Voting by Axioms

MSc Thesis *(Afstudeerscriptie)*

written by

Marie Christin Schmidtlein
(born 24th May, 1998 in St. Wendel, Germany)

under the supervision of **Prof. Dr. Ulle Endriss**, and submitted to the Examinations Board in partial fulfillment of the requirements for the degree of

MSc in Logic

at the *Universiteit van Amsterdam.*

Date of the public defense:	**Members of the Thesis Committee:**
28th June, 2022	Dr. Benno van den Berg *(chair)*
	Arthur Boixel, MSc
	Prof. Dr. Ulle Endriss
	Dr. Davide Grossi

Institute for Logic, Language and Computation

Abstract

How should we decide on the outcome of an election? Social choice theory offers many voting rules to answer this question, but also establishes various impossibility results showing that no single ideal rule exists. During recent years, researchers have developed a method for using axioms, i.e., desirable properties for voting rules, to justify or explain why a certain voting outcome is appropriate in a given scenario. This can be used to argue in favor of or against a specific voting rule's behavior and shows that axioms can prescribe which outcome is to be returned in a certain situation. This thesis is dedicated to developing a novel decision procedure, Voting by Axioms, that takes decisions purely based on preselected axioms. That is, in each scenario, the voting rule returns the outcome that is justified or forced by the axioms that we care most about. The construction process sheds light on what axioms are and how to formalize them and we include a thorough analysis of the voting rule as well as an evaluation and possible generalizations of the framework in this thesis.

Acknowledgments

Special thanks go to my supervisor Ulle Endriss. You did not hesitate for a second when I approached you about writing a thesis under your supervision. Instead, you believed in my skills from the beginning, you let me choose the topic that seemed most interesting to me, and you supported me until the defense and beyond. I could not have asked for a better supervision: You gave me freedom to explore my own ideas, you showed interest in the project and encouraged me to keep going, you met me at eye-level which allowed us to have insightful involved technical and philosophical discussions, you were always available for consultations and at all times provided prompt, valuable feedback. In this way, you guided me throughout the whole process, which turned out to be really enjoyable and stimulating, and enabled me to write an interesting, many-faceted thesis. For anyone considering to work with Ulle, take this as a 5-star review!

I am grateful for everyone else who accompanied me throughout this Master's program — teachers, committee members, mentors, fellow students. Thank you for all the interesting and thought-provoking discussions, for sharing your knowledge and expertise, for showing me the world of logic.

Lastly, I thank all my loved ones — family, friends, flatmates, companions — *Herzensmenschen*. For always being by my side, for believing in me and enduring my frustrations, for listening to my worries and sharing your own, for making me laugh, showing your appreciation and for celebrating with me. I draw strength and take heart from your affection and could never have done it without!

Contents

Mathematical Definitions and Notational Conventions

We will use the following basic mathematical concepts and notational conventions throughout this thesis.

Relations and Orders

- A *binary relation* R over a set S is a set $R \subseteq S \times S$. We write xRy iff $(x, y) \in R$.
- A binary relation is *reflexive* if it satisfies for all $x \in S$ that xRx holds.
- A binary relation R is *transitive* in case for all $x, y, z \in S$, if xRy and yRz holds, then also xRz is true.
- A *(weakly) connected* binary relation is such that for any two distinct elements $x \neq y$ in S, we have xRy or yRx.
- Call a binary relation R *strongly connected* or *total* if any two elements in S (not necessarily distinct) are comparable, i.e., for all $x, y \in S$ we have xRy or yRx.
- A binary relation R is called *irreflexive* in case for any $x \in S$, we do not have xRx.
- A binary relation R is *anti-symmetric* in case for any two elements $x, y \in S$, if xRy and yRx is the case, then $x = y$ holds.
- A *(strict linear) order* or *ranking* is a binary relation that is irreflexive, transitive and connected.
- We call a binary relation that is reflexive and transitive a *(weak) preorder*. An irreflexive and transitive relation is called a *strict preorder*.
- An order is called *partial* if we do not assume that it is connected, i.e., there could be elements x, y that are *incomparable*, so neither xRy nor yRx holds. That is, it is an irreflexive, transitive binary relation.

- An order is called *weak* if we do not assume that it is irreflexive and if it is strongly connected, i.e., there could be elements x, y that we are *indifferent* about, so both xRy and yRx holds. This means, it is a transitive, strongly connected relation or, equivalently, a total preorder.
- The set of all possible (strict) rankings R over a set S is

$$\mathcal{L}(S) := \{R \subseteq S \times S \mid R \text{ is irreflexive, transitive and connected}\}.$$

- For a set S, we denote its maximal elements with regards to a given (weak or strict) preorder R by

$$\max_R(S) := \{x \in S \mid \text{for all } y \in S, \text{ if } yRx, \text{ then } xRy\}$$

  If R is irreflexive and connected, i.e., a strict ranking, and S is a finite set, then this set is always a singleton. By abuse of notation, we denote the unique maximal element by $\max_R(S)$ in this case.

Sets and Functions

- Let $\mathcal{P}(S)$ denote the power set of a set S, i.e., the set of all possible subsets of S. We denote the power set with the empty set removed, i.e., the set of all non-empty subsets of S, by $\mathcal{P}_+(S) := \mathcal{P}(S) \setminus \{\emptyset\}$.
- The set union of a collection of sets $\mathcal{S}$ is defined as $\bigcup \mathcal{S} := \{x \in S \mid S \in \mathcal{S}\}$.
- For two sets S, S', write $S \sqcup S'$ for their disjoint union, i.e., write this instead of $S \cup S'$ in case $S \cap S' = \emptyset$ holds.
- We denote the set of functions from a set S to a set S' by $S \to S'$.
- Notice that we can view a function $F : S \to S'$ as a set of pairs

$$\{(x, F(x)) \mid x \in S\} \subseteq S \times S'.$$

  This allows us, for two functions $F_1 : S_1 \to S_1'$ and $F_2 : S_2 \to S_2'$, with disjoint domains S_1, S_2, to define their union $F_1 \sqcup F_2$, a function defined on the union of the two domains $S_1 \sqcup S_2$, as the set

$$\{(x, F_1(x)) \mid x \in S_1\} \sqcup \{(y, F_2(y)) \mid y \in S_2\}.$$

- For two sets of functions $\mathcal{F}_1, \mathcal{F}_2$ defined on disjoint domains, i.e., $\mathcal{F}_1 \subseteq S_1 \to S_1'$ and $\mathcal{F}_2 \subseteq S_2 \to S_2'$ with $S_1 \cap S_2 = \emptyset$, define their product $\mathcal{F}_1 \otimes \mathcal{F}_2$ as the set of functions $\{F_1 \sqcup F_2 : (S_1 \sqcup S_2) \to (S_1' \cup S_2') \mid (F_1, F_2) \in \mathcal{F}_1 \times \mathcal{F}_2\}$.
- We can restrict a function $F : S_1 \to S'$ to a subdomain $S_2 \subseteq S_1$ to obtain a function $F{\restriction_{S_2}} : S_2 \to S'$ given by $\{(x, F(x)) \mid x \in S_2\}$.

Chapter 1

Introduction

How does one take decisions as a group? Electing a president, settling on a restaurant to go to with friends, selecting games to be played at a party — there are plenty of situations in which a collective decision has to be made. But how can we do this in a sensible way such that the process is fair and takes the individual preferences into account? This is the central question in social choice theory. Many voting rules, ways to decide what the outcome should be, given the voter's preferences, have been suggested in this research field. Think of the *Plurality*, *Antiplurality*, *Borda* or *Copeland* rules (Zwicker, 2016). But which one should we use? Social choice scientists answer this question by proposing a multitude of so-called *axioms*, desirable properties that a decision procedure should possess, as a means of assessing and comparing voting rules. These axioms capture characteristics of a favorable voting rule, so the more of them are satisfied by a given rule, the better. So then, we should just choose the voting rule that fulfills all or at least most of these principles. Unfortunately, social choice research has found various impossibility results stating that many of the suggested axioms are incompatible, i.e., not satisfiable together. A famous result by Gibbard (1973) and Satterthwaite (1975), for example, shows that if there are more than two alternatives, every rule is either manipulable or a dictatorship. This means that there exists no single best voting rule that has all properties that we want a decision procedure to exhibit. This is part of the reason why so many different voting rules exist and are applied in practice.

Thus, to take a collective decision, one seemingly acceptable voting rule needs to be chosen. This choice might appear arbitrary and policymakers should be able to explain to the electorate why the picked rule takes good decisions. Ideally, they would use axioms, norms that are intuitive and socially accepted, to justify the rule's behavior. Rather than stating which (few) properties the voting rule satisfies overall, one should be able to explain in every given situation, why it is reasonable to go along with the rule's outcome. Boixel and Endriss (2020) enable exactly this with their theory about how axioms can justify assigning a certain outcome in a given situation. More precisely, they say that some given axioms justify an outcome in case assigning the outcome is a necessary

consequence of satisfying the axioms. This means that if we want a rule with the given property, it must return this specific outcome in the given situation, since, otherwise, it would violate the axioms. In this way, axioms can prescribe or force a certain behavior of a rule. Thus, instead of picking some voting rule and using axioms to justify its behavior, as previously suggested, could we not decide on axioms that we are most interested in first, and determine the outcome that is forced by the axioms for each scenario directly? This is the objective of this thesis. We want to use the property of axioms to prescribe a rule's behavior to construct a voting rule whose outcomes are fully determined and justified by underlying axioms. This means, we first settle on axioms or axiom sets that we care most about, and then define a procedure that, for each possible situation independently, searches for an outcome justified or forced by the given axioms. We call this *Voting by Axioms*. The idea for this kind of voting rule originates from Boixel and Endriss (2020, Example 3) and will be made precise over the course of this thesis. Besides motivating and defining the framework and the voting rule, we are going to analyze them and suggest possible generalizations.

Why is this interesting to look at? In this thesis, we are going to focus on a voting rule that is grounded in axioms. More precisely, Voting by Axioms is a procedure such that for each possible voting scenario, its outcome is justified by axioms that we care about. In this respect, the resulting rule is superior to standard voting rules, for which we have to search for explanations afterwards and might not even find justifications in terms of our most liked axioms. In Voting by Axioms, however, we can control which axioms the rule should be governed by. It is a way of systematizing the justification of outcomes since Boixel and Endriss (2020) only look at specific situations whereas the Voting by Axioms rule is a determined procedure defined for all possible situations.

Another advantage of the voting rule that we are going to construct is that it can be seen as a way of avoiding impossibility results. Generally, these theorems have a bad connotation since they ruin our hope that a distinguished voting rule with many good properties can exist. They state that certain axioms together are unsatisfiable, which usually results in discarding or weakening one or multiple of the axioms. In Voting by Axioms, we generally specify multiple axioms or axiom sets with a prioritization among them that will be used to define the voting rule. Therefore, if an axiom set is unsatisfiable, we can instead consider all its satisfiable subsets and thereby still include all axioms in our procedure (assuming that all axioms in the set are individually satisfiable). Thus, although the Voting by Axioms rule cannot satisfy all axioms occurring in an impossibility result, it can still respect them or factor all of them in.

Further, we can view the Voting by Axioms rule as a method for automatically finding the best justification in each situation. Instead of searching, for a given scenario and specified axioms, for all possible justifications and then manually choosing the best one, we can start by listing all axiom sets that we are interested in and rank them by appeal, and the rule that we are going to construct will always choose the justification in terms of the best possible axiom

set. For instance, we could say that every axiom used in a justification comes with a certain cost, e.g., the cognitive capacity that it takes the listener to comprehend the axiom. A natural objective would be to find the justification with the lowest cost, i.e., which is easiest to deliver to the audience.

Besides the perks that the Voting by Axioms rule itself provides, the construction of the rule and the model that it lives in give valuable insights into what axioms are and what we demand from a good decision procedure for voting. First, in order to include axioms as formal objects into the model, we need to understand their role and significance. Part of this analysis is discussing multiple ways of classifying axioms, i.e., putting them into categories according to their structure, function or complexity. Usually, we view axioms as one condition that a voting rule either satisfies or does not satisfy. In this thesis, we change perspective and rather take them to be a multitude of conditions, so-called *instances*. Are we only content if we can satisfy all instances of an axiom or does it suffice if our rule respects most conditions that the axiom imposes? For instance, for the Condorcet principle (see Table 2.1), it might suffice if in the extreme cases, where the vast majority of voters prefers one alternative in pairwise comparison to every other alternative, this alternative wins solely. However, in cases where the pairwise contest is close to being tied for most alternatives, we have less grounds to impose the Condorcet principle and, therefore, might accept a breaking of the principle. This yields a more nuanced view on the satisfiability of axioms.

Even further, we are required to rank and compare axioms, to prioritize with regards to their normative appeal or cognitive demand. Undergoing this process will require policymakers to deal with the characteristics of decision procedures and help them becoming aware of the values they want to protect. For instance, they need to decide whether to care more about protecting the outcome from being manipulated or about every voter having the same influence. And is it worth imposing complex principles that the average voter cannot even wrap their head around? In this way, the thoughts that go into constructing Voting by Axioms can shape the societal discussion about voting.

Moreover, we are going to showcase methods that help determining whether sufficiently many axioms are included in our corpus to justify an outcome in every possible voting scenario. Thus, Voting by Axioms assists in extracting problematic situations in which it is hard to decide on an outcome. We can then look at these scenarios and think about how we would intuitively decide and whether this follows an overarching principle that we want to require. This can motivate new definitions of axioms.

And, lastly, the notion of justifying or forcing an outcome yields a new way of assessing the logical strength or complexity of an axiom. Namely, we can check for an axiom, in how many scenarios it prescribes an outcome, or to what extent it predetermines the behavior of a voting rule.

Contribution and Related Work. We are going to define a novel procedure on how to decide a vote by specifying governing principles first in this thesis. For

this, we thoroughly discuss what axioms are and how to formalize them. In the course of this, we come up with methods for making sense of an axiom when we only know which rules satisfy the axiom, not which thought the axiom expresses. In particular, we define procedures for determining which scenarios the axiom speaks about and for extracting all the conditions that an axiom imposes, and thereby obtain a formulation of the axiom in some language. Based on this, we define a language-independent hierarchy or classification of axioms. Then, we carry out a complete social choice theoretic analysis of the defined Voting by Axioms rule, including an axiomatic and a complexity analysis. We also suggest metrics for analyzing voting rules and axiom sets, and show how the developed framework can be generalized.

As previously mentioned, our work is built upon the theory of axiomatic justification of outcomes in voting by Boixel and Endriss (2020). The idea to use axioms to argue about or explain the behavior of voting rules originates from Cailloux and Endriss (2016) who showed how to justify the outcomes of the Borda rule with axioms. Boixel and Endriss (2020) generalized this and came up with a formal description of justifications (featuring a normative and an explanatory component) together with a computation method using constraint programming for finding such explanations. In further work, the complexity of this approach was analyzed (Boixel & de Haan, 2021; Peters, Procaccia, Psomas, & Zhou, 2020) and more feasible methods for finding justifications in terms of axioms were developed (Boixel, Endriss, & de Haan, 2022; Nardi, Boixel, & Endriss, 2022). Further, a paper by Suryanarayana, Sarne, and Kraus (2022) explores in an experiment, what method of justifying social choice mechanism outcomes is most effective in practice. All this ties in with the growing need for methods yielding explanations for the behavior or the decisions of intelligent systems. Explicability is seen as a key factor in building trustworthy systems (EU High-Level Expert Group on AI, 2019), which is why a whole field of explainable AI (XAI) emerged (Adadi & Berrada, 2018).

Outline. The second chapter of this thesis focuses on defining the framework and the Voting by Axioms rule. This includes a presentation of a standard model for voting theory, a discussion of axioms, debating how to formalize them, what to do when we only know an axiom's extension (i.e., the set of rules that satisfy the axiom) and how to divide them into smaller axioms, called axiom instances, and ends with a formal definition of the Voting by Axioms rule. In Chapter 3, we first present criteria for when and ways to find out whether the defined rule is indeed a well-defined voting rule. Afterwards, we work out conditions for when the voting rule satisfies some of the underlying axioms and we define a function measuring to what extent the axioms are satisfied. The last aspect of the analysis focuses on computational complexity of the defined procedure. Chapter 4 explores possible extensions or generalizations of the defined model, for instance allowing for a weak or partial order over the underlying axiom sets or deriving an order over them from an order over the axioms themselves. The

conclusion in Chapter 5 summarizes and evaluates the introduced approach and states future research questions.

Chapter 2

Constructing the Framework

In this chapter, we are going to make the idea of voting based on a selection of axioms precise. In the field of logic, we often aim at formally representing an existing phenomenon. Whether it is getting to the bottom of what knowledge and belief are to define epistemic logic, or solving philosophical paradoxes such as the Lottery Paradox (Kyburg Jr, 1961) with the help of formal descriptions. This process of going from an intuitive idea to a suitable formal model is involved and a key contribution of this thesis. In the first step, it requires us to thoroughly analyze the given phenomenon, to identify the key factors and properties that a model should reproduce. Next, creativity and out-of-the-box thinking are needed to come up with formalisms that have one or multiple of the desired characteristics. As the third step, the proposed models have to be compared to each other and pitted against the identified requirements. This process is rarely linear and consists of multiple iterations of the aforementioned, weaving in insights gathered along the way. In the case of Voting by Axioms, this chiefly includes stating precisely what axioms are and then constructing a voting rule within the standard setup for voting that does justice to the idea of being based on given axioms.

This chapter introduces the standard framework for voting in social choice theory, discusses the question what axioms and axioms instances are and how to represent them, and presents the constructed rule for Voting by Axioms.

2.1 Voting Theory

What is studied in voting theory is a decision making process by a group of agents among multiple alternatives. Voting theory is an integral part of social choice theory and has widely been studied over the past centuries (Marquis de Condorcet, 1785; Arrow, 1951; Arrow, Sen, and Suzumura, 2002, 2011). We want to present a commonly used framework for voting which features ranked

ballots to capture the voters' declared preferences and which determines one or multiple winning alternatives. In this thesis, we adopt a variation of the model used by Boixel and Endriss (2020). A thorough introduction to the history, questions and methods of voting theory was written by Zwicker (2016).

In voting, a group of *voters* provide their preferences over various *alternatives* to elect a subset of them. We denote the finite set of all voters (the *universe*) by $N^* := \{1, \dots, n\}$ and the finite set of alternatives by $X := \{1, \dots, m\}$. The format for voters to voice their opinion is by ordering the alternatives from best to worst. So every voter's preference is represented by a *ranking*, a strict linear order, over all alternatives. Therefore, a voter's *ballot* is an element of the set of all possible rankings over X, denoted by $\mathcal{L}(X)$.

To systematize voting, we need to be able to tell for each voting scenario, in which some voters in the universe cast ballots, which alternatives should win. First, we want to define such voting scenarios. For a specified *electorate* $N \subseteq N^*$, a *profile* contains one ballot for each voter in the electorate. The set of all profiles for a given electorate is $N \to \mathcal{L}(X)$ containing functions R that assign to each voter i their submitted ranking $R_i := R(i)$. We can express that in profile R, voter i prefers alternative x over alternative y by writing $x \, R_i \, y$.[1] Further, we will use a compressed notation for rankings, writing voter i's ballot as $x_1 x_2 \dots x_m$ instead of $x_1 R_i x_2 R_i \cdots R_i x_m$ for alternatives $x_i \in X$. Suppressing the electorate in the notation, we can represent profiles in a compact way as vectors of ballots, e.g.,

$$\begin{array}{l} 1 \, R_1 \, 2 \, R_1 \, 3 \\ 2 \, R_2 \, 3 \, R_2 \, 1 \\ 1 \, R_3 \, 3 \, R_3 \, 2 \end{array} \quad \approx \quad (123, 231, 132).$$

For a given profile R, we will refer to its electorate by N_R. We obtain the set of all possible profiles, i.e., for all possible electorates,

$$\mathcal{L}(X)^+ := \bigcup_{N \in \mathcal{P}_+(N^*)} (N \to \mathcal{L}(X)).$$

We want to define procedures that decide for every voting scenario what the outcome should be. A *voting rule* is a function $F : \mathcal{L}(X)^+ \to \mathcal{P}_+(X)$ that assigns an *outcome* to every possible profile. Such a rule is *irresolute* since it allows for multiple alternatives to win. Notice that we exclude the empty set from the set of possible outcomes since we want the voting rule to always elect at least one winner. We use set-theoretic notation and indicate the set of all voting rules as $\mathcal{L}(X)^+ \to \mathcal{P}_+(X)$. In the following, we give some examples of commonly used voting rules (Zwicker, 2016).

Example 1. Positional scoring rules *assign a score s_i to an alternative x whenever a voter puts x in the i-th position of their ranking. Under these rules, the*

[1] In social choice theory, some authors denote a weak preference relation by R and its strict counterpart by P. Since we only consider strict preferences here, we do not follow this convention and let R be a strict linear order.

alternatives with the highest cumulative score win. For instance, the plurality rule *assigns* 1 *point to the highest-ranked alternative and* 0 *points to all others. Another example is the* Borda rule *that awards* $m-1$ *points to the highest-ranked alternative,* $m-2$ *to the second highest, down to* 0 *points to the lowest-ranked alternative.*

Another class of rules are Condorcet extensions*, that is, rules that satisfy the Condorcet principle (see Table 2.1). For example, the* Copeland rule *computes all pairwise majority contests (i.e., for alternatives* x, y*, it compares* $|\{i \in N_R \mid x\ R_i\ y\}|$ *to* $|\{i \in N_R \mid y\ R_i\ x\}|$*) and assigns* 1 *for each win,* 0 *for each loss and some fixed number in* $[0, 1]$ *whenever a tie occurs. Then, the alternatives with the highest score win.*

In principle, one does not have to restrict oneself to finitely many voters or alternatives. There are interpretations of an infinite universe as a way of dealing with uncertainty (represented by infinitely many possible states) among a set of voters (Mihara, 1997), or as capturing the indefinite future, taking into account not only voters currently alive but also future generations (Koopmans, 1960). Similarly, one could let voters decide about the width of a street or the number of years that a law should stay in effect, offering them infinitely many alternatives to choose from. For the sake of simplicity and since we are interested in implementing our model, we are going to require, in this thesis, that all sets are finite.

2.2 Axioms and Axiom Instances

Axioms are the basic ingredient for the voting rule that we want to define. So far we described them as desirable properties that a decision procedure for voting should possess. Therefore, usually, we view axioms as principles speaking *about* the voting model that we are working with. For our undertaking, however, we need them to be formal objects *within* our framework. This section is dedicated to understanding and describing axioms. This includes presenting multiple ways to classify them, suggesting intensional and extensional definitions of axioms and showing how we can split one axiom into multiple smaller ones.

2.2.1 The Nature of Axioms

The axiomatic method is a key technique in the field of social choice theory to analyze and characterize voting rules (Plott, 1976; Thomson, 2001; Zwicker, 2016). Although they are used frequently, it is rarely scrutinized what exactly an axiom is. In short, we take it to be a normative, desirable principle that voting rules should comply with. Plott (1976) offers a more nuanced account by presenting three perspectives on the nature of axioms.

- First, recall that the role of a voting rule is to uncover the social preference, that what society wants, from the ballots. This requirement, the concept of "social preference" in itself, directs demands at what can be considered

an admissible voting rule. For instance, we expect it to choose the "best" outcome and want it to respect and be in touch with the individual preferences. In this sense, axioms are "a type of minimal expectation about system performance." (Plott, 1976, p.520).

- However, referring to major impossibility theorems such as Arrow's Theorem (Arrow, 1951), Plott notes that even those minimal requirements are often not satisfiable together. He states that "[a]lmost anything we say and/or anyone has ever said about what society wants or should get is threatened with internal inconsistency" (Plott, 1976, p.512). Accepting that there is no ideal procedure, the second take on what axioms constitute, is that they should constrain the behavior of voting rules with the aim of finding acceptable rules rather than exactly corresponding to social preference.

- We can push this idea further and view axioms as features or parts constituting a voting rule. This third perspective is tied to characterization results in social choice which express necessary and satisfactory conditions for a rule to coincide with a specific voting rule or to belong to a certain class of rules. In this sense, axioms are basic principles from which we can build up voting rules and that can be used to compare rules to each other.

In the literature, axioms are commonly taken to be intuitive, philosophically, economically or practically motivated requirements concerning the decision procedure. They help finding solutions that are desirable in the sense of mirroring how problems are solved in the real world or how they should be solved (Thomson, 2001). For instance, we might want every alternative or candidate to have a chance to win and to be treated equally. While some considered axioms are rather technical and mathematically motivated, many others stem from laws, tradition, history or common sense. Think of anonymity, which represents the principle "one person, one vote", or of unanimity, which says that if everyone most prefers the same alternative, then it should be the sole winner. Thomson (2001) extracts eight main functions that axioms perform. These include guaranteeing efficiency, symmetry, consistency, informational simplicity and implementability.

The following standard axioms serve as examples and will be used on a number of occasions throughout this thesis. We take F to be a generic voting rule and consider profiles R, R' with electorate N, unless otherwise specified.

Table 2.1: Overview of standard axioms

Anonymity (ANO)	*When renaming the agents, the outcome does not change.*
	If for some permutation $\sigma : N^* \to N^*$, we have $N_{R'} = \sigma(N_R)$ and $R'_i = R_{\sigma(i)}$ for all $i \in N_{R'}$, then $F(R') = F(R)$.
Neutrality (NEU)	*All alternatives are treated equally.*
	If for some permutation $\sigma : X \to X$, we have $R'_i = \sigma(R_i)$ for all $i \in N$,[2] then $F(R') = \sigma(F(R))$.
Pareto Principle (PAR)	*A Pareto dominated alternative (i.e, there is another alternative that everyone prefers to the given one) should not be chosen.*
	If $\{i \in N \mid x \, R_i \, y\} = N$ for some x, then $y \notin F(R)$.
Unanimity (UNA)	*If all voters rank the same alternative highest, then this should be the sole winner.*
	If there is some x^* such that $\{i \in N \mid x^* \, R_i \, y\} = N$ for all y with $y \neq x^*$, then $F(R) = \{x^*\}$.
Condorcet Principle (CON)	*If one alternative wins in a pairwise contest against all others, it should be the sole winner.*
	If for some x^*, for all $y \in X \setminus \{x^*\}$, we have $\lvert\{i \in N \mid x^* \, R_i \, y\}\rvert > \lvert N\rvert/2$, then $F(R) = \{x^*\}$.
Reinforcement (REI)	*If the outcomes of two elections with disjoint electorates intersect, then the merged election should output the intersection.*[3]
	If $N_R \cap N_{R'} = \emptyset$ and $F(R) \cap F(R') \neq \emptyset$, then $F(R \cup R') = F(R) \cap F(R')$.[4]
Cancellation (CAN)	*If all pairwise contests result in a tie, then all alternatives should win.*

[2]Here, for a permutation $\sigma : X \to X$ and an order $R_i \in \mathcal{L}(X)$, denote by $\sigma(R_i)$ the order which is given by $x \, \sigma(R_i) \, y$ iff $\sigma^{-1}(x) \, R_i \, \sigma^{-1}(y)$.

[3]Note that, in our model, reinforcement differs from its original formulation by Young (1974) since we restrict attention to a finite universe and electorates within this universe. Young's model, however, comprises infinitely many voters, and so imposes conditions on what would happen if voters outside our considered universe were to vote. Boixel and Endriss (2020) give a more detailed comparison of these versions of the axiom.

[4]Formally, we view the profile functions R and R' as sets of pairs and by taking their union, we obtain a profile given by a function $N_R \sqcup N_{R'} \to \mathcal{L}(X)$.

	If $\lvert\{i \in N \mid x\ R_i\ y\}\rvert = \lvert\{i \in N \mid y\ R_i\ x\}\rvert$ for all $x, y \in X$, then $F(R) = X$.
Faithfulness (FAI)	*If there is only one voter, then their highest-ranked alternative should be the winner.*
	If $N = \{i\}$, then $F(R) = \{\max_{R_i}(X)\}$.
Positive Responsiveness (PR)	*If the support for a winning alternative increases, then it should become the sole winner.*
	If $x^* \in F(R)$ and $R' \neq R$ is such that for all $y, z \in X \setminus \{x^*\}$, we have $\lvert\{i \in N \mid y\ R_i\ z\}\rvert = \lvert\{j \in N \mid y\ {R_j}'\ z\}\rvert$ and $\lvert\{i \in N \mid x^*\ R_i\ y\}\rvert \leq \lvert\{j \in N \mid x^*\ {R_j}'\ y\}\rvert$, then $F(R') = \{x^*\}$.

We introduced axioms as normative principles or desirable properties that we want a voting rule to satisfy. While this is an intuitive description of the term "axiom", we are yet to state precisely what mathematical object(s) should correspond to this concept. We want to show that this is not a trivial task since multiple appealing formalizations exist.

2.2.2 Axioms as Formal Objects

One way to formalize axioms is to translate their natural language descriptions into a formal language. For instance, the anonymity axiom says "All voters should be treated equally" or shorter "One person, one vote". We can express it in first-order logic as the sentence

$$\text{"}\forall \text{ bijections } \sigma : N^* \to N^*\ \forall R, R' \text{ of same electorate size } (\forall i\,(R'_i = R_{\sigma(i)}) \to F(R') = F(R))\text{"}.$$

We will see in Section 3.1.1 that we can even translate axioms into a propositional language in our framework. This suggests defining axioms as syntactic objects in some formal language. In logic, we often differentiate between the *intensional* and the *extensional* definition of a concept or object, or in Frege's (1892) words between "Sinn" and "Bedeutung". Whereas the former is based on the meaning of the term, the latter determines it on grounds of what it designates or which entities it refers to. In the case of axioms, the intension is given by the thought that the axiom expresses, i.e., the normative principle that it captures. This meaning is preserved when defining an axiom in terms of its formulation in some language. The counterpart is given by the extension of an axiom, that is, the set of all voting rules satisfying the principle.

Definition 2.1. For an axiom A, its *extension* (or *interpretation*) is given by $\mathbb{I}(A) := \{F \in \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F \text{ satisfies } A\}$. Similarly, for any set of axioms $\mathcal{A}$, we define $\mathbb{I}(\mathcal{A})$ to be the set of all voting rules that satisfy all axioms in $\mathcal{A}$ simultaneously, i.e., $\mathbb{I}(\mathcal{A}) := \bigcap_{A \in \mathcal{A}} \mathbb{I}(A)$.

An important observation is that two intensionally distinct axioms can have the same extension. That is, although there are two different motivations or norms, the rules that satisfy these coincide. For instance, if we consider a model with only one voter, faithfulness, the Pareto principle and the Condorcet principle all coincide. However, their meaning or leading principle is very different: whereas faithfulness directly speaks about the outcome of one-voter profiles, the Pareto principle excludes dominated alternatives from the outcome and the Condorcet principle requires electing the Condorcet winner. So should we say that these are different axioms? Or should we consider them to be the same?

We will see in Section 2.3 that in the basic definition of justifying or forcing outcomes and of Voting by Axioms, the only information about an axiom that we make use of is its extension, the set of rules satisfying the axiom. Thus, for succinctness and simplicity, an alternative approach is to identify axioms with their extension, forgetting about their intentional meaning. However, this conciseness comes at a price. Namely, it is problematic since it presupposes that we actually know the extension of every axiom. In practice however, especially when working with a large domain with many voters and alternatives, the set of possible voting rules becomes huge. Note that the cardinality is given by

$$|\mathcal{L}(X)^+ \to \mathcal{P}_+(X)| = |\mathcal{P}_+(X)|^{|\mathcal{L}(X)^+|} = (2^m - 1)^{(m!+1)^n - 1}.$$

Therefore, for many axioms, it is infeasible to list all rules that are consistent with the axiom, not to mention computational difficulties. But if we do not know the extension of an axiom, we are unable to use it in a framework that defines axioms extensionally.

But even if we decide to keep the intensional meaning of an axiom in our formal description, there are still decisions to be taken. We could associate with each axiom exactly one formula. But which formal language should we use to express the axiom? Promising candidates are first-order and propositional logic. If we want to generalize to a setting where the universe N^* can be infinite, do we need infinite formulas? To circumvent this, we could break down a single axiom into many smaller axioms, together forming the whole axiom. For instance, we could split unanimity into many axioms, one for each profile in which one candidate has unanimous support, saying that under this specific profile, the candidate must be the single winner. This means breaking up the general principle governing the axiom into its individual cases. In this way, we could also define an axiom as a set of syntactic objects, the set of the axiom's *instances*.

This is one possible motivation for the concept of axiom instances. Before we dive deeper into what they are and how to define them, we want to present a way of classifying axioms. We already saw a functional distinction between axioms due to Thomson (2001). Further, he sets "one-problem axioms", considering only one or a few profiles at a time, apart from axioms with "full coverage", imposing a condition across the whole domain (Thomson, 2001, p.353). The most profound account of sorting axioms into groups was given by Fishburn (2015). He introduces three main categories of axioms: structural, existential and universal axioms.

- *Structural axioms* describe what the domain, i.e., the set of profiles on which a voting rule is defined, looks like. They are sometimes also called domain-restricting axioms because we already chose the set $\mathcal{L}(X)^+$ as the function domain, which gets reduced in size by the axiom. This kind of axiom works on a meta-level and concerns the structure of the voting model rather than imposing requirements on which outcome the rule should return. Examples are the axiom "There are at least three alternatives" or "All voters have mutually distinct preferences". In this thesis, we do not consider these to be axioms since they have to be hard-wired into the voting model.

- The second type of axiom are *universal axioms*. These are axioms that only contain universal quantifiers, no existential quantifiers. They are often phrased as conditionals, i.e., "If ... is the case, then F has the property ...". This is the most extensive and central class of axioms according to Fishburn. He subdivides it into *intraprofile* and *interprofile axioms*. The former kind only speaks of one profile at a time whereas the latter connects multiple profiles. Standard examples of interprofile axioms are anonymity, neutrality or reinforcement. We can further decompose the class of intraprofile axioms into *active* and *passive axioms*. This is similar to Thomson's (2001) differentiation between one-problem and full coverage axioms. While active axioms apply only to a subset of all profiles and are usually of the form "For all profiles such that ..., it is the case that ...", passive axioms apply generically to all profiles. Examples for the former kind are the Condorcet principle or unanimity. A passive axiom, on the other hand, might say that the rule should be resolute and settle on a single winning alternative for each profile.

- The remaining axioms, those that contain an existential quantifier, are named *existential axioms*. They stipulate the existence of some voter, outcome or profile with certain properties. Examples are surjectivity, stating that every possible outcome is chosen under some profile, or the no-dummy axiom, requiring that for every voter, we can find a profile such that if they unilaterally change their vote, the outcome changes, i.e., every voter has some influence.

As we have seen, an axiom need not have implications for all profiles in the domain. Therefore, with every axiom A, we want to associate a set $\mathbb{P}(A)$ of profiles that it speaks about or affects. If the axiom is given in some formal language, its definition is straightforward. Nardi (2021) follows this approach in his thesis. We will see, however, that we can also define it without a syntactic form at hand. But what do we mean by an axiom "speaking about" a profile when no underlying language is given? Intuitively, an axiom speaks about a profile if it imposes some condition on what the outcome under this profile can be. This means, there should be some outcome that a rule satisfying the axiom is not allowed to give back (possibly dependent on other profiles' outcomes). On the contrary, if an axiom does not speak about a profile, a voting rule

should be allowed to assign any outcome to it, independent of the other profiles' outcomes. To summarize, on the set of profiles that an axiom speaks about, the voting rule's restriction has to belong to a certain subset of all possible functions. Further, for any rule in this subset, every possible extension to the whole domain must also belong to the axiom's interpretation. This motivates the following definition. For a set of profiles S, we write S^c for the relative complement $\mathcal{L}(X)^+ \setminus S$.

Definition 2.2. Let the set of profiles that *axiom A speaks about* $\mathbb{P}(A)$ be the smallest set of profiles S such that we can view the interpretation $\mathbb{I}(A)$ as a product of a set of functions on S and all functions on its complement, i.e., $\mathbb{I}(A) = \mathcal{F} \otimes (S^c \to \mathcal{P}_+(X)) = \{F : \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F{\restriction_S} \in \mathcal{F}\}$ for some $\mathcal{F} \subsetneq (S \to \mathcal{P}_+(X))$.

This definition captures that on $\mathbb{P}(A)$, the axiom imposes conditions, restricting what the function can be, and on the complement, the voting rule can be any function. Importantly, it is a property of the product operator that the outcomes of profiles in $\mathbb{P}(A)$ are assigned independently from the outcomes on profiles that A does not speak about, since any function in $\mathcal{F}$ gets extended with every possible function defined on $\mathbb{P}(A)^c$. We need to check that it is always possible to split the extension into a product of that form and that there is indeed a single smallest set S achieving this.

Proposition 2.3. *The set $\mathbb{P}(A)$ is well-defined, i.e., for every axiom A, there exists a unique smallest set of profiles S such that $\mathbb{I}(A) = \mathcal{F} \otimes (S^c \to \mathcal{P}_+(X))$ for $\mathcal{F} \subsetneq (S \to \mathcal{P}_+(X))$.*

Proof. We want to show existence and uniqueness of such a set S.

For existence, notice that we can always write the extension of a non-trivial axiom (i.e., with $\mathbb{I}(A) \neq \mathcal{L}(X)^+ \to \mathcal{P}_+(X)$) as a trivial product $\mathbb{I}(A) \otimes \emptyset$. Similarly, for a trivial axiom, we can write its extension as $\emptyset \otimes \mathcal{L}(X)^+ \to \mathcal{P}_+(X)$. Thus, there does always exist such a set $S \in \{\mathcal{L}(X)^+, \emptyset\}$ that allows to split the extension into a product.

To prove uniqueness, we show that if two distinct sets S and S' satisfy $\mathbb{I}(A) = \mathcal{F} \otimes (S^c \to \mathcal{P}_+(X)) = \mathcal{F}' \otimes (S'^c \to \mathcal{P}_+(X))$, then their intersection also has the property of splitting the extension into a product. Without loss of generality, we may assume that $S \cap S' \neq \emptyset$. Because if the sets were disjoint, then $S' \subseteq S^c$ would hold. From this, by the product splitting property of S, we could infer that voting rules in $\mathbb{I}(A)$ restricted to S' could be any function. This would contradict the definition of S' being a domain on which the acceptable functions are a strict subset of all possible functions. So the two sets do indeed intersect. Notice that for any subdomain of a domain, we can write the set of all functions on the whole domain as a product of all functions on the subdomain and on its complement. Therefore, together with $(S' \setminus S) \sqcup (S \cup S')^c = S^c$, we have

$$\begin{aligned}\mathbb{I}(A) &= \mathcal{F} \otimes ((S' \setminus S) \to \mathcal{P}_+(X)) \otimes ((S \cup S')^c \to \mathcal{P}_+(X)) \\ &= \mathcal{F}' \otimes ((S \setminus S') \to \mathcal{P}_+(X)) \otimes ((S \cup S')^c \to \mathcal{P}_+(X)).\end{aligned}$$

We can combine these, and use that $(S' \setminus S) \sqcup (S \setminus S') \sqcup (S \cup S')^c = (S \cap S')^c$, to find a set $F'' \subsetneq S \cap S' \to \mathcal{P}_+(X)$ such that

$$\begin{aligned}\mathbb{I}(A) &= \mathcal{F}'' \otimes ((S' \setminus S) \to \mathcal{P}_+(X)) \otimes ((S \setminus S') \to \mathcal{P}_+(X)) \otimes ((S \cup S')^c \to \mathcal{P}_+(X)) \\ &= \mathcal{F}'' \otimes ((S \cap S')^c \to \mathcal{P}_+(X)).\end{aligned}$$

Thus, we have shown that $S \cap S'$ is such that we can represent the extension of A as a product of functions as required. From this, it follows that the unique smallest set with the product splitting property is the intersection of all sets S that split the interpretation into a product as required. □

Given a formal language, we may define the set $\mathbb{P}(A)$ as the set of all profiles occurring in the formal specification of A. This set is always a (not necessarily strict) superset of the extensionally defined set of profiles that A speaks about. This is immediate since A can only restrict the outcome under a profile by including an explicit statement featuring the profile. However, the formal statement of A in the given language may contain redundancies or trivial statements, e.g., a tautology $R = R$. The set derived from the language-based definition of $\mathbb{P}(A)$ would include such profiles and is therefore not necessarily minimal. In this light, the language-independent definition appears to be superior due to its ability to include only those profile that the axiom does indeed impose non-trivial conditions on.

For a set of axioms $\mathcal{A}$, we define $\mathbb{P}(\mathcal{A})$ to be given by $\bigcup_{A \in \mathcal{A}} \mathbb{P}(A)$. Notice that this convention preserves the product splitting property of the set $\mathbb{P}(\cdot)$, i.e., $\mathbb{P}(\mathcal{A})$ is the smallest set of profiles S such that $\mathbb{I}(\mathcal{A}) = \mathcal{F} \oplus (S^c \to \mathcal{P}_+(X))$ holds for some $\mathcal{F} \subsetneq S \to \mathcal{P}_+(X)$.

In the social choice literature, axioms are often taken do be desiderata, properties that help set apart good voting rules from bad ones. A large part of results in the field are so-called impossibility results. They are usually interpreted as negative results, stating that already a handful of desirable axioms together are not satisfiable. We want to offer a new perspective on axioms in this thesis. In Voting by Axioms, instead of requiring that our derived voting rule fully satisfies all axioms that we decided on, we try to merely fulfill certain necessary conditions imposed by the axioms. That is to make sure that our rule does not completely go against the axioms. So rather than focusing on how the rule behaves globally, we want to ensure that, locally, the requirements that a set of axioms places on some profile are satisfied. This will be discussed in greater detail once we formally define the procedure for Voting by Axioms. The shift in perspective that we suggest is from viewing an axiom as absolute, global, desirable property that we are only content with if it is satisfied on the whole domain, towards considering it to be a multitude of conditions applying only to a subdomain that we aim to fulfill a great part of.

2.2.3 Dividing Axioms into Instances

But how should we change our binary view on (dis-)satisfaction into a more sensitive measure capturing to what degree a voting rule satisfies a set of ax-

ioms? The key idea is to break one axiom up into multiple components and then determining how many of these are satisfied. Thomson (2001) too argues in favor of considering a multitude of logically independent, concise axioms when aiming for a positive result. They should not be redundant and "each axiom should preferably embody only one specific aspect of the general idea" (Thomson, 2001, p.339). This is our starting point for defining *axiom instances*. We want these to be (atomic) particles, components or substatements of the axiom, together making up the whole axiom. One way to get there is viewing the axiom as general normative principle and taking every application or actualization of this principle to be an instance. For example, anonymity says that if two profiles are the same modulo renaming of the agents, then the outcomes should agree. One instance of this principle would be any specific choice of two profiles, e.g., "$F(123, 231, 312) = F(231, 123, 312)$". With this notion at hand, we can calculate how many instances of an axiom are satisfied by a voting rule.

Notice that it might be difficult to define a general procedure on how to obtain instances for an axiom due to the different structure and nature of axioms. Intraprofile axioms yield a natural notion of instance because they look at every profile in isolation. So for each profile, take the imposed condition to be an instance. For interprofile axioms, on the other hand, profiles cannot be viewed independently from each other. It does not make sense to ask whether the axiom is satisfied on a single profile. Further, this idea might not be suited for existential axioms. The problem is that they access the whole domain at once (since one has to check for all profiles whether the negation holds) and, therefore, do not give rise to a notion of locality or partial satisfaction. If an existential axiom does contain universal quantification, however, we might be able to split it up into multiple existential instances.

In the following, we want discuss three ways to define axiom instances: Either we can require for axioms to be given in terms of their instances, or we can define axioms syntactically in some language and derive instances from this representation, or, alternatively, we can define axioms extensionally and define a procedure for obtaining instances.

The first one of these possibilities is easiest to work with since we leave it to the user to manually choose the most natural division of axioms into instances. Similarly, we can view instances as atomic axioms (in the style of algebraic axioms, see Section 3.2.2) and take axioms to be sets or complex expressions built up from these instances. In that way, we view instances not as realizations of general principles, but as building blocks to create axioms from. In both of these cases, we take instances to be the foundational concept from which we derive axioms. In the course of this thesis, we will see that it matters that the separation of axioms into instances is rather fine- than coarse-grained and that it is comparable for different axioms. Thus, a more systematic approach for obtaining instances is desirable. This leads us to consider ways of including the generation of instances for a given axiom into the framework.

Ways to obtain instances from a syntactic object are as follows: For a propositional formula, transform it into conjunctive normal form (CNF), then take each disjunctive clause to be an instance. For a first-order formula, transform

it into prenex normal form, then take each instantiation of the leading universal quantifiers to be an instance. However, these syntax-based notions of axiom instances are language-dependent and might not always correspond to what we would intuitively take to be a particle of the axiom. For example, the natural propositional logic formulations of reinforcement (see Table 3.1) and of many existential axioms, e.g., surjectivity, are not in CNF. Although they can of course be transformed into CNF, we do not expect the resulting formulas to be intuitive specifications of the axioms.

If an axiom is defined extensionally, i.e., if we only know which rules satisfy the axiom, we try to deduce its instances as the principles common among all rules. However, it is not a priori clear how to extract these principles from the extension. Are these principles not part of the intension, the meaning of an axiom? Again, for intraprofile axioms, that force an outcome on each profile that they speak about, this is straightforward since for the associated profiles, all rules in the extension return the same outcome. It is more difficult to extract a conditional, speaking about multiple profiles at a time, from the extension. Consider for instance the axiom "If $F(R) = O$, then $F(R') = O'$". Its extension will contain rules that return O given R but also rules that return any other outcome for R. The first type of rule will return O' for R', whereas the second type can give back any outcome under R'. To uncover the dependency between the two profiles, we would need to look at the image of (R, R') under all voting rules in the axiom's extension, to find that all outcome tuples are possible besides (O, O''), where $O'' \neq O'$. Since an axiom can speak about arbitrarily many profiles, this means that we need to inspect all possible combinations of profiles in the domain. So we suggest a procedure for deriving instances from the extension of an axiom by extracting, step by step, the intraprofile, two-profile, up to all-profile/existential conditions restricting the outcomes of voting rules satisfying the axiom. Fix an enumeration of profiles $\mathcal{L}(X)^+ = \{R_1, R_2, \dots, R_{|\mathcal{L}(X)^+|}\}$.

- Given some axiom A, start by, for all profiles R_i, defining instances "The outcome under R_i lies in $A(R_i)$", in case $A(R_i) := \{F(R_i) \mid F \in \mathbb{I}(A)\}$ is a strict subset of $\mathcal{P}_+(X)$.

- Next, we consider all pairs of profiles (R_i, R_j) with $i < j$. If $A(R_i, R_j) := \{(F(R_i), F(R_j)) \mid F \in \mathbb{I}(A)\} \subsetneq A(R_i) \times A(R_j)$, then we add the instance "For all $(O, O') \in (A(R_i) \times A(R_j)) \setminus A(R_i, R_j)$, if R_i returns O, then R_j does not return O'". The idea is that since we already extracted all intraprofile conditions, we know that the outcome of the pair must lie in $A(R_i) \times A(R_j)$. If we cannot find a voting rule in A's extension for every of these outcome pairs, assigning these outcomes to the profiles, this means that there exist further conditions that A imposes on the pair. Thus, we add one instance stating that all these outcome pairs should be excluded.

- For any finite tuple of profiles $(R_{i_1}, R_{i_2}, \dots)$, let

$$A(R_{i_1}, R_{i_2}, \dots) := \{(F(R_{i_1}), F(R_{i_2}), \dots) \mid F \in \mathbb{I}(A)\}.$$

Continue this process by considering larger tuples as follows: For all $k \leq |\mathcal{L}(X)^+|$, consider k-tuples of profiles $(R_{i_1}, \ldots, R_{i_k}) \in (\mathcal{L}(X)^+)^k$ such that $i_1 < i_2 < \cdots < i_k$. If $A(R_{i_1}, \ldots, R_{i_k})$ is a strict subset of

$$\bigcap_{\ell \leq k} \left\{ (O_1, \ldots, O_k) \middle| \begin{array}{l} O_\ell \in A(R_{i_\ell}), (O_1, \ldots, O_{\ell-1}, O_{\ell+1}, \ldots O_k) \in \\ A(R_{i_1}, \ldots, R_{i_{\ell-1}}, R_{i_{\ell+1}}, \ldots, R_{i_k}) \end{array} \right\},$$

then add the instance "For all outcome tuples $(O_1, \ldots O_k)$ contained in the big intersection but not in $A(R_{i_1}, \ldots, R_{i_k})$, if $R_{i_j} \mapsto O_j$ for all $j < k$, then R_{i_k} does not return O_k".

This means, we check for a given k-tuple which outcome tuples are still allowed, taking into account all restrictions already imposed on subtuples. Notice that the maximum conditions that we have at stage k are all conditions speaking about the behavior on $k-1$ many profiles. Now for each k-tuple, there are $\binom{k}{k-1} = k$ many possibilities to obtain conditions on the tuple. Thus, for each way of forming the k-tuple out of a $(k-1)$-subtuple and one more profile, we need to check which conditions we can derive from the $(k-1)$-tuple on the k-tuple. We want to take *all* these conditions into account simultaneously, so the allowed tuples are the ones allowed in all ways of forming the k-tuple from a $(k-1)$-tuple. That is, they lie in the intersection of all the allowed outcomes w.r.t. some $(k-1)$-condition. We then add one instance that excludes all those outcome tuples that would be allowed to the best of our current knowledge, but that no voting rule satisfying A actually returns on the k-tuple.

Notice that when setting $k = 2$ in the general case, we recover the instances that we defined for pairs. We will see that this procedure does indeed yield instances for the axiom A according to some minimal requirements on pages 24 to 25.

Example 2. *Consider the reinforcement axiom. This axiom does not impose any conditions on what outcome(s) a single profile or two profiles together should return. Thus,* $\mathrm{REI}(R) = \mathcal{L}(X)^+$ *for all profiles* R*, and similarly,* $\mathrm{REI}(R, R') = \mathcal{L}(X)^+ \times \mathcal{L}(X)^+$ *for all* (R, R')*. So at stage* $k = 3$*, according to our current knowledge, all outcome triplets would be allowed. However, we see that, given a triplet of profiles* (R_1, R_2, R_3) *such that* $N_{R_3} = N_{R_1} \sqcup N_{R_2}$*, no voting rule in* $\mathbb{I}(\mathrm{REI})$ *returns a triplet* (O_1, O_2, O_3)*, where* $O_3 \neq O_1 \cap O_2 \neq \emptyset$*. Thus, for each such profile triplet, we obtain an instance excluding these outcome triplets. If we consider larger tuples, however, no further restrictions will be imposed, since the reinforcement axiom only speaks about three profiles at a time. Notice that each instance corresponds exactly to stating, for one choice of three profiles where one is the disjoint union of the two others, that if the outcomes of the two smaller profiles intersect, this intersection should be the outcome of the larger profile.*

Without having settled on what exactly an axiom instance is in general, we introduce the following notational conventions.

Definition 2.4. If A' is an axiom instance of A, we write $A' \lhd A$. Further, if A' is an instance of some axiom in the axiom set $\mathcal{A}$, we write $A' \lhd \mathcal{A}$.

With an axiom instance A' we may associate its *extension* (or *interpretation*) $\mathbb{I}(A')$ consisting of all voting rules that satisfy the instance. Further, we write $\mathbb{P}(A')$ for the set of profiles that the instance speaks about.

We have come to understand that a general definition of axiom instance is hard to find. So without fixing one particular way of deriving instances, we can define the following necessary requirements for the definition of axiom instances, inspired by Boixel and Endriss (2020).

(Axiom)	Every axiom instance A' of an axiom A is an axiom itself.
(Segmentation)	The extension of A is equal to the intersection of all instance extensions, i.e., $\mathbb{I}(A) = \bigcap_{A' \lhd A} \mathbb{I}(A')$. In particular, $\mathbb{I}(A') \supseteq \mathbb{I}(A)$ holds.
(Substatement)	The set of profiles that A speaks about is the union of sets of profiles that its instances speak about, i.e., $\mathbb{P}(A) = \bigcup_{A' \lhd A} \mathbb{P}(A')$.

Recall that we want the separation of an axiom into instances to be a division of the axiom's statement into multiple parts. The first minimal condition ensures that axioms and their instances are of the same kind — they both set conditions for voting rules. The second requirement formalizes the idea that instances split up the original axiom, i.e., they all follow from the axiom and taken together, they form the whole axiom. Together with (Segmentation), the third requirement stresses that instances should be parts of the original axiom statement, not just arbitrary weakenings. This is, an instance should only contain pieces of information and requirements that the axiom itself includes. Rooted in how we treat axioms phrased in a given formal language, we therefore require that instances, as substatements of the axiom, talk only about the same or a subset of the profiles that the original axiom speaks about.

Example 3. *We can split unanimity into m axioms, one for each alternative x saying "If everyone ranks x as their most preferred alternative, then $\{x\}$ should be the winning set.". It would not be acceptable to replace the instance for $x = 2$ by a disjunction "(The outcome under the profile $(12\dots m)$ should be $\{2\}$ and the outcome under $(12\dots m, m12\dots m-1)$ should be $\{1,2\}$),* or *(if everyone ranks 2 as their most preferred alternative, then $\{2\}$ should be the winning set)". Notice that the disjunctive statement is a weakening of the aforementioned instance for $x = 2$, and therefore a weakening of unanimity. We still obtain the extension of unanimity when intersecting all extensions of the instances since the instance for $x = 1$ contradicts the first disjunct. However, unanimity never mentions the profile $(12\dots m, m12\dots m-1)$ and so, we would not consider the disjunctive statement an instance of unanimity.*

This shows that (Substatement) does not already follow from (Segmentation). The crucial point to observe is that, only because one interpretation is a subset of the other, does not mean that the representation of the extension as a product of sets of functions stays intact.

Example 3 (continued)**.** *For unanimity, we have*

$$\mathbb{P}(\text{UNA}) = \mathcal{U} := \{R \mid \exists x \in X \, \forall i \, \forall y \neq x \, (x \, R_i \, y)\},$$

which allows us to write $\mathbb{I}(\text{UNA}) = \{f_{una}\} \otimes (\mathcal{U}^c \to \mathcal{P}_+(X))$*, where* $f_{una} : \mathcal{U} \to \{\{x\} \mid x \in X\}$ *is the function that satisfies* $f_{una}(R) = \{x\}$ *iff* $x \, R_i \, y$ *for all* i *and all* $y \neq x$*. Consider* A' *with* $\mathbb{I}(A') := \mathbb{I}(\text{UNA}) \cup \{\mathbf{1}\}$*, where* $\mathbf{1}(R) = \{1\}$ *for all* R*. We clearly have* $\mathbb{I}(\text{UNA}) \subseteq \mathbb{I}(A')$ *but note that the extension of* A' *can only be split trivially into* $\mathbb{I}(A') \otimes \emptyset$ *since adding* $\mathbf{1}$ *means speaking about the outcome on every profile. In other words,* $\mathbb{P}(\text{UNA}) = \mathcal{U} \not\supset \mathcal{L}(X)^+ = \mathbb{P}(A')$*.*

The aforementioned three conditions arise immediately from our conception of the term "instance" as one out of multiple substatements, together forming the whole axiom. Additionally, with regards to our framework and the further use of axiom instances in this thesis, we want to stipulate the following condition called (Non-Redundancy).

> Axiom instances are not redundant, i.e., the set of instances is minimal with respect to the aforementioned three conditions. In particular, if $\mathcal{A}'$ is a set of instances for A, then for every subset $\mathcal{A}'' \subsetneq \mathcal{A}'$, the extension is strictly bigger than that of A, so $\mathbb{I}(\mathcal{A}'') \supsetneq \mathbb{I}(\mathcal{A}') = \mathbb{I}(A)$.

There are weaker conditions that one could consider. For instance, that the extensions of instances should be mutually distinct or, as suggested by Boixel and Endriss (2020), that there should only be finitely many instances. Notice that if we looked at an infinite electorate, it would be quite natural to have infinitely many instances. To allow for this generalization, we do not adopt the latter condition. Mutual distinctness ensures computability in our finite framework since there are only finitely many distinct supersets of an axiom's extension, thus only finitely many possible instances. However, recall that we want to divide an axiom into instances, different parts of the whole statement, in order to measure partial or gradual satisfaction of an axiom. This is, we want to count how many parts of an axiom are fulfilled. For this quantitative measure to be meaningful, it is key that different axiom instances actually capture disjoint ideas, i.e., that they are redundance-free. This is the reason for imposing this additional condition.

We want to undergird the validity and relevancy of the minimal requirements for instances that we put forward by showing that they are satisfied by all described concrete ways to derive instances.

Suppose we work in a propositional language and take the disjunctive clauses of a formula in CNF to be the instances. Then, each instance is itself a sentence in the language, thus, is an axiom. By the semantics of conjunction, a voting

rule is contained in the extension of the whole axiom whenever it satisfies all disjunctive clauses, i.e., whenever it lies in the extension of each instance. Trivially, since every disjunctive clause is a subformula of the axiom, the instance speaks only about profiles occurring in the axiom. Thus, all three minimal requirements are satisfied. However, if we consider an arbitrary CNF, it may contain multiple equivalent clauses with the same extension. So, to additionally satisfy the (Non-Redundancy) requirement, we should take the disjunctive clauses and derive a minimal subset that has the axiom's extension to obtain an acceptable set of instances instead.

Similarly, suppose we work in a first-order language and we take an instance to be an instantiation of all leading universal quantifiers in the axiom's formulation. These instantiations are clearly axioms themselves. Since they represent one way of making the universal quantification true, their extension is a superset of the axiom's extension. Lastly, since the universally quantified formula speaks about all objects of some kind and an instantiation speaks solely about one of these objects, the set of profiles that the instance speaks about is a subset of the profiles that the original axiom imposes conditions on. Again, we might need to disregard certain instantiations to guarantee non-redundancy.

Lastly, recall the extension-based procedure for deriving instances defined on pages 20 to 21. We want to show that this method indeed yields instances.

- Notice that all expressions that we called instances are indeed axioms conditioning the outcome of a voting rule.[5]

- Since we obtained the conditions from analyzing the voting rules in the extension of the axiom, their extensions are clearly supersets of the extension of the axiom. Notice also that by considering all possible combinations of profiles, we exhausted all possible restrictions that the axiom could impose. Thus, there cannot exist a voting rule that satisfies all derived instances but not the axiom itself, since this would mean that the voting rule contradicts a condition imposed by A. But this condition must be among the instances already. Thus, the requirement (Segmentation) is met.

- Regarding (Substatement), note that the instances capture exactly when the axiom disallows a certain outcome to be returned. Thus, an instance cannot speak about a profile, i.e., restrict what outcome can be assigned, if the axiom itself does not. Similarly, if the axiom speaks about a profile, there must be an instance that reflects this since we exhausted all possible conditions.

- Regarding (Non-Redundancy), notice that we already eliminated redundant conditions by extracting those with fewer profiles first and only checking for *additional* restrictions when considering larger tuples. This is because an instance on k profiles corresponds to *excluding* those outcomes

[5] Strictly speaking, the axiom and the instances need to be the same kind of formal object. So if the axiom is defined by its extension, then for each extracted condition, define an instance to be the set of voting rules satisfying this condition.

that would be allowed when taking into account all previously imposed conditions (speaking about up to $k-1$ profiles), but that no voting rule satisfying the axiom actually returns. Clearly, an instance on a smaller tuple does not imply any of the conditions on a larger tuple (since we only add an instance if there are strictly less admissible outcomes than expected from all conditions on smaller tuples). But also, an instance derived from a tuple does not imply any instance derived from a subtuple (since we exclude only outcomes that have not been excluded on lower stages already). Further, we required the tuples to have ascending indices to not consider the same profile combination multiple times. Thus, the instances are indeed non-redundant.

A Classification of Axioms. Notice that this method yields a new way of classifying axioms. Recall that the classification of axioms by Fishburn (2015) was very vague. Although, intuitively, we can grasp what is meant by a universal versus an existential axiom, how can we tell what kind of quantification occurs in an axiom if we do not even have a formal language given? What does it mean that existential axioms are "based primarily on existential qualifiers", but universal axioms may "use such qualifiers in a secondary manner" (Fishburn, 2015, p.180)? We claim that with the given procedure, we can define a sharp classification of axioms, somewhat similar to Fishburn's and we introduce a hierarchy of axioms.

We say that A is a *k-profile axiom* (or an axiom *of rank k*) if, in the extension-based instance division, the last instance is derived from a k-tuple. This means that no condition imposed by the axiom speaks about more that k profiles at a time. We want to call 1-profile axioms *intraprofile axioms* and, similarly, for $k > 1$, a k-profile axiom is named *interprofile axiom*. Importantly, these do not coincide with Fishburn's notions.

At first, it might seem that tuples for $k < |\mathcal{L}(X)^+|$ correspond to Fishburn's universal axioms and $|\mathcal{L}(X)^+|$-profile axioms correspond to existential axioms, since if we existentially quantify across the whole domain, we need to guarantee that, when looking *at all profiles*, the negation is not true. The problem with this naming convention of axioms is that we can also have universal axioms speaking about all profiles, e.g., "If all even-numbered electorate profiles are assigned to the same outcome, then the rule is constant", or axioms existentially quantifying over a subdomain and therefore only speaking about a subset of all profiles, e.g., "There exists a profile in which everyone votes the same and for which all alternatives win". Thus, the axioms that Fishburn takes to be intra- and interprofile axioms, obtain the same label in our classification. However, there can be axioms that Fishburn would not include in these classes, due to existential quantification, which we still name that way. Note that the suggested distinction between k-profile axioms for $k < |\mathcal{L}(X)^+|$ and $|\mathcal{L}(X)^+|$-profile axioms is also somewhat similar to what Thomson (2001) calls one-problem axioms and full coverage axioms, respectively. Recall that he names those axioms that do not speak about all profiles one-problem axioms and says that an axiom has

full coverage whenever it imposes some non-trivial condition on every profile. Thomson describes the latter kind as those axioms that can possibly uniquely characterize a voting rule by themselves if they force an outcome on all profiles. While a $|\mathcal{L}(X)^+|$-profile axiom definitely has full coverage, for all other profiles, the rank of an axiom does not determine whether the axiom is a one-problem axiom or not. Think for instance of a passive intraprofile axiom in Fishburn's classification, which speaks about all profiles, one at a time. This is a full coverage intraprofile axiom. Thus, our classification differs both from Fishburn's and Thomson's approach.

Instead of making our classification dependent on an axiom's formulation in some language, in which case well-definedness would depend on logically equivalent formulas being classified as the same kind of axiom, our hierarchy is based purely on the axiom's extension and the behavior of the voting rules satisfying the axiom. This hierarchy uncovers the structural complexity of an axiom. The higher the rank of an axiom is, the more interdependencies between profiles it imposes. In the instance extraction process, we see that axioms with a higher rank can, in general, be more restrictive, i.e., yield a smaller extension, than those with lower rank, since they impose additional conditions that were not expressible in terms of smaller tuples.

In Voting by Axioms, we need 1-profile axioms to establish a set of profiles that outcomes are forced on. As a next step, we have to include k-profile axioms, where k is less or equal than the size of the set of profiles that the intraprofile axioms force an outcome on, for standing a chance at well-definedness. These axioms of rank k can then lead to forced outcomes on other profiles. If necessary, we can then also include axioms of higher rank.

Example 4. *Consider the following standard axioms.*

- *Unanimity is a* 1*-profile axiom since it specifies for all profiles with unanimous support independently what the outcome should be.*
- *Anonymity is a* 2*-profile axiom specifying for each two profiles that can be obtained from one another by renaming the agents, that the outcome should be the same. Technically, this will require for each profile that all profiles obtained from it by renaming the agents should have the same outcome. Since there are* $|N^*|!$ *many possible renamings, the outcome might prescribe that up* $|N^*|!$ *many profiles should have the same outcome. However, this requirement follows from the conditions imposed on pairs. So if one of the* $|N^*|!$ *profiles returns a different outcome from the others, then it already contradicts the condition that paired with one of these other profiles, it should return the same outcome.*
- *Surjectivity is a* $|\mathcal{L}(X)^+|$*-profile axiom since for a given outcome, if for any* $(|\mathcal{L}(X)^+| - 1)$*-tuple of profiles, none of them is assigned to the outcome, then the remaining profile needs to be assigned to the profile.*

We introduced multiple ways of deriving instances from an axiom. We can alter how fine-grained we want this division into instances to be. The most

extreme but useless version of this is splitting the axiom into instances that each exclude exactly one voting rule, i.e., for every $F \notin \mathbb{I}(A)$, consider the instance "The voting rule does not coincide with the rule F". Notice, however, that the set of profiles that the instance speaks about is the set of all profiles in this case. So only if $\mathbb{P}(A) = \mathcal{L}(X)^+$ is the case, will this be an acceptable definition of instances with regards to (Substatement).

It seems quite natural to split up instances further and further until we obtain the smallest substatements of the axiom with respect to the profiles that they speak about. This means that if we were to divide the instances even further and thereby strictly reducing the set of occurring profiles, then the resulting conditions would no longer be weakenings of the original axiom. In the case of anonymity, for example, it makes sense to split the axiom into instances for each pair of profiles R, R', where R' is obtained from R by renaming the agents, that express "The outcome under R and R' must coincide". The only way to generate a new instance that speaks about strictly less profiles is a condition only applying to exactly one of the two profiles, at the same time allowing any arbitrary outcome under the other of the two profiles. Any rule satisfying this reduced condition but not assigning the same outcome to both profiles is not anonymous. Thus, the extension of the reduced condition is not a superset of the extension of anonymity. So this violates the second minimal requirement, (Segmentation), for instances.

However, this degree of granularity is not always the most natural one. If the axiom is existential, for example, we can never split it up into instances that are not themselves existential and therefore speak about all profiles simultaneously. Nonetheless, we might want to split the axiom up into instances. If it contains a universal quantification over outcomes or voters, for instance, we want to break up the axiom accordingly. Examples are the axioms surjectivity and no-dummy, which naturally split into "The outcome under some profile is O." for every $O \in \mathcal{P}_+(X)$ and "There exist two profiles that only differ in i's vote that have different outcomes." for every voter i, respectively. This is another reason for why there might not be a general procedure applying to all axioms to derive instances.

We explored what the term "axiom" refers to in the field of social choice theory and that we can define axioms either as syntactic objects or purely extensionally. We saw multiple common examples of axioms and ways to classify axioms according to their structure or function. Further, we challenged the view that axioms are normative principles that are either completely satisfied or not. Instead, we suggested dividing them into multiple axiom instances, of which only a part can be satisfied. Formalizing the notion of an axiom instance turned out to be difficult in all generality. We suggested definitions for first-order and propositional logic and a procedure based on the extension of an axiom. Beyond that, we worked out necessary requirements as a baseline for deriving instances. With this deepened understanding of axioms and their instances, we can now use them as central objects in defining Voting by Axioms.

2.3 Voting by Axioms

It is the objective of this thesis to define a decision procedure that respects given axioms. In contrast to first defining a voting rule and then using the axiomatic method to assess its quality, we want to start by specifying axioms that we care about and then define a voting rule, profile by profile, returning outcomes that are justified by these axioms. In this section, we will explain what it means for an axiom to *justify* or *force* an outcome in the sense of Boixel and Endriss (2020) and, based on this, define a method for deriving a voting rule from a collection of axiom sets.

The idea behind the justification of outcomes in voting is using axioms to answer the question why a certain outcome should be assigned in a given situation. Axioms are normative properties that society decided a reasonable voting rule should exhibit. Therefore, it is natural to use them as arguments in an explanation or to pit the behavior of a voting rule against them. We will say that an axiom justifies picking some outcome if it leaves no other choice, i.e., if choosing any other outcome would contradict the axiom. In this case, we say that the axiom forces the outcome and this occurs exactly if all voting rules satisfying the axiom give back this very outcome for the profile. Hence, if an axiom forces an outcome, assigning this outcome can be viewed as a necessary condition in order for the axiom to be satisfied.

Definition 2.5. Given some profile $R \in \mathcal{L}(X)^+$, an axiom set $\mathcal{A}$ with $\mathbb{I}(\mathcal{A}) \neq \emptyset$ *forces* (or *justifies*) an outcome $O \in \mathcal{P}_+(X)$ if for all $F \in \mathbb{I}(\mathcal{A})$, it holds that $F(R) = O$.

For a set of axioms $\mathcal{A}$, we denote the set of profiles that it forces an outcome on by $Forc(\mathcal{A}) := \{R \in \mathcal{L}(X)^+ \mid \mathbb{I}(\mathcal{A}) \neq \emptyset \text{ and } \mathcal{A} \text{ forces an outcome on } R\}$.

Whenever we say that an axiom or a set of axioms forces an outcome, we presuppose that its interpretation is non-empty, even if not explicitly stated. Note that we give a shortened and simplified account of what it means for axioms to justify an outcome in this thesis. We only consider what Boixel and Endriss (2020) call the *normative basis* of a justification, the set of axioms that forces an outcome. Additionally, they specify a minimal set of instances of these axioms that forces the outcome, which corresponds to a precise *explanation*. Rather than referring to broad principles, this explanatory component will contain the precise instances relating to the profile in question. Such a normative basis and explanation together make up a justification. For a more detailed account, we refer to the paper by Boixel and Endriss (2020).

Since one set of axioms rarely justifies an outcome on every profile (this would mean that exactly one rule satisfies all the axioms jointly), we will use a set of sets of axioms to define the voting rule. So given a collection of sets of axioms $\mathbb{A}$ and a strict ranking $\succ$ over it, we want to derive a voting rule that returns outcomes justified by the axioms in $\mathbb{A}$, while using the relation $\succ$ to prioritize. The definition is straightforward. Namely, for each profile, assign the outcome that is forced by the highest-ranked axiom set in the collection that forces some outcome on the profile.

Definition 2.6. The *Voting by Axioms* rule *derived from* a non-empty collection of non-empty sets of axioms $\mathbb{A}$ and an order $\succ \in \mathcal{L}(\mathbb{A})$ over it $F_{(\mathbb{A},\succ)}$ assigns $F_{(\mathbb{A},\succ)}(R) = O$ iff there exists some $\mathcal{A} \in \mathbb{A}$ that forces O given R and such that for all $\mathcal{A}' \in \mathbb{A}$ with $\mathcal{A}' \succ \mathcal{A}$, the set of axioms $\mathcal{A}'$ does not force any outcome given R.

Let us look at an example of how to determine the Voting by Axioms rule.

Example 5. *Consider the ordered collection* $\mathbb{A}$ *given by* $\{\text{CAN}\} \succ \{\text{NEU}, \text{FAI}\} \succ \{\text{ANO}, \text{NEU}, \text{PR}\}$ *and a setup with* $m = 2$ *alternatives and* $n = 3$ *voters. Notice that* CAN *only forces an outcome on profiles* $(12, 21)$ *and* $(21, 12)$ *(independent of the electorate). Among the remaining profiles,* FAI *forces an outcome on all profiles with only one voter. Neutrality by itself does not force any outcome and considering* FAI *and* NEU *together does not extend the set of profiles that an outcome is forced on beyond the one-voter profiles. Recall that by May's Theorem,* ANO*,* NEU *and* PR *characterize the simple majority rule for two alternatives. Thus, these axioms force an outcome (namely, the outcome of the simple majority rule) on all remaining profiles. Thus, the voting rule* $F_{(\mathbb{A},\succ)}$ *is well-defined. It will assign the following outcomes (this determines the whole rule since the rule is anonymous):*

$$
\begin{array}{lll}
(12) \mapsto \{1\} & (12,12) \mapsto \{1\} & (12,12,12) \mapsto \{1\} \\
(21) \mapsto \{2\} & (12,21) \mapsto \{1,2\} & (12,12,21) \mapsto \{1\} \\
 & (21,12) \mapsto \{1,2\} & (12,21,21) \mapsto \{2\} \\
 & (21,21) \mapsto \{2\} & (21,21,21) \mapsto \{2\}
\end{array}
$$

It might be unfeasible in situations with a large corpus of axioms and many sets thereof to determine a ranking over all sets of axioms. Instead, as a special case, we can rank the axioms in the corpus and from this generate a ranking over all possible sets of axioms. That is, for a set of axioms $\mathcal{A}$, we can consider $\mathbb{A} = \mathcal{P}_+(\mathcal{A})$ and we can lift a given ordering $>$ over the axioms in $\mathcal{A}$ to a ranking $\succ$ over the non-empty subsets of $\mathcal{A}$, i.e., over $\mathbb{A}$. By slight abuse of notation, we denote the resulting voting rule by $F_{(\mathcal{A},>)}$. To make use of this, we need a method to, from a set of axioms $\mathcal{A}$ with a ranking over its axioms, i.e., $> \in \mathcal{L}(\mathcal{A})$, derive a ranking $\succ$ over subsets of $\mathcal{A}$, i.e., $\succ \in \mathcal{L}(\mathcal{P}_+(\mathcal{A}))$. One way of doing this, introduced by Pattanaik and Peleg (1984), is by saying that one set of axioms is more desirable than another one if its most preferred axiom is better than the other set's highest-ranked one.

Definition 2.7. The *lexicographic maximax ranking* given $\mathcal{A}$ and $>$ is constructed as follows. Set $\{A\} \succ \{A'\}$ for all axioms $A, A' \in \mathcal{A}$ with $A > A'$ and recursively define for $\mathcal{A}', \mathcal{A}'' \subseteq \mathcal{A}$:

$$
\begin{aligned}
\mathcal{A}' \succ \mathcal{A}'' \text{ iff either } & \max_>(\mathcal{A}') > \max_>(\mathcal{A}'') \\
\text{or } & \left(\begin{array}{l} \max_>(\mathcal{A}') = \max_>(\mathcal{A}'') \text{ and} \\ \Big(|\mathcal{A}''| > |\mathcal{A}'| = 1 \text{ or } \mathcal{A}' \setminus \{\max_>(\mathcal{A}')\} \succ \mathcal{A}'' \setminus \{\max_>(\mathcal{A}'')\} \Big) \end{array} \right)
\end{aligned}
$$

We will consider further examples and motivations for lifting rankings over axioms to rankings over axiom sets in Section 4.1.

This chapter laid the foundation for Voting by Axioms by introducing the framework that we work in and by motivating our choice of definitions. Next, we need to check if or when the defined procedure is well-defined and good-natured and whether it achieves what we want it to.

Chapter 3

Analyzing Voting by Axioms

We proposed a method for obtaining a voting rule that justifies its outcomes with underlying axioms. First, we have to examine whether the suggested definition actually yields a well-defined voting rule. We will see that this depends on the collection of axiom sets that we take as a basis. In this chapter, we want to unfold what well-definedness means in this context and what methods we can use to test for it. After this, we want to assess to what extent the defined voting rule succeeds at respecting the chosen axioms, in which cases it has good properties and how difficult it is to calculate the rule.

3.1 Well-Definedness

In the last chapter, we defined what it means for axioms to force an outcome on a given profile and used this to define a decision procedure. This was done profile by profile. More precisely, for every profile, we identified the highest-ranked axiom set that forces an outcome and then assigned this outcome to the profile. There are two immediate questions regarding this procedure. Can we always find some set of axioms in the collection that forces an outcome on a given profile? Is there always a unique outcome forced by an axiom set or could there be multiple? Only if we can answer "yes" to both of these questions, will the described method yield a well-defined voting rule in general. This is, it determines a function on the set of all profiles, assigning exactly one outcome to each of them.

Let us turn to the second question first. We claim that if a set of axioms forces an outcome on a given profile, then it cannot force another outcome on the same profile. The reason is that forcing an outcome means, by definition, that all voting rules return the same outcome on the given profile. This can only be true of one outcome. This can also be seen as an immediate consequence of the result by Boixel and Endriss (2020, Theorem 1) that there cannot exist

justifications for different outcomes based on the same axioms. Since forcing an outcome corresponds to being a normative basis for a justification, we can conclude the following proposition.

Proposition 3.1. *For a given profile R, it is impossible that a set of axioms $\mathcal{A}$ with $\mathbb{I}(\mathcal{A}) \neq \emptyset$ forces two distinct outcomes O_1 and O_2.*

Proof. For the sake of contradiction, assume that $\mathcal{A}$ forces two different outcomes $O_1 \neq O_2 \in \mathcal{P}_+(X)$. This means that $\mathbb{I}(\mathcal{A}) \subseteq \{F \in \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F(R) = O_1\}$ and $\mathbb{I}(\mathcal{A}) \subseteq \{F \in \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F(R) = O_2\}$. Thus, the set $\mathbb{I}(\mathcal{A})$ is a subset of the intersection

$$\{F \in \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F(R) = O_1\} \cap \{F \in \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F(R) = O_2\}.$$

Clearly, the intersection of these two sets is empty since $O_1 \neq O_2$. Hence, $\mathbb{I}(\mathcal{A}) = \emptyset$. This contradicts our assumption of $\mathcal{A}$ having a non-empty extension. Thus, if $\mathcal{A}$ forces an outcome given R, it forces exactly one outcome. □

This means that, for a fixed profile, if we can find a maximal set according to $\succ$ that forces some outcome, it forces exactly one outcome. Notice that if this was not the case, the rule $F_{(\mathbb{A},\succ)}$ would not be well-defined since we would need to further specify how it should deal with ties, i.e., with the case in which multiple outcomes are forced by the same set. In this light, we may equivalently define the Voting by Axioms rule via

$$F_{(\mathbb{A},\succ)}(R) = O \text{ iff } F(R) = O \text{ for any } F \in \mathbb{I}(\max_{\succ}\{\mathcal{A} \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A})\}). \quad (*)$$

Recall that $\succ$ is a strict linear order. Since, further, $\mathbb{A}$ is finite, there exist a unique minimal and a unique maximal element. The same holds for any non-empty subset of $\mathbb{A}$. For this maximal element, which is a set of axioms, we know by Proposition 3.1, that it forces exactly one outcome given R. Hence, every function F contained in its interpretation outputs the same outcome on R. Therefore, the given expression is well-defined and it can easily be seen that it captures the same condition as given in Definition 2.6.

Next, we turn to the first question. For the voting rule to be well-defined on the whole domain $\mathcal{L}(X)^+$, it remains to check that, or under what conditions, for all profiles $R \in \mathcal{L}(X)^+$, we can find a set of axioms $\mathcal{A} \in \mathbb{A}$ that forces some outcome given R. It is easy to see that this is not the case for every arbitrary collection of axiom sets.

Example 6. *Consider the collection $\mathbb{A}$ that only contains the set $\{\text{UNA}\}$. As seen earlier, the unanimity axiom* UNA *only speaks about, i.e., imposes conditions on, profiles where one alternative has unanimous support. That is, $\mathbb{P}(\text{UNA}) = \{R \mid \exists x \in X\, \forall i\, \forall y \neq x\, (x\ R_i\ y)\}$. Since unanimity is an intraprofile axiom, prescribing an outcome for every profile that it speaks about, we deduce that it forces an outcome exactly on all profiles in $\mathbb{P}(\text{UNA})$. This means that on every other profile, the Voting by Axioms rule is undefined due to the lack of other axiom sets that could force an outcome on the remaining profiles.*

So for which collections is the rule well-defined and how can we efficiently check for a given collection whether it is? We defined for an axiom set an associated set of profiles that the axioms force an outcome on, denoted by $\mathit{Forc}(\mathcal{A})$ (see Definition 2.5). Similarly, for a collection of sets of axioms $\mathbb{A}$, denote the set of profiles that it forces an outcome on as

$$\mathit{Forc}(\mathbb{A}) := \bigcup_{\mathcal{A}\in\mathbb{A}} \mathit{Forc}(\mathcal{A}).$$

Using this notation, we want to find out when $\mathit{Forc}(\mathbb{A}) = \mathcal{L}(X)^+$ is the case. In particular, while the assigned outcomes might differ, whether the derived voting rule is well-defined only depends on the collection $\mathbb{A}$, not on the order $\succ$.

Naively, we could verify well-definedness as follows: Compute the set of profiles on which a set $\mathcal{A}$ forces an outcome on $\mathit{Forc}(\mathcal{A})$ for all sets $\mathcal{A} \in \mathbb{A}$ and then check if every profile $R \in \mathcal{L}(X)^+$ is contained in one of these sets. We can improve this by reducing the number of sets of axioms that we need to consider. Notice that if $\mathcal{A}_1 \subseteq \mathcal{A}_2$ holds, then we have $\mathbb{I}(\mathcal{A}_1) \supseteq \mathbb{I}(\mathcal{A}_2)$. As a result, if both sets are satisfiable and $\mathcal{A}_1$ forces an outcome O on a profile, then $\mathcal{A}_2$ also forces O. Hence, with a decreasing number of axioms, it gets increasingly difficult to force an outcome due to the rising number of voting rules to be considered. Conversely, with an increasing number of axioms, we have higher chances of forcing an outcome, but if we consider too many, the extension jumps to the empty set, rendering the axioms unsatisfiable together.

Proposition 3.2. *If $\mathcal{A}_1 \subseteq \mathcal{A}_2$ and $\mathbb{I}(\mathcal{A}_2) \neq \emptyset$, then $\mathit{Forc}(\mathcal{A}_1) \subseteq \mathit{Forc}(\mathcal{A}_2)$.*

Proof. Suppose $\mathcal{A}_1$ forces $O \in \mathcal{P}_+(X)$ given R. That is, for all $F \in \mathbb{I}(\mathcal{A}_1) \neq \emptyset$, it is the case that $F(R) = O$. Since $\mathcal{A}_1 \subseteq \mathcal{A}_2$, we have $\mathbb{I}(\mathcal{A}_1) \supseteq \mathbb{I}(\mathcal{A}_2)$. But so any $F \in \mathbb{I}(\mathcal{A}_2) \neq \emptyset$ also satisfies $F \in \mathbb{I}(\mathcal{A}_1)$ and, thus, $F(R) = O$. Hence, $\mathcal{A}_2$ forces O given R. In particular, if $\mathcal{A}_1$ forces an outcome given R, then so does $\mathcal{A}_2$. Since R was arbitrary, we can conclude $\mathit{Forc}(\mathcal{A}_1) \subseteq \mathit{Forc}(\mathcal{A}_2)$. □

We can make use of this connection between subsets of axiom sets to remove redundant sets from the procedure. For a collection of sets of axioms $\mathbb{A}$, consider its subcollection of sets with non-empty extension $\{\mathcal{A} \in \mathbb{A} \mid \mathbb{I}(\mathcal{A}) \neq \emptyset\}$. We are interested in its maximal elements with respect to set-inclusion (there might be multiple ones since it is a partial order) $\max_{\supseteq}\{\mathcal{A} \in \mathbb{A} \mid \mathbb{I}(\mathcal{A}) \neq \emptyset\}$. Then by Proposition 3.2, we know that

$$\mathit{Forc}(\mathbb{A}) = \mathit{Forc}(\max_{\supseteq}\{\mathcal{A} \in \mathbb{A} \mid \mathbb{I}(\mathcal{A}) \neq \emptyset\}).$$

Recall that the well-definedness of $F_{(\mathbb{A},\succ)}$ only depends on the collection and not on the order. Thus, to check whether some outcome is forced by a set in the collection for every profile, it suffices to check if there is a forcing set among the set-inclusion maximal satisfiable sets of the collection. This is especially relevant in the special case of lifting a ranking over a corpus of axioms to a ranking over all possible axiom sets. Note that the number of possible sets of axioms built from the corpus is exponential in number of axioms belonging to

the corpus. This is a large number, so instead of checking for each of the axiom sets whether it forces an outcome on a profile, we can restrict attention to the maximal satisfiable sets.

We can weaken the aforementioned and formulate a necessary but not sufficient condition for well-definedness in terms of profiles that the involved axioms speak about. Namely, in order for Voting by Axioms to be well-defined, for every profile, some axiom in the collection should speak about it. Otherwise, no condition is imposed on the outcome of this profile — in particular, no condition that would force an outcome. That is, as a first step to check for well-definedness, we could test whether it holds that

$$\mathbb{P}(\mathbb{A}) := \underbrace{\bigcup_{\mathcal{A}\in\mathbb{A}} \mathbb{P}(\mathcal{A})}_{\supseteq Forc(\mathbb{A})} = \mathcal{L}(X)^+.$$

Integral for verifying well-definedness is being able to determine whether a set of axioms forces an outcome on some profile. While the brute-force approach to this is checking whether all voting rules satisfying the axioms give back the same outcome under the profile, this is generally not feasible. Recall that with large parameters, we obtain an extensive number of possible voting rules, making it difficult to list all voting rules belonging to an axiom's extension. Thus, we want to find ways that are computable and do not make use of an axiom's extension to determine whether an outcome is forced. We will present two ways of implementing the problem in the following — one based in propositional logic and a more proof-theoretic approach featuring a tableau-type calculus.

There exist also other approaches, e.g., one by Boixel and Endriss (2020) solving the problem by checking whether a constraint network is satisfiable. In this, to check whether outcome O is forced on profile R by some axioms, they use the axiom instances and the statement that the outcome under the profile is not O as constraints, which means that if the network is unsatisfiable, the outcome O is forced on the profile by the instances. Due to limited scalability of this approach, a different method based on graphs was introduced (Nardi, 2021; Nardi et al., 2022). In this, justifications are represented as paths in a graph, whose nodes are given by profiles that are connected via edges linking all profiles that an axiom instance speaks about. Since not all paths correspond to well-formed justifications, the task is to verify if any of the paths starting at the profile in question indeed represents a justification.

3.1.1 Forcing as Logical Consequence

The first more workable approach for determining whether a set of axioms forces an outcome on a given profile that we want to consider is translating the axioms into propositional formulas and examining the logical consequences of the axioms. The premise of this method is that if a set of axioms justifies or forces an outcome, then assigning this outcome is a necessary logical consequence of the axiom. So by means of logic, we should be able to deduce this from the

axiom's formal specification. We choose to go with propositional logic since it is expressive enough to capture voting and axioms and, at the same time, can easily be implemented for computation. SAT-solvers, which check whether some propositional formula is satisfiable (Biere, Heule, van Maaren, & Walsh, 2009), have been successfully employed in the field of computational social choice for finding new theorems and automatically proving or verifying them. An introduction and review of this method can be found in a book chapter by Geist and Peters (2017).

Nardi (2021) uses a similar way of formulating axioms in terms of a propositional language in order to derive an algorithm for generating justifications of outcomes. Consider the propositional atoms

$$P := \{p_{R,x} \mid R \in \mathcal{L}(X)^+ \text{ and } x \in X\} \cup \{\top, \bot\},$$

where we interpret $p_{R,x}$ as "alternative x is among the winners, given profile R".[6] Notice that since we work with a finite universe N^* and finitely many alternatives X, there are only finitely many propositional atoms in P. We then use the standard connectives to build a propositional language for $p \in P$

$$\mathcal{L} ::= p \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi \mid \varphi \to \varphi.$$

Notice that universal or existential quantification over voters, alternatives, outcomes and profiles can be expressed by finite conjunctions or disjunctions, respectively, since there are only finitely many of each of these objects. An important question is whether every axiom can be expressed in the language $\mathcal{L}$. This question can be read two ways: Given a formulation of the axiom in a natural or formal language, is there an appropriate translation to our propositional language? Or, alternatively, for every possible axiom extension, can we find a formula in the propositional language that exactly these voting rules satisfy? The first question is difficult to answer since it depends on what we consider an "appropriate translation". Generally, natural language descriptions tend to be more vague, requiring to use additional concepts when translating into a formal language. For instance, consider the informal description of anonymity as "all voters are treated equally" and observe that its formal counterpart makes use of permutations, specific profiles and their outcomes (see Table 2.1). Focusing on the latter interpretation of the question, we claim that for each possible extension, we can find an axiom expressed in terms of the propositional language $\mathcal{L}$ with exactly that extension. More precisely, if we have $\mathbb{I}(A)$ given, we may express the axiom by giving, for each $F \in \mathbb{I}(A)$, a full characterization in terms of propositions, i.e.,

$$\bigvee_{F \in \mathbb{I}(A)} \bigwedge_{R \in \mathcal{L}(X)^+} \Big(\bigwedge_{x \in F(R)} p_{R,x} \wedge \bigwedge_{y \in X \setminus F(R)} \neg p_{R,y} \Big).$$

It is immediate that this axiom has the same extension as A since every disjunct singles out one voting rule in $\mathbb{I}(A)$. In the following, we will assume that

[6] Instead of using propositional atoms $p_{R,x}$, we could also use atoms $q_{R,O}$ for $O \in \mathcal{P}_+(X)$, where $p_{R,x}$ corresponds to $\bigvee_{x \in O} q_{R,O}$, as featured in a paper by Cailloux and Endriss (2016).

axioms A are formulas in the language $\mathcal{L}$. An encoding of the standard axioms introduced in Table 2.1 can be found in Table 3.1, all besides REI in CNF.

Table 3.1: Propositional encoding of standard axioms

ANO	$\bigwedge_{R\in\mathcal{L}(X)^+}\bigwedge_{\sigma:N^*\overset{1:1}{\to}N^*}\bigwedge_{R'=\sigma(R)}\bigwedge_{x\in X}(\neg p_{R,x}\vee p_{R',x})$ [7]
NEU	$\bigwedge_{R\in\mathcal{L}(X)^+}\bigwedge_{\sigma:X\overset{1:1}{\to}X}\bigwedge_{R'=\sigma(R)}\bigwedge_{x\in X}(\neg p_{R,x}\vee p_{R',\sigma(x)})$ [7]
PAR	$\bigwedge_{y\in X}\bigwedge_{x\in X\setminus\{y\}}\bigwedge_{R:\forall i(xR_iy)}\neg p_{R,y}$
UNA	$\bigwedge_{x\in X}\bigwedge_{R:\forall y\neq x(\{i\mid xR_iy\}=N)} outcome(R,\{x\})$
CON	$\bigwedge_{x\in X}\bigwedge_{R:\forall y\neq x(\lvert\{i\mid xR_iy\}\rvert>\lvert N\rvert/2)} outcome(R,\{x\})$
REI	$\bigwedge_{R,R'\in\mathcal{L}(X)^+}\bigwedge_{R'':N_{R''}=N_R\sqcup N_{R'}}\Big(\bigwedge_{y\in X}(\neg p_{R,y}\vee\neg p_{R',y})\Big)$ $\vee\Big(\bigwedge_{x\in X}(\neg p_{R,x}\vee\neg p_{R',x}\vee p_{R'',x})\wedge(\neg p_{R'',x}\vee p_{R,x})\wedge(\neg p_{R'',x}\vee p_{R',x})\Big)$
CAN	$\bigwedge_{R:\forall x\neq y(\{i\mid xR_iy\}=\lvert N\rvert/2)} outcome(R,X)$
FAI	$\bigwedge_{x\in X}\bigwedge_{R:N_R=\{i\},\forall y\neq x(xR_iy)} outcome(R,\{x\})$
PR	$\bigwedge_{R\in\mathcal{L}(X)^+}\bigwedge_{x\in X}\bigwedge_{R'\in P_R}\Big((\neg p_{R,x}\vee p_{R',x})\wedge\bigwedge_{y\neq x}(\neg p_{R,x}\vee\neg p_{R',y})\Big)$, where $P_R:=\{R'\neq R\mid\forall y,z(\lvert\{i\mid y\,R_i\,z\}\rvert=\lvert\{j\mid y\,R_j{}'\,z\}\rvert$ and $\lvert\{i\mid x\,R_i\,y\}\rvert\leq\lvert\{j\mid x\,R_j{}'\,y\}\rvert)\}$

Note: Please refer to the definition of the formula $outcome(R,O)$ on page 37.

So far we have only specified a language, syntax, an assortment of meaningless symbols and given intuitive explanations for how to interpret them. Now, we want to formally define the semantics for this logic by giving conditions for when a voting rule F makes a formula φ in $\mathcal{L}$ true; denote this by $F\vDash\varphi$. Equivalently, we could say that F lies in the interpretation of φ, defined by $\mathbb{I}(\varphi):=\{F\mid F\vDash\varphi\}$, the set of voting rules that satisfy φ.

$F\vDash\top$ always
$F\vDash\bot$ never
$F\vDash p_{R,x}$ iff $x\in F(R)$
$F\vDash\neg\varphi$ iff $F\nvDash\varphi$
$F\vDash\varphi\vee\psi$ iff $F\vDash\varphi$ or $F\vDash\psi$
$F\vDash\varphi\wedge\psi$ iff $F\vDash\varphi$ and $F\vDash\psi$
$F\vDash\varphi\to\psi$ iff $F\vDash\varphi$ implies $F\vDash\psi$

[7] Please refer to Table 2.1 for the exact meaning of $R'=\sigma(R)$ in the two cases where σ is a permutation of voters or alternatives.

Given this notion of truth, we can define when a set of formulas Σ *logically entails* a formula ψ, write $\Sigma \vDash \psi$. We define this to be the case if all voting rules F that make all formulas in Σ true, also make ψ true, i.e., whenever $F \vDash \varphi$ is the case for all $\varphi \in \Sigma$, then also $F \vDash \psi$ holds. We call a set of formulas Σ satisfiable if there exists some F with $F \vDash \varphi$ for all $\varphi \in \Sigma$, so $\mathbb{I}(\varphi) \neq \emptyset$ or, equivalently, if $\Sigma \nvDash \bot$.

Next, we want to find a criterion in terms of logical consequence for when a set of axioms forces an outcome. This should be the case exactly if assigning the outcome is entailed by the axiom. In other words, a satisfiable set of axioms $\mathcal{A}$ forces an outcome $O \in \mathcal{P}_+(X)$ given R iff $\mathcal{A} \vDash \mathit{outcome}(R, O)$, where

$$\mathit{outcome}(R, O) := \bigwedge_{x \in O} p_{R,x} \wedge \bigwedge_{x \in X \setminus O} \neg p_{R,x}.$$

This formula determines for all alternatives x, whether they should win or lose given R, yielding a unique outcome set O. Thus, for a satisfiable set of axioms $\mathcal{A}$, we can write

$$\mathit{Forc}(\mathcal{A}) = \{R \in \mathcal{L}(X)^+ \mid \mathcal{A} \vDash \mathit{outcome}(R, O) \text{ for some } O \in \mathcal{P}_+(X)\}.$$

In this way, one can make use of a SAT-solver to determine whether a set of axioms $\mathcal{A}$ forces an outcome on a given profile R. To this end, we want to find out if it is possible for a voting rule to satisfy $\mathcal{A}$ and not give back O, i.e., satisfy $\neg \mathit{outcome}(R, O)$. If the answer to this is "no", then the outcome O is forced by $\mathcal{A}$ on R. Since a SAT-solver is unaware of the concept of a voting rule, we still need to encode this as a propositional formula. The constraint here is that a voting rule is a truth assignment of the propositional atoms, choosing at least one winner per profile. This corresponds to the formula

$$\mathit{atLeastOne} := \bigwedge_{R \in \mathcal{L}(X)^+} \bigvee_{x \in X} p_{R,x}.$$

Thus, for a satisfiable set of axioms $\mathcal{A}$, the formula that we want the SAT-solver to examine is

$$\mathit{atLeastOne} \wedge \Big(\bigwedge_{A \in \mathcal{A}} A \Big) \wedge \neg \mathit{outcome}(R, O)$$

for every possible outcome O. By Proposition 3.1, we know that there can at most be one such O for which the formula is unsatisfiable.

Strictly speaking, this approach still requires us to check for every voting rule in the extension of an axiom set whether it satisfies some formula. So is this procedure any better than or different from the brute-force method? Note that while in the worst case, one needs to check all possible truth assignments to find out whether a formula is satisfiable, a lot of efforts went into developing more efficient algorithms for SAT-checking. For instance, many SAT-solvers require the formula to be in CNF since the simple structure allows for more basic, efficient algorithms. The first breakthrough came with the DPPL resolution framework, considerably reducing the number of propositional atoms to be

checked (Davis, Logemann, & Loveland, 1962; Davis & Putnam, 1960). Other heuristics have been explored, many of which aim at reducing the search space of truth assignments as much as possible, or which detect classes of formulas that a quicker algorithm exists for. Biere et al. (2009) give a detailed account on the developments of Boolean satisfiability. Taking this into account, we can conclude that the propositional logic encoding, using a state-of-the-art SAT-solver, is an efficient way of determining whether the Voting by Axioms rule is well-defined.

3.1.2 Detecting Forcing via Tableaux

One way of telling whether the Voting by Axioms rule is well-defined is by computing the sets $\mathit{Forc}(\mathcal{A})$ for sets of axioms $\mathcal{A}$ in the collection $\mathbb{A}$. While we described what these sets contain with help of the logical consequence relation in a propositional logic in Section 3.1.1 and this is enough to hand the problem over to a SAT-solver, we now want to describe a more direct, algorithmic, (human) computable method. Boixel et al. (2022) introduced a calculus based on the tableau method to prove that a given set of axiom instances explains or justifies a certain outcome.[8] While the authors designed the calculus to find one among possibly many explanations for assigning a specific outcome, we want to use the proof system to find out whether a set of axioms forces any outcome on a profile. In the following, we present a slightly adjusted version of the calculus, in which we test whether a set of axioms (rather than a set of axiom instances) forces an outcome.

The main goal of the calculus is to check whether a given set of axioms is consistent with the requirement that a voting rule must return a specific outcome under some given profile. To express this, we need to formally represent the statement that a rule should assign a certain outcome.

Definition 3.3. An *outcome statement* is a tuple $s = (R, \mathcal{O})$, where R is any profile in $\mathcal{L}(X)^+$ and $\mathcal{O}$ is a set of outcomes $O \in \mathcal{P}_+(X)$.

The *interpretation* (or *extension*) $\mathbb{I}(s)$ of an outcome statement consists of all voting rules that make the the statement true.

We interpret an outcome statement s as the requirement that a voting rule should return an outcome inside $\mathcal{O}$ for the profile R. Thus, for every rule $F \in \mathbb{I}(s)$, we have $F(R) \in \mathcal{O}$. We can also consider a set of outcome statements S and define accordingly

$$\mathbb{I}(S) := \bigcap_{(R,\mathcal{O})\in S} \{F : \mathcal{L}(X)^+ \to \mathcal{P}_+(X) \mid F(R) \in \mathcal{O}\}.$$

This looks familiar from the propositional logic representation in Section 3.1.1. Notice that an outcome statement $s = (R, \mathcal{O})$ corresponds exactly to the formula

[8] A good starting point to learn about tableau calculi for propositional logic is the book chapter by D'Agostino (1999).

$\bigvee_{O \in \mathcal{O}} outcome(R, O)$, or equivalently, with the alternative propositional atoms defined in Footnote 6, to $\bigvee_{O \in \mathcal{O}} q_{R,O}$.

The objective of this calculus is to prove that, for some given set of axioms $\mathcal{A}$ and a set of outcomes statements S, the intersection $\mathbb{I}(\mathcal{A}) \cap \mathbb{I}(S)$ is empty. This means that there is no voting rule satisfying all axioms in $\mathcal{A}$ that simultaneously makes all outcome statements in S true. Boixel et al. (2022) explain how this can be used to obtain impossibility results, to explain why some voting rule violates an axiom or to find an explanation for assigning a certain outcome. We focus on the third use case.

For this, we set $S := \{(R, \mathcal{P}_+(X) \setminus \{O\})\}$, which corresponds to checking whether $\mathcal{A}$ forces the outcome O given the profile R. We claim that for a satisfiable set of axioms $\mathcal{A}$, if $\mathbb{I}(\mathcal{A}) \cap \mathbb{I}(S) = \emptyset$ holds for this choice of S, then $\mathcal{A}$ forces O given R. This, again, is in correspondence to how we proceeded in the case of propositional logic. Namely, we try to find out whether there is a rule satisfying $\mathcal{A}$ but not assigning the outcome O to profile R. If there is not, i.e., if $\mathcal{A}$ and S are inconsistent, then all voting rules in $\mathbb{I}(\mathcal{A})$ must assign O to R, i.e., the axioms force the outcome. For this to work out, or more precisely, for being able to conclude from the inconsistency of $\mathcal{A}$ and S that $\mathcal{A}$ forces O, it is crucial that we consider a satisfiable axiom set $\mathcal{A}$ (recall from Definition 2.5 that it was a precondition for forcing that the axiom set is consistent). Although the basic idea is the same, instead of handing the problem over to a SAT-solver, we will now describe a calculus to answer this question directly.

A *tableau* is a rooted tree, in which every node is a set of outcome statements. There is an initial starting node from which a tree is constructed by repeatedly applying *expansion rules* that add one or two new child nodes below one of the leaf nodes of the current tableau. To show that $\mathbb{I}(\mathcal{A}) \cap \mathbb{I}(S) = \emptyset$ holds, we define the root node to be $S_0 := S$. The expansion rules are going to add outcome statements that are implied by instances of the axioms in $\mathcal{A}$ to the tableau. We continue this process until a contradiction in the outcome statements is apparent or until we exhausted all instances. In this way, we logically chain axiom instances together either until we obtain an argument for why the axioms imply the outcome statements or until we run into a contradiction, either among the axioms (this cannot be the case if $\mathcal{A}$ is satisfiable) or between the axioms and one of the outcome statements in the root node. The possible rules to be applied to a node in the tableau are the following:

- **Axiom-driven expansion rule:** For any axiom instance $A' \triangleleft \mathcal{A}$ and any profile $R \in \mathbb{P}(A')$, a branch ending in a node S' can be extended by adding

  $$S'' := S' \cup \{(R, \mathcal{O})\} \text{ with } \mathcal{O} := \{F(R) \mid F \in \mathbb{I}(S') \cap \mathbb{I}(A')\}$$

  as its child node, given that $\mathbb{I}(S'') \subsetneq \mathbb{I}(S')$.

- **Branching rule:** If the leaf node S' of a branch contains an outcome statement of the form $(R, \mathcal{O}_1 \sqcup \mathcal{O}_2)$ for non-empty sets $\mathcal{O}_1, \mathcal{O}_2$, then we can add two child nodes to S', one for each set of outcomes. That is, we add $S'_1 := S' \setminus \{s\} \cup \{(R, \mathcal{O}_1)\}$ and $S'_2 := S' \setminus \{s\} \cup \{(R, \mathcal{O}_2)\}$ to the

tableau. To apply this rule, we may assume that S' contains the trivial statement $(R, \mathcal{P}_+(X))$.

- **Simplification rule:** If a branch ends in a node S' that contains multiple outcome statements for the same profile R, for instance, $s_1 = (R, \mathcal{O}_1)$ and $s_2 = (R, \mathcal{O}_2)$, then we may add the node $S'' := S' \setminus \{s_1, s_2\} \cup \{(R, \mathcal{O}_1 \cap \mathcal{O}_2)\}$ to the tableau.

The first rule adds, for a specific profile, the constraints that an axiom instance in $\mathcal{A}$ imposes on it in terms of an outcome statement. It combines the outcome statements that are already present at the current node with one more condition implied by $\mathcal{A}$, which might lead to an outcome statement of the form $(R, \emptyset)$, if the axiom instance is in contradiction to the present outcome statements. We call a leaf node *inconsistent* if it contains at least one such inadmissible outcome statement. The branching rule makes a case distinction for a present outcome statement $(R, \mathcal{O})$ by separating the set of allowed outcomes $\mathcal{O}$ into two disjoint subsets. The third rule, in contrast, allows to merge multiple outcome statements about the same profile and, thereby, helps making inconsistencies apparent.

We continue applying expansion rules to a tableau until we cannot apply any rule to any of the leaf nodes anymore. This will happen eventually since whenever we apply a rule, either the extension strictly decreases from parent to child node (for the axiom-driven expansion rule and the branching rule) or the node size, meaning the size of the set of outcome statements corresponding to the node, decreases (for the simplification rule). Since all sets are finite, we can only finitely many times apply such rules. If we cannot apply any rules anymore, we call the tableau *saturated*. We say that such a tableau is *closed* if, for each of its branches, every leaf node is inconsistent. Otherwise, call it *open*. For a given set of axioms $\mathcal{A}$ and an initial set of outcome statements S, we call a tableau constructed as above a tableau *rooted in* S and *licensed by* $\mathcal{A}$.

In their paper, Boixel et al. (2022) prove that the tableau calculus is correct for determining whether there is a voting rule satisfying both a given set of axioms and a given set of outcome statements, i.e., it is a sound and complete system. In other words, there exists no such voting rule if and only if we can find a proof (that is, a closed tableau) showing that the intersection of the interpretations is empty. They also suggest a way of implementing the procedure using answer set programming (ASP), a declarative programming paradigm suited for search problems based in logic programming (Brewka, Eiter, & Truszczyński, 2011).

How can we make use of this proof system in determining whether the Voting by Axioms rule $F_{(\mathbb{A},\succ)}$ is well-defined? In the beginning of Section 3.1, we concluded that it suffices to compute the set of profiles that an outcome is forced on $Forc(\cdot)$ for set-inclusion maximal satisfiable sets in $\mathbb{A}$. The strategy now is to determine the set of profiles that an outcome is forced on for one of the set-inclusion maximal satisfiable sets $\mathcal{A}_0$ by applying the calculus for each profile R. For one outcome O at a time, we will use $S = \{(R, \mathcal{P}_+(X) \setminus \{O\})\}$ as root node to construct a tableau licensed by $\mathcal{A}_0$ and check whether it is closed until

we find an outcome that is forced given R or until we exhausted all outcomes. We can then continue in the same manner to check for the next set-inclusion maximal set $\mathcal{A}_1$, whether and on which profiles in $\mathcal{L}(X)^+ \setminus \mathit{Forc}(\mathcal{A}_0)$ it forces an outcome. We continue in this fashion until either no non-forced profiles are left or until we run out of axiom sets $\mathcal{A}_i \in \max_{\supseteq}\{\mathcal{A} \in \mathbb{A} \mid \mathbb{I}(\mathcal{A}) \neq \emptyset\}$ to consider. In the former case, the rule is well-defined, in the latter, it is not.

This method requires calculating many tableaux and it would be more elegant to conflate them into a single tableau. However, with the given system, this is not possible. Since the tableau checks for satisfiability of statements, it answers the question whether a voting rule with all considered properties exists or not. But if we do not want to loop over outcomes to determine whether on a given profile an axiom set forces an outcome, this is not possible by asking whether there exists *at least one* voting rule with certain properties, as just one single question. Instead, the suitable question to ask is whether there exist *at least two* voting rules satisfying the axioms, assigning two different outcomes to the profile. Thus, even when focusing on a single profile, there is not one voting rule that can answer the question whether some outcome is forced — this problem requires inspecting multiple rules. Similarly, since in Voting by Axioms we look at each profile independently, it is not helpful to check for multiple profiles at the same time whether a specific outcome combination is forced.

Observe that while we assumed in the calculus that a notion of axiom instance is given, we did not use any information about how exactly axiom instances have been defined or derived. We only assumed that a division of axiom into instances is fixed and we know their interpretations and which profiles they speak about. Thus, whatever way of obtaining instances we choose, if we find a closed tableau for this one, we know, in general, that an outcome is forced by the set of axioms. We can then use this information and forget about the exact determination of axiom instances.

This concludes our study of well-definedness of the Voting by Axioms rule. We have stated precisely what this notion depends on in this setting and offered two computable methods for checking this. Now, assuming that the rule is well-defined, we want to examine its behavior and work out conditions for when it has good properties.

3.2 Axiomatic Analysis

It is now time to analyze the voting rule that we defined. The standard way to evaluate and compare voting rules is the axiomatic method (Plott, 1976; Thomson, 2001; Zwicker, 2016). This means testing for different axioms, whether the rule satisfies them. Applied to our setting, we are interested in finding out, if or under what conditions the rule $F_{(\mathbb{A},\succ)}$ satisfies axioms occurring in $\mathbb{A}$. Ideally, the derived rule would satisfy all axioms in the collection or at least all the ones that were used to force an outcome in the construction of the rule. However, this will only very rarely be the case. The main problem is that we defined the rule profile by profile, forgetting about most interprofile conditions that the

axioms impose. We pointed out that if an axiom forces an outcome, assigning this outcome is a necessary condition for satisfying the axiom. In other words, this outcome statement is a weakening of the axiom. In general, this condition is not satisfactory for a rule to satisfy the original axiom, i.e., it could be a strict weakening of the axiom. We want to work out in which cases we can deduce that the Voting by Axioms rule fulfills axioms occurring in the collection of axiom sets. In principle, $F_{(\mathbb{A},\succ)}$ can also satisfy axioms that do not occur in $\mathbb{A}$. But this cannot be established in general and should rather be checked for a concrete rule $F_{(\mathbb{A},\succ)}$ using the standard tools of the axiomatic method.

3.2.1 Respecting Axiom Instances

The first special case that we want to consider is that of intraprofile axioms. They work well together with Voting by Axioms since they, too, look at each profile independently. So in this case, the derived rule does not only respect a subset of the conditions imposed by the axiom, but it fulfills the whole axiom.

Theorem 3.4. *If A is an intraprofile axiom and it is contained in the maximal forcing set in $\mathbb{A}$ according to $\succ$ for every profile, i.e., if for all $R \in \mathbb{P}(A)$, we have $A \in max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}$, then $F_{(\mathbb{A},\succ)}$ satisfies A.*

Proof. Notice that to satisfy an intraprofile axiom, it suffices to check for each profile that it speaks about in isolation, whether the outcome assigned to it is among the outcomes permitted by the axiom. For an arbitrary profile $R \in \mathbb{P}(A)$, we know that $A \in \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}$, so by definition of the Voting by Axioms rule, $F_{(\mathbb{A},\succ)}(R) = F(R)$ for every $F \in \mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}) \subseteq \mathbb{I}(A)$. In particular, there exists some $F \in \mathbb{I}(A)$ with $F_{(\mathbb{A},\succ)}(R) = F(R)$. This shows that the outcome under the derived voting rule is admissible with respect to A. Since R was an arbitrary profile that the axiom speaks about, the same holds across all of $\mathbb{P}(A)$. Because the axiom does not impose any conditions on all other profiles, we conclude that $F_{(\mathbb{A},\succ)}$ satisfies A. □

Notice that it is not enough if A is merely contained in the maximal satisfiable set of the collection. The reason for this is that an intraprofile axiom does not necessarily force an outcome on all profiles that it speaks about. Importantly, if the axiom does not force an outcome on a profile, we cannot infer that the axiom does not speak about this profile at all. Consider, for instance, the Pareto principle which says for specific profiles that a certain alternative is not among the winners. In general, this still allows for multiple different outcomes to be assigned to the profile. Thus, if we only know that A is contained in the maximal satisfiable set, then it might happen that this set does not force an outcome on some profile but another set in the collection that does not contain A does force an outcome. Then, this outcome is assigned to the profile under $F_{(\mathbb{A},\succ)}$. But this might not be an admissible outcome with respect to A, e.g., if it contains some alternative that should not win according to A. In this case, the Voting by Axioms rule would violate the intraprofile axiom A.

Further, we want to stress that the proposition does, in general, not apply if we consider an interprofile axiom.

Example 7. *Let* $\mathbb{A}$ *be given by* $\{\text{CAN}, \text{REI}\} \succ \{\mathbf{1}, \text{REI}\}$*, where the axiom* **1** *says that the voting rule gives back* $\{1\}$ *for all profiles. First, notice that both sets of axioms are satisfiable, by the constant functions assigning* X *and* $\{1\}$*, respectively. For this choice of* $\mathbb{A}$*, the derived rule is well-defined since the second set of axioms forces an outcome on every profile. Trivially, reinforcement is contained in every maximal forcing set. Observe that* $\{\text{CAN}, \text{REI}\}$ *does not force any outcome on the profiles* $(213, 132)$ *and* $(231, 312)$*. Therefore,* $F_{(\mathbb{A},\succ)}(213, 132) = F_{(\mathbb{A},\succ)}(231, 312) = \{1\}$*. So then, the intersection* $F_{(\mathbb{A},\succ)}(213, 132) \cap F_{(\mathbb{A},\succ)}(231, 312)$ *is just* $\{1\}$*. However, by cancellation, the Voting by Axioms rule* $F_{(\mathbb{A},\succ)}$ *returns* X *for the profile* $(213, 132, 231, 312) = (213, 132) \cup (231, 312)$*, instead of* $\{1\}$*. This shows that the derived function does not satisfy reinforcement, which is an interprofile axiom, although it is contained in all sets in the collection.*

The issue that we uncovered in the previous example is that for interprofile axioms, the Voting by Axioms rule only guarantees that on every profile, a generally feasible outcome is assigned, not how these outcomes relate to each other. More precisely, for an interprofile instance A', the Voting by Axioms rule assigns to each $R \in \mathbb{P}(A')$ an outcome O such that $F(R) = O$ for some $F \in \mathbb{I}(A')$. However, it disregards whether the derived rule itself lies in $\mathbb{I}(A')$. So next, we want to consider a case, in which for all instances of an axiom, we can guarantee that the Voting by Axioms rule satisfies them.

Theorem 3.5. *If for an axiom* A *it holds that for all instances* $A' \triangleleft A$*, there exists some* $\mathcal{A} \in \mathbb{A}$ *with* $A' \triangleleft \mathcal{A}$ *such that for all profiles* $R \in \mathbb{P}(A')$*, we have* $max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\} = \mathcal{A}$*, then the rule* $F_{(\mathbb{A},\succ)}$ *satisfies* A*.*

Proof. In order to tell whether the derived rule $F_{(\mathbb{A},\succ)}$ satisfies the axiom A, we need to check if it satisfies all instances $A' \triangleleft A$. By assumption, we know that for a given instance A', there is a set of axioms $\mathcal{A}$ such that for all profiles $R \in \mathbb{P}(A')$, all rules $F \in \mathbb{I}(\mathcal{A}) \subseteq \mathbb{I}(A')$ have the same outcome $F(R) = F_{(\mathbb{A},\succ)}(R)$. Fix one such rule F and observe that a voting rule satisfies some instance A' if and only if its restriction to $\mathbb{P}(A')$ satisfies the instance. Since $F_{(\mathbb{A},\succ)}{\restriction}_{\mathbb{P}(A')} = F{\restriction}_{\mathbb{P}(A')} \in \mathbb{I}(A')$, the Voting by Axioms rule satisfies A'. Since $A' \in A$ was an arbitrary instance, the derived rule $F_{(\mathbb{A},\succ)}$ satisfies A. □

Again, notice that we did not use any information about axiom instances besides the minimal requirements defined in Section 2.2. That is, no matter how exactly we define axiom instances or which way of splitting an axiom into multiple instances we use, the result holds.

3.2.2 Using Characterization Results

The social choice literature is full of characterization results, e.g., May's (1952) Theorem or Young's (1975) characterization of positional scoring rules. These

are theorems stating that a voting rule satisfies a certain set of axioms exactly if it coincides with a specified voting rule or if it lies in some class of rules. In other words, such a theorem yields necessary and satisfactory conditions for a rule to agree with one particular voting rule or to belong to a special class of voting rules. We expect this kind of result to be helpful in the context of Voting by Axioms since it tells us that the given axioms induce a certain behavior (meaning that it is more likely that outcomes are forced by the axioms) and that if the derived rule follows this behavior, it will satisfy the axioms (meaning that it is more likely to obtain a good rule). In the following, we will show that, under a few additional conditions, if we give high priority to a characterizing axiom set in our collection, the derived Voting by Axioms rule will be the characterized rule or lie in the characterized class.

General Characterization Results. We say that a set of axioms $\mathcal{A}$ *uniquely characterizes* a voting rule F if $\mathbb{I}(\mathcal{A}) = \{F\}$. It is immediate by Proposition 3.2 that if a set $\mathcal{A}$ uniquely characterizes a function F, then any satisfiable superset of $\mathcal{A}$ also characterizes F. First, observe that if $\mathbb{A}$ contains a set of axioms that uniquely characterizes some voting rule, then the derived rule $F_{(\mathbb{A},\succ)}$ is well-defined. This is because this set of axioms forces an outcome on every profile, so there exists a maximal outcome-forcing set for every profile in the collection. If we can guarantee that the characterizing set of axioms (or its supersets) appear(s) sufficiently high in the ranking of $\mathbb{A}$, we can even conclude that the derived voting rule is precisely the characterized one.

Theorem 3.6. *If a set of axioms $\mathcal{A}$ uniquely characterizes a voting rule F, then for any collection of sets of axioms $\mathbb{A}$, and for any order $\succ \in \mathcal{L}(\mathbb{A})$ such that for every R, we have $\mathcal{A} \subseteq max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}$, the derived rule $F_{(\mathbb{A},\succ)}$ is the rule F itself.*

Proof. By definition, for each profile R, the function $F_{(\mathbb{A},\succ)}$ assigns that outcome to the profile that is forced by the highest-ranked forcing set in the collection. Because $\mathcal{A}$ is a subset of $\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}$, by Proposition 3.2, we know that $\mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}) \subseteq \{F\}$ for every R. Since the maximal forcing set is satisfiable (this is a precondition of forcing an outcome), we know that equality must hold. Thus, using the alternative definition from $(*)$, we may infer that $F_{(\mathbb{A},\succ)}(R) = F(R)$ for every R. This is what we wanted to show. □

Notice that it suffices to assume that $\mathcal{A}$ is contained in the highest-ranked set of the collection that forces some outcome on some profile, i.e., that $\mathcal{A} \subseteq \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid Forc(\mathcal{A}') \neq \emptyset\}$ holds. That is the case because it implies that this set uniquely characterizes F and, for each profile, is the highest-ranked forcing axiom set. Further, we can strengthen this result a little bit. To this end, call $\mathcal{A}$ *uniquely and minimally characterizing* for F if it uniquely characterizes F and if it is minimal with this property, i.e., for all $\mathcal{A}' \subsetneq \mathcal{A}$, the interpretation $\mathbb{I}(\mathcal{A}')$ contains at least two rules. Notice that there may exist multiple sets of axioms that all uniquely and minimally characterize the same voting rule

F. And given some characterizing set of axioms, we can always find at least one minimally characterizing subset. Recall that what we actually used in the proof to deduce that the unique function satisfying all axioms in the maximal forcing sets is F, was that $\mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}) \subseteq \{F\}$ holds for all profiles R. We derived this from the assumption that $\mathcal{A} \subseteq \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ is true. However, it suffices to assume something weaker, namely that there is some minimally characterizing subset of axioms $\mathcal{A}' \subseteq \mathcal{A}$ for F such that $\mathcal{A}'$ is a subset of $\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ for all profiles.

We want to note that it is sometimes also possible to go the other way round. That is, knowing that Voting by Axioms is well-defined, we can in some cases deduce that the derived rule is uniquely characterized by a set of axioms. The crucial condition here is that all axioms that we used to define the outcomes of the rule $F_{(\mathbb{A},\succ)}$ must be satisfiable together. This means that the set containing all these axioms forces an outcome on every profile (since for every profile, we know that some subset of it does). Thus, if $\mathbb{A}$ is such that Voting by Axioms is well-defined, and further, we consider a ranking $\succ$ such that $\bigcup_{R\in\mathcal{L}(X)^+} \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ is satisfiable, then this union uniquely characterizes the rule $F_{(\mathbb{A},\succ)}$.

We now want to look at the case when axioms determine a class of voting rules. We say that a set of axioms $\mathcal{A}$ *uniquely characterizes* a non-empty class of rules $\mathcal{F}$, if a rule satisfies all the axioms in $\mathcal{A}$ if and only if it belongs to $\mathcal{F}$. In other words, we require that $\mathbb{I}(\mathcal{A}) = \mathcal{F}$. In this case, we want to find a condition telling us when the derived voting rule lies in the class $\mathcal{F}$. The problem here is that if axioms characterize a class of rules, this does not mean that they force any outcome on any profile. For instance, the class of positional scoring rules that Young (1975) characterized only forces an outcome on symmetric profiles. Thus, on the remaining profiles, other sets in the collection will determine the outcome of the Voting by Axioms rule, making it improbable for the rule to coincide with a rule in the characterized class. However, if we include further axioms in the collection that help single out one voting rule from the characterized class, we can guarantee that the derived rule lies in the class.

Theorem 3.7. *If a set of axioms $\mathcal{A}$ uniquely characterizes a class of voting rules $\mathcal{F}$, then for any collection of sets of axioms $\mathbb{A}$, and for any order $\succ \in \mathcal{L}(\mathbb{A})$ such that for every R, we have $\mathcal{A} \subseteq \mathit{max}_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ and the intersection $\bigcap_{R\in\mathcal{L}(X)^+} \mathbb{I}(\mathit{max}_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\})$ is non-empty, $F_{(\mathbb{A},\succ)}$ lies in $\mathcal{F}$.*

Let us clarify this statement. The first condition guarantees that, for every profile, its outcome under the derived rule agrees with the outcome of some rule in $\mathcal{F}$. The second condition additionally ensures that there is one single rule that satisfies the axioms $\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ for all profiles $R \in \mathcal{L}(X)^+$ simultaneously. Since, for each profile, one of these sets (supersets of $\mathcal{A}$) forces an outcome, it means that the union forces an outcome on all profiles and, thereby, uniquely characterizes one single rule inside of $\mathcal{F}$. This is exactly the Voting by Axioms rule, so it lies in $\mathcal{F}$. We will now provide the full formal proof of the statement.

Proof. Since $\mathcal{A} \subseteq \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ is the case, we also have $\mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}) \subseteq \mathbb{I}(\mathcal{A}) = \mathcal{F}$. Thus, taking into account that the intersection of extensions is nonempty, we can conclude that

$$\bigcap_{R \in \mathcal{L}(X)^+} \mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}) \cap \mathcal{F} \neq \emptyset.$$

But also notice that rules in $\bigcap_{R \in \mathcal{L}(X)^+} \mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\})$ force an outcome on every profile, thus the intersection is given by $\{F\}$ for some $F \in \mathcal{F}$. In particular, $F \in \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ for every R. Hence, by the alternative definition of $F_{(\mathbb{A},\succ)}$ given in $(*)$, we have $F_{(\mathbb{A},\succ)}(R) = F(R)$ for every R, meaning that $F_{(\mathbb{A},\succ)} = F \in \mathcal{F}$. This concludes the proof. □

Similarly as before, we can define minimally characterizing sets of axioms for a class of rules. This allows us to weaken the assumption that $\mathcal{A}$ is a subset of $\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ for every profile R to there being some minimally characterizing subset of axioms $\mathcal{A}' \subseteq \mathcal{A}$ for $\mathcal{F}$ that is a subset of $\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}$ for every profile.

Notice that Theorem 3.6 can be derived from Theorem 3.7, because if $\mathcal{A}$ characterizes a rule F, this means that is characterizes the class $\mathcal{F} := \{F\}$. The second condition of Theorem 3.7 is vacuously satisfied in this case since we have $\mathbb{I}(\max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A}')\}) = \{F\}$ for every profile.

Domain-restricted Characterization Results. In social choice theory, we are not always able to provide a full characterization of a voting rule. For instance, May's Theorem, which characterizes the simple majority rule, only applies in case there are exactly two alternatives (May, 1952). Or Moulin (1980) showed that strategyproofness and the axiom "tops-only" (i.e., a rule only uses the information which alternative is ranked highest) characterize the class of min-max voting rules on the domain of single-peaked preferences (Weymark, 2011). We might even want to consider subproblems of voting, e.g., voting on combinatorial domains such as committee elections or group planning problems (Lang & Xia, 2016). We want to be able to also make use of such partial characterization results in the context of Voting by Axioms.

It might occur that a theorem states that a certain rule, defined as a function on a subset of the whole domain, is the only rule satisfying some particular set of axioms. We want to use this theorem to deduce what our Voting by Axioms rule (defined on the whole domain) looks like on this subset of the domain.

For this, we need to make sure that the axioms are phrased in terms of voting rules defined on the whole domain $\mathcal{L}(X)^+ \to \mathcal{P}_+(X)$ and that they are satisfiable on the whole domain. If A is an axiom that only applies to voting rules defined on a subdomain $P \subseteq \mathcal{L}(X)^+$, this can easily be achieved by requiring for all voting rules in $\mathcal{L}(X)^+ \to \mathcal{P}_+(X)$ that their restriction to P satisfies A. For instance, if we view faithfulness as an axiom on the domain of single-voter profiles, it would say "For all profiles, the highest-ranked alternative wins". To make it an axiom on the whole domain, we transform it to "For all single-voter

profiles, the highest-ranked alternative wins". In this case, if the original axiom is satisfiable, so is the axiom constructed on the whole domain. It can also be the case, however, that the axioms appearing in the partial characterization result describe principles applying to profiles beyond the subdomain P. In this case, it is not a priori clear that, only because there is a voting rule satisfying all axioms on P, any rule would satisfy the axioms across all profiles. For such axioms to play any role in Voting by Axioms and to force an outcome, we need them to be satisfiable though. If they are not globally satisfiable, we may use a trick and replace the axioms by copies of themselves, but only keeping those conditions that apply to profiles in P. These replacement axioms are then satisfiable on the whole domain.

Proposition 3.8. *Let $\mathcal{A}$ with $\mathbb{I}(\mathcal{A}) \neq \emptyset$ be a set of axioms uniquely characterizing some voting rule F on a subdomain $P \subseteq \mathcal{L}(X)^+$, i.e., $\{F'\restriction_P \mid F' \in \mathbb{I}(\mathcal{A})\} = \{F\}$. If $\mathcal{A} \in \mathbb{A}$, then $Forc(\mathbb{A}) \supseteq P$ is the case. If further, for every $R \in P$, we have $\mathcal{A} \subseteq max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in Forc(\mathcal{A}')\}$, then the derived rule $F_{(\mathbb{A},\succ)}$ coincides with F on P.*

The proof is analogous to the one of Theorem 3.6, just restricting attention to the profiles in P. This is possible since the Voting by Axioms rule is defined for each profile independently.

We saw that characterization results are very useful to obtain a well-defined and well-behaving Voting by Axioms rule. But how do we find such results? We want to show for a subclass of axioms how to obtain an axiom set that characterizes a voting rule.

Using Algebraic Axioms. An interesting subclass of axioms to consider is that of *algebraic axioms*. The main idea is to define basic or atomic axioms, which are simple and represent the minimal possible restrictions that normative principles can impose, and to build up more complex axioms from them. Due to the specific structure that these axioms exhibit, we have a more direct way of checking whether they force an outcome, given some profile. This idea is due to Kaminski (2004).

Definition 3.9. There are three kinds of *basic algebraic axioms*:

- A *basic stationary axiom* is of the form "The outcome under R is O." for some fixed $R \in \mathcal{L}(X)^+$ and $O \in \mathcal{P}_+(X)$
- A *basic variance axiom* says "If R_1 is assigned to O_1, then R_2 is assigned to O_2." for some fixed $R_1, R_2 \in \mathcal{L}(X)^+$ and $O_1 \neq O_2 \in \mathcal{P}_+(X)$.
- A *basic invariance axiom* is given by "If the outcome under R_1 is O, then the same is true for R_2." for some fixed $R_1, R_2 \in \mathcal{L}(X)^+$ and $O \in \mathcal{P}_+(X)$.

An *algebraic axiom* is a set of basic algebraic axioms. Any such axiom that consists only of basic stationary (variance, invariance) axioms is also called a stationary (variance, invariance) axiom.

Notice that axioms featured in the social choice literature that are algebraic are usually conjunctions of multiple basic algebraic axioms. For instance, unanimity is a stationary axiom specifying "For all R such that $R_i = x$ for all voters i and some $x \in X$, the outcome under R is $\{x\}$". It consists of $m \cdot (2^n - 1)$ basic stationary axioms, one for each x and every electorate of every size. We usually refer to these basic axioms as *instances* of the compound axiom. Examples for invariance and variance axioms are anonymity and neutrality, respectively.

We can easily express these axioms in terms of the propositional language $\mathcal{L}$ that we defined in Section 3.1.1. A basic stationary axiom corresponds to "$outcome(R, O)$". A basic variance axiom is formalized as "$outcome(R_1, O_1) \rightarrow outcome(R_2, O_2)$" and, similarly, a basic invariance axioms is expressed by "$outcome(R_1, O) \rightarrow outcome(R_2, O)$". A complex algebraic axiom then consists of conjunctions of these basic axioms.

Regarding Voting by Axioms, stationary axioms directly stipulate on certain profiles which outcome should be assigned, thus, they force outcomes. So when deciding on a collection $\mathbb{A}$, more specifically on a set of axioms $\mathcal{A}$ in it, the role of a stationary axiom is to contribute a set of profiles that an outcome is forced on. These axioms should be accompanied by (in-)variance axioms since these can increase the domain of profiles that an outcome is forced on $Forc(\mathcal{A})$ by applying Modus Ponens. We will now see that if we have enough stationary axioms to begin with and suitable (in-)variance axioms to extend the forcing of outcomes across the whole domain, we obtain a uniquely characterizing set of axioms.

This is the idea behind the following theorem by Kaminski (2004, Theorem 1) which tells us how to obtain an axiom set (to include in our collection $\mathbb{A}$) that forces an outcome on every profile. We need the following terminology in order to phrase it. Observe how any voting rule partitions the set of all profiles into cells on which the outcome is constant. More precisely, define the partition imposed by $F \in \mathcal{L}(X)^+ \rightarrow \mathcal{P}_+(X)$ as

$$Part(F) = \{P \subseteq \mathcal{L}(X)^+ \mid P = F^{-1}(O) \neq \emptyset \text{ for some } O \in \mathcal{P}_+(X)\}.$$

For any profile R and a given partition, we denote by $[R]$ the cell in the partition that R belongs to. We call $\{R_1, \ldots, R_{|Part(F)|}\}$ a selection from $Part(F)$ if it includes exactly one representative of every cell in the partition.

Theorem 3.10. *Let $\{R_1, \ldots, R_k\}$ be a selection from $Part(F^*)$. A set of algebraic axioms $\mathcal{A}$ uniquely characterizes a voting rule F^* if and only if it implies the following three axioms:*

(A_1) *The outcome under R_1 is $F^*(R_1)$*

(A_2) *For all $i = 2, \ldots, k$, if R_1 is assigned to $F^*(R_1)$, then R_i is assigned to $F^*(R_i)$*

(A_3) *For all $i = 1, \ldots, k$ and all $R \in \mathcal{L}(X)^+$ such that $R \in [R_i]$, the outcome under R is the same as under R_i*

In words, the first axiom forces the outcome $F^*(R_1)$ given the profile R_1. The second axiom extends this to say that for all profiles R_i, the outcome $F^*(R_i)$ is forced. The third axiom completes this by saying that the function has to be constant on the cells in $Part(F^*)$. Together, this forces any rule satisfying the axioms to be identical to F^*.

This means that, in particular, if a set of axioms $\mathcal{A}$ implies the axioms A_1, A_2, A_3, then the derived voting rule $F_{(\mathbb{A},\succ)}$ is well-defined for all collections containing either $\mathcal{A}$ or a superset of $\mathcal{A}$ with non-empty extension. Further, by Theorem 3.6, if the characterizing axiom set is ranked sufficiently high, we can deduce that the derived function $F_{(\mathbb{A},\succ)}$ is F^* itself.

A shortcoming of this theory is that it does not take into account all axioms, so the class of all algebraic axioms is a strict subclass of all set-theoretic axioms, given by all possible interpretation sets $\mathbb{I}(A) \subseteq (\mathcal{L}(X)^+ \to \mathcal{P}_+(X))$. For example, the Pareto Principle is not an algebraic axiom since it does not specify full outcome sets but only says for some alternatives if they should or should not be among the winners. Further, any axiom that talks about more than two profiles at once, e.g., reinforcement, is not an algebraic axiom in the defined sense.

One way of generalizing the idea of algebraic axioms is to allow for the disjunction of basic axioms. In this way, one can allow for axioms of the form "If ..., then x is among the winners of R." This would correspond to the disjunction of basic stationary axioms for R and all outcomes O that include x. However, this makes the calculus much more complicated since we need to make case distinctions. We have seen a systematic approach for how to handle these statements and how to find out whether a set of axioms forces an outcome, namely the tableau calculus by Boixel et al. (2022), see Section 3.1.2.

We showed that in rare cases, some of the axioms that we take as a basis are inherited by the Voting by Axioms rule. This means that, seen as a general voting rule across the whole domain, it seldom satisfies desirable principles. Nonetheless, we want to emphasize that we should not only measure the quality of the rule in terms of satisfaction of global principles. Recall that it is the very idea behind justifying outcomes by axioms to look at profiles independently and allow that only a part of the conditions of an axiom is satisfied. So if whether the Voting by Axioms rule overall satisfies or does not satisfy the underlying axioms is not suitable for assessing its quality, how else can we measure its performance?

3.3 Metrics

In this section, we want to present two metrics that help evaluating Voting by Axioms. The first kind measures to what extent a voting rule satisfies a given set of axioms, whereas the second one quantifies how close an axiom set is to forcing an outcome on a given profile.

Satisfaction Metrics. As mentioned earlier, we want to define a notion of gradual satisfaction. Instead of having a binary measure of satisfaction, we

want to come up with a more fine-grained metric.[9] With the notion of axiom instances at hand, a natural metric arises. For a given axiom set, a *satisfaction metric* measures for each voting rule, to what extent (as a number between 0 and 1) the axioms are satisfied.

We can define a satisfaction metric for a set of axioms $\mathcal{A}$ with axiom instances A' such that at least one instance is satisfiable as $d^{\text{sat}}_{\mathcal{A}} : (\mathcal{L}(X)^+ \to \mathcal{P}_+(X)) \to [0,1]$, given by

$$d^{\text{sat}}_{\mathcal{A}}(F) := \frac{|\{A' \lhd \mathcal{A} \mid F \in \mathbb{I}(A')\}|}{|\{A' \mid A' \lhd \mathcal{A}\}|}.$$

This function calculates the ratio between the number of instances that a given voting rule satisfies and the total number of instances. Notice that $d^{\text{sat}}_{\mathcal{A}}(F) = 1$ is the case if and only if F satisfies $\mathcal{A}$. Further, if we say that every axiom consists of exactly one instance (namely the axiom itself), then we obtain a metric stating how many of the axioms are completely satisfied. We can also let $\mathcal{A}$ be a singleton $\{A\}$ to find out to what extent the axiom A is satisfied by a rule.

This is a simple approach. However, satisfaction degrees for different sets of axioms are hard to compare due to different numbers of axiom instances. For an intraprofile axiom acting on a single profile, for instance, many voting rules achieve a satisfaction rate of 1 and the co-domain of the metric is just $\{0,1\}$. For a complex interprofile axiom, on the other hand, the number of axiom instances can be large, yielding a co-domain extensive in quantity, which makes it harder for a voting rule to reach a rate of 1. In short, there is a qualitative difference for a rule to reach a satisfaction score of $x \in [0,1]$ between different sets of axioms. It is, thus, important to aim for a similar level of granularity across all axioms in the instances division.

While we proposed a natural way of assessing how much a set of axioms is satisfied by a rule, is it also a good metric? Besides the described problems of comparability, the defined metric does not take into account the logical strength of an axiom instance. If an instance's extension consists of exactly one voting rule, then this is an axiom that is very difficult to satisfy. And, conversely, if most voting rules belong to the extension, this is a sign that the axiom is rather easy to satisfy. To reflect this in the metric, we might want to give more weight to the former instance than to the latter. Taking this into account, we could alternatively define the satisfaction metric as

$$d^{\text{sat}}_{\mathcal{A}}(F) := \frac{\sum_{A' \lhd \mathcal{A}, F \in \mathbb{I}(A')} |\mathbb{I}(A')^c|}{\sum_{A' \lhd \mathcal{A}} |\mathbb{I}(A')^c|},$$

where $\mathbb{I}(A')^c := (\mathcal{L}(X)^+ \to \mathcal{P}_+(X)) \setminus \mathbb{I}(A')$. In this case, again, the instances that F satisfies are being counted against all instances, but this time weighted by how difficult it is to satisfy the instance. Here, we presuppose that the fewer

[9] In mathematics, the term "metric" refers to a bivariate function that measures the distance between two objects and satisfies the identity of indiscernibles, symmetry and the triangle inequality. In this thesis, we use the term in an informal way to denote a function that quantifies how close an object is to achieving a maxim on a scale from 0 to 1.

rules adhere to an axiom, the more difficult it is to satisfy the axiom. So, the more rules do not satisfy the instance A', i.e., are contained in $\mathbb{I}(A')^c$, the harder it is to satisfy A'. We divide by the sum of all numbers of extension complements in order to normalize, i.e., so that the weights across all instances still sum up to 1.

Notice that this definition manages to level out the differences among the instances of one axiom. However, it still does not ensure comparability between different axioms. That is, it does not take into account the differences in terms of logical strength between the axioms. A further improvement could be achieved by

$$d^{\text{sat}}_{\mathcal{A}}(F) := \frac{1}{\sum_{A\in\mathcal{A}} |\mathbb{I}(A)^c|} \sum_{A\in\mathcal{A}} |\mathbb{I}(A)^c| \cdot \frac{\sum_{A'\lhd A, F\in\mathbb{I}(A')} |\mathbb{I}(A')^c|}{\sum_{A'\lhd A} |\mathbb{I}(A')^c|}$$

In this metric, we calculate for each axiom in the axiom set, how many of its instances are satisfied (where the instances are weighted according to their logical strength). These scores are then averaged and in this, each axiom, too, obtains a weight according to its logical strength. The metric still depends on the choice of how to divide one axiom up into multiple instances though. That is, although it might balance out differences in strength between different axioms and instances, how sensitive it is overall depends on how fine-grained the instance division is. Recall that an axiom with only one instance can reach a score of 1 much more easily than an axiom with many instances.

So while the precise numbers are not informative, notice that the metric is still monotonic, i.e., if we fix the instances for axioms, then if one rule satisfies more instances than another, its satisfaction score is higher. This means that, as long as we keep the division of axioms into instances constant, we can compare the performance of multiple rules relative to each other.

Example 8. *We want to compute the Voting by Axioms rule derived from the ranked collection* $\{\text{FAI}, \text{ANO}\} \succ \{\text{CON}, \text{NEU}\} \succ \{\text{CAN}\} \succ \{\mathbf{1}\}$ *in a setting with* 3 *alternatives and* 2 *voters. For all single-voter profiles, the maximal forcing set in the collection is* $\{\text{FAI}, \text{ANO}\}$*, which forces the highest-ranked alternative of the voter to be the single winner. For all other profiles, this axiom set does not force an outcome, so we need to consider the other axiom sets in the collection. The second highest-ranked set* $\{\text{CON}, \text{NEU}\}$ *forces the unique Condorcet winner to be the sole winner, whenever it exists. This leaves profiles where all pairwise majority contests are tied (e.g.,* $(123, 321)$*) and profiles in which two alternatives have the same winning scores under pairwise majority contests (e.g.,*$(123, 312)$*). The cancellation axiom takes care of the first kind and assigns* X *to all of them, and the axiom* $\mathbf{1}$*, stating that the outcome under all profiles is* $\{1\}$*, forces an outcome on all remaining profiles. Thus, the Voting by Axioms rule is well-defined (which it would not be if we excluded the axiom set* $\{\mathbf{1}\}$*!).*

Let us calculate the satisfaction scores of the individual axioms. First, notice that faithfulness, anonymity, neutrality, the Condorcet principle and cancellation are completely satisfied, i.e., yield a satisfaction score of 1*. For the*

intraprofile axiom **1**, *we can simply count how many of the* 48 *profiles return the outcome* $\{1\}$. *There are* 4 *profiles that faithfulness forces this outcome on,* 4 *profiles where* 1 *is the Condorcet winner and there are* 18 *profiles that the last axiom set determines the outcome* **1** *on for the Voting by Axioms rule. This yields a satisfaction score of* $\frac{26}{48} \approx 0.54$. *Note that, in this case, all three definitions of the satisfaction metric return the same outcome, if we define one instance per profile for the axiom* **1**.

Forcing metrics. Besides measuring, ex post, the performance of the Voting by Axioms rule, we might also be interested, ex ante, in how to choose good axiom sets for the collection. One criterion, of course, is that the sets of axioms should be satisfiable since, otherwise, they cannot force any outcome by definition. But now suppose that the collection, in its current state, does not force any outcome on profile R. How can we find an axiom set that does? Trivially, we can always add a set of axioms that uniquely characterizes some rule to the collection since it forces an outcome on every profile. But this might involve introducing new axioms to the collection that we are not really interested in. How can we alter one of the existing axiom sets to obtain a set that forces an outcome on R? The strategy is, if we could measure how close a set is to forcing an outcome, to start with the most promising axiom set and add or strengthen axioms until the set prescribes an outcome on R, using the metric to guide the process. This is the motivation behind introducing a metric capturing to what extent an axiom set restricts what outcome can be assigned to a given profile.

We want to suggest two different ways of gauging how close an axiom set is to forcing an outcome on a given profile. Both *forcing metrics* are functions assigning to each satisfiable axiom set a value in $[0, 1]$. The first one divides 1 by the number of different admissible outcomes with regards to the axioms on the given profile. Formally, define it via

$$d_R^{forc}(\mathcal{A}) := \frac{1}{|\{F(R) \mid F \in \mathbb{I}(\mathcal{A})\}|}.$$

Note that we have $d_R^{forc}(\mathcal{A}) = 1$ precisely if $\mathcal{A}$ forces an outcome given R.

For the second one, let

$$d'^{forc}_R(\mathcal{A}) := \frac{\max\{|S| : S \subseteq \mathbb{I}(\mathcal{A}), \text{ for all } F_1, F_2 \in S, \text{ we have } F_1(R) = F_2(R)\}}{|\mathbb{I}(\mathcal{A})|}.$$

If we partition $\mathbb{I}(\mathcal{A})$ into cells such that all rules inside a cell assign the same outcome to R, then the function d'^{forc}_R takes the ratio of the size of the largest cell versus the size of the whole extension of $\mathcal{A}$. This expresses, proportionally, how much support the outcome that the plurality of rules satisfying $\mathcal{A}$ return has. Again, we have $d'^{forc}_R(\mathcal{A}) = 1$ if and only if $\mathcal{A}$ forces an outcome given R.

As mentioned before, if we are looking for a set that forces an outcome on R, we would first calculate the scores of all axiom sets in the collection under a forcing metric. We can then pick the one that we need to alter the least in order to force an outcome, i.e., the one with the highest score, and try to add

axioms or strengthen present axioms that reduce the extension in size but that are still compatible with the original axiom set. During this process, we can always assess how close we are to forcing an outcome by calculating the forcing metric for the altered axiom set. We continue in this fashion until we reach a forcing score of 1.

These functions could also be used for other purposes. For instance, we could use the them to derive characterization results. For instance, we can start with a promising set of axioms that forces an outcome on many profiles. This set is close to uniquely characterizing a voting rule, but on a few profiles, it still allows for multiple outcomes. We can start with one of these profiles and measure how close the axioms are to forcing an outcome. Dependent on that, we can either add axioms or strengthen one of the given axioms to force an outcome on this profile. We can then check for the resulting set, whether it forces an outcome on all profiles. If not, we repeat the operation. We follow this procedure until we end up with an axiom set that forces an outcome on every profile. This is then a uniquely characterizing axiom set for the resulting rule.

3.4 Computational Complexity

In this thesis and in constructing our model, we aimed at keeping problems feasible to implement and to compute. Now we want see if we succeeded, that is, we want to shed light on how computationally hard it is to compute the Voting by Axioms rule. Before taking a closer look, our intuition should tell us already that this is a difficult undertaking. Let us start by quantifying how many objects our model contains.

- n voters in the universe, yielding $\sum_{k=1}^{n} \binom{n}{k} = 2^n - 1$ possible electorates
- m alternatives, yielding $2^m - 1$ possible outcomes
- $m!$ possible rankings of alternatives, yielding $\sum_{k=1}^{n} \binom{n}{k} m!^k = (m!+1)^n - 1$ many possible profiles (each voter in each electorate gets to pick between $m!$ rankings)
- $(2^m - 1)^{((m!+1)^n - 1)}$ many possible voting rules
- $2^{(2^m-1)^{((m!+1)^n-1)}}$ many possible axiom extensions

Since most of our procedures work profile by profile, the number of possible profiles is an important factor for their complexity. The number of profiles lies in $\mathcal{O}(m!^n)$, meaning that it grows exponentially in n and factorially in m. Notice that factorial growth is faster than exponential for large parameters, but in voting, we usually consider setups with only a few alternatives and a large electorate. Therefore, the parameter to primarily focus on should be the number of voters n. To illustrate how problematic exponential and superexponential complexity is, we calculate the number of profiles for different parameters n

and m. Since the model itself is complex, we expect long runtimes for our calculations.

m \ n	2	4	6
2	8	80	728
4	624	390624	$\approx 244 \cdot 10^6$
6	519840	$\approx 270 \cdot 10^9$	$\approx 140 \cdot 10^{15}$

Table 3.2: Number of profiles depending on n and m

Boixel and de Haan (2021) have analyzed the complexity of checking whether there exists a justification for a given profile. This problem is closely related to the question of whether there exists an axiom set that forces an outcome on a profile, since we search for the normative basis of a justification. However, they use the full notion of a justification in voting introduced by Boixel and Endriss (2020) which additionally imposes the condition that, to explain an outcome, a set of axiom instances must be minimal with the property of forcing the outcome. That is, Boixel and de Haan determine both the explanation (the minimal set of instances forcing an outcome) and the normative basis (the axioms that these instances stem from) of a justification. Another difference is that they use a many-sorted first-order language, allowing to quantify over alternatives, voters and profiles, to encode axioms. The authors come to the conclusion that determining whether some outcome can be justified on a given profile lies in $\mathrm{EXP}^{\mathrm{NP}}$ and is NEXP-hard and even verifying whether suggested instances yield a justification lies in NEXP $\wedge$ coNEXP and is also NEXP-hard. This sets the tone for this section. We will show that we can slightly improve on these results in our setting, when searching for forced outcomes instead of explanatory justifications. However, together with the insights from the previous paragraph, we expect exponential or superexponential complexity.

We want to break up the task of calculating the Voting by Axioms rule into multiple subproblems. First, define the problem of finding out, whether a certain outcome is forced by an axiom on a given profile, i.e., of verifying whether an axiom set indeed forces a given outcome on a profile.

Check if Outcome is Forced (Check-Forc)

Input : Set of voters N, set of alternatives X, profile $R \in \mathcal{L}(X)^+$, satisfiable axiom A as propositional formula in $\mathcal{L}$, outcome $O \in \mathcal{P}_+(X)$

Question : Does A force O given R?

Note that this also works for finite sets of axioms $\mathcal{A}$ by setting $A := \bigwedge_{A' \in \mathcal{A}} A'$.

Proposition 3.11. Check-Forc *is* coNP*-complete w.r.t. the number of propositional atoms occurring in the specification of* A*.*

Proof. To show that a problem is coNP-complete, it suffices to prove that its complement is NP-complete. Because this means that the dual problem lies in NP (so the problem is in coNP) and any input for the NP problem can be transformed in polynomial time into an input for the dual problem (so any coNP-problem can be solved by taking the inverse outcome of the dual problem with the transformed input). We denote the dual problem by Check-NotForc, which answers the question whether the given outcome is not forced by a given axiom A on a given profile R.

First, we need to show that Check-NotForc belongs to NP. It is enough to show that it is an instance of an NP-problem, since this means that it can be computed in polynomial time by a nondeterministic Turing machine. Recall from Section 3.1.1 that an axiom forces O on R iff $\mathit{atLeastOne} \wedge A \wedge \neg \mathit{outcome}(R, O)$ is unsatisfiable. So, to answer Check-NotForc, we need to solve the problem Sat for the formula $\mathit{atLeastOne} \wedge A \wedge \neg \mathit{outcome}(R, O)$ and Sat is an NP-problem with respect to the number of propositional letters occurring in the given formula. It remains to show that the input to Sat is polynomial w.r.t. the letters in A. If we show this, then Check-NotForc can be computed in polynomial time by a nondeterministic Turing machine.

Unfortunately, all profiles occur in the formula *atLeastOne*. Thus, there are possibly exponentially or superexponentially many atoms in this formula. To circumvent this and to obtain a formula polynomial in the letters occurring in A, we may replace the disjunction over all profiles in *atLeastOne* with a disjunction solely over the profiles that A speaks about and over profile R, call this formula $\mathit{atLeastOne}'$. Note that $\mathit{atLeastOne} \wedge A \wedge \neg \mathit{outcome}(R, O)$ is logically equivalent to $\mathit{atLeastOne}' \wedge A \wedge \neg \mathit{outcome}(R, O)$ since A and $\mathit{outcome}(R, O)$ only impose conditions on the profiles occurring in $\mathit{atLeastOne}'$, so the additional clauses in the original formula *atLeastOne* are always satisfiable since we are free to assign whatever outcome we please on these profiles. Therefore, solving Check-NotForc boils down to solving Sat for $\mathit{atLeastOne}' \wedge A \wedge \neg \mathit{outcome}(R, O)$. Note that there can be at most as many profiles as propositional letters occurring in A, so there can be at most $m \cdot (|\mathit{Prop}(A)| + 1)$ many letters in $\mathit{atLeastOne}'$ (this would be the case if all atoms in A correspond to different profiles distinct from the given profile R; note that there exist m atoms per profile $p_{R,x}$). Thus, the input to Sat is polynomial with regards to the number of letters in A, and so, Check-NotForc lies in NP.

For NP-hardness, it suffices to give a polynomial time reduction from an NP-complete problem to Check-NotForc. Again, consider Sat, which answers the question whether a propositional formula φ is satisfiable. We already described that Check-NotForc answers the question if the formula $\mathit{atLeastOne} \wedge A \wedge \neg \mathit{outcome}(R, O)$ is satisifable. So if we manage to express our formula φ in terms of an axiom and outcome, we are done. Our goal is to associate the formula φ with a satisfiable axiom A and to choose an outcome and profile in such a way that φ is satisfiable if and only if A does not force this outcome on the profile.

To this end, given the propositional letters $\mathit{Prop}(\varphi) = \{p_1, \ldots, p_k\}$ occurring in φ, let m be the smallest number such that $m! \geq k + 1$ and set $n = 1$.

Our voting model then contains as many profiles as distinct ballots, i.e., $m!$ many. Fix an enumeration of the profiles $\mathcal{L}(X)^+ = \{R_1, R_2, \ldots, R_{m!}\}$ and identify the propositional atoms p_i with $p_{R_i,1}$ for $i = 1, \ldots, k$. This means that we can now express φ in terms of the propositional letters $p_{R,x}$, denote this formula by $A_\varphi := \varphi[p_i/p_{R_i,1}]$. In this formula, we replace every occurrence of a propositional atom p_i with $p_{R_i,1}$ for all $i = 1, \ldots k$. So A_φ expresses whatever A said, but now speaks about whether alternative 1 wins or looses in profiles $R_1, \ldots, R_k$. Consider the formula $A := A_\varphi \vee \mathit{outcome}(R_{k+1}, X)$. Notice that this formula is always satisfiable because X is an acceptable outcome, so $\mathit{outcome}(R_{k+1}, X)$ is satisfiable. We will treat A as our axiom and consider Check-NotForc for the profile R_{k+1} and outcome X (the choice of outcome is arbitrary).

We claim that Check-NotForc comes out true if and only if Sat comes out true. Recall that Check-NotForc returns "yes" if A does not force X on R_{k+1}, i.e., if $A \not\models \mathit{outcome}(R_{k+1}, X)$. This, in turn, is the case iff $\mathit{atLeastOne} \wedge A \wedge \neg\mathit{outcome}(R_{k+1}, X)$ is satisfiable, that is, there exists a voting rule that satisfies the axiom A but that does not return X given R_{k+1}. It remains to check that the aforementioned conjunction, used to solve Check-NotForc, is logically equivalent to A_φ. We write $\approx$ for logical equivalence and obtain

$$
\begin{aligned}
&\mathit{atLeastOne} \wedge A \wedge \neg\mathit{outcome}(R_{k+1}, X) \\
&\approx (A_\varphi \vee \mathit{outcome}(R_{k+1}, X)) \wedge (\neg\mathit{outcome}(R_{k+1}, X) \wedge \mathit{atLeastOne}) \\
&\approx (A_\varphi \wedge \neg\mathit{outcome}(R_{k+1}, X) \wedge \mathit{atLeastOne}) \\
&\qquad \vee (\mathit{outcome}(R_{k+1}, X) \wedge \neg\mathit{outcome}(R_{k+1}, X) \wedge \mathit{atLeastOne}) \\
&\approx A_\varphi \wedge \neg\mathit{outcome}(R_{k+1}, X) \wedge \mathit{atLeastOne}
\end{aligned}
$$

Notice that since the propositional letters occurring in A_φ correspond to different profiles, they are independent from one another and, further, the formula is consistent with *atLeastOne* because A_φ only restricts for alternative 1 whether it should or should not be among the winners. Similarly, the formula $\mathit{outcome}(R_{k+1}, X)$ refers only to the profile R_{k+1}, which renders it independent from A_φ, which speaks about profiles $R_1, \ldots, R_k$. It remains to realize that $\mathit{atLeastOne} \wedge \neg\mathit{outcome}(R_{k+1}, X)$ is always satisfiable, e.g. by assigning the outcome $\{1\}$ to every profile. Thus, the conjunction $(\mathit{atLeastOne} \wedge \neg\mathit{outcome}(R_{k+1}, X)) \wedge A_\varphi$ is satisfiable iff A_φ is satisfiable. Since we merely renamed the propositional letters, this is the case exactly if φ is satisfiable. Together, Check-NotForc for parameters A, R_{k+1} and $O = X$ returns "yes" whenever Sat returns "yes" for φ. This transformation is polynomial since we only rename the atoms in φ.

We showed that Check-NotForc has NP-membership and is NP-hard, which together means that it is an NP-complete problem. Therefore, its dual problem Check-Forc is coNP-complete. □

This problem loosely corresponds to the Check-Just problem in Boixel and de Haan (2021) for the case of quantifier-free formulas which is shown to be DP-complete, where DP := NP∧coNP. Since we do not need to verify minimality of

the axiom set with regards to forcing an outcome (which is the NP-component of the problem), this is in line with our result about coNP-completeness.

We can use this basic subproblem to find out whether an axiom forces any outcome on a given profile.

Exists Forced Outcome (Exists-Forc)

Input : Set of voters N, set of alternatives X, profile $R \in \mathcal{L}(X)^+$, satisfiable axiom A as propositional formula in $\mathcal{L}$

Question : Is there an outcome $O \in \mathcal{P}_+(X)$ such that A forces O given R?

First, notice that this problem does not correspond to what is called Exists-Just by Boixel and de Haan (2021). Whereas their problem ranges over all instance sets for a fixed outcome and determines whether there exists some justification for the outcome on the profile, our problem loops over outcomes to determine for a given axiom, whether it forces one of them on the profile. The brute-force algorithm to answer this question is to solve Check-Forc for A and R and all possible outcomes O until we find a forced outcome or until we exhausted all outcomes. In the worst case, however, this means solving Check-Forc $2^m - 1$ many times. This is exponential in the number of alternatives, so in general hard to compute. Since it is solvable in exponential time by using a coNP-oracle, Exists-Forc lies in EXP^{coNP}. Recall, though, that in voting, we usually work with a small, fixed number of alternatives, rendering this algorithm feasible in most cases. We might be able to accelerate this by extracting from the axiom the outcomes that it speaks about, or better, the outcomes that are allowed by the axiom. We know that the forced outcomes across the domain must be a subset of these, so this restricts the search space of outcomes. It might be difficult to obtain this information though. Besides this, we do not expect there to exist other heuristics that help speed up the process since, if a specific outcome is not forced, this generally does not tell us anything about whether another outcome is forced.

The next more complex problem is checking whether Voting by Axioms is well-defined for a given collection of sets of axioms. This means solving Exists-Forc for every possible profile and a given collection $\mathbb{A}$.

Check Well-Definedness of Voting by Axioms (Check-WDef)

Input : Set of voters N, set of alternatives X, collection of axiom sets $\mathbb{A}$ each given as one propositional formula in $\mathcal{L}$

Question : Is the rule $F_{(\mathbb{A},\succ)}$ well-defined for any ranking $\succ \in \mathcal{L}(\mathbb{A})$?

This problem first requires us to identify the satisfiable sets within the collection. Then, in the worst case, we need to check for every profile and every satisfiable axiom set whether the set forces an outcome on the profile. This could be the case if, for all profiles, only one axiom set forces an outcome. If all sets are satisfiable, this means that Exists-Forc is called $|\mathcal{L}(X)^+| \cdot |\mathbb{A}|$ many times. For this problem, we have to check for each profile independently whether an

outcome is forced, so we cannot change anything about the factor $|\mathcal{L}(X)^+|$. However, recall from Section 3.1 that it suffices to restrict attention to set-inclusion maximal sets in the collection. Further, as mentioned in Section 3.1, as a preliminary test, we could check whether every profile is spoken about by some profile. We might also be able to find heuristics helping to identify the most promising order of axiom sets for each profile to check Exists-Forc for. This could be done by checking which axiom sets speak about the profile and how often. Or we could optimize the order, in which we check the profiles, ranking them by the number of axioms or instances that speak about them and going from the least coverage to highest profile coverage, to increase the chances of finding profiles that no outcome is forced on quickly. Nonetheless, this means that, in the worst case, if for every axiom set only the last axiom set that we try out forces an outcome, and only the last outcome that we test is the forced outcome, we need to call Check-Forc $|\mathbb{A}| \cdot ((m! + 1)^n - 1) \cdot (2^m - 1)$ many times. This is exponential both in n and m. So even when working with a small number of alternatives m, the computation of this is hard, because we need to call a coNP-oracle exponentially many times.

While there might exist quicker algorithms than the brute-force methods that we presented, due to the high complexity of the framework itself, there do not exist subexponential algorithms with respect to n for computing the Voting by Axioms rule. The benefit of looking at each profile independently, however, is that we do not actually have to calculate the whole rule in advance, but we could first obtain the ballots and then calculate the outcome for this very profile only.

We have seen in this chapter that Voting by Axioms does not work for every possible collection of axioms. We introduced two frameworks that help determine whether the voting rule is well-defined for a given collection and, further, stated conditions for when the Voting by Axioms rule satisfies the axioms that it is based on. We saw that this is rarely the case but for all other cases, we suggested a metric that helps quantifying, to what extent axioms are satisfied by the voting rule. Lastly, we analyzed the computational complexity and found that computing the defined rule is computationally hard.

Chapter 4

Extensions of the Framework

In Section 2.3, we presented a simple version of Voting by Axioms based on a collection of axiom sets $\mathbb{A}$ with a strict ranking over it. We chose to require a complete, strict order because this is easiest to work with since the order yields a unique maximal axiom set. Supplying such a ranking might not be feasible in practice though due to bounded rationality and limited cognitive capacity of the system's user. First, notice that it is easier for humans to compare axioms rather than axiom sets. So we want to discuss how we can lift preferences over axioms to preferences over sets of axioms. Further, if $\mathbb{A}$ contains many sets of axioms, it might be difficult to arrange all of them in a coherent, strict ranking, either because one finds two sets incomparable or because one is indifferent between two sets. This is the motivation behind examining what happens in case we are given a weak or partial order over $\mathbb{A}$. Last, we present an alternative voting rule that is not based on axioms in the sense that it justifies each profile's outcome with the them, but that aims at satisfying the axioms as much as possible.

4.1 Lifting Orders

As motivated before, one reason for considering preferences over axioms rather than preferences over axiom sets is that comparing sets of objects is a much more cognitively difficult task for humans. Scontras, Graff, and Goodman (2012) suggest two models of how group comparison might work in humans: either one compares the objects contained in the groups one by one (*point-wise comparison*), or one uses some aggregation function, e.g., the mean, to assign one overall value per group that is then compared (*collective comparison*). In both cases, it is clear that comparing sets of objects is cognitively more demanding than just comparing objects themselves. Another motivation arises from viewing Voting by Axioms as a method to find the best (w.r.t. some criterion) justification among many for every profile. Instead of just ranking the axioms themselves,

we can assign a numerical value to each of them, corresponding to its utility, cost or complexity, and calculate a score for every axiom set. By minimizing this score, we can then determine the best justification. We will present two types of liftings in this section, one corresponding loosely to point-wise comparison, and the other one following the idea of collective comparison.

We have already seen one way of lifting an order over axioms to an order over sets of axioms in Definition 2.7, the lexicographic maximax ranking, which aims at optimizing the most preferred axioms in the set. Notice that in this method, at every stage, we compared one object to another one (namely, the best axioms of both sets). Therefore, we take this to be a point-wise comparison. This lifting was proposed by Pattanaik and Peleg (1984) as an extension of preference orders, that is, a lifting satisfying $\{A\} \succ \{A'\}$ iff $A > A'$. In their paper, they characterize this lifting as the unique lifting satisfying *neutrality*, *dominance*, *top independence* and *disjoint independence*.[10] This is part of a whole branch of research that defines and axiomatically analyzes extensions of preference orders, i.e., liftings from rankings over objects to rankings over sets of objects that respect the object ranking, see the discussion paper by Barberà et al. (2004). However, when using the lexicographic maximax lifting, some larger axiom sets are preferred to smaller ones simply because their best axiom is preferred to the smaller set's best axiom. In particular, we know that dominance is satisfied, which entails that if A is strictly preferred to all alternatives in $\mathcal{A}'$, then $\mathcal{A}' \cup \{A\} \succ \mathcal{A}'$. This also uncovers the downside of this lifting to rank some larger sets higher. Notice that, in our case, neutrality (i.e., the lifting treats all alternatives equally), dominance (i.e., adding a very good alternative is an improvement, adding a very bad one is a worsening) and independence (i.e., there are no interdependencies between the axioms influencing the ranking) are not suitable properties that we want our lifting to satisfy. This is because axioms are objects that can be logically related to each other and that have different logical strength, normative appeal as well as different motivations.

Instead, notice that, generally, smaller sets of axioms are preferable if we view them as explanations of an outcome since larger sets lead to more complex explanations. In general, we want to use only as many axioms for a justification as are absolutely needed. We could phrase this as an axiom for liftings called *antimonotonicity w.r.t. set inclusion*, requiring that $\mathcal{A}' \subsetneq \mathcal{A}''$ implies $\mathcal{A}' \succ \mathcal{A}''$. This expresses that if we add more axioms to a given justification, the longer explanation is less desirable. We might also be interested in having a lifting that satisfies *simple top- and bottom monotonicity*, i.e., if $A > A' > A''$ holds, then we have $\{A, A''\} \succ \{A', A''\}$ and $\{A, A'\} \succ \{A, A''\}$.[11] This says that, given one axiom, if we have the choice to add one of two other axioms that are either both more or less preferred than the original axiom, then adding the better one out of the two is preferable to adding the worse one. In a way, this captures

[10] We follow the version of the theorem cited in the paper by Barberà, Bossert, and Pattanaik (2004, Theorem 11).

[11] These axioms are called type 1 and type 2 simple dominance by Bossert, Xu, and Pattanaik (2000) originally, but we go with the names given by Barberà et al. (2004).

that for sets of the same size, we aim at including the best possible axioms in the set.

We suggest a different lifting that still aims at accommodating as many good axioms as possible but which gives preference to smaller sets. Therefore, it will satisfy all three suggested axioms. More precisely, it first groups the axiom sets by cardinality, ranking smaller sets higher, and within the groups, applies maximax to determine the order. Denote the axiom ranked on spot i within a ranking $>$ over an axiom set $\mathcal{A}'$ by $\mathcal{A}'(i)$, i.e., there exist $i-1$ many axioms $A \in \mathcal{A}'$ such that $A > \mathcal{A}'(i)$ and $|\mathcal{A}'| - i$ many axioms $A \in \mathcal{A}'$ with $\mathcal{A}'(i) > A$.

Definition 4.1. The *shortlex maximax ranking* given $\mathcal{A}$ and $>$ is constructed as follows. Set $\{A\} \succ \{A'\}$ for all axioms with $A > A'$ and recursively define

$$\begin{aligned}\mathcal{A}' \succ \mathcal{A}'' \text{ iff either } &|\mathcal{A}'| < |\mathcal{A}''| \\ &\text{or } \Big(|\mathcal{A}'| = |\mathcal{A}''| \text{ and} \\ &\quad \exists i \leq |\mathcal{A}'| \text{ s.t. } (\forall j < i\,(\mathcal{A}'(j) = \mathcal{A}''(j)) \text{ and } \mathcal{A}'(i) > \mathcal{A}''(i))\Big).\end{aligned}$$

It is easy to see that the shortlex maximax lifting satisfies both simple top- and bottom monotonicity since for sets of the same size, it works by comparing the best axioms in the sets. Further, it satisfies antimonotonicity w.r.t. set inclusion since it generally prefers smaller sets to larger ones.

Is this the suitable lifting to consider for sets of axioms? This is difficult to answer in general. One could argue, for instance, that although two axioms are individually not very desirable, paired together they become acceptable, and this could violate simple top- or bottom monotonicity. Think, for instance of anonymity and neutrality that have a similar governing principle, meaning that if we understood one of them, we easily can make sense of the second one. Moreover, it could be the case that an explanation using three very simple axioms is preferable to an explanation using two axioms, one simple and one very difficult one that most people cannot make sense of. This means, it might be desirable for the lifting to also satisfy *additive representability*, i.e., there exist utility functions u assigning to each axiom a real number such that $\mathcal{A}' \succ \mathcal{A}''$ iff $\sum_{A \in \mathcal{A}'} u(A) \geq \sum_{A \in \mathcal{A}''} u(A)$ (Barberà et al., 2004). This allows us to attribute a weight or utility with every axiom, which quantifies how much one axiom is preferred to another one. This inspires the following kind of lifting based on cost, the dual concept to utility.

Recall that we motivated Voting by Axioms by claiming that we can use it to find the best possible justification for each profile. One way of making this idea precise is by assuming that every axiom comes with some fixed cost, e.g., based on its logical strength, its complexity or its cognitive charges. We then want to aggregate the cost to derive an order on the sets of these axioms. This time, the goal is to minimize the cost, i.e., only using as many axioms as are strictly necessary for forcing an outcome.

Formally, suppose $\mathcal{A}$ is a set of axioms together with a cost function $c : \mathcal{A} \to \mathbb{R}_{\geq 0}$. For a subset of axioms $\mathcal{A}'$, calculate its cost via $c(\mathcal{A}') = \sum_{A \in \mathcal{A}'} c(A)$.

This allows us to derive a non-strict order $\succeq$ over $\mathcal{P}_+(\mathcal{A})$, the subsets of $\mathcal{A}$. This comparison strategy is a collective one since the cost of an axiom set is not a property of any of the individual axioms, but a compound property.

Definition 4.2. The *cost-based ranking* for $\mathcal{A}$ and $c : \mathcal{A} \to \mathbb{R}_{\geq 0}$ is defined for all non-empty $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{A}$ via $\mathcal{A}_1 \succeq \mathcal{A}_2$ iff $c(\mathcal{A}_1) \leq c(\mathcal{A}_2)$.

Notice that we obtain a weak order from this definition, meaning that there might be multiple equally preferred sets of axioms. We can add any method of tie-breaking, e.g., lexicographic tie-breaking, to turn the obtained order into a strict linear order. Also notice that, instead of summing up the cost of the individual axioms, one could use other aggregation functions to attribute a value for comparison to axiom sets. For instance, one could calculate the product of individual costs for each axiom set.

Example 9. *Consider the axioms introduced in Table 2.1. Additionally, let* SUR *be the surjectivity axiom, stating that for each outcome in $\mathcal{P}_+(X)$, we can find a profile that gets assigned to the outcome. Further, the axiom* CON^+ *is defined as "If there is no Condorcet winner, then everyone wins", where a Condorcet winner is an alternative x that wins every pairwise majority contest. Let $\mathcal{A} := \{\mathrm{ANO}, \mathrm{NEU}, \mathrm{PAR}, \mathrm{CON}, \mathrm{CON}^+\mathrm{REI}, \mathrm{FAI}, \mathrm{SUR}\}$ and define a cost function*

$$\begin{aligned} c(\mathrm{ANO}) = c(\mathrm{NEU}) = c(\mathrm{FAI}) &= 1, \\ c(\mathrm{PAR}) = c(\mathrm{CON}) = c(\mathrm{CON}^+) &= 2, \\ c(\mathrm{REI}) = c(\mathrm{SUR}) &= 3. \end{aligned}$$

We obtain $\{\mathrm{CON}\} \succ \{\mathrm{FAI}, \mathrm{REI}\}$. Thus, if we search for a justification for the profile $(123, 132)$, we find that the Condorcet principle by itself forces the outcome $\{1\}$. This is easier to explain than saying that, by faithfulness, if both voters were to vote in a separate election, the outcome would be $\{1\}$ in both cases, so by reinforcement, the outcome in the joint election must be $\{1\}$ as well. This is an example for when Voting by Axioms can help choose the most economical justification for one specific *outcome among many justifications.*

This procedure can also help find the justification with lowest cost when justifications for different *outcomes are available. Again, consider a setup with* 3 *alternatives and* 2 *voters. Note that if we assume the Condorcet principle, then surjectivity, together with the Pareto principle, anonymity and neutrality, requires that on the profiles without a Condorcet winner, both non-Pareto-dominated alternatives win, e.g., $F(123, 231) = \{1, 2\}$. But* CON^+ *forces $F(123, 231) = X$. The Voting by Axioms rule would return X for this profile since the cost-based ranking yields $\{\mathrm{CON}^+\} \succ \{\mathrm{CON}, \mathrm{ANO}, \mathrm{NEU}, \mathrm{PAR}, \mathrm{SUR}\}$.*

In this example, we saw that using a cost-based ranking to find the cheapest justification sometimes helps identifying the best justification for a given outcome among multiple available justifications but, in general, it finds the justification with the lowest cost out of all justifications of some outcome. This is a useful mechanism if we truly are only concerned about minimizing the justification's cost. If, however, it is the case that, among all available justifications

for a profile, a great share plead for the same outcome and there is one justification for another outcome that has lowest cost, the Voting by Axioms rule would still select the outcome with the cheapest justification. One could argue, in this case, that since the other outcome is supported by multiple distinct justifications, there exist more grounds to choosing this outcome.

We saw that, when using the cost-based approach, we generally obtain a weak preference order, meaning that multiple sets can be equally preferred. Our suggestion was to simply settle on a tie-breaking algorithm and turn the order into a strict ranking. In the next section, we want to examine more thoroughly how to deal with non-strict and partial preference rankings.

4.2 Weak and Partial Orders of Axiom Sets

Recall that in the standard setup for Voting by Axioms, we need to provide a collection of axiom sets $\mathbb{A}$ and a strict ranking over it. This allows us to assign to a profile that outcome imposed by the single maximal forcing set of axioms in the collection. While this makes it simple to define a voting rule based on the collection, notice that, in practice, it can be infeasible to submit a complete, strict order over all axiom sets. Many axioms have been defined in social choice theory (e.g., see Plott, 1976) and there are exponentially many possible sets formed with these axioms. So the sheer quantity can overwhelm the user of our model. Further, as previously mentioned, it is a difficult task for humans to compare two sets of objects. So the user might also fail at some instances and adjudge two sets incomparable. This leads us to loosening the requirements imposed on the ranking over the collection $\mathbb{A}$.

A strict ranking was defined to be an irreflexive, transitive, connected binary relation. We can generalize this by also allowing transitive and strongly connected relations $\succeq$, that is, total preorders. We call this a *weak* ranking. This means that we can compare any two sets of axioms, but we allow for multiple sets to be equally preferable. The consequence of this is that ties between multiple maximal forcing axiom sets can occur. These are sets of axioms that are ranked at least as high as all other sets. In this case, tie-breaking would solve the problem and the solution would be acceptable in most cases since all maximal sets are equally preferred and we would be content with any of these.

Qualitatively different from this is the problem of incomplete preferences. That is, we can allow for the order $\succ$ on the collection of axiom sets $\mathbb{A}$ to be *partial*. This means that we drop the connectedness requirement, assuming only irreflexivity and transitivity. The impact of this is that two sets of axioms $\mathcal{A}$ and $\mathcal{A}'$ in $\mathbb{A}$ no longer have to be comparable, so it might be the case that neither $\mathcal{A} \succ \mathcal{A}'$ nor $\mathcal{A}' \succ \mathcal{A}$ holds. This is expected to happen in practice since it might be difficult for two disjoint and (seemingly) unrelated sets of axioms to determine which one is more preferable. A natural measure to compare two sets of axioms is logical consequence, i.e., testing whether one set's extension is contained in the other one's. If the axiom sets contain independent axioms of varying logical strength and that are motivated differently (e.g., philosophically

vs. technically), the user might have difficulties finding grounds to stipulate that one set is preferred over the other.

The problem that arises from using partial orders is the same as for weak orders, namely that multiple maximal forcing sets of axioms can exist for one profile. The key difference, however, is that we can no longer state that they are equally preferred and that it, therefore, does not matter which one we choose. The sets that the tie occurs between are the sets that are ranked highest within one of the connected components of the ordering. The point is that we cannot prioritize between the different connected components, and thus, cannot easily arrange the components such that we obtain a strict ranking. In this case, tie-breaking is not a suitable solution.

Further, notice that we can also combine both of these generalizations and allow for $\succeq$ to be any preorder. This means that the order could both be weak and incomplete.

Let us explore two other ways of handling the tie between multiple maximal forcing outcomes. One option is to let the derived voting rule $F_{(\mathbb{A},\succeq)}$ be *irresolute* in the sense that it is a function $\mathcal{L}(X)^+ \to \mathcal{P}_+(\mathcal{P}_+(X))$, possibly returning multiple sets of outcomes in $\mathcal{P}_+(X)$. In this case, we define the Voting by Axioms rule as $F_{(\mathbb{A},\succeq)}(R) = \mathcal{O}^R_{\succeq}$, where

$$\mathcal{O}^R_{\succeq} := \{O \mid \text{there is } \mathcal{A} \in \max_{\succeq}\{\mathcal{A} \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A})\}) \text{ that forces } O\}$$

Since we already defined voting rules to be irresolute, i.e., they need not always return exactly one winner, adding another layer of irresoluteness is undesirable.

So, alternatively, we can opt for giving back the set of all *possible winners*, that is, all alternatives that occur in an outcome forced by at least one of the maximal forcing sets of axioms. This means, the derived voting rule is still a function in $\mathcal{L}(X)^+ \to \mathcal{P}_+(X)$ and defined via $F_{(\mathbb{A},\succeq)}(R) = \bigcup \mathcal{O}^R_{\succeq}$, where $\mathcal{O}^R_{\succeq}$ is the set of outcomes forced by $\succeq$-maximal forcing sets of axioms, as previously defined. In this context, one can also introduce the notion of a *necessary winner*, which is any alternative in X that wins in all outcomes forced by maximal forcing sets, i.e, which lies in the intersection of $\mathcal{O}^R_{\succeq}$. Konczak and Lang (2005) introduced these notions as solution concepts for incomplete preferences, the idea being that if we consider extensions of the given ordering, which could be obtained by further incoming information, a winner is possible if it wins in some refinement (i.e., at this point it is still possible for the alternative to win), and it is necessary if it wins in all refinements (i.e., we need not know the complete preference order to decide this).

Notice that it is even more difficult in this case, to predict the behavior of the Voting by Axioms rule, than in the standard setting. We can adapt the results from Section 3.2, e.g., by assuming that the characterizing axiom set is contained in all maximal forcing sets for Theorem 3.6. However, these theorems will only rarely be applicable. Notice that all axioms that prescribe what set should be the outcome under some profile (e.g., all algebraic axioms, see Section 3.2.2) do not pair well with the idea of possible winners, as this notion merges multiple outcomes. Thus, such an axiom can only be satisfied if possible and necessary

winners coincide. On the other hand, axioms that specify that some alternative should or should not be among the winners of some profile might be compatible with these notions, e.g., the Pareto principle. This is because if an axiom says that an alternative should win, it suffices to inspect the necessary winners, and conversely, for an axiom that says than an alternative should loose, we can check whether it occurs among the possible winners to check whether the axiom is satisfied. A simple observation is that if a satisfiable axiom only consists of instances that exclude an alternative from the outcome under a profile, then if it is contained in one of the maximal forcing sets for each profile that it speaks about, the Voting by Axioms rule assigning possible winners satisfies the axiom. Yet still, for all other axioms it is difficult to conclude, just from the information that the axiom is contained in some highly-ranked set in the collection, that it is satisfied by the Voting by Axioms rule.

We can also make use of the concepts of possible and necessary winners for the purpose of *preference elicitation*. This describes procedures that incrementally gather information about the user's or voters' preferences until there is enough information to take a clear decision. Boutilier and Rosenschein (2016, Section 5) present an overview of preference elicitation algorithms and their complexity. In our case, we explained that it can be a difficult or even unfeasible task to supply a strict linear order over the collection of axiom sets $\mathbb{A}$. Instead of requiring the user to supply a complete order over all sets, the system could repeatedly prompt the user to decide for a pair or selection of axiom sets, which one should be ranked highest, until the constructed (partial) order is informative enough to determine the derived Voting by Axioms rule.

For this, we will assume that Voting by Axioms is well-defined on the collection $\mathbb{A}$. Remember that this means that on each profile, at least one set in the collection forces an outcome. The preference elicitation procedure starts by retrieving a partial and possibly weak order $\succeq_0$ over $\mathbb{A}$ from the user (this could be skipped by setting $\succeq_0 = \emptyset$). We then compute possible and necessary winners for each profile. Recall that for a preorder $\succeq_0$ over $\mathbb{A}$ and for some profile R, we defined the set of all $\succeq_0$-maximal forced outcomes, given R, as $\mathcal{O}^R_{\succeq_0}$. The possible winners are then given by $\bigcup \mathcal{O}^R_{\succeq_0}$, whereas the (possibly empty) set of necessary winners is defined as $\bigcap \mathcal{O}^R_{\succeq_0}$. If the two sets coincide, i.e., if all $\succeq_0$-maximal forcing axiom sets force the same outcome on R, then we resolved the tie, we can assign this outcome to the profile and we do not need any additional information regarding this profile. If there is a profile R on which possible and necessary winners do not yet coincide, then we ask the user for more information regarding the ordering of the subcollection $\max_{\succeq_0}\{\mathcal{A} \in \mathbb{A} \mid R \in \mathit{Forc}(\mathcal{A})\}$. Ideally, the user would provide a complete, strict ranking over this subcollection. Alternatively, ask the user to specify their most preferred set in this collection. The extended order that the user's input yields is denoted by $\succeq_1$. We proceed with this method until possible and necessary winners coincide on all profiles. This will happen eventually since, in the worst case, we continue with this procedure until we obtain a strict linear order. Then there is exactly one maximal forcing axiom set for each profile.

While this is a sensible procedure for preference elicitation, notice that for each $\succeq_i$, we basically check whether Voting by Axioms is well-defined with this choice of ranking. With regards to the complexity results from Section 3.4, this is a computationally hard procedure.

4.3 Towards a Satisfaction-Maximizing Voting Rule

When hearing the term "Voting by Axioms", two possible interpretations come to mind. The first one is what the rule defined in Section 2.3 is based on, i.e., a voting rule that justifies its outcomes with given axioms. The second one describes a voting rule that aims at satisfying axioms as much as possible. Notice that these are two different objectives. Whereas in the first case, we extract specific intraprofile conditions imposed by the axiom and make sure to satisfy these, in the second case, we aim to maximize the number of axioms' conditions that are satisfied. Recall that for the derived rule $F_{(\mathbb{A},\succ)}$, we proceeded profile by profile. We now want to suggest a procedure that focuses on axiom instances instead.

On the one hand, this procedure can help us, if we have a set of axioms that is inconsistent, to find a voting rule that at least satisfies as many of the imposed conditions as possible. On the other hand, if we start with a satisfiable axiom set, the method can help us to choose one voting rule from the axioms' extension that additionally possesses as many other desirable properties as possible. Again, we start off with a collection of axiom sets $\mathbb{A}$ and a strict ranking $\succ$ over it. This time, we proceed set by set, maximizing the number of instances that the rule satisfies. This means, we start with the highest-ranked axiom set in $\mathbb{A}$ and determine the largest satisfiable subsets of its instances. In the next step, we refine this by determining the largest satisfiable sets of instances that contain one of the previously obtained sets of instances and, additionally, maximize the number of instances from the second-highest axiom set in $\mathbb{A}$ that are satisfied. We continue this procedure for all sets in the collection.

For a collection $\mathbb{A}$ with a strict linear order $\succ$ over it, label the i-th set in the ranking as $\mathbb{A}(i)$, e.g., $\mathbb{A}(1) = \max_{\succ}(\mathbb{A})$. For each set of axioms $\mathcal{A}$, fix a set of instances $\mathit{Inst}(\mathcal{A}) := \{A' \mid A' \triangleleft \mathcal{A}\}$. We define a partial order over sets based on their cardinality, i.e., $S_1 \geq S_2$ iff $|S_1| \geq |S_2|$. Formally, the algorithm works as follows:

- Determine the largest satisfiable subsets of $\mathbb{A}(1)$'s instances, i.e.,

$$\mathcal{I}_1 := \max_{\geq}\{\mathcal{A}' \in \mathcal{P}_+(\mathit{Inst}(\mathbb{A}(1))) \mid \mathbb{I}(\mathcal{A}') \neq \emptyset\}.$$

  Note that in general there are multiple maximal sets of instances.

- Next, we determine the biggest satisfiable sets that additionally satisfy instances from the second-highest-ranked set in the collection. That is,

$$\mathcal{I}_2 := \max_{\geq}\{\mathcal{A}' = \mathcal{A}_1 \cup \mathcal{A}_2 \mid \mathcal{A}_1 \in \mathcal{I}_1, \mathcal{A}_2 \in \mathcal{P}_+(\mathit{Inst}(\mathbb{A}(2))), \mathbb{I}(\mathcal{A}') \neq \emptyset\}.$$

- In general, for $k < |\mathbb{A}|$, we define the next set of instances given $\mathcal{I}_k$ as

$$\mathcal{I}_{k+1} := \max_{\geq}\{\mathcal{A}' = \mathcal{A}_1 \cup \mathcal{A}_2 \mid \mathcal{A}_1 \in \mathcal{I}_k, \mathcal{A}_2 \in \mathcal{P}_+(Inst(\mathbb{A}(k+1))), \mathbb{I}(\mathcal{A}') \neq \emptyset\}.$$

  We continue this method either until $|\mathbb{I}(\mathcal{A}')| = 1$ for all $\mathcal{A}' \in \mathcal{I}_k$ for some k (which means that we singled out rules) or until we have iterated through all sets in the collection.

Note that if the highest-ranked set of axioms is satisfiable, then the derived rule satisfies all these axioms. In each step, considering the next highest-ranked axiom set in the collection, we try to additionally impose as many of the set's instances as possible. In general, we do not take into account which axiom the satisfied instances belong to. If we want to try to satisfy the individual axioms as much as possible, we should include these as singletons in the collection.

With the presented procedure, we might still end up with multiple maximal instance sets and multiple voting rules contained in their extensions. This is similar to the situation when our standard Voting by Axioms rule is not well-defined since on some profile no outcome is forced by any set in the collection. In this case, there remains a choice to be made. Adding more axiom sets might help single out one voting rule.

Further, it is important to note that this approach very much depends on the definition of the axiom instances. The more fine-grained the division of axioms into instances is, the higher is the degree of satisfaction achievable with the defined method. This is very similar to what we said about satisfaction metrics in Section 3.3. Note that the described procedure identifies the rules with the highest score under the first definition of satisfaction metric. But notice that if one axiom is split into dozens of instances and another one is divided merely into two, this could mean that the instances of the second axiom have smaller extensions and are, thus, harder to satisfy. This will lead to the defined rule maximizing the number of satisfied instances of the first axiom rather than of the second one. This is to say that this method may not lead to an optimal result that maximizes satisfiability and guarantees fairness between all involved axioms. Nonetheless, it can give some indication to which are the more desirable rules from within a set of voting rules.

Chapter 5

Conclusion

Based on the idea that one can take axioms into account for deciding what the outcome in a given voting scenario should be, we developed a voting rule that justifies all its outcomes in this manner with preselected axioms (Voting by Axioms). We started with a standard model from voting theory with ordinal preferences and a finite number of voters and alternatives. Since we wanted axioms not only to be objects on a metalevel *speaking about* the rules in the model, but rather to be formal objects *included in* the model, we contemplated about their nature and possible ways to formalize them. We found that axioms, in the field of social choice theory, are usually taken to be normative principles or desirable properties of a decision procedure, which stem from the function of such a procedure to uncover the social preference of the voters. On the formal side, we had the choice between an intensional definition, taking axioms to be formulas in a formal language, or a purely extensional definition, defining axioms as the set of voting rules that satisfy them. While for constructing our Voting by Axioms rule we only needed the extension of axioms, we saw that to split an axiom into instances and for computational feasibility, formal descriptions of axioms, e.g., in a propositional language, are desirable. However, we developed methods of making sense of an extensionally given axiom, e.g., by defining the set of profiles that an axiom speaks about or by stepwise extracting the conditions that an axiom imposes. Based on the latter, we defined a new, language-independent hierarchy of axioms.

An important shift of perspective was to not view axioms as principles that are either completely satisfied or not, but rather to see them as a multitude of conditions that we try to satisfy as many as possible of. This motivated the notion of an axiom instance as one part or subcondition of an axiom. Another perspective on instances was to view them as atomic axioms that we use as building blocks to construct more complex axioms. Otherwise, we showed that it is difficult, in general, to define a procedure to recover axiom instances from a given axiom. The main issue was to find the right degree of granularity, i.e., an even division into conditions of similar strength that represent realizations of the overarching principle. Thus, rather than settling on one definition of axiom

instances, we defined four necessary conditions that a set of instances for an axiom should satisfy.

With possible formalizations of axioms and instances at hand, we then defined Voting by Axioms as a rule that, for a given ranked collection of axioms, returns for every profile that outcome forced by the highest-ranked set in the collection that forces an outcome on the profile. Our analysis of this voting rule started by stating conditions for when it is well-defined. We introduced two systems, viewing forcing as an outcome statement logically following from an axiom set. One solved this question with a propositional logic encoding of the axioms, and we suggested using a SAT-solver to detect forcing. The other one used a tableau calculus to find the answer by constructing tableaux that turn out open or closed and thereby detect whether an outcome is forced.

Thereafter, we described special cases in which the resulting Voting by Axioms rule satisfies one or multiple of the axioms in the underlying collection. Namely, in the case of intraprofile axioms, if all profiles that are connected via an axiom have the same maximal forcing set, and if an axiom set uniquely characterizing some voting rule appears sufficiently high in the collection. These are rare cases in which the derived rule globally satisfies desirable properties. However, we stressed that we are less interested in complete satisfaction of principles, and, rather, aim at satisfying certain intraprofile conditions implied by the axioms, and possibly more instances. To measure how well the rule performs in this respect, we developed satisfaction metrics that count how many of the axioms' instances are satisfied by a rule. Further, we suggested two functions that measure how close an axiom set is to forcing an outcome on a profile.

As a last step in our analysis, we looked at the computational complexity of constructing the Voting by Axioms rule for a given ranked collection of axiom sets. Due to the high complexity of the model itself with exponentially many profiles and outcomes, we concluded that, although deciding whether some given outcome is forced by an axiom on a profile is coNP-complete, the more complex task of building the whole rule lies in $\mathrm{EXP}^{\mathrm{coNP}}$.

After establishing Voting by Axioms, we looked at possible extensions of the framework catered to theoretical and practical applications of the model. This included describing how one can derive a ranking over all possible subsets of an axiom set from an order over the axiom set itself. This was relevant both to account for cognitive limitations of the system user, and to allow for using the Voting by Axioms rule to find the best justification among multiple ones. Another approach to simplifying the user's task to rank the axiom sets in the collection was to consider weak or partial rankings. While we introduced the notion of possible winners to define a procedure in this case, we concluded that it is difficult to infer from this that the rule satisfies any of the underlying axioms. Lastly, we suggested a different method of deriving a voting rule from a collection of axiom sets, this time aiming to satisfy the axioms as much as possible.

Evaluation. Did we do justice to our objective of defining a decision procedure governed by axioms? We defined a voting rule that justifies its outcomes with given axioms. This means isolating each profile and only focusing on conditions imposed by the axioms on this very profile. In this way, we extract conditions that are necessary for the axioms to be fulfilled, i.e., that are minimal requirements for still standing a chance at satisfying the axioms. While this guarantees a minimal coherence of the voting rule with the axioms, it is generally not enough to conclude that the rule satisfies the axioms. Therefore, we obtain a voting rule that is sensible locally on each profile, but whose global behavior is unpredictable. We saw that, as a result, we only obtain a voting rule with desirable properties if the axioms themselves completely prescribe a voting rule's behavior. In particular, the defined rule is usually not strategyproof or robust since any change in the profile might lead to a completely different outcome, forced by a different set of axioms. Thus, from the standard point of view in social choice, where we determine the quality of a voting rule by how many and which axioms it satisfies, the Voting by Axioms rule is generally futile.

We escaped this criticism by challenging the criterion of satisfaction of axioms as the sole indicator for a voting rule's goodness. Instead, we suggested that it may already indicate fine quality if a rule satisfies a subset of all conditions imposed by an axiom. Not all profiles might even be relevant or realistic in a given election (e.g., if there exist a lot of possible profiles and voters, it is unlikely that everyone submits the same ballot) and, as a result, it might not matter so much how a voting rule behaves on them and whether conditions imposed on them are satisfied. This justifies restricting attention to a subset of conditions imposed by an axiom. The major benefit of focusing on these intraprofile conditions, specifying that an outcome is forced, is that we obtain an explainable, transparent procedure. For each outcome, by construction, we have a justification with axioms that we care about available. The voting rule is no longer a black box governed by some mathematical formula, but a human-understandable process, grounded in socially accepted normative principles. Besides, the idea behind the justification of outcomes in voting was a departure from voting rules assessed by global axioms, focusing instead on decisions in single situations. The axiomatic results in Section 3.2 tell us exactly when these two approaches coincide, i.e., for which axioms the local considerations taken together yield the global view.

Moreover, although we paid attention to keeping the model implementable, we discovered that computation is hard. Even calculating the outcome of the Voting by Axioms rule for one specific profile takes exponential time with a coNP-oracle. We also suggested many notions that make use of the extension of an axiom, e.g., the set of profiles that an axioms speaks about $\mathbb{P}(A)$ (Definition 2.2), the introduced axiom hierarchy (Section 2.2.3) or the satisfaction and forcing metrics (Section 3.3). This is problematic, as mentioned previously, since the extension of an axiom, the list of all voting rules satisfying it, is not an object suitable for implementation. The issue is the large number of profiles in $\mathcal{O}(m!^n)$ that yields an even larger number of voting rules. Thus, many of the mentioned ideas are more of mathematical interest than application-oriented.

However, the key mechanism in Voting by Axioms, detecting whether a

specified outcome is forced by an axiom on a profile, is coNP-complete and can be determined by a SAT-solver. This is a well-developed method with many heuristics that yields an acceptable computation performance. Further, computing the outcome of the derived voting rule on one profile is feasible as long as we work with a sufficiently small number of alternatives. Thus, whereas the extensions and special tools are hard to compute, applying the Voting by Axioms rule itself has a reasonable runtime.

Future Research. The developed framework prompts further questions. Regarding the last point of criticism, it would be interesting to search for extension-independent algorithms for the aforementioned notions. We have established upper complexity bounds for Voting by Axioms and argued that likely no significantly more efficient algorithm exists. It would still be desirable to prove a hardness result for Exists-Forc (Section 3.4) and to come up with heuristics for the described problems to obtain algorithms tailored and optimized for Voting by Axioms.

Further, our axiomatic analysis of the derived rule only took into account the axioms contained in the collection that the rule is based on. One could develop other axioms designed especially for Voting by Axioms, e.g., how the rule should behave if an axiom is added, taken away or exchanged. There might also exist other special cases, e.g., when restricting attention to a subclass of axioms, in which we can conclude that the Voting by Axioms rule globally satisfies some axiom.

We did not settle ultimately on what formal object should correspond to an axiom and how to define instances for a given axiom. Further research could compare possible representations of axioms, trying to identify the most suitable one for Voting by Axioms which is expressive enough to capture all (relevant) axioms, succinct enough to be easily implementable, for which identifying the forcing conditions is simple and which, at the same time, yields a natural notion of axiom instance.

Extensions of the framework were only briefly discussed. Depending on the use case, other liftings of orders should be taken into account and may be characterized by order lifting axioms. When searching for the best justification in terms of cost, we saw that the Voting by Axioms rule will assign the outcome that is forced by the axiom set with lowest cost, even if most axiom sets plead for a different outcome. For this application especially, one could develop an algorithm that finds the simplest, yet most backed justification. Besides, our inspection of weak and partial orders led us to consider possible and necessary winners. More work in this area could uncover conditions for when the information about possible and necessary winners is enough to conclude that the Voting by Axioms rule satisfies certain axioms. Lastly, the satisfaction-maximizing procedure rarely singled out one voting rule, so a method needs to be developed on how to decide among the voting rules remaining at the end of the procedure.

Furthermore, an experiment similar to the one by Suryanarayana et al. (2022) would be helpful to understand, which justifications are well-understood

by humans and which impact the choice of the Voting by Axioms rule has on the voting behavior of the electorate. These insights could help in ranking axiom sets or in defining order liftings, and would give a further indication of the quality of the Voting by Axioms rule.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.

Arrow, K. (1951). *Social choice and individual values*. John Wiley & Sons.

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

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

Barberà, S., Bossert, W., & Pattanaik, P. (2004). Ranking sets of objects. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of Utility Theory* (pp. 893–977). Springer.

Biere, A., Heule, M., van Maaren, H., & Walsh, T. (Eds.). (2009). *Handbook of satisfiability*. IOS Press.

Boixel, A., & de Haan, R. (2021). On the complexity of finding justifications for collective decisions. In *Proceedings of the the 35th AAAI Conference on Artificial Intelligence (AAAI-2021)* (pp. 5194–5201).

Boixel, A., & Endriss, U. (2020). Automated justification of collective decisions via constraint solving. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2020)* (pp. 168–176).

Boixel, A., Endriss, U., & de Haan, R. (2022). A calculus for computing structured justifications for election outcomes. In *Proceedings of the the 36th AAAI Conference on Artificial Intelligence (AAAI-2022)*.

Bossert, W., Xu, Y., & Pattanaik, P. (2000). Choice under complete uncertainty: Axiomatic characterizations of some decision rules. *Economic Theory*, *16*, 295–312.

Boutilier, C., & Rosenschein, J. S. (2016). Incomplete information and communication in voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, & A. D. Procaccia (Eds.), *Handbook of Computational Social Choice* (pp. 223–258). Cambridge University Press.

Brewka, G., Eiter, T., & Truszczyński, M. (2011). Answer set programming at a glance. *Communications of the ACM*, *54*(12), 92–103.

Cailloux, O., & Endriss, U. (2016). Arguing about voting rules. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)* (pp. 287–295).

D'Agostino, M. (1999). Tableau methods for classical propositional logic. In

M. D'Agostino, D. M. Gabbay, R. Hähnle, & J. Posegga (Eds.), *Handbook of Tableau Methods* (pp. 45–123). Springer Netherlands.

Davis, M., Logemann, G., & Loveland, D. (1962). A machine program for theorem-proving. *Communications of the ACM*, *5*(7), 394–397.

Davis, M., & Putnam, H. (1960). A computing procedure for quantification theory. *Journal of the ACM*, *7*(3), 201–215.

EU High-Level Expert Group on AI. (2019). *Ethics guidelines for trustworthy artificial intelligence.* `https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419`. (accessed on 27th May, 2022)

Fishburn, P. C. (2015). *The theory of social choice*. Princeton University Press.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, *100*(1), 25–50.

Geist, C., & Peters, D. (2017). Computer-aided methods for social choice theory. In U. Endriss (Ed.), *Trends in Computational Social Choice* (pp. 249–267). AI Access.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*(4), 587–601.

Kaminski, M. (2004). Social choice and information: The informational structure of uniqueness theorems in axiomatic social theories. *Mathematical Social Sciences*, *48*, 121–138.

Konczak, K., & Lang, J. (2005). Voting procedures with incomplete preferences. In *Proceedings of the IJCAI-05 Workshop on Advances in Preference Handling* (pp. 196–201).

Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica*, *28*(2), 287–309.

Kyburg Jr, H. E. (1961). *Probability and the logic of rational belief*. Wesleyan University Press.

Lang, J., & Xia, L. (2016). Voting in combinatorial domains. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, & A. D. Procaccia (Eds.), *Handbook of Computational Social Choice* (pp. 197–222). Cambridge University Press.

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

May, K. (1952). A set of independent necessary and sufficient conditions for simple majority decision. *Econometrica*, *20*, 680.

Mihara, H. R. (1997). Arrow's theorem and Turing computability. *Economic Theory*, *10*(2), 257–276.

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

Nardi, O. (2021). *A graph-based algorithm for the automated justification of collective decisions* (Master's thesis). University of Amsterdam, ILLC.

Nardi, O., Boixel, A., & Endriss, U. (2022). A graph-based algorithm for the automated justification of collective decisions. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2022)* (pp. 935–943).

Pattanaik, P. K., & Peleg, B. (1984). 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.

Peters, D., Procaccia, A. D., Psomas, A., & Zhou, Z. (2020). Explainable voting. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS-2020)* (Vol. 33, pp. 1525–1534).

Plott, C. R. (1976). Axiomatic social choice theory: An overview and interpretation. *American Journal of Political Science*, *20*(3), 511–596.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions. *Journal of Economic Theory*, *10*(2), 187–217.

Scontras, G., Graff, P., & Goodman, N. D. (2012). Comparing pluralities. *Cognition*, *123*(1), 190–197.

Suryanarayana, S. A., Sarne, D., & Kraus, S. (2022). Justifying social-choice mechanism outcome for improving participant satisfaction. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2022)* (pp. 1246–1255).

Thomson, W. (2001). On the axiomatic method and its recent applications to game theory and resource allocation. *Social Choice and Welfare*, *18*(2), 327–386.

Weymark, J. A. (2011). A unified approach to strategy-proofness for single-peaked preferences. *SERIEs*(2), 529–550.

Young, H. P. (1974). An axiomatization of Borda's rule. *Journal of Economic Theory*, *9*(1), 43–52.

Young, H. P. (1975). Social choice scoring functions. *SIAM Journal on Applied Mathematics*, *28*(4), 824–838.

Zwicker, W. S. (2016). Introduction to the theory of voting. In F. Brandt, V. Conitzer, U. Endriss, J. Lang, & A. D. Procaccia (Eds.), *Handbook of Computational Social Choice* (pp. 23–56). Cambridge University Press.