Strategyproof social choice for restricted domains

Botan, S.

Publication date
2021
Document Version
Final published version

Institute for Logic, Language and Computation

Strategyproof Social Choice for Restricted Domains

Sirin Botan

Strategyproof Social Choice for Restricted Domains

Sirin Botan

Strategyproof Social Choice for Restricted Domains

Sirin Botan

Strategyproof Social Choice for Restricted Domains

ILLC Dissertation Series DS-2021-11

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

These investigations were supported, in part, by the Dutch Research Council (NWO) in the context of the *Collective Information* project funded under the VICI scheme (grant number 639.023.811).

Cover design by Karin Fischnaller.
Printed and bound by GVO drukkers & vormgevers B.V.

ISBN: 978-94-6332-803-6

Strategyproof Social Choice for Restricted Domains

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 26 november 2021, te 13.00 uur

door

Sirin Botan

geboren te Oslo

Promotiecommisie

Promotor:	prof. dr. U. Endriss	Universiteit van Amsterdam
Co-promotor:	dr. R. de Haan	Universiteit van Amsterdam
Overige leden:	prof. dr. F. Brandt	Technische Universität München
	prof. dr. P. Faliszewski	Akademia Górniczo-Hutnicza im. Stanisława Staszica
	dr. D. Grossi	Universiteit van Amsterdam
	prof. dr. ing. R.A.M. van Rooij	Universiteit van Amsterdam
	prof. dr. G. Schäfer	Universiteit van Amsterdam
	prof. dr. M. Slavkovik	Universitetet i Bergen

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Contents

Acknowledgments

In the past few years, whenever I was feeling less than motivated, I would imagine writing my acknowledgements—not always fully convinced I would ever need to do it for real. When the time came, I put off writing them until the very last minute. This, in combination with my less than stellar memory, pretty much guarantees I will have forgotten to mention at least one person. Rest assured I probably like you just fine despite the omission.

First, I want to profusely thank my supervisor Ulle Endriss. I had a pretty rough go of it the first few years of my PhD and I would not be sitting here writing these acknowledgements if it were not for the encouragement I received from Ulle. It's been an honour and a pleasure to learn from him these past seven (!!) years. I could go on forever about what an excellent supervisor Ulle is, and how the standards he sets for himself and those around him have made me better at just about everything (except baking). Most importantly, though, Ulle is just the coolest, funniest guy around and I will miss him being part of my workdays.

Second, I want to thank Ronald de Haan, my second supervisor, for his constant support, helpful feedback, and endless supply of puns. Ronald is one of the most caring people I know, and I feel so fortunate to have him in my corner.

I also owe a lot to the ILLC office for their help throughout the years with both big problems and small. In particular, I want to thank Tanja and Jenny who, in addition to being excellent at their jobs, have also supported me on a more personal level during my years here.

Six very nice people have graciously agreed to be part of my committee, for which I am very grateful. Thank you to Felix Brandt, Piotr Faliszewski, Davide Grossi, Robert van Rooij, Guido Schäfer, and Marija Slavkovik. I deeply admire and am intimidated by you all.

A huge part of my life in Amsterdam has been the COMSOC group: Ronald, Zoi, Arthur, Simon, Adrian, Arianna, Julian, Jan, and Oliviero. Thank you all for making every group meeting something to look forward to. As the saying goes, the real PhD is the friends you made along the way. Zoi, I don't think I can put into words how much your friendship has meant to me over the past years. It has been so fun sharing this experience with you. Arthur, thank you for letting me keep distracting you in the office, and contributing with some stellar distractions yourself. I am so proud of everything you've achieved, Arthur, truly. How lucky for me that I got to grow alongside you. Simon, thank you for making yourself my friend as quickly as you did. You have been the calm in many of my storms this past year and for that I am beyond grateful. Arianna, I really can't imagine how this past year would have been without your presence across the hall. Thank you for being my influencer and for giving the best advice.

To Anna, Arianna, and Lwenn, thank you for providing daily magic during the pandemic. There really is nothing like the warmth of a good group chat. To Julian, thank you for the years of Wednesday crosswords, culminating in our tournament entry. It's been great sharing this hobby with you.

Finally, thank you to my friends and family from all over the world for providing encouragement and distraction when it was sorely needed. In particular I want to say zor spas to my dad who has been my biggest fan since day one, and my mum who has always had my back (and made really sure I never forgot). I also want to extend a personal thank you to my friend Kat, who has been unwavering in her support. I don't have a sister, but I imagine her friendship comes as close to it as I'd willingly endure.

A special thanks goes to Karin who made my retro dreams come true by designing the cover of this book.

Chapter 1

Introduction

"The board decided...", "The council does not approve...", "Our preference is...". Collective decisions are so ubiquitous in most of our daily lives that we can read statements like these without batting an eye. But there is a lot lurking under the surface when we talk about collective preferences or opinions. What does it mean, for example, that the council does not approve of something if they are a group made up of possibly disagreeing individuals? These questions, and similar ones, are explored in the field known as *social choice theory* (McLean and Urken, 1995; Arrow et al., 2002). The typical social choice problem is arguably the aggregation of individual preferences over a set of alternatives into a collective preference (or a winning alternative). For example, everyone in the council might rank a set of proposals and desire to choose one proposal among several options. In other words, they are looking to aggregate their preferences into a collective decision.

The field of social choice theory is co-parented by two fathers, Nicolas de Condorcet and Jean-Charles de Borda, whose famous disagreement in the late 1700s highlighted that there are contrasting methods for aggregating preferences. While Condorcet argued that an alternative that beats every competitor in a pairwise majority contest should always be chosen as the winner, Borda argued that a better method for determining the best alternative was to sum over scores based on the position of alternatives in each individual's ranking. This disagreement made a case for the systematic study of aggregation methods, and built the foundation for social choice theory as a discipline. Just a few centuries later, Kenneth Arrow built upon this foundation with his eponymous impossibility theorem (Arrow, 1950). Not long after this, Gibbard and Satterthwaite (independently) showed, in essence, that all reasonable voting methods are susceptible to strategic manipulation by voters (Gibbard, 1973; Satterthwaite, 1975). This result—the

Gibbard-Satterthwaite Theorem—is often listed together with Condorcet's observation of majority cycles and Arrow's impossibility theorem as one of the classical results in social choice. The theorem's inclusion in this exclusive list highlights the prominence of strategic manipulation in the social choice literature.

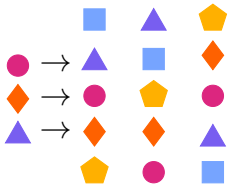
Strategic manipulation is also the topic of this thesis. We will be visiting three frameworks that formalise what it means for a group of agents to hold an opinion, elect a committee that represents them, or decide on their preferred candidate among several options. Our particular focus is on these agents themselves and when our choices for *how* to reach collective decisions affect their incentives and abilities to vote strategically.

1.1 Strategic Manipulation

Before really diving in, let's start with an example of strategic voting.

1.1. Example (Strategic Manipulation in Voting). Suppose three agents want to choose a shape among five alternatives. We represent their preferences below with each agent's shapes ordered from top to bottom, starting with their favourite and ending with their least preferred. For example, the first agent's top choice is the square while their least preferred shape is the pentagon.

The agents decide to use the Copeland rule to aggregate their preferences in order to choose a winner. This rule looks at how many times an alternative wins a pairwise majority contest and elects those alternatives that beat the highest number of other candidates. If all agents report their truthful preferences, the Copeland rule would tell them that the winning shape is the triangle. The first agent however, really wants the square to win and sees an opportunity—by submitting an untruthful ranking, she can force this outcome. If she shuffles her preference to place the triangle below the circle and diamond (represented here to the left of her initial preference), the rule will choose the square as the sole winner.



This example demonstrates that the Copeland rule (as well as many other voting methods) sometimes incentivises agents to submit untruthful preferences, as doing so may result in a more preferred outcome. △

In general, we'd like our aggregation methods to be *strategyproof*—meaning they do not incentivise agents to be untruthful as in the example above. Strategyproofness guarantees that the outcome of the aggregation does indeed reflect

the opinions of the individual agents as much as possible. It also ensures that we do not place the burden on the agents to figure out what the best strategy is in terms of what preferences or opinions to submit. Truth-telling is by definition a (weakly) dominant strategy when the aggregation method being used is strategyproof.

Throughout this thesis, we look at strategyproofness in three areas of social choice theory. We will first look at the typical voting framework for electing a single winner by taking into account agents' rankings over a set of alternatives. We will then see multiwinner elections where the goal is to elect a set of winners (or a *committee*) based on agents' approvals and disapprovals of candidates. Finally, we will explore the framework of judgment aggregation, where agents give their opinion on a set of possibly interconnected binary issues, and we aim to select a consistent judgment over these issues. For the purposes of this introduction, we stick to the aggregation of preference rankings in voting as our running example, though we note that the broad ideas we discuss also appear in the other two frameworks, as we will see when moving through the chapters.

When strategic behaviour comes up within the context of social choice, it mainly refers to possible manipulation of collective outcomes—can a voter or a group of voters sway the outcome of an election by misrepresenting their preferences when they submit their ballot(s)? As the Gibbard-Satterthwaite Theorem indicates, the answer to this question is often, unfortunately, a resounding yes. Strategic manipulation is difficult to completely avoid. This difficulty is not just present in voting theory; similar negative results also exist for multiwinner voting and judgment aggregation. Because of the pervasiveness of this issue in social choice, there are a host of different methods in the literature for finding settings where these negative results do not apply. A well-known and often used approach is to consider specific types of input to the aggregation method—so-called *restricted domains*. A voting method (or voting rule) takes as input a collection of preference rankings that we call a preference *profile*. We can examine these profiles to find common structures, or identify particular types of profiles where difficulties in aggregation can be avoided. Focusing on specific domains can often yield positive results, including results related to strategyproofness. For example, if we restrict our attention to the domain of unanimous profiles—meaning those where all voters submit the same preference ranking—it is easy to see that any reasonable voting rule should simply output this same ranking, and that no agent has an incentive to misreport their preference if they all agree on the ranking of the alternatives. This thesis builds on these ideas in order to tease out when we are actually able to establish some level of immunity to strategic manipulation, and—a related question—when the lack of strategyproofness can undermine some of the properties of the aggregation methods we study.

First, as we will often be speaking of domains and restricted domains, let's clarify exactly what we mean by this. In general, an aggregation method that accepts all inputs is said to have a universal domain. Often we only allow "rational"

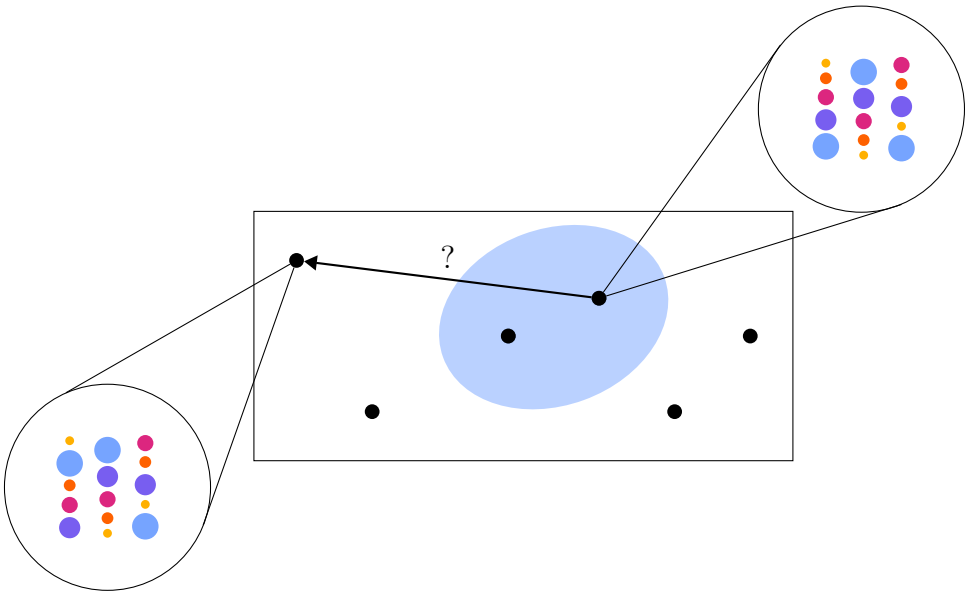
inputs—for example, individual agents cannot submit a cyclic preference—but do not put any further conditions on the profiles. When we look at an aggregation method with a restricted domain, this simply means that the aggregation method only accepts inputs that fall within that domain. Often we will speak of restricted domains without reference to a particular aggregation method.

Among the domains that are commonly encountered, the best-known is the set of *single-peaked* preference profiles (Black, 1948), which we will use as our example here. A preference profile is single-peaked if the agents can agree on some ordering of the alternatives such that each agent's preferences are in line with the proximity to their top alternative in this ordering—for all the alternatives that come before (after) the agent's top choice in the ordering, her preferences coincide with the distance from her top choice. A natural domain restriction will often define profiles that can be expected to appear in real-life collective decision making. For example, it is reasonable to assume in many cases that political preferences are single-peaked along a left-to-right axis. In voting, we know that if we are aggregating preferences that are single-peaked along the same order, we do not have to worry about Condorcet cycles in the majority preference.[1] If agents are truthful, and truthful preferences more often than not fall within the single-peaked domain, then we can be reassured that such cycles in the collective preference order will not often appear in practice. We also know that if we restrict the input of the aggregation function to single-peaked profiles, then we get strategyproofness "for free" within this domain (Moulin, 1980). Of course, this requires that we really only consider profiles within the restricted domain—agents cannot move outside the domain even when voting strategically, because the voting rule will not accept these profiles as input. But restricting the input to the voting rule is arguably inadvisable in settings where many, but not all, "truthful" profiles fall within the domain in question. This is the problem we examine: can we use previously identified domains of profiles to establish positive strategyproofness results, without actually restricting the input to the voting rule? Or do we create new incentives for manipulation on profiles within a certain domain, where such incentives would not exist if the input to the voting rule were to be restricted to profiles in this domain only?

1.2. Example (Manipulating from Restricted Domain). Consider the following illustration, where each black dot represents a profile and the blue area represents the profiles that are within our favourite restricted domain—for example, single-peakedness. We have zoomed in on a particular single-peaked three-agent profile (upper right) where each column represents an agent's preference order. For example, the first agent has ranked the smallest size (yellow circle) at the top and the largest (blue circle) at the bottom. We can see that this profile is single-

[1] A Condorcet cycle occurs when, for example, a majority of agents prefer an alternative a to b, b to c, and c to a, despite all individual agents having acyclic preferences over these same alternatives.

peaked along the small-to-large axis. Imagine these three agents have given us their preferences over t-shirt sizes because they are going to place an order for group t-shirts. They get a considerable discount when ordering the same size for all three, so must therefore choose one size based on everyone's preferences. It makes sense then, that the first agent (who is a size small) prefers the smallest size, and likes the options less and less the further away they are from her true size.



Existing results on strategyproofness can be restated as follows: agents do not have any incentive to manipulate between profiles within the blue area. Our question is whether it is possible to manipulate from inside the blue area to outside it. For example, can the first agent benefit from misreporting her preferences as in the bottom left profile (where she claims she'd go for the largest size if she cannot have the smallest size—which is the size that properly fits her). While this may be an unusual preference, we do not want to disallow such inputs as they can in theory appear even in truthful profiles (an agent might choose an oversized shirt over one that *almost* fits, if they cannot have their true size). △

As we have hinted above, our examination of strategyproofness relative to a particular type of profile also has an *axiomatic* motivation. There are axioms, or normative properties of voting rules, that directly relate to a specific type of profile. The most prominent of these is *Condorcet consistency*. This axiom states that if a profile admits a Condorcet winner—a candidate that beats every other candidate in a pairwise majority contest—then this candidate should be the sole winner in this profile. Next to scoring rules (the more Bordaesque class), Condorcet-consistent voting rules make up the bulk of rules studied. It is of interest, therefore, to see how this axiom interacts with strategic manipulation

by voters. Looking back at Example 1.1, we can see that the triangle (the chosen shape in the "truthful" profile) is a Condorcet winner. Yet the first agent is able to manipulate in a way that undermines the Condorcet consistency of the Copeland rule.

1.2 Our Motivation

The main question we ask in this thesis is the following:

Is it possible to manipulate *from* a profile in a "well-behaved" domain to one outside the domain in question?[2]

We ask this with a dual motivation in mind. The first is to establish strategyproofness on particular domains where we know manipulation within the domain is not possible. In this way we argue that we do not "lose" the existing safeguard against manipulation on these profiles, even when we allow the full domain of profiles as input. The second is to understand how strategyproofness (or lack thereof) on these domains can interact with axiomatic properties of our aggregation methods. Let us explore in further detail why we ask this question, and what kind of results we are chasing, while keeping the voting framework as our example.

Simply put, our goal is to establish strategyproofness on domains where we know manipulation *within* the domain is not possible. The particular domains we look at are known to be well-behaved and, in a sense, *natural* domains to examine. For example, the likelihood of a Condorcet cycle is arguably quite low in large elections (Gehrlein, 2006; Regenwetter et al., 2006), meaning profiles with these cycles are not likely to appear in practice. This reasoning can be used to motivate restricting the domain of voting rules to only those profiles with a Condorcet winner. If these are the profiles that will realistically show up in any case, why risk the trouble of allowing others? Our work here pushes against this idea of restricting the input to any aggregation method in a twofold manner:

- First, if we assume *most* profiles that appear in practice have a Condorcet winner, then establishing that no manipulation can occur in those profiles amounts to showing that manipulation will be the exception and not the rule—even when all profiles are allowed as input to the voting rule. Only when the truthful profile does *not* have a Condorcet winner will anyone possibly be able to successfully shift the outcome in their favour. However, if the truthful profile does not have a Condorcet winner and we restrict the input to the voting rule, agents will be forced to misreport their preferences by necessity.

[2] In Chapter 5 we also consider, for judgment aggregation rules, whether it is possible to manipulate from outside the domain in.

- Second, if it turns out that *all* truthful profiles have a Condorcet winner, then showing no manipulation can occur in those profiles amounts to showing that we do not create any additional incentives for manipulation by allowing all profiles as input (as opposed to restricting the domain). This would mean there is no reason to restrict the input to the voting rule. If the truthful profile is in the Condorcet domain, that will certainly be the reported profile, even when we allow agents the freedom to move to any other profile.

Let us now look at our axiomatic motivation, using the axiom of Condorcet consistency as our example. Our aim is to provide a more fine-grained way of distinguishing between Condorcet-consistent voting rules in terms of how strategic manipulation interacts with the existence of Condorcet winners in the reported profile when we know the truthful profile has a Condorcet winner. How much can we throw at a Condorcet-consistent rule before deserving Condorcet winners start losing elections?

- If a Condorcet winner exists and the rule we use is Condorcet-consistent, establishing that no manipulation can occur in profiles with a Condorcet winner would tell us that no agent has any incentive to manipulate in a way that "dethrones" a Condorcet winner.

Chapters 3, 4, and 5 each focus on our "main question" in a different framework within the area of social choice theory.

1.3 Thesis Overview

In each chapter of this thesis we focus on a particular domain, and a particular class of aggregation methods. The classes of methods we look at are particularly salient representatives within each framework.

Preferences. We dedicate Chapter 2 to *preferences*. Studying strategyproofness boils down to looking at whether agents can bring about "more preferred" outcomes for themselves. Therefore, preferences are central in our work. Many times we have to make decisions about how agents' preferences are structured. For example, if you tell me only which foods you like and which you dislike, or give me a ranking of a set of dishes, I may not be able to determine your favourite three-course meal without making some extra assumptions. A key aspect of this chapter is the topic of *preference extensions*. The aggregation rules we consider are *irresolute*, meaning they do not break ties for us. We will therefore discuss how to lift preferences over winners (or committees, or judgments) to preferences over sets of winners (or committees, or judgments)—a necessary evil when considering strategyproofness of irresolute rules. All the strategyproofness results

throughout this thesis will be relative to a certain preference extension, or a class of extensions.

Voting. Chapter 3 looks at (single-winner) voting. This is arguably the simplest framework we consider in this thesis. Each agent submits a strict preference ordering over a set of alternatives, and a voting rule takes these preference orders and returns a winner. We consider the class of (weighted) *tournament solutions*—voting rules which only require the majority relation as input. The domain we focus on in this chapter is the set of profiles where a Condorcet winner exists. For tournament solutions, we ask the question "will these voting rules return the Condorcet winner whenever one exists, also under the assumption that agents will behave strategically?" Our main result here ties the preservation of Condorcet winners to the decisiveness of the voting rule. Knowing that indecisiveness is required for robustness, we go about establishing positive results for more indecisive tournament solutions. This chapter is based on Botan and Endriss (2021).

Approval-Based Multiwinner Voting. Chapter 4 considers approval-based multiwinner voting. In this framework, each agent distinguishes between the alternatives they approve—their *approval set*—and those which they do not approve. An aggregation method takes this input and returns a set of alternatives, often called a *committee*. In our setting, we are looking for committees of a fixed size k. The domain of interest in this chapter is the set of *party-list profiles*, where any two approval sets either coincide or are disjoint. The class of rules we examine are the well-known *Thiele methods*. In this chapter we consider three types of manipulation actions, among them the existing notion of *free-riding*—agents relying on the popularity of some candidates they like and instead putting their weight behind others. Happily, we are able to establish quite a few positive results for the whole class of Thiele methods. The key result in this chapter is that Thiele rules are immune to free-riding on party-list profiles for a large class of preference extensions. This chapter is based on Botan (2021).

Judgment Aggregation. Chapter 5 discusses strategyproofness in judgment aggregation, the most general framework we consider in this thesis. Here, agents give their opinion on several binary issues represented by formulas of propositional logic. An aggregation rule returns the collective opinion over these issues based on the opinions of the individual agents. We consider *majoritarian* rules in this chapter, meaning those that return the outcome of the majority on all issues whenever doing so results in a logically consistent collective opinion. The natural domain to consider for these rules—and indeed the one we do consider—is the set of profiles that result in a consistent majority opinion. This chapter considers the most notable majoritarian judgment aggregation rules, from Slater to Dodgson. We obtain a range of strategyproofness results. Our strongest results are for the

class of *additive majority rules*, which includes the well-known Kemeny and Slater rules. This chapter is based on Botan and Endriss (2020).

In Chapter 6 we summarise our results and discuss directions for possible future work.

Chapter 2

Lifting Preferences

The aggregation methods we examine throughout this thesis are *irresolute*—meaning they do not always return a single winner. To reason about agents' incentives for strategic manipulation, we need to *lift* their preferences over objects to preferences over *sets of objects*. We do this by means of a *preference extension*—a function that takes a ranking over objects and returns a ranking over sets of objects. Here, "object" can take on a different meaning based on the framework. In Chapter 3, we look at voting, and an object is a single candidate—the winner. For the multiwinner elections we study in Chapter 4, these objects are sets of candidates. In Chapter 5, where we study judgment aggregation, the objects are sets of propositional logic formulas.

In general, we can interpret preferences over sets of objects in two distinct ways. The first is to see the sets as bundles of objects the agent will receive, meaning the set itself is the final outcome. The second is to see them as a set of mutually exclusive alternatives, one of which will be chosen in the end. We will consider only this second interpretation as this is what fits with our motivation. After all, our goal in the end is to choose one winner from a set of tied winners.

Throughout this thesis, we use $\succeq$ to speak about preferences over objects, and $\overset{\circ}{\succeq}$ to speak about the corresponding preferences over sets of objects. Let us look at an example to illustrate the problem ahead of us.

2.1. Example. The princess of Arrovia has declared that she would like to have dinner at one of the country's finest restaurants tonight. Her trusty assistant knows that the princess prefers Arrovian food to Thai to Italian. He must choose between two restaurants; one which serves either Arrovian or Italian cuisine, depending on which chef is available (though the diners do not know in advance who will be working, or how the work schedule is made), and one which serves

Thai cuisine every day. Which restaurant does the princess prefer? If the princess wants to guarantee that she at least gets her second choice, she may prefer the Thai restaurant. However, if she would risk ending up with Italian for a shot at some delicious Arrovian food, she may prefer the Arrovian-Italian one. Her assistant cannot be sure which the princess prefers based just on the information he has about her food preferences. △

2.1 Ranking Sets of Objects

The question of how to lift preferences over objects to preferences over sets of objects is not a new one. The study of preference extensions from an axiomatic point of view gained traction with the impossibility theorem of Kannai and Peleg (1984). They demonstrated that an extension cannot simultaneously satisfy two quite weak conditions—a dominance axiom and an independence axiom. This axiomatic work was continued by, among others, Barberà and Pattanaik (1984), Barberà et al. (1984), Fishburn (1984a), Bossert et al. (2000), and Pattanaik and Peleg (1984). Much of this work was directly inspired by the Kannai-Peleg Theorem, and featured similar impossibility results, as well as axiomatic characterisations of concrete extensions. More recently the preference lifting problem was studied by Geist and Endriss (2011), who used a SAT-solver to automatically generate a large number of impossibility results, as well as by Maly et al. (2019) who show one can circumvent the Kannai-Peleg impossibility when considering only certain families of sets (rather than ranking all sets of objects). For a thorough review of the problem of lifting preferences from an axiomatic viewpoint we refer the reader to Barberà et al. (2004).

The preference lifting problem predates the Kannai-Peleg Theorem. Much of the interest in the question of how to lift preferences was spurred on by the Gibbard-Satterthwaite Theorem (Gibbard, 1973; Satterthwaite, 1975). While the Gibbard-Satterthwaite Theorem deals a blow to *resolute* voting rules—which always return a single winner—it says nothing about irresolute rules. Many extensions were defined out of necessity by those studying strategic aspects of irresolute voting rules. Notable examples of extensions defined with this goal in mind are those by Fishburn (1972), Gärdenfors (1976), Kelly (1977), and Barberà (1977). In the aftermath of Gibbard-Satterthwaite, we saw many similar impossibility results for irresolute rules (Duggan and Schwartz, 2000; Gärdenfors, 1976; Kelly, 1977; Barberà, 1977). These results differ from Gibbard-Satterthwaite in how they define manipulability as they by necessity must make assumptions about agents' preferences over sets of alternatives. Moving to the irresolute setting also allowed for some more positive results. To give one example, Gärdenfors (1976) identifies two strategyproof voting rules for the Gärdenfors extension. The voting rule that returns all alternatives ranked first by at least one agent—also known as the omninomination rule—is in fact strategyproof for Gärdenfors preferences.

The same holds for the rule returning the Condorcet winner when one exists and the whole set of alternatives when one does not—sometimes called the Condorcet rule. In more recent years, we have seen a renewed interest in the strategyproofness of irresolute voting rules. We will go into more detail on this in Chapter 3.

2.2 The Formal Framework

We now define the notation we will use for preferences and preference extensions throughout this thesis. Recall that we use $\succeq$ to speak about preferences over objects, and use $\overset{\circ}{\succeq}$ to speak about the corresponding preferences over sets of objects. A preference order is a binary relation that is reflexive, transitive, antisymmetric, and connex. For a preference order $\succeq$ over a set X, we define the corresponding strict order $\succ$, and indifference relation $\sim$ in the usual way. We say $a \succ b$ if $a \succeq b$ and $b \not\succeq a$, and $a \sim b$ if $a \succeq b$ and $b \succeq a$. The same holds for preference orders over sets of objects: $A \overset{\circ}{\succ} B$ if $A \overset{\circ}{\succeq} B$ and $B \overset{\circ}{\not\succeq} A$, and $A \overset{\circ}{\sim} B$ if $A \overset{\circ}{\succeq} B$ and $B \overset{\circ}{\succeq} A$.

A *preference extension* is a function e mapping any given preference relation $\succeq$ over objects—elements of X—to a relation $e(\succeq)$ over sets of objects—nonempty subsets of X:

$$e(\succeq) = \overset{\circ}{\succeq}$$

While $\succeq$ is a complete (though possibly weak) order, $\overset{\circ}{\succeq}$ is not necessarily complete, meaning some sets may be incomparable under certain preference extensions. For notational simplicity, we will occasionally omit any reference to the specific extension e and speak about $\overset{\circ}{\succeq}$, rather than $e(\succeq)$. We will sometimes go even further, and refer to $\overset{\circ}{\succeq}$ as an extension. We hope the reader can forgive this slight abuse of notation.

We require, for all preference extensions, that $a \succeq b$ implies $\{a\} \overset{\circ}{\succeq} \{b\}$, and $a \succ b$ implies $\{a\} \overset{\circ}{\succ} \{b\}$. This requirement simply dictates that e stays faithful to the agent's preferences over objects when comparing singleton sets and corresponds to the extension rule of Barberà et al. (2004).

Finally, we say an agent i has *e-preferences* if $\overset{\circ}{\succeq}_i = e(\succeq_i)$, meaning her preference ranking $\succeq_i$ over objects is extended to a preference order $\overset{\circ}{\succeq}_i$ over sets of objects according to e.

Conditions on Extensions. We say an extension is *reflective* if

$$A \overset{\circ}{\succ} B \text{ implies that there is some } a \in A \text{ and } b \in B \text{ such that } a \succ b,$$

and *strongly reflective* if

$$A \overset{\circ}{\succ} B \text{ implies there is some } a \in A \text{ and } b \in B \text{ s.t. } a \succ b \text{ and } \{a, b\} \not\subseteq A \cap B.$$

Reflectiveness simply formalises the idea that the preferences over sets should reflect the preferences over object that they are extending. Strong reflectiveness is (clearly) a strengthening of reflectiveness, and is similar in spirit to the l-extension of Kruger and Terzopoulou (2020). Finally, we say a preference extension is *weakly pessimistic* if $X \mathrel{\mathring{\succ}} Y$ implies that there exists some $y \in Y$ such that $x \succeq y$ for all $x \in X$.

Defining Extensions. We now define some well-known extensions from the literature that will appear throughout the thesis. Let A and B be two sets such that $A \cup B \subset X$, and let $\succeq$ be a preference order over X. We first define the *optimistic extension*, which we refer to as e_o and the *pessimistic extension*—which we refer to as e_p. Let $\succeq^o = e_o(\succeq)$ and $\succeq^p = e_p(\succeq)$.

- $A \succeq^o B$ if and only if there is some $a \in A$ s.t. $a \succeq b$ for all $b \in B$.
- $A \succeq^p B$ if and only if for all $a \in A$ there exists some $b \in B$ s.t. $a \succeq b$.

Simply put, the optimistic extension looks only at the "best" object in each set—according to $\succeq$—and compares them, while the pessimistic extension looks only at the "worst" object in each set. We say an agent has optimistic preferences (or is an optimistic agent) if their preferences are extended according to the optimistic extension. This extension has a very natural interpretation, as is implied by the name; when comparing two sets, an optimistic agent operates under the assumption that their top choice within each set will be chosen as the final outcome. The pessimistic extension has a similar interpretation—a pessimistic agent operates under the assumption that their bottom choice will always be chosen as the final outcome.

2.2. Proposition. *Both e_o and e_p are strongly reflective.*

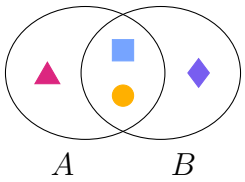
Proposition 2.2 follows directly from the definition of these extensions.

We now define three of the most well-studied extensions that we mentioned in the introduction to this chapter. The *Kelly extension* (Kelly, 1977)—which we refer to as e_k, The *Fishburn extension* (Fishburn, 1972)—which we refer to as e_f, and the *Gärdenfors extension* (Gärdenfors, 1976), which we refer to as e_g. We say an agent has *Kelly preferences* if her preferences over objects are extended to sets of objects according to the Kelly extension. We define Fishburn and Gärdenfors preferences analogously. Let $\succeq^k = e_k(\succeq)$, $\succeq^f = e_f(\succeq)$, and $\succeq^g = e_g(\succeq)$. The Kelly and Fishburn extensions are defined as follows:

- $A \succeq^k B$ if and only if $a \succeq b$ for all $a \in A$ and all $b \in B$.
- $A \succeq^f B$ iff $a \succeq b \succeq c$ for all $a \in A \setminus B, b \in A \cap B$ and $c \in B \setminus A$.

A set A is (weakly) preferred to B under the Kelly extension if and only if all elements of A are (weakly) preferred to all elements of B. A common interpretation of this extension is that agents have no idea about the tie-breaking mechanism that will be used to choose a winner from each set. Agents only express a preference when they know that the chosen winner from A will be preferred to the chosen winner from B, no matter what tie-breaking mechanism is used. The Fishburn extension is similar to Kelly, though it treats those elements that appear in both sets differently. So the objects in only A are preferred to those in both A and B, which are in turn preferred to those only in B.

2.3. Example. Let use compare the Kelly and Fishburn extensions. Suppose we have the following preference ranking over shapes: ▲ ≻ ■ ≻ ● ≻ ♦.



Here we can see that it is not the case that $A \succeq^k B$ (nor is it the case that $B \succeq^k A$) as we have an element of A, mainly ●, that is not preferred to all elements of B—■ ≻ ●. However, we do have that $A \succ^f B$ as ▲ is (strictly) preferred to both ● and ■, which are in turn both (strictly) preferred to ♦. Note that any two sets with an intersection that contains more than one element cannot be compared under the Kelly extension if we are extending strict preferences. △

Finally, the Gärdenfors extension is defined as follows:

- $A \succeq^g B$ if and only if one of the following three conditions is satisfied:

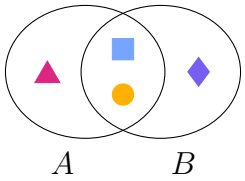
  (i) $A \subset B$ and $a \succeq b$ for all $a \in A$ and $b \in B \setminus A$

  (ii) $B \subset A$ and $a \succeq b$ for all $a \in A \setminus B$ and $b \in B$

  (iii) Neither $A \subset B$ nor $B \subset A$, and $a \succeq b$ for all $a \in A \setminus B$ and $b \in B \setminus A$

The Gärdenfors extension dictates that, if one set is to be preferred over another, then new elements added should be preferred to those already in the initial set. Similarly, the elements removed should be less preferred.

2.4. Example. Let use compare the Fishburn and Gärdenfors extensions with an example. We compare the same sets A and B from Example 2.3, though we are extending a different order over shapes. Suppose we have the following preference ranking over shapes: ■ ≻ ● ≻ ▲ ≻ ♦.



Clearly it is not the case that $A \succeq^f B$—the shapes in the intersection of the two sets are not preferred to ▲. However, we do have that $A \succ^g B$, as ▲ $\succ$ ♦. △

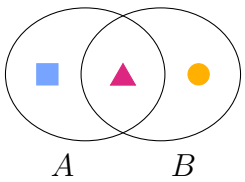
We now state two propositions that follow directly from the definitions above, and show the relative strength of the extensions.

2.5. Proposition. *$A \succ^k B$ implies $A \succ^f B$.*

2.6. Proposition. *$A \succ^f B$ implies $A \succ^g B$.*

These connections do not exists between the three "named" extensions—e_k, e_f, and e_g—and the optimistic and pessimistic extensions. We can see this in the following example where we extend weak preferences.

2.7. Example. Suppose we have the following (weak) preference ranking over shapes: ■ $\sim$ ▲ $\succ$ ●. We compare the two sets A and B below.



In this case $A \succ^k B$ (and therefore also $A \succ^f B$ and $A \succ^g B$). However, since the top alternatives in both sets are equally preferred, the optimistic extension will be indifferent between them. A similar example can be constructed for the pessimistic extension. △

The three extensions e_k, e_f, and e_g all satisfy strong reflectiveness and weak pessimism. This follows from simple examination of the axioms and the definition of the extensions.

2.8. Proposition. *The Kelly, Fishburn, and Gärdenfors extensions are strongly reflective.*

2.9. Proposition. *The Kelly, Fishburn, and Gärdenfors extensions are weakly pessimistic.*

What does all this mean for strategyproofness results that are related to some preference extension? Propositions 2.5 and 2.6 tell us that if we can show strategyproofness under the Gärdenfors extension, this immediately gives us strategyproofness for Kelly and Fishburn. Thus, a result using Gärdenfors is stronger than one using Fishburn or Kelly. On the other hand, a strategyproofness result for either the optimistic or pessimistic extension does not imply Kelly, Fishburn or Gärdenfors strategyproofness. Propositions 2.2, 2.8, and 2.9 tell us that if we can show strategyproofness for reflective, strongly reflective, or weakly pessimistic preferences, this immediately gives us strategyproofness under all three extensions—Kelly, Fishburn, and Gärdenfors.

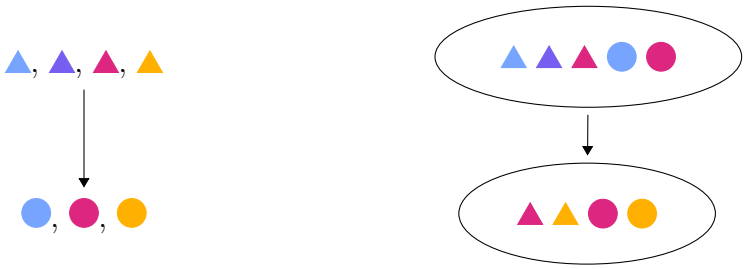
General Gärdenfors Preferences. We now define a larger class of preference extensions which includes both the Gärdenfors and optimistic preference extensions (as well as the Fishburn and Kelly preference extensions). We will use this class for strategyproofness results in Chapter 4. We say a preference extension is a *general Gärdenfors preference extension* in case that $A \mathrel{\mathring{\succ}} B$ holds only if one of the following holds:

(*i*) $A \not\subset B$ and there exists $a \in A \setminus B$ and $b \in B$ such that $a \succ b$.

(*ii*) $A \subset B$, $a \succeq b$ for all $a \in A$ and $b \in B \setminus A$, and there exists $a \in A$ and $b \in B \setminus A$ such that $a \succ b$.

Note that while the Gärdenfors preference extension is one specific preference extension, general Gärdenfors extensions are a class of preference extensions, of which the Gärdenfors extension is a member. We give an example below of preferences that fall into the class of general Gärdenfors preferences that are not captured by the specific extensions we have mentioned.

2.10. Example (General Gärdenfors Preference). Suppose we have an agent i who prefers all triangles to all circles (represented on the left). We can define the following preferences over sets of shapes: If A is a subset of B, $A \mathrel{\mathring{\succ}_i} B$ if and only if condition (ii) above is satisfied. The agent only wants to move to a subset—thereby excluding some possibilities without adding new ones—when certain guarantees are met. Otherwise, the agent prefers the outcome with the best ratio of triangles to circles.



With such preferences, agent i would prefer a set of three triangles and two circles to one with two of each (represented on the right). This is an example of a type of preference that can be captured by the class of general Gärdenfors preferences. In this specific instance, the Gärdenfors extension would not be able to compare the two sets, as each set contains an element that is strictly preferred to an element (only) in the other; ▲ $\succ_i$ ●, and ▲ $\succ_i$ ●. An optimistic agent would be indifferent between the two outcomes. △

As we move through the chapters, we will continue to encounter the notions and extensions we have defined here.

Chapter 3

Preserving Condorcet Winners in Voting

Voting theory is used to aggregate individual preferences with the aim of choosing a winner from a set of alternatives. In this setting, each agent (or voter) submits their strict preference order over a set of alternatives and a voting rule (or social choice function) is used to find the collectively "most preferred" alternative. There are many important considerations when deciding on what type of voting rule to use. Does it treat all voters the same? Does it unfairly favour some alternatives? Will it ever choose an alternative that is disliked by all the voters?

One central concern when choosing a voting rule is whether it is possible for agents to strategically manipulate the outcome in their favour. In other words, can a voter lie and ensure that a more preferred alternative will win compared to when she submits her truthful preference? In Chapter 1 we saw some examples of strategic manipulation and briefly mentioned the Gibbard-Satterthwaite Theorem, a central result in social choice that concerns precisely this kind of strategic behaviour. We now formally state the theorem. Recall that a voting rule is *resolute* when it always returns a singleton. A resolute voting rule is *nonimposed* if every alternative wins in some profile, and it is a *dictatorship* when the top alternative of the same agent—the dictator—is always chosen as the winner.

3.1. Theorem (Gibbard, 1973; Satterthwaite, 1975). *Every resolute, nonimposed, and strategyproof voting rule is a dictatorship.*

In the aftermath of the Gibbard-Satterthwaite Theorem, we have seen similar impossibility results for irresolute rules. These results differ in how they define manipulability as they must make a choice about how to define agents' preferences over sets of alternatives. We present the most well-known such result. Duggan

and Schwartz (2000) generalise the Gibbard-Satterthwaite Theorem to irresolute rules. They do so for the optimistic and pessimistic preference extensions that we saw in Chapter 2. First note that an irresolute rule is *nonimposed* if every singleton is returned as the winning set in some profile. Further, an irresolute rule is *weakly dictatorial* is there is some agent whose top alternative is always included in the outcome. We can now state the theorem.

3.2. Theorem (Duggan and Schwartz, 2000). *Any nonimposed irresolute social choice function that is strategyproof under both the optimistic and pessimistic extension must be weakly dictatorial.*

Although impossibilities abound, shifting focus away from resoluteness has also led to positive results regarding the strategyproofness of social choice functions. As we stated in Chapter 2, one of example of this is Gärdenfors (1976), who identified two strategyproof functions for the Gärdenfors extension. Building on these earlier results, Brandt (2015) characterises the pairwise social choice functions that are strategyproof under the Kelly preference extension. Among these are the bipartisan set and the minimal covering set. We discuss in Section 3.3.1 the connection between our work and Kelly-strategyproofness. Brandt and Brill (2011) further add to these results and find sufficient conditions for strategyproofness under the stronger Fishburn and Gärdenfors preferences as well, thereby identifying further social choice functions that are strategyproof for each of the three extensions.

In addition to focusing on irresolute rules, another well-employed strategy to deal with strategic manipulation that is relevant for our purposes is to consider a restricted domain for the social choice function. Among these domain restrictions, the best-known is the single-peaked domain of Black (1948), which we saw an example of in Chapter 1. Many such restricted domains ensure strategyproofness for Condorcet extensions given that the voting rule only allows profiles from the domain as input. While our motivation in this chapter is primarily of an axiomatic nature, strategyproofness on profiles with a Condorcet winner—or domains that are subsets of this set of profiles—is also appealing from a practical viewpoint as real-world elections have a high probability of giving us a Condorcet winner (Gehrlein, 2006; Gehrlein and Lepelley, 2011).

There are also many examples of positive results relative to more fine-grained axioms that focus on—and limit—the type of manipulation performed by the agent. In preference aggregation, both Sato (2013) and Bossert and Sprumont (2014) obtain positive results for strategyproofness when considering specific manipulations based on the distance between and agent's initial preferences and the outcome. More recently, Kruger and Terzopoulou (2020) have found voting rules that are strategyproof to manipulation by adding, removing, or swapping alternatives in agents' (incomplete) preference orders. In some sense, our method of considering strategyproofness on a particular set of profiles is in line with this

approach as we are focusing on a specific type of manipulation—only those manipulations that start from a profile with a Condorcet winner.

Similar positive results have been obtained by considering voters' ignorance of others' preference as an informational barrier (Conitzer et al., 2011; Reijngoud and Endriss, 2012; Osborne and Rubinstein, 2003). Another successful approach has been to argue for the computational hardness of computing a possible manipulation strategy as a barrier to manipulation (Bartholdi et al., 1989; Conitzer and Walsh, 2016).

Our particular focus in this chapter is strategic manipulation of Condorcet extensions—voting rules that will return the Condorcet winner as the unique winner whenever such an alternative exists. Condorcet extensions have long held a prominent place in social choice theory, and for good reason. Going against the opinion of the majority is generally frowned upon. However, the definition of a Condorcet extension does not take into account possible manipulation by the voters. A profile where all agents vote truthfully may have a Condorcet winner, but this alternative may not end up in the set of winners if agents are acting strategically and the reported profile differs from the truthful one. We examine exactly when the lack of strategyproofness affects whether we can trust that a Condorcet extension will give us all "true" Condorcet winners, even under the assumption that agents might vote strategically. By doing this, we distinguish between rules that do not incentivise manipulation from profiles with Condorcet winners, and those that do. We call the former *robust Condorcet extensions*, and they are the Condorcet extensions that always return the Condorcet winner whenever the *truthful* profile has one. Further, we highlight the relationship between the decisiveness of a social choice function and whether it incentivises agents to unseat a Condorcet winner via manipulation. Decisiveness as a concept turns out to be a relevant consideration for strategyproofness of irresolute rules, in particular for Kelly-strategyproofness (Brandt et al., 2021). Though their technical definition for indecisiveness is different from ours, we both speak about the size of the set returned by voting rules.

The idea of preserving Condorcet winners has also been examined in the setting of probabilistic social choice. Hoang (2017) shows that *maximal lotteries* (Fishburn, 1984b) are strategyproof on profiles with Condorcet winners when based on the majority relation. Brandl et al. (2018) extend this result to all maximal lotteries.

This chapter is organised as follows. We introduce the framework and relevant literature in Section 3.1. In Section 3.2 we present impossibility results that tell us many Condorcet extensions cannot be robust. We first show that no Condorcet-consistent tournament solution can be robust for all preference extensions. We then redirect our search to look for rules that can satisfy robustness for at least some extensions. Our main result tells us this is not a possibility for a large class of more decisive rules. In Section 3.3 we examine the connection between Kelly-strategyproofness and robustness, and present a number of sufficient conditions

for a Condorcet extension to be robust. We identify several attractive social choice functions that are robust Condorcet extensions for a large class of preferences. This includes, in particular, the minimal extending set (Brandt, 2011) and its coarsenings. We conclude in Section 3.4.

3.1 The Model

Let A be a finite set of *alternatives*, and $N = \{1, \dots, n\}$ a finite set of *agents*. A *preference profile* $\boldsymbol{P} = (\succ_1^{\boldsymbol{P}}, \dots, \succ_n^{\boldsymbol{P}})$ is a vector of strict linear orders over A, where $\succ_i^{\boldsymbol{P}}$ is the *preference relation* of agent i in the profile $\boldsymbol{P}$. We write $\succeq_i^{\boldsymbol{P},e}$ to denote agent i's preference relation extended according to the preference extension e—i.e., $\succ_i^{\boldsymbol{P},e} = e(\succ_i^{\boldsymbol{P}})$. For two profiles $\boldsymbol{P}$ and $\boldsymbol{P'}$, and agent $i \in N$, we write $\boldsymbol{P} =_{-i} \boldsymbol{P'}$, and say they are *$i$-variants*, if $\succ_j^{\boldsymbol{P}} = \succ_j^{\boldsymbol{P'}}$ for all $j \in N \setminus \{i\}$. $\mathcal{L}(A)$ denotes the set of all linear orders over A, and $\mathcal{L}(A)^n$ denotes the set of all profiles for n agents. We write $N_{aa'}^{\boldsymbol{P}} = \{i \in N \mid a \succ_i a'\}$ to denote the set of agents who prefer alternative a to a'.

For a profile $\boldsymbol{P}$, $\succeq^{\boldsymbol{P}}$ (with asymmetric part $\succ^{\boldsymbol{P}}$) is the *majority relation* for $\boldsymbol{P}$ and is defined such that $a \succeq^{\boldsymbol{P}} a'$ if and only if $|\{i \in N \mid a \succ_i^{\boldsymbol{P}} a'\}| \geq |\{i \in N \mid a' \succ_i^{\boldsymbol{P}} a\}|$, for all $a, a' \in A$. An alternative $a \in A$ is a *Condorcet winner* in profile $\boldsymbol{P}$ if it defeats every other alternative in a pairwise majority contest, meaning $a \succ^{\boldsymbol{P}} a'$ for all $a' \in A \setminus \{a\}$. We define $\mathcal{D}_{Condorcet}$ as the set of all profiles with a Condorcet winner. We say a relation $\succeq$ over A is *complete* if for all $a, b \in A$ it is the case that $a \succeq b$ or $b \succeq a$. A relation $\succ$ is *connex* if $a \succ b$ or $b \succ a$ for all distinct $a, b \in A$. Note that the majority relation of any profile is complete, and any individual preference relation is connex. We write $\mathrm{top}(\succ)$ to denote the maximal element of the strict linear order $\succ$.

An irresolute *social choice function* (SCF) f is a mapping from profiles to nonempty subsets of alternatives:

$$f : \mathcal{L}(A)^n \to 2^A \setminus \{\emptyset\}$$

To avoid having to break majority ties, we will most of the time focus on SCFs for odd n. Of course we do not strictly speaking need odd n, but rather that profiles input to the function have a strict majority relation. We will explicitly mention when we talk about profiles with an even number of agents. A SCF f is a *Condorcet extension*, or is *Condorcet-consistent*, if it returns (only) the Condorcet winner whenever one exists.

For irresolute SCF, the size of the set of winning alternatives is an important consideration. All things being equal, it is preferable that the SCF does not outsource the decision-making to a tie-breaking mechanism, but rather does most of the work of selecting a winner itself. More simply put, we would rather a SCF return small sets more often than it returns large ones. Of course, an irresolute SCF cannot always avoid tie-breaking—a single winner will often need to be

chosen from the outcome using some tie breaking rule—but the SCF should ideally break as many ties as possible before an outside tie-breaking method is applied. As an example of a rule that returns quite large sets, take the rule that returns the Condorcet winner if one exists, and returns the whole set of alternatives otherwise. While this is clearly a Condorcet extension, it is a very *indecisive* rule, as it often results in many ties in the outcome.

Recall that a resolute SCF always returns a singleton. In order to quantify one aspect of how decisive an irresolute rule is, we define a weaker notion of resoluteness. We say f is *weakly resolute* if there exists a profile $\boldsymbol{P} \in \mathcal{L}(A)^n \setminus \mathcal{D}_{Condorcet}$ for which $|f(\boldsymbol{P})| = 1$. So, a rule is weakly resolute if it *sometimes* returns a singleton for a profile without a Condorcet winner. For some SCFs, we can directly compare how decisive they are relative to each other. A SCF f is a *refinement* of f' if for all profiles $\boldsymbol{P} \in \mathcal{L}(A)^n$ it is the case that $f(\boldsymbol{P}) \subseteq f'(\boldsymbol{P})$, meaning f always returns a subset of f'. If f is a refinement of f', we say f' is a *coarsening* of f. If a rule is a refinement of another, it is clearly the more decisive of the two.

3.1.1 Tournaments and Tournament Solutions

A *tournament* $\boldsymbol{T}$ is a pair $(S, \succ^{\boldsymbol{T}})$, where S is a set of nodes (or alternatives) and $\succ^{\boldsymbol{T}}$ is an asymmetric and connex relation over S, which we call the *dominance relation* of the tournament. For a set S, we denote by $\mathcal{T}(S)$ all tournaments on S.

For a tournament $\boldsymbol{T} = (S, \succ^{\boldsymbol{T}})$, we say an alternative $a \in S$ *dominates* $a' \in S$ in the tournament $\boldsymbol{T}$ if $a \succ^{\boldsymbol{T}} a'$. The *dominion* of a in $\boldsymbol{T}$ is defined as $D_{\boldsymbol{T}}(a) = \{a' \in S \mid a \succ^{\boldsymbol{T}} a'\}$, the set of alternatives it dominates. The set of *dominators* of a in $\boldsymbol{T}$ is defined as $\overline{D}_{\boldsymbol{T}}(a) = \{a' \in S \mid a' \succ^{\boldsymbol{T}} a\}$, the set of alternatives that dominate it. For $S' \subseteq S$, we define the restriction $\succ^{\boldsymbol{T}}_{S'} = \{(a, a') \in S' \times S' \mid a \succ^{\boldsymbol{T}} a'\}$, which is $\succ^{\boldsymbol{T}}$ restricted to the set S'. A *subtournament* of $\boldsymbol{T} = (S, \succ^{\boldsymbol{T}})$ is a tournament $(S', \succ^{\boldsymbol{T}}_{S'})$ where $S' \subseteq S$. Thus, a subtournament of $\boldsymbol{T}$ is a subset of the nodes in $\boldsymbol{T}$, together with the edges between those nodes. We say $\pi : S \to S'$ is an isomorphism between two tournaments $\boldsymbol{T} = (S, \succ^{\boldsymbol{T}})$ and $\boldsymbol{T'} = (S', \succ^{\boldsymbol{T'}})$ if π is a bijection, and $a \succ^{\boldsymbol{T}} a' \Leftrightarrow \pi(a) \succ^{\boldsymbol{T'}} \pi(a')$ for all $(a, a') \in S \times S$.

We say a profile $\boldsymbol{P} \in \mathcal{L}(A)^n$ *induces* tournament $\boldsymbol{T} = (A, \succ^{\boldsymbol{T}})$ if $\succ^{\boldsymbol{P}} = \succ^{\boldsymbol{T}}$. So a profile induces a tournament if they range over the same alternatives, and the strict part of the majority relation is exactly the dominance relation of the tournament. Note that if a profile induces a tournament, the strict component of the majority relation of that profile must be connex. As tournaments do not speak about agents, we cannot directly talk about two tournaments being i-variants for some agent $i \in N$. Instead, we say two tournaments $\boldsymbol{T} = (A, \succ^{\boldsymbol{T}})$ and $\boldsymbol{T'} = (A, \succ^{\boldsymbol{T'}})$ are *single-agent variants*, and write $\boldsymbol{T} =_{-1} \boldsymbol{T'}$, if there exist a set of agents N and profiles $\boldsymbol{P}, \boldsymbol{P'} \in \mathcal{L}(A)^n$, for $n = |N|$, such that $\boldsymbol{P} =_{-i} \boldsymbol{P'}$ for some agent $i \in N$, and the profiles $\boldsymbol{P}$ and $\boldsymbol{P'}$ induce the tournaments $\boldsymbol{T}$ and $\boldsymbol{T'}$, respectively.

We say an element $a \in S$ is the *Condorcet winner* of the tournament $\boldsymbol{T} = (S, \succ^{\boldsymbol{T}})$ if $\overline{D}_{\boldsymbol{T}}(a) = \emptyset$. This corresponds to the notion of a Condorcet winner of a profile; if a tournament has a Condorcet winner, that alternative will be the Condorcet winner in any profile that induces this tournament. We write $\mathcal{T}_{Condorcet}$ to refer to the set of tournaments that have a Condorcet winner. A *tournament solution* F is a mapping from tournaments to sets of alternatives, that does not distinguish between isomorphic tournaments:

$$F : \mathcal{T}(S) \to 2^S \setminus \{\emptyset\}$$

So $F(\boldsymbol{T'}) = \{\pi(a) \mid a \in F(\boldsymbol{T})\}$ if π is an isomorphism between $\boldsymbol{T}$ and $\boldsymbol{T'}$. For ease of reading, we sometimes write $F(\succ^{\boldsymbol{T}})$ to mean $F(S, \succ^{\boldsymbol{T}})$ when S is clear from context.

A SCF f is *equivalent to* a tournament solution F if $f(\boldsymbol{P}) = F(A, \succ^{\boldsymbol{P}})$ for all $\boldsymbol{P} \in \mathcal{L}(A)^n$. Note that the majority relation of this profile $\boldsymbol{P}$ must be a strict order, as the SCF f is defined for odd n only. In a slight muddling of terminology, we will refer to SCFs that are equivalent to tournament solutions as *tournament-solution SCFs*.

Tournament solutions roughly correspond to Fishburn's C1 functions (Fishburn, 1977), which require only the information in the majority graph to determine the winners. More precisely, tournament-solution SCFs correspond to *neutral* C1 functions. A SCF satisfies *neutrality* if for any profile $\boldsymbol{P}$ and any permutation $\pi : A \to A$ it is the case that $f(\pi(\boldsymbol{P})) = \pi(f(\boldsymbol{P}))$. This is because tournament solutions do not distinguish between isomorphic tournaments, and therefore do not favour any alternatives over others. While it is sometimes assumed that tournament solutions are Condorcet-consistent, we will not impose this restriction. We do, however, only consider Condorcet extensions in this chapter.

3.1.2 Weighted Tournaments

Some of our results also extend to *weighted* tournament solution SCFs. These correspond to neutral C2 functions under Fishburn's classification (Fishburn, 1977), and require the strength of the majorities as input to determine the winner. First, let's define what we mean by a weighted tournament. A weighted tournament is a pair (S, W) where S is a set of nodes, and $W : S \times S \to \mathbb{Z}$ is a weight function such that $W(a, a) = 0$. Alternatively, we can think of this as a graph where each edge has a weight. For any profile $\boldsymbol{P}$ we can define a corresponding weighted tournament. We first define the majority margin $m^{\boldsymbol{P}}(a, a')$ of an alternative a over another alternative a', relative to a profile $\boldsymbol{P}$. The majority margin tells us the difference between the number of agents who prefer a to a' and the number of agents who prefer a' to a:

$$m^{\boldsymbol{P}}(a, a') = |\{i \in N \mid a \succ_i^{\boldsymbol{P}} a'\}| - |\{i \in N \mid a' \succ_i^{\boldsymbol{P}} a\}|$$

The weighted tournament corresponding to $\boldsymbol{P}$ is then $\boldsymbol{T} = (A, W)$ where W is such that $W(a, a') = m^{\boldsymbol{P}}(a, a')$ for all $a, a' \in A$ such that $a \neq a'$. Note that this entails $W(a, a) = 0$.

We say a weighted tournament solution F' is the *weighted counterpart* to a tournament solution F if the two functions agree whenever all edges with positive weight have the same weight (and as a consequence all edges with negative weight have the same weight). More formally, for a tournament $\boldsymbol{T} = (A, \succ^{\boldsymbol{T}})$ and weighted tournament $\boldsymbol{T'} = (A, W)$ where $W(a, a') = k$ for some $k > 0$ if and only if $a \succ^{\boldsymbol{T}} a'$, it is the case that $F(\boldsymbol{T}) = F'(\boldsymbol{T'})$.

3.1.3 Particular Tournament Solutions

We have already seen the Copeland rule in action in Chapter 1, but we now define the *Copeland tournament solution*—which corresponds to the Copeland SCF—formally. The Copeland score of an alternative in a tournament is the number of other alternatives it dominates. The Copeland rule selects those alternatives with the highest Copeland scores (Copeland, 1951). Thus, Copeland selects those alternatives with the largest dominion.

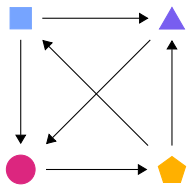
$$F_{\text{Cop}}(\boldsymbol{T}) = \operatorname*{argmax}_{a \in A} |D_{\boldsymbol{T}}(a)|$$

Another prominent SCF is the Slater rule. For a tournament $\boldsymbol{T}$, we define the *Slater set* as $\text{Sla}(\boldsymbol{T}) = \operatorname*{argmax}_{\succ \in \mathcal{L}(A)} |\succ \cap \succ^{\boldsymbol{T}}|$. The *Slater tournament solution* is then defined as follows.

$$F_{\text{Sla}}(\boldsymbol{T}) = \{\text{top}(\succ) \mid \succ \in \text{Sla}(\boldsymbol{T})\}$$

The Slater rule returns the top element of those linear orders that are closest to $\succ^{\boldsymbol{T}}$. In other words, Slater "flips" the smallest number of edges in the tournament until we reach a linear order, and then returns the maximal alternative of the linear orders reached by a minimal number of flips.

3.3. Example (Copeland and Slater). Let $\boldsymbol{T}$ be the tournament below.



We first determine the winners under the Copeland rule. Both ■ and ⬟ are each dominated by a single other alternative and they are the only such alternatives. Thus, these are our Copeland winners. Note however that, while we can reverse

the edge (■,⬟) to obtain a Condorcet winner, the resulting tournament (bottom left, with reversed edge represented as dashed) will still have a 3-cycle comprising ●,⬟, and ▲. On the other hand, the tournament resulting from reversing the edge (●, ⬟) (bottom right) is cycle-free.

We conclude that ■ is not a winner under Slater, while ⬟ is, and is in fact the only winner. △

We now move on to two SCFs with slightly more complex definitions. First, we will need the notion of a maximal transitive subtournament. A tournament $\boldsymbol{T'} = (S', \succ^{\boldsymbol{T'}})$ is a *maximal transitive subtournament* of $\boldsymbol{T} = (S, \succ^{\boldsymbol{T}})$ if

1. $\boldsymbol{T'}$ is a subtournament of $\boldsymbol{T}$,
2. $\succ^{\boldsymbol{T'}}$ is a transitive relation, and
3. there is no other transitive subtournament $(S'', \succ^{\boldsymbol{T''}})$ of $\boldsymbol{T}$ such that $S' \subset S''$.

We write $\hat{\boldsymbol{T}}$ to denote the set of all maximal transitive subtournaments of tournament $\boldsymbol{T}$. Note that if a tournament $\boldsymbol{T}$ has a Condorcet winner, it will be the maximal element of *all* maximal transitive subtournaments.[1] The *Banks set* (Banks, 1985) is the set of maximal elements of all maximal transitive subtournaments of a tournament:

$$F_{\text{Ba}}(\boldsymbol{T}) = \{\text{top}(\succ_S^{\boldsymbol{T}}) \mid (S, \succ_S^{\boldsymbol{T}}) \in \hat{\boldsymbol{T}}\}.$$

Because the Condorcet winner will top all maximal transitive subtournaments, the Banks set is a Condorcet extension.

A set $S \subseteq A$ is a F_{Ba}*-stable set* of a tournament $\boldsymbol{T}$ if $a \notin F_{\text{Ba}}(S \cup \{a\}, \succ_{S\cup\{a\}}^{\boldsymbol{T}})$ for all $a \in A \setminus S$. A F_{Ba}-stable set of a tournament T is *minimal* if there is no F_{Ba}-stable set S' of T such that $S' \subset S$. The *minimal extending set* $F_{\text{ME}}(\boldsymbol{T})$ (Brandt, 2011) of a tournament $\boldsymbol{T}$ is the union of all minimal F_{Ba}-stable sets of $\boldsymbol{T}$:

$$F_{\text{ME}}(\boldsymbol{T}) = \bigcup\{S \subseteq A \mid S \text{ is a minimal } F_{\text{Ba}}\text{-stable set of } \boldsymbol{T}\}.$$

We give an example to shed some light on these definitions.

[1] As the existence of a Condorcet winner does not imply no cycles are present, there may indeed be several maximal transitive subtournaments.

3.4. Example. In the tournament $\boldsymbol{T}$ below, the two maximal transitive subtournaments are indicated using darker edges. It is clear that the subtournaments are transitive, and they are both maximal; adding the last alternative will break transitivity. From examining these subtournaments, we can see that $F_{\text{Ba}}(\boldsymbol{T}) = \{\blacksquare, \blacktriangle\}$.

This tournament has two minimal F_{Ba}-stable sets: $\{\blacksquare, \blacktriangle, \bullet\}$—because $\pentagon \notin F_{\text{Ba}}(\boldsymbol{T})$, and $\{\blacksquare, \blacktriangle, \pentagon\}$—because $\bullet \notin F_{\text{Ba}}(\boldsymbol{T})$. F_{ME} will output the union of these sets: $F_{\text{ME}}(\boldsymbol{T}) = \{\blacksquare, \blacktriangle, \pentagon, \bullet\}$. Note that the set $\{\blacktriangle, \pentagon, \bullet\}$ is not F_{Ba}-stable, as $\blacksquare \in F_{\text{Ba}}(\boldsymbol{T})$. △

3.1.4 Particular Weighted Tournament Solutions

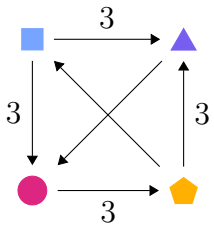
We now define the Kemeny rule (Kemeny, 1959) as our first example of a weighted tournament solution and an example of a rule that is a weighted counterpart of a tournament solution—Slater. More precisely, we define the equivalent Kemeny SCF. For a profile $\boldsymbol{P}$, we define the *Kemeny set* as $\text{Kem}(\boldsymbol{P}) = \underset{\succ \in \mathcal{L}(A)}{\text{argmax}} \sum_{i \in N} |\succ \cap \succ_i|$.

The *Kemeny SCF* is then defined as follows.

$$f_{\text{Kem}}(\boldsymbol{P}) = \{\text{top}(\succ) \mid \succ \in \text{Kem}(\boldsymbol{P})\}$$

Where Slater tries to flip the smallest number of edges needed to reach a linear order, the Kemeny rule looks to reach a linear order by flipping a set of edges with the smallest sum of weights possible. It then returns the top element of those orders.

3.5. Example (Kemeny). Let $\boldsymbol{T}$ be the weighted tournament below, which is a weighted counterpart to the tournament we saw in Example 3.3. We omit the weights for edges where the weight equals 1.



Recall that $\pentagon$ is the sole winner under Slater and that we arrive at this winner after reversing an edge with weight 3 (dashed edge in weighted tournament on the left below).

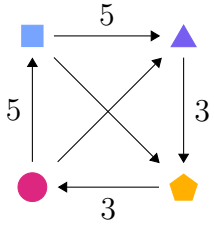
However, there is a linear order that requires us to reverse two edges each with weight 1 (dashed edges in weighted tournament on the right below). As Kemeny tries to minimise the weight of edges reversed until arriving at a linear order, we see that the rule will return ■ as the sole winner. △

A second well-known weighted tournament solution is the *ranked pairs* rule F_{RP} (Tideman, 1987). Intuitively, F_{RP} orders pairs of alternatives by the strength of the majority margins, then iteratively builds a linear order, starting with pairs of alternatives with a high majority margin. If we look at a weighted tournament, this amounts to adding (one of) the edge(s) with highest weight. We continue adding edges from highest to lowest weight whenever adding an edge does not create a cycle. We now formally define this iterative process. For any weighted tournament $\boldsymbol{T}$, and any order $p_1, \ldots, p_{\binom{m}{2}}$, where p_k is an edge in $\boldsymbol{T}$ such that that $W(p_k) \geq W(p_{k+1})$ for $k \in [1, \binom{m}{2} - 1]$, ranked pairs proceeds as follows. Let $\succ_0 = \emptyset$. At step k, where $p_k = (a, b)$:

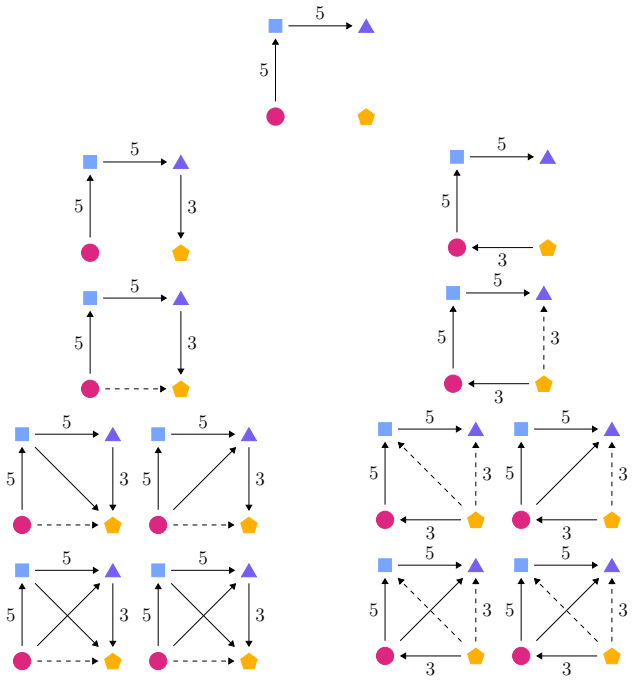
- $\succ_k = \succ_{k-1} \cup \{(a, b)\}$ if $\succ_{k-1} \cup \{(a, b)\}$ is acyclic,
- $\succ_k = \succ_{k-1} \cup \{(b, a)\}$ if $\succ_{k-1} \cup (a, b)\}$ contains a cycle.

This process will terminate after step $\binom{m}{2}$, and $\text{top}(\succ_{\binom{m}{2}}) \in F_{\text{RP}}(\boldsymbol{T})$. Note that because two edges may have the same weight, there can be several orders of edges that satisfy our requirement. Thus, the iterative process may result in different orders depending on which edge we look at first. The definition for ranked pairs becomes much clearer with an example.

3.6. Example (Ranked Pairs). Consider the weighted tournament below. We omit weights where they equal 1. We want to determine the winners under the ranked pairs rule.



We start the iterative process by adding the two edges with weight 5 as they do not create a cycle. We then choose one of the edges with weight 3. This leads to a "branching" depending on which edge we choose. We use dashed lines to indicate those edges that we have reversed to prevent cycles.



We see that the iterative process returns two unique linear orders. This gives us two winners: ● and ⬟. △

3.1.5 Robust Condorcet Extensions

Recall that we use $\overset{\circ}{\succeq}_i^{\boldsymbol{P}}$ to refer to agent i's preferences over sets of candidates. We say an irresolute SCF f is *Condorcet-manipulable* by agent i in profile $\boldsymbol{P}$ if there exists another profile $\boldsymbol{P'} =_{-i} \boldsymbol{P}$ such that $f(\boldsymbol{P'}) \overset{\circ}{\succeq}_i^{\boldsymbol{P}} f(\boldsymbol{P})$ and $\boldsymbol{P} \in \mathcal{D}_{Condorcet}$. We are now ready to present our central definition.

A SCF f is a *robust Condorcet extension* under a preference extension e if f is Condorcet-consistent and not Condorcet-manipulable in any profile $\boldsymbol{P} \in \mathcal{D}_{Condorcet}$ by any agent $i \in N$ with preferences $\overset{\circ}{\succeq}_i^{\boldsymbol{P}} = e(\succ_i^{\boldsymbol{P}})$.

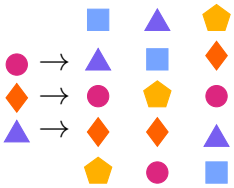
We sometimes write that a SCF is *robust* to mean that it is a robust Condorcet extension, as robustness is a property of Condorcet extensions. So a SCF is robust under a certain preference extension, if it is a Condorcet extension, and it is not Condorcet-manipulable by any agent whose preferences over alternatives have been extended to sets of alternatives according to that extension.

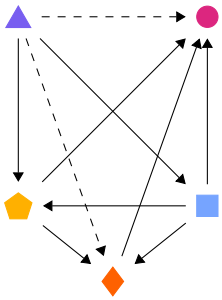
While robustness is a weak strategyproofness requirement, it also speaks about how well a rule can preserve Condorcet winners. A robust Condorcet extension

ensures that, if the truthful profile has a Condorcet winner, then it is a weakly dominant strategy for all agents to report their true preferences, thus ensuring that no Condorcet winner loses that designation because of strategic manipulation. A robust Condorcet extension therefore ensures that profiles with Condorcet winners are, in a sense, stable. We give an example of a Condorcet manipulation of the *Copeland* SCF to demonstrate what failure of robustness looks like.

3.7. Example. Recall the example in Chapter 1 where three agents want to choose a shape among five alternatives. We represent their preferences below with each agent's shapes ordered from top to bottom, starting with their favourite and ending with their least preferred. For example, the first agent's top choice is ■ while their least preferred shape is ⬟.The corresponding tournament is shown on the right.

We use the Copeland rule to aggregate their preferences in order to choose a winner. If all agents report their truthful preferences, the winning shape is ▲. As we saw in Chapter 1 however, if the first agent shuffles her preference to place ▲ below both ● and ♦ (represented here to the left of her initial preference), ■ will be the sole winner.





As the Copeland winner is the alternative with the smallest number of incoming edges in the majority graph, ■ would be the lone Copeland winner if agent 1 misreports her preferences, meaning, Copeland incentivises a Condorcet-manipulation in this profile. △

While the truthful profile in Example 3.7 has a Condorcet winner, Copeland is not guaranteed to return this alternative as the winner (or even among them) unless we assume all agents vote truthfully. In the same scenario, a robust Condorcet extension would ensure no agent would have an incentive to misreport her preferences, and thus ensure that the truthful profile and the reported profile coincide.

3.1.6 Relation to Domain Restrictions

One consequence of a Condorcet extension being robust is that it is not manipulable on profiles that fall within several known domain restrictions. This is because many restrictions are subsets of $\mathcal{D}_{Condorcet}$. We define two of the most well known

such restrictions here—one based on an ordering of the alternatives and one based on an ordering of the agents. These domain restrictions are examples of types of structures that may arise naturally in many contexts where agents are asked to rank a set of alternatives.

- Given a linear order $\triangleright$ over A, we say $\succ_i$ is *single-peaked* with respect to $\triangleright$ if for all alternatives $a, a' \in A$ such that either $\text{top}(\succ_i) \triangleright a \triangleright a'$ or $a' \triangleright a \triangleright \text{top}(\succ_i)$, we have $a \succ_i a'$. A profile $\boldsymbol{P}$ is *single-peaked* whenever there exists a linear ordering $\triangleright$ of the alternatives such that for every $i \in N$, $\succ_i$ is single-peaked with respect $\triangleright$.

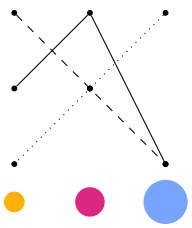
- Given a linear order $\triangleright$ of the agents, a profile $\boldsymbol{P}$ is *single-crossing* with respect to $\triangleright$ if for every pair of alternatives $a, a' \in A$ we have that $N^{\boldsymbol{P}}_{aa'}$ and $N^{\boldsymbol{P}}_{a'a}$ are intervals of the order $\triangleright$.

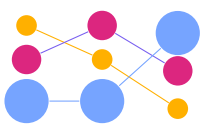
We give an example of a profile that is both single-peaked and single-crossing.

3.8. Example. Recall the example from Chapter 1 where agents gave their preferences over t-shirt sizes. We examine a variant of this scenario here. Three agents have expressed their preferences over t-shirt sizes—small, medium, and large—in the profile below, as they are going to place an order for group t-shirts. For example, agent 1 prefers a size small to medium to large.

$$\begin{aligned} &\text{Agent 1:} && S \succ M \succ L \\ &\text{Agent 2:} && M \succ S \succ L \\ &\text{Agent 3:} && L \succ M \succ S \end{aligned}$$

This profile is single-peaked along the small-to-large axis. In the figure on the left we have represented the t-shirt sizes by circles—a small yellow circle, a medium pink circle, and a large blue circle. Each agent's preferences can be placed such that there is a single-peak. The dashed line represents agent 1's preferences where the peak is the small size. The solid black line (agent 2's preferences) has a single peak at the medium. The dotted line represents agent 3's preferences where the peak is the large size. The profile is also single-crossing with respect to the order Agent 1 $\triangleright$ Agent 2 $\triangleright$ Agent 3. In the figure on the right each agent's preferences are represented by a single column—for example the leftmost column is agent 1's preferences—small over medium over large. We can see that for any two sizes, the lines representing them only cross a single time (take for example the blue and yellow lines which cross when we move from agent 2 to agent 3). This means that all agents who prefer blue over yellow will form an interval of the order.





Of course we can imagine this same structure popping up in scenarios that concern more serious topics. For example, agents preferences over tax rates might be single peaked with regard to the low-to-high axis. △

Restricting the possible input to a social choice function can be a way to avoid majority cycles as well as manipulation. All single-peaked and single-crossing profiles have a Condorcet winner. Therefore, any positive result for the Condorcet domain will hold for these domains as well. If we are considering a scenario where the profile is expected to be single-peaked or single-crossing, for example, robustness of a SCF can tell us that manipulation is unlikely to be a central concern.

3.2 Failure of Robustness

As decisive social choice functions are often preferred to indecisive ones, the existence of a weakly resolute tournament solution that is also a robust Condorcet extension would be a welcome result. Unfortunately, we will see that this goal is not achievable. We present a first impossibility result, showing there is no function that can meet our robustness requirement for all preference extensions.

3.9. Proposition. *For $m \geq 3$ and $n > 1$, no (weighted) tournament-solution SCF is robust under* all *preference extensions.*

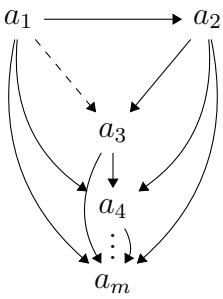
Proof: Let $A = \{a_1, \ldots, a_m\}$ and let f be a Condorcet-consistent SCF equivalent to the weighted tournament solution F, and let n be odd. We show that there must exist some preference over sets of alternatives such that F fails robustness. To that end, suppose agent 1's preferences over sets of alternatives $\overset{\circ}{\succeq}_1$ are such that $X \overset{\circ}{\succ}_1 Y$ if

- $|X| > 1$ and $|Y| = 1$, or
- $X = \{x\}$ and $Y = \{y\}$ for some $x, y \in A$ s.t. $x \succ y$.

These preferences satisfy our requirements for preference extensions.

Let $\boldsymbol{P}$ be the profile shown below, with the induced weighted tournament $\boldsymbol{T}$ depicted below that. We have omitted the weights, but it is easy to see that all edges have weight 1.

	agent 1	agent 2	agent 3	$\frac{n-3}{2}$ agents	$\frac{n-3}{2}$ agents
	a_4	a_1	a_3	a_1	a_m
	$\vdots$	a_2	a_1	a_2	a_{m-1}
	a_m	a_3	a_2	a_3	a_{m-2}
	a_2	a_4	a_4	a_4	a_{m-3}
$a_3 \rightarrow$	a_1	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$a_1 \rightarrow$	a_3	a_m	a_m	a_m	a_1



Note that $W(a_1, x) = 1$ for all $x \in A \setminus \{a_1\}$, $W(a_2, x) = 1$ for all $x \in A \setminus \{a_1, a_2\}$, and $W(a_3, x) = 1$ for all $x \in A \setminus \{a_1, a_2, a_3\}$. As f is a Condorcet extension, $f(\boldsymbol{P}) = \{a_1\}$.

Let $\boldsymbol{P'} =_{-1} \boldsymbol{P}$, where $a_3 \succ_1^{\boldsymbol{P'}} a_1$ (depicted on the left of the profile), meaning agent 1 reverses the edge (a_1, a_3) in the induced tournament $\boldsymbol{T'}$ by reversing the order of these alternatives in her ranking. Note that this is the case because $W(a_1, a_3) = 1$ meaning a single agent has the power to reverse this edge in the tournament. The weighted tournament $\boldsymbol{T'}$, induced by $\boldsymbol{P'}$, consists of a 3-cycle where all edges in the cycle have the same weight. This means any tournament that results from permuting these three alternatives would be isomorphic to $\boldsymbol{T'}$. The three alternatives a_1, a_2, and a_3 must therefore be treated symmetrically by F and any SCF f corresponding to F. In other words, either $a_1, a_2, a_3 \in f(\boldsymbol{P'})$ or $a_1, a_2, a_3 \notin f(\boldsymbol{P'})$. There are two cases we need to consider:

- $|f(\boldsymbol{P'})| > 1$. If this is the case, we know $f(\boldsymbol{P'}) \mathrel{\mathring{\succ}_i} f(\boldsymbol{P})$.
- $|f(\boldsymbol{P'})| = 1$. If this is the case, the single winner cannot be in $\{a_1, a_2, a_3\}$. But for any other alternative x, agent 1 strictly prefers x to a_1.

In either case, agent 1 will be able to successfully perform a Condorcet-manipulation in the profile $\boldsymbol{P}$, meaning f cannot be robust. □

We can state an equivalent result for even n. As we are not guaranteed that the (weighted) majority is a tournament when n is even—given that there may be ties in the majority relation, we state the result for SCFs directly.

3.10. Proposition. *For $m \geq 3$ and n being even, no neutral, Condorcet-consistent C2 SCF is robust under* all *preference extensions.*

Proof: Let $A = \{a_1, \ldots, a_m\}$ and let f be a neutral, Condorcet-consistent C2 SCF. Suppose agent 1's preferences over sets of alternatives $\mathring{\succeq}_1$ are such that $X \mathring{\succ}_1 Y$ if and only if one of the following holds:

- $|X| > 1$ and $|Y| = 1$, or
- $X = \{x\}$ and $Y = \{y\}$ for some $x, y \in A$ s.t. $x \succ y$.

These preferences are the same as those in the proof of Proposition 3.9.

Let $\boldsymbol{P}$ be the profile shown below.

agent 1	$\frac{n}{2}-1$ agents	$\frac{n}{2}$ agents
a_m	a_1	a_m
$\vdots$	$\vdots$	$\vdots$
a_1	a_m	a_1

Note that a_m is a Condorcet winner in $\boldsymbol{P}$, so $f(\boldsymbol{P}) = \{a_m\}$. Let $\boldsymbol{P'}$ be the profile where agent 1 reverses her preference order, and no other agent changes their preferences. Note that when agent 1 reverses her submitted order, all majority margins will be exactly 0, meaning all alternatives will be tied, and the resulting weighted majority graph will simply be the complete graph where all edges have weight 0. Thus $f(\boldsymbol{P'}) = A$, meaning $f(\boldsymbol{P'}) \mathrel{\mathring{\succ}_i} f(\boldsymbol{P})$. So as was the case in Proposition 3.9, agent 1 will be able to successfully manipulate in a profile with a Condorcet winner, meaning f cannot be robust. □

As no SCFs can be robust for *all* preference extensions, we redirect our search to those that may be robust for *some* preference extension. We first recall a result by McGarvey (1953), which we will use to prove the main result of this section. We include the proof for the sake of completeness.

3.11. Theorem (McGarvey, 1953). *Let A be a set of alternatives, and let $\geq$ be a complete relation over A. Then there is a profile $\boldsymbol{P} \in \mathcal{L}(A)^n$ for some even n such that $\succeq^{\boldsymbol{P}} = \geq$, and if $a > b$, there are $\frac{n}{2}+2$ agents ranking a over b in $\boldsymbol{P}$.*

Proof: For a set of alternatives A and a relation $\geq$ (with strict component $>$) over A, the profile $\boldsymbol{P}$ is constructed for an even number of agents $N = \{i_{ab}, j_{ab} \mid (a,b) \in >\}$ as follows. For every pair of alternatives such that $a > b$, there are two voters i_{ab} and j_{ab}, with the following preferences:

$$a \succ^{\boldsymbol{P}}_{i_{ab}} b \succ^{\boldsymbol{P}}_{i_{ab}} x_1 \succ^{\boldsymbol{P}}_{i_{ab}} \cdots \succ^{\boldsymbol{P}}_{i_{ab}} x_{|A|-2} \text{ and}$$

$$x_{|A|-2} \succ^{\boldsymbol{P}}_{j_{ab}} \cdots \succ^{\boldsymbol{P}}_{j_{ab}} x_1 \succ^{\boldsymbol{P}}_{j_{ab}} a \succ^{\boldsymbol{P}}_{j_{ab}} b,$$

Here $\{x_1, \ldots x_{|A|-2}\} = A \setminus \{a, b\}$. For each agent in $N \setminus \{i_{ab}, j_{ab}\}$ who prefers a over b, there will be exactly one corresponding agent who prefers b over a, meaning in the profile $\boldsymbol{P}$ exactly $\frac{n}{2}+2$ agents prefer a to b. As this holds for any pair of alternatives, it is clear that $\succeq^{\boldsymbol{P}} = \geq$. □

We now show that weakly resolute rules fail robustness for all preference extensions, and further, that they are the only rules that do so. This strengthens an

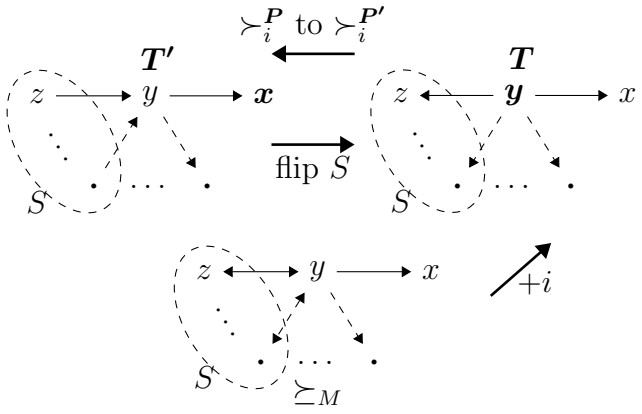


Figure 3.1: Tournaments $\boldsymbol{T}$ and $\boldsymbol{T'}$—with winners marked in bold—and relation $\succeq_M$ from the proof of Theorem 3.12. Ties are represented by bidirectional arrows.

observation from Brandt et al. (2016) stating that all weakly resolute rules are Kelly-manipulable.

3.12. Theorem. *A tournament-solution SCF is weakly resolute if and only if it fails robustness under all preference extensions.*

Proof: For the right-to-left direction we prove the contrapositive. That is, we suppose f is a tournament-solution SCF that fails weak resoluteness and show it must be robust under some preference extension. To see that this must be the case, note that any rule failing weak resoluteness never returns singletons outside $\mathcal{D}_{Condorcet}$. This means the preference extension ranging only over singletons would never (strictly) favour a larger set over the singleton set with the Condorcet winner.[2] As f will always return a set larger than a singleton outside the Condorcet domain, Condorcet-manipulation under these preferences is not possible, thereby making f robust under this preference extension.

For the left-to-right direction, let A be our set of alternatives. Suppose f is a weakly resolute tournament-solution SCF, equivalent to a tournament solution F. We show it is possible for an agent to manipulate f from a profile with a Condorcet winner under an arbitrary preference extension e, meaning f cannot be robust under any preference extension.

We first define two tournaments $\boldsymbol{T}$ and $\boldsymbol{T'}$, which we will show are single-agent variants. As F is equivalent to a weakly resolute SCF, there is some tournament $\boldsymbol{T'} = (A, \succ^{\boldsymbol{T'}}) \in \mathcal{T}(A) \setminus \mathcal{T}_{Condorcet}$ such that $F(\boldsymbol{T'}) = \{x\}$ for an alternative $x \in A$. As x is not a Condorcet winner in $\boldsymbol{T'}$, there must be some $y \in A$ such

[2]Note that all relevant manipulations here would be between singleton outcomes, meaning we are taking advantage of strategyproofness of the Condorcet, or majority, rule (Campbell and Kelly, 2003).

that $y \succ^{\boldsymbol{T'}} x$, and by the same reasoning, there must be at least one alternative $z \in A$ such that $z \succ^{\boldsymbol{T'}} y$. We conclude that the nodes $\{x, y, z\}$ and the edges $(y, x), (z, y)$ must be present in $\boldsymbol{T'}$. For a visual representation, see Figure 3.1.

Let $S = \overline{D}_{\boldsymbol{T'}}(y)$ be the dominators of y in $\boldsymbol{T'}$. We define a second tournament $\boldsymbol{T} = (A, \succ^{\boldsymbol{T}})$, where $y \succ^{\boldsymbol{T}} a$ for all $a \in S$, and $\succ^{\boldsymbol{T}}$ agrees with $\succ^{\boldsymbol{T'}}$ on all other pairs of alternatives. In other words, we simply reverse all incoming edges of y in $\boldsymbol{T'}$ to obtain $\boldsymbol{T}$. Note that this makes y a Condorcet winner in $\boldsymbol{T}$, meaning $F(\boldsymbol{T}) = \{y\}$.

We now show that $\boldsymbol{T}$ and $\boldsymbol{T'}$ are single-agent variants. We start by constructing a profile $\boldsymbol{P}$ that induces $\boldsymbol{T}$. To this end, consider a complete relation $\succeq_M$ (with strict component $\succ_M$, and symmetric component $\sim_M$) over A, such that $\succeq_M \;=\; \succ^{\boldsymbol{T}} \cup \{(a, y) \mid a \in S\}$. This means $\succeq_M$ and $\succ^{\boldsymbol{T}}$ agree on all pairs of alternatives except those for which $\boldsymbol{T}$ and $\boldsymbol{T'}$ differ. In those cases, $\succeq_M$ gives a tie between the alternatives. By Theorem 3.11, we know there exists a profile $\boldsymbol{P^*} = (\succ_1^{\boldsymbol{P^*}}, \ldots, \succ_n^{\boldsymbol{P^*}})$ with majority relation $\succeq_M$. Further, we know that we can construct $\boldsymbol{P^*}$ with an even number of agents n, such that for any $a, a' \in A$, where $a \succ_M a'$, there are exactly $\frac{n}{2} + 2$ agents who prefer a to a' in $\boldsymbol{P^*}$. We use $\boldsymbol{P^*}$ to construct the profile $\boldsymbol{P}$. Let $\boldsymbol{P} = (\succ_1^{\boldsymbol{P^*}}, \ldots, \succ_n^{\boldsymbol{P^*}}, \succ_i^{\boldsymbol{P}})$, where $x \succ_i^{\boldsymbol{P}} y \succ_i^{\boldsymbol{P}} a$ for all $a \in A \setminus \{x, y\}$.

To see that $\boldsymbol{P}$ induces tournament $\boldsymbol{T}$, note that for any pair of alternatives (a, a'), either

(i) $a \succ_M a'$—meaning $a \succ^{\boldsymbol{T}} a'$, and $\frac{n}{2} + 2$ prefer a to a' in $\boldsymbol{P^*}$, or

(ii) $a \sim_M a'$—meaning $a' = y$, and $a \in S$ (or vice versa).

If (i) is the case, a majority of agents in $\boldsymbol{P}$ will prefer a to a' regardless of agent i's preferences; $\frac{n}{2} + 2$ agents still form a strict majority of $n + 1$ agents. If on the other hand (ii) is the case, we know from agent i's preferences that $y \succ_i^{\boldsymbol{P}} a$. As these alternatives were tied in $\boldsymbol{P^*}$, adding agent i to the profile breaks these ties in favour of y, so a majority of agents in $\boldsymbol{P}$ will now prefer y to a. Thus the only differences between $\succeq_M$ and $\succeq^{\boldsymbol{P}}$ relate to the pairs on which $\succeq_M$ and $\succeq^{\boldsymbol{T}}$ differ. As the changes agree with $\succ^{\boldsymbol{T}}$, this makes $\succ^{\boldsymbol{T}} \;=\; \succ^{\boldsymbol{P}}$, meaning $\boldsymbol{P}$ induces $\boldsymbol{T}$. As $F(\boldsymbol{T}) = \{y\}$, we can conclude that $f(\boldsymbol{P}) = \{y\}$.

It now remains to construct a profile $\boldsymbol{P'}$ such that $\boldsymbol{P} =_{-i} \boldsymbol{P'}$ and $\boldsymbol{P'}$ induces $\boldsymbol{T'}$. Let $\boldsymbol{P'} = (\succ_1^{\boldsymbol{P}}, \ldots, \succ_n^{\boldsymbol{P}}, \succ_i^{\boldsymbol{P'}})$, and $x \succ_i^{\boldsymbol{P'}} a \;\; \succ_i^{\boldsymbol{P'}} y$, for all $a \in A \setminus \{x, y\}$, meaning agent i moves y to the bottom of their ranking. Clearly, $\boldsymbol{P'}$ is an i-variant of $\boldsymbol{P}$. In the tournament induced by $\boldsymbol{P'}$, it must be the case that the edges (a, y) for all $a \in S$ are present as the majority on these alternatives is dictated by agent i (and all other edges remain as they were in $\boldsymbol{T}$). As these edges correspond exactly to those where $\boldsymbol{T}$ and $\boldsymbol{T'}$ differ, $\boldsymbol{P'}$ must induce $\boldsymbol{T'}$, and as $F(\boldsymbol{T'}) = \{x\}$ we conclude $f(\boldsymbol{P'}) = \{x\}$.

Finally, let $\succeq_i^{\boldsymbol{P},e}$ be agent i's true preferences over sets of alternatives, extended according to e. It is immediately clear, as both outcomes are singletons

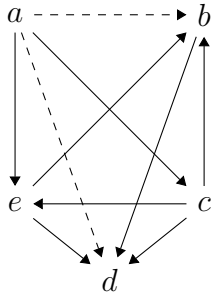
and $x \succ_i^{\boldsymbol{P}} y$, that $f(\boldsymbol{P'}) \succ_i^{\boldsymbol{P},e} f(\boldsymbol{P})$. As $\boldsymbol{P}$ has a Condorcet winner, this constitutes a Condorcet-manipulation, meaning f cannot be robust under preference extension e. □

We note that Theorem 3.12 applies to two of the most prominent Condorcet extensions—Copeland, and Slater. We now show that we can extend this result to all weighted counterparts of Slater, among them the Kemeny rule and the ranked pairs rule.

3.13. Proposition. *Any Condorcet-consistent weighted counterpart to Slater fails robustness.*

Proof: Let $\boldsymbol{P}$ be the 3-agent profile to the left with the corresponding weighted tournament to the right. Let f be a Condorcet-consistent weighted tournament solution SCF that corresponds to Slater.

Agent 1: $c \succ a \succ b \succ d \succ e$
Agent 2: $a \succ c \succ e \succ d \succ b$
Agent 3: $e \succ b \succ d \succ a \succ c$



Note that while we have omitted the weights in the graph, every edge in the tournament has weight 1. Because a is the Condorcet winner, $f(\boldsymbol{P}) = \{a\}$. Let $\boldsymbol{P'}$ be the profile where agent 1 submits the order $c \succ b \succ d \succ a \succ e$—meaning she drops a below both b and d, and all other agents submit the same order as in $\boldsymbol{P}$. Note that by doing so, agent 1 reverses edges (a, b) and (a, d) in the tournament, meaning there is no longer a Condorcet winner, though all weights remain 1. It is easy to see that c is the Slater winner in the equivalent (unweighted) tournament—reversing the edge (c, a) results in a linear order, and this is the only order that is reachable with a single flip. As all weights in the weighted tournament are 1, any weighted counterpart to Slater will therefore also return c, so $f(\boldsymbol{P'}) = \{c\}$. This means that $f(\boldsymbol{P'}) \mathrel{\mathring{\succ}}_1 f(\boldsymbol{P})$, making this a successful manipulation for agent 1 from a profile with a Condorcet winner. So f cannot be robust. □

Theorem 3.12 ceases to hold for even n, as the following shows. We will need the following notion. An alternative $a \in A$ is a *weak Condorcet winner* in profile $\boldsymbol{P}$ if for all $a' \in A$ we have $a \succeq^{\boldsymbol{P}} a'$. We now give an example of a Condorcet-consistent C1 SCF that is robust for some preference extension.

3.14. Example (Robust C1 SCF for even n). We define a C1 SCF f and a preference extension e such that $\mathring{\succ}_i = e(\succ_i)$ for some agent i, and f is robust under

e. Let f be the following SCF. If there is a Condorcet winner a, then f returns a, and only a. If there are any weak Condorcet winners, f returns the set of all Condorcet winners. Otherwise, f returns A. We define e as follows: $X \overset{\circ}{\succ}_i Y$ if and only if $X = \{x\}$, $Y = \{y\}$, and $x \succ_i y$.

Let $\boldsymbol{P}$ be an arbitrary n-agent profile with a Condorcet winner a. Because a is a Condorcet winner, we know that $W(a, x) \geq 2$ for any $x \in A \setminus \{a\}$. Let $\boldsymbol{P'}$ be an i-variant of $\boldsymbol{P}$. Clearly agent i cannot change her submitted preference in a way that creates some alternative $a' \in A$ such that $W(a', a) > 0$. This means a remains a weak Condorcet winner in $\boldsymbol{P'}$. So it must be the case that $a \in f(\boldsymbol{P'})$. Thus it cannot be the case that $f(\boldsymbol{P'}) \overset{\circ}{\succ}_i f(\boldsymbol{P})$ by definition of $\overset{\circ}{\succ}_i$ So f must be a robust Condorcet extension under e. △

3.3 Robust Tournament Solutions

In this section, we present our robustness results for several tournament-solution SCF, and their coarsenings. Our results hold for all weakly pessimistic extensions.

3.3.1 Relation to Kelly-Strategyproofness

While strategyproofness for, say, Gärdenfors preferences, is not easily satisfied, there are several tournament-solution SCFs that have been shown to be strategyproof for the Kelly preference extension (Kelly, 1977). Recall from Chapter 2 the definition of the Kelly extension: For any two sets X and Y in $2^A \setminus \{\emptyset\}$, $X \succ^k Y$ if and only if $x \succeq y$ for all $x \in X$ and all $y \in Y$, and there exists an $x \in X$ and a $y \in Y$ such that $x \succ y$. We say a SCF f is *Kelly-strategyproof* if no agent with Kelly preferences can successfully manipulate.

A SCF satisfying Gärdenfors-strategyproofness implies it also satisfies Kelly-strategyproofness, as the former must exclude more cases of manipulation to be satisfied. However, as robustness only requires taking into account comparisons where at least one singleton set is present, we can use strategyproofness results for Kelly preferences to show robustness for all weakly pessimistic extensions, including Gärdenfors.

3.15. Proposition. *If a Condorcet-consistent SCF f is Kelly-strategyproof, then it is a robust Condorcet extension under any weakly pessimistic preference extension.*

Proof: Let f be a rule that is Kelly-strategyproof and let e be a weakly pessimistic extension. That is, for any two profiles $\boldsymbol{P}$ and $\boldsymbol{P'}$, and any agent $i \in N$, if $\boldsymbol{P'} =_{-i} \boldsymbol{P}$, then $f(\boldsymbol{P'}) \not\succ_i^{\boldsymbol{P},k} f(\boldsymbol{P})$. Suppose $\boldsymbol{P}$ has a Condorcet winner, meaning $f(\boldsymbol{P}) = \{a\}$ for some $a \in A$. Because $f(\boldsymbol{P'}) \not\succ_i^{\boldsymbol{P},k} f(\boldsymbol{P})$, it cannot be the case that all elements of $f(\boldsymbol{P'})$ are preferred to a. So either

$f(\boldsymbol{P'}) = f(\boldsymbol{P})$ or there is some $a' \in f(\boldsymbol{P'})$ s.t. $a \succ_i^{\boldsymbol{P}} a'$. It is then immediate from the definition of a weakly pessimistic extension that $f(\boldsymbol{P'}) \not\succ_i^{\boldsymbol{P},e} f(\boldsymbol{P})$. □

Proposition 3.15 shows that we get robustness under weakly pessimistic preferences "for free" for Condorcet extensions known to be Kelly-strategyproof, such as the bipartisan set and the minimal covering set (Brandt, 2015).

3.3.2 Minimal Extending Set & Beyond

In this section we show that robustness for Condorcet extensions diverges from Kelly-strategyproofness, as we can find rules that satisfy the former while failing the latter. The minimal extending set F_{ME} is one of several tournament solutions that can be defined based on this notion of stability. The top cycle for example, is the union of all minimal CNL-stable sets (Brandt, 2011), where CNL is the tournament solution returning the set of all Condorcet nonlosers—meaning alternatives with at least one outgoing edge. The minimal extending set is Kelly-manipulable. However, we show it is still robust under weakly pessimistic preferences, and extend this result to all coarsenings of F_{ME}.

3.16. Theorem. F_{ME} is a robust Condorcet extension under all weakly pessimistic preference extensions.

Proof: For a set of alternatives A, and a set of agents N, let $\boldsymbol{P} =_{-i} \boldsymbol{P'}$ be i-variant profiles for an agent $i \in N$. Let $\boldsymbol{P}$ be such that $x \in A$ is the Condorcet winner in $\boldsymbol{P}$. Let $\boldsymbol{T} = (A, \succ^{\boldsymbol{T}})$ and $\boldsymbol{T'} = (A, \succ^{\boldsymbol{T'}})$ be the (single-agent variant) tournaments induced by $\boldsymbol{P}$ and $\boldsymbol{P'}$, respectively.

We assume $F_{\text{ME}}(\boldsymbol{T'}) \neq F_{\text{ME}}(\boldsymbol{T})$.[3] Because of this, we know $\overline{D}_{\boldsymbol{T'}}(x)$ is nonempty, as the two outcomes cannot differ if x remains a Condorcet winner in $\boldsymbol{T'}$. As $\boldsymbol{P} =_{-i} \boldsymbol{P'}$, any changes going from $\boldsymbol{T}$ to $\boldsymbol{T'}$ must be counter to agent i's preferences. This implies $x \succ_i^{\boldsymbol{P}} a$ for all $a \in \overline{D}_{\boldsymbol{T'}}(x)$. So, all alternatives in $\overline{D}_{\boldsymbol{T'}}(x)$ are worse than x to agent i.

We want to show that there is some minimal F_{Ba}-stable set S of $\boldsymbol{T'}$, such that $S \cap \overline{D}_{\boldsymbol{T'}}(x) \neq \emptyset$. This would guarantee the existence of an alternative $a \in \overline{D}_{\boldsymbol{T'}}(x)$ that is also in $F_{\text{ME}}(\boldsymbol{T'})$, precluding agent i with weakly pessimistic preferences from preferring this outcome to $F_{\text{ME}}(\boldsymbol{T})$.

So suppose for contradiction that S is a minimal F_{Ba}-stable set of $\boldsymbol{T'}$ such that $S \cap \overline{D}_{\boldsymbol{T'}}(x) = \emptyset$. The only way this can be the case is if $S \subseteq D_{\boldsymbol{T'}}(x) \cup \{x\}$. We consider two cases.

Case 1: Suppose $x \notin S$. As x dominates all alternatives in $D_{\boldsymbol{T'}}(x)$, it will dominate all alternatives in S, as $S \subseteq D_{\boldsymbol{T'}}(x)$. This means x is a Condorcet winner in the tournament $(S \cup \{x\}, \succ^{\boldsymbol{T'}}_{S \cup \{x\}})$, and thus, $x \in F_{\text{Ba}}(\succ^{\boldsymbol{T'}}_{S \cup \{x\}})$. So S cannot be a F_{Ba}-stable set, contradicting our assumption that it is a minimal one.

[3] If no such single-agent variants exist, robustness of the rule would immediately follow.

Case 2: Suppose instead $x \in S$. To reach our contradiction, we want to show there exists an alternative $a \in \overline{D}_{\boldsymbol{T'}}(x)$ such that $a \in F_{\text{Ba}}(\succ^{\boldsymbol{T'}}_{S\cup\{a\}})$—which would imply S is not F_{Ba}-stable. We use an algorithm proposed by Hudry (2004) to find such an alternative $a \in F_{\text{Ba}}(\succ^{\boldsymbol{T'}}_{S\cup\{a\}})$. We start at step 1 with a transitive subtournament of $(S \cup \{a\}, \succ^{\boldsymbol{T'}}_{S\cup\{a\}})$. Let $\boldsymbol{S}_1 = (\{x, a\}, \succ^{\boldsymbol{T'}}_{\{x,a\}})$, for some $a \in \overline{D}_{\boldsymbol{T'}}(x)$. We label all remaining elements of S—which are all elements of $D_{\boldsymbol{T'}}(x)$—in any order from 2 to $|S|$. At step k, we look at the alternative labelled k, and add it to the tournament $\boldsymbol{S}_{k-1}$ to create $\boldsymbol{S}_k$, if it does not break transitivity to do so. As a dominates x, and x dominates all $a' \in D_{\boldsymbol{T'}}(x)$, adding any alternative outside the dominion of a will break transitivity, as it will create a 3-cycle. Thus, at any step, an alternative $a' \in D_{\boldsymbol{T'}}(x)$ will only be added to the tournament if $a \succ^{\boldsymbol{T'}} a'$. When the algorithm terminates after iterating through all alternatives, we will be left with a subtournament $\boldsymbol{S}_{|S|}$ of $(S\cup\{a\}, \succ^{\boldsymbol{T'}}_{S\cup\{a\}})$. It is easy to see the resulting tournament will be transitive, and it will indeed be a maximal transitive subtournament of $(S\cup\{a\}, \succ^{\boldsymbol{T'}}_{S\cup\{a\}})$, as no further alternatives can be added to the tournament without breaking transitivity. Importantly, the maximal element of the resulting subtournament will be a, meaning $a \in F_{\text{Ba}}(\succ^{\boldsymbol{T'}}_{S\cup\{a\}})$. Thus, S cannot be an F_{Ba}-stable set, which contradicts our assumption that it is a minimal one.

As we have shown that no subset of $D_{\boldsymbol{T'}}(x) \cup \{x\}$ can be a F_{Ba}-stable set of $\boldsymbol{T'}$, any minimal F_{Ba}-stable set must contain at least one element of $\overline{D}_{\boldsymbol{T'}}(x)$, meaning it cannot be the case that $F_{\text{ME}}(\boldsymbol{T'}) \succ_i^{\boldsymbol{P},e} F_{\text{ME}}(\boldsymbol{T})$ when e is a weakly pessimistic preference extension. □

In terms of decisiveness, F_{ME} is among the more decisive tournament solutions that fail weak resoluteness, as it is a refinement of several prominent tournament solutions, including the top cycle and the Banks set (Brandt et al., 2017). We now show that all coarsenings of a robust SCF inherit the robustness property.

3.17. Lemma. *If a Condorcet-consistent SCF f is robust under weakly pessimistic preferences, then all Condorcet-consistent coarsenings of f are robust under weakly pessimistic preferences.*

Proof: Let f be a SCF that is robust under weakly pessimistic preferences, and let f' be a Condorcet-consistent coarsening of f. Let $\boldsymbol{P}$ be a profile with a Condorcet winner a. Note that $f(\boldsymbol{P}) = f'(\boldsymbol{P}) = \{a\}$ as both are Condorcet extensions.

Suppose $\boldsymbol{P'}$ is an i-variant of $\boldsymbol{P}$ for some agent $i \in N$. Because f is robust under weakly pessimistic preferences, either (i) there must be some $a' \in f(\boldsymbol{P'})$ such that $a \succ_i^{\boldsymbol{P}} a'$, or (ii) $f(\boldsymbol{P}) = f(\boldsymbol{P'})$.

If (i) is the case, then as $f(\boldsymbol{P'}) \subseteq f'(\boldsymbol{P'})$, a' is also an element of $f'(\boldsymbol{P'})$. As $f'(\boldsymbol{P}) = \{a\}$, we know $a' \in f'(\boldsymbol{P'}) \setminus f'(\boldsymbol{P})$, meaning by definition of a weakly pessimistic extension that it cannot be the case that $f'(\boldsymbol{P'}) \succ_i^{\boldsymbol{P},e} f'(\boldsymbol{P})$ for any weakly pessimistic extension e.

If (ii) is the case, we know a must also be the Condorcet winner in $\boldsymbol{P'}$ as f cannot satisfy weak resoluteness if it is robust under any preference extension, and therefore does not return singletons outside the Condorcet domain. Since f' is also a Condorcet extension, we know $f'(\boldsymbol{P'}) = \{a\}$, meaning $f'(\boldsymbol{P'}) \not\succ_i^{\boldsymbol{P},e} f'(\boldsymbol{P})$. □

3.18. Corollary. *Condorcet-consistent coarsenings of* F_{ME} are robust under weakly pessimistic preferences.

Corollary 3.18 follows from Lemma 3.17 and Theorem 3.16, and it establishes the robustness of the Banks set. Note that Corollary 3.18 is not restricted to tournament-solution SCFs, but holds for all Condorcet-consistent SCFs.

3.4 Summary

In this chapter, we have introduced the strategyproofness-related notion of a robust Condorcet extension. We have argued that Condorcet extensions that are robust are preferable to those that are not, as we can trust that they will return true Condorcet winners when they exists. We have introduced an axiom—weak resoluteness—and shown that no weakly resolute tournament solution can be a robust Condorcet extension. Finally, we have shown that the minimal extending set is a robust Condorcet extension under all weakly pessimistic preferences, and have extended this result to all coarsenings of F_{ME}.

We have argued that in lieu of searching for fully strategyproof rules, a fruitful endeavour is to explore immunity against more specific manipulations that may interact with, and compromise, other desirable properties satisfied by manipulable social choice functions. We have scratched the surface in this chapter, but have limited our exploration to robustness of irresolute rules in general, and (weighted) tournament solutions in particular. These are, of course, only a small class of all Condorcet extensions, and it remains to be seen if similar results can be obtained for other classes.

Chapter 4

Strategyproofness on Party-List Profiles in Multiwinner Voting

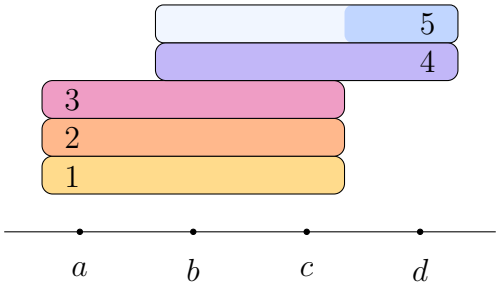
In multiwinner voting, agents vote on a set of candidates with the goal of electing a committee, or a subset of the candidates (Faliszewski et al., 2017). Applications for multiwinner voting rules range from parliamentary elections, to determining a list of nominees for an award, to online recommender systems. In this chapter we study strategyproofness of approval-based multiwinner voting rules (Kilgour, 2010). In this setting, each agent is asked to provide a subset of candidates that she approves of, and a set of winning candidates is chosen based on the approvals of the agents. We will look at rules that return committees of a fixed size k, which is a standard assumption in the literature. In many settings, however, it can also make sense to allow for committees of various sizes (Faliszewski et al., 2020; Kilgour, 2016), for example as an alternative to arbitrarily breaking ties.

As with other areas of social choice theory, an important aspect of studying multiwinner voting rules is determining their susceptibility to strategic manipulation. In recent years, we have seen impossibility results for approval-based multiwinner voting rules demonstrating that strategyproof rules are difficult to come by if we would like them to ensure some level of proportional representation. Peters (2018) establishes that no resolute approval-based rule—one that always returns a single winning committee—can be both proportional and strategyproof, even for very weak notions of proportionality and strategyproofness. Kluiving et al. (2020) show that an impossibility still remains when moving to irresolute rules. Our aim in this chapter is to examine whether there are any domain-specific "escape routes" for these impossibility results in the approval-based multiwinner voting framework. We devote focus in particular to a type of manipulation that is known as *free-riding* (Hylland, 1992; Schulze, 2004) (or subset-manipulation).

This is a simple and often successful way of manipulating multiwinner elections. Free-riding describes when an agent omits some alternative from their set of approved candidates, and in doing so, obtains a better outcome for herself. We also study what we call superset-manipulation, and disjoint-set-manipulation, defined analogously. We dip our toe in the water with an example of free-riding. Here we look at proportional approval voting (PAV). PAV maximises the total utility of agents, where an agent's utility for a committee containing ℓ of her approved candidates is determined by the following formula.

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{\ell}$$

4.1. Example (Manipulating by Free-Riding). Consider the profile depicted below. Here, agents $1, 2$, and 3 all approve the candidates a, b and c, while voters 4 and 5 approve candidates b, c, and d. Suppose we want to elect a committee comprising three candidates. In this profile, PAV will elect the committee $\{a, b, c\}$ as the unique winning committee.



If the last voter, agent 5 drops b and c from her approval set (represented in the lighter blue colour) and submits the approval set $\{d\}$, however, the unique winner will be $\{b, c, d\}$—her most preferred committee. The candidates b and c have enough support without agent 5, and dropping them from her approval set results in the inclusion of candidate d as PAV attempts to ensure all voters are represented in the outcome. Thus, agent 5 has an incentive to manipulate in this profile by submitting a subset of her truthful approval set. △

Because strategyproofness results for approval-based multiwinner voting have largely been negative, we consider weakening requirements to identify cases where we can obtain positive results. We do this by considering strategyproofness on a particular type of input—so-called party-list profiles. Intuitively, these are profiles where each candidate belongs to a single party, and agents approve of parties as a whole rather than any subset of the candidates. Our focus here is to examine whether manipulation is possible from a party-list profile to *any* other profile, not just those in the party-list domain. In particular, we look at a class of multiwinner voting rules known as *Thiele methods* (Thiele, 1895; Janson, 2016).

As we've seen in Chapter 3, honing in on a more well-behaved domain of profiles is a tried-and-true method for obtaining strategyproofness. Luckily for us, such restricted domains have already been studied for approval-based profiles. For example, Elkind and Lackner (2015) study a number of novel restrictions for this setting, including the party-list domain. We will consider several of these restrictions in more detail in Section 4.2.4.

Why do we care about strategyproofness on party-list profiles in particular? In many multiwinner elections, particularly parliamentary elections, the system used is a "closed-party" system that does not allow voters the freedom to pick and choose candidates from across parties. Party-list profiles can also appear in settings with lower stakes. A restaurant that only allows you to choose a fixed three-course meal for the table, for example, rather than creating your own from among all possible dishes is perhaps unknowingly creating "parties" for diners to choose among (if each dish only appears in one three-course meal). As we argued in Chapter 1, establishing barriers to manipulation on party-list profiles is appealing as it provides an argument against restricting the input to the voting rule. If your voting rule is strategyproof on party-lists you can do away with restricting the input to the rule without incurring *needless risk* of strategic voting. If the "true" profile is a party-list profile and we use a voting rule that is strategyproof on party-list profiles, it will not matter whether we restrict the domain of the voting rule. The reported profile will be the same, even under the assumption that agents behave strategically. If the "true profile" is not a party-list profile the reported profile will obviously differ if we only allow party-list profiles as input as the true profile falls outside that domain. We know agents cannot fully express their true opinion, and thus, the outcome cannot claim to reflect the true opinions of voters. With no domain restriction, agents can fully express their opinion on the candidates, though of course we run the risk of the outcome being the result of a manipulation by a voter—even with a rule that is strategyproof on party-list profiles. Of course, we cannot know ahead of time whether the party-list domain is in fact expressive enough. Strategyproofness on party-list profiles guarantees that we only risk strategic manipulation in the unrestricted case if the domain restriction would not allow voters to express their true opinion in the first place. Let us demonstrate this idea with an example.

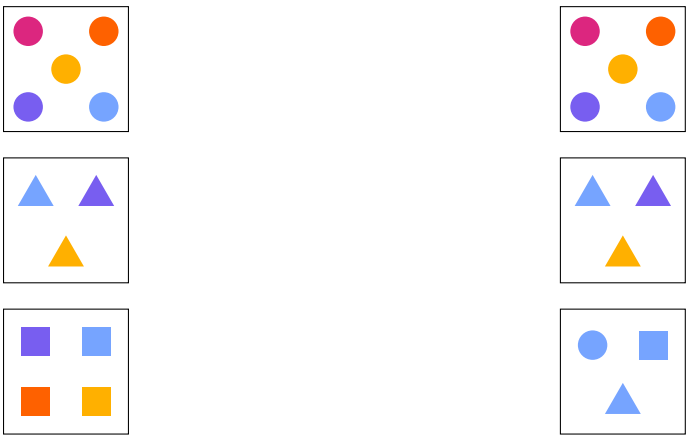
4.2. Example. The country of Arrovia is holding a multiwinner election. They have three parties—the circle party, the square party, and the triangle party. They will use an approval-based multiwinner voting rule f that is strategyproof on party-list profiles, and are deciding whether to restrict the domain of f. Through opinion polling they have determined that there are two possible "truthful" profiles of approval ballots, though they are not sure which is the real one.

Let's first look at the profile on the left—what we've called 'possible world A'. Here the first agent approves all circles, the second agent approves all triangles,

and the third agent approves all squares. We assume that these are the truthful opinions of the agents in this possible world.

- Suppose we restrict the domain of f, in this case to the party-list domain. Because this is a party-list profile, we know that we will not run into any problems by doing this.

- Suppose we do not restrict the domain of the voting rule. Then, because f is strategyproof on party-list profiles, we can exclude the possibility of manipulation.



A *B*

Thus, there is no benefit to restricting the input to f in this possible world.

Let us now look at the right profile—'possible world B'. Again we suppose this is a profile of truthful opinions in this possible world. Here the first agent approves all circles, the second agent approves all triangles, and the third agent approves all blue shapes. Note that this is not a party-list profile. The bottom agent approves candidates from all parties (and there is no way to redefine the parties to make this a party-list profile).

- Suppose we do not restrict the domain of f. Then it is possible that at least one agent will have an incentive to manipulate. The reported profile *may*, therefore, differ from this truthful one.

- Suppose we do restrict the domain of f to the party-list domain in an attempt to prevent manipulation. Because this is a not party-list profile, we can say with certainty that at least one agent *must* report an approval set that is not their truthful one.

This example shows that restricting the domain of f will prevent manipulation, but at the cost of limiting the voters ability to express their true opinion. Of course we cannot know ahead of time which world is the real one. But if Arrovia

uses a voting rule that is strategyproof on party-list profiles, the two systems—restricted vs. unrestricted domain—would produce the same outcome if the truthful profile is a party-list one. And *if* in fact the party-list ballots are not expressive enough, the outcome of the non-restricted election will potentially better reflect Arrovians' interests (after all, there is no guarantee someone will manipulate on all non party-list profiles). △

As mentioned, we will discuss three types of manipulation in this chapter. We devote Section 4.3 to *free-riding*, and present strategyproofness results for a large class of preferences. We then study *superset-strategyproofness* and *disjoint-set-strategyproofness* in Section 4.4. While these types of manipulation are not as well studied, they are natural extensions of the idea of free-riding. For superset- and disjoint-set-strategyproofness, we are able to obtain results that are independent of the preference extension. For optimistic agents, we are able to show that Thiele rules are fully strategyproof on party-list profiles. We present these results in Section 4.5. Sections 4.4 and 4.5 together demonstrate the interplay between the choice of preference extension and the strength of the strategyproofness axioms we use. Our results for superset-strategyproofness hold for all strongly reflective preference extensions, as this is a somewhat weak strategyproofness axiom. In contrast, our positive result in Section 4.5 hold for a much stronger strategyproofness axiom, but in turn pertains only to one particular preference extension.

4.1 Preliminaries

Let $\mathcal{C}$ be a finite set of *candidates*, and $N = \{1, \dots, n\}$ a finite set of *agents* or *voters*—we will use these terms interchangeably. A *profile* $\boldsymbol{A} = (A_1, \dots, A_n)$ is a vector of *approval sets*, where $A_i \subseteq \mathcal{C}$ is the set of candidates approved by agent i in the profile $\boldsymbol{A}$. The set of *supporters* $N_a^{\boldsymbol{A}}$ of a candidate a in profile $\boldsymbol{A}$ is the set of agents who approve it—$N_a^{\boldsymbol{A}} = \{i \in N \mid a \in A_i\}$. We write $\mathcal{P}(\mathcal{C})$ to denote all subsets of $\mathcal{C}$—in other words, all possible approval sets—and $\mathcal{P}(\mathcal{C})^n$ to denote the set of all profiles for n agents. We write $\mathcal{P}_k(\mathcal{C})$ to mean the set of all k-size subsets of $\mathcal{C}$. We will often call these subsets *committees*. For two profiles $\boldsymbol{A}$ and $\boldsymbol{A'}$, and an agent $i \in N$, we write $\boldsymbol{A} =_{-i} \boldsymbol{A'}$—and say they are *$i$-variants*—if $A_j = A'_j$ for all $j \in N \setminus \{i\}$. We restrict our attention in this chapter to manipulation on party-list profiles. A profile $\boldsymbol{A}$ is a *party-list profile* if for all $i, j \in N$, either $A_i = A_j$ or $A_i \cap A_j = \emptyset$.

4.1.1 Approval-Based Multiwinner Voting Rules

We define voting rules relative to an outcome of size k. An (irresolute) *approval-based k-committee rule* f takes as input a profile $\boldsymbol{A}$ and returns a set $f(\boldsymbol{A})$ of

k-sized committees—or *k-committees*. Formally f is a function from profiles to k-sized subsets of $\mathcal{C}$:

$$f : \mathcal{P}(\mathcal{C})^n \to 2^{\mathcal{P}_k(\mathcal{C})} \setminus \{\emptyset\}$$

A *resolute* rule is one that always returns singletons. We will sometimes refer to these functions as *multiwinner voting rules* (or simply voting rules), mainly when k is clear from context. Note that our definition differs from the common approach of using a profile and target committee size as input to the voting rule. In this alternate framework, it is not necessary to define a separate voting rule for each k. Our results hold for either framework, as they do not depend on the value of k.

The rules we will examine in this chapter are *Thiele methods* (Thiele, 1895; Janson, 2016). Given a vector of weights $\boldsymbol{w} = (w_1, w_2, \dots)$, we define the *utility*

$$u_i^{\boldsymbol{A}}(C, \boldsymbol{w}) = \sum_{x=1}^{|A_i \cap C|} w_x$$

of agent i for committee C, given the approval set A_i. The $\boldsymbol{w}$-score of a committee C in a profile $\boldsymbol{A}$ is

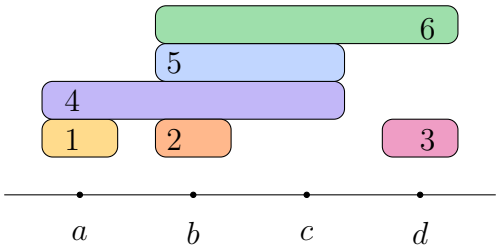
$$u_N^{\boldsymbol{A}}(C, \boldsymbol{w}) = \sum_{i \in N} u_i^{\boldsymbol{A}}(C, \boldsymbol{w}).$$

When the weight vector $\boldsymbol{w}$ is clear from context, we will omit it from the notation and simply write $u_i^{\boldsymbol{A}}(C)$. A k-committee rule $f_{\boldsymbol{w}}$ is a *Thiele method* (or Thiele rule) if for a vector of nonnegative weights $\boldsymbol{w} = (w_1, w_2, \dots)$, where $w_1 = 1$ and $w_j \geq w_{j+1}$, and a profile $\boldsymbol{A}$, the rule $f_{\boldsymbol{w}}$ includes a committee C in the outcome if and only if C is a k-committee with a maximal $\boldsymbol{w}$-score in $\boldsymbol{A}$.[1] Thiele methods are based on the notion of diminishing returns for the agents—the second representative gained does not increase the agents' utility as much as the first representative. Thus, tasked with deciding whether to give one agent a second representative or to give another agent their first, Thiele rules will never opt for the former over the latter. Let us now define three well-known Thiele methods.

- *Approval voting* (AV) is the Thiele method defined by weight vector $(1, \dots, 1)$.
- *Approval-based Chamberlin-Courant* (approval-based CC) is defined by weight vector $(1, 0, \dots, 0)$.
- *Proportional approval voting* (PAV) is the Thiele method defined by the weight vector $(1, \frac{1}{2}, \frac{1}{3}, \dots)$.

[1] Note that while we require that $w_1 = 1$, we can always rescale any weight vector where this is not the case given that all weights are non-negative.

4.3. Example (Outcomes of AV, approval-based-CC, and PAV). Suppose we are looking for a committee of size 2. Consider the following profile of approval ballots:



AV will return the committee $C = \{b, c\}$ as candidate b has 4 approvals, candidate c has 3 approvals, and both a and d have only 2 approvals each. So $u_N^{\boldsymbol{A}}(C) = 0 + 1 + 0 + 2 + 2 + 2 = 7$, and this is the highest possible sum of utilities.

Approval-CC on the other hand, will return the committees $C_1 = \{a, b\}$ and $C_2 = \{b, d\}$ as these committees "represent" 5 agents each, while any other committee of size 2 "represents" only 4 agents. Let's check the utilities for the committee C_1. Note that the utility of each agent can be either 0 or 1. So $u_N^{\boldsymbol{A}}(C_1) = 1 + 1 + 0 + 1 + 1 + 1 = 5$. This is indeed the highest value we can get. If we try to find a committee of size 2 that also represents agent 3, we have to omit either a or b, each of which are the only possible representative for voters 1 and 2, respectively.

Finally, PAV will return the committees $C_1 = \{a, b\}$, $C_2 = \{b, c\}$ and $C_3 = \{b, d\}$, which each have a "PAV score" of 5.5. For example $\{a, b\}$ gives voter 4 utility 1.5—two of her approved candidates are included and so $u_i^{\boldsymbol{A}}(C_1) = 1 + \frac{1}{2}$—and gives voters $1, 4, 2$ and 6 each utility 1, summing to a score of 5.5. It is easy to check that any other committee of size 2 will have a lower score. For example, the committee $\{a, c\}$ has score 4.5, again voter 4 gets utility 1.5, but there are now only 3 other voters who each get utility 1—voters 1, 4 and 6. △

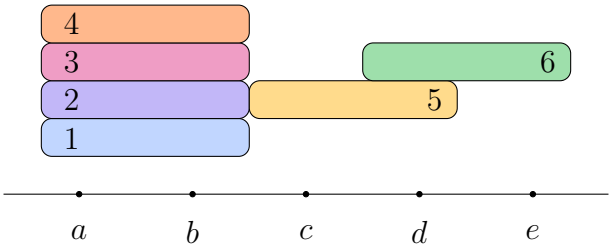
4.1.2 Proportionality and Voter Representation

A selling point for Thiele methods is the fact that they emphasise voter representation. PAV has been shown by Aziz et al. (2017) to satisfy a particularly strong proportional representation axiom—extended justified representation, and is therefore of special interest. Here we formulate these representation axioms for irresolute rules by requiring that every committee in the outcome meet the required conditions, and state their results for irresolute rules. We can do this because these particular results do not depend in any way on the tie-breaking rule—meaning all committees in the outcome satisfy the relevant axioms.

For any profile $\boldsymbol{A}$ and a positive integer $\ell \leq k$, we say a committee $C \in \mathcal{P}_k(\mathcal{C})$ provides ℓ-representation in $\boldsymbol{A}$ if there is no subset $N' \subseteq N$ of voters such that $|N'| \geq \ell \cdot \frac{n}{k}$ and $|\bigcap_{i \in N'} A_i| \geq \ell$, but $|C \cap A_i| < \ell$ for all voters $i \in N$. A rule f

satisfies *justified representation* if for any profile $\boldsymbol{A}$ we have that every $C \in f(\boldsymbol{A})$ provides 1-representation. More simply put, justified representation requires that if if k candidates are to be selected, then each group of size $\frac{n}{k}$ should have at least one "representative". So if a sufficiently large portion of the electorate manage to agree on at least one candidate they want, then we cannot not leave every member of that group unrepresented. We say f satisfies *extended justified representation* if for any profile $\boldsymbol{A}$ we have that every $C \in f(\boldsymbol{A})$ provides ℓ-representation for all ℓ, $1 \leq \ell \leq k$. Extended justified representation strengthens justified representation by requiring that if a large enough group of agents agree on ℓ candidates, then for every committee in the outcome there should be at least one agent in the group who gets ℓ representatives—meaning ℓ of this agent's approved candidates are included in the committee.

4.4. Example. Consider the approval profile $\boldsymbol{A}$ below, and let f be some 3-committee rule such that $f(\boldsymbol{A}) = \{\{a, c, e\}\}$.



In order to check whether this committee provides 1-representation (and thus whether f could satisfy justified representation) we need to check every subset of voters of size $\frac{6}{3}$ to see if they are given proper representation. First, as agents 1 through 4 agree on a (and b), they are a cohesive group deserving of a representative. As $A_1 \cap \{a, c, e\} = \{a\}$, this group is properly represented in the outcome. Then, agents 5 and 6 agree on candidate d, and $A_6 \cap \{a, c, e\} = \{e\}$, so this group is also properly represented. We can see that justified representation is not violated in this instance.

Can f possibly satisfy extended justified representation? A simple counterexample tells us this is not possible. Agents 1 through 4 agree on two candidates, a and b. In order for f to satisfy 2-representation, there must be at least one agent among the four who has at least two approved candidates in the outcome. But we can easily see that this is not the case, for each agent in the set, only one of their approved candidates appears in the outcome. Thus f does not satisfy extended justified representation. △

We now state two results by Aziz et al. (2017) that pertain to these axioms.

4.5. Theorem (Aziz et al., 2017). *For weight vectors $\boldsymbol{w}$ such that $w_1 = 1$ and $w_j \leq \frac{1}{j}$, the Thiele rule $f_{\boldsymbol{w}}$ satisfies justified representation.*

4.6. THEOREM (Aziz et al., 2017). *PAV is the only Thiele rule with $w_1 = 1$ that satisfies extended justified representation.*

Theorem 4.5 covers both PAV and approval-CC. These results confirm the intuition that Thiele rules are attempting to achieve some type of proportional representation, and that PAV is particularly successful in this endeavour. While this is good news for representative democracy, we'll see that it does not bode well in terms of strategyproofness.

4.2 Strategyproofness in Multiwinner Voting

Strategic manipulation in multiwinner elections has been studied from several angles. Most relevant for us is the axiomatic study initiated by Lackner and Skowron (2018) who confirm that most approval-based multiwinner voting rules—with the exception of AV—are susceptible to strategic manipulation. In contrast with other axiomatic work on strategyproofness in multiwinner voting (see for example, Peters (2018)), they do not assume resoluteness. Lackner and Skowron (2018) study three axioms—independence of irrelevant alternatives, monotonicity, and SD-strategyproofness (a strategyproofness axiom rooted in the notion of stochastic dominance (Bogomolnaia and Moulin, 2001)). They find that independence and monotonicity each exclude a certain type of manipulation. Their monotonicity axiom is closely related to the notion of free-riding. Informally, monotonicity requires that an extra approval for a candidate that is already in a winning committee never results in that committee going from winning to losing. Thus, monotonicity implies that an agent cannot get a committee to go from losing to winning by removing a candidate in the committee from their approval set. Lackner and Skowron (2018) show that most Thiele rules fail this axiom, which is in some sense expected, as we know these rules are susceptible to free-riding. They do show however, that Thiele rules satisfy their independence axiom.

Further from our focus, Yang and Wang (2018) study strategic aspects of multiwinner voting relative to various graph-based restrictions on the winning committees. Among these restrictions is the often-seen assumption that committees must be of a fixed size k. Laslier and Van der Straeten (2016) study strategic voting of multiwinner approval voting in a probabilistic setting. Bredereck et al. (2016) examine the related notion of bribery in a multiwinner setting. On the more computational side, Bartholdi and Orlin (1991) show, for example, that determining whether there exists a possible manipulation of Single Transferable Vote is NP-complete, establishing that computational complexity can be a barrier to manipulation also for multiwinner voting. More recently, the computational complexity of strategic manipulation in multiwinner voting has been studied by Aziz et al. (2015), Bredereck et al. (2018), and Meir et al. (2008). As an example of a more negative result, Obraztsova et al. (2013) give polynomial-time

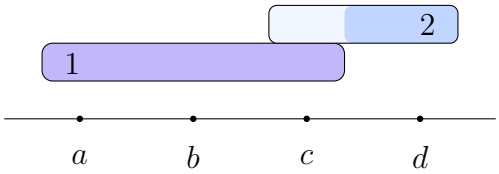
algorithms for manipulation of multiwinner scoring rules. On the related notion of robustness, Bredereck et al. (2017) examine how robust the outcome of multiwinner voting rules are to small changes in the input. While they interpret this as the possibility of mistakes made by agents when submitting their preferences, we can also think of these small perturbations as strategic actions by the voters. Gawron and Faliszewski (2019) study robustness in approval-based multiwinner rules, and Misra and Sonar (2019) study robustness in restricted domains.

4.2.1 Preferences and Manipulability

As agents submit approval sets, we need to explicitly specify their preference ranking over committees. We define agent i's preferences over k-committees $\succeq_i$ (with strict part $\succ_i$) as follows: For two k-committees C and C', it is the case that $C \succeq_i C'$ if and only if $|A_i \cap C| \geq |A_i \cap C'|$. We write $C \sim_i C'$ if $C \succeq_i C'$ and $C' \succeq_i C$—meaning agent i is indifferent between the two sets. Agents will never be confronted with committees of different sizes, so such committees are considered incomparable. Note that $\succeq_i$ is defined relative to A_i, which is what we take to be the agent's truthful approval set.

Given a preference order $\mathring{\succeq}_i$ over outcomes (i.e., sets of committees), we say an irresolute rule f is *manipulable* by agent i in the profile $\boldsymbol{A}$ if there exists another profile $\boldsymbol{A'} =_{-i} \boldsymbol{A}$ such that $f(\boldsymbol{A'}) \mathring{\succ}_i f(\boldsymbol{A})$. A rule is *strategyproof* under the preference extension e if it is not manipulable in any profile by any agent with preferences $\mathring{\succeq}_i = e(\succeq_i)$. Of the three Thiele rules we have defined, approval voting is strategyproof (but fails our proportionality axioms). Both PAV and approval-CC are manipulable. We have already seen an example of PAV being manipulable. Here we also demonstrate a manipulation of approval-CC.

4.7. Example. Let $k = 3$ and consider the 2-agent profile below.



Even in this very simple profile, manipulation of approval-CC is possible. First, note that any size-3 subset of the candidates will represent both agents, and will therefore be included in the outcome—$f(\boldsymbol{A}) = \{\{a, b, c\}, \{a, c, d\}, \{a, b, d\}, \{b, c, d\}\}$ Note however, that if agent 2 (untruthfully) approves only candidate d, then any committee that does not include d will no longer represent all agents. So the outcome now becomes $f(\boldsymbol{A}) = \{\{a, c, d\}, \{a, b, d\}, \{b, c, d\}\}$. For agent 2 all the remaining committees are (weakly) preferred to the committee $\{a, b, c\}$ that is no longer in the outcome, and there is a committee that is strictly preferred. This would therefore be a

successful manipulation if agent 2's preferences are extended according to, for example, the Kelly preference extension. △

This susceptibility to manipulation is not limited to PAV and approval-CC. The negative results we have seen affect many other Thiele rules (when we consider the domain of all profiles).

4.2.2 Impossibilities

Peters (2018) has shown that there are no resolute approval-based multiwinner voting rules that can simultaneously satisfy even very weak strategyproofness and proportionality axioms. Before we state his result, we need to define three axioms for resolute voting rules. A rule f satisfies...

- ...*strategyproofness* if for any profile $\boldsymbol{A}$ and i-variant $\boldsymbol{A'}$ such that $A'_i \subseteq A_i$, we do not have $f(\boldsymbol{A'}) \cap A_i \supset f(\boldsymbol{A}) \cap A_i$.
- ...*proportionality* if for any party-list profile $\boldsymbol{A}$ where some singleton approval set $\{a\}$ appears in the profile at least $\frac{n}{k}$ times, we have that $a \in f(\boldsymbol{A})$.
- ...*weak efficiency* if for any profile $\boldsymbol{A}$ where we have $|\bigcup_{i \in N} A_i| \geq k$ and there is some $a \in \mathcal{C}$ s.t. $a \notin A_i$ for all i, it is not the case that $a \in f(\boldsymbol{A})$.

4.8. Theorem (Peters, 2018). *There exists no resolute approval-based multiwinner voting rule that satisfies strategyproofness, proportionality, and weak efficiency.*

Note that this strategyproofness axiom is not only weakened by considering subset manipulations, but also by looking at a particular type of set-based preferences. Here, an agent prefers the outcome $f(\boldsymbol{A'})$ only if this committee adds additional candidates from the agent's approval set, and removes none of the agent's approved candidates that were already present in $f(\boldsymbol{A})$.

Among the customary responses to such an impossibility result in social choice is to consider what happens in the irresolute case. Kluiving et al. (2020) did just that for irresolute approval-based multiwinner rules. They obtain a similar result using the following axioms for irresolute rules. Recall from Chapter 2 that a set A is strictly "Kelly-preferred" to a set B if all elements of A are weakly preferred to all elements of B, and at least one element of A is strictly preferred to an element of B. A voting rule is Kelly-manipulable if an agent with Kelly preferences can bring about a more preferred outcome by submitting an untruthful vote. A rule f is...

- ...*Kelly-strategyproof* if it is not manipulable in any profile by an agent with Kelly preferences.

- ...*minimally proportional* if for any party-list profile $\boldsymbol{A}$ and any candidate $a \in \mathcal{C}$, if some singleton approval set $\{a\}$ appears in the profile at least $\frac{n}{k}$ times, then $a \in C$ for all $C \in f(\boldsymbol{A})$.
- ...*Pareto efficient* if for any profile $\boldsymbol{A}$ and any two committees C, C' where $C \succeq_i C'$ for all $i \in N$ and $C \succ_i C'$ for some $i \in N$, it is not the case that $C' \in f(\boldsymbol{A})$.

4.9. Theorem (Kluiving et al., 2020). *There exists no irresolute approval-based multiwinner voting rule that is minimally proportional, Pareto efficient, and Kelly-strategyproof.*

How do these results relate to our project in this chapter? We do two things to circumvent the impossibility by Peters (2018). The first is to consider irresolute rules—which we know from Kluiving et al. (2020) is not a solution on its own. The second is to consider only manipulations that occur from a profile of a certain type—party-list profiles. In fact, we can show that simply considering the restriction to party-list profiles alone would also not solve our problem in the case of resolute rules. As we will see in Section 4.3, it is still possible to manipulate the resolute PAV rule on a party-list profile. The more surprising part of our results is that we are able to establish strategyproofness on party-list profiles for irresolute rules, and indeed are able to do so i) for rules that satisfy *Pareto efficiency* (such as PAV) and ii) for several more permissive preference extensions (compared to Kelly).

4.2.3 Types of Manipulation

Our focus is domains on which proportional multiwinner rules can be immune to certain types of manipulation by certain types of agents. We now define the three types of manipulation we will consider in this chapter: subset-manipulation—or free-riding—superset-manipulation, and disjoint-set-manipulation. We will then define their corresponding strategyproofness axioms.

Given a rule f and two profiles $\boldsymbol{A}$ and $\boldsymbol{A'}$ such that $\boldsymbol{A} =_i \boldsymbol{A'}$ for some agent i with preference order $\overset{\circ}{\succeq}_i$ over outcomes, we say an agent is able to *free-ride* if $A'_i \subset A_i$ and $f(\boldsymbol{A'}) \overset{\circ}{\succ}_i f(\boldsymbol{A})$. Free-riding is particularly relevant for rules that attempt to achieve some level of representation for all voters. A free-rider often omits a popular candidate from their approvals so this candidate does not count toward their own representation. A rule f is *superset-manipulable* by agent i if $A'_i \supset A_i$ and $f(\boldsymbol{A'}) \overset{\circ}{\succ}_i f(\boldsymbol{A})$. Finally, a rule f is *disjoint-set-manipulable* if $A'_i \cap A_i = \emptyset$ and $f(\boldsymbol{A'}) \overset{\circ}{\succ}_i f(\boldsymbol{A})$.

A rule f is is *immune to free-riding* under the preference extension e if no agent with preferences $\overset{\circ}{\succeq}_i = e(\succeq_i)$ can free-ride in any profile. It is *superset-strategyproof* under the preference extension e if no agent with preferences $\overset{\circ}{\succeq}_i = e(\succeq_i)$ can superset-manipulate in any profile, and is *disjoint-set-strategyproof* under the

preference extension e if no agent with preferences $\overset{\circ}{\succeq}_i = e(\succeq_i)$ can disjoint-set-manipulate in any profile.

4.2.4 Manipulation on Restricted Domains

Recall that a profile $\boldsymbol{A}$ is a *party-list profile* if for all $i, j \in N$, either $A_i = A_j$ or $A_i \cap A_j = \emptyset$. The main notion we explore in this chapter is strategyproofness relative to the domain of party-list profiles. We say a rule is *strategyproof on party-list profiles*—or *party-list-strategyproof*—under a preference extension e if no agent with preferences $\overset{\circ}{\succeq}_i = e(\succeq_i)$ can manipulate on a party-list profile.

As we touched on in the introduction to this chapter, party-list profiles have practical relevance for multiwinner elections. There are also other natural ways we can restrict the domain of approval profiles. Domain restrictions for dichotomous preferences have not gotten as much attention compared to the myriad restricted domains that exist for ranked preference profiles. Nevertheless, several such domains have been studied by Elkind and Lackner (2015). Our goal in this chapter is to find domains where agents have no incentive to manipulate. As such, these domains are excellent candidates for our project. We now define four of the twelve domain restrictions for approval profiles defined by Elkind and Lackner (2015) and show how PAV fails strategyproofness on these domains. This result holds for all domains studied by Elkind and Lackner (2015), with the exception of party-list profiles.

- A profile $\boldsymbol{A}$ satisfies *voter extrema interval* if there exists an ordering $\triangleright$ of the agents in N such that for every candidate $a \in \mathcal{C}$, we have that for all $i \in N_a^{\boldsymbol{A}}$ and all $i' \notin N_a^{\boldsymbol{A}}$, it is the case that $i \triangleright i'$.

- A profile $\boldsymbol{A}$ satisfies *voter interval* if there exists an ordering $\triangleright$ of the agents such that for every $a \in \mathcal{C}$, we have that $N_a^{\boldsymbol{A}} = \{i \in N \mid i_\ell \triangleright i \triangleright i_r\}$ for some $i_\ell, i_r \in N$.

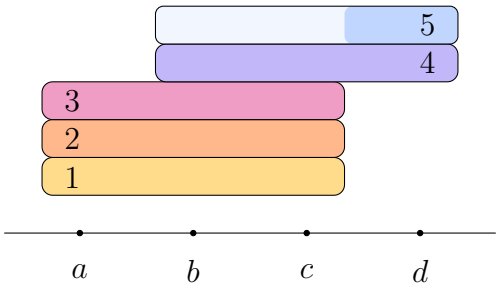
- A profile $\boldsymbol{A}$ satisfies *candidate extrema interval* if there exists an ordering $\triangleright$ of the candidates in $\mathcal{C}$ such that for every agent $i \in N$, we have that for all $a \in A_i$ and all $b \notin A_i$, it is the case that $a \triangleright b$.

- A profile $\boldsymbol{A}$ satisfies *candidate interval* if there exists an ordering $\triangleright$ of the candidates in $\mathcal{C}$ such that for every agent $i \in N$, we have that for all $A_i = \{a \in C \mid \ell \triangleright a \triangleright r\}$ for some $\ell, r \in \mathcal{C}$.

We can show with a simple counterexample that strategyproofness for the class of Thiele rules is not attainable for any of these domains, and that this is the case independent of the preference extension. Because the domains studied by Elkind and Lackner (2015), with the exception of the party-list domain, are defined by weaker conditions than the four domains above, this counterexample also applies

to them. This is a strong indication that the party-list domain is indeed the most fruitful avenue for us to explore.

4.10. Example (Strategyproofness for Other Domains). Recall the profile from Example 4.1.



First, note that in the profile represented above, both the agents and the candidates are represented in an ordering that shows they satisfy each of the four restrictions above.[2] As we saw in Example 4.1, when agent 5 submits her truthful approval set, PAV elects $\{a, b, c\}$ as the unique winning committee. If she submits a subset of her true approval set she obtains an outcome, $\{b, c, d\}$, that she strictly prefers to $\{a, b, c\}$. Both outcomes are single committees, making the preference extension irrelevant. So PAV cannot be immune to free-riding on these domains, or on any that contain them. △

4.3 Free-Riding

Recall that free-riding describes when an agent submits a subset of their truthful approval set, in order to obtain a better outcome. Recall also, from Chapter 2, that general Gärdenfors preferences are a class of preference extensions that generalise the concept behind the Gärdenfors extension. We show that free-riding on party-list profiles is not possible for agents with general Gärdenfors preferences. However, we will also see quite reasonable scenarios where such manipulation, even on party-list profiles, remains possible. For example, immunity to free-riding is the strategyproofness notion used by Peters (2018), meaning his impossibility result holds even for this limited type of manipulation.

Our first order of business is to establish two lemmas. We write $\succeq_i'$ to mean agent i's preference order over committees, under the assumption that A_i' is the agent's truthful approval set—i.e., $C' \succeq_i' C$ if $|A_i' \cap C'| \geq |A_i' \cap C|$. Lemma 4.11 identifies cases where the preference orders $\succeq_i$ and $\succeq_i'$ coincide.

[2] As noted, the profile also falls under the other restrictions considered by Elkind and Lackner (2015) that are more permissive that the party-list restriction, meaning they define domains that are supersets of the party-list domain.

4.11. Lemma. *For profiles $\boldsymbol{A} =_{-i} \boldsymbol{A'}$ and a Thiele rule f such that $C \in f(\boldsymbol{A})$ and $C' \in f(\boldsymbol{A'})$, $C' \succ_i C$ implies $C' \succ'_i C$.*

Proof: Let f be a Thiele rule defined by weight vector $\boldsymbol{w}$, let $\boldsymbol{A}$ and $\boldsymbol{A'}$ be i-variants for some agent i. Suppose further that $C \in f(\boldsymbol{A})$, $C' \in f(\boldsymbol{A'})$, and $C' \succ_i C$. Our aim is to show that $C' \succ'_i C$.

As $C \in f(\boldsymbol{A})$, we know the $\boldsymbol{w}$-score of C in profile $\boldsymbol{A}$ must be at least as high as that of C'. We write this as follows, separating the utility of agent i from the utilities of agents in $N \setminus \{i\}$:

$$u_i^{\boldsymbol{A}}(C) \; + \sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A}}(C) \; \geq \; u_i^{\boldsymbol{A}}(C') \; + \sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A}}(C')$$

By assumption, we know that $u_i^{\boldsymbol{A}}(C) < u_i^{\boldsymbol{A}}(C')$, so in order for the $\boldsymbol{w}$-score of C to be at least as high as that of C', agents in $N \setminus \{i\}$ must, in a sense, collectively prefer C to C':

$$\sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A}}(C) \;\; > \sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A}}(C')$$

Because $\boldsymbol{A} =_{-i} \boldsymbol{A'}$, the utility of an agent in $N \setminus \{i\}$ for any committee remains the same relative to $\boldsymbol{A}$ and $\boldsymbol{A'}$—in other words, $u_j^{\boldsymbol{A}}(C) = u_j^{\boldsymbol{A'}}(C)$ and $u_j^{\boldsymbol{A}}(C') = u_j^{\boldsymbol{A'}}(C')$ for all $j \in N \setminus \{i\}$. Thus, we have that:

$$\sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A'}}(C) \;\; > \sum_{j \in N \setminus \{i\}} u_j^{\boldsymbol{A'}}(C') \tag{i}$$

Finally, $C' \in f(\boldsymbol{A'})$ implies that $\boldsymbol{w}$-score of C' in profile $\boldsymbol{A'}$ is at least as high as that of C. This, together with Equation (i) implies $u_i^{\boldsymbol{A'}}(C') > u_i^{\boldsymbol{A'}}(C)$—or $C' \succ'_i C$—as desired. □

We now prove a slightly more technical lemma that we will utilise multiple times throughout this chapter. Lemma 4.12 establishes the existence of a particular committee C^* in the initial outcome, whenever an agent attempts free-riding. We construct this committee C^* by replacing any candidates from a winning committee C that are in A_i but not A'_i with candidates from A'_i that are in C. We do this until either C^* contains enough candidates, or until all candidates from A'_i are in C^*.

4.12. Lemma. *Let f be a Thiele rule. For a party-list profile $\boldsymbol{A}$, agent i, and an i-variant $\boldsymbol{A'} =_{-i} \boldsymbol{A}$ where $A'_i \subset A_i$, it is the case that $C \in f(\boldsymbol{A})$ implies that there is some k-committee $C^* \in f(\boldsymbol{A})$ such that*

- *$C^* \sim_i C$, and*
- *for $C' \in f(\boldsymbol{A'})$ we have that $C' \succ_i C$ implies $(A_i \cap C^*) \subseteq A'_i$.*

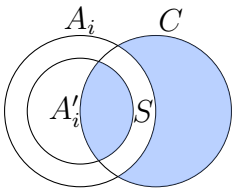


Figure 4.1: Approval sets A_i and A'_i in relation to committee C. Blue area is $C_{\texttt{start}}$. To construct C^*, we add candidates from A'_i until the set reaches desired size or we run out of candidates.

Proof: Let $S = (A_i \cap C) \setminus A'_i$. This is the set of all candidates that A_i and C agree on, except those also approved by A'_i. We will be building up C^* by starting with $C_{\texttt{start}} = C \setminus S$ and adding candidates until we reach a committee of size k. See Figure 4.1 for a visual representation of these sets and how they relate to each other. We add candidates to $C_{\texttt{start}}$ as follows:

- If $|A'_i \setminus C_{\texttt{start}}| \geq |S|$—meaning if there are enough candidates in A'_i to fill the $|S|$ "open spots"—we add candidates from $A'_i \setminus C_{\texttt{start}}$ until we reach a committee C^* such that $|C^*| = k$.
- Otherwise, we add all candidates in A'_i to the committee. We then fill the remaining "open slots" with candidates from S until we reach a committee C^* of size k.

Because $\boldsymbol{A}$ is a party-list profile, it is clear from the construction of C^* that $u_j^{\boldsymbol{A}}(C^*) = u_j^{\boldsymbol{A}}(C)$ for all $j \in N$—as C and C^* only differ on alternatives in A_i, and the two committees are of equal size. So we already know that $C \sim_i C'$. As $C \in f(\boldsymbol{A})$, it must therefore also be the case that $C^* \in f(\boldsymbol{A})$.

In order to prove the second part of the statement, suppose $C' \in f(\boldsymbol{A'})$ and $C' \succ_i C$. Suppose further $|A'_i \setminus C_{\texttt{start}}| \not\geq |S|$. This means we have exhausted all candidates in A'_i when building C^*, and so $A'_i \subseteq C^*$. Because A'_i is contained in C^*, we know that $C^* \succeq'_i C'$. However as $C^* \in f(\boldsymbol{A})$ and $C' \succ_i C^*$, Lemma 4.11 tells us that $C' \succ'_i C^*$, which is, of course, a contradiction.

As it cannot be the case that $C' \succ_i C^*$ and $|A'_i \setminus C_{\texttt{start}}| \not\geq |S|$, it remains only to show that $|A'_i \setminus C_{\texttt{start}}| \geq |S|$ implies $(A \cap C^*) \subseteq A'$. If $|A'_i \setminus C_{\text{start}}| \geq |S|$, then we know that all candidates added to C_{start} to create C^* must come from A'_i. In other words, we know that $C^* \setminus C_{\texttt{start}} \subseteq A'_i$—so $A_i \cap (C^* \setminus C_{\texttt{start}}) \subseteq A'_i$. Additionally, the candidates that remain in $C_{\texttt{start}} \cap A_i$ are only those that are also in A'_i—so $(C_{\texttt{start}} \cap A_i) \subseteq A'_i$. Putting this together, we can see that $(A_i \cap C^*) \subseteq A'_i$ Thus, we have shown that $C' \in f(\boldsymbol{A'})$ and $C' \succ_i C$ implies $(A_i \cap C^*) \subseteq A'_i$. □

We can now, with the help of these two lemmas, prove a result pertaining to free-riding. Broadly, Proposition 4.13 establishes two things. If free-riding brings about a "more preferred" committee in the manipulated outcome, then i) that committee will already have been in the initial outcome, and ii) this committee will be accompanied by a "less preferred" committee in the manipulated outcome. We will take advantage of both these facts, separately, in results that build on Proposition 4.13.

4.13. Proposition. *Let f be a Thiele rule. Given an agent $i \in N$, profiles $\boldsymbol{A} =_{-i} \boldsymbol{A'}$—where $\boldsymbol{A}$ is a party-list profile, and approval sets $A'_i \subset A_i$: if $C' \succ_i C$ for committees $C' \in f(\boldsymbol{A'})$ and $C \in f(\boldsymbol{A})$, then there exists a committee C^* such that $C \sim_i C^*$, and $\{C^*, C'\} \subseteq f(\boldsymbol{A}) \cap f(\boldsymbol{A'})$.*

Proof: Let f be a Thiele rule. Suppose we have two profiles $\boldsymbol{A}$ and $\boldsymbol{A'}$—where $\boldsymbol{A}$ is a party-list profile—and an agent i such that $\boldsymbol{A} =_{-i} \boldsymbol{A}$, and $A'_i \subset A_i$. Suppose further that we have committees $C' \in f(\boldsymbol{A'})$ and $C \in f(\boldsymbol{A})$, such that $C' \succ_i C$. As $\boldsymbol{A}$ is a party-list profile, Lemma 4.12 tells us there must be some $C^* \in f(\boldsymbol{A})$ such that $C^* \sim_i C$ and $(A_i \cap C^*) \subseteq A'_i$.

We first show that $C^* \in f(\boldsymbol{A'})$. Note that $(A_i \cap C^*) \subseteq A'_i$ and $A'_i \subset A_i$ implies that $|A_i \cap C^*| = |A'_i \cap C^*|$. We also know that $A_j = A'_j$ for all agents $j \neq i$. So we conclude that $|A_j \cap C^*| = |A'_j \cap C^*|$ for all $j \in N$. We can express this in terms of agents' utilities.

$$u_N^{\boldsymbol{A}}(C^*) = u_N^{\boldsymbol{A'}}(C^*) \tag{ii}$$

Because $C^* \in f(\boldsymbol{A})$, it must hold that $u_N^{\boldsymbol{A}}(C^*) \geq u_N^{\boldsymbol{A}}(C')$. This together with Equation (ii) implies:

$$u_N^{\boldsymbol{A'}}(C^*) \geq u_N^{\boldsymbol{A}}(C')$$

Additionally, as $A'_i \subset A_i$, we know that $|A_i \cap C'| \geq |A'_i \cap C'|$. Since all other agents submit the same approval set in both profiles, we have $u_N^{\boldsymbol{A}}(C') \geq u_N^{\boldsymbol{A'}}(C')$, which means:

$$u_N^{\boldsymbol{A'}}(C^*) \geq u_N^{\boldsymbol{A'}}(C')$$

As f is a Thiele rule, and $C' \in f(\boldsymbol{A'})$, this implies that $C^* \in f(\boldsymbol{A'})$.

We show that $C' \in f(\boldsymbol{A})$ in a similar manner. First, since $C' \in f(\boldsymbol{A'})$, by definition of f we know that $u_N^{\boldsymbol{A'}}(C') \geq u_N^{\boldsymbol{A'}}(C^*)$. We use Equation (ii) again to conclude $u_N^{\boldsymbol{A'}}(C') \geq u_N^{\boldsymbol{A}}(C^*)$. Since $\boldsymbol{A} =_{-i} \boldsymbol{A'}$ and $A'_i \subset A_i$, we have $u_N^{\boldsymbol{A}}(C') \geq u_N^{\boldsymbol{A'}}(C')$, which implies $u_N^{\boldsymbol{A}}(C') \geq u_N^{\boldsymbol{A}}(C^*)$. As f is a Thiele rule and $C^* \in f(\boldsymbol{A})$, it must also be the case that $C' \in f(\boldsymbol{A})$.

So we have $\{C^*, C'\} \subseteq f(\boldsymbol{A}) \cap f(\boldsymbol{A'})$, as desired. □

While Proposition 4.13 does not explicitly reference any preferences over outcomes, we still see a hint of what's to come. We build on Proposition 4.13 in several of our strategyproofness results. Theorem 4.15 below is the first of these. We first state a corollary that follows from Proposition 4.13 alone.

4.14. Corollary. *On party-list profiles, Thiele methods are immune to free-riding by optimistic agents.*

We are now ready to state the main result of this section. Theorem 4.15 is a strategyproofness result that pertains to a specific class of preference extensions.

4.15. Theorem. *On party-list profiles, Thiele methods are immune to free-riding by agents with general Gärdenfors preferences.*

Proof: Let f be a Thiele rule defined by weight vector $\boldsymbol{w}$. Suppose we have a party-list profile $\boldsymbol{A}$, and profile $\boldsymbol{A'}$ such that $\boldsymbol{A} =_{-i} \boldsymbol{A'}$, and $A'_i \subset A_i$. Let $\mathring{\succeq}_i$ be the preference order $\succeq_i$ extended according to some general Gärdenfors extension. We want to show that $f(\boldsymbol{A'}) \not\mathring{\succ}_i f(\boldsymbol{A})$. To this end, suppose we have committees $C \in f(\boldsymbol{A})$ and $C' \in f(\boldsymbol{A'})$ such that $C' \succ_i C$. If no such committees exist, the desired result immediately follows. Proposition 4.13 then tells us that $C' \in f(\boldsymbol{A})$. Thus, if $f(\boldsymbol{A'}) \not\subset f(\boldsymbol{A})$, it cannot be the case that $f(\boldsymbol{A'}) \mathring{\succ}_i f(\boldsymbol{A})$ for any general Gärdenfors preference—as $f(\boldsymbol{A'}) \mathring{\succ}_i f(\boldsymbol{A})$ implies there is some $C' \in f(\boldsymbol{A'}) \setminus f(\boldsymbol{A})$ that is strictly preferred by agent i to some $C \in f(\boldsymbol{A})$.

So suppose $f(\boldsymbol{A'}) \subset f(\boldsymbol{A})$. From Lemma 4.12 we know there exists a committee $C^* \in f(\boldsymbol{A})$ such that $C^* \sim_i C$, and $(A_i \cap C^*) \subset A'_i$. Note that because $A'_i \subset A_i$ we also have that $|A'_i \cap C^*| = |A_i \cap C^*|$. To prove our claim, we need to consider two cases.

Case 1: Suppose $A_i \subseteq C'$—meaning C' is one of agent i's top choices. We show that this implies $C' \notin f(\boldsymbol{A'})$, reaching a contradiction. We know that $u_N^{\boldsymbol{A'}}(C^*) = u_N^{\boldsymbol{A}}(C^*)$ as $|A'_i \cap C^*| = |A_i \cap C^*|$ and $\boldsymbol{A} =_{-i} \boldsymbol{A'}$. As $\{C^*, C'\} \subseteq f(\boldsymbol{A})$, we also know that $u_N^{\boldsymbol{A}}(C^*) = u_N^{\boldsymbol{A}}(C')$, so we can conclude:

$$u_N^{\boldsymbol{A'}}(C^*) = u_N^{\boldsymbol{A}}(C') \tag{iii}$$

Finally, as $A_i \subseteq C'$ by assumption, we know that $|A_i \cap C'| = |A_i|$. As $A'_i \subset A_i$, this implies that $|A'_i \cap C'| < |A_i \cap C'|$, which, as $\boldsymbol{A}$ and $\boldsymbol{A'}$ are i-variants, implies that $u_N^{\boldsymbol{A'}}(C') < u_N^{\boldsymbol{A}}(C')$. Using Equation (iii), we can thus conclude that $u_N^{\boldsymbol{A'}}(C') < u_N^{\boldsymbol{A'}}(C^*)$, meaning $C' \notin f(\boldsymbol{A'})$. So we have reached a contradiction.

Case 2: Suppose instead that $A_i \not\subseteq C'$. We then need to consider two subcases.

- If for all $a \in A_i \setminus A'_i$ we have $a \in C'$, then this means that $A_i \setminus C' \subseteq A'_i$. Because $A'_i \subset A_i$, there must exist candidates $a \in A_i \cap C'$ such that $a \notin A'_i$, and $b \in A_i \setminus C'$ such that $b \in A'_i$. Let $\tilde{C} = C' \setminus \{a\} \cup \{b\}$. Consider that for all $j \neq i$, $\tilde{C} \sim'_i C'$—candidates a and b are either both accepted by j or both rejected. For agent i, clearly $\tilde{C} \succ'_i C'$, so we have that

  $$u_N^{\boldsymbol{A'}}(\tilde{C}) > u_N^{\boldsymbol{A'}}(C')$$

  This implies $C' \notin f(\boldsymbol{A'})$, contradiction our initial assumption.

- Suppose instead that there exists some $a \in A_i \setminus A'_i$ such that $a \notin C'$. Because Lemma 4.11 tells us that $|A'_i \cap C'| > |A'_i \cap C|$, we know that $A'_i \cap C'$ is nonempty. Thus there must also exist an alternative $b \in A'_i \cap C'$, and clearly $b \in A_i$. We construct a committee $\tilde{C} = C' \setminus \{b\} \cup \{a\}$. Because the $\boldsymbol{w}$-score of the two committees C' and $\tilde{C}$ are the same in $\boldsymbol{A}$, and we know that $C' \in f(\boldsymbol{A})$, it must also be the case that $\tilde{C} \in f(\boldsymbol{A})$. However, $\tilde{C} \notin f(\boldsymbol{A'})$, as $C' \sim'_j \tilde{C}$ for all $j \neq i$, and $C' \succ'_i \tilde{C}$, meaning C' has a strictly higher $\boldsymbol{w}$-score in $\boldsymbol{A'}$. So $\tilde{C} \in f(\boldsymbol{A}) \setminus f(\boldsymbol{A'})$. We know from Proposition 4.13 that there is some $C^* \in f(\boldsymbol{A'})$ s.t. $C^* \sim_i C$. As $\tilde{C} \sim_i C'$, we know that $\tilde{C} \succ_i C^*$. Therefore it cannot be the case that $f(\boldsymbol{A'}) \mathrel{\mathring{\succ}}_i f(\boldsymbol{A})$.

Thus we can conclude that for any general Gärdenfors preference we have that $f(\boldsymbol{A'}) \not\mathrel{\mathring{\succ}}_i f(\boldsymbol{A})$, as desired. □

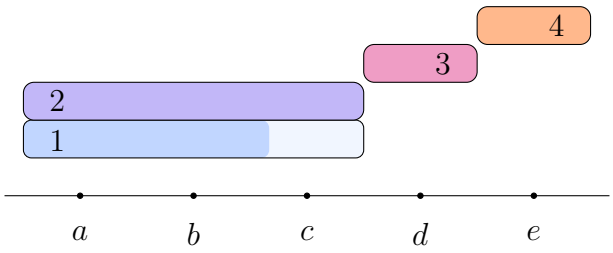
The following is an immediate consequence of Theorem 4.15, and covers three well-known preference extensions.

4.16. Corollary. *On party-list profiles, Thiele methods are immune to free-riding for the Gärdenfors, Fishburn, and Kelly preference extensions.*

Theorem 4.15 should be interpreted as a positive result. As we have emphasised, free-riding is a simple method of manipulation, and is a very natural way to vote strategically in approval elections. Excluding this type of strategising paints a hopeful picture. Additionally, our result holds for a large class of preferences. This class includes many natural extensions that have received much attention in the social choice literature. Of course, the theorem does not hold for *all* possible ways of extending preferences. We now give an example of a specific preference extension that is not captured by our definition of general Gärdenfors preferences, and show free-riding on party-list profiles becomes possible under this extension.

4.17. Example (Preferences not Covered by Theorem 4.15). Let $\boldsymbol{A}$ be the profile depicted below where agents 1 and 2 approve of the candidates a, b and c, while agent 3 approves of only d, and agent 4 approves of only e. Clearly, $\boldsymbol{A}$ is a party-list profile.

Let $k = 3$. We use proportional approval voting to demonstrate that a manipulation is possible in this profile for agent 1. The outcome under PAV comprises nine committees in total; six with two candidates from agent 1's approval set: $\{a, b, d\}, \{a, b, e\}, \{a, c, d\}, \{a, c, e\}, \{b, c, d\}, \{b, c, e\}$, and three with a single candidate each from agent 1's approval set: $\{a, d, e\}, \{b, d, e\}, \{c, d, e\}$. Suppose agent 1 prefers smaller sets to larger ones, provided that for any C in the larger set, there is some C' in the smaller set such that $C \sim_1 C'$.



Consider what happens when agent 1 drops c from their judgment set (represented in lighter blue). Because committees containing c will now have a lower $\boldsymbol{w}$-score, the new outcome will contain a total of four committees; two committees $\{a, b, d\}, \{a, b, e\}$ with two of agent 1's approved candidates, and two—$\{a, d, e\}, \{b, d, e\}$—each with a single candidate from A_1. We can see from agent 1's preferences that she would prefer the second (manipulated) outcome in this case. Thus, agent 1 has an incentive to free-ride in this profile. △

4.18. Remark. Example 4.17 also demonstrates that our results do not hold for resolute rules that break ties according to a linear order over committees. Suppose for example that the tie-breaking order $>$ is such that $\{c, d, e\} > \{a, b, d\}$, and $\{a, b, d\} > \{a, b, e\} > \{a, d, e\} > \{b, d, e\}$. Then the outcome in the first profile is $\{c, d, e\}$, while the outcome in the second is $\{a, b, d\}$—an improvement for agent 1.

Recall from Section 4.2.4 that we were not able to establish immunity to free-riding for other known restricted domains. We want to note here that our results do not simply hold on party-lists because it is a relatively strong restriction. Similar restrictions (such as those we defined in Section 4.2.4) do not have the same properties that enable strategyproofness. Among the natural restrictions discussed in the literature, the party-list domain is indeed the largest domain we can find where we are able to establish some barriers to manipulation for Thiele rules as a class.

4.4 Superset- and Disjoint-set-Manipulation

We will now show strategyproofness for two additional types of manipulation, superset-strategyproofness and disjoint-set-strategyproofness. This result holds for all strongly reflective preference extensions, including, of course, all general Gärdenfors preferences.

4.19. Theorem. *On party-list profiles, Thiele methods are immune to superset-manipulation and disjoint-set-manipulation for all strongly reflective preference extensions.*

Proof: Let $\boldsymbol{A}$ be a party-list profile, and f a Thiele rule defined by a weight vector $\boldsymbol{w}$. Suppose there is a profile $\boldsymbol{A'}$, and committees C and C' such that

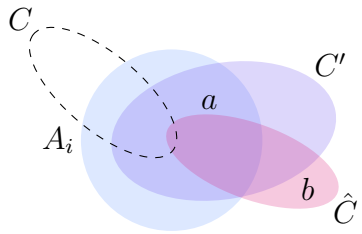


Figure 4.2: Representation of committees C, C' and $\hat{C}$, and candidates a and b used in proof of Theorem 4.19.

$\boldsymbol{A} =_{-i} \boldsymbol{A'}$, $C \in f(\boldsymbol{A})$, and $C' \in f(\boldsymbol{A'})$. Suppose also that either $A'_i \supset A_i$, or $A_i \cap A'_i = \emptyset$. We want to show that $C' \succ_i C$ implies $\{C, C'\} \subseteq f(\boldsymbol{A}) \cap f(\boldsymbol{A'})$. This is enough to establish superset and disjoint-set-strategyproofness for all strongly reflective preferences. With this goal in mind, suppose $C' \succ_i C$.

We first show that $C' \in f(\boldsymbol{A})$. Because $\boldsymbol{A}$ is a party-list profile and $C' \succ_i C$, there must be some $\hat{C} \in f(\boldsymbol{A})$ such that $\hat{C} \sim_i C$ and $(A_i \cap C') \supset (A_i \cap \hat{C})$, where C and $\hat{C}$ only differ on alternatives in A_i. To see why this is the case, note that the differences among C and $\hat{C}$ pertain only to candidates in A_i, and as such $\hat{C}$ will have the same $\boldsymbol{w}$-score as C in $\boldsymbol{A}$. We use the fact that $\hat{C} \in f(\boldsymbol{A})$ to show $C' \in f(\boldsymbol{A})$. For a visual representation of these committees and the candidates we will reference, see Figure 4.2.

Let $a \in A_i$ be an alternative such that $a \in C' \setminus \hat{C}$. Such an alternative must exist as $C' \succ_i \hat{C}$. Because $\boldsymbol{A}$ is a party-list profile, and $|C'| = |\hat{C}|$, we know there must be some party with (strictly) fewer representatives in C' than in $\hat{C}$. In other words, there exists some alternative $b \notin A_i$ such that $b \in \hat{C} \setminus C'$ and $|A_j \cap \hat{C}| > |A_j \cap C'|$ for all $j \in N_b^{\boldsymbol{A}}$. We define a k-committee $C_1 = \hat{C} \setminus \{b\} \cup \{a\}$.

Our immediate goal is to show $C_1 \in f(\boldsymbol{A})$. Note that the $\boldsymbol{w}$-score of C_1 in $\boldsymbol{A}$ cannot be higher than that of $\hat{C}$, as this would imply $\hat{C} \notin f(\boldsymbol{A})$. Because the two committees differ only with regard to alternatives a and b, we can express this as follows:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|} \geq \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|+1}$$

Similarly, the committee $C' \setminus \{a\} \cup \{b\}$ cannot have a higher $\boldsymbol{w}$-score than C' in $\boldsymbol{A'}$, so it must hold that:

$$\sum_{j \in N_a^{\boldsymbol{A'}}} w_{|A'_j \cap C'|} \geq \sum_{j \in N_b^{\boldsymbol{A'}}} w_{|A'_j \cap C'|+1}$$

We want to connect the two inequalities above. We know that $|A_j \cap C'| > |A_j \cap \hat{C}|$ for all $j \in N_a^{\boldsymbol{A}}$. For all $j \neq i$ this immediately tells us $|A'_j \cap C'| > |A_j \cap \hat{C}|$, as $\boldsymbol{A} =_{-i} \boldsymbol{A'}$. For agent i, we know either $A'_i \supset A_i$, or $A'_i \cap A_i = \emptyset$. If $a \in A'_i$,

then it must be that $A'_i \supseteq A_i$, meaning $|A'_i \cap \hat{C}| \geq |A_i \cap \hat{C}|$. From Lemma 4.11 we know that $C' \succ_i \hat{C}$, $\hat{C} \in f(\boldsymbol{A})$, and $C' \in f(\boldsymbol{A'})$ implies $|A'_i \cap C'| > |A'_i \cap \hat{C}|$, so we can conclude that $|A'_j \cap C'| > |A_j \cap \hat{C}|$ for all $j \in N_a^{\boldsymbol{A'}}$. Because $\boldsymbol{w}$ is a non-increasing weight vector, this implies $w_{|A'_j \cap C'|} \leq w_{|A_j \cap \hat{C}|+1}$ for all $j \in N_a^{\boldsymbol{A'}}$. Given that $N_a^{\boldsymbol{A}} \supseteq N_a^{\boldsymbol{A'}}$, we can conclude that:

$$\sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|+1} \geq \sum_{j \in N_a^{\boldsymbol{A'}}} w_{|A'_j \cap C'|}$$

We can now build the following chain of inequalities:

$$\begin{aligned} \sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|} &\geq \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|+1} \\ &\geq \sum_{j \in N_a^{\boldsymbol{A'}}} w_{|A'_j \cap C'|} \\ &\geq \sum_{j \in N_b^{\boldsymbol{A'}}} w_{|A'_j \cap C'|+1} \end{aligned} \tag{iv}$$

Recall that for all $j \in N_b^{\boldsymbol{A}}$, it is the case that $|A_j \cap \hat{C}| > |A_j \cap C'|$, which—as $A_j = A'_j$—is equivalent to $|A_j \cap \hat{C}| > |A'_j \cap C'|$. This implies $w_{|A_j \cap \hat{C}|} \leq w_{|A'_j \cap C'|+1}$. As $N_b^{\boldsymbol{A}} \subseteq N_b^{\boldsymbol{A'}}$, we then have that:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|} \leq \sum_{j \in N_b^{\boldsymbol{A'}}} w_{|A'_j \cap C'|+1} \tag{v}$$

Equations (iv) and (v) together imply that our chain of inequalities "collapses", meaning we get:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|} = \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap \hat{C}|+1}$$

This can only be the case if $C_1 \in f(\boldsymbol{A})$.

Finally, to see that $C_1 \in f(\boldsymbol{A})$ implies that $C' \in f(\boldsymbol{A})$, consider the following. We know that C_1 is created by adding one candidate from C' and removing one candidate that is not in C'. If $|\hat{C} \setminus C'| = 1$, then $C_1 = C'$ and we are done. Otherwise, note that $|C_1 \setminus C'| = |\hat{C} \setminus C'| - 1$, meaning C_1 is one candidate closer to C' than $\hat{C}$. Importantly, we also know the following:

(i) $C' \succ_i C_1$,

(ii) $(A_i \cap C') \supset (A_i \cap C_1)$, and

(iii) $C_1 \in f(\boldsymbol{A})$.

Thus, we can use the same argument we used to show that $C_1 \in f(\boldsymbol{A})$ to show there is some committee $C_2 \in f(\boldsymbol{A})$ such that $|C_2 \setminus C'| = |\hat{C} \setminus C'| - 2$. We can repeat this argument until we reach a committee $C_m \in f(\boldsymbol{A})$ where $m = |\hat{C} \setminus C'|$, meaning $C_m = C'$.

We now show that $C \in f(\boldsymbol{A'})$. We omit some details as the proof proceeds in a similar fashion as above. Note that as $C' \in f(\boldsymbol{A})$, and $\boldsymbol{A}$ is a party-list profile, we know there must be some $\hat{C}' \in f(\boldsymbol{A})$ such that $|A_j \cap C'| = |A_j \cap \hat{C}'|$ for all $j \in N$, and $A_i \cap \hat{C}' \supset A_i \cap C$, where C' and $\hat{C}'$ only differ on alternatives in A_i. This is because $C' \in f(\boldsymbol{A})$, and $\hat{C}'$ will have the same $\boldsymbol{w}$-score as C' in $\boldsymbol{A}$. Because C' and $\hat{C}'$ only differ on alternatives in A_i, and both are k-sized committees, it must be the case that $|A'_i \cap \hat{C}'| = |A'_i \cap C'|$. As $\boldsymbol{A} = \boldsymbol{A'}$, we know $|A'_j \cap \hat{C}'| = |A_j \cap C'|$ for all $j \neq i$ as well, so we can conclude that $\hat{C}' \in f(\boldsymbol{A'})$—as it would have the same $\boldsymbol{w}$-score as C' in $\boldsymbol{A'}$.

We repeat a similar argument as we did when showing $C' \in f(\boldsymbol{A})$. We have some $a' \in A_i$ such that $\hat{C}' \setminus C$, and some $b' \in C \setminus \hat{C}'$ such that for all $j \in N_b^{\boldsymbol{A}}$ we have $|A_j \cap C| > |A_j \cap \hat{C}'|$. Let $C'_1 = C' \setminus \{b\} \cup \{a\}$. Arguing in almost exactly the same way as above, we show that:

$$\sum_{j \in N_b^{\boldsymbol{A'}}} w_{|A'_j \cap C'|+1} = \sum_{j \in N_a^{\boldsymbol{A'}}} w_{|A'_j \cap C'|}$$

Thus $\hat{C}'_1 \in f(\boldsymbol{A'})$. We can again repeat this argument to show that $C \in f(\boldsymbol{A'})$.

So we have shown that for any $C' \in f(\boldsymbol{A'})$ and $C \in f(\boldsymbol{A})$, if $C' \succ_i C$, then $\{C, C'\} \subseteq f(\boldsymbol{A}) \cap f(\boldsymbol{A'})$, meaning it cannot be the case that $f(\boldsymbol{A'}) \succ_i^e f(\boldsymbol{A})$ for any strongly reflective extension e. □

Corollary 4.20 follows from Theorem 4.15 and Theorem 4.19.

4.20. Corollary. *On party-list profiles, Thiele methods are superset-strategyproof, disjoint-set-strategyproof, and immune to free-riding for general Gärdenfors preferences.*

Our results paint a positive picture for Thiele rules on party-list profiles as they rule out all three types of manipulation considered in this chapter for a large class of preferences. Importantly, we are also able to establish some level of strategyproofness for these rules that does not depend much on the choice of preference extension. The class of strongly reflective preferences is large, and arguably includes many extensions of interest.

4.5 Optimistic Agents

We obtain our strongest result for the optimistic preference extension, as we can establish full strategyproofness for Thiele methods on party-list profiles.

We will be working with profiles which are not party-list profiles, but are, informally speaking, one agent away from a party-list profile. We write $\boldsymbol{A}_{-i}$ to mean the profile $\boldsymbol{A}$ with the approval set of agent i removed. An agent i casts a *separable vote* in a profile $\boldsymbol{A}$ if for all agents $j \in N$ either $A_i \subseteq A_j$ or $A_i \cap A_j = \emptyset$.

We now show that optimistic agents have no incentive to superset-manipulate. We will use this to establish our full strategyproofness result for optimistic agents.

4.21. Lemma. *Let f be a Thiele rule, and $\boldsymbol{A}$ a profile such that $\boldsymbol{A}_{-i}$ is a party-list profile, and A_i is a separable vote in $\boldsymbol{A}$. Given an agent $i \in N$, and a profile $\boldsymbol{A'} =_{-i} \boldsymbol{A}$ such that $A'_i \supset A_i$: if $C' \succ_i C$ for all $C \in f(\boldsymbol{A})$, then $C' \notin f(\boldsymbol{A'})$.*

Proof: Let f be a Thiele rule defined by weight vector $\boldsymbol{w}$. Suppose we have $\boldsymbol{A}$ and $\boldsymbol{A'}$—where $\boldsymbol{A}_{-i}$ is a party-list profile, and A_i is a separable vote in $\boldsymbol{A}$—and an agent $i \in N$ such that $\boldsymbol{A} =_{-i} \boldsymbol{A}$, and $A'_i \supset A_i$. Let $C \in f(\boldsymbol{A})$ be among the most preferred committees for agent i in $f(\boldsymbol{A})$, and suppose there exists a k-committee C', such that $C' \succ_i C$. We want to show that $C' \notin f(\boldsymbol{A'})$.

We identify two candidates relevant for our purposes. Because $C' \succ_i C$, we know there must exist at least one candidate $a \in A_i \cap (C' \setminus C)$. We know that there is at least one party with strictly fewer representatives in C' than in C, as they are both committees of the same size. More formally, because $\boldsymbol{A}_{-i}$ is a party-list profile, there must also exist some candidate $b \in C \setminus (C' \cap A_i)$, such that for all $j \in N_b^{\boldsymbol{A}}$, it is the case that $|A_j \cap C| > |A_j \cap C'|$. If this were not the case, then $C' \succeq_j C$ for all $j \neq i$ and $C' \succ_i C$, meaning $C \notin f(\boldsymbol{A})$, contradicting our initial assumption.

First, we want to show that $A_j \cap C = A_i \cap C$ for all $j \in N_a^{\boldsymbol{A}}$. We know that $A_i \subseteq A_j$ for all $j \in N_a^{\boldsymbol{A}}$ (and $A_j = A_{j'}$ for $j, j' \in N_a^{\boldsymbol{A}} \setminus \{i\}$). Recall that $a \notin C$. Suppose $(A_j \setminus A_i) \cap C \neq \emptyset$, meaning there is some alternative $x \in A_j \setminus A_i$ such that $x \in C$. Then the committee $C \setminus \{x\} \cup \{a\}$ will have a $\boldsymbol{w}$-score higher than C in the profile $\boldsymbol{A}$. As this implies $C \notin f(\boldsymbol{A})$, we can conclude that $(A_j \setminus A_i) \cap C = \emptyset$. In other words, we have that $A_j \cap C = A_i \cap C$ for all $j \in N_a^{\boldsymbol{A}}$. As $A_i \subseteq A_j$ for all $j \in N_a^{\boldsymbol{A}}$, we know $|A_j \cap C'| \geq |A_i \cap C'|$. As $|A_i \cap C'| > |A_i \cap C|$ by assumption, this implies that $|A_j \cap C'| > |A_j \cap C|$. As $\boldsymbol{w}$ is a non-increasing weight vector, we this implies that $w_{|A_j \cap C'|} \leq w_{|A_j \cap C|+1}$ for all $j \in N_a^{\boldsymbol{A}}$. Similarly we know $|A_j \cap C| > |A_j \cap C'|$ for all $j \in N_b^{\boldsymbol{A}}$. Thus we have:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap C'|+1} \geq \sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap C|}$$

$$\sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap C|+1} \geq \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap C'|}$$

We also claim the following:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap C|} > \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap C|+1}$$

To see why this is the case, note that if it did not hold, then the committee $(C \setminus \{b\}) \cup \{a\}$ would have a $\boldsymbol{w}$-score at least as high as C, and thus would be among the winning committees in $f(\boldsymbol{A})$. This would clearly contradict our assumption that $C \in f(\boldsymbol{A})$ is one of the most preferred outcomes for agent i in $f(\boldsymbol{A})$, as $(C \setminus \{b\}) \cup \{a\} \succ_i C$. Putting together the above, we conclude that:

$$\sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap C'|+1} > \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap C'|} \tag{vi}$$

We can now show $C' \notin f(\boldsymbol{A'})$. Let $\tilde{C} = (C' \setminus \{a\}) \cup \{b\}$ be a k-committee. We calculate the $\boldsymbol{w}$-score of $\tilde{C}$ in $\boldsymbol{A'}$.

$$u_N^{\boldsymbol{A'}}(\tilde{C}) = u_N^{\boldsymbol{A'}}(C') - \sum_{j \in N_a^{\boldsymbol{A}}} w_{|A_j \cap C'|} + \sum_{j \in N_b^{\boldsymbol{A}}} w_{|A_j \cap C'|+1}$$

Taken together with Equation (vi), the above implies that $\tilde{C}$ has a $\boldsymbol{w}$-score strictly higher than C' in $\boldsymbol{A'}$, meaning $C' \notin f(\boldsymbol{A'})$. □

We use Proposition 4.13, which speaks only about free-riding, and Lemma 4.21, which pertains to superset-manipulation, to prove the following Theorem for optimistic agents.

4.22. Theorem. *Thiele methods are party-list-strategyproof for optimistic agents.*

Proof: Let $\boldsymbol{A}$ be a party-list profile, and let $\boldsymbol{A'}$ be a profile such that $\boldsymbol{A} =_{-i} \boldsymbol{A'}$. Suppose there exists some C' such that $C' \succ_i C$ for any $C \in f(\boldsymbol{A})$. Let f be a Thiele rule. We show that $C' \notin f(\boldsymbol{A'})$.

Suppose for contradiction that $C' \in f(\boldsymbol{A'})$. We construct a third, intermediate, profile $\boldsymbol{A^*}$ where $\boldsymbol{A^*} =_{-i} \boldsymbol{A}$ and $A_i^* = A_i \cap A_i'$. Note that $\boldsymbol{A^*}_{-i}$ is a party-list profile, and A_i^* is a separable vote in $\boldsymbol{A^*}$. We assume $A_i^* \neq A_i$—if this were not the case, then $f(\boldsymbol{A}) = f(\boldsymbol{A^*})$, and Lemma 4.21 alone would be enough to establish our claim.

As $C' \in f(\boldsymbol{A'})$ and $A_i' \supset A_i^*$, we know from Lemma 4.21 that there exists some $C^* \in f(\boldsymbol{A^*})$ such that $C^* \succeq_i C'$. If this were not the case, then C' would be strictly preferred to all committees in $f(\boldsymbol{A^*})$, and so $C' \notin f(\boldsymbol{A'})$, which would be a contradiction. As $C^* \succeq_i C'$, we know that $C^* \succ_i C$. Because $A_i^* \subset A_i$, we can use Proposition 4.13 to show that $C^* \in f(\boldsymbol{A^*})$ implies $C^* \in f(\boldsymbol{A})$. This contradicts our assumption that $C' \succ_i C$ for all $C \in f(\boldsymbol{A})$ as $C^* \succeq_i C'$. □

Note that Theorem 4.22 makes no assumptions about how agents may manipulate, as any possible manipulation amounts to an agent first removing some candidates from their approval set (possibly none), and subsequently adding new candidates (again, possibly none). As the optimistic preference extension is a very natural and intuitive extension, Theorem 4.22 is a very welcome result.

4.6 Summary

In this chapter, we have studied strategyproofness of Thiele methods on party-list profiles. In particular, we focused on three types of manipulation: free-riding, superset-manipulation, and disjoint-set-manipulation. We have shown for general Gärdenfors preferences that it is not possible to manipulate Thiele rules in any of the three manners we considered—subset-manipulation, superset-manipulation, and disjoint-set-manipulation—on party-list profiles. For superset and disjoint-set-manipulation, this holds for all strongly reflective preference extensions. We have also shown that Thiele methods are fully strategyproof on party-list profiles for optimistic agents.

Strategyproofness on party-list profiles spells good news for many applications of multiwinner voting. Our results also suggest that focusing on specific domains or profiles may be a fruitful avenue of study for establishing further strategyproofness results. Our particular focus here has been on approval-based rules. We are hopeful that similar methods may also yield strategyproofness results for multiwinner rules that aggregate preference rankings.

Chapter 5

Majoritarianism and Strategyproofness in Judgment Aggregation

Our playground in this chapter is the world of judgment aggregation—a rich framework for analysing all kinds of multiagent decision making scenarios (List and Pettit, 2002; Grossi and Pigozzi, 2014). Judgment aggregation (JA) is an area in the broader field of social choice theory concerned with aggregating individual opinions on a set of possibly interconnected issues. In judgment aggregation we model the views held by individual agents as sets of propositional formulas. We then design rules for aggregating such judgments into a collective judgment that adequately represents the views held by the group. This framework generalises preference aggregation as traditionally studied in social choice theory (Dietrich and List, 2007a) and is closely related to the field of belief merging as long studied in AI (Everaere et al., 2017). We revisit a classical example from legal theory due to Kornhauser and Sager (1993), both to demonstrate the framework in action and to show a classic problem that has arguably played a large part in inspiring the study of JA as a framework. This example is dubbed the *doctrinal paradox* as Kornhauser and Sager were concerned specifically with aggregation of opinions related to legal doctrine. A more general version of this example, which abstracts away from any specific context or application, is known as the *discursive dilemma* (List and Pettit, 2002).

5.1. Example (The Doctrinal Paradox). We are back in the kingdom of Arrovia. A defendant is facing prosecution and their fate will be determined by three judges. The job of the judges is to determine whether the defendant is liable for a breach of contract. The Arrovian legal doctrine is clear: a defendant is liable if and only if there has been a breach of the language in the contract, and the contract itself is legally valid. Determining the validity of the contract

and whether it has been breached is ultimately up to the three judges. Each judge has their opinion on the issues at hand—represented in the table below. Judge 1 for example, believes the contract is valid and that it has been breached. Consequently she finds the defendant is indeed liable. Given the opinions of the three judges, how should they determine the final judgment?

	Valid	Breached	Liable	(Valid + Breached) $\leftrightarrow$ Liable
Judge 1	Yes	Yes	Yes	Yes
Judge 2	Yes	No	No	Yes
Judge 3	No	Yes	No	Yes

One option is to ask each judge whether they think the contract is valid, and whether it was breached. Since a majority of the judges do believe that a valid contract has been breached, we might want to conclude that the defendant is liable. However, if we ask each judge directly about the defendant's liability, a majority of judges will say the defendant is not liable. This discrepancy—or "paradox"—occurs because looking at the majority on each issue may give us an inconsistent collective opinion. Indeed, as a group, the three judges believe (i) that the contract is valid, (ii) that it was breached, and (iii) that the defendant is not liable—an outcome that is not consistent with the legal doctrine they all support. △

The fact that taking the *propositionwise majority* can result in an inconsistent collective opinion is at the heart of much of the JA literature. This problem has implications for topics ranging from the computational complexity of determining the outcome of aggregation to the strategyproofness of aggregation rules. Our focus in this chapter is on this last topic: the susceptibility of judgment aggregation rules to strategic manipulation by self-interested agents. It should come as no surprise to you by now, dear reader, that finding strategyproof aggregation rules in JA is not a simple goal. As is the case in many areas of social choice, there is a tradeoff in judgment aggregation between strategyproofness and certain requirements on the output of the aggregation. By a combination of well-known results, we know that it is essentially impossible to design an aggregation rule that simultaneously guarantees an outcome that is *logically consistent* and *immune to strategic manipulation* (Dietrich and List, 2007c; Dokow and Holzman, 2010; List and Pettit, 2002; Nehring and Puppe, 2007). Let us revisit Example 5.1 to demonstrate what exactly we mean by strategic manipulation in judgment aggregation.

5.2. Example (Strategic Manipulation). Suppose the three judges from Example 5.1 decide to use the *premise-based* aggregation rule. That is, they will check the majority opinion on whether the contract is valid, and whether it has been breached. They will then use this to determine the liability of the defendant.

Judge 3 however, knows his colleagues well, and can guess what their opinions will be. He does a quick calculation and realises that if he submits his truthful opinion, they will be convicting who he believes to be an innocent person. Judge 3 therefore claims to his colleagues that he believes the contract has not been breached. His quick thinking pays off, the judges now collectively believe the contract is valid, but not that it has been breached. The defendant goes free thanks to judge 3! △

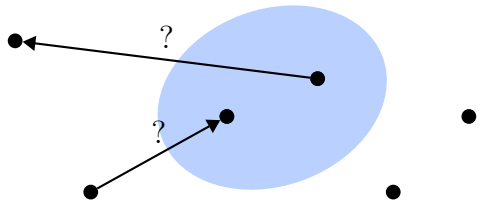
In this chapter we put forward a novel notion of strategyproofness, which requires immunity to strategic manipulation only in certain well-defined situations—namely when either the truthful profile of individual judgments *or* the profile a would-be manipulator is trying to reach are *majority-consistent*—meaning the outcome of the propositionwise majority is consistent.[1] We argue that this type of strategyproofness offers a reasonable compromise for aggregation rules one may want to use in practice. In this chapter we study the most commonly encountered majoritarian rules in judgment aggregation and prove that several important rules—including the Kemeny rule and the Slater rule—are strategyproof in this sense. This new notion of *domain-strategyproofness* is particularly useful when trying to improve upon aggregation rules that are known to be (fully) strategyproof but that can guarantee consistent outcomes only on a restricted domain (as is the case for the majority rule). In such a case, a rule that is guaranteed to always return consistent outcomes and that is strategyproof relative to that same domain is an attractive alternative. Indeed, if a rule is strategyproof for a restricted domain then this tells us two things. First, if the truthful profile is in the domain, then no agent has an incentive to manipulate. Second, if the profile that results after all judgments have been submitted is in the domain, then we can be certain that the profile reported cannot have been the result of strategic manipulation. Note that, as in previous chapters, while we work with a restricted domains, agents may attempt to manipulate to, or from, any profile both within and outside the domain.

The axiom—and associated domain—we look at in this chapter is *majoritarianism*. Rules that are majoritarian agree with the propositionwise majority whenever possible. While the outcome of the propositionwise majority may sometimes be inconsistent, its use as a benchmark in JA cannot be disputed—if the majority opinion is consistent, we usually do not want to go against it. Many domain restrictions in JA fall within the domain of profiles with a consistent majority. Thus, these are natural scenarios we might encounter in application and strategyproofness on these profiles is a positive result. Our results are again an argument against restricting the domain of the aggregation rule.

Recall our safe "island" from Chapter 1. Whereas we until now have only examined the question of whether an agent can manipulate *from* a profile in

[1] Note that this differs from our approach in Chapters 3 and 4—here we also examine whether it is possible to manipulate *to* the domain in question.

the "safe" domain, we will now also look at whether an agent can manipulate *to* a profile in the domain of interest (represented below with the second arrow pointing in to the domain).



5.3. Example. It is getting close to the winter holidays in Arrovia. The king is running an opinion poll to determine what the citizens want at the winter fair. The poll is quite complex and many of the possible activities and potential food stalls are interconnected. For example, it is not possible to have both a gingerbread stall and a cinnamon cookie stall as the baker can only run one. The king knows this is a problem he should solve with judgment aggregation. Knowing that aggregating using the propositionwise majority will often result in inconsistent collective opinions, he decides to use an aggregation rule that is majoritarian *and* strategyproof relative to the domain of majority-consistent profiles. When the results of the poll come back, he is delighted to see that the result is consistent. The king can announce the winter fair plans with a clear conscience, knowing that no manipulation could have occurred. △

This chapter is an examination of strategyproofness of majoritarian rules relative to profiles where the propositionwise majority returns a consistent outcome. In Sections 5.1 and 5.2 we go over the judgment aggregation framework and define the aggregation rules we will study. Section 5.3 is devoted to the topic of strategyproofness. Here we go over relevant work from the literature as well as further motivate our choice to study a restricted domain by presenting a negative strategyproofness result for profiles outside our well-behaved domain. In Section 5.4 we present the main positive strategyproofness result of this chapter that pertain to the class of rules knows as *additive majority rules*. In Section 5.5 we study coarsenings of additive majority rules. While the strategyproofness results we present here are not as strong as those of the preceding section, we see that these rules still manage to put up some barriers against manipulation. Finally, we examine the Dodgson rule in Section 5.6, which turns out to be the most manipulable rule we study.

5.1 Preliminaries

As always, let $N = \{1, \dots, n\}$ be our finite set of *agents*. We will assume that n is odd to avoid having to make tie-breaking decisions when computing the

outcome of the propositionwise majority. We have a (nonempty) set of formulas of propositional logic $\Phi = \Phi^+ \cup \Phi^-$, called the *agenda*, where Φ^+ is a set of nonnegated formulas, and $\Phi^- = \{\neg\varphi \mid \varphi \in \Phi^+\}$.

A *judgment* J is a subset of Φ. We use $\mathcal{J}(\Phi) \subseteq 2^\Phi$ to denote the set of all judgments that are (logically) *consistent* as well as *complete*—meaning they include one of φ and $\neg\varphi$ for every $\varphi \in \Phi^+$. Observe that any consistent judgment will also be *complement-free*, meaning that it cannot include both φ and $\neg\varphi$ for any $\varphi \in \Phi^+$. Any element of $\mathcal{J}(\Phi)$ is a permissible judgment J_i for an agent $i \in N$. This amounts to requiring agents are rational in that they always submit a complete and consistent judgment. We write $J =_\varphi J'$ to mean that judgments J and J' agree on formula φ.

The *Hamming distance* between two judgments J and J' in $\mathcal{J}(\Phi)$ is defined as $H(J, J') := |J \setminus J'| = |J' \setminus J|$. Thus, $H(J, J')$ is the number of elements in Φ^+ on which J and J' disagree. We say that judgment J' is *between* J and J'', if $J \cap J'' \subseteq J' \subseteq J \cup J''$. For example, if $J = \{\varphi, \psi\}$, $J' = \{\neg\varphi, \psi\}$ and $J'' = \{\varphi, \neg\psi\}$, then J is between J' and J'' but J' is not between J and J''. Observe that $J \cap J'' \subseteq J'$ if and only if $J' \subseteq J \cup J''$ in case all three judgments are both complete and complement-free.

A *profile* $\boldsymbol{J} = (J_1, \ldots, J_n) \in \mathcal{J}(\Phi)^n$ is a vector of individual judgments, one for each agent in N. For any such profile $\boldsymbol{J}$ and any $\varphi \in \Phi$, the set $N_\varphi^{\boldsymbol{J}} := \{i \in N \mid \varphi \in J_i\}$ is the set of *supporters* of proposition φ, with $n_\varphi^{\boldsymbol{J}} := |N_\varphi^{\boldsymbol{J}}|$. The *majority judgment* associated with a given profile $\boldsymbol{J}$ is defined as $m(\boldsymbol{J}) := \{\varphi \in \Phi \mid n_\varphi^{\boldsymbol{J}} > \frac{n}{2}\}$. We will sometimes refer to $m(\boldsymbol{J})$ as the outcome of the (propositionwise) majority. We say that profiles $\boldsymbol{J}$ and $\boldsymbol{J'}$ are *i-variants*, and we write $\boldsymbol{J} =_{-i} \boldsymbol{J'}$, if $J_j = J'_j$ for all agents $j \neq i$ (and possibly $J_i \neq J'_i$ for agent i).

Let $\mathcal{M}(\Phi, n) \subseteq \mathcal{J}(\Phi)^n$ be the domain of all profiles for a given agenda and a given number of agents for which the majority outcome is consistent: $\mathcal{M}(\Phi, n) := \{\boldsymbol{J} \mid m(\boldsymbol{J}) \not\models \bot\}$. If Φ and n are clear from context, we simply write $\mathcal{M}$.

5.2 Judgment Aggregation Rules

Intuitively, a *judgment aggregation rule* is a function that maps any given profile to the collective judgment of the group. We restrict our attention to aggregation rules that, for any given profile of complete and consistent judgments, will only return collective judgments that are complete and consistent. As we saw in the introduction, the majority rule—which returns $m(\boldsymbol{J})$ for any given profile $\boldsymbol{J}$—does *not* meet this requirement. While we have been speaking of *a* collective judgment, in practice most natural rules are *irresolute*—meaning that they allow for the possibility of ties between several collective judgments and thus require a tie-breaking mechanism to settle on a single outcome. This remains the case even for an odd number of agents. With all this in mind, formally, an aggregation rule is a function

$$f : \mathcal{J}(\Phi)^n \to 2^{\mathcal{J}(\Phi)} \setminus \{\emptyset\}.$$

So f takes a profile and returns a (nonempty) set of judgments. Note, again, that we require that f returns a complete and consistent outcome in order to qualify as an aggregation rule.

5.2.1 Majority-Preserving Rules

Our focus is on *majoritarian* (or majority-preserving) rules. A rule f is majoritarian if $f(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$ for all profiles $\boldsymbol{J}$ such that $m(\boldsymbol{J})$ is consistent. Thus, such a rule returns the outcome of the propositionwise majority when it is consistent, and does something else when it is not. Majoritarian rules constitute the bulk of well-studied rules in judgment aggregation (Lang et al., 2017), and are a natural starting point for studying domain-specific strategyproofness—by definition there is a subdomain of profiles where they are known to be "well-behaved". Importantly, the majority is in many cases the ideal, "democratic", outcome. Thus, its use as a benchmark is not just for technical reasons, but also normative ones.

We now define the majoritarian rules we study in this chapter. Often aggregation rules are categorised based on how much information they require to determine the outcome. Specifically, we distinguish between rules that are based on the majoritarian set—similar to tournament solutions in voting—and rules based on the weighted majoritarian set. We will also see an example of an aggregation rule that requires the actual judgments given in the profile in order to determine the winning judgments—the so-called *Dodgson rule*.

Rules based on the majoritarian set only need the outcome of the propositionwise majority in order to determine the collective opinion. Meaning they need only the information given by $m(\boldsymbol{J})$. So $f(\boldsymbol{J}) = f(\boldsymbol{J'})$ if $m(\boldsymbol{J}) = m(\boldsymbol{J'})$. Within this class, one rule reigns supreme over the rest—the *Slater* rule f_{Sla}.

$$f_{\text{Sla}}(\boldsymbol{J}) = \operatorname*{argmin}_{J \in \mathcal{J}(\Phi)} H(J, m(\boldsymbol{J}))$$

Slater returns those consistent judgments that maximise the number of propositions on which they agree with a majority of the agents, without differentiating between majorities of different strengths. Clearly, f_{Sla} generalises the Slater rule familiar from preference aggregation (Slater, 1961). It also known under several other names, such as the *endpoint rule* (Miller and Osherson, 2009) and *maxcard Condorcet rule* (Lang et al., 2017).

A second prominent rule based on the majoritarian set is the *maximal Condorcet rule*, f_{Con}—also known as the Condorcet admissible set (Nehring et al., 2014). For a set of formulas $S \subseteq \Phi$, a set $S' \subseteq S$ is a *maximally consistent* subset of S if and only if (i) S' is consistent and (ii) there is no consistent set S'' such that $S' \subset S'' \subseteq S$. Let $C(J)$ denote the set of all maximally consistent subsets

of the judgment J, and let $S^+ = \{J \in \mathcal{J}(\Phi) \mid J \supseteq S\}$. The maximal Condorcet rule is defined as follows:

$$f_{\text{Con}}(\boldsymbol{J}) = \{J^+ \mid J \in C(m(\boldsymbol{J}))\}$$

We give an example to illustrate these two aggregation rules.

5.4. Example. Let $\boldsymbol{J}$ be the five-agent profile below.

	$p \wedge r$	$p \wedge s$	q	$p \wedge q$
1 agent	Yes	Yes	Yes	Yes
2 agents	No	No	Yes	No
2 agents	Yes	Yes	No	No
$m(\boldsymbol{J})$	Yes	Yes	Yes	No

It is easy to see that $m(\boldsymbol{J})$ can be made consistent by negating only one proposition in the set—for example $(p \wedge q)$—meaning any judgment returned by the Slater rule must 'flip' only one proposition. Thus $f_{\text{Sla}}(\boldsymbol{J}) = \{\{(p\wedge r), (p\wedge s), q, (p\wedge q)\}, \{(p \wedge r), (p \wedge s), \neg q, \neg(p \wedge q)\}\}$. For the maximal Condorcet rule, things are a bit more complicated. We have a majority judgment $m(\boldsymbol{J}) = \{(p \wedge r), (p \wedge s), q, \neg(p \wedge q)\}$, so the set of maximally consistent subsets of the majority is $C(m(\boldsymbol{J})) = \{\{(p\wedge r), (p\wedge s), \neg(p\wedge q)\}, \{(p\wedge r), (p\wedge s), q\}, \{q, \neg(p\wedge q)\}\}$. Given this, we can calculate the outcome of the maximal Condorcet rule, $f_{\text{Con}}(\boldsymbol{J}) = \{\{(p \wedge r), (p\wedge s), q, (p\wedge q)\}, \{(p\wedge r), (p\wedge s), \neg q, \neg(p\wedge q)\}, \{\neg(p\wedge r), \neg(p\wedge s), q, \neg(p\wedge q)\}\}$. We can now easily see that $f_{\text{Sla}}(\boldsymbol{J}) \neq f_{\text{Con}}(\boldsymbol{J})$. △

Rules based on the weighted majoritarian set look at the size of the majorities in order to determine the winning judgments. The most prominent such rule is the *Kemeny* rule f_{Kem}.

$$f_{\text{Kem}}(\boldsymbol{J}) = \operatorname*{argmin}_{J \in \mathcal{J}(\Phi)} \sum_{i \in N} H(J, J_i)$$

We can think of the Kemeny rule as returning those judgments that minimise the average Hamming distance to the judgments in the profile. This rule generalises the well-known Kemeny rule for preference aggregation (Kemeny, 1959) and is also known under a number of other names, notably the *distance-based rule* (Pigozzi, 2006), *median rule* (Nehring et al., 2014), and *prototype rule* (Miller and Osherson, 2009).

5.5. Example. Consider again the five-agent profile $\boldsymbol{J}$ from Example 5.4. Note that we have three out of five agents accepting the first three formulas, while four out of the five reject $p\wedge q$. Thus, we know that $\{(p\wedge r), (p\wedge s), q, (p\wedge q)\} \notin f_{\text{Kem}}(\boldsymbol{J})$ as we would need to reject a formula accepted by four out of five agents to reach this judgment (and we know there is another, smaller, majority we can go

against). In fact $f_{\text{Kem}}(\boldsymbol{J}) = \{(p \wedge r), (p \wedge s), \neg q, \neg(p \wedge q)\}$ as we can reach this outcome by going against the three-agent majority on q.[2] △

A second well-known rule based on the weighted majority is the *ranked agenda* rule f_{RA}. Ranked agenda is a generalisation of the ranked pairs voting rule (Tideman, 1987). Intuitively, f_{RA} orders the propositions by the size of their support, then iteratively determines the truth value of each proposition, starting with those propositions that have a larger support and setting each proposition to true when possible—meaning when this does not create inconsistencies. We now formally define this iterative process. Let $m = |\Phi|$. For any profile $\boldsymbol{J}$, and any order $\varphi_1, \ldots, \varphi_m$ of propositions in Φ such that $n^{\boldsymbol{J}}_{\varphi_k} \geq n^{\boldsymbol{J}}_{\varphi_{k+1}}$ for $k \in [1, m-1]$, ranked agenda proceeds as follows. Let $S_0 = \emptyset$. At step k:

- $S_k = S_{k-1} \cup \{\varphi_k\}$ if $S_{k-1} \cup \{\varphi_k\}$ is consistent,
- $S_k = S_{k-1} \cup \{\neg\varphi_k\}$ if $S_{k-1} \cup \varphi_k\}$ is not consistent.

This process will terminate after step m, and $S_m \in f_{\text{RA}}(\boldsymbol{J})$. Note that because two propositions may have an equal number of supporters, there can be several orders that satisfy our requirement. We can think of this as the iterative process, in a sense, "branching" if φ_k and φ_{k+1} have the same number of supporters.

Finally, we define the *leximax* rule f_{lex} (Everaere et al., 2014; Nehring and Pivato, 2019). Given a profile $\boldsymbol{J}$ and a judgment $J \in \mathcal{J}(\Phi)$, we define a dominance relation $\succ_{\boldsymbol{J}}$ such that $J \succ_{\boldsymbol{J}} J'$ if and only if there is some $k \in \{\lceil \frac{n}{2} \rceil, \ldots, n\}$ such that

- $|\{\varphi \in \Phi \mid n^{\boldsymbol{J}}_{\varphi} = k\} \cap J| > |\{\varphi \in \Phi \mid n^{\boldsymbol{J}}_{\varphi} = k\} \cap J'|$—the judgment J accepts a larger number of propositions with k supporters than J'—and,
- for all $k' > k$ we have that $|\{\varphi \in \Phi \mid n^{\boldsymbol{J}}_{\varphi} = k'\} \cap J| = |\{\varphi \in \Phi \mid n^{\boldsymbol{J}}_{\varphi} = k'\} \cap J'|$—for formulas with more than k supporters, the two sets do not differ in the number of formulas they accept.

$f_{\text{lex}}(\boldsymbol{J})$ is defined as the set of judgments that are not $\succ_{\boldsymbol{J}}$-dominated.

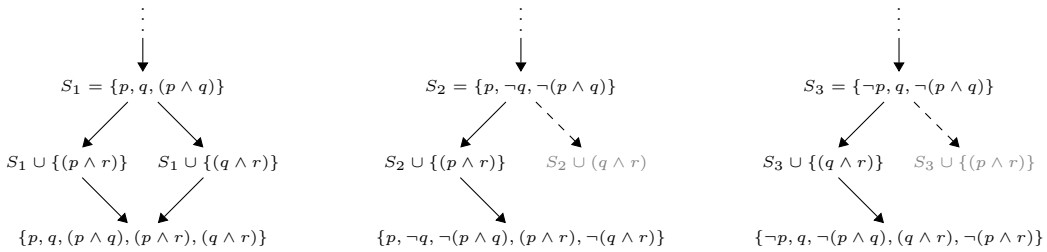
$$f_{\text{lex}}(\boldsymbol{J}) = \{J \mid J \succeq_{\boldsymbol{J}} J' \text{ for all } J' \in \mathcal{J}(\Phi)\}$$

Ranked agenda and leximax are similar in that they both prioritise large majorities over small one, and iteratively add propositions to a set, starting from those with highest support. Ranked agenda, however, does not break ties by "looking ahead" as leximax does. When no two propositions have equal support, the two rules will output the same judgments. We give an example of a profile where the two rules return different outcomes to demonstrate exactly why and how they differ.

[2] While here we have $f_{\text{Kem}}(\boldsymbol{J}) \subseteq f_{\text{Sla}}(\boldsymbol{J})$, this is not always the case. Kemeny may in some cases go against the majority on a larger number of formulas than Slater if the majorities supporting these formulas is small enough.

5.6. Example. Take the following nine-agent profile $\boldsymbol{J}$.

	p	q	$p \wedge q$	$p \wedge r$	$q \wedge r$
3 agents	Yes	No	No	Yes	No
3 agents	No	Yes	No	No	Yes
2 agents	Yes	Yes	Yes	Yes	Yes
1 agent	Yes	Yes	Yes	No	No
$m(\boldsymbol{J})$	Yes	Yes	No	Yes	Yes



Let us first order these formulas by the size of their support: $n^{\boldsymbol{J}}_p = n^{\boldsymbol{J}}_q = n^{\boldsymbol{J}}_{\neg(p \wedge q)} = 6$, and $n^{\boldsymbol{J}}_{(p \wedge r)} = n^{\boldsymbol{J}}_{(q \wedge r)} = 5$. Both f_{RA} and f_{lex} start with the formula with largest support and they both "add formulas" based on the size of their support as long as adding the formula to the set does not cause inconsistencies. First, note that any judgment in the outcome (for both RA and leximax) will include two of the formulas accepted by a 6-to-3 majority—p, q, and $\neg(p \wedge q)$. There are three of these sets as we see in the first "tier" of the graph above, S_1, S_2 and S_3. We then add formulas with a smaller majority margin. Dashed edges show where we are unable to add a formula due to inconsistencies. It is clear from the picture that ranked agenda will return all three leaves—$\{p, q, (p \wedge q), (p \wedge r), (q \wedge r)\}$, $\{p, \neg q, \neg(p \wedge q), (p \wedge r), \neg(q \wedge r)\}$, and $\{\neg p, q, \neg(p \wedge q), (q \wedge r), \neg(p \wedge r)\}$.

For leximax however, we can see that the last two leaves are both $\succ_{\boldsymbol{J}}$-dominated by the leftmost leaf. The leftmost judgment accepts two formulas with a 5-to-4 majority, whereas the other two accept only one such formula. For formulas with larger majorities, the two sets each accept two formulas. Thus we can conclude that $f_{\mathrm{lex}}(\boldsymbol{J}) = \{p, q, (p \wedge q), (p \wedge r), (q \wedge r)\}$. △

Finally, we examine a rule based on elementary changes in the profile. In order to define this, we first define the Hamming distance *between two profiles* $\boldsymbol{J}$ and $\boldsymbol{J'}$ as $H_P(\boldsymbol{J}, \boldsymbol{J'}) := \sum_{i \in N} H(J_i, J'_i)$. The *Dodgson rule* (for odd n) is defined as follows:

$$f_{\mathrm{Dod}}(\boldsymbol{J}) = \{\, m(\boldsymbol{J'}) \mid \operatorname*{argmin}_{\boldsymbol{J'} \in \mathcal{M}(\Phi, n)} H_P(\boldsymbol{J}, \boldsymbol{J'}) \,\}$$

This rule is also known as the *minimal-profile-change rule* (Lang et al., 2017) and as the *"full" distance-based rule* (Miller and Osherson, 2009). The Dodgson rule chooses those judgments that can be reached by making the smallest number

of atomic changes to the profile before reaching a consistent majority, where an atomic change consists in changing the judgment of a single agent on a single formula (and its negation). Dodgson then returns the majority outcome of these profiles.

We will see that among these rules, Slater, Kemeny and leximax fall within a class of rules called *additive majority rules*, and ranked agenda and maximal Condorcet are closely related to rules in this class. Additive majority rules will be our main focus in this chapter, and this is also the class of rules for which we are able to establish the strongest positive results. We do not consider the class of scoring rules in judgment aggregation defined by Dietrich (2014), or distance-based rules that are not majoritarian. This is because our results are, at their core, based on the domain where the majority outcome is consistent. Majoritarian rules are by definition exactly those that are well-behaved on this particular domain.

5.2.2 Additive Majority Rules

In this section we review the large family of aggregation rules called *additive majority rules*. This family includes some of the most important aggregation rules discussed in the literature, notably the Kemeny rule and the Slater rule. We first define and review this family of rules in some detail. We then show that its most prominent exponents are *not* fully strategyproof, before proving that nevertheless all rules in the family are strategyproof on profiles with a consistent majority.

Recall that our rules are defined for odd n. A judgment aggregation rule f is an *additive majority rule* (AMR) if there exists a non-decreasing *gain function* $g : [0, n] \to \mathbb{R}$ with $g(k) < g(k')$ for any $k < \frac{n}{2}$ and $k' \geq \frac{n}{2}$ such that, for any profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$, the following condition is satisfied:

$$f(\boldsymbol{J}) \quad = \quad \operatorname*{argmax}_{J \in \mathcal{J}(\Phi)} \sum_{\varphi \in J} g(n_\varphi^{\boldsymbol{J}})$$

The family of additive majority rules was first identified by Nehring and Pivato (2019). Here we have slightly adapted their original definition to our needs: on the one hand, we only consider rules that weight all formulas equally, and on the other, we consider a slightly larger family of gain functions g. Because of this, Slater is an AMR by our definition, but falls outside of the class according to theirs. Additive majority rules are based on the weighted majoritarian set, meaning that for each formula φ in the agenda the rule only looks at *how many* agents have φ in their judgment. Rules within this family differ only in how much they prioritise large majorities over small ones. Nehring and Pivato (2019) call this the *elasticity* of the gain function. Elasticity quantifies how far we can "stretch" the size of the majority before the aggregation rule will "snap", and change the outcome it returns. On one end of this spectrum lie rules for which

the size of the majority does not play a large (or even any) role; on the other end, we find rules that prioritise large majorities over small ones. Observe that the requirement of $g(k) < g(k')$ for $k < \frac{n}{2}$ and $k' \geq \frac{n}{2}$ ensures that every AMR is majority-preserving.

The additive majority rules include three of the most studied majority-preserving rules in judgment aggregation. Let us now give their definitions in terms of a gain function. The first is the Kemeny rule f_{Kem}, defined by the simplest of gain functions:

$$g(x) \quad = \quad x$$

The Slater rule f_{Sla} is defined by the following gain function:

$$g(x) \quad = \quad \begin{cases} 0 & \text{if } 0 \leq x < \frac{n}{2} \\ 1 & \text{if } \frac{n}{2} \leq x \leq n \end{cases}$$

Thus, f_{Sla} rule considers all formulas accepted by a majority of agents as equal, and tries to respect as many of these majorities as is possible without violating consistency. In particular, it will not distinguish between a unanimously accepted formula and one accepted by just $\lceil \frac{n}{2} \rceil$ agents.

A third AMR of prominence is the leximax rule f_{lex}. Recall that leximax gives maximal preference to stronger majorities, meaning that it orders the formulas in the agenda in terms of the number of agents supporting them and then tries to accept as many formulas supported by a given number of agents as possible before moving on to formulas with fewer supporters. The leximax rule is the AMR with the following gain function:

$$g(x) \quad = \quad |\Phi|^x$$

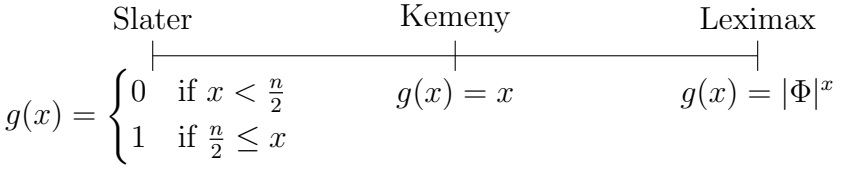
Leximax lands on the opposite side of the elasticity spectrum compared to Slater; while Slater does not distinguish at all between small majorities and large ones, f_{lex} will never prioritise any number of small majorities over a single large one. For example, it will choose a single formula accepted by n agents, over $|\Phi| - 1$ formulas each accepted by $n - 1$ agents.

It is clear that our definitions of Slater and Kemeny in terms of the gain function are equivalent to the standard definitions of these rules that we gave in Section 5.2.1. The same holds, of course, for leximax. Our definition ensures that if $n_{\varphi}^{\boldsymbol{J}} > n_{\psi}^{\boldsymbol{J}}$, then no matter the support for any other formulas in Φ, if adding φ to the outcome will not break consistency, then ψ will never be chosen over φ.

The class of additive majority rules includes many more rules of practical interest. Let us highlight two further examples, characterised by the following gain functions:

$$g(x) = \sum_{k=1}^{x} \frac{1}{k} \qquad g(x) = x \sum_{k=0}^{x} \epsilon^k \text{ for } \quad \epsilon \ll 1$$

Figure 5.1: AMRs on the elasticity spectrum.



The first rule falls somewhere between Slater and Kemeny in terms of elasticity; like Kemeny, it distinguishes between small and large majorities, but the "marginal returns" gained from additional support diminish as majorities grow larger. The second rule is very close to the Kemeny rule, but will prioritise large majorities slightly more. The rule can be seen as a way to break ties between Kemeny outcomes; it gives extra importance to larger majorities only insofar as this can be helpful in differentiating between outcomes that othewise would be considered equally appealing.

5.3 Strategyproofness in Judgment Aggregation

As we've more than hinted, it is essentially impossible to design an aggregation rule that is immune to manipulation by strategic agents while also ensuring that the rule will always return an outcome that is logically consistent. In this section we will formalise what we mean by this, review some of the literature on strategyproofness in JA, and delve into the formal definition of our new domain-specific strategyproofness axiom.

5.3.1 Preferences and Manipulability

We have so far avoided explicitly discussing agents' preferences over judgments, but to discuss strategyproofness and incentives for manipulation, we need to define an agent's preferences. Since agents hold and submit judgments rather than rankings over possible outcomes, we cannot directly reason about their preferences over these judgments. Still, following Dietrich and List (2007c), we will assume that an agent's preferences over outcomes are related to their truthfully held judgments and that we can glean at least some information about their preferences by extrapolating from those judgments. Specifically, we assume that an agent's most preferred outcome is their own truthful judgment. In many cases it makes sense to also assume that agents like outcomes less the further away they are from their true judgment, according to some notion of distance.

An agent i with true judgment $J_i \in \mathcal{J}(\Phi)$ is said to have *closeness-respecting preferences* if $J \cap J_i \supseteq J' \cap J_i$ implies $J \succeq_i J'$ for all $J, J' \in \mathcal{J}(\Phi)$. We

focus on a special case of closeness-respecting preferences based on the Hamming distance: agent i has *Hamming preferences* in case $J \succeq_i J'$ if and only if $H(J, J_i) \leq H(J', J_i)$. In this chapter, we will only consider agents with Hamming preferences over judgments, unless otherwise stated. Assuming that agents have Hamming preferences amounts to assuming that they care equally about every proposition in the agenda. This is a strong assumption that will not be justified in all circumstances, but in the absence of domain-specific information about preferences it is arguably the most natural way to proceed. Hamming preferences have indeed been the dominant choice in the literature on strategic behaviour in judgment aggregation to date (Baumeister et al., 2017). They have also been used to analyse strategic manipulation of social welfare functions (Bossert and Storcken, 1992; Athanasoglou, 2016).

5.3.2 Strategyproof Aggregation Rules

Let us now recall the standard definition of strategyproofness that we will use to review a well-known result showing that designing rules of practical interest that are strategyproof in this sense is essentially impossible. Let $\boldsymbol{J}$ be a profile such that J_i is agent i's truthful judgment, inducing her preference order $\succeq_i$ over judgments. Then a *resolute* aggregation rule f is *manipulable* by agent i in profile $\boldsymbol{J}$, if there exists a profile $\boldsymbol{J}' =_{-i} \boldsymbol{J}$ such that $f(\boldsymbol{J}') \succ_i f(\boldsymbol{J})$. An aggregation rule is *strategyproof* if it is not manipulable by any agent in any profile $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$.

The study of strategic manipulation in judgment aggregation was initiated by Dietrich and List (2007c). They showed that only rules belonging to a very narrowly defined family—the most attractive representatives of which are the majority rule and other so-called quota rules that accept a given proposition whenever a certain number of agents do—are immune to strategic manipulation. These rules, however, are inadequate for many applications, because they cannot guarantee the consistency of outcomes.

5.7. Theorem (Dietrich and List, 2007c). *A resolute judgment aggregation rule is strategyproof for all closeness-respecting preferences if and only if it is independent and monotonic.*

The axiom of independence requires that deciding whether f will accept φ is possible by only considering how the individual agents judge φ, while monotonicity requires that additional support for an accepted proposition φ never gets φ rejected. Formally, f is independent and monotonic if and only if $N_\varphi^{\boldsymbol{J}} \subseteq N_\varphi^{\boldsymbol{J}'}$ implies $\varphi \in J \Rightarrow \varphi \in J'$ for $f(\boldsymbol{J}) = \{J\}$ and $f(\boldsymbol{J}') = \{J'\}$ (Botan et al., 2016). Both axioms feature prominently in impossibility theorems, which essentially show that any rule that satisfies them is bound to return inconsistent outcomes for some profiles (Dokow and Holzman, 2010; List and Pettit, 2002; Nehring and Puppe,

2007). Among the standard aggregation rules, the only ones that satisfy both independence and monotonicity are the quota rules (Dietrich and List, 2007b). Although this class of rules can guarantee strategyproofness for a large family of preferences, they do not always return a consistent outcome and thus, arguably, are of little practical interest. This is why Theorem 5.7 must be interpreted as a negative result. Indeed, the characterisation result by Dietrich and List (2007c) is most often viewed under a similar lens as the Gibbard-Sattertwaite Theorem: it suggests that there are no attractive rules that are strategyproof.

Prior work aimed at addressing this dilemma has identified various fruitful directions. We mention three particularly successful approaches. One way of circumventing the conflict between strategyproofness and a consistent majority is to identify *restricted domains* of profiles of individual judgments for which better performance of certain rules, notably the majority rule, can be guaranteed (List, 2003). We will go into further detail on several existing domain restrictions in Section 5.3.5. A more recent approach has been to analyse the computational complexity of the manipulation problem as a means of providing *complexity barriers* against unwanted behaviour (Endriss et al., 2012; de Haan, 2017; Baumeister et al., 2017). Finally, positive results can be found by studying the extent to which limiting access to relevant information may serve as an effective *informational barrier* against manipulation (Terzopoulou and Endriss, 2019). The route we take in this chapter—which is to offer a more fine-grained analysis of the concept of strategyproofness itself—is complementary to these approaches. Another natural approach to overcoming the lack of strategyproof rules is to restrict attention to strategyproofness for Hamming preferences only, rather than strategyproofness for *all* closeness-respecting preferences. Unfortunately we will see in Section 5.3.3 that for the most well-known majority-preserving rules this also is not attainable.

The most closely related body of research to our own work concerns the manipulation of *social welfare functions* that map profiles of preference orders to collective preference orders. Bossert and Storcken (1992) were the first to study this problem and suggested to model the preferences of agents over preference orders in terms of the *Kendall-tau distance* between orders.[3] While guaranteeing strategyproofness in this model is generally impossible (Bossert and Storcken, 1992; Athanasoglou, 2016), Bossert and Sprumont (2014) obtain positive results for a weak form of strategyproofness that considers only manipulations which bring about an outcome that is *between* an agent's true preference order and the current outcome. This means that the outcome of any manipulation must be comprised *only* of atomic changes in the agent's favour. They find that several important preference aggregation rules, among them the Kemeny and the Slater rule, are strategyproof in this sense. Sato (2015) presents several refinements

[3]This corresponds to using the *Hamming distance* to model preference over judgments, which is what we will do in this chapter.

of these results. Athanasoglou (2016) considers strategyproofness for Hamming preferences in preference aggregation, and shows that for a sufficient number of alternatives, several of the rules considered by Bossert and Sprumont (2014) fail this stronger strategyproofness requirement.

5.3.3 Hamming Strategyproofness

As we've seen, Theorem 5.7 excludes the possibility of Kemeny, Slater, or leximax being strategyproof for all closeness-respecting preferences. It leaves open, however, the possibility that they are strategyproof for Hamming preferences. Indeed Kemeny and Slater, whose standard distance-based definitions are closely tied to the Hamming distance, seem to be promising candidates for rules that are strategyproof in this sense. We are now going to see that this is not the case, and that all three rules are manipulable on the full domain for a sufficiently large agenda.

Athanasoglou (2016) shows for social welfare functions that both Kemeny and Slater are manipulable for all preference extensions, when the number of alternatives exceeds three. As any preference profile can be embedded into judgment aggregation (Endriss, 2016), and as the outcomes of the Kemeny and Slater judgment aggregation rules will agree with their social welfare function counterparts in the preference aggregation domain, we obtain the following result. While we restrict the set of complete and consistent judgments without explicitly spelling out the formulas in the agenda themselves, we note that, by a result of Dokow and Holzman (2010), it is possible to construct an agenda with these structural properties (and we can, conveniently, abstract away from the specifics of Φ). We will use this fact several times in this chapter.

5.8. Proposition (Athanasoglou, 2016). *The Kemeny rule and the Slater rule are manipulable under all preference extensions.*

Proof: Let the agenda Φ be such that the following are the only complete and consistent judgments:

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
J_1	No	No	No	Yes	No	No
J_2	No	Yes	Yes	No	No	No
J_3	Yes	Yes	Yes	Yes	Yes	Yes
J_4	Yes	No	Yes	Yes	No	No
J_5	Yes	No	No	Yes	No	Yes

Let $\boldsymbol{J}$ be the following profile:

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
J_1	No	No	No	Yes	No	No
J_2	No	Yes	Yes	No	No	No
J_3	Yes	Yes	Yes	Yes	Yes	Yes
$m(\boldsymbol{J})$	No	Yes	Yes	Yes	No	No

Let $\boldsymbol{J'} =_{-1} \boldsymbol{J}$, and suppose now agent 1 submits $J_1' = J_4$ in profile $\boldsymbol{J'}$ and no one else changes their judgment. It is a (frustrating but) simple matter to check that $f_{\text{Kem}}(\boldsymbol{J}) = f_{\text{Sla}}(\boldsymbol{J}) = \{J_2\}$ and $f_{\text{Kem}}(\boldsymbol{J'}) = f_{\text{Sla}}(\boldsymbol{J'}) = \{J_4\}$. First, note that for both profiles $\boldsymbol{J}$ and $\boldsymbol{J}$, all majorities are two to one, meaning Slater and Kemeny must coincide in these profiles. We can check the distance from the majority outcome to each of the sets $J_1, \dots J_5$, which will give us the outcomes $\{J_2\}$ and $\{J_4\}$, respectively.

Because $J_2 \succ_i J_4$—and both outcomes are singletons—we see that for any preference extension where $a \succ_i b$ implies $\{a\} \mathrel{\mathring{\succ}}_i \{b\}$, both Slater and Kemeny are manipulable by agents with Hamming preferences over judgments. □

We now show that the same holds for the leximax rule.

5.9. Proposition. *The leximax rule is manipulable under all preference extensions.*

Proof: Let $\boldsymbol{J}$ be the profile below (borrowed from Lang et al. (2017)), with $\Phi^+ = \{p \wedge r, p \wedge s, q, p \wedge q, t\}$ and 16 agents, including one distinguished agent i:

	$p \wedge r$	$p \wedge s$	q	$p \wedge q$	t
6 agents	Yes	Yes	Yes	Yes	Yes
7 agents	No	No	Yes	No	No
2 agents	Yes	Yes	No	No	Yes
J_i	Yes	Yes	No	No	Yes
Maj	Yes	Yes	Yes	No	Yes

We first note the support for the formulas in the agenda Φ:

$$n_q^{\boldsymbol{J}} = 13 \qquad n_{\neg(p\wedge q)}^{\boldsymbol{J}} = 10 \qquad n_{p\wedge r}^{\boldsymbol{J}} = n_{p\wedge s}^{\boldsymbol{J}} = n_t^{\boldsymbol{J}} = 9$$

It is clear then that $f_{\text{lex}}(\boldsymbol{J}) = J = \{\neg(p \wedge r), \neg(p \wedge s), q, \neg(p \wedge q), t\}$—if we iteratively add formulas to our set, we can see that q, and $\neg(p \wedge q)$ must be included, implying that both $(p \wedge r)$ and $(p \wedge s)$ must be excluded. Finally, including t does not create any inconsistency. Let now $\boldsymbol{J'}$ be an i-variant of $\boldsymbol{J}$ where $J_i' = \{p \wedge r, p \wedge s, q, p \wedge q, t\}$—meaning the grey cells in the table above are "flipped". Then:

$$n_{\neg(p\wedge q)}^{\boldsymbol{J'}} = n_{p\wedge r}^{\boldsymbol{J'}} = n_{p\wedge s}^{\boldsymbol{J'}} = n_t^{\boldsymbol{J'}} = 9$$

Rejecting $p \wedge q$ will therefore no longer maximise gain, and simple calculation tells us $f_{\text{lex}}(\boldsymbol{J'}) = J'_i$. As agent i has Hamming preferences and $H(J_i, J) = 3 > 2 = H(J_i, J)$, we know $J'_i \succ_i J$, which implies $f_{\text{lex}}(\boldsymbol{J'}) \mathrel{\mathring{\succ}_i} f_{\text{lex}}(\boldsymbol{J})$ for any preference extension $\mathring{\succeq}_i$. □

We can see that strategyproofness on the full domain is is too demanding a property. It is unattainable for the salient additive majority rules, even when we restrict attention to Hamming preferences and are free to choose any preference extension.

5.3.4 Domain-Strategyproofness

Our proposal in this chapter is to consider a carefully weakened notion of strategyproofness, parametrised by some domain $\mathcal{D}$ of profiles of individual judgments. Under this novel notion of strategyproofness we require immunity to manipulation only in two situations: when the truthful profile belongs to $\mathcal{D}$, or when the profile the manipulating agent might deviate to belongs to $\mathcal{D}$. While this notion is related to the idea of imposing a restriction on the domain on which the aggregation rule is defined (List, 2003), we do not actually impose any such restriction in our work.

Our specific focus is on the domain $\mathcal{M}$ of profiles that guarantee consistent outcomes under the majority rule. A rule that is $\mathcal{M}$-strategyproof will be immune to manipulation in all those cases in which the (strategyproof) majority rule would return a consistent outcome (and thus would be useable at all), while also returning consistent outcomes for all other profiles.

Let $\boldsymbol{J} \in \mathcal{J}(\Phi)^n$ be a profile, with J_i being agent i's truthful judgment. Let $\succeq_i$ be agent i's preference order over judgments, and $\mathring{\succeq}_i$ her preference order over sets of judgments. Then f is *manipulable* by agent i in profile $\boldsymbol{J}$, if there exists a profile $\boldsymbol{J'} =_{-i} \boldsymbol{J}$ such that $f(\boldsymbol{J'}) \mathrel{\mathring{\succ}_i} f(\boldsymbol{J})$. An *irresolute* aggregation rule is *strategyproof* under a given preference extension e if it is not manipulable by any agent i where $\mathring{\succeq}_i = e(\succeq_i)$. We say that f is *$\mathcal{D}$-manipulable* by agent i in $\boldsymbol{J}$ if there exists another profile $\boldsymbol{J'} =_{-i} \boldsymbol{J}$ such that $f(\boldsymbol{J'}) \mathrel{\mathring{\succ}_i} f(\boldsymbol{J})$ and at least one of $\boldsymbol{J}$ and $\boldsymbol{J'}$ belong to $\mathcal{D}$. If only $\boldsymbol{J'}$ belongs to $\mathcal{D}$, we say agent i can manipulate *to* $\mathcal{D}$. If only $\boldsymbol{J}$ belongs to $\mathcal{D}$, we say agent i can manipulate *from* $\mathcal{D}$. A rule is called *$\mathcal{D}$-strategyproof* under a preference extension e if it is not $\mathcal{D}$-manipulable by any agent $i \in N$ where $\mathring{\succeq}_i = e(\succeq_i)$.

The main notion of strategyproofness we will investigate in this chapter is $\mathcal{M}$-strategyproofness, or *majority-strategyproofness*. Note that a rule being majority-preserving does not guarantee $\mathcal{M}$-strategyproofness. For example, a rule that outputs the majority judgment if consistent and otherwise outputs a fixed judgment clearly is majority-preserving but not $\mathcal{M}$-strategyproof. $\mathcal{M}$-strategyproofness of a majority-preserving rule guarantees that the majority outcome *will* in fact be preserved, even under the assumption that agents will manipulate if they have

an incentive to do so. Such a rule would also guarantee that the set of "manipulable" profiles are a subset of the profiles resulting in an inconsistent outcome when using the (strategyproof) majority rule, as any manipulation must be between profiles where the majority rule would result in an inconsistent outcome. Thus, there is a sense in which $\mathcal{M}$-strategyproof rules will minimise the regret of the mechanism designer; if we—as the mechanism designer—care to a great extent about consistency and non-manipulability, it will never be preferable to use the majority rule over an $\mathcal{M}$-strategyproof majority-preserving rule that can guarantee consistency.

5.3.5 Restricted Domains

Before moving on, let us highlight the close connection between strategyproofness and *domain restrictions* in the literature (Dietrich and List, 2010). We briefly review four of the central domain restrictions in judgment aggregation, and examine how exactly they relate to our notion of domain-strategyproofness—and the domain $\mathcal{M}$.

Utilising a domain restriction amounts to restricting the potential input of an aggregation rule to a set of well-behaved profiles. Domain-strategyproofness similarly exploits the well-behavedness of a (sub)domain of profiles, but does so without restricting the actual input to the aggregation rule. There are several domain restrictions in judgment aggregation that are used to obtain positive results—mainly, that a consistent majority outcome can be guaranteed on these domains. Thus, many known domain restrictions will give positive results in terms of domain-strategyproofness as well, as they are built to guarantee a consistent majority outcome within the domain. We define four known domain restrictions in judgment aggregation. Common among them all is that they rely on an ordering, either of the agents or of the agenda—similar to, say, the single-peakedness condition we've seen for preference profiles (Black, 1948).

The first restriction we consider is *unidimensional alignment* (List, 2003). Unidimensional alignment requires that the agents accepting a proposition are either all to the left, or all to right of those rejecting that proposition. A second restriction based on an ordering of the agents is *unidimensional orderedness* (Dietrich and List, 2010). Unidimensional orderedness is a slightly weaker condition that only requires the agents accepting a proposition are adjacent to each other. Often this ordering of agents is thought of as placing each agent in a position on, say, a political spectrum. Finally, *single-plateauedness* and *single-canyonedness* (Dietrich and List, 2010) both rely on an ordering of the agenda. One example of such an ordering might be an order of gift ideas for a friend from least to most expensive. In a single-plateaued profile each agent would give a price-interval that indicates how much they are willing to spend. In a single-canyoned profile agents might be seen as wanting either to go all out and spend a lot of money on

the gift (within some range), or otherwise just buy a cheap item (again within some low-cost range).

- A profile $\boldsymbol{J}$ is *unidimensionally aligned* whenever there exists a linear ordering $\triangleright$ of the agents such that for every $\varphi \in \Phi$, we have that for all $i \in N^{\boldsymbol{J}}_{\varphi}$ and all $i' \in N^{\boldsymbol{J}}_{\neg\varphi}$, it is the case that $i \triangleright i'$.
- A profile $\boldsymbol{J}$ is *unidimensionally ordered* whenever there exists a linear ordering $\triangleright$ of the agents such that for every $\varphi \in \Phi$, we have that $N^{\boldsymbol{J}}_{\varphi} = \{i \in N \mid i_\ell \triangleright i \triangleright i_r\}$ for some $i_\ell, i_r \in N$.
- A profile $\boldsymbol{J}$ is *single-plateaued* whenever there exists a linear ordering $\triangleright$ of the agenda such that for all $i \in N$ we have $A_i = \{\varphi \in \Phi \mid \varphi_\ell \triangleright \varphi \triangleright \varphi_r\}$.
- A profile $\boldsymbol{J}$ is *single-canyoned* whenever there exists a linear ordering $\triangleright$ of the agenda such that for all $i \in N$ we have $A_i = \Phi \setminus \{\varphi \in \Phi \mid \varphi_\ell \triangleright \varphi \triangleright \varphi_r\}$.

Common among all these domain restrictions in judgment aggregation is that they enable the majority to return a logically consistent outcome for any profile in the domain, meaning that the set of unidimensionally aligned profiles, for example, is indeed a subset of $\mathcal{M}$.

5.10. Theorem (List, 2003; Dietrich and List, 2010). *For odd n, if $\boldsymbol{J}$ is unidimensionally ordered, unidimensionally aligned, single-plateauded, or single-canyoned, then $m(\boldsymbol{J})$ is consistent.*

This is particularly nice as these restrictions have an intuitive explanation and describe a particular structure that we may see in a real-life profile of judgments. Thus, we know that the domain of profiles with a consistent majority include many natural structures that can arise in practice. We also know that an agent cannot manipulate a majority-preserving rule within the domain of unidimensionally aligned profiles; domain-strategyproofness looks to strengthen such results by excluding the possibility of any incentive for manipulation from or to, for example, a unidimensionally aligned profile.

5.4 Majority-Strategyproofness of Additive Majority Rules

While we cannot guarantee strategyproofness on the full domain, it turns out that $\mathcal{M}$-strategyproofness is attainable for Hamming preferences and a large class of preference extensions. Before presenting our main result, we prove three technical lemmas. The first establishes a relation between majority outcomes in two profiles that are i-variants, and the second links the notion of betweenness to the Hamming distance.

5.11. Lemma. *For profiles $\boldsymbol{J} =_{-i} \boldsymbol{J'}$, $m(\boldsymbol{J})$ is between J_i and $m(\boldsymbol{J'})$.*

Proof: As all judgments involved are complete and complement-free, we simply need to show $m(\boldsymbol{J}) \subseteq J_i \cup m(\boldsymbol{J'})$. Take any $\varphi \in m(\boldsymbol{J})$. Suppose $\varphi \notin J_i$. If $J'_i =_\varphi J_i$, then $N_\varphi^{\boldsymbol{J'}} = N_\varphi^{\boldsymbol{J}}$, so $\varphi \in m(\boldsymbol{J'})$. But if $J'_i \neq_\varphi J_i$, then $\varphi \in J'_i$ and $n_\varphi^{\boldsymbol{J'}} > n_\varphi^{\boldsymbol{J}}$, so again $\varphi \in m(\boldsymbol{J'})$. □

The following lemma is a somewhat well-known fact—implicit in the work of Duddy and Piggins (2012) for example, who prove the equivalent statement for preference orders. We give a proof for the context of judgment aggregation in the interest of completeness.

5.12. Lemma. *If for complete and complement-free judgment sets J, J', J'', it is the case that J' is between J and J'', then we have that $H(J, J'') = H(J, J') + H(J', J'')$.*

Proof: By definition of betweenness, $J' \subseteq J \cup J''$. To see that

$$H(J', J) + H(J', J'') = |(J'' \setminus J \cup J \setminus J'') \cap J'|$$

note that for any $\varphi \in J'$, there are three cases we need to consider: either $\varphi \in J \setminus J'$, or $\varphi \in J' \setminus J$, or $\varphi \in J \cap J'$. If $\varphi \in J \cap J'$, this means that considering φ does not add to the Hamming distance from J' to J nor to the Hamming distance from J to J''. Thus we only need to consider the first two of three possible cases in order to find the sum of the two Hamming distances. In other words, we can simply count the number of times J and J'' disagree on formulas in J'.

Since $H(J', J) + H(J', J'')$ is the Hamming distance between J and J'' restricted only to the formulas present in J', this distance cannot exceed $H(J, J'')$, meaning it must be the case that $H(J, J') + H(J', J'') \leq H(J, J'')$. This together with the triangle inequality, $H(J, J'') \leq H(J, J') + H(J', J'')$, proves the claim. □

Our final lemma establishes a relationship between majority outcomes and the outcomes of an AMR, in terms of the Hamming distance. By definition, the Slater rule satisfies the property in Lemma 5.13. We show that the same is true for any AMR when restricting our scope to i-variants. This will be useful for proving $\mathcal{M}$-strategyproofness for the class as a whole.

5.13. Lemma. *Let f be an additive majority rule and let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ for some agent i, and such that $m(\boldsymbol{J'})$ is consistent. Then $H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \geq H(m(\boldsymbol{J}), J^*)$ for all $J^* \in f(\boldsymbol{J})$.*

Proof: Let g be the non-decreasing gain function defining f and fix an arbitrary judgment set $J^* \in f(\boldsymbol{J})$. Let $k = H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$ and $k' = H(m(\boldsymbol{J}), J^*)$. So we need to show that $k \geq k'$.

We first derive a constraint on k. Observe that agent i can change the majority outcome for a formula φ under profile $\boldsymbol{J}$ only in case $n_\varphi^{\boldsymbol{J}}$ is equal to either $\lfloor \frac{n}{2} \rfloor$ or $\lceil \frac{n}{2} \rceil$. With this in mind, we can write the total gain for formulas $\varphi \in m(\boldsymbol{J'})$ under profile $\boldsymbol{J}$ as follows:

$$\begin{aligned}
&\sum_{\varphi \in m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \sum_{\varphi \in m(\boldsymbol{J'}) \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; - \sum_{\varphi \in m(\boldsymbol{J}) \setminus m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \; k \cdot g(\lfloor \tfrac{n}{2} \rfloor) \; - \; k \cdot g(\lceil \tfrac{n}{2} \rceil)
\end{aligned}$$

Next, we derive a similar constraint on k'. Let us compute the total gain for formulas $\varphi \in J^*$ under the same profile $\boldsymbol{J}$:

$$\begin{aligned}
&\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; - \sum_{\varphi \in m(\boldsymbol{J}) \setminus J^*} g(n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; - \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} g(n - n_\varphi^{\boldsymbol{J}}) \\
&= \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \sum_{\varphi \in J^* \setminus m(\boldsymbol{J})} \big[g(n_\varphi^{\boldsymbol{J}}) - g(n - n_\varphi^{\boldsymbol{J}}) \big]
\end{aligned}$$

As g is a non-decreasing function, $g(n_\varphi^{\boldsymbol{J}}) - g(n - n_\varphi^{\boldsymbol{J}})$ is non-decreasing in $n_\varphi^{\boldsymbol{J}}$. Hence, given that the maximal value that $n_\varphi^{\boldsymbol{J}}$ can take for any $\varphi \notin m(\boldsymbol{J})$—and thus for any $\varphi \in J^* \setminus m(\boldsymbol{J})$—is $\lfloor \frac{n}{2} \rfloor$, the last sum in the equation above is at most equal to $k' \cdot \big[g(\lfloor \frac{n}{2} \rfloor) - g(n - \lfloor \frac{n}{2} \rfloor) \big] = k' \cdot \big[g(\lfloor \frac{n}{2} \rfloor) - g(\lceil \frac{n}{2} \rceil) \big]$. So we obtain:

$$\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \; \leq \sum_{\varphi \in m(\boldsymbol{J})} g(n_\varphi^{\boldsymbol{J}}) \; + \; k' \cdot \big[g(\lfloor \tfrac{n}{2} \rfloor) - g(\lceil \tfrac{n}{2} \rceil) \big]$$

Finally, let us combine the constraints on k and k' that we have derived. Recall that, by assumption, $m(\boldsymbol{J'})$ is a consistent judgment set. So it is available as a potential outcome under profile $\boldsymbol{J}$. Thus, the score of J^*, one of the *actual* outcomes under $\boldsymbol{J}$, must be at least as high as that of $m(\boldsymbol{J'})$:

$$\sum_{\varphi \in J^*} g(n_\varphi^{\boldsymbol{J}}) \; \geq \sum_{\varphi \in m(\boldsymbol{J'})} g(n_\varphi^{\boldsymbol{J}})$$

Putting everything together, and keeping in mind that $g(\lfloor \frac{n}{2} \rfloor) - g(\lceil \frac{n}{2} \rceil) < 0$, we obtain $k \geq k'$ as claimed. □

We can now combine the three lemmas to get our main result.

5.14. Theorem. *Additive majority rules are $\mathcal{M}$-strategyproof under all reflective preference extensions.*

Proof: Let f be the AMR defined by the non-decreasing gain function g, and let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ for some agent i, and J_i is agent i's truthful opinion. We need to show that, if $m(\boldsymbol{J})$ or $m(\boldsymbol{J'})$ is consistent, then it cannot be the case that $f(\boldsymbol{J'}) \succeq_i f(\boldsymbol{J})$ for any reflective preference extension $\overset{\circ}{\succeq}_i$. In other words, we need to show for all $J \in f(\boldsymbol{J})$ and $J' \in f(\boldsymbol{J'})$ that $J' \not\succ_i J$.

From Lemmas 5.11 and 5.12 together, we obtain:

$$H(J_i, m(\boldsymbol{J'})) = H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \tag{i}$$

Note that if both $m(\boldsymbol{J})$ and $m(\boldsymbol{J'})$ are consistent, then as f is majority-preserving, $f(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$ and $f(\boldsymbol{J'}) = \{m(\boldsymbol{J'})\}$. Any possible manipulation between these profiles would therefore imply a possible manipulation of the majority rule. However, Theorem 5.7 tells us no manipulation of the majority rule is possible. Thus, we need only consider the following two cases.

Case 1: For inconsistent $m(\boldsymbol{J})$ and consistent $m(\boldsymbol{J'})$, Lemma 5.13 says that for any outcome $J^* \in f(\boldsymbol{J})$, it is the case that $H(m(\boldsymbol{J}), J^*) \leq H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$. We need to show that $H(J_i, J^*) \leq H(J_i, m(\boldsymbol{J'}))$—or that $J^* \succeq_i m(\boldsymbol{J'})$. With this in mind, take an arbitrary judgment set $J^* \in f(\boldsymbol{J})$. Combining the triangle inequality with Lemma 5.13 and (i), we get (ii):

$$\begin{aligned} H(J_i, J^*) &\leq H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), J^*) \\ &\leq H(J_i, m(\boldsymbol{J})) + H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \\ &= H(J_i, m(\boldsymbol{J'})) \end{aligned} \tag{ii}$$

In other words, for any $J^* \in f(\boldsymbol{J})$ and the unique $J' = m(\boldsymbol{J'}) \in F(\boldsymbol{J'})$, we have that $J^* \succeq_i J'$. So for any reflective preference extension $\overset{\circ}{\succeq}_i$ it cannot be the case that $f(\boldsymbol{J'}) \overset{\circ}{\succ}_i f(\boldsymbol{J})$.

Case 2: For consistent $m(\boldsymbol{J})$ and inconsistent $m(\boldsymbol{J'})$, we know by Lemma 5.13 that $H(m(\boldsymbol{J'}), J^*) \leq H(m(\boldsymbol{J}), m(\boldsymbol{J'}))$ for any $J^* \in f(\boldsymbol{J'})$. We now need to show that $H(J_i, J^*) \geq H(J_i, m(\boldsymbol{J}))$—or that $m(\boldsymbol{J'}) \succeq_i J^*$.

Take an arbitrary judgment set $J^* \in F(\boldsymbol{J'})$. We again use the triangle inequality, Lemma 5.13, and (i) to get (iii):

$$\begin{aligned} H(J_i, J^*) &\geq H(J_i, m(\boldsymbol{J'})) - H(m(\boldsymbol{J'}), J^*) \\ &\geq H(J_i, m(\boldsymbol{J'})) - H(m(\boldsymbol{J}), m(\boldsymbol{J'})) \\ &= H(J_i, m(\boldsymbol{J})) \end{aligned} \tag{iii}$$

In other words, for any $J^* \in f(\boldsymbol{J'})$ and the unique $J = m(\boldsymbol{J}) \in f(\boldsymbol{J})$, we have that $J \succeq_i J^*$. So again, for any reflective preference extension $\mathring{\succeq}_i$ it cannot be the case that $f(\boldsymbol{J'}) \mathring{\succ}_i f(\boldsymbol{J})$.

Taking these cases together we have shown that $f(\boldsymbol{J'}) \not\mathring{\succ}_i f(\boldsymbol{J})$ as desired, meaning agent i cannot successfully manipulate by submitting any untruthful judgment J_i' in place of J_i. □

5.15. Corollary. *The Kemeny, Slater, and leximax rules are $\mathcal{M}$-strategyproof under all reflective preference extensions.*

The majority-strategyproofness of additive majority rules presents a strong argument for their use in lieu of the majority rule. They offer an alternative that guarantees consistency, and ensures that the majority will be preserved in all cases. Importantly they also offer the post-aggregation "check" for majority consistent outcomes, meaning it is possible to recognise cases where no manipulation can have occurred, thereby ensuring we can trust the outcome.

The following proposition tells us that $\mathcal{M}$-strategyproofness of additive majority rules does not hold for closeness-respecting preferences in general.

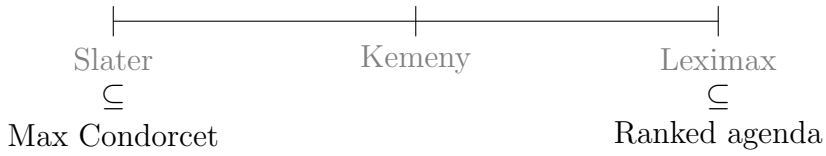
5.16. Proposition. *For any AMR f there is some closeness-respecting preference under which f is not $\mathcal{M}$-strategyproof.*

Proof: Let $\boldsymbol{J}$ be the profile below—we assume Φ is such that J_1, J_2, J_3 are the only consistent judgments—and suppose J_1 is agent 1's truthful opinion. Suppose further that agent 1 has closeness-respecting preferences, but $J_3 \succ_1 J_2$, meaning she does not have Hamming Preferences. Note that this is possible only because $J_1 \cap J_3 \not\subseteq J_1 \cap J_2$. Note further that because the strength of all majorities is equal (a two-to-one majority), all AMRs will agree with the Slater rule on this profile.

	φ_1	φ_2	φ_3	φ_4	φ_5
J_1	Yes	Yes	No	No	No
J_2	No	Yes	No	Yes	Yes
J_3	Yes	No	Yes	Yes	Yes
$m(\boldsymbol{J})$	Yes	Yes	No	Yes	Yes

We calculate the distance from the majority to each possible outcome and find that $f_{\text{Sla}}(\boldsymbol{J}) = \{J_2\}$. Suppose $\boldsymbol{J'} = (J_1', J_2, J_3)$, and $J_1' = J_3$. Then, $f_{\text{Sla}}(\boldsymbol{J'}) = \{m(\boldsymbol{J'})\} = \{J_3\}$. Since i prefers this outcome to J_2, she will have an incentive to manipulate *to* the majority. Since any AMR will agree with Slater both in $\boldsymbol{J}$ and $\boldsymbol{J'}$, this is enough to prove our claim. □

Figure 5.2: Coarsenings of known AMRs.



Results for Other Domains

Let us briefly review how our results relate to the domain restrictions in Section 5.3.4, the most prominent example of which is unidimensional alignment. Let $\mathcal{U}(\Phi, n)$ be the domain of unidimensionally aligned profiles for Φ and n. As $\mathcal{U}(\Phi, n) \subseteq \mathcal{M}(\Phi, n)$ (List, 2003), we immediately obtain:

5.17. Corollary. *Additive majority rules are $\mathcal{U}$-strategyproof under all reflective preference extensions.*

Clearly, this result is not unique to unidimensionally aligned profiles, but holds for any domain restriction in judgment aggregation that guarantees a consistent majority—including the domain of unidimensionally ordered, single-canyoned, and single-plateaued profiles.

5.5 Coarsenings of Additive Majority Rules

In this section we first examine two rules, the maximal Condorcet rule and the ranked agenda rule. These rules are related to the additive majority rules in that they will always return a superset of the outcome of some AMR. It turns out that this particular relationship affords these rules a certain level of protection against manipulation.

We say that an aggregation rule f' is a *coarsening* of a rule f (and f is a *refinement* of f') if $f'(\boldsymbol{J}) \supseteq f(\boldsymbol{J})$ for all profiles $\boldsymbol{J}$. We will examine two such coarsenings of AMRs—the maximal Condorcet rule and the ranked agenda rule, both defined in Section 5.2.1. Their relationship with the AMRs we have seen in this chapter are shown in Figure 5.2. We first show that our strategyproofness results for AMRs do not extend to these rules—we can find a coarsening of an AMR that is manipulable under some preference extension, both to and from $\mathcal{M}$. We will then see that their proximity to the additive majority rules means that they are nevertheless afforded some level of immunity to manipulation. More specifically, we are going to show that for the Kelly preference extension, these rules are $\mathcal{M}$-strategyproof.

Observe that f_{Sla} is a refinement of f_{Con} in that $f_{\text{Sla}}(\boldsymbol{J}) \subseteq f_{\text{Con}}(\boldsymbol{J})$ for all profiles $\boldsymbol{J}$. This is clear from the definition of f_{Sla} given in Section 5.2.1 as the rule

that selects the maximal consistent subset of the majority in terms of cardinality.[4] We are going to show that for the Kelly extension, f_{Con} is $\mathcal{M}$-strategyproof. We first state some weaker strategyproofness results for pessimistic and optimistic agents.

5.18. Proposition. *For any coarsening f of an additive majority rule, there exists some reflective preferences under which f fails $\mathcal{M}$-strategyproofness—both from or to $\mathcal{M}$.*

Proof: *Manipulation* to *Majority:* Let $\boldsymbol{J}$ be the profile below, where J_1 is agent 1's truthful opinion, she has pessimistic preferences, and $\boldsymbol{J'} =_{-1} \boldsymbol{J}$ is such that $J_1' = \{p, \neg q, \neg(p \wedge q), p \wedge r\}$.

	p	q	$p \wedge q$	$p \wedge r$
J_1	Yes	Yes	Yes	Yes
J_2	Yes	No	No	Yes
J_3	No	Yes	No	No
$m(\boldsymbol{J})$	Yes	Yes	No	Yes

We can see that $f_{\text{Con}}(\boldsymbol{J}) = \{\{p, q, p \wedge q, p \wedge r\}, \{p, \neg q, \neg(p \wedge q), p \wedge r\}, \{\neg p, q, \neg(p \wedge q), \neg(p \wedge r)\}\}$, and since $m(\boldsymbol{J'})$ is consistent, $f_{\text{Con}}(\boldsymbol{J'}) = \{\{p, \neg q, \neg(p \wedge q), p \wedge r\}\}$. As $\{p, \neg q, \neg(p \wedge q), p \wedge r\} \succ_1 \{\neg p, q, \neg(p \wedge q), \neg(p \wedge r)\}$, agent 1 can successfully manipulate from from $\boldsymbol{J}$ to $\boldsymbol{J'}$—meaning *to* the majority.

Manipulation from *Majority:* Let $\boldsymbol{J}$ be the profile below, suppose J_1 is agent 1's truthful opinion and suppose that she is optimistic. Further, let $\boldsymbol{J'} =_{-1} \boldsymbol{J}$ be the profile which differs only in that agent 1 submits $J_1' = \{a, b, c, \neg d, (a \wedge \neg d) \rightarrow (b \wedge c)\}$.

	p	q	r	s	$(p \wedge \neg s) \rightarrow (q \wedge r)$
J_1	Yes	Yes	Yes	Yes	Yes
J_2	Yes	No	No	Yes	Yes
J_3	No	No	No	No	Yes
$m(\boldsymbol{J})$	Yes	No	No	Yes	Yes

As $m(\boldsymbol{J})$ is consistent, $f_{\text{Con}}(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$. For $\boldsymbol{J'}$, the majority, $m(\boldsymbol{J'}) = \{p, \neg q, \neg r, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\}$, is not consistent. It is simple to confirm $\{p, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\} \in C(m(\boldsymbol{J'}))$, and thus that $J^* = \{p, q, r, \neg s, (p \wedge \neg s) \rightarrow (q \wedge r)\} \in f_{\text{Con}}(\boldsymbol{J'})$. We calculate the distances from J_1 to find that $J^* \succ_1 m(\boldsymbol{J})$. As there exists some strictly better outcome in $f_{\text{Con}}(\boldsymbol{J'})$, agent 1 can manipulate maximal Condorcet *from* majority.□

[4] In fact this relationship holds between f_{Con} and f_{Kem} as well as between f_{Con} and f_{lex} (Lang et al., 2017), meaning f_{Con} is a coarsening of several AMRs.

All hope for these coarsenings is not lost, however. While the proof of Proposition 5.18 shows that pessimistic agents can manipulate a coarsening of an AMR to majority, and optimistic agents can manipulate from majority, these rules still provide some level of protection against manipulation for both pessimistic and optimistic agents.

5.19. Proposition. *For any coarsening of an additive majority rule a pessimistic agent cannot manipulate* from *majority.*

Proof: Let f' be a coarsening of an AMR f. Further, let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ and $f'(\boldsymbol{J}) = \{m(\boldsymbol{J})\}$. Suppose for contradiction that there is a pessimistic agent i, with truthful opinion J_i, who can manipulate from $\boldsymbol{J}$ to $\boldsymbol{J'}$. Then $J' \succ_i m(\boldsymbol{J})$ for all $J' \in f'(\boldsymbol{J'})$. As $f(\boldsymbol{J'}) \subseteq f'(\boldsymbol{J'})$, this would constitute a successful manipulation of f by a pessimistic agent, which contradicts Theorem 5.14, as f is an AMR. □

5.20. Proposition. *For any coarsening of an additive majority rule an optimistic agent cannot manipulate* to *majority.*

Proof: Let f' be a coarsening of an AMR f. Further, let $\boldsymbol{J}$ and $\boldsymbol{J'}$ be two profiles such that $\boldsymbol{J} =_{-i} \boldsymbol{J'}$ and $f'(\boldsymbol{J'}) = \{m(\boldsymbol{J'})\}$. Suppose for contradiction that there is an optimistic agent i, with truthful opinion J_i, who can manipulate from $\boldsymbol{J}$ to $\boldsymbol{J'}$. Then $m(\boldsymbol{J'}) \succ_i J^*$ for all $J^* \in f'(\boldsymbol{J})$. As $f(\boldsymbol{J}) \subseteq f'(\boldsymbol{J})$, this would constitute a successful manipulation of f by an optimistic agent, which contradicts Theorem 5.14, as f is an AMR. □

We can now use Proposition 5.19 and Proposition 5.20 to prove the following theorem.

5.21. Theorem. *Coarsenings of additive majority rules are $\mathcal{M}$-strategyproof under the Kelly extension.*

Proof: Let f be a coarsening of an AMR. By definition, if an optimistic agent cannot manipulate a rule to the majority, then an agent with Kelly preferences cannot either. This, together with Proposition 5.20, shows that agents with Kelly preferences cannot manipulate f *to* majority. Similarly, if a pessimistic agent cannot manipulate a rule from the majority, then an agent with Kelly preferences cannot either. This, together with Proposition 5.19 shows that agents with Kelly preferences cannot manipulate f *from* majority. Putting these together establishes $\mathcal{M}$-strategyproofness of coarsening of AMRs under the Kelly extension. □

Thus, while a pessimistic or optimistic agent might manipulate a coarsening of an AMR—say the maximal Condorcet rule—these rules do benefit from their relationship with AMRs in terms of $\mathcal{M}$-strategyproofness for agents with Kelly preferences.

Due to the aforementioned relationship between maximal Condorcet and Slater, we get the following result.

5.22. Corollary. *The maximal Condorcet rule is $\mathcal{M}$-strategyproof under the Kelly extension.*

While f_{RA} is not itself an AMR, it is a coarsening of the leximax rule as we have that $f_{\text{lex}}(\boldsymbol{J}) \subseteq f_{\text{RA}}(\boldsymbol{J})$ for all profiles $\boldsymbol{J}$ (Lang et al., 2017). Therefore, our result also applies for the ranked agenda rule.

5.23. Corollary. *The ranked agenda rule is $\mathcal{M}$-strategyproof under the Kelly extension.*

Thus, both maximal Condorcet and ranked agenda benefit from their proximity to known additive majority rules.

5.6 The Dodgson Rule

We conclude our examination by straying even further from the additive majority rules. Recall that the Dodgson rule returns the majority outcome of profiles that can be reached by making the smallest number of atomic changes to the initial profile before reaching a consistent majority. This is clearly a majority-preserving rule, but it is not an AMR. Indeed, it also lacks the strategyproofness properties of the previous majority-preserving rules examined in this chapter.

5.24. Proposition. *The Dodgson rule fails $\mathcal{M}$-strategyproofness for all preference extensions.*

Proof: Let Φ be an agenda with $|\Phi^+| = 10$. Consider the profile $\boldsymbol{J}$ below, with J_1 being agent 1's true judgment:

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}
J_1	No	Yes	Yes	No	No	No	No	No	No	No
J_2	No	No	No	Yes	No	Yes	Yes	No	Yes	Yes
J_3	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No
$m(\boldsymbol{J})$	No	No	No	No	No	Yes	Yes	No	Yes	No

Suppose that—besides J_1, J_2, and J_3 appearing $\boldsymbol{J}$—the only other judgments that are consistent are J_4, J_5, J_6, and J_7 shown below.

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	φ_8	φ_9	φ_{10}
J_4	No	No	No	No	No	Yes	Yes	No	Yes	No
J_5	Yes	Yes	No	Yes	No	No	No	No	No	No
J_6	No	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes
J_7	No	Yes	No	Yes	No	Yes	Yes	No	No	No

As the majority outcome is consistent, $f_{\text{Dod}}(\boldsymbol{J}) = \{m(\boldsymbol{J})\} = \{J_4\}$.

Let $\boldsymbol{J'}$ be an i-variant of $\boldsymbol{J}$ with $J_1' = J_5$, making $m(\boldsymbol{J'})$ inconsistent. We calculate the relevant distances between the allowed judgments and the judgments in the profile $\boldsymbol{J'}$, shown in the following table.

	J_1	J_2	J_3	J_4	J_5	J_6	J_7
J_2	7	0	4	2	6	**1**	2
J_3	7	4	0	2	8	5	4
J_5	3	6	7	6	0	5	4

We see that the minimal number of atomic changes we can make to the profile $\boldsymbol{J}'$—while ensuring all input judgments are consistent—is 1, as $H(J_2, J_6) = 1$. For all other relevant pairwise comparisons of admissible judgments, the Hamming distance between them is 2 or greater. Indeed, replacing J_2 with J_6 will result in profile $\boldsymbol{J}^* = (J_5, J_6, J_3)$, with a consistent majority outcome. Thus $f_{\text{Dod}}(\boldsymbol{J}') = \{m(\boldsymbol{J}^*)\} = \{J_7\}$. As $J_7 \succ_1 J_4$, it must be the case that $f(\boldsymbol{J}') \mathrel{\mathring{\succ}}_1 f(\boldsymbol{J})$ for any preference extension (as we are comparing singleton sets), making this a successful manipulation from $\mathcal{M}$. □

Recall that strategyproofness *from* $\mathcal{M}$ ensures that a majority-preserving rule agrees with the majority, even when accounting for possible strategic manipulation.

For agents with Kelly (and pessimistic) preferences, the following example shows manipulation is possible both to and from $\mathcal{M}$. Thus, for this type of agent, Dodgson will also fail to provide the post-aggregation guarantee that no manipulation has occurred.

5.25. Example (Dodgson Manipulation *to* Majority). Let $\boldsymbol{J}$ be the profile below, where J_1 is agent 1's true judgment, and suppose her preferences are extended according to the Kelly extension. Let Φ be an agenda such that J_1, J_2, and J_3 are the only consistent judgments.

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6
J_1	No	No	No	Yes	No	No
J_2	No	No	Yes	No	Yes	Yes
J_3	Yes	Yes	Yes	Yes	Yes	Yes
$m(\boldsymbol{J})$	No	No	Yes	Yes	Yes	Yes

Note that the majority outcome is not consistent. It is easy to check that $f(\boldsymbol{J}) = \{J_2, J_3\}$. Now let $\boldsymbol{J'}$ be an i-variant of $\boldsymbol{J}$, where $J'_1 = J_2$. Then $f_{\mathrm{Dod}}(\boldsymbol{J'}) = m(\boldsymbol{J'}) = J_2$, as agent 1 prefers J_2 over J_3. As she has Kelly preferences, we have $f_{\mathrm{Dod}}(\boldsymbol{J'}) \mathrel{\mathring{\succ}_1} f_{\mathrm{Dod}}(\boldsymbol{J})$ which is a successful manipulation *to* $\mathcal{M}$. △

The Dodgson rule exemplifies that by no means all majority-preserving rules are associated with some level of $\mathcal{M}$-strategyproofness. Not only does it fail $\mathcal{M}$-strategyproofness for the types of agents we have considered in this chapter, but it is highly susceptible to manipulation by agents with an even wider variety of preferences. This means it presents ample opportunities for manipulation.

5.7 Summary

In this chapter, we have introduced a novel weakening of strategyproofness, which we called domain-strategyproofness. We have argued that in the absence of full strategyproofness, domain-strategyproofness often offers a sufficiently strong barrier against manipulation. We have focused in particular on the majority-consistent domain, and examined majority-preserving aggregation rules, showing varying levels of strategyproofness for several prominent rules from the judgment aggregation literature. Our results make a strong case for the use of additive majority rules, a class of rules that includes both the Kemeny rule and the Slater rule.

Chapter 6

Conclusion

In Chapter 1 we outlined the main question addressed in this thesis:

Is it possible to manipulate *from* a profile in a "well-behaved" domain to one outside the domain in question?

With this question in mind, we set out to identify natural "well-behaved" domains in three different settings. We looked at the Condorcet domain in voting, the party-list domain in multiwinner voting, and the majoritarian domain in judgment aggregation. Throughout the thesis, we made strides towards answering our question in all three settings and we found a mix of positive and negative answers. In Chapter 3 our answers were largely negative as we excluded the possibility that many well-known and axiomatically convincing rules can be robust Condorcet extensions. In Chapters 4 and 5 we were luckier and obtained strategyproofness results for large classes of natural rules—Thiele methods in multiwinner voting and additive majority rules in judgment aggregation.

We now briefly summarise our results and discuss some of the connections between the three main chapters of the thesis. We will then discuss some possible future directions.

6.1 Looking Back

After setting the stage in **Chapter 1**, we moved on to the topic of preferences and preference extensions in **Chapter 2**. Having briefly gone over the history of the problem of *lifting preferences*, we outlined axioms for extensions and defined specific extensions that have parameterised our results throughout the rest of the thesis. After this prep work, we delved into the three main chapters.

In **Chapter 3** we looked at the connection between Condorcet consistency and strategic manipulation. Specifically, we examined Condorcet-consistent (weighted) tournament solutions and strategyproofness on profiles where a Condorcet winner exists. We defined the notion of a robust Condorcet extension as a means of distinguishing between Condorcet-consistent rules that do not incentivise manipulation on profiles with Condorcet winners and those that do. We first established that no weighted tournament solution (and thus no tournament solution) can satisfy robustness for all preference extensions. This held true for all neutral C2 social choice functions, for both an odd and even number of agents. Knowing this, we went looking for tournament solutions that may be robust for *some* preference extensions. Our characterisation result told us that robustness and weak resoluteness are incompatible for all preference extensions. In addition, the result also extended to several well-known weighted tournament solutions such as Kemeny and ranked pairs (this result held for an odd number of agents, but for an even number of agents we were able to find a neutral Condorcet-consistent C1 rule that did in fact satisfy robustness). As a consequence of this impossibility result, we narrowed our search to more indecisive rules. After outlining the connections between Kelly-strategyproofness and robustness—which helped us establish robustness for several well-known Condorcet-consistent tournament solutions—we saw our main positive result of this chapter. We established robustness under all weakly pessimistic extensions for the minimal extending set and all its coarsenings, which included rules that are not Kelly-strategyproof.

In **Chapter 4** we shifted our focus to multiwinner voting rules. In particular we looked at strategyproofness on party-list profiles for the class of Thiele methods. Our interest in Thiele methods was in part based on several impossibility results that highlighted the incompatibility of proportionality and strategyproofness. As Thiele rules as a class aim for some level of proportional representation, this was a natural place to look for possibilities by weakening the strategyproofness requirement. For approval-based elections, we identified the party-list domain as the most fruitful area to hunt for positive results. We had several moving parts to play with here: for all results in this chapter, we specified what type of manipulation—of the three types we looked at—as well as which preference extension(s) the result holds for. We looked at three types of manipulation—free-riding, superset-manipulation, and disjoint-set-manipulation. We showed that Thiele methods are immune to free-riding on party-list profiles for the class of general Gärdenfors preferences. For superset-manipulation and disjoint-set-manipulation, the corresponding result was stronger in that it held for all strongly reflective preference extensions—a larger class of preferences. We also identified a specific preference extension—the optimistic extension—where Thiele methods are fully strategyproof on party-list profiles. Our results in this chapter highlight the trade-off between the strength of the strategyproofness axioms and the preference extensions. For a stronger strategyproofness axiom we were only able to show it was satisfied relative to a specific class of extensions. When weak-

ening the strategyproofness requirement further (by, for example, considering only free-riding) we found results that held for a much larger class of preferences.

Finally, in **Chapter 5** we looked at majoritarian rules in judgment aggregation and studied strategyproofness on profiles with a consistent majority outcome. We studied the most prominent majoritarian judgment aggregation rules and were able to establish varying levels of strategyproofness on profiles with a consistent majority outcome. We first rid ourselves of the notion that we might be able to get strategyproofness results by considering manipulation by agents with Hamming preferences without playing around with any specific domains. After settling on the domain of profiles with a consistent majority outcome, we moved from good to bad in this chapter. Our first—and strongest—result pertained to additive majority rules. We showed that rules in this class are majority-strategyproof for all reflective preferences. This result applies to prominent aggregation rules such as Slater, Kemeny, and leximax. We promptly moved on to coarsenings of additive majority rules. For this class, we were able to establish strategyproofness for the Kelly extension. The main representative rules in this class were the ranked agenda rule and the maximal Condorcet rule. Finally, we examined the Dodgson rule, and showed that it fails majority-strategyproofness for all preference extensions.

As the observant (and not so observant) reader will have noticed, many of the judgment aggregation rules we study in Chapter 5 are counterparts to the voting rules we studied in Chapter 3. More accurately, they are counterparts to the preference aggregation rules that give rise to these voting rules. For example, the Slater preference aggregation rule returns all Slater orders, while the voting rule we studied returned only the top elements of these orders. In addition to this, the judgment aggregation counterparts to preference profiles with a Condorcet winner will always return a consistent majority outcome. Because we can embed any preference aggregation problem into the judgment aggregation framework, it stands to reason that the results we state in Chapter 5 for judgment aggregation rules also hold for preference aggregation rules. It is interesting to note that, while our results from Chapter 5 must hold for preference aggregation rules like Slater and Kemeny, this is not in fact the case for the Slater and Kemeny *voting rules*, which are not robust on profiles with a Condorcet winner. This means that while it is indeed possible for an agent to change the top element of the orders output by, say, the Slater rule, this improvement must be accompanied by changes against an agent's preferences further down the order. This also hints at unexplored territories in the more general judgment aggregation framework. Our positive results held for Hamming preferences, but for more limited preferences (agents who care about only one formula for example), our findings in Chapter 3 indicate that these strategyproofness results will break down, even if intuitively these preferences seem like they would be more conducive to positive results.

6.2 Looking Forward

Before closing, we reflect on some directions for future research.

As we know by now, strategyproofness results rely heavily on the type of preferences under consideration. While an agent with a certain type of preferences may be incentivised to misreport her preferences or opinions, another may find truth-telling is the best strategy. Pragmatism has forced us to make some—arguably justifiable—choices regarding what types of preferences we ascribe to our agents. In Chapter 4 we assumed agents' preferences in the approval-based multiwinner setting are based solely on the number of approved alternatives the agent has in common with the outcome. In Chapter 5 we assumed agents have Hamming preferences over judgments. While both choices are reasonable, they are not the only possibilities.

We also made choices about which frameworks to study in this thesis. While we have considered three different settings in this thesis, they all have in common that the output is a "collective decision". However, domain-specific strategyproofness axioms are likely to lead to interesting results also in other areas of social choice, where the end goal is not a collective decision in this sense. Consider for example, the area of *matching* where we know that there is no algorithm that guarantees stable outcomes while also being strategyproof (Roth, 1982). A domain-specific weakening of strategyproofness can help give insight into when and how it is possible for agents to manipulate a matching algorithm that guarantees stable outcomes. We know there are domain restrictions in this setting, such as top-dominance, where stability and strategyproofness manage to coexist (Alcalde and Barberà, 1994). The area is seemingly ripe for an analysis similar to the one we have conducted here: can agents manipulate from a profile satisfying top-dominance to one that does not? And what effect does this have on the "guarantee" of stability?

In general, this thesis is yet another argument for considering a wider range of strategyproofness axioms. Weakening strategyproofness is a clear way to avoid impossibility results. This approach contributes to a clearer understanding of when and why manipulation occurs. This in turn can inform decisions about how to go about aggregating preferences or opinions. For example, if we find that a particular voting rule is strategyproof under partial information, we may want to use that rule in cases where we know the voters do not know anything about others' preferences. If we are in a setting where the votes are likely to be single-peaked, we might want to use a rule that incentivises truth-telling on those profiles. If we know agents are hesitant to make big sweeping changes to their opinion, we may only need to safeguard against, for example, swaps of adjacent alternatives in an agent's ranking. By considering many different notions of strategyproofness—among them, our domain-specific notions—we are able to paint a much clearer picture of when manipulation occurs, and why strategyproofness fails in some cases and not others.

Bibliography

José Alcalde and Salvador Barberà. Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, 4 (3):417–435, 1994. (page 102)

Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–46, 1950. (page 1)

Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura, editors. *Handbook of Social Choice and Welfare*, volume 1. Elsevier, 2002. (page 1)

Stergios Athanasoglou. Strategyproof and efficient preference aggregation with Kemeny-based criteria. *Games and Economic Behavior*, 95:156–167, 2016. (page 81, 82, 83)

Haris Aziz, Serge Gaspers, Joachim Gudmundsson, Simon Mackenzie, Nicholas Mattei, and Toby Walsh. Computational aspects of multi-winner approval voting. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2015. (page 51)

Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2):461–485, 2017. (page 49, 50, 51)

Jeffrey S. Banks. Sophisticated voting outcomes and agenda control. *Social Choice and Welfare*, 1(4):295–306, 1985. (page 26)

Salvador Barberà. Manipulation of social decision functions. *Journal of Economic Theory*, 15(2):266–278, 1977. (page 12)

Salvador Barberà and Prasanta K. Pattanaik. Extending an order on a set to the power set: some remarks on Kannai and Peleg's approach. *Journal of Economic Theory*, 32(1):185–191, 1984. (page 12)

Salvador Barberà, Charles R. Barrett, and Prasanta K. Pattanaik. On some axioms for ranking sets of alternatives. *Journal of Economic Theory*, 33(2): 301–308, 1984. (page 12)

Salvador Barberà, Walter Bossert, and Prasanta K. Pattanaik. Ranking sets of objects. In *Handbook of Utility Theory*, pages 893–977. Springer, 2004. (page 12, 13)

John J. Bartholdi and James B. Orlin. Single transferable vote resists strategic voting. *Social Choice and Welfare*, 8(4):341–354, 1991. (page 51)

John J. Bartholdi, Craig A. Tovey, and Michael A. Trick. The computational difficulty of manipulating an election. *Social Choice and Welfare*, 6(3):227–241, 1989. (page 21)

Dorothea Baumeister, Jörg Rothe, and Ann-Kathrin Selker. Strategic behavior in judgment aggregation. In Ulle Endriss, editor, *Trends in Computational Social Choice*, chapter 8, pages 145–168. AI Access, 2017. (page 81, 82)

Duncan Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948. (page 4, 20, 86)

Anna Bogomolnaia and Hervé Moulin. A new solution to the random assignment problem. *Journal of Economic theory*, 100(2):295–328, 2001. (page 51)

Walter Bossert and Yves Sprumont. Strategy-proof preference aggregation: Possibilities and characterizations. *Games and Economic Behavior*, 85:109–126, 2014. (page 20, 82, 83)

Walter Bossert and Ton Storcken. Strategy-proofness of social welfare functions: The use of the Kemeny distance between preference orderings. *Social Choice and Welfare*, 9(4):345–360, 1992. (page 81, 82)

Walter Bossert, Prasanta K. Pattanaik, and Yongsheng Xu. Choice under complete uncertainty: axiomatic characterizations of some decision rules. *Economic Theory*, 16(2):295–312, 2000. (page 12)

Sirin Botan. Manipulability of Thiele methods on party-list profiles. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 2021. (page 8)

Sirin Botan and Ulle Endriss. Majority-strategyproofness in judgment aggregation. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2020. (page 9)

Sirin Botan and Ulle Endriss. Preserving Condorcet winners under strategic manipulation. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, 2021. (page 8)

Sirin Botan, Arianna Novaro, and Ulle Endriss. Group manipulation in judgment aggregation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2016. (page 81)

Florian Brandl, Felix Brandt, and Christian Stricker. An analytical and experimental comparison of maximal lottery schemes. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018. (page 21)

Felix Brandt. Minimal stable sets in tournaments. *Journal of Economic Theory*, 146(4):1481–1499, 2011. (page 22, 26, 39)

Felix Brandt. Set-monotonicity implies Kelly-strategyproofness. *Social Choice and Welfare*, 45(4):793–804, 2015. (page 20, 39)

Felix Brandt and Markus Brill. Necessary and sufficient conditions for the strategyproofness of irresolute social choice functions. In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge (TARK)*, 2011. (page 20)

Felix Brandt, Markus Brill, and Paul Harrenstein. Tournament solutions. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 3. Cambridge University Press, 2016. (page 35)

Felix Brandt, Paul Harrenstein, and Hans Georg Seedig. Minimal extending sets in tournaments. *Mathematical Social Sciences*, 87:55–63, 2017. (page 40)

Felix Brandt, Martin Bullinger, and Patrick Lederer. On the indecisiveness of kelly-strategyproof social choice functions. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2021. (page 21)

Robert Bredereck, Piotr Faliszewski, Rolf Niedermeier, and Nimrod Talmon. Complexity of shift bribery in committee elections. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 2016. (page 51)

Robert Bredereck, Piotr Faliszewski, Andrzej Kaczmarczyk, Rolf Niedermeier, Piotr Skowron, and Nimrod Talmon. Robustness among multiwinner voting

rules. In *Proceedings of the 10th International Symposium on Algorithmic Game Theory (SAGT)*, 2017. (page 52)

Robert Bredereck, Andrzej Kaczmarczyk, and Rolf Niedermeier. On coalitional manipulation for multiwinner elections: Shortlisting. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018. (page 51)

Donald E. Campbell and Jerry S. Kelly. A strategy-proofness characterization of majority rule. *Economic Theory*, 22(3):557–568, 2003. (page 35)

Vincent Conitzer and Toby Walsh. Barriers to manipulation in voting. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 6. Cambridge University Press, 2016. (page 21)

Vincent Conitzer, Toby Walsh, and Lirong Xia. Dominating manipulations in voting with partial information. In *25th AAAI Conference on Artificial Intelligence (AAAI)*, 2011. (page 21)

Arthur H. Copeland. A "reasonable" social welfare function. University of Michigan Seminar on Applications of Mathematics to the Social Sciences, 1951. (page 25)

Franz Dietrich. Scoring rules for judgment aggregation. *Social Choice and Welfare*, 42(4):873–911, 2014. (page 78)

Franz Dietrich and Christian List. Arrow's Theorem in judgment aggregation. *Social Choice and Welfare*, 29(1):19–33, 2007a. (page 69)

Franz Dietrich and Christian List. Judgment aggregation by quota rules: Majority voting generalized. *Journal of Theoretical Politics*, 19(4):391–424, 2007b. (page 82)

Franz Dietrich and Christian List. Strategy-proof judgment aggregation. *Economics & Philosophy*, 23(3):269–300, 2007c. (page 70, 80, 81, 82)

Franz Dietrich and Christian List. Majority voting on restricted domains. *Journal of Economic Theory*, 145(2):512–543, 2010. (page 86, 87)

Elad Dokow and Ron Holzman. Aggregation of binary evaluations. *Journal of Economic Theory*, 145(2):495–511, 2010. (page 70, 81, 83)

Conal Duddy and Ashley Piggins. A measure of distance between judgment sets. *Social Choice and Welfare*, 39(4):855–867, 2012. (page 88)

John Duggan and Thomas Schwartz. Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare*, 17(1):85–93, 2000. (page 12, 19, 20)

Edith Elkind and Martin Lackner. Structure in dichotomous preferences. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*, 2015. (page 45, 55, 56)

Ulle Endriss. Judgment aggregation. In Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors, *Handbook of Computational Social Choice*, chapter 17. Cambridge University Press, 2016. (page 83)

Ulle Endriss, Umberto Grandi, and Daniele Porello. Complexity of judgment aggregation. *Journal of Artificial Intelligence Research*, 45:481–514, 2012. (page 82)

Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. Counting votes for aggregating judgments. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1177–1184, 2014. (page 76)

Patricia Everaere, Sébastien Konieczny, and Pierre Marquis. An introduction to belief merging and its links with judgment aggregation. In Ulle Endriss, editor, *Trends in Computational Social Choice*, chapter 7, pages 123–143. AI Access, 2017. (page 69)

Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. In Ulle Endriss, editor, *Trends in Computational Social Choice*, chapter 2, pages 27–47. AI Access, 2017. (page 43)

Piotr Faliszewski, Arkadii Slinko, and Nimrod Talmon. Multiwinner rules with variable number of winners. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI)*, 2020. (page 43)

Peter C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, 3(1):18–40, 1972. (page 12, 14)

Peter C. Fishburn. Condorcet social choice functions. *SIAM Journal on applied Mathematics*, 33(3):469–489, 1977. (page 24)

Peter C. Fishburn. Comment on the Kannai-Peleg impossibility theorem for extending orders. *Journal of Economic Theory*, 32(1):176–179, 1984a. (page 12)

Peter C. Fishburn. Probabilistic social choice based on simple voting comparisons. *The Review of Economic Studies*, 51(4):683–692, 1984b. (page 21)

Peter Gärdenfors. Manipulation of social choice functions. *Journal of Economic Theory*, 13(2):217–228, 1976. (page 12, 14, 20)

Grzegorz Gawron and Piotr Faliszewski. Robustness of approval-based multiwinner voting rules. In *Proceedings of the 6th International Conference on Algorithmic Decision Theory (ADT)*, 2019. (page 52)

William V. Gehrlein. *Condorcet's Paradox*. Springer, 2006. (page 6, 20)

William V. Gehrlein and Dominique Lepelley. *Voting Paradoxes and Group Coherence*. Springer, 2011. (page 20)

Christian Geist and Ulle Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011. (page 12)

Allan Gibbard. Manipulation of voting schemes: A general result. *Econometrica*, 41(4):587, 1973. (page 1, 12, 19)

Davide Grossi and Gabriella Pigozzi. *Judgment Aggregation: A Primer*. Morgan & Claypool Publishers, 2014. (page 69)

Ronald de Haan. Complexity results for manipulation, bribery and control of the Kemeny judgment aggregation procedure. In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2017. (page 82)

Lê Nguyên Hoang. Strategy-proofness of the randomized Condorcet voting system. *Social Choice and Welfare*, 48(3):679–701, 2017. (page 21)

Olivier Hudry. A note on "Banks winners in tournaments are difficult to recognize" by G.J. Woeginger. *Social Choice and Welfare*, 23(1):113–114, 2004. (page 40)

Aanund Hylland. Proportional representation without party lists. In Raino Malnes and Arild Underdal, editors, *In Rationality and Institutions : Essays in Honour of Knut Midgaard on the Occasion of his 60th Birthday*, pages 126–153. Universitetsforlaget, 1992. (page 43)

Svante Janson. Phragmén's and Thiele's election methods. Technical Report arXiv:1611.08826, 2016. (page 44, 48)

Yakar Kannai and Bezalel Peleg. A note on the extension of an order on a set to the power set. *Journal of Economic Theory*, 32(1):172–175, 1984. (page 12)

Jerry S. Kelly. Strategy-proofness and social choice functions without singlevaluedness. *Econometrica*, pages 439–446, 1977. (page 12, 14, 38)

John G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959. (page 27, 75)

D. Marc Kilgour. Approval balloting for multi-winner elections. In *Handbook on Approval Voting*, pages 105–124. Springer, 2010. (page 43)

D. Marc Kilgour. Approval elections with a variable number of winners. *Theory and Decision*, 81(2):199–211, 2016. (page 43)

Boas Kluiving, Adriaan de Vries, Pepijn Vrijbergen, Arthur Boixel, and Ulle Endriss. Analysing irresolute multiwinner voting rules with approval ballots via SAT solving. In *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI)*, 2020. (page 43, 53, 54)

Lewis A. Kornhauser and Lawrence G. Sager. The one and the many: Adjudication in collegial courts. *California Law Review*, 81(1):1–59, 1993. (page 69)

Justin Kruger and Zoi Terzopoulou. Strategic manipulation with incomplete preferences: Possibilities and impossibilities for positional scoring rules. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2020. (page 14, 20)

Martin Lackner and Piotr Skowron. Approval-based multi-winner rules and strategic voting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018. (page 51)

Jérôme Lang, Gabriella Pigozzi, Marija Slavkovik, Leendert van der Torre, and Srdjan Vesic. A partial taxonomy of judgment aggregation rules and their properties. *Social Choice and Welfare*, 48(2):327–356, 2017. (page 74, 77, 84, 93, 95)

Jean-François Laslier and Karine Van der Straeten. Strategic voting in multi-winner elections with approval balloting: a theory for large electorates. *Social Choice and Welfare*, 47(3):559–587, 2016. (page 51)

Christian List. A possibility theorem on aggregation over multiple interconnected propositions. *Mathematical Social Sciences*, 45:1–13, 2003. (page 82, 85, 86, 87, 92)

Christian List and Philip Pettit. Aggregating sets of judgments: An impossibility result. *Economics and Philosophy*, 18(1):89–110, 2002. (page 69, 70, 81)

Jan Maly, Miroslaw Truszczyński, and Stefan Woltran. Preference orders on families of sets—when can impossibility results be avoided? *Journal of Artificial Intelligence Research*, 66:1147–1197, 2019. (page 12)

David C. McGarvey. A theorem on the construction of voting paradoxes. *Econometrica*, 21(4):608–610, 1953. (page 34)

Iain McLean and Arnold Urken, editors. *Classics of Social Choice*. University of Michigan Press, 1995. (page 1)

Reshef Meir, Ariel D. Procaccia, Jeffrey S. Rosenschein, and Aviv Zohar. Complexity of strategic behavior in multi-winner elections. *Journal of Artificial Intelligence Research*, 33:149–178, 2008. (page 51)

Michael K. Miller and Daniel Osherson. Methods for distance-based judgment aggregation. *Social Choice and Welfare*, 32(4):575–601, 2009. (page 74, 75, 77)

Neeldhara Misra and Chinmay Sonar. Robustness radius for Chamberlin-Courant on restricted domains. In *Proceedings of the International Conference on Current Trends in Theory and Practice of Informatics (SOFSEM)*, 2019. (page 52)

Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35 (4):437–455, 1980. (page 4)

Klaus Nehring and Marcus Pivato. Majority rule in the absence of a majority. *Journal of Economic Theory*, 183:213–257, 2019. (page 76, 78)

Klaus Nehring and Clemens Puppe. The structure of strategy-proof social choice—Part I: General characterization and possibility results on median spaces. *Journal of Economic Theory*, 135(1):269–305, 2007. (page 70, 81)

Klaus Nehring, Marcus Pivato, and Clemens Puppe. The Condorcet set: Majority voting over interconnected propositions. *Journal of Economic Theory*, 151:268–303, 2014. (page 74, 75)

Svetlana Obraztsova, Yair Zick, and Edith Elkind. On manipulation in multi-winner elections based on scoring rules. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, 2013. (page 51)

Martin J. Osborne and Ariel Rubinstein. Sampling equilibrium, with an application to strategic voting. *Games and Economic Behavior*, 45(2):434–441, 2003. (page 21)

Prasanta K. Pattanaik and Bezalel Peleg. An axiomatic characterization of the lexicographic maximin extension of an ordering over a set to the power set. *Social Choice and Welfare*, 1(2):113–122, 1984. (page 12)

Dominik Peters. Proportionality and strategyproofness in multiwinner elections. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS)*, pages 1549–1557, 2018. (page 43, 51, 53, 54, 56)

Gabriella Pigozzi. Belief merging and the discursive dilemma: An argument-based account to paradoxes of judgment aggregation. *Synthese*, 152(2):285–298, 2006. (page 75)

Michel Regenwetter, Bernard Grofman, Ilia M. Tsetlin, and A.A.J. Marley. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press, 2006. (page 6)

Annemieke Reijngoud and Ulle Endriss. Voter response to iterated poll information. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 635–644, 2012. (page 21)

Alvin E. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982. (page 102)

Shin Sato. A sufficient condition for the equivalence of strategy-proofness and nonmanipulability by preferences adjacent to the sincere one. *Journal of Economic Theory*, 148(1):259–278, 2013. (page 20)

Shin Sato. Bounded response and the equivalence of nonmanipulability and independence of irrelevant alternatives. *Social Choice and Welfare*, 44:133–149, 2015. (page 82)

Mark Allen Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975. (page 1, 12, 19)

Markus Schulze. Free riding. *Voting Matters*, 18, 2004. (page 43)

Patrick Slater. Inconsistencies in a schedule of paired comparisons. *Biometrika*, 48(3–4):303–312, 1961. (page 74)

Zoi Terzopoulou and Ulle Endriss. Strategyproof judgment aggregation under partial information. *Social Choice and Welfare*, 53(3):415–442, 2019. (page 82)

Thorvald Nicolai Thiele. Om flerfoldsvalg. *Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger*, page 415–441, 1895. (page 44, 48)

Thorwald Nicolaus Tideman. Independence of clones as a criterion for voting rules. *Social Choice and Welfare*, 4(3):185–206, 1987. (page 28, 76)

Yongjie Yang and Jianxin Wang. Multiwinner voting with restricted admissible sets: Complexity and strategyproofness. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, 2018. (page 51)

Index

Abstract

This thesis examines strategic manipulation in three areas of social choice theory—single-winner voting, multiwinner voting, and judgment aggregation. It is widely accepted that *strategyproofness* often does not play nice with other axioms. While we would like our aggregation methods to be *strategyproof*—meaning no agent has an incentive to misreport her preferences or opinions—strategic manipulation is difficult to avoid, no matter what specific framework we consider.

A well-known and often used approach is to consider only specific types of input to the aggregation method—so-called *restricted domains*. Our approach here is to consider manipulation *on* profiles that fall within certain restricted domains where existing results tell us manipulation within the domain is not possible. In general we ask whether agents can manipulate *from* a profile in a "well-behaved" domain to one outside the domain in question.

By showing that an aggregation method is strategyproof in this sense, we show that allowing all inputs will not create unnecessary possibilities for manipulation. Thus, our work is an argument against restricting the domain of the aggregation method. We also aim to understand how strategyproofness on these domains can interact with the axiomatic properties of our aggregation methods.

Chapters 3, 4, and 5 each focus on this larger question in a different framework within the area of social choice. In each chapter we focus on a particular domain, and a particular class of aggregation methods. Chapter 3 looks at (single-winner) voting, focusing on the domain of profiles with a *Condorcet winner*. Chapter 4 considers approval-based multiwinner voting, where we study strategyproofness on *party-list profiles*, where any two approval sets either coincide or are disjoint. Chapter 5 discusses strategyproofness of *majoritarian* judgment aggregation rules on profiles with a consistent majority.

Samenvatting

In deze dissertatie onderzoeken we strategische manipulatie in drie gebieden van de socialekeuzetheorie: verkiezingen met één winnaar (*single-winner voting*), verkiezingen met meerdere winnaars (*multiwinner voting*), en het aggregeren van oordelen (*judgment aggregation*). Het is breed geaccepteerd dat strategiebestendigheid vaak niet goed samen gaat met andere axioma's. Hoewel we graag zouden willen dat aggregatiemethoden strategiebestendig zijn—wat betekent dat geen van de individuele agents een belang heeft bij het geven van een oneerlijke weergave van haar voorkeuren of meningen—is strategische manipulatie lastig om te voorkomen, ongeacht welk specifiek raamwerk we beschouwen.

Een welbekende en vaak gebruikte aanpak is om alleen bepaalde soorten invoer voor de aggregatiemethoden te beschouwen—zogenaamde *beperkte domeinen*. Onze aanpak in deze dissertatie is om manipulatie *op* profielen te beschouwen die in bepaalde van zulke beperkte domeinen vallen waarvoor bestaande resultaten laten zien dat manipulatie binnen deze domeinen niet mogelijk is. In het algemeen stellen we de vraag of agents kunnen manipuleren *vanaf* een profiel in een "zich goedgedragend" domein naar een profiel buiten het domein in kwestie.

Door te laten zien dat een aggregatiemethode strategiebestendig is in deze betekenis laten we zien dat het toestaan van alle mogelijke invoer geen onnodige mogelijkheden voor manipulatie creëert. Daarmee vormt ons werk een argument tegen het beperken van het domein voor aggregatiemethoden. We streven er ook naar om te begrijpen hoe strategiebestendigheid op deze domeinen en andere axiomatische eigenschappen van aggregatiemethoden op elkaar inwerken.

Hoofdstukken 3, 4 en 5 richten zich ieder op deze grotere vraag binnen een ander raamwerk in het gebied van de socialekeuzetheorie. In elk hoofdstuk richten we ons op een bepaald domein en een bepaalde klasse van aggregatiemethoden. In Hoofdstuk 3 bekijken we verkiezingen (met één winnaar) en richten we ons

op het domein met profielen die een *Condorcetwinnaar* hebben. In Hoofdstuk 4 beschouwen we op goedkeuring gebaseerde verkiezingen met meerdere winnaars en bestuderen we strategiebestendigheid op *partijlijstprofielen*, waar elke twee verzamelingen van goedkeuringen ofwel samenvallen, ofwel een lege doorsnede hebben. In Hoofdstuk 5, ten slotte, behandelen we strategiebestendigheid van *op meerderheid gerichte* oordeelaggregatiemethoden op profielen met een consistent meerderheidsoordeel.

Titles in the ILLC Dissertation Series:

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zoé Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong Lê**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory

ILLC DS-2017-04: **Giovanni Cinà**
Categories for the working modal logician

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
Communication and Computation: New Questions About Compositionality

ILLC DS-2017-06: **Peter Hawke**
The Problem of Epistemic Relevance

ILLC DS-2017-07: **Aybüke Özgün**
Evidence in Epistemic Logic: A Topological Perspective

ILLC DS-2017-08: **Raquel Garrido Alhama**
Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence

ILLC DS-2017-09: **Miloš Stanojević**
Permutation Forests for Modeling Word Order in Machine Translation

ILLC DS-2018-01: **Berit Janssen**
Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs

ILLC DS-2018-02: **Hugo Huurdeman**
Supporting the Complex Dynamics of the Information Seeking Process

ILLC DS-2018-03: **Corina Koolen**
Reading beyond the female: The relationship between perception of author gender and literary quality

ILLC DS-2018-04: **Jelle Bruineberg**
Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems

ILLC DS-2018-05: **Joachim Daiber**
Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation

ILLC DS-2018-06: **Thomas Brochhagen**
Signaling under Uncertainty

ILLC DS-2018-07: **Julian Schlöder**
Assertion and Rejection

ILLC DS-2018-08: **Srinivasan Arunachalam**
Quantum Algorithms and Learning Theory

ILLC DS-2018-09: **Hugo de Holanda Cunha Nobrega**
Games for functions: Baire classes, Weihrauch degrees, transfinite computations, and ranks

ILLC DS-2018-10: **Chenwei Shi**
Reason to Believe

ILLC DS-2018-11: **Malvin Gattinger**
New Directions in Model Checking Dynamic Epistemic Logic

ILLC DS-2018-12: **Julia Ilin**
Filtration Revisited: Lattices of Stable Non-Classical Logics

ILLC DS-2018-13: **Jeroen Zuiddam**
Algebraic complexity, asymptotic spectra and entanglement polytopes

ILLC DS-2019-01: **Carlos Vaquero**
What Makes A Performer Unique? Idiosyncrasies and commonalities in expressive music performance

ILLC DS-2019-02: **Jort Bergfeld**
Quantum logics for expressing and proving the correctness of quantum programs

ILLC DS-2019-03: **András Gilyén**
Quantum Singular Value Transformation & Its Algorithmic Applications

ILLC DS-2019-04: **Lorenzo Galeotti**
The theory of the generalised real numbers and other topics in logic

ILLC DS-2019-05: **Nadine Theiler**
Taking a unified perspective: Resolutions and highlighting in the semantics of attitudes and particles

ILLC DS-2019-06: **Peter T.S. van der Gulik**
Considerations in Evolutionary Biochemistry

ILLC DS-2019-07: **Frederik Möllerström Lauridsen**
Cuts and Completions: Algebraic aspects of structural proof theory

ILLC DS-2020-01: **Mostafa Dehghani**
Learning with Imperfect Supervision for Language Understanding

ILLC DS-2020-02: **Koen Groenland**
Quantum protocols for few-qubit devices

ILLC DS-2020-03: **Jouke Witteveen**
Parameterized Analysis of Complexity

ILLC DS-2020-04: **Joran van Apeldoorn**
A Quantum View on Convex Optimization

ILLC DS-2020-05: **Tom Bannink**
Quantum and stochastic processes

ILLC DS-2020-06: **Dieuwke Hupkes**
Hierarchy and interpretability in neural models of language processing

ILLC DS-2020-07: **Ana Lucia Vargas Sandoval**
On the Path to the Truth: Logical & Computational Aspects of Learning

ILLC DS-2020-08: **Philip Schulz**
Latent Variable Models for Machine Translation and How to Learn Them

ILLC DS-2020-09: **Jasmijn Bastings**
A Tale of Two Sequences: Interpretable and Linguistically-Informed Deep Learning for Natural Language Processing

ILLC DS-2020-10: **Arnold Kochari**
Perceiving and communicating magnitudes: Behavioral and electrophysiological studies

ILLC DS-2020-11: **Marco Del Tredici**
Linguistic Variation in Online Communities: A Computational Perspective

ILLC DS-2020-12: **Bastiaan van der Weij**
Experienced listeners: Modeling the influence of long-term musical exposure on rhythm perception

ILLC DS-2020-13: **Thom van Gessel**
Questions in Context

ILLC DS-2020-14: **Gianluca Grilletti**
Questions & Quantification: A study of first order inquisitive logic

ILLC DS-2020-15: **Tom Schoonen**
Tales of Similarity and Imagination. A modest epistemology of possibility

ILLC DS-2020-16: **Ilaria Canavotto**
Where Responsibility Takes You: Logics of Agency, Counterfactuals and Norms

ILLC DS-2020-17: **Francesca Zaffora Blando**
Patterns and Probabilities: A Study in Algorithmic Randomness and Computable Learning

ILLC DS-2021-01: **Yfke Dulek**
Delegated and Distributed Quantum Computation

ILLC DS-2021-02: **Elbert J. Booij**
The Things Before Us: On What it Is to Be an Object

ILLC DS-2021-03: **Seyyed Hadi Hashemi**
Modeling Users Interacting with Smart Devices

ILLC DS-2021-04: **Sophie Arnoult**
Adjunction in Hierarchical Phrase-Based Translation

ILLC DS-2021-05: **Cian Guilfoyle Chartier**
A Pragmatic Defense of Logical Pluralism

ILLC DS-2021-06: **Zoi Terzopoulou**
Collective Decisions with Incomplete Individual Opinions

ILLC DS-2021-07: **Anthia Solaki**
Logical Models for Bounded Reasoners

ILLC DS-2021-08: **Michael Sejr Schlichtkrull**
Incorporating Structure into Neural Models for Language Processing

ILLC DS-2021-09: **Taichi Uemura**
Abstract and Concrete Type Theories

ILLC DS-2021-10: **Levin Hornischer**
Dynamical Systems via Domains: Toward a Unified Foundation of Symbolic and Non-symbolic Computation