Voting by Axioms

Marie Christin Schmidtlein
ILLC, University of Amsterdam
The Netherlands

Ulle Endriss
ILLC, University of Amsterdam
The Netherlands

ABSTRACT

We develop an approach for collective decision making from first principles. In this approach, rather than using a—necessarily imperfect—voting rule to map any given scenario where individual agents report their preferences into a collective decision, we identify for every concrete such scenario the most appealing set of normative principles (known as axioms in social choice theory) that would entail a unique decision and then implement that decision. We analyse some of the fundamental properties of this new approach, from both an algorithmic and a normative point of view.

KEYWORDS

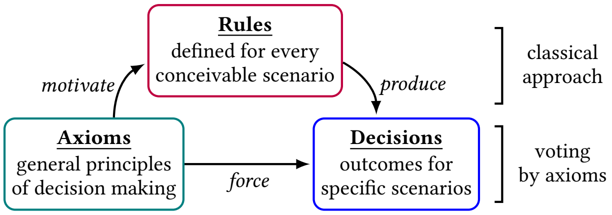
Computational Social Choice; Decision Making; Multiagent Systems

ACM Reference Format:
Marie Christin Schmidtlein and Ulle Endriss. 2023. Voting by Axioms. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 9 pages.

1 INTRODUCTION

There is a well-known mismatch between, on the one hand, seminal results in social choice theory—the principled study of decision making in groups—saying that it is essentially impossible to design an adequate rule for mapping the preferences of individuals into a collective decision [2, 3, 19, 31] and, on the other hand, our everyday experience of "making things work", often by using pragmatic methods—such as the infamous plurality rule—we know to be flawed. In fact, this pragmatic approach is not entirely without scientific justification. Results in behavioural social choice have shown that the problematic scenarios ultimately responsible for the mathematically enticing but otherwise discouraging findings of social choice theory are very rare in practice [30]. But too often the misguided take-away from this observed mismatch is to throw the baby out with the bathwater and to ignore the deep insights about sound and normatively grounded decision making provided by social choice theory altogether.

Instead, in this paper we put forward an approach to collective decision making that is grounded in the axiomatic method of social choice theory [3, 11, 28, 37] but that accounts for the fact that it is impossible to design a "perfect" voting rule that will produce a suitable decision for every conceivable profile of preferences reported by the members of a group. While in classical social choice theory axioms, i.e., formal renderings of normative principles, are used to motivate acceptable voting rules (that can then be applied in any concrete situation we might encounter), in our approach we take decisions from first principles—by appealing *directly* to axioms when proposing a decision in a given situation:



Our approach of *"voting by axioms"* is inspired by a remark in recent work of Boixel and Endriss [9] on explainable decision making who introduce an approach to justifying the outcome of an election by providing a step-by-step explanation of how that outcome is entailed by an appealing set of axioms. They suggest that their approach could also be used to decide such an election in the first place [9, Example 3], but do not develop the idea any further. In this paper, we formalise this idea and introduce the notion of a collection of sets of axioms—ranked from most to least desirable—*forcing* an outcome for a given profile R of preferences. Here, a single set $\mathcal{A}$ of axioms forces a given outcome O on R if every voting rule F that satisfies all the axioms in $\mathcal{A}$ would produce O. When that is the case and when we find $\mathcal{A}$ normatively appealing, then $\mathcal{A}$ provides the perfect justification for choosing O. But often this will not be the case. Then, if we have available a ranked collection of several such sets of axioms, we can see whether the next best set might force an outcome, and so forth. So we end up taking a decision that is suggested by the best possible set of normative principles available to us that actually speaks to the situation at hand.

While our approach, in principle, is relevant to any kind of decision making scenario, in practice it is most suited to high-stakes situations where a fairly small group of agents need to choose between a fairly small number of alternatives and where we require any decision taken to stand on sound normative grounds. Agents here could be human beings who are assisted by decision support technology implementing our approach, or they could be autonomous software agents acting on behalf of human stake-holders.

Related work. In methodological terms, our approach owes much to the development of the axiomatic method in social choice theory [28, 37], starting with the seminal work of Arrow [2]. More specifically, as previously mentioned, our approach is inspired by work of Boixel and Endriss [9] on explainable decision making and thus has links to recent work in this area by several authors [5, 8, 10, 13, 22, 24, 27, 29, 34, 35]. While contributions in that literature tend to focus on the task of generating human-readable explanations for *why* a given decision is forced by a given set of axioms (or, more generally, by a given set of assumptions), our concern here is more basic and fundamental and we ask *whether* that decision is forced in the first place. Finally, there are connections to recent work on using SAT solvers to support the generation of

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

proofs for impossibility theorems in social choice theory [18], an approach pioneered by Tang and Lin [36], because one can use the same kind of encoding of axioms into propositional logic to develop practical implementations of our approach on top of a SAT solver.

Contribution. We develop the *voting by axioms* approach to collective decision making by motivating, formalising, and analysing the concept of axioms *forcing* a collective decision on a given profile of individual preferences. In particular, we establish the computational complexity of deciding whether a given set of axioms is sufficiently strong to force an outcome on a given profile, and we show how the basic algorithmic task of computing a forced outcome can be relegated to a state-of-the-art SAT solver. We furthermore elucidate how *voting by axioms* relates to the classical approach of using axioms to motivate voting rules by establishing conditions under which our approach coincides with applying a reasonable voting rule that satisfies some given axioms.

Roadmap. The remainder of this paper is structured as follows. After recalling some basic notions from social choice theory in Section 2, we develop the *voting by axioms* approach in Section 3. We then analyse this approach further: Section 4 is devoted to algorithmic and Section 5 to axiomatic results. Section 6 concludes. For a further discussion of the approach and additional results we refer to the Master's thesis of the first author [32].

2 PRELIMINARIES

In this section we recall a number of fundamental concepts from social choice theory [3] and fix our notation for speaking about scenarios of collective decision making.

2.1 Basic Notational Conventions

Throughout, for any given finite set S, we use use $\mathcal{P}(S) := \{S' \mid S' \subseteq S\}$ to refer to its powerset and $\mathcal{P}_+(S) := \mathcal{P}(S) \setminus \{\emptyset\}$ to refer to the set of all its nonempty subsets.

We furthermore use $\mathcal{L}(S)$ to denote the set of all strict linear orders on S. These are binary relations that are irreflexive, transitive and connected (so can be used to strictly rank the elements of S from best to worst). For any $\succ \in \mathcal{L}(S)$ and $S' \subseteq S$, we use $\max_{\succ} S'$ to refer to the unique maximal element in S' with respect to $\succ$.

2.2 Voting with Variable Electorates

We work with a standard model of voting with variable electorates commonly used in social choice theory [3], where the set of individuals expressing preferences at any given time may vary. Let $N^* = \{1, \ldots, n\}$ be a finite set of *agents*. We also refer to N^* as the *universe*. Given a set of *alternatives* $X = \{1, \ldots, m\}$, the goal of voting is to determine a favourable subset of these alternatives based on the voters' preferences. For a given *electorate* $N \subseteq N^*$ of agents expressing a preference in a given situation, a *profile* (or *scenario*) R is a function that assigns to each agent $i \in N$ a (reported) preference $R_i \in \mathcal{L}(X)$. Taking into account all possible electorates, we denote the set of all profiles by $\mathcal{L}(X)^+$.

A *voting rule* is is a function $F : \mathcal{L}(X)^+ \to \mathcal{P}_+(X)$ that maps each profile R to an *outcome* $O \subseteq X$. Note that an outcome is a set of (tied-for-best) alternatives rather than a single alternative, so voting rules are *irresolute* in general. Many such rules have been discussed in the literature, including among others the plurality rule, the Borda rule, and the Copeland rule [39].

2.3 Axioms and their Semantics

A central concept in voting theory are so-called *axioms*, normative principles describing what we demand from a good or sensible decision procedure. They are used to analyse voting rules since we take a good voting rule to be one that satisfies a lot of these desirable principles [28, 37, 39]. We will use the following common axioms throughout the paper for concrete examples:

- Anonymity: If profile R' can be obtained from R by renaming the agents, then both should be assigned the same outcome.
- Neutrality: If R' can be obtained from R by renaming the alternatives, then the outcome for R' should be obtainable from the outcome for R via the same renaming process.
- Pareto: If profile R is such that all agents prefer alternative x to y, then y should not be part of the outcome.
- Condorcet: If an alternative x^* wins all pairwise majority contests, then the outcome should be $\{x^*\}$.
- Reinforcement: If the outcomes for two profiles with disjoint electorates have a nonempty intersection, then that intersection should be the outcome when the union of both electorates report preferences.[1]
- Cancellation: If all pairwise contests result in ties, then the outcome should be the set of all alternatives.

Throughout this paper, when talking about axioms more abstractly, we use A and A' to refer to individual axioms, $\mathcal{A}$ and $\mathcal{A}'$ to refer to sets of axioms, and $\mathbb{A}$ to refer to collections of sets of axioms.

While the definitions of specific well-known axioms we sketched above are sufficiently precise for our present purposes, in other situations it can be important to have available a general means to formally define the semantics of axioms. One possible route to take is to encode axioms in a suitable logical language; we are going to briefly discuss this approach in Section 4.1. But the most general approach is to simply equate an axiom A with the set of voting rules F that satisfy A.[2] So, following Boixel and Endriss [9], we define the *interpretation* (or *extension*) $\mathbb{I}(A)$ of an axiom A as the set containing exactly those voting rules F that satisfy A. Similarly, for a set of axioms $\mathcal{A}$, its interpretation is given by the voting rules satisfying all axioms in $\mathcal{A}$ simultaneously, i.e., $\mathbb{I}(\mathcal{A}) = \bigcap_{A \in \mathcal{A}} \mathbb{I}(A)$. We call $\mathcal{A}$ *nontrivial* if $\mathbb{I}(\mathcal{A}) \neq \emptyset$, i.e., if there exists at least one voting rule that satisfies all of the axioms in $\mathcal{A}$.[3]

3 AXIOMATIC FORCING OF DECISIONS

In this section we develop and discuss a formal definition of the core concept we introduce in this paper: the concept of an outcome on a given profile of preferences being *forced* by a corpus of sets

[1] In the original formulation of Reinforcement due to Young [38] electorates are not restricted to subsets of a fixed universe N^*. Imposing this restriction, as we do here, leads to a weaker variant of the axiom, first proposed by Boixel and Endriss [9].

[2] We take this opportunity to stress once more that, for the purposes of of this paper, we are interested in developing a principled approach to collective decision making that does not rely on the application of a voting rule to a given profile of preferences. But voting rules still are a convenient means for fixing the formal semantics of axioms, even though this is not the only way of defining axioms. Indeed, none of the definitions for specific axioms given earlier involved any reference to the concept of voting rule.

[3] Famous examples for axiom sets that are trivial are those involved in so-called *impossibility theorems*, such as the Gibbard-Satterthwaite Theorem [19, 31].

of axioms, ranked from most to least desirable. We begin with the more basic case of a single set of axioms forcing an outcome.

3.1 Simple Forcing

Suppose we are presented with a profile R and a set $\mathcal{A}$ of axioms on which to base our decision which outcome to choose for R. Sometimes the axioms in $\mathcal{A}$ might allow us to exclude certain alternatives from the outcome (e.g., when Pareto is applicable), and sometimes $\mathcal{A}$ might even fully determine—or *force*—a specific outcome O. Let us make this idea formally precise.

Definition 1 (Forcing). *We say that a nontrivial axiom set $\mathcal{A}$ forces an outcome $O \in \mathcal{P}_+(X)$ on a given profile $R \in \mathcal{L}(X)^+$ if every voting rule satisfying the axioms in $\mathcal{A}$ would return that outcome:*

$$F(R) = O \text{ for all } F \in \mathbb{I}(\mathcal{A}).$$

Recall that $\mathcal{A}$ being nontrivial means that $\mathbb{I}(\mathcal{A}) \neq \emptyset$. The case of trivial axiom sets is explicitly excluded from our definition of forcing, because a trivial axiom set would vacuously force every conceivable outcome O on every conceivable profile R.

We write ForcedProf($\mathcal{A}$) for the set of all profiles on which *some* outcome is forced by $\mathcal{A}$. If $R \in$ ForcedProf($\mathcal{A}$), we write ForcedOut($\mathcal{A}, R$) for the unique outcome forced by $\mathcal{A}$ for R.

Example 1. Consider the axiom Condorcet. For any profile with a *Condorcet winner*, i.e., an alternative that is preferred by a majority in every pairwise contest with another alternative, the Condorcet winner should be the single winning alternative. In other words, if R has a Condorcet winner x^*, then Condorcet forces the outcome $\{x^*\}$ on R, i.e., ForcedOut(Condorcet, R) = $\{x^*\}$. △

The concept of forcing is closely related to that of *axiomatic justifications* for election outcomes introduced by Boixel and Endriss [9]. What they call a justification for choosing a given outcome has two parts: (i) a set $\mathcal{A}$ of axioms, called the *normative basis*, such that choosing any other outcome would be incompatible with $\mathcal{A}$, and (ii) a step-by-step *explanation* presented in terms of concrete instances of those axioms illustrating why that is so. So if $\mathcal{A}$ forces O on profile R, it is a normative basis for choosing O for R.

Let us now establish some basic structural properties of our notion of forcing. The first is a simplified version of a result by Boixel and Endriss [9, Theorem 1]. It demonstrates that our notion of forcing is well-defined: for a given axiom set, in any given situation, there only ever is (at most) one outcome that is forced.

Observation 1. *On any given profile R, it is impossible for a nontrivial axiom set $\mathcal{A}$ to force two distinct outcomes O and O'.*

This is the case because forcing the outcome O requires that all voting rules satisfying the axiom set $\mathcal{A}$ return this outcome for profile R. Since voting rules return exactly one outcome for each profile, the same cannot hold for O' as well.

Next we show that, if we add new axioms to a given axiom set, the range of profiles on which an outcome is forced will increase—as long as we do not add so many axioms as to render it trivial.

Proposition 2. *For any two nontrivial axiom sets $\mathcal{A}$ and $\mathcal{A}'$ with $\mathcal{A} \subseteq \mathcal{A}'$, it is the case that* ForcedProf($\mathcal{A}$) $\subseteq$ ForcedProf($\mathcal{A}'$).

Proof. Take any profile $R \in$ ForcedProf($\mathcal{A}$). So for some outcome O it holds for all $F \in \mathbb{I}(\mathcal{A})$ that $F(R) = O$. Since $\mathcal{A} \subseteq \mathcal{A}'$ holds, we have $\mathbb{I}(\mathcal{A}') \subseteq \mathbb{I}(\mathcal{A})$. Thus, in particular for all voting rules $F \in \mathbb{I}(\mathcal{A}')$ it holds that $F(R) = O$, i.e., $\mathcal{A}'$ forces O on R. Therefore, we have $R \in$ ForcedProf($\mathcal{A}'$). Since R was chosen arbitrarily, we may infer that ForcedProf($\mathcal{A}$) $\subseteq$ ForcedProf($\mathcal{A}'$) is true. □

A nontrivial set $\mathcal{A}$ of axioms is said to *characterise* a (unique) voting rule F if F is the only voting rule that satisfies all of the axioms in the set, i.e., if $\mathbb{I}(\mathcal{A}) = \{F\}$. Some of the most important results in the literature on social choice theory are characterisation theorems establishing relationships of this kind [see, e.g. 23, 38]. The following simple result shows how such characterisation theorems can be restated in terms of our notion of forcing.

Proposition 3. *An axiom set $\mathcal{A}$ characterises a unique voting rule if and only if $\mathcal{A}$ forces some outcome on every profile:*

$$\text{ForcedProf}(\mathcal{A}) = \mathcal{L}(X)^+.$$

Proof. If $\mathcal{A}$ characterises some voting rule F^*, then $\mathbb{I}(\mathcal{A}) = \{F^*\}$ is the case and so, trivially, for every $F \in \mathbb{I}(\mathcal{A})$ and every R, we have $F(R) = F^*(R)$. By definition, this means that on every profile R, the axiom set $\mathcal{A}$ forces some outcome, namely $F^*(R)$.

Conversely, suppose that ForcedProf($\mathcal{A}$) = $\mathcal{L}(X)^+$ holds. This means that for every profile R, all $F \in \mathbb{I}(\mathcal{A})$ return $F(R) =$ ForcedOut($\mathcal{A}, R$). But this means that $\mathbb{I}(\mathcal{A})$ only consists of the unique voting rule that maps $R \mapsto$ ForcedOut($\mathcal{A}, R$) for all profiles. In other words, $\mathcal{A}$ characterises this rule. □

In the rare cases where an appealing axiom set $\mathcal{A}$ is known to characterise some voting rule F, our approach of looking for an outcome that is forced by our axioms collapses to the standard approach of simply applying voting rule F to the profile at hand. Still, even then, being able to argue that outcome O is forced by a set of appealing axioms is more satisfying than simply knowing that it is returned by a given voting rule. More problematically, for some appealing axiom sets $\mathcal{A}$ one might want to work with there simply will not be a unique voting rule that directly corresponds to $\mathcal{A}$. This follows from the sparsity of characterisation results—as well as the pervasiveness of impossibility results [2, 19, 31]—in the literature. To account for this, let us broaden the notion of forcing.

3.2 Ranked Forcing

We now generalise the fundamental idea of forcing by working with a ranking of several axiom sets rather than a single such set. This will allow us, for any given profile R, to always look for the highest-ranked set that actually does force an outcome on R.

Let us now make this idea precise. A *ranked axiom corpus* is a pair $\langle \mathbb{A}, \succ \rangle$ consisting of a collection $\mathbb{A}$ of axiom sets and a strict linear order $\succ$ declared on this collection $\mathbb{A}$. We say that $\langle \mathbb{A}, \succ \rangle$ is *nontrivial* if every axiom set $\mathcal{A}$ in $\mathbb{A}$ is nontrivial.

Definition 2 (Ranked Forcing). *We say that a nontrivial ranked axiom corpus $\langle \mathbb{A}, \succ \rangle$ forces an outcome $O \in \mathcal{P}_+(X)$ on a given profile $R \in \mathcal{L}(X)^+$ if at least one axiom set $\mathcal{A} \in \mathbb{A}$ forces some outcome on R and if O is the outcome forced by the top-ranked such axiom set in $\mathcal{A} \in \mathbb{A}$:*

$$O \ := \ \text{ForcedOut}(\max_{\succ}\{\mathcal{A} \in \mathbb{A} \mid R \in \text{ForcedProf}(\mathcal{A})\},\ R).$$

We extend our notation for plain forcing to the case of ranked forcing in the natural manner by writing ForcedOut($\mathbb{A}, \succ, R$) for the outcome that $\langle \mathbb{A}, \succ \rangle$ forces on profile R, and $\textsc{ForcedProf}(\mathbb{A}, \succ)$ for the set of all profiles on which *some* outcome is forced by $\langle \mathbb{A}, \succ \rangle$.

By Proposition 3, $\mathbb{A}$ including an axiom set $\mathcal{A}$ that characterises a unique voting rule is a sufficient (but not a necessary) condition for $\langle \mathbb{A}, \succ \rangle$ forcing an outcome on every profile. So by placing a set $\mathcal{A}$ that characterises some rule (which might not be ideal but offers a decent level of quality) at the bottom of the ranked axiom corpus, we can ensure that there always is an outcome being forced.

Example 2. Let $\langle \mathbb{A}, \succ \rangle$ be a corpus of the form $\{\textsc{Condorcet}\} \succ \mathcal{A}$, where $\mathcal{A}$ is some axiom set characterising the well-known *Borda rule* [38, 39]. It is easy to see that $\{\textsc{Condorcet}\}$ forces an outcome on all profiles that have a *Condorcet winner*, i.e., an alternative that beats all others in pairwise majority contests, and that it does not force an outcome on any other profile. Since $\{\textsc{Condorcet}\}$ is the top-ranked set in the corpus, on Condorcet profiles the corpus $\langle \mathbb{A}, \succ \rangle$ forces the singleton containing the Condorcet winner. On all other profiles, by assumption, $\langle \mathbb{A}, \succ \rangle$ forces the same outcome as would be returned by the Borda rule. Overall, we end up with the same form of decision making as has been proposed by Duncan Black back in 1958, who argued we should choose the Condorcet winner when it exists and otherwise use the Borda rule [7]. △

Also note that by placing an axiom set $\mathcal{A}$ that characterises a unique rule F at the very top of the ranked axiom corpus, *voting by axioms* reduces to simply applying F to produce outcomes. In this sense our approach can be seen as generalising the classical approach of social choice theory, where we first use axioms to motivate voting rules and then apply those voting rules to concrete profiles.

We have not yet commented on the question where $\langle \mathbb{A}, \succ \rangle$ might come from. It seems natural to assume that we might start out with a large set of candidate axioms and then use some or all of the subsets of that set to populate $\mathbb{A}$. But supplying, from scratch, a complete and strict ranking over the sets in $\mathbb{A}$ might be infeasible in practice. First, comparing sets of items is more complex a task for humans than comparing items themselves and, second, the number of axiom sets might be exponential in the number of axioms, yielding an extensive number of items to rank. To overcome this difficulty, an alternative approach is to provide a ranking $\triangleright$ on single axioms and to then *lift* it to a ranking $\succ$ on the sets in $\mathbb{A}$. There are myriad ways of how to lift an order on objects to an order on sets of objects, and there is a large literature on how to axiomatically characterise such *preference extensions* [4]. In our context, there are two natural objectives when lifting a ranking: First, all else being equal, smaller axiom sets should be preferred over larger sets since we want to use only as many axioms to force an outcome as are absolutely needed. Second, axiom sets with highly-ranked axioms should be preferred over those with less desirable axioms. The following example illustrates one way of achieving these objectives.

Example 3. Let $\triangleright$ be a strict linear order over a set of axioms $\mathcal{A}^*$ and let $\mathbb{A}$ be an axiom corpus only featuring axioms from $\mathcal{A}^*$. We define the *shortlex maximax ranking*[4] $\succ$ over $\mathbb{A}$ as follows: $\mathcal{A} \succ \mathcal{A}'$ holds if and only if either $|\mathcal{A}| < |\mathcal{A}'|$, or in case both sets have the same size, there is an index i such that the top i axioms in both sets are the same but the $(i+1)$st-highest ranked axiom in $\mathcal{A}$ is preferred to the one in $\mathcal{A}'$ with respect to $\triangleright$.

For instance, consider the axioms $\textsc{Par} \triangleright \textsc{Con} \triangleright \textsc{Ano}$. This order would be lifted to this order on sets: $\{\textsc{Par}\} \succ \{\textsc{Con}\} \succ \{\textsc{Ano}\} \succ \{\textsc{Par}, \textsc{Con}\} \succ \{\textsc{Par}, \textsc{Ano}\} \succ \{\textsc{Con}, \textsc{Ano}\} \succ \{\textsc{Par}, \textsc{Con}, \textsc{Ano}\}$. △

Another way of generating a ranking $\succ$ on a collection $\mathbb{A}$ of axiom sets would be to associate each axiom A with a cost $c(A)$, calculate for each axiom set the sum of the costs of the axioms it contains, and then rank the sets from cheapest to most expensive. Here, $c(A)$ might reflect the cost of persuading someone to accept A. For example, we might expect that most people are more likely to accept Pareto than, say, Cancellation. To fit our definition of a ranked axiom corpus with a strict ranking $\succ$, care would have to be taken when defining the cost function c, so as to ensure that no two axiom sets end up having the exact same total cost.[5]

Example 4. Again, consider the three axioms from the previous example, now with assigned costs $c(\textsc{Par}) = 1$, $c(\textsc{Con}) = 3$, $c(\textsc{Ano}) = 5$. We obtain the following costs (displayed in the induced order $\succ$ from most to least preferred): $c(\{\textsc{Par}\}) = 1$, $c(\{\textsc{Con}\}) = 3$, $c(\{\textsc{Par}, \textsc{Con}\}) = 4$, $c(\{\textsc{Ano}\}) = 5$, $c(\{\textsc{Par}, \textsc{Ano}\}) = 6$, $c(\{\textsc{Con}, \textsc{Ano}\}) = 8$, $c(\{\textsc{Par}, \textsc{Con}, \textsc{Ano}\}) = 9$. △

3.3 Extensions

As mentioned earlier, it might be a challenge to arrange all sets of an axiom corpus in a coherent, strict ranking. This is so not only because of the possibly very large number of axiom sets involved, but also because one may find two sets incomparable or one may be indifferent between two axiom sets.

To address this concern, one could extend our approach to allow for corpora $\mathbb{A}$ with *weak* or *incomplete orders* $\succeq$. The former are binary relations that are transitive and strongly connected (so allow for clusters of equally preferred elements), while the latter need not be connected (i.e., they allow for two sets to be incomparable). In these cases there not always is a single top-ranked axiom set forcing an outcome, but there could be multiple (either equally-ranked or incomparable) ones. We then can use the solution concept of *possible winners* [21] to define ranked forcing. The idea is that instead of returning the winning set, we determine which alternatives could be in the winning set according to current information. If the ranking were to be refined at a later point in time, the outcome would then be a subset of the set of possible winners.

Formally, a nontrivial (weakly or incompletely) ranked axiom corpus $\langle \mathbb{A}, \succeq \rangle$ forces an outcome $O \in \mathcal{P}_+(X)$ on a given profile $R \in \mathcal{L}(X)^+$ if O is the union of the outcomes forced by the top-ranked axiom sets in $\mathbb{A}$ that force some outcome:

$$O := \bigcup_{\mathcal{A} \in \mathbb{A}^*} \textsc{ForcedOut}(\mathcal{A}, R),$$

where $\mathbb{A}^* := \max_{\succeq}\{\mathcal{A} \in \mathbb{A} \mid R \in \textsc{ForcedProf}(\mathcal{A})\}$ is the set of maximal axiom sets in the corpus with respect to $\succeq$ that force some outcome on profile R.

We are not going to consider this possible extension of our approach any further in the remainder of this paper.

[4] This takes the *lexicographic maximax ranking* by Pattanaik and Peleg [25] as the basis and then analogously applies the principle of *shortlex* [see, e.g. 33] to give priority to shorter sets over larger sets.

[5] One pragmatic way of achieving that no two axioms sets have the same total cost would be to first define a cost function mapping axioms to natural numbers and to then refine that function by adding for each axiom A a small real number chosen uniformly at random from the interval $[0, \varepsilon]$, for a suitably small constant ε.

4 COMPUTATIONAL CONSIDERATIONS

Now that we have motivated and formally introduced the notion of forcing and explained how it can be used for taking collective decisions, in this section we discuss the design of algorithms for determining the outcome forced in a given situation. We begin by showing how SAT solvers can be utilised to this end by encoding axioms in a simple propositional language. We then sketch the limitations of any algorithm designed to determine forced outcomes by establishing the computational complexity of this problem.

Due to the similarities between the forcing of outcomes and axiomatic justifications for outcomes [9] we had noted earlier, there will be close links both to the design of practical algorithms for justifying election outcomes [24] and to the computational complexity of computing axiomatic justifications [8].

4.1 Forcing as Satisfiability Solving

Recall that an axiom set $\mathcal{A}$ forces an outcome O on a profile R if every voting rule F that satisfies $\mathcal{A}$ returns O when applied to R. In other words, proposing any outcome other than O for R would be inconsistent with the axioms in $\mathcal{A}$. So we can think of the task of proving that O is the right outcome as the task of proving that $\mathcal{A}$ together with the assumption that O is *not* the outcome is logically inconsistent. To operationalise this idea, we need to find a way of encoding axioms into a suitable logic. If we succeed, we can make use of state-of-the-art tools for automated reasoning to handle the task of determining forced outcomes. This is precisely the approach we take. There, by now, is much precedent for this approach in the literature on computational social choice, where encodings of axioms into propositional logic have been used so as to be able to utilise modern SAT solvers [6] to reason about scenarios of collective decision making. Most such contributions of this kind have been concerned with offering computational support for proving impossibility theorems [see, e.g., 12, 17, 18, 26, 36], but Nardi et al. [24] utilised this idea to design a practically viable algorithm for computing axiomatic justifications for election outcomes.

So let us define a propositional logic to reason about voting scenarios that allows us to speak about which alternatives win for a given profile. The models of formulas in this logic are given by voting rules. Similar encodings are common in the literature on using SAT solvers to prove impossibility theorems, starting with the work of Tang and Lin [36], and have also been used, for instance, by Cailloux and Endriss [13] and Nardi et al. [24] in the context of computing justifications for election outcomes.

We use propositional variables of the form $p_{R,x}$, where R is the name of a profile and x is the name of an alternative, to express that for profile R alternative x belongs to the outcome. We use the usual connectives to define a propositional language $\mathcal{L}$ of formulas φ:

$$\varphi \quad ::= \quad p_{R,x} \mid \neg\varphi \mid \varphi \vee \varphi \mid \varphi \wedge \varphi$$

Note that a voting rule F can be represented in this language by taking a conjunction over all profiles and alternatives and including either the positive or negative literal, depending on whether the alternative is contained in the voting rule's outcome for the profile in question or whether it is not.[6] Thus, every conceivable axiom A can be expressed in our language $\mathcal{L}$ by taking the disjunction over all the conjunctions corresponding to voting rules in $\mathbb{I}(A)$.

Example 5. The axiom Pareto can be encoded in $\mathcal{L}$ as follows:

$$\bigwedge_{y \in X} \ \bigwedge_{x \in X \setminus \{y\}} \ \bigwedge_{R : \forall i.(x,y) \in R_i} \neg p_{R,y}.$$

Here, the third conjunction operator is intended to range over all profiles R for which it is the case that every agent i who expresses a preference in R ranks x above y. △

Next, let us define the semantics for $\mathcal{L}$, by specifying under which circumstances a given voting rule F satisfies a given formula φ. We say that an atomic formula $p_{R,x}$ is true for the voting rule F (or that F is a model for $p_{R,x}$), denoted by $F \models p_{R,x}$, if and only if $x \in F(R)$ is the case. We recursively extend the truth conditions for the connectives in the familiar way [see, e.g. 14, Chapter 1]. Then we can define the interpretation of a formula φ as the set of all voting rules that φ is true for, i.e., $\mathbb{I}(\varphi) := \{F \mid F \models \varphi\}$. We call a formula φ *satisfiable* if there exists some rule that is a model for it, i.e., if $\mathbb{I}(\varphi) \neq \emptyset$. Note that if φ describes an axiom A, then φ being satisfiable corresponds to A being nontrivial. We say that a set of formulas Σ *logically entails* a formula φ, denoted by $\Sigma \models \varphi$, if for all rules F with $F \models \psi$ for all $\psi \in \Sigma$ it also is the case that $F \models \varphi$.

Now we are in a position to express the notion of forcing as a condition formulated in terms of logical consequence:

Observation 4. *A nontrivial set of axioms $\mathcal{A}$ encoded as formulas in $\mathcal{L}$ forces an outcome O on profile R if and only if*

$$\mathcal{A} \ \models \ \bigwedge_{x \in O} p_{R,x} \ \wedge \ \bigwedge_{x \in X \setminus O} \neg p_{R,x}.$$

By a slight abuse of notation, we will refer to the righthand formula by $p_{R,O}$. It expresses that for profile R outcome O is assigned.

Note that asking whether $\mathcal{A}$ logically entails assigning a specific outcome O is equivalent to determining whether $\mathcal{A}$ together with the condition that O is *not* assigned to R is unsatisfiable. Therefore, we can use a SAT solver to determine whether a nontrivial axiom set $\mathcal{A}$ forces an outcome O. However, in standard propositional logic the models are given by arbitrary valuations as opposed to voting rules. So to feed our problem into a standard SAT solver, we need to add to every formula the requirement that the model is a well-defined voting rule, i.e., that for every R, at least one alternative must be chosen. This corresponds to the following formula:

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

To summarise, a nontrivial axiom set $\mathcal{A}$ forces O on R if and only if SAT returns *false* for the formula $\mathit{atLeastOne} \wedge \left(\bigwedge_{A \in \mathcal{A}} A\right) \wedge \neg p_{R,O}$.

We might also want to characterise—using our logic—when $R \in$ ForcedProf($\mathcal{A}$) is the case, i.e., we might want to be able to tell whether $\mathcal{A}$ forces any outcome at all on R. In view of Observation 4, to force an outcome, $\mathcal{A}$ will have to determine for each alternative whether it should belong to the outcome. Formally, for each x either $p_{R,x}$ or $\neg p_{R,x}$ must be a logical consequence of $\mathcal{A}$. This provides a clear way of computing the forced outcome. Conversely, $\mathcal{A}$ does not force an outcome on R if we can find an alternative for which the axiom does not determine whether it should be in the outcome.

[6] Observe that here we rely on our assumption that the universe N^* is finite, which ensures that any such formula will have finite length.

Observation 5. *A nontrivial axiom set $\mathcal{A}$ does not force any outcome on R if and only if there exists an $x \in X$ such that both*

$$\mathcal{A} \not\models p_{R,x} \quad \textit{and} \quad \mathcal{A} \not\models \neg p_{R,x}.$$

Note that, by definition of the logical consequence relation, this means that there exist voting rules $F_1, F_2 \in \mathbb{I}(\mathcal{A})$ such that $x \in F_1(R)$ and $x \notin F_2(R)$. So since these two rules return distinct outcomes on profile R, no outcome is forced.

Again, we can transform this into a SAT problem by checking for each alternative x whether both $\mathit{atLeastOne} \wedge \left(\bigwedge_{A \in \mathcal{A}} A\right) \wedge p_{R,x}$ and $\mathit{atLeastOne} \wedge \left(\bigwedge_{A \in \mathcal{A}} A\right) \wedge \neg p_{R,x}$ are satisfiable formulas. This provides us with an algorithm for deciding whether a given nontrivial axiom set forces an outcome on a given profile. This algorithm amounts to up to $2m$ calls to a SAT solver (where m is the number of alternatives). If for some alternative both SAT calls succeed, then no outcome is being forced. Otherwise, for each pair of calls exactly one will succeed. Then, from the answers returned by the SAT solver, we can construct the forced outcome in question.

4.2 Computational Complexity

While we have been able to outline a pragmatic approach for computing forced outcomes that exploits the availability of advanced SAT solving technology, it nevertheless is clear that this is a very demanding task. To better understand this problem, we now analyse its computational complexity. We focus on the most fundamental question one can ask in our context, namely whether or not there exists a forced outcome for a given axiom set and a given profile:

ExistsForced

Input: $\mathcal{L}$-encoding of nontrivial axiom set $\mathcal{A}$, profile $R \in \mathcal{L}(X)^+$
Question: Is it the case that $R \in$ ForcedProf$(\mathcal{A})$?

Thus, when presented with a set $\mathcal{A}$ of axioms, all encoded in propositional logic, and a profile R, we ask whether $\mathcal{A}$ forces *some* outcome on R.[7] Unsurprisingly, this problem is computationally intractable. The following result establishes its precise complexity.

Theorem 6. *The problem* ExistsForced *is coNP-complete in the combined size of the $\mathcal{L}$-encodings of the axioms in the set provided.*

Proof. Given that a voting rule satisfies (the $\mathcal{L}$-encodings of) all axioms in $\mathcal{A}$ if and only if it satisfies their conjunction, w.l.o.g. we may assume that $\mathcal{A}$ consists of a single axiom A.

To show coNP-completeness of ExistsForced, we prove NP-completeness of the dual problem: decide whether—for given $\mathcal{A} = \{A\}$ and R—it is *not* the case that $R \in$ ForcedProf$(\{A\})$.

To establish membership in NP for the dual problem, we need to show that there is a way of providing a polynomial-time-verifiable certificate for the claim that A does not force an outcome on R. In principle, a suitable certificate could consist of two voting rules $F_1, F_2 \in \mathbb{I}(A)$ and an alternative x with $x \in F_1(R)$ and $x \notin F_2(R)$, showing that there are two rules that both satisfy A and that return distinct outcomes for R. But for this to be a suitable form of providing certificates, we still need to demonstrate that such a certificate can be represented using a number of bits that is polynomial in the size of (the $\mathcal{L}$-encoding of) A and that we can verify $x \in F_1(R)$ and $x \notin F_2(R)$ in polynomial time.

Recall that there is a one-to-one correspondence between voting rules in $\mathbb{I}(A)$ and models of the formula $\mathit{atLeastOne} \wedge A$. So a concrete way of providing a voting rule as part of a certificate is to provide the corresponding model. Verifying $x \in F_1(R)$ and $x \notin F_2(R)$ then reduces to one simple lookup of the truth value for the variable $p_{R,x}$ in each of the two models, so certainly can be done in polynomial time. What about the size of the certificate? If the encoding of A mentions all propositional letters in the language (as is the case, for instance, for Anonymity), then the size of any of its models is at most that of the size of A—and thus surely polynomial. But for axioms (such as Condorcet) that only refer to a small subset of all possible pairs of profiles and alternatives, this is less obvious (and simply false in some extreme cases). But for such axioms A we can use the following trick. We know that, if the encoding of A never mentions p_{R^*,x^*}, then A does not rule out the possibility that in profile R^* alternative x^* is part of the outcome. So we can restrict attention to rules F_1 and F_2 with $x^* \in F_1(R^*)$ and $x^* \in F_2(R^*)$ for all profiles R^* and alternatives x^* for which p_{R^*,x^*} does not occur in the encoding of A. By using the convention that for any description of a model that does not mention a specific variable the intended meaning is that that variable is set to *true*, we obtain a way of representing rules that requires a number of bits that is polynomial in the size of A. It now remains to be shown that we can always verify in polynomial time that a given model really satisfies $\mathit{atLeastOne} \wedge A$. Such a model-checking operation clearly can be performed in time polynomial in the size of $\mathit{atLeastOne} \wedge A$, but once again this size might be super-exponential in the size of A. But now we can use the same kind of trick as before and simply work with a shortened version of $\mathit{atLeastOne}$ that omits every conjunct $(p_{R^*,1} \vee \cdots \vee p_{R^*,m})$ for which R^* is never mentioned in the encoding of A. This completes the proof of NP-membership.

Next, we show NP-hardness for the dual problem via a polynomial-time reduction from SAT. Suppose we are given a formula φ and want to check its satisfiability. Let $\{p_1, \ldots, p_k\}$ be the set of propositional variables occurring in φ. Set $n = 1$ and define m to be the smallest natural number such that $m! \geq k + 1$ (meaning that $m \geq 2$). Our voting model then contains as many profiles as there are distinct ballots, namely $m!$. Fix an enumeration of the profiles $\mathcal{L}(X)^+ = \{R_1, R_2, \ldots, R_{m!}\}$ and identify the variables p_i with $p_{R_i,1}$ for $i = 1, \ldots, k$. We can now express φ in terms of (some of) the variables $p_{R,x}$. Denote this formula by $A_\varphi := \varphi[p_i / p_{R_i,1}]$. So we replace every occurrence of a variable p_i in φ by $p_{R_i,1}$ for all $i = 1, \ldots k$. Now A_φ expresses whatever φ said, but now speaks about whether alternative 1 wins or looses in profiles $R_1, \ldots, R_k$.

Consider the formula $A := A_\varphi \vee \bigwedge_{x \in X} p_{R_{k+1},x}$, which is satisfiable because the second disjunct always is (it corresponds to assigning the full outcome X to profile R_{k+1}). We now treat A as our axiom and consider the dual problem of ExistsForced for profile R_{k+1}. Recall that A does not force any outcome if and only if we can find an alternative y such that both $\mathit{atLeastOne} \wedge A \wedge p_{R_{k+1},y}$ and $\mathit{atLeastOne} \wedge A \wedge \neg p_{R_{k+1},y}$ are satisfiable. Observe that the first formula is logically equivalent to $(\mathit{atLeastOne} \wedge A_\varphi \wedge p_{R_{k+1},y}) \vee (\mathit{atLeastOne} \wedge \bigwedge_{x \in X} p_{R_{k+1},x})$, which is satisfiable since the second disjunct is. The second formula is logically equivalent to $(\mathit{atLeastOne} \wedge A_\varphi \wedge$

[7] We do not include N^* and X as part of the input of ExistsForced, because this information is implicit when we specify a profile R, which we can think of as a partial function from N^* into the set of preferences over X.

$\neg p_{R_{k+1},y}) \vee (atLeastOne \wedge \neg p_{R_{k+1},y} \wedge \bigwedge_{x\in X} p_{R_{k+1},x})$, which in turn is equivalent to just $atLeastOne \wedge A_\varphi \wedge \neg p_{R_{k+1},y}$.

So deciding the dual problem of ExistsForced (with inputs $\{A\}$ and R_{k+1}) boils down to deciding the satisfiability of $atLeastOne \wedge A_\varphi \wedge \neg p_{R_{k+1},y}$ for all y, aiming at finding one such alternative y for which the formula is satisfiable. But note that the given formula is satisfiable—for any choice of y—if and only if A_φ is. Indeed, $atLeastOne$ can always be satisfied by making $p_{R_\ell,2}$ true for every profile R_ℓ with $\ell \neq k+1$ and making $p_{R_{k+1},y'}$ true for $y' \neq y$. This ensures that the literal $\neg p_{R_{k+1},y}$, too, can always be satisfied since it refers only to profile R_{k+1}, which renders it independent from A_φ. Finally, since we merely renamed the variables, A_φ is satisfiable if and only if φ is satisfiable. This completes the reduction. □

One practical take-away from Theorem 6 is that ExistsForced can be decided with a single call to a SAT solver. This observation might seem to be in conflict with our earlier discussion of a practical algorithm for computing forced outcomes, which involved $2m$ calls to a SAT solver. To understand that, in fact, there is no such conflict, note that the $2m$ calls of our algorithm could in principle be reduced to a single call by renaming all the variables in each of the $2m$ formulas involved and then, for each alternative, constructing the conjunction of both formulas associated with it, and finally taking the disjunction of all these conjoined formulas. This is possible because, to determine that a given axiom set does not force any outcome for a given profile, we do not need to be able to retrieve the individual results for the $2m$ SAT calls but only need to know the result for this combined call.

How should we interpret our complexity result? Boixel and de Haan [8] study the complexity of closely related problems arising in the context of the justification and explanation of election outcomes and show these problems to be hard for complexity classes that, under the usual complexity-theoretic assumptions, are located above coNP. Specifically, for the problem of deciding existence of a justification they prove Σ_2^p-completeness when axioms are encoded in propositional logic (albeit using a somewhat different encoding than we do). So in view of these known results, Theorem 6 should be interpreted as surprising and as good news. The difference in complexity is due to the fact that in our model we do not require a step-by-step explanation for *why* a given outcome is forced.

Having said this, it also is important to not over-interpret our findings. It still is the case that the $\mathcal{L}$-encoding of certain axioms can be huge, and the upper bound established by Theorem 6 of course holds only relative to the size of this encoding.

Finally, while our SAT-based algorithm has the advantage of being completely general, there is room for the development of special-purpose algorithms tailored to specific axioms. For instance, Cailloux and Endriss [13] and Peters et al. [27] have done so for the justification problem with axioms characterising the Borda rule [38]. A class of structurally simple axioms, where such a tailored approach seems particularly promising, are the so-called *algebraic axioms* introduced by Kaminski [20]. There are three kind of basic algebraic axioms (stationary, variance, invariance) from which more complex axioms can be built. While stationary axioms express that a certain outcome is forced on a specified profile, (in-)variance axioms are conditionals stating that, if a special outcome is assigned to a given profile, then some particular outcome needs to be assigned to another profile. It is conceivable that for those axioms there exist more efficient algorithms to determine whether an outcome is forced on a profile. This boils down to answering the question of whether there is a "path" of axioms, starting with a stationary one, followed by conditionals such that the antecedent of one axiom is exactly the consequent of the next one, ending with a consequent that yields a forcing statement about the target profile.

5 AXIOMATIC ANALYSIS

In this section we want to analyse the decision making process developed earlier and present interesting instances thereof. As part of this analysis we consider the case in which ranked forcing induces a voting rule. We then use the axiomatic method to determine when this rule itself satisfies some of the given axioms.

5.1 Intraprofile Axioms

For our first axiomatic result, we identify a family of axioms that are particularly well-behaved in the context of *voting by axioms*. This is the family of what Fishburn [16, Chapter 14] calls the *intraprofile axioms*. Intraprofile axioms have a particularly simple structure in that they only speak about conditions on outcomes "one profile at a time". In our logical encoding, they can be formulated as conjunctions of conditions that each relate to just one single profile.

Example 6. Pareto is an intraprofile axiom since, whenever the axiom requires alternative y to not be part of the outcome for a profile R due to y being dominated by x, this condition can be verified by considering R in isolation. Condorcet and Cancellation also are easily seen to be intraprofile axioms.

On the other hand, Reinforcement is not an intraprofile axiom, because each instance speaks about three profiles, imposing an interdependency, namely that the outcome of the unified election must be the intersection of the disjoint elections' outcomes. Anonymity and Neutrality also are not intraprofile axioms. △

While the notion of intraprofile axiom is used extensively in social choice theory, and while its intuitive meaning is unambiguous, we in fact are not aware of a formal, language-independent definition anywhere in the literature. The following definition closes this gap.

Definition 3 (Intraprofile axioms). *We say that an axiom A is an intraprofile axiom if and only if it is the case that*

$$\mathbb{I}(A) = \bigcap_{R\in\mathcal{L}(X)^+} \{F \mid F(R) \in \{F'(R) \mid F' \in \mathbb{I}(A)\}\}.$$

The set of rules F occurring on the righthand side of the equation is the set of all voting rules that are consistent with the intraprofile conditions that A imposes for profile R. An arbitrary (non-intraprofile) axiom could impose further conditions on how the outcomes of multiple profiles should depend on one another. The equation above, however, requires that no conditions beyond the intraprofile conditions are imposed by the axiom.

We now show that intraprofile axioms are particularly attractive and natural to use for *voting by axioms* since, if such an axiom forces outcomes (possibly jointly with other axioms), not only will every forced outcome be consistent with the axiom but the procedure (across all such profiles) will be as well. Note that, if an axiom requires that the outcome of one profile is dependent on the outcome of another profile, this interdependency might not be preserved when two different axiom sets in our corpus end up forcing outcomes on these profiles. In such a situation, the two forced decisions jointly might not be consistent with the axiom. This problem does not occur for intraprofile axioms since they do not allow for such interdependencies. This is why we can show that, for a given intraprofile axiom A, it is the case that the behaviour of *voting by axioms* across all profiles where A belongs to the top-ranked axiom set forcing an outcome coincides with the behaviour of a voting rule that satisfies A.

Theorem 7. *Given a nontrivial ranked axiom corpus* $\langle \mathbb{A}, \succ \rangle$ *and an intraprofile axiom* A*, there exists a voting rule* $F \in \mathbb{I}(A)$ *such that for all profiles* R *with* $A \in \max_{\succ}\{\mathcal{A} \in \mathbb{A} \mid R \in \textsc{ForcedProf}(\mathcal{A})\}$*, we have* $F(R) = \textsc{ForcedOut}(\mathbb{A}, \succ, R)$*.*

Proof. We want to construct a voting rule that satisfies A and that assigns $\textsc{ForcedOut}(\mathbb{A}, \succ, R)$ whenever A lies in the top-ranked set that forces an outcome on R. It follows from the definition of an intraprofile axiom that a voting rule F satisfies A if and only if, for every profile R, we have $F(R) \in \{F'(R) \mid F' \in \mathbb{I}(A)\}$. Note that, if R is such that A lies in the top-ranked forcing set, then the forced outcome is consistent with A, i.e., $\textsc{ForcedOut}(\mathbb{A}, \succ, R) \in \{F'(R) \mid F' \in \mathbb{I}(A)\}$. Thus, any voting rule that assigns $\textsc{ForcedOut}(\mathbb{A}, \succ, R)$ on such profiles R and any outcomes allowed by the axiom on the other profiles satisfies what we were looking for. □

As suggested before, this theorem is not true for all axioms since there may be multi-profile conditions imposed by the axiom that the forcing disregards. For example, Anonymity *per se* does not determine which alternatives should win for a given profile but it requires that profiles differing only by renaming the agents should have the same outcome. Thus, outcome O might be forced on profile R by one axiom set $\mathcal{A}$ including Anonymity while, on profile R' obtained from R by renaming agents, a completely unrelated outcome O' is forced by another axiom set $\mathcal{A}'$ that also contains Anonymity. Then there would be rules F and F' satisfying Anonymity with $F(R) = O$ and $F'(R') = O'$, yet there is no single anonymous voting rule that will map both these profiles to the given outcomes.

5.2 Induced Rules and Characterisation Results

If $\textsc{ForcedProf}(\mathbb{A}, \succ) = \mathcal{L}(X)^+$, so if the ranked axiom corpus $\langle \mathbb{A}, \succ \rangle$ is sufficiently rich to force an outcome on every possible profile, we can think of ranked forcing as defining a voting rule.

Definition 4 (Induced voting rule). *Let* $\langle \mathbb{A}, \succ \rangle$ *be a nontrivial ranked axiom corpus that forces an outcome on every possible profile. Then the voting rule* $F_{\langle \mathbb{A}, \succ \rangle}$ *induced by* $\langle \mathbb{A}, \succ \rangle$ *is defined as follows:*

$$F_{\langle \mathbb{A}, \succ \rangle} : R \mapsto \textsc{ForcedOut}(\mathbb{A}, \succ, R)$$

We already explained (in Section 3.2) that, if a voting rule F is characterised by $\mathcal{A}$ and we place that axiom set at the top of a ranked axiom corpus, then *voting by axioms* is equivalent to applying F to take decisions. We now want to present a result that further refines this simple insight by establishing that, if characterising axioms are placed high enough in a corpus, then the voting rule induced by ranked forcing will be the characterised rule itself. In particular, the induced voting rule will satisfy the characterising axioms.

Theorem 8. *If an axiom set* $\mathcal{A}$ *uniquely characterises a voting rule* F*, so if* $\mathbb{I}(\mathcal{A}) = \{F\}$*, then for any nontrivial ranked axiom corpus* $\langle \mathbb{A}, \succ \rangle$ *such that we have* $\mathcal{A} \subseteq \max_{\succ}\{\mathcal{A}' \in \mathbb{A} \mid R \in \textsc{ForcedProf}(\mathcal{A}')\}$ *for every profile* R*, the induced voting rule* $F_{\langle \mathbb{A}, \succ \rangle}$ *coincides with the characterised rule* F*.*

Proof. By Proposition 3, we know that $\mathcal{A}$ forces an outcome on every profile. Since, further, $\mathcal{A}$ is contained in the top-ranked forcing set of the corpus for every profile, we know that, for every R, we have $\textsc{ForcedOut}(\mathbb{A}, \succ, R) = \textsc{ForcedOut}(\mathcal{A}, R) = F(R)$. By definition of induced voting rules, this yields $F_{\langle \mathbb{A}, \succ \rangle} = F$. □

As a final observation, we note that the notion of induced voting rule also allows us to formulate a somewhat stronger variant of Theorem 7: If an intraprofile axiom A is contained in the top-ranked axiom set forcing an outcome for every profile R, then the voting rule induced by the given ranked axiom corpus satisfies A.

6 CONCLUSION

We introduced a novel approach to collective decision making from first principles. Instead of using a—necessarily imperfect—voting rule, we proposed to use axioms to determine and justify outcomes in voting scenarios. By using a collection of multiple axiom sets, this approach allows us to involve many (even mutually inconsistent) axioms in the decision process. At the same time it must be noted that this method is only ever as good as the axioms it uses. The decisions taken will be appropriate only if the axioms are.

We presented one way of implementing the framework to compute outcomes based on forcing. The encoding is done in a simple propositional logic that is expressive enough to capture every possible axiom (though it can be difficult to parse encoded axioms due to the high number of propositional variables). This allows us to use a SAT solver to determine forced outcomes. Nonetheless, the complexity result we obtained suggests that applying *voting by axioms* remains a challenging task in practice.

We further explained how our approach can be seen as an extension of the classical approach of social choice theory by showing that *voting by axioms* can sometimes be represented by a voting rule. We highlighted two cases in which our procedure enjoys particularly attractive properties.

Future work should be dedicated to making it easier to use *voting by axioms* in practice. One aspect of this would be to develop a formal language for encoding axioms that is more compact and that lends itself more easily to presenting axioms to users in human-readable form. Some steps in this direction have been taken by Boixel and de Haan [8] and more broadly in the literature on modelling social choice scenarios in mathematical logic [see, e.g., 1, 15]. Another aspect would be to develop heuristics leading to faster algorithms to determine forcing, for instance, along the lines of recent work by Nardi et al. [24]. Related to this point, it also would be interesting to study special classes of axioms, such as the algebraic axioms of Kaminski [20], for which forcing is easier to determine. Finally, more research should be done on how to support users to construct a suitable corpus of axioms and, specifically, a suitable ranking of sets of axioms drawn from that corpus. The *shortlex maximax ranking* and the cost-based approach briefly discussed in this paper are just two of many possible ways of doing this.

ACKNOWLEDGMENTS

We would like to thank our three anonymous AAMAS reviewers for their helpful feedback on our work.

REFERENCES

[1] Thomas Ågotnes, Wiebe van der Hoek, and Michael Wooldridge. 2011. On the Logic of Preference and Judgment Aggregation. *Autonomous Agents and Multiagent Systems* 22, 1 (2011), 4–30.

[2] Kenneth J. Arrow. 1951. *Social Choice and Individual Values*. John Wiley & Sons.

[3] Kenneth J. Arrow, Amartya K. Sen, and Kotaro Suzumura (Eds.). 2002. *Handbook of Social Choice and Welfare*. Vol. 1. Elsevier.

[4] Salvador Barberà, Walter Bossert, and Prasanta K. Pattanaik. 2004. Ranking Sets of Objects. In *Handbook of Utility Theory*, Salvador Barberà, Peter J. Hammond, and Christian Seidl (Eds.). Springer, 893–977.

[5] Khaled Belahcene, Christophe Labreuche, Nicolas Maudet, Vincent Mousseau, and Wassila Ouerdane. 2019. Comparing Options with Argument Schemes Powered by Cancellation. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI-2019)*. 1537–1543.

[6] Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh (Eds.). 2009. *Handbook of Satisfiability*. IOS Press.

[7] Duncan Black. 1958. *The Theory of Committees and Elections*. Cambridge University Press.

[8] Arthur Boixel and Ronald de Haan. 2021. On the Complexity of Finding Justifications for Collective Decisions. In *Proceedings of the the 35th AAAI Conference on Artificial Intelligence (AAAI-2021)*. 5194–5201.

[9] Arthur Boixel and Ulle Endriss. 2020. Automated Justification of Collective Decisions via Constraint Solving. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2020)*. 168–176.

[10] Arthur Boixel, Ulle Endriss, and Ronald de Haan. 2022. A Calculus for Computing Structured Justifications for Election Outcomes. In *Proceedings of the the 36th AAAI Conference on Artificial Intelligence (AAAI-2022)*. 4859–4866.

[11] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice*. Cambridge University Press.

[12] Felix Brandt and Christian Geist. 2016. Finding Strategyproof Social Choice Functions via SAT Solving. *Journal of Artificial Intelligence Research* 55 (2016), 565–602.

[13] Olivier Cailloux and Ulle Endriss. 2016. Arguing about Voting Rules. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)*. 287–295.

[14] Dirk van Dalen. 2004. *Logic and Structure* (4th ed.). Springer.

[15] Ulle Endriss. 2011. Logic and Social Choice Theory. In *Logic and Philosophy Today*, Amitabha Gupta and Johan van Benthem (Eds.). Vol. 2. College Publications, 333–377.

[16] Peter C. Fishburn. 1973. *The Theory of Social Choice*. Princeton University Press.

[17] Christian Geist and Ulle Endriss. 2011. Automated Search for Impossibility Theorems in Social Choice Theory: Ranking Sets of Objects. *Journal of Artificial Intelligence Research* 40 (2011), 143–174.

[18] Christian Geist and Dominik Peters. 2017. Computer-Aided Methods for Social Choice Theory. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access, 249–267.

[19] Allan Gibbard. 1973. Manipulation of Voting Schemes: A General Result. *Econometrica* 41, 4 (1973), 587–601.

[20] Marek Kaminski. 2004. Social Choice and Information: The Informational Structure of Uniqueness Theorems in Axiomatic Social Theories. *Mathematical Social Sciences* 48 (2004), 121–138.

[21] Kathrin Konczak and Jérôme Lang. 2005. Voting Procedures with Incomplete Preferences. In *Proceedings of the IJCAI-2005 Workshop on Advances in Preference Handling*. 196–201.

[22] Sarit Kraus, Amos Azaria, Jelena Fiosina, Maike Greve, Noam Hazon, Lutz Kolbe, Tim-Benjamin Lembcke, Jörg P. Müller, Soren Schleibaum, and Mark Vollrath. 2020. AI for Explaining Decisions in Multi-Agent Environments. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI-2020)*. 13534–13538.

[23] Kenneth O. May. 1952. A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision. *Econometrica* 20 (1952), 680.

[24] Oliviero Nardi, Arthur Boixel, and Ulle Endriss. 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)*. 935–943.

[25] Prasanta K. Pattanaik and Bezalel Peleg. 1984. An Axiomatic Characterization of the Lexicographic Maximin Extension of an Ordering Over a Set to the Power Set. *Social Choice and Welfare* 1 (1984), 113–122.

[26] Dominik Peters. 2018. Proportionality and Strategyproofness in Multiwinner Elections. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2018)*. 1549–1557.

[27] Dominik Peters, Ariel D. Procaccia, Alexandros Psomas, and Zixin Zhou. 2020. Explainable Voting. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS-2020)*, Vol. 33. 1525–1534.

[28] Charles R. Plott. 1976. Axiomatic Social Choice Theory: An Overview and Interpretation. *American Journal of Political Science* 20, 3 (1976), 511–596.

[29] Ariel D. Procaccia. 2019. Axioms Should Explain Solutions. In *The Future of Economic Design*, Jean-François Laslier, Hervé Moulin, M. Remzi Sanver, and William S. Zwicker (Eds.). Springer, 195–199.

[30] Michel Regenwetter, Bernard Grofman, Ilia Tsetlin, and Anthony A. J. Marley. 2006. *Behavioral Social Choice: Probabilistic Models, Statistical Inference, and Applications*. Cambridge University Press.

[31] Mark A. Satterthwaite. 1975. Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory* 10, 2 (1975), 187–217.

[32] Marie Christin Schmidtlein. 2022. *Voting by Axioms*. Master's thesis. ILLC, University of Amsterdam.

[33] Michael Sipser. 2012. *Introduction to the Theory of Computation* (3rd ed.). Cengage Learning.

[34] Sharadhi Alape Suryanarayana, David Sarne, and Sarit Kraus. 2022. Explainability in Mechanism Design: Recent Advances and the Road Ahead. In *Proceedings of the 19th European Conference on Multiagent Systems (EUMAS-2022)*. Springer, 364–382.

[35] Sharadhi Alape Suryanarayana, David Sarne, and Sarit Kraus. 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)*. 1246–1255.

[36] Pingzhong Tang and Fangzhen Lin. 2009. Computer-aided Proofs of Arrow's and other Impossibility Theorems. *Artificial Intelligence* 173, 11 (2009), 1041–1053.

[37] William Thomson. 2001. On the Axiomatic Method and its Recent Applications to Game Theory and Resource Allocation. *Social Choice and Welfare* 18, 2 (2001), 327–386.

[38] H. Peyton Young. 1974. An Axiomatization of Borda's Rule. *Journal of Economic Theory* 9, 1 (1974), 43–52.

[39] William S. Zwicker. 2016. Introduction to the Theory of Voting. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press, 23–56.