election is calculated by the **average** of all allocation proposals, i.e. the sum of the allocation proposals divided by the number of votes cast.
The two projects are a **Bike Workshop** (supervised self-help workshop) and the **Campus Garden** (opportunity to grow and harvest fruits and vegetables). Your allocation proposal therefore consists of two amounts of money: the amount to be donated to the Bike Workshop and the amount to be donated to the Campus Garden. Both amounts must add up to 100 Euro.
There will be no individual payment. The total amount of 100 Euro will be donated to the two projects according to the voting results. The election is held online and takes about 5 minutes. If you would like to participate in the election, please click here: https://www.survio.com/survey/d/d1

In addition to your group, there are other groups that will vote on the allocation of 100 Euro on the two projects. Every group consists of 17 participants and votes on the usage of 100 Euro each. For the realization of the allocation proposal of each group, at least one vote per group is required.
The election will run until 17.07.2017. If you are interested in the result of the election, please enter your e-mail address after the election. Your answers will be treated anonymously and cannot be assigned to the e-mail address. Once again, this is the link to the election: https://www.survio.com/survey/d/d1

Best regards,

the Chair of Economic Theory of KIT

Median Treatment

The result of the election is calculated by the **median** of all allocation proposals, i.e. the vote that is in the middle position after sorting all votes in ascending order is elected. If the number of votes cast is even, the median is calculated from the mean value of the two middle allocation proposals.

The Voting Process

See Fig. 15.

Fig. 15 Vote

The Questionnaire

See Figs. 16, 17, 18, 19 and 20.

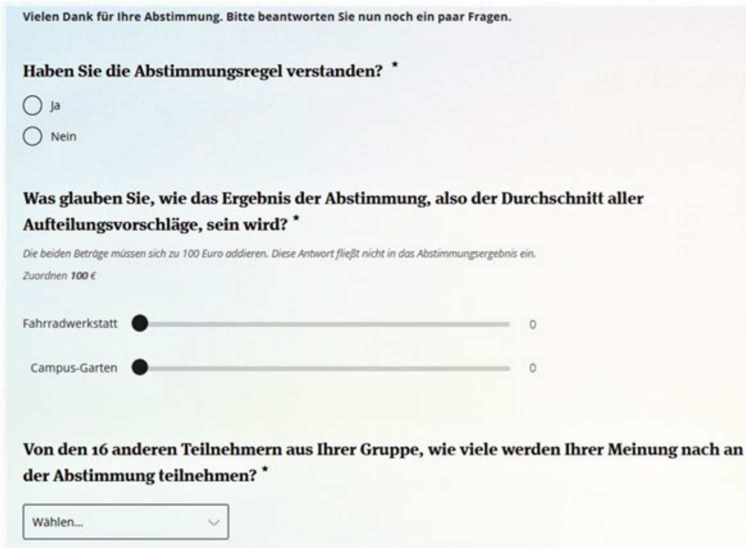
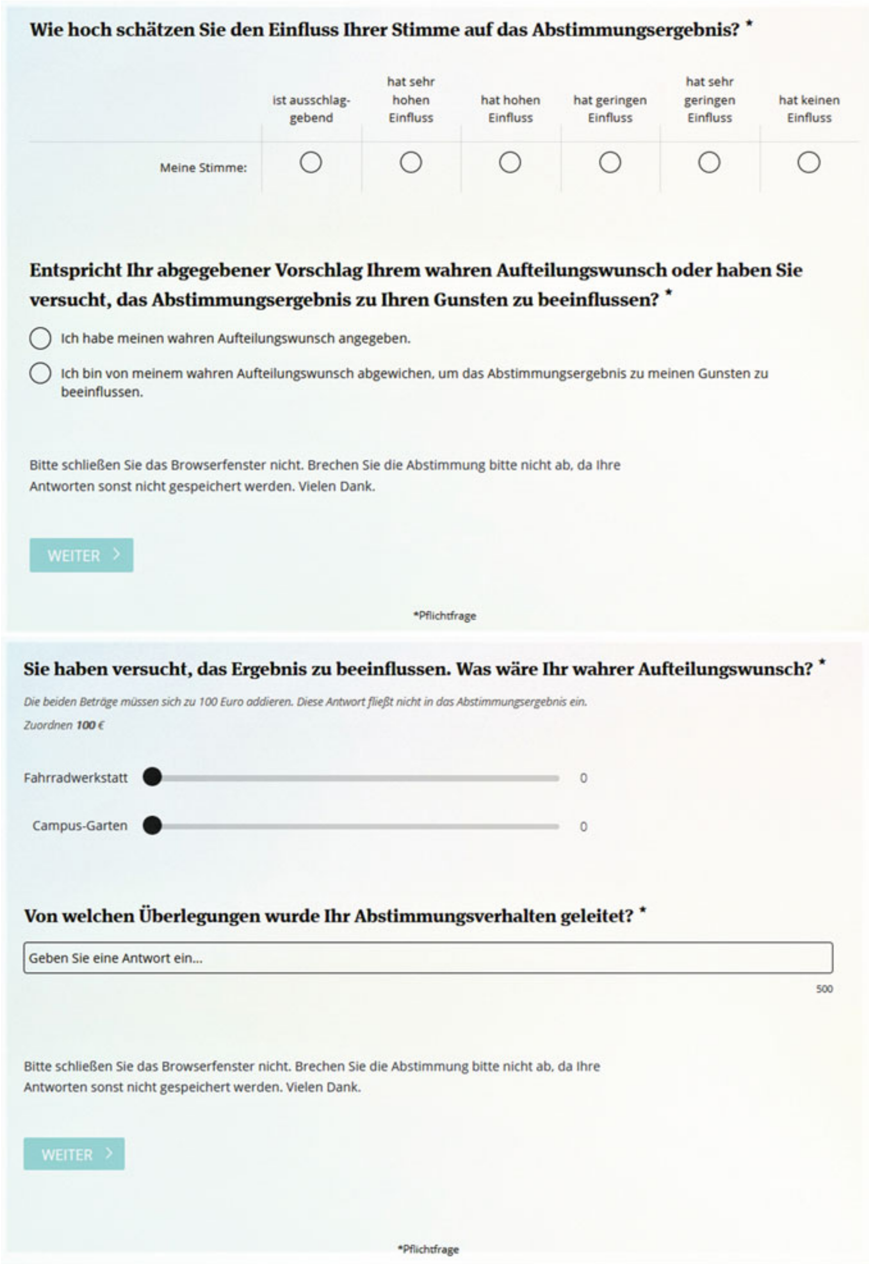


Fig. 16 Questions on the vote I



Wie hoch schätzen Sie den Einfluss Ihrer Stimme auf das Abstimmungsergebnis? *

	ist ausschlag-gebend	hat sehr hohen Einfluss	hat hohen Einfluss	hat geringen Einfluss	hat sehr geringen Einfluss	hat keinen Einfluss
Meine Stimme:	○	○	○	○	○	○

Entspricht Ihr abgegebener Vorschlag Ihrem wahren Aufteilungswunsch oder haben Sie versucht, das Abstimmungsergebnis zu Ihren Gunsten zu beeinflussen? *

○ Ich habe meinen wahren Aufteilungswunsch angegeben.

○ Ich bin von meinem wahren Aufteilungswunsch abgewichen, um das Abstimmungsergebnis zu meinen Gunsten zu beeinflussen.

Bitte schließen Sie das Browserfenster nicht. Brechen Sie die Abstimmung bitte nicht ab, da Ihre Antworten sonst nicht gespeichert werden. Vielen Dank.

WEITER >

*Pflichtfrage

Sie haben versucht, das Ergebnis zu beeinflussen. Was wäre Ihr wahrer Aufteilungswunsch? *

Die beiden Beträge müssen sich zu 100 Euro addieren. Diese Antwort fließt nicht in das Abstimmungsergebnis ein.

Zuordnen ***100 €***

Fahrradwerkstatt 0

Campus-Garten 0

Von welchen Überlegungen wurde Ihr Abstimmungsverhalten geleitet? *

Geben Sie eine Antwort ein...

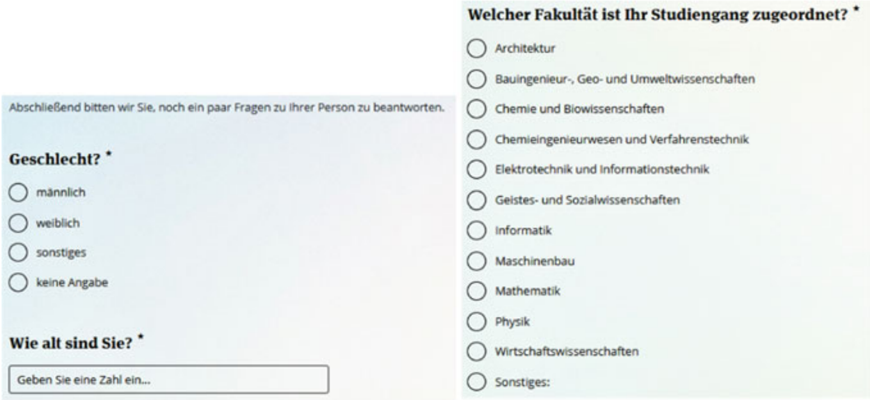
500

Bitte schließen Sie das Browserfenster nicht. Brechen Sie die Abstimmung bitte nicht ab, da Ihre Antworten sonst nicht gespeichert werden. Vielen Dank.

WEITER >

*Pflichtfrage

Fig. 17 Questions on the vote II



Abschließend bitten wir Sie, noch ein paar Fragen zu Ihrer Person zu beantworten.

Geschlecht? *

- männlich
- weiblich
- sonstiges
- keine Angabe

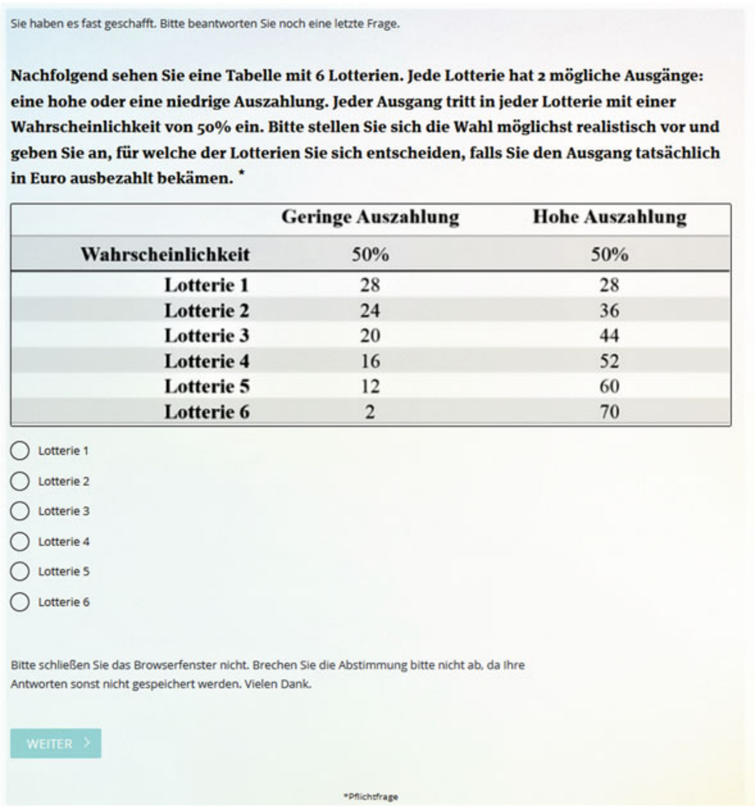
Wie alt sind Sie? *

Geben Sie eine Zahl ein...

Welcher Fakultät ist Ihr Studiengang zugeordnet? *

- Architektur
- Bauingenieur-, Geo- und Umweltwissenschaften
- Chemie und Biowissenschaften
- Chemieingenieurwesen und Verfahrenstechnik
- Elektrotechnik und Informationstechnik
- Geistes- und Sozialwissenschaften
- Informatik
- Maschinenbau
- Mathematik
- Physik
- Wirtschaftswissenschaften
- Sonstiges:

Fig. 18 Demographic questions



Sie haben es fast geschafft. Bitte beantworten Sie noch eine letzte Frage.

Nachfolgend sehen Sie eine Tabelle mit 6 Lotterien. Jede Lotterie hat 2 mögliche Ausgänge: eine hohe oder eine niedrige Auszahlung. Jeder Ausgang tritt in jeder Lotterie mit einer Wahrscheinlichkeit von 50% ein. Bitte stellen Sie sich die Wahl möglichst realistisch vor und geben Sie an, für welche der Lotterien Sie sich entscheiden, falls Sie den Ausgang tatsächlich in Euro ausbezahlt bekämen. *

	Geringe Auszahlung	Hohe Auszahlung
Wahrscheinlichkeit	50%	50%
Lotterie 1	28	28
Lotterie 2	24	36
Lotterie 3	20	44
Lotterie 4	16	52
Lotterie 5	12	60
Lotterie 6	2	70

- Lotterie 1
- Lotterie 2
- Lotterie 3
- Lotterie 4
- Lotterie 5
- Lotterie 6

Bitte schließen Sie das Browserfenster nicht. Brechen Sie die Abstimmung bitte nicht ab, da Ihre Antworten sonst nicht gespeichert werden. Vielen Dank.

WEITER >

*Pflichtfrage

Fig. 19 Risk preferences

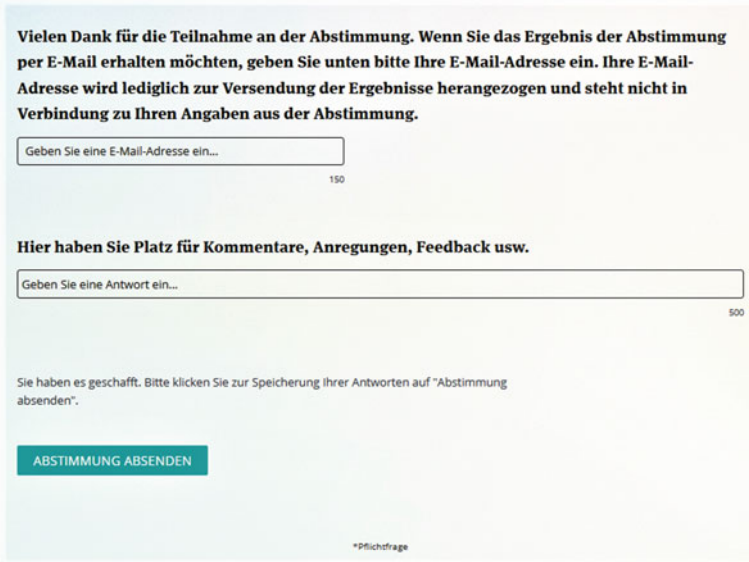


Fig. 20 Concluding remarks

References

Agranov, M., Goeree, J. K., Romero, J., & Yariv, L. (2017). What makes voters turn out: The effects of polls and beliefs. *Journal of the European Economic Association, 16*(3), 825–856.

Aldrich, J. H. (1993). Rational choice and turnout. *American Journal of Political Science, 37*(1), 246.

Arzumanyan, M., & Polborn, M. K. (2017). Costly voting with multiple candidates under plurality rule. *Games and Economic Behavior, 106*, 38–50.

Barzel, Y., & Silberberg, E. (1973). Is the act of voting rational? *Public Choice, 16*(1), 51–58.

Black, D. S. (1948). On the rationale of group decision-making. *Journal of Political Economy, 56*, 23–34.

Blais, A. (2000). *To vote or not to vote*. University of Pittsburgh Press.

Blais, A., Pilet, J.-B., van der Straeten, K., Laslier, J.-F., & Héroux-Legault, M. (2014). To vote or to abstain? An experimental test of rational calculus in first past the post and PR elections. *Electoral Studies, 36*, 39–50.

Block, V. (2014). *Single-peaked preferences—Extensions, empirics and experimental results* (Ph.D. thesis). Karlsruhe Institute of Technology (KIT).

Bock, O., Baetge, I., & Nicklisch, A. (2014). hroot: Hamburg registration and organization online tool. *European Economic Review, 71*, 117–120.

Börgers, T. (2004). Costly voting. *American Economic Review, 94*(1), 57–66.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization, 87*, 43–51.

Dave, C., Eckel, C. C., Johnson, C. A., & Rojas, C. (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty, 41*(3), 219–243.

Dhillon, A., & Peralta, S. (2002). Economic theories of voter turnout. *The Economic Journal, 112*(480), F332–F352.

Downs, A. (1957). *An economic theory of democracy*. Harper & Brothers.

Duffy, J., & Tavits, M. (2008). Beliefs and voting decisions: A test of the pivotal voter model. *American Journal of Political Science, 52*(3), 603–618.
Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior, 23*(4), 281–295.
Enos, R. D., & Fowler, A. (2014). Pivotality and turnout: Evidence from a field experiment in the aftermath of a tied election. *Political Science Research and Methods, 2*(2), 309–319.
Ferejohn, J. A., & Fiorina, M. P. (1974). The paradox of not voting: A decision theoretic analysis. *American Political Science Review, 68*(2), 525–536.
Ferejohn, J. A., & Fiorina, M. P. (1975). Closeness counts only in horseshoes and dancing. *The American Political Science Review, 69*(3), 920–925.
Foster, C. B. (1984). The performance of rational voter models in recent presidential elections. *The American Political Science Review, 78*(3), 678–690.
Gelman, A., King, G., & Boscardin, W. J. (1998). Estimating the probability of events that have never occurred: When is your vote decisive? *Journal of the American Statistical Association, 93*(441), 1–9.
Grillo, A. (2017). Risk aversion and bandwagon effect in the pivotal voter model. *Public Choice, 172*(3–4), 465–482.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review, 92*(5), 1644–1655.
Krasa, S., & Polborn, M. K. (2009). Is mandatory voting better than voluntary voting? *Games and Economic Behavior, 66*(1), 275–291.
Kurz, S., Maaser, N., & Napel, S. (2017). On the democratic weights of nations. *Journal of Political Economy, 125*, 1599–1634.
Ledyard, J. O. (1984). The pure theory of large two-candidate elections. *Public Choice, 44*(1), 7–41.
Levine, D. K., & Palfrey, T. R. (2007). The paradox of voter participation? A laboratory study. *American Political Science Review, 101*(1), 143–158.
Lindner, T. (2011). *Zur Manipulierbarkeit der Allokation öffentlicher Güter: Theoretische Analyse und Simulationsergebnisse*. KIT Scientific Publishing.
Louis, P., Núñez, M., & Xefteris, D. (2022). Trimming extreme reports in preference aggregation. In *University of Cyprus Working Papers in Economics 12-2019*. University of Cyprus Department of Economics.
Luce, R. D., & Raiffa, H. (1957). *Games and Decisions*. Wiley.
Marchese, C., & Montefiori, M. (2011). Strategy versus sincerity in mean voting. *Journal of Economic Psychology, 32*(1), 93–102.
Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice, 35*(4), 437–455.
Müller, M. (2022). *Strategy-proofness and participation in collective decisions* (Ph.D. Dissertation). Karlsruhe Institute of Technology. (KIT)
Müller, M., & Puppe, C. (2020). Strategy-proofness and responsiveness imply minimal participation. In *KIT Working Paper Series in Economics* (Vol. 138).
Osborne, M. J., Rosenthal, J. S., & Turner, M. A. (2000). Meetings with costly participation. *American Economic Review, 90*(4), 927–943.
Palfrey, T. R., & Rosenthal, H. (1983). A strategic calculus of voting. *Public Choice, 41*(1), 7–53.
Palfrey, T. R., & Rosenthal, H. (1985). Voter participation and strategic uncertainty. *American Political Science Review, 79*(1), 62–78.
Puppe, C., & Rollmann, J. (2021). Mean versus median voting in multi-dimensional budget allocation problems—A laboratory experiment. *Games and Economic Behavior, 130*, 309–330.
Renault, R., & Trannoy, A. (2005). Protecting minorities through the average voting rule. *Journal of Public Economic Theory, 7*(2), 169–199.
Reynaud, A., & Couture, S. (2012). Stability of risk preference measures: Results from a field experiment on French farmers. *Theory and Decision, 73*(2), 203–221.
Riker, W. H., & Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review, 62*(1), 25–42.

Rollmann, J. (2020). *Voting over resource allocation: Nash equilibria and costly participation* (Ph.D. thesis). Karlsruhe Institute of Technology (KIT)

Settle, R. F., & Abrams, B. A. (1976). The determinants of voter participation: A more general model. *Public Choice, 27*(1), 81–89.

Silberman, J., & Durden, G. (1975). The rational behavior theory of voter participation. *Public Choice, 23*(1), 101–108.

Tullock, G. (1967). *Toward a mathematics of politics*. University of Michigan Press.

The Office Makes the Politician

David Stadelmann

Abstract This paper explores behavioral changes regarding the political representation of voters by leveraging data from the Swiss Houses of Parliament. Although politicians in both houses are elected in the same geographical constituencies, public expectations would stipulate that politicians serving in the Council of States should focus relatively more on the preferences of their geographical constituency than on the nation, while the opposite should be the case for politicians serving in the National Council. I provide empirical evidence for such directional behavioral changes after politicians are elected. The evidence is consistent with the existence of an incentive effect of the office itself which acts on politicians to fulfill public expectations. Such an incentive effect, termed a "Thomas Becket incentive", would be complementary to the established relevance of elections as a selection and incentive device.

1 Introduction

Why do political representatives seek to correspond to voters' preferences and public expectations *after* elections? A common answer is that they do so because voters select politicians well in elections or, because (re-)election constraints incite politicians to consider voters' preferences next to their self-interest (e.g., Downs, 1957; Mueller, 2003; Persson & Tabellini, 2000). Accordingly, elections can be seen as a selection and incentive device (e.g., Lee, 2004). From this point of view, the electoral system shapes the way voters select politicians, and it influences electoral incentives (e.g., Cox, 1997; Duverger, 1954; Dow, 2011; Lijphart, 1994; Powell, 2000; Stadelmann et al., 2019).

The present contribution tries to extend the dichotomy of elections as a selection and incentive device by providing indicative evidence for a complementary view to explain why politicians may align their decisions with voters' preferences. I argue that serving in an elected office may act as an incentive in itself to fulfill public

D. Stadelmann (✉)
University of Bayreuth, Bayreuth, Germany
e-mail: david.stadelmann@uni-bayreuth.de

S. Kurz et al. (eds.), *Advances in Collective Decision Making*, Studies in Choice and Welfare, https://doi.org/10.1007/978-3-031-21696-1_22

expectations that are associated with the office.[1] This complementary view stresses the relevance of public expectations of the office itself as a potential behavioral incentive. I call this incentive a *Thomas Becket incentive*. Hillman (2009) provides a brief argument for such an incentive for politicians while Issing (2006) highlights such an incentive for central bankers.

The motivation behind the *Thomas Becket incentive* is the following: In 1162 Thomas Becket, a confidant and the Lord Chancellor of the English King Henry II, became the Archbishop of Canterbury. Henry II himself arguably desired Thomas Becket to hold this office. However, instead of supporting the King in his disagreement with Pope Alexander III in Rome, Thomas Becket dutifully served as an Archbishop and fulfilled the expectations of the new office he held. As the Archbishop, he was expected to represent the church and not the King. Indeed, he took the side of the church against Henry II, which resulted in a confrontational stance with the King. Eight years later, Henry II supposedly ordered his knights to free him of his former companion. Thomas Becket was slain in 1170. He became canonized by Pope Alexander III (see Barlow, 1986 for a biography on Thomas Becket). Hence, a *Thomas Becket incentive* should be seen as an incentive to fulfill the duty prescribed by the office, independently of selection or other incentives that may matter too, of course. The term *Thomas Becket incentive* shall therefore refer to a situation where officials detach themselves (to some degree) from their earlier office to satisfy the expectations of their new office.[2]

I provide indicative evidence for a *Thomas Becket incentive* in terms of congruence with voters of the same politicians who change office from the National Council to the Council of States of the Swiss Houses of Parliament. Swiss referenda allow me to observe the revealed preferences of the majority of each electoral district as well as of the national majority. Thereby, a natural measure of congruence of politicians with their constituency and the nation emerges (e.g., Stadelmann et al., 2013). The Swiss constitution and public expectations would stipulate that politicians in the Council of States *should* represent their geographical constituency rather than the nation. The opposite *should* hold for politicians in the National Council. These *normative* expectations are (of course) not legally binding.[3] Nevertheless, I show that they are fulfilled for politicians changing from the National Council to the Council of States: The same politician behaves differently with respect to preferences of constituents and nation when changing from one office to another, independently of selection or incentives induced through the electoral system as well as independently of personal characteristics and interest group affiliations.

[1] The view does not imply at all that selection and re-election incentives are irrelevant (see Portmann et al., 2022). But they do not need to be the only explanation for different behavior of politicians with respect to the representation of voters' preferences after politicians are elected to an office.

[2] In public and political discourses, references to the "dignity of an office" or the expectation that an "office will change the person" are frequently made.

[3] Article 161(1) of Federal Constitution forbids binding voting instructions.

The remainder of the paper is structured as follows: A description of the institutional setting is given in Sect. 2. Section 3 presents the empirical strategy and data. Section 4 provides indicative evidence consistent with the existence of a *Thomas Becket incentive*. I offer concluding remarks in Sect. 5.

2 Institutional Setting

2.1 National Council and Council of States

Switzerland has a bicameral parliament comprised of a Lower House (National Council, "Nationalrat" in German) and an Upper House (Council of States, "Ständerat" in German). Politicians to both houses are elected in 26 geographical constituencies (electoral districts), called the Cantons. Elections for the National Council and the Council of States take place on the same date, since 1851 always on a Sunday in October. For the Council of States, there are additional run-off elections, usually in November.

Elected politicians in both offices serve for four-year terms. The National Council has 200 members, and the Council of States has 46 members. Politicians in the National Council are elected under a proportional electoral system, while politicians in the Council of States are elected under a two-round majority-plurality system.[4] Apart from the electoral system, formal election requirements and prerogatives in the two offices are identical. The National Council and the Council of States have the same legislative power. Members of both elected offices decide on exactly the same laws and constitutional amendments. Legislative proposals have to be approved by majorities of both offices. As the Council of States is smaller, serving there is usually regarded as more prestigious.

Final roll call votes take place at the end of a parliamentary session. They are recorded by an electronic voting system since 1996 for members of the National Council, see Portmann et al. (2012). Thus, 1996 is the year where the sample of observations starts. There has been no electronic voting system for the Council of States until 2014. Since winter 2006, a camera records its sessions (see Stadelmann et al., 2014, 2019). The camera recordings allow the identification of individual voting behavior also of the members of the Council of States.[5]

[4] Differences induced by the two electoral systems have been analyzed by Stadelmann et al. (2019), and recently by Portmann et al. (2022), where the former focuses on parties, and the latter contribution shows that selection does not matter for the representation of the geographical constituency.

[5] In some cases, individual votes cannot be observed due to a too slow movement of the camera during the voting phase (see discussion in the appendix to Stadelmann et al. 2019).

2.2 Referendum Decision

Switzerland features a system of direct democracy. Thus, parliamentary proposals do not directly turn into law. Citizens may challenge parliamentary decisions in a referendum. Referendum decisions permit voters to judge legislative proposals and rank them against the status quo. Thereby, referendum decisions can be seen as revealed preferences of voters for policies (see Brunner & Ross, 2010; Brunner et al., 2011; Frey, 1994; Hessami, 2016; Noam, 1980; Portmann et al., 2012). Different types of referendum instruments are explained in greater detail in Stadelmann et al. (2013), Portmann (2014), (2022), and the subsequent empirical analysis will always control for referendum types as most of the existing literature employing Swiss referenda does.

Swiss referendum choices are implemented and entail policy consequences. There are no quorum requirements and consequently no strategic incentives to abstain.[6] Referendum results are available for every geographical constituency and the nation as a whole. Thus, it is possible to analyze the preferences of the majority in each constituency as well as the preferences of the majority of the nation. Referendum decisions take place frequently (Stadelmann & Torgler, 2013). They cover a wide range of political issues.

All information on the topics and results of referendum decisions are provided by the Swiss Parliamentary Services. Referendum titles, official booklets, deliberations, etc., in three of the four official Swiss languages can be found on the Website of the Swiss Parliament (https://www.parlament.ch/de/services/volksabstimmungen, accessed April 10, 2022)

2.3 Congruence with Geographical Constituency and with the Nation

The law and constitutional proposals presented to voters in referendum decisions are word-for-word identical to the proposals on which politicians decided in their parliamentary roll call votes. Thereby, I can observe what individual politicians decide and what the majority in their geographical constituency, as well as the nation, decide on the identical policy proposal. By matching individual decisions of politicians in the National Council and the Council of States with referendum decisions of their constituency, I directly compare whether the choices of politicians correspond to the preferences of their geographical constituency as well as to the preferences of the nation.

To measure congruence with the geographical constituency, I analyze whether a politician in the National Council or the Council of States corresponds to the preferences of the majority of her geographical constituents or not (see, e.g., Stadel-

[6] Hizen (2021) provides a discussion on strategic abstention.

mann, 2013). Thereby, I obtain a binary, natural measure of congruence between an individual politician's choices in both offices and the revealed preferences of her geographical constituency, called *Congruence with geographical constituency*.

Similarly, the institutional setting allows measuring congruence with national preferences. As referendum decisions reveal the preferences of the constituency, they also reveal the preferences of the national majority of voters. By matching the parliamentary decisions of individual politicians in both elected offices with the revealed preferences of the national majority, I obtain a measure for congruence with the nation, called *Congruence with nation*.

Generally speaking, merging roll call votes and referendum decisions to obtain a measure of congruence can now be said to be a relatively established practice in literature (see, e.g., Barcelo, 2019; Brunner et al., 2013; Garrett, 1999; Giger & Kluver, 2016; Matsusaka, 2010, 2017; Portmann et al., 2012, 2022; Potrafke, 2013; Stadelmann et al., 2019). Using such a measure of congruence has relevant advantages as decisions of politicians and voters are observed on identical legislative proposals. Thereby, numerous difficulties regarding comparisons are avoided which normally arise when the political decisions and voters' preferences are measured on different scales (see Achen, 1977; Gerber & Lewis, 2004; Matsusaka, 2010; Powell, 2009). Nevertheless, such a direct measure of congruence can only be established in countries or polities with referenda. As the main intention of this paper is to highlight the potential existence of a *Thomas Becket incentive*, the external validity of the two measures of congruence that I employ is not a direct or fundamental issue. Moreover, *Congruence with geographical constituency* and *Congruence with nation* obtain some external validity as politicians do not know in advance what their geographical constituency or the national voter majority actually wants, such that they have to employ other means to *predict* voters' preferences when voting in parliament (see Brunner et al., 2013; Garrett, 1999; Stadelmann et al., 2013).

My two measures of congruence, such as *Congruence with geographical constituency* and *Congruence with nation*, are both specific for individual politicians. They will permit me to distinguish whether politicians who are elected from the National Council to the Council of States tend to align themselves *differently* with their geographical constituency relative to the nation after being elected to a different office.

3 Empirical Strategy

3.1 Beyond Electoral Incentives

Portmann et al. (2022) have shown that incentives do not only dominate selection but that there is no role for selection, at least in the Swiss context. However, it is unclear which specific type of incentive matters for politicians when representing voters' preferences. The present paper aims to provide evidence that politicians change their behavior with regard to voters' preferences *after* changing from one elected office to another elected office, *and* that this observable behavioral change is consistent with public expectations, independently of electoral selection and incentives of the respective office that the politicians hold. That is, I aim to explore the existence of a *Thomas Becket incentive*.

In the Swiss context, public expectations as well as electoral incentives predict that the politicians serving in the National Council *will* correspond to a lower degree to the preferences of their geographical constituency than politicians serving in the Council of States. The electoral districts for politicians of the National Council and the Council of States are identical. Thus, the electoral incentives and re-election constraints of politicians in both elected offices are to focus on their constituency. Or in other words, regarding their own individual utility, self-interested politicians have no reason to represent the nation but only their geographical constituency.

To provide indicative evidence of a *Thomas Becket incentive*, I will show that any behavioral change of politicians who change office goes beyond the incentives induced by the electoral system itself. Put differently, independently of the electoral incentives, politicians in the National Council should represent the national voters *relatively* more than their geographical constituency. Members of the Council of States should represent their geographical constituency *relatively* more than the nation.

Normatively, the Swiss constitution stipulates that politicians of the National Council are supposed to be representatives of the nation, while politicians of the Council of States represent their Canton, i.e., their geographical constituency. More precisely, Article 149(1) states that the National Council is composed of representatives of "the people" while Article 150(1) states that the Council of States is composed of representatives of "the cantons". This normative view is consistent with the German, French, and Italian translations for the names of the respective houses, e.g., the "Nationalrat" (National Council) refers to the "nation", while the "Ständerat" (Council of States) refers to the "states"/"cantons". Thus, the constitutional article appeals to the office itself. Evidently, any constitutional article is open to interpretation but public expectations toward members of the National Council and politicians of the Council of States are different.

3.2 Estimation Equation

As outlined above, I distinguish the variables *Congruence with geographical constituenc*y and C*ongruence with nation*. This allows me to investigate whether politicians who change from the National Council to the Council of States change their behavior with respect to representation of the constituency versus the nation after being elected, that is, I can exclude classical selection issues. I test whether the same politicians rather represent the preferences of their constituency as opposed to the preferences of the nation after changing from one office to the other. In practice, preferences of nation and constituencies correlate, i.e., Swiss citizens will tend to hold similar preferences regarding certain policies independently of which constituency they live in, but exceptions and diverging views occur.

To perform the analysis, I proceed as follows: The two variables *Congruence with geographical constituency* and C*ongruence with nation* jointly form a new single vector as a dependent variable called *Congruence with nation or constituency*, such that each individual legislative decision of a politician is associated with two levels of congruence. An indicator variable *representing geographical constituency* serves as an independent variable and equals 1 if the dependent variable measures congruence with the constituency for the respective politician. I estimate the following model with an interaction term between *politician changed office* and *representing geographical constituency*:

$$\begin{gathered}(\textit{Congruence with nation or constituency})_{ir} = \alpha + \beta_1(\textit{Politician changed office})_{ir} \\ +\beta_2(\textit{Representing geographical constituency})_{ir} \\ +\beta_3(\text{Politician changed office})_{ir}{}^{*}(\textit{Representing geographical constituency})_{ir} \\ +\gamma_i + \boldsymbol{X}_{ir}\delta + \epsilon_{ir}\end{gathered}$$

The index *i* denotes a politician, and *r* denotes the legislative decision, respectively, the referendum decision. *Representing geographical constituency* indicates whether the constituency is on average better represented than the nation. I control for time-variant legislator characteristics such as time served in parliament,[7] among others in the matrix $\boldsymbol{X}_{ir}$ (see Table 1 for descriptive statistics below). γ_i represents politician specific fixed effects. Additional fixed effects are entered in different regressions. The estimation is carried out as a linear probability model, and estimated standard errors are clustered for politicians. Employing a logit or probit model would yield qualitatively and quantitatively similar results.

The coefficients β_1 and β_3 are of interest for my interpretations: $\beta_1$indicates whether politicians who changed from the National Council to the Council of States (*politician changed office*) have generally higher congruence levels if *representing geographical constituency* is set to zero. The coefficient of the interaction term, β_3, indicates whether politicians changing office additionally increase (or decrease) their congruence levels by more closely representing the majority of their geographical constituency. As I include politician specific fixed effects, β_1 and β_3 allow me to investigate the change in a politician's behavior toward the geographical constituency and the nation, respectively, when the politician changes from the National Council to the Council of States.

In line with Portmann et al. (2022), I expect $\beta_1 > 0$ due to a standard incentive effect, that is, politicians who change office from the National Council to the Council of States have generally higher congruence levels after their change due to electoral

[7] The relevance of career paths has been highlighted, for example, by Pickard (2021).

incentives. More importantly, I expect $\beta_3 > 0$ due to a *Thomas Becket incentive*, that is, I expect a *differentially* higher congruence with the geographical constituency than the nation *after* politicians have changed office.

3.3 Identifying Assumptions

The empirical setting outlined above is intended to capture the *differential* change in the behavior of politicians toward their geographical constituency' preferences and the national preferences when changing from the National Council to the Council of States. Thus, the same person must be observed in her respective roles in both offices. The setting ensures this as politician specific fixed effects are employed such that the same person is indeed observed in both elected offices. However, other identification assumptions need to be defended.

I do not observe that politicians decide on the *identical* legislative proposal *twice* (once in the National Council and once in the Council of States). Politicians of the National Council and the Council of States are only elected in the same constituencies, and they decide on identical proposals that affect their geographical constituency and the nation. Looking at the policy proposals that are decided in a referendum, there is no evident reason to believe that the National Council and the Council of States treat them any differently *after* politicians changed office, especially when also controlling for time fixed effects. There has been no institutional change over time that would affect the two offices differently regarding how politicians behave toward their constituency. Thus, politicians remaining in the National Council can be seen as a reasonable comparison group for those that change to the Council of States.

Elections lead to changes in the composition of both offices: Some politicians who previously served in the National Council may not serve anymore. They may have withdrawn from Parliament or they were not reelected. Note that the use of politician specific fixed effects in the empirical specification considers all time-invariant characteristics of old *and* new politicians. Consequently, factors like gender (we observe an increase in the number of women in parliament over time), party affiliations (there has been a rise in right party affiliations of politicians over time), valance (there has been a stronger focus on personal traits), etc., are captured, even if the composition of both elected offices changes after elections. It may be argued that newly elected politicians in both offices need a learning period. To account for such a learning period, I include the variables *first year in parliament* and *first term in parliament*. Of course, I also control for *time in parliament*.

Parties play a role even if they are not allowed to legally instruct politicians how to vote. The influence of parties is captured indirectly through politician specific fixed effects. The relevance of party recommendations is thereby controlled for as the same politician from the same party is observed in both offices. There are some politicians who changed parties during the period of analysis. Excluding those politicians does not change the results or interpretations.

The size of the National Council and the Council of States are different such that debates and, consequently, the voting behavior of politicians may be affected. Indeed, in the Swiss context, the Council of States is sometimes seen as a "Chambre de Réflexion" where debates are less heated. Differences regarding the style of debates may induce generally higher (or lower) congruence levels of all politicians in the Council of States. However, they do not explain why congruence with the geographical constituency should increase more strongly than congruence with the nation, which is the expectation that $\beta_3 > 0$ due to the existence of a *Thomas Becket incentive*.

While identification assumptions cannot be proven, there are strong reasons in my view to believe that there is no selection of specific referendum decisions or groups of politicians, which would affect the same politicians' congruence with their geographical constituency vs the nation *differently*, when changing from the National Council to the Council of States.

3.4 Data

The dataset explored in this paper corresponds to the main dataset recently employed by Portmann et al. (2022). It comprises 156 legislative decisions with subsequent referenda for the years 1996 to 2015, covering the 45th to the 49th legislature of the Swiss Houses of Parliament. The sample of politicians who have served in the National Council includes 547 politicians, who made a total of 28,308 individual legislative decisions. This leads to (2 * 28,308 =) 56,616 observed matches between politicians' decisions and their geographical constituents (*Congruence with geographical constituency*) and the national majority (*Congruence with nation*).

Out of 547 politicians, 32 have served both in the National Council and the Council of States over the time period analyzed. After having served in the National Council, these politicians were elected to the Council of States. I observe their congruence toward their geographical constituency as well as the nation in both offices. The sample of politicians who changed elected office is comprised of 32.3% women, and their average time in Parliament is 8.9 years (time in National Council and Council of States combined). This broadly corresponds to the overall sample average where 26.1% of decisions are made by female politicians, and the average time being in Parliament when deciding on legislative proposals is 6.5 years (see Table 1).

Time-invariant controls for individual politicians (e.g., gender, education, etc.) are not included in the dataset and the estimations as they are automatically captured by politician specific fixed effects. Time-variant politician specific controls were collected from the official homepage of the Swiss Parliament where short biographies for each politician are available (see Portmann et al., 2012, who started the data collection, and Portmann 2014 for detailed descriptions). Interest group affiliations[8] and

[8] Swiss legislators must disclose their affiliations with interest groups such as executive board seats in companies and foundations, committee memberships, counseling activities, and other activities

Table 1 Descriptive statistics of main dataset

	Obs.	Mean	SD	Min.	Max.
Dependent variable					
Congruence of politicians with nation and constituency	56,616	0.6474	0.4778	0	1
Main variable of interest					
Politician changed office	56,616	0.0272	0.1626	0	1
Representing geographical constituency	56,616	0.5	0.5	0	1
Time-variant politician specific controls					
Time in parliament (years)	56,616	6.498	5.1448	0.0066	34.283
First year in parliament	56,616	0.1296	0.3359	0	1
First term in parliament	56,616	0.3853	0.4867	0	1
Interest group affiliations					
Number of section interest groups	56,616	2.9716	3.9081	0	59
Number of cause interest groups	56,616	2.5019	3.1104	0	28
Number of regional interest groups	56,616	0.1543	0.5306	0	8

Notes Unweighted statistics represented. See Portmann et al. (2022) for sources and details.

their classification stem from Portmann et al. (2022) which are based on information from the Swiss Parliamentary Services and a large additional data collection effort by the authors. All time-variant descriptive statistics are provided in Table 1.

4 Public Expectations and Behavior in Office

4.1 Serving in Office as an Incentive in Itself

Table 2 presents the main results which can be interpreted as consistent with the view that a Thomas Becket incentive acts on politicians.

Specification (1) shows that politicians who are elected from the National Council to the Council of States represent their geographical constituency as well as the nation more closely than politicians who remain in the National Council. The variable *Politician changed office* has a positive and statistically significant coefficient, that is, $\beta_1 > 0$. The estimates suggest an increase in overall congruence of about 4.9% points when politicians change from one elected office to another. This effect is consistent with Portmann et al. (2022) who highlights the relevance of incentives over selection. The variable *representing geographical constituency* is not statistically significant, suggesting that politicians who remain in the National Council represent

for lobby groups according to the federal law (Art. 11, Parlamentsgesetz). The Swiss Parliamentary Services collect this information (see Gava, 2017, Péclat & Puddu, 2017). The register frequently attracts media attention. Following the literature, we group them into sectional (*#Sectional*) and cause groups *(#Cause*) (see Giger & Kluver, 2016; Stadelmann et al., 2016).

Table 2 Representing geographical constituencies more closely than the nation when changing office—A Thomas Becket incentive in action

	(1)	(2)	(3)
Dependent variable: Congruence of politicians with nation and constituency			
Politician changed office	**0.0495**** **(0.0234)**	**0.0436*** **(0.0234)**	**0.0524**** **(0.0225)**
Representing geographical constituency	0.0033 (0.0036)	0.0033 (0.0036)	0.0033 (0.0037)
Politician changed office * Representing geographical constituency	**0.0383*** **(0.0206)**	**0.0388*** **(0.0206)**	**0.0392*** **(0.0207)**
Politician FEs	Yes	Yes	Yes
Referendum type FEs	No	Yes	Yes
Time FE	No	Yes	Yes
Time-variant politician specific controls	No	Yes	Yes
Interest group affiliations	No	No	Yes
n. Obs.	56,616	56,616	56,616
R^2	0.1053	0.1056	0.106

Notes ***, **, and * indicate a mean significance level of <1%, 1–5%, and 5–10%, respectively. Linear probability models are estimated and standard errors are clustered for MPs. All estimations include an intercept. The indicator variable "politician changed office" takes the value of 1 if an individual MP changes from the National Council to the Council of States. The identifier "representing geographical constituency" takes the value of 1 if the dependent variable refers to the preferences of geographical constituency and 0 if it refers to the preferences of the nation. "Politician FEs" capture all time-invariant politician specific heterogeneity (e.g., gender, education, etc.), "Referendum type FEs" capture all referendum specific heterogeneity (e.g., turnout, salience, etc.), and "Time FEs" capture all parliamentary year specific heterogeneity (e.g., composition of parliament, parties in parliament). "Time-variant controls" include "time in parliament", "time in parliament squared", "first year in parliament", and "first term in parliament". "Interest group affiliations" include the "number of section interest groups", "number of cause interest groups", and "number of regional interest groups"

the nation and the constituency equally well.[9] The interaction term between *politician changed office* and *representing geographical constituency* is positive, that is, $\beta_3 > 0$. It reveals that politicians who changed elected office have after their change even higher congruence levels with the preferences of their constituency, that is, they represent their geographical constituency better than the nation. Thus, they put an additional emphasis on their geographical constituency, thereby corresponding to their constitutional task and the public expectations. The *additional* congruence with their constituency's preferences is with 3.8% points quantitatively relevant.

[9] Interestingly, the insignificant coefficient of the variable *representing geographical constituency* might be interpreted as consistent with a Thomas Becket incentive too: politicians from the National Council fulfill their constitutional task to some degree, i.e., they do not focus more strongly on the constituency in which they are elected than on the nation. If a Thomas Becket incentive played no role, one might expect that the coefficient for representing geographical constituency should be positive and significant. An insignificant coefficient indicates that politicians represent national citizens who do not vote for them.

It is consistent with the view that politicians change their behavior after changing elected office in a directional way according to the public expectations: It is consistent with the existence of a Thomas Becket incentive.

In specifications (2) and (3), referendum type fixed effects, time fixed effects, time-variant politician specific controls, and interest group affiliations are added. All results remain statistically significant and quantitatively similar, that is, $\beta_1 > 0$ and $\beta_3 > 0$. Thus, if a politician changes from the National Council to the Council of States, congruence in general increases as would be expected due to the electoral system, $\beta_1 > 0$. At the same time, the politician who changes office puts relatively more weight on the preferences of the geographical constituency than the nation, $\beta_3 > 0$, supporting the view that a Thomas Becket incentive is relevant.

4.2 No Changes Prior to Being Elected

Table 3 provides further indirect evidence for a Thomas Becket incentive. Here, I restrict the analysis to two separate subsamples, namely a sample of members of the National Council only (specifications 1 and 2) and the Council of States only (specifications 3 and 4). As the focus is not anymore on the act of changing office, the indicator variable is now called *Politician is office changer*[10], that is, it indicates for specifications (1) and (2) whether a politician will be elected from the National Council to the Council of States (comparison group equals other members of the National Council). For specification (3) and (4), it indicates whether a politician was elected from the National Council to the Council of States (comparison group equals other members of the Council of States).

If a Thomas Becket incentive matters, I should not observe that politicians who will change office represent their geographical constituency differently from the nation in comparison to other politicians of the National Council. Thus, the interaction between politician is office changer, and representing geographical constituency should be statistically insignificant and close to zero. This is precisely what is observed in specifications (1) and (2). Thus, future members of the Council of States—while still serving in the National Council—behave statistically identically to other members of the National Council. They do not show any higher congruence level with their constituency that elects them to the Council of States. Thus, they fulfill the task of their current office in a similar way as other members of the National Council.

Specifications (3) and (4) show that once *having been elected* to the Council of States, politicians have lower congruence levels with the nation (between −3.6 and −3.2% points) than politicians who were not in the National Council before serving in the Council of States. The results are marginally statistically significant at the 10-%-level. At the same time, the interaction term between politicians who have changed office and the indicator for representing the geographical constituency is

[10] The inclusion of politician specific fixed effects is impossible. I include party fixed effects to hold constant for ideology instead.

Table 3 No effects prior to being elected: Evidence prior to and after change of office

	(1)	(2)	(3)	(4)
Dependent variable: Congruence of politicians with nation and constituency				
Politician is office changer	**0.0101 (0.0103)**	**0.0090 (0.0097)**	**−0.0320* (0.0194)**	**−0.0361* (0.0217)**
Representing geographical constituency	0.0027 (0.0036)	0.0028 (0.0037)	−1.4e−15 (0.0123)	−1.3e−15 (0.0123)
Politician is office changer * Representing geographical constituency	**0.0104 (0.0217)**	**0.0106 (0.0219)**	**0.0416* (0.0245)**	**0.0418* (0.0249)**
Party FEs	Yes	Yes	Yes	Yes
Referendum type FEs	No	Yes	No	Yes
Time FE	No	Yes	No	Yes
Time-variant politician specific controls	No	Yes	No	Yes
Interest group affiliations	No	Yes	No	Yes
n. Obs.	55,078	55,078	4172	4172
R^2	0.0886	0.0891	0.092	0.0926
Dataset	**Lower house members**		**Upper house members**	

Notes ***, **, and * indicate a mean significance level of <1%, 1–5%, and 5–10%, respectively. Linear probability models are estimated, and standard errors are clustered for MPs. All estimations include an intercept. The indicator variable "politician changed office" takes the value of 1 if an individual MP changes from the National Council to the Council of States. The identifier "representing geographical constituency" takes the value of 1 if the dependent variable refers to the preferences of geographical constituency and 0 if it refers to the preferences of the nation. "Politician FEs" capture all time-invariant politician specific heterogeneity (e.g., gender, education, etc.), "referendum type FEs" capture all referendum specific heterogeneity (e.g., turnout, salience, etc.), and "time FEs" capture all parliamentary year specific heterogeneity (e.g., composition of parliament, parties in parliament). "Time-variant controls" include "time in parliament", "time in parliament squared", "first year in parliament", and "first term in parliament". "Interest group affiliations" include the "number of section interest groups", "Number of cause interest groups", and "number of regional interest groups"

positive and marginally significant at the 10-%-level, that is, politicians who changed elected office have higher congruence levels with their constituency of about 4.16–4.18% points. Thus, if anything, the results in specifications (3) and (4) suggest that politicians who have changed elected office may be more eager in representing their constituency rather than the nation in comparison to other members of the Council of States. This result suggests that they may take their constitutional task of representing the constituency more to heart consistent with a Thomas Becket incentive. Adding up the two coefficients results in a statistically insignificant overall difference between politicians who changed elected office and those in the Council of States who did not which is close to zero.

5 Conclusions

This article endeavours to provide a complementary view to the prevalent argument that elections act as selection or incentive devices to align politicians with the preferences of voters. I provide empirical evidence consistent with the view that politicians seek to correspond to public expectations *after* achieving a political office. The results support the view that public expectations of holding an office might be of relevance, a topic that has been largely neglected in the economic literature. Politicians who are elected to office may also correspond to the public expectations of the office they hold independently of selection and electoral incentives.

The indicative evidence that I provide is consistent with the story of the historical person of Thomas Becket, who dutifully served in the office as Lord Chancellor and later in the office of Archbishop. Thomas Becket changed his behavior to fulfill the respective office, the office itself did not change. My results are supportive of the existence of a *Thomas Becket incentive* in politics, such that one may argue that the office makes the politician. A *Thomas Becket incentive* relates to how rules of appropriateness and rule-driven behavior matter in the context of formally organized political institutions, see March and Olsen (2008). Evidently, this view does not exclude that standard election incentives and selection are relevant too. Politicians in my analysis do not perfectly satisfy the expectations of their office, that is, they do not perfectly correspond to the preferences of their geographical constituency after changing office.

At a first glance, the existence of a *Thomas Becket incentive* suggests that the elections as a selection device are less important. That interpretation would be premature at this point. In fact, it might be argued that voters specifically select politicians with the expectation that they will change *once* in office, that is, elections serve as a selection device for politicians who fulfill the tasks of the office. Once these politicians change to another office, they subsequently fulfill the task of the other office. The empirical setting analyzed in this paper does not allow me to exclude such a type of selection effect. Still, I would consider this line of argumentation as somewhat semantic: Practically, it means that selection is based on a *belief* of voters in the existence of a *Thomas Becket incentive* of the office itself. More generally, this debate may be related to classical debates in political representation such as the delegate vs trustee view (e.g., Pitkin, 1967).

In policy discussions, we tend to hear arguments of the type "once in office, the person will adapt". These arguments are essentially referring to a potential *Thomas Becket incentive*. The results are, to my knowledge, the first to provide empirical evidence which can be seen as consistent with the existence of such a type of incentive for congruence of politicians with the preferences of their constituency. I believe that they provide a complement to the dichotomy of elections as a selection and incentive device. It is worthwhile to extend the existing dichotomy by allowing for the potential existence of a *Thomas Becket incentive*, to collect further evidence for its existence, and to integrate it into future theoretical research on institutions and electoral systems.

References

Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science, 21*(4), 805–815.

Barceló, J. (2019). The clarity of the majority's preference moderates the influence of lobbying on representation. *Party Politics, 25*(3), 349–357.

Barlow, F. (1986). *Thomas Becket*. University of California Press.

Brunner, E. J., Ross, S. L., & Washington, E. (2011). Economics and policy preferences: Causal evidence of the impact of economic conditions on support for redistribution and other ballot proposals. *Review of Economics and Statistics, 93*(3), 888–906.

Brunner, E. J., & Ross, S. L. (2010). Is the median voter decisive? Evidence from referenda voting patterns. *Journal of Public Economics, 94*(11–12), 898–910.

Brunner, E. J., Ross, S. L., & Washington, E. L. (2013). Does less income mean less representation? *American Economic Journal: Economic Policy, 5*(2), 53–76.

Cox, G. W. (1997). *Making votes count: Strategic coordination in the world's electoral systems*. Cambridge University Press.

Dow, J. K. (2011). Party-system extremism in majoritarian and proportional electoral systems. *British Journal of Political Science, 41*, 341–361.

Downs, A. (1957). *An economic theory of democracy*. Harper & Brothers.

Duverger, M. (1954). *Political parties: Their organization and activity in the modern state*. Wiley.

Frey, B. S. (1994). Direct democracy: Politico-economic lessons from Swiss experience. *American Economic Review, 84*(2), 338–342.

Garrett, T. A. (1999). A test of shirking under legislative and citizen vote: The case of state lottery adoption. *Journal of Law and Economics, 42*(1), 189–208.

Gava, R., Varone, F., Mach, A., Eichenberger, S., Christe, J., & Chao-Blanco, C. (2017). Interests groups in parliament: Exploring MPs' interest affiliations (2000–2011). *Swiss Political Science Review, 23*(1), 77–94.

Gerber, E. R., & Lewis, J. B. (2004). Beyond the median: Voter preferences, district heterogeneity, and political representation. *Journal of Political Economy, 112*(6), 1364–1383.

Giger, N., & Klüver, H. (2016). Voting against your constituents? How lobbying affects representation. *American Journal of Political Science, 60*(1), 190–205.

Hessami, Z. (2016). How do voters react to complex choices in a direct democracy? Evidence from Switzerland. *Kyklos, 69*(2), 263–293.

Hillman, A. L. (2009). *Public finance and public policy: Responsibilities and limitations of government*. Cambridge University Press.

Hizen, Y. (2021). A referendum experiment with participation quorums. *Kyklos, 74*(1), 19–47.

Issing, O. (2006). Central bank independence—Economic and political dimensions. *National Institute Economic Review, 196*(1), 66–76.

Lee, D. S., Moretti, E., & Butler, M. J. (2004). Do voters affect or elect policies? Evidence from the U.S. House. *The Quarterly Journal of Economics, 119*(3), 807–859.

Lijphart, A. (1994). *Electoral systems and party systems*. Oxford University Press.

March, J. G., & Olsen, J. P. (2008). The logic of appropriateness. In R. E. Goodin, M. Moran, & M. Rein (Eds.), *The Oxford handbook of public policy*. Oxford University Press.

Matsusaka, J. G. (2010). Popular control of public policy: A quantitative approach. *Quarterly Journal of Political Science, 5*(2), 133–167.

Matsusaka, J. G. (2017). When do legislators follow constituent opinion? Evidence from matched roll call and referendum votes. In *New Working Paper Series No. 9*. Stigler Center for the Study of the Economy and the State, University of Chicago Booth School of Business.

Mueller, D. C. (2003). *Public choice* (3rd ed.). Cambridge University Press.

Noam, E. M. (1980). The efficiency of direct democracy. *Journal of Political Economy, 88*(4), 803–810.

Péclat, M., & Puddu, S. (2017). Swiss politicians' ties: A comprehensive dataset. *Swiss Political Science Review, 23*(2), 175–190.

Persson, T., & Tabellini, G. (2000). *Political economics: Explaining economic policy*. MIT Press.
Pickard, H. (2021). The impact of career politicians: Evidence from US Governors. *Kyklos, 74*(1), 103–125.
Pitkin, H. F. (1967). *The concept of representation*. University of California Press.
Portmann, M. (2014). *Parliamentary representation of citizens' preferences: Explaining the differences between parliamentarians' votes and popular referendum results, Vol. 15 of Neue Studien zur politischen Ökonomie*. Nomos.
Portmann, M., Stadelmann, D., & Eichenberger, R. (2012). District magnitude and representation of the majority's preferences: Quasi-experimental evidence from popular and parliamentary votes. *Public Choice, 151*(3–4), 585–610.
Portmann, M., Stadelmann, D., & Eichenberger, R. (2022). Incentives dominate selection—Chamber legislators are driven by electoral rules and voter preferences. *Journal of Economic Behavior & Organization, 193*, 353–366.
Potrafke, N. (2013). Evidence on the political principal-agent problem from voting on public finance for concert halls. *Constitutional Political Economy, 24*(3), 215–238.
Powell, G. B. (2000). *Elections as instruments of democracy: Majoritarian and proportional visions*. Yale University Press.
Powell, G. B. (2009). The ideological congruence controversy: The impact of alternative measures, data, and time periods on the effects of election rules. *Comparative Political Studies, 42*(12), 1475–1497.
Stadelmann, D., Portmann, M., & Eichenberger, R. (2013). Quantifying parliamentary representation of constituents' preferences with quasi-experimental data. *Journal of Comparative Economics, 41*(1), 170–180.
Stadelmann, D., Portmann, M., & Eichenberger, R. (2014). Full transparency of politicians' actions does not increase the quality of political representation. *Journal of Experimental Political Science, 1*, 16–23.
Stadelmann, D., Portmann, M., & Eichenberger, R. (2016). How lobbying affects representation: Results for majority-elected politicians. *The B.E. Journal of Economic Analysis and Policy, 16*(4), 20160040
Stadelmann, D., Portmann, M., & Eichenberger, R. (2019). Preference representation and the influence of political parties in majoritarian vs. proportional systems: An empirical test. *British Journal of Political Science, 49*(1), 181–204
Stadelmann, D., & Torgler, B. (2013). Bounded rationality and voting decisions over 160 years: Voter behavior and increasing complexity in decision-making. *PLoS ONE, 8*(12), e84078.