Arguing about Voting Rules

Olivier Cailloux
Université Paris-Dauphine
PSL Research University
CNRS, UMR 7243, LAMSADE
75016 Paris, France
olivier.cailloux@dauphine.fr

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam
P.O. Box 94242
1090 GE Amsterdam, The Netherlands
ulle.endriss@uva.nl

ABSTRACT

When the members of a group have to make a decision, they can use a voting rule to aggregate their preferences. But which rule to use is a difficult question. Different rules have different properties, and social choice theorists have found arguments for and against most of them. These arguments are aimed at the expert reader, used to mathematical formalism. We propose a logic-based language to instantiate such arguments in concrete terms in order to help people understand the strengths and weaknesses of different voting rules. Our approach allows us to automatically derive a justification for a given election outcome or to support a group in arguing over which voting rule to use. We exemplify our approach with an in-depth study of the Borda rule.

Keywords

Social Choice Theory; Argumentation; Decision Support

1. INTRODUCTION

When the members of a committee need to make a decision, they can use a *voting rule* to aggregate their individual preferences over the available alternatives to arrive at a suitable collective choice. The normative and mathematical analysis of such voting rules is part of *social choice theory* [1], and their algorithmic properties are studied in *computational social choice* [6]. The significant interest amongst AI researchers in social choice theory in recent years, initially sparked by the relevance of the theory to AI applications in areas such as recommender systems, multiagent systems, and search technologies, has opened up several entirely new perspectives on the old problem of voting and has led to novel synergies with a variety of fields in AI and computer science, such as algorithms, knowledge representation, and machine learning. In this paper, we propose to explore a new such synergy and show how to fruitfully apply ideas from automated reasoning and principles of argumentation as studied in AI to a new kind of problem in voting.

There are many different voting rules: *Plurality*, *Veto*, *Borda*, *Copeland*, *Approval*, and so forth [18]. Each of them satisfies certain appealing properties, but none is perfect. Multiple arguments in favour and against different rules have been put forward in the literature, starting with the famous dispute between Condorcet and Borda in the 18th century [13]. However, these arguments are dispersed in the specialised literature and are often developed in a highly formal and abstract manner. It therefore is difficult, if not impossible, for an untrained individual to understand them. This means that the members of our committee can hardly have an informed discussion about which voting rule to use. We would like to enable such discussions, by making arguments regarding voting rules understandable to non-experts and by providing tools for generating and applying those arguments in concrete situations.

In this paper, we make two contributions towards this long-term goal of enabling informed argumentation about voting rules between non-expert users. First, we develop a general framework for modelling arguments for and against specific outcomes of a voting rule, given a concrete election instance. This framework allows us to represent many important arguments, either new or taken from the literature, and either highly specific or in the general and abstract form of *axioms* encoding high-level properties. Because the framework instantiates these arguments on concrete examples, it does not require the audience to understand the axioms in their full generality. Nevertheless, an argument in our framework can still be general in the sense of being applicable to any concrete election instance. Importantly, our framework is not tailored to defend a specific rule: it permits the use of arguments in favour of different voting rules. As a second contribution, we instantiate our framework by providing an algorithm for generating arguments justifying the outcome recommended by the *Borda rule* for any given election. The technique we use builds on the axiomatisation of that rule developed by Young [21].

Example 1. To illustrate what we ultimately aim for, consider an election with three alternatives, $\{a, b, c\}$, and three voters. Voters v_1 and v_2 both prefer a to b to c, while voter v_3 preferes c to b to a. The situation is summarised in Figure 1 in the form of preference profile $\boldsymbol{R}$. Which alternative wins this election depends on the voting rule used. The Veto rule, for instance, recommends electing the alternative that is ranked at the bottom least often, i.e., it would elect alternative b. The Borda rule, on the other hand, awards 2 points every time an alternative is ranked first, 1 point every time it is ranked second, and 0 points every time it is ranked last, i.e., under the Borda rule alternative a would win (with $2+2+0$ points against $1+1+1$ for b and $0+0+2$ for c). So which alternative is the "right" winner? What we envision is a system that would be able to automatically generate an

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

$$\boldsymbol{R} = \begin{array}{ccc} v_1 & v_2 & v_3 \\ a & a & c \\ b & b & b \\ c & c & a \end{array}, \quad \boldsymbol{R}_1 = \begin{array}{c} v_1 \\ a \\ b \\ c \end{array}, \quad \boldsymbol{R}_2 = \begin{array}{cc} v_2 & v_3 \\ a & c \\ b & b \\ c & a \end{array}.$$

Figure 1: The profiles used in the introductory example. Each column represents the preference ordering of one voter.

easy-to-understand sequence of arguments for justifying, for instance, that alternative a is the deserving winner. Such a system might initiate the following dialogue.

System: Consider election $\boldsymbol{R}_1$, involving only voter v_1 (see also Figure 1). Do you agree that a, enjoying unanimous support, should win this election?

User: Yes, of course.

System: Now consider election $\boldsymbol{R}_2$, involving only voters v_2 and v_3. Do you agree that, for symmetry reasons, the outcome should be a three-way tie?

User: Yes, that sounds reasonable.

System: Observe that when we combine $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, we obtain our election of interest, namely $\boldsymbol{R}$. Do you agree that in this combined election, as there was a three-way tie in $\boldsymbol{R}_2$, $\boldsymbol{R}_1$ should be used to decide the winner?

User: Yes, I do.

System: To summarise, you agree that a should win for $\boldsymbol{R}$.

The reader familiar with the axiomatic method in social choice theory may recognise some of the standard axioms satisfied by the Borda rule at the core of two of these arguments (namely Pareto dominance and reinforcement). We will formally introduce these axioms in Section 2. If the user disagrees with one of the steps, the system might try another strategy of arguing in favour of a. Alternatively, we might also ask our system to generate a sequence of arguments to justify that b should win. $\triangle$

In this paper, we do not address the rendering of such arguments in natural language. Rather, we address the challenge of automatically generating the arguments themselves, expressed in a simple logic-based language. Our framework offers a general solution to the problem of representing such arguments to justify any given outcome for any given election. Of course, a given user will only find some of the arguments that can be represented in principle convincing in practice. For any "natural" voting rule, one should expect that there will be (by virtue of its naturalness) a convincing set of arguments that can be used to justify the outcomes recommend by that rule. The challenge then is to automatically generate a concrete sequence of such arguments for a given outcome to be justified. We provide a solution to this algorithmic problem for the case of the Borda rule.

The remainder of this paper is organised as follows. Section 2 introduces a logic for specifying reasonableness criteria (i.e., axioms) for voting rules. We prove the logic to be complete and demonstrate how it can be used to justify an election outcome. In Section 3 we provide an algorithm for justifying outcomes returned by the Borda rule for arbitrary elections. While our main technical contributions concern the challenge of justifying a given election outcome, in Section 4 we briefly explore other further applications of our approach to richer forms of argumentation about voting rules. Section 5 concludes with a discussion of related work.

2. GENERAL FRAMEWORK

In this section we introduce a formal model of voting rules for variable electorates, we show how to describe such rules and their properties in a simple logical language, and we then use this language to develop a framework for reasoning and arguing about voting rules.

2.1 Voting Rules

We begin by introducing what is essentially the standard formal model of voting familiar from social choice theory [9, 18], with varying sets of voters.

Let $\mathcal{A}$, with $m = |\mathcal{A}|$, be the finite set of *alternatives*. Let $\mathcal{P}_\emptyset(\mathcal{A})$ denote the powerset of $\mathcal{A}$, excluding the empty set. We use the letters $A \subseteq \mathcal{A}$ and $\alpha \subseteq \mathcal{P}_\emptyset(\mathcal{A})$ to designate subsets of alternatives and sets of subsets of alternatives, respectively. We model *preferences* as (strict) linear orders (transitive, irreflexive, and connected binary relations) over $\mathcal{A}$. Let $\mathcal{N}$ be the infinite universe of potential *voters*. A *profile* $\boldsymbol{R}$ is a mapping from a finite subset of voters $N_{\boldsymbol{R}} \subseteq \mathcal{N}$ to linear orders over $\mathcal{A}$. For technical reasons, we allow $N_{\boldsymbol{R}}$ to be empty, in which case we call $\boldsymbol{R}$ the *null profile*. Let $\boldsymbol{\mathcal{R}}$ denote the set of all non-null profiles. A *voting rule* f maps each non-null profile $\boldsymbol{R}$ to a non-empty subset of $\mathcal{A}$, the set of (tied) election winners for the profile in question.

Given a profile $\boldsymbol{R}$, let $\overline{\boldsymbol{R}}$ be the profile consisting of the reverses of the linear orders found in $\boldsymbol{R}$. For two profiles $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ defined over disjoint sets of voters, we define their *sum* $\boldsymbol{R}_1 \oplus \boldsymbol{R}_2$ as the profile $\boldsymbol{R}_1 \cup \boldsymbol{R}_2$. (Note that the union of two functions, considered as sets of input-output pairs, defined over disjoint domains, is itself a well-defined function.) In this paper, we will only use addition of profiles in contexts where the identities of the voters do not matter. Therefore, we also define addition over profiles that are not defined over disjoint sets of voters, the addition then being preceded by an arbitrary renaming of the voters of the second profile. Formally, given two profiles $\boldsymbol{R}_1, \boldsymbol{R}_2$ with $N_{\boldsymbol{R}_1} \cap N_{\boldsymbol{R}_2} \neq \emptyset$, define s as an arbitrary injection mapping every voter $i \in N_{\boldsymbol{R}_2}$ to a voter $s(i) \in \mathcal{N} \setminus N_{\boldsymbol{R}_1}$; define $t(\boldsymbol{R})$ as the profile $\{ (s(i), P) \mid (i, P) \in \boldsymbol{R} \}$; and define $\boldsymbol{R}_1 \oplus \boldsymbol{R}_2 = \boldsymbol{R}_1 \cup t(\boldsymbol{R}_2)$. E.g., for $\boldsymbol{R} = \{ (i, (a, b)) \}$, $\boldsymbol{R} \oplus \boldsymbol{R}$ is $\{ (i, (a, b)), (i', (a, b)) \}$, with $i' \neq i$ an arbitrary voter. A profile $\boldsymbol{R}$ may be multiplied by a natural number $k \in \mathbb{N}$, defined in the natural way as repeated addition with copies of itself and denoted by $k\boldsymbol{R}$. Multiplying a profile by zero yields the null profile. Throughout this paper, natural numbers are taken to include zero.

2.2 Logical Language and Axioms

To formally describe voting rules we will make use of the language of propositional logic over the set of atomic propositions $\{ p_{\boldsymbol{R},A} \mid \boldsymbol{R} \in \boldsymbol{\mathcal{R}},\ \emptyset \subset A \subseteq \mathcal{A} \}$. This set includes one atom for every possible non-null profile $\boldsymbol{R}$ and every possible non-empty subset A of alternatives. The language $\mathcal{L}$ is the set of all formulæ that can be formed using these atoms and the propositional connectives $\neg, \wedge, \vee$, and $\rightarrow$ as well as the special propositions $\top$ and $\bot$, in the usual manner [19]. A *literal* is an atom or its negation; a *clause* is a disjunction of literals.

The semantics of $\mathcal{L}$ is defined as follows. Given a voting rule f, the *model* v_f assigns the value `T` (*true*) to the atom

$p_{\boldsymbol{R},A}$ if $f(\boldsymbol{R}) = A$ and the value F (*false*) otherwise. That is, $p_{\boldsymbol{R},A}$ is true if f chooses A as the set of winners whenever the voters vote as in profile $\boldsymbol{R}$. The definition of v_f extends to the whole set of formulæ using the usual semantics of propositional logic. We say that v_f *satisfies* a set of formulæ *iff* it assigns the value T to every formula in the set.

To make the semantics of the atoms explicit in the language, we from now on write $[\boldsymbol{R} \mapsto A]$ instead of $p_{\boldsymbol{R},A}$. We also write $[\boldsymbol{R} \overset{\in}{\mapsto} \alpha]$, for any non-empty $\alpha \subseteq \mathcal{P}_{\emptyset}(\mathcal{A})$, as a shorthand for $\bigvee_{A\in\alpha}[\boldsymbol{R} \mapsto A]$. We will refer to such clauses involving only one profile, i.e., formulæ specifying the possible sets of winners for a given profile, as *uni-profile clauses*.

We can express familiar as well as new axioms of social choice theory in our language. We call any such rendering of an axiom in $\mathcal{L}$ an *$\mathcal{L}$-axiom*. Formally, an $\mathcal{L}$-axiom is simply a set of formulæ. Here are some examples for $\mathcal{L}$-axioms.

Dom *Dominance* postulates that a Pareto-dominated alternative (i.e., an alternative to which some other alternative is preferred by every voter) should not win. The formulæ are, for each $\boldsymbol{R} \in \mathcal{R}$, $[\boldsymbol{R} \overset{\in}{\mapsto} \mathcal{P}_{\emptyset}(U_{\boldsymbol{R}})]$, where $U_{\boldsymbol{R}}$ is the set of alternatives that are not Pareto-dominated in $\boldsymbol{R}$.

Anon *Anonymity* asks for symmetry w.r.t. voters: for all $\boldsymbol{R} \in \mathcal{R}$, $\emptyset \subset A \subseteq \mathcal{A}$, $N' \subseteq \mathcal{N}$, bijections $\sigma : N' \to N_{\boldsymbol{R}}$, anonymity requires $[\boldsymbol{R} \mapsto A] \to [(\boldsymbol{R} \circ \sigma) \mapsto A]$.

Cond This axiom says that, if there is a *Condorcet winner* (an alternative beating all other alternatives in one-on-one majority contests), then it should be the sole winner: thus, for each profile $\boldsymbol{R}$ with Condorcet winner a, it requires $[\boldsymbol{R} \mapsto \{ a \}]$.

Reinf *Reinforcement* requires that, when joining two profiles for which the winning sets have a non-empty intersection, the resulting profile must have that intersection as the only winners: for each $\boldsymbol{R}_1, \boldsymbol{R}_2, N_{\boldsymbol{R}_1} \cap N_{\boldsymbol{R}_2} = \emptyset, A_1 \cap A_2 \neq \emptyset$, reinforcement imposes the formula $([\boldsymbol{R}_1 \mapsto A_1] \wedge [\boldsymbol{R}_2 \mapsto A_2]) \to [\boldsymbol{R}_1 \oplus \boldsymbol{R}_2 \mapsto A_1 \cap A_2]$.

SymCanc *Symmetric cancellation* says that, when a profile consists of a linear order and its inverse, then the only reasonable outcome is the full set of alternatives: for each such profile $\boldsymbol{R}$, this axiom thus requires $[\boldsymbol{R} \mapsto \mathcal{A}]$.

Reinforcement, also known as *consistency* in the literature, was introduced by Young [21]. Like dominance and the Condorcet principle, it is one of the classical axioms considered in social choice theory [9]. SymCanc is an *ad hoc*, but intuitively appealing, axiom we will use in Example 2.

An $\mathcal{L}$-axiom may also be limited to capturing what an adequate behaviour is on a few specific cases, or even just a single specific case. As an example, let us inspect the argument put forward by Fishburn [8, p. 544] against the Condorcet principle. Consider the profile $\boldsymbol{R}_F$ shown in Figure 2, involving 9 alternatives and 101 voters.[1] Observe that w is the Condorcet winner, as it is preferred to any other alternative by 51 out of 101 voters. Yet, it is intuitively appealing to postulate that alternative a is in fact a more deserving winner of this election. This may be seen by looking at the numbers of times alternatives a and w obtain a given rank (also displayed in Figure 2).

[1]Fishburn explains his argument without giving a fully worked out example. The profile used here is taken from `http://rangevoting.org/FishburnAntiC.html`.

	number of voters					
	31	19	10	10	10	21
1	a	a	f	g	h	h
2	b	b	w	w	w	g
3	c	c	a	a	a	f
4	d	d	h	h	f	w
5	e	e	g	f	g	a
6	w	f	e	e	e	e
7	g	g	d	d	d	d
8	h	h	c	c	c	c
9	f	w	b	b	b	b

	w	a
1	0	50
2	30	0
3	0	30
4	21	0
5	0	21
6	31	0
7	0	0
8	0	0
9	19	0

Figure 2: The profile Fishburn uses to argue against the Condorcet property; and the number of voters placing alternative w or a at a given rank.

FvsC The *Fishburn-versus-Condorcet $\mathcal{L}$-axiom* is defined as the formula $[\boldsymbol{R}_F \mapsto \{ a \}]$.

2.3 Reasoning about Voting Rules

Now that we have a logical language for describing the outcomes of a voting rule for different profiles in place, we want to be able to reason about statements in this language. To this end, we first fix some additional terminology regarding the relationship between $\mathcal{L}$-axioms and voting rules.

Definition 1. An *$\mathcal{L}$-axiomatisation* is a set of $\mathcal{L}$-axioms. A voting rule f *conforms* to the $\mathcal{L}$-axiomatisation J iff v_f satisfies all $\mathcal{L}$-axioms j in J. An $\mathcal{L}$-axiomatisation J is *consistent* iff some voting rule conforms to it. J *characterises* f iff f is the only voting rule conforming to it.

Given a set of assumptions of what makes a good voting rule, expressed in the form of an $\mathcal{L}$-axiomatisation, we want to be able to decide whether a given claim about a given set of alternatives being the deserving winners for a given profile logically follows from those assumptions. In other words, we want to be able to justify election outcomes in terms of a given $\mathcal{L}$-axiomatisation. The next definition fixes the semantics of what it means for such a claim to be valid.

Definition 2. Consider an $\mathcal{L}$-axiomatisation J and a formula φ in our language. We say that φ is a *valid claim* given J iff $v_f(\varphi) = \texttt{T}$ for all voting rules f conforming to J.

We use the term 'claim' instead of 'formula' when we want to emphasise that a formula is used to make a point about specific voting rules. As our proposal is aimed at making arguments as easy to understand as possible, we suggest to restrict claims to uni-profile clauses, which have an easily interpretable meaning. Our results are general however. Note that if J is inconsistent, then all claims are vacuously valid.

We can now define a formal proof system to allow us to establish whether a given claim is valid. Let us first define κ, representing our domain knowledge. It is the set of all formulæ of the form $[\boldsymbol{R} \mapsto A_1] \wedge [\boldsymbol{R} \mapsto A_2] \to \bot$, for all profiles $\boldsymbol{R}$ and $\emptyset \subset A_1 \neq A_2 \subseteq \mathcal{A}$ (saying that a voting rule f cannot select more than one set of winners), plus all formulæ of the form $[\boldsymbol{R} \overset{\in}{\mapsto} \mathcal{P}_{\emptyset}(\mathcal{A})]$ (saying that f must select at least one set of winners). Thus, κ encodes the requirement of f being a function. We now define a proof of a claim φ grounded

$$\boldsymbol{R} = \begin{array}{cccc} a & b & a & c \\ b & c & b & b \\ c & a & c & a \end{array}, \quad \boldsymbol{R}_D = \begin{array}{cc} a & b \\ b & c \\ c & a \end{array}, \quad \boldsymbol{R}_S = \begin{array}{cc} a & c \\ b & b \\ c & a \end{array}.$$

Figure 3: The profiles used in Example 2. Each column represents the preference ordering of one voter.

in J as a demonstration that φ can be inferred from J and κ, i.e., that $(\bigcup J) \cup \kappa \vdash \varphi$. Natural deduction [19], which is widely regarded as producing proofs of good readability, is particularly suited to this purpose, but any other system that is sound and complete for propositional logic could be used as well.

Definition 3. A *proof* of claim φ grounded in $\mathcal{L}$-axiomatisation J is a natural deduction proof for $(\bigcup J) \cup \kappa \vdash \varphi$.

For the purposes of presenting examples, in this paper, we will take certain shortcuts and omit the detailed derivation of simple facts about propositional logic. We will justify such steps as being inferred 'by propositional reasoning' (PR), together with a reference to the premises used. What is important in view of our ultimate goal of justifying election outcomes to users is that any such propositional reasoning step can be decomposed into a sequence of basic steps in a natural deduction proof, which can then be translated into an argument in natural language that can be explained to a non-expert user [2, 14, 20].

Example 2. We prove below, on the basis of $\mathcal{L}$-axioms Dom, SymCanc, and Reinf defined earlier, that the profile $\boldsymbol{R}$ of Figure 3 must have as winners either $\{a\}$, $\{b\}$, or $\{a, b\}$, i.e., that c should not win. Each line consists of a formula we have shown to be true, followed by the justification for that proof step. The profiles $\boldsymbol{R}_D$ and $\boldsymbol{R}_S$ are also defined in Figure 3. Note that $\boldsymbol{R} = \boldsymbol{R}_D \oplus \boldsymbol{R}_S$.

1. $[\boldsymbol{R}_D \overset{\in}{\mapsto} \{\{a\}, \{b\}, \{a, b\}\}]$ (Dom)
2. $[\boldsymbol{R}_S \mapsto \{a, b, c\}]$ (SymCanc)
3. $([\boldsymbol{R}_D \mapsto \{a\}] \wedge [\boldsymbol{R}_S \mapsto \{a, b, c\}]) \to [\boldsymbol{R} \mapsto \{a\}]$ (Reinf)
4. $([\boldsymbol{R}_D \mapsto \{b\}] \wedge [\boldsymbol{R}_S \mapsto \{a, b, c\}]) \to [\boldsymbol{R} \mapsto \{b\}]$ (Reinf)
5. $([\boldsymbol{R}_D \mapsto \{a, b\}] \wedge [\boldsymbol{R}_S \mapsto \{a, b, c\}]) \to [\boldsymbol{R} \mapsto \{a, b\}]$ (Reinf)
6. $[\boldsymbol{R}_D \mapsto \{a\}] \to [\boldsymbol{R} \mapsto \{a\}]$ (PR from 2 & 3)
7. $[\boldsymbol{R}_D \mapsto \{b\}] \to [\boldsymbol{R} \mapsto \{b\}]$ (PR from 2 & 4)
8. $[\boldsymbol{R}_D \mapsto \{a, b\}] \to [\boldsymbol{R} \mapsto \{a, b\}]$ (PR from 2 & 5)
9. $[\boldsymbol{R}_D \mapsto \{a\}] \vee [\boldsymbol{R}_D \mapsto \{b\}] \vee [\boldsymbol{R}_D \mapsto \{a, b\}]$ (rewrite 1)
10. $[\boldsymbol{R} \mapsto \{a\}] \vee [\boldsymbol{R} \mapsto \{b\}] \vee [\boldsymbol{R} \mapsto \{a, b\}]$ (PR from 6–9)
11. $[\boldsymbol{R} \overset{\in}{\mapsto} \{\{a\}, \{b\}, \{a, b\}\}]$ (rewrite 10)

Each of these steps is simple enough to be rendered in natural language, so as to be presented to a non-expert user, just as in Example 1. For instance, steps 2 and 3 directly correspond to steps also present in Example 1, while step 6 might be explained by pointing out that when two premises imply a conclusion, then that conclusion is implied by the first premise alone once we have established that the second premise is in fact true. △

Remark 1. It is important to understand that two $\mathcal{L}$-axioms may be equivalent, logically speaking, while leading to proofs that differ in terms of how easy or difficult they are to understand for a human reader. In our proposal, it is important to choose $\mathcal{L}$-axioms not only according to what they entail (their logical power), but also according to the ease of understanding them. This is similar to the general goal of axiomatising a function: we search for axioms that have, as much as possible, an intuitive content. In our case, however, an $\mathcal{L}$-axiomatisation is good if it strikes an appropriate balance between the lengths of proofs it produces and the intuitiveness of the concrete instantiations of the $\mathcal{L}$-axioms it contains. As an illustration, Reinf could be changed in order to shorten the proof of Example 2. A modified Reinf would say, for example, that a profile associated with a set of possible sets of winners, when added to a profile that has the full set $\mathcal{A}$ as the winners, must still be associated with the same set of possible sets of winners. This axiom would yield, in a single step, that $[\boldsymbol{R}_D \overset{\in}{\mapsto} \{\{a\}, \{b\}, \{a, b\}\}] \wedge [\boldsymbol{R}_S \mapsto \{a, b, c\}]) \to [\boldsymbol{R} \overset{\in}{\mapsto} \{\{a\}, \{b\}, \{a, b\}\}]$.

We now want to show that, with our definition of proofs, we can prove only and all claims that indeed are valid.

Theorem 1 (Completeness). *For any $\mathcal{L}$-axiomatisation J and any claim φ, there exists a proof of φ grounded in J iff φ is valid given J.*

Proof. The theorem follows from the (soundness and) completeness of natural deduction for classical propositional logic [19], together with the fact that there exists a bijection between voting rules conforming to J and models satisfying J and κ. Indeed, for each f conforming to J, the corresponding model v_f satisfies J and κ. Now take any v satisfying J and κ. As the model satisfies κ, we can define a rule f from that model, taking $f(\boldsymbol{R}) = A$ when the model says $[\boldsymbol{R} \mapsto A]$ is true. Because $v = v_f$, f conforms to J. □

Thus, while our logical language permits us to speak about voting rules by making arbitrary claims about the possible sets of winners for a given profile, we now have a proof system in place for deriving any valid such claim from a given axiomatisation provided in the same language. The renderings of the axioms themselves may be long and unwieldy (e.g., Dom explicitly lists all undominated alternatives for every profile), but the concrete proofs produced nevertheless can be expected to be relatively simple and human-readable (as seen in Example 2). Finding the right concrete profiles (e.g., $\boldsymbol{R}_D$ and $\boldsymbol{R}_S$ in Example 2) to use in a proof may be hard, but reading an existing proof is easy. In Section 3 we will address this challenge of actually producing proofs.

3. JUSTIFYING BORDA OUTCOMES

The *Borda rule* is one of the most important voting rules in the literature [18]. Under this rule, an alternative a earns as many points from a given voter as that voter ranks other alternatives below a. The Borda score of an alternative is the sum of points it earns in this manner; the alternatives with the highest Borda score win. For our purposes, it will be convenient to use the following alternative definition.

Definition 4. Given a profile, the *beta score* of an alternative is the sum of the numbers of alternatives it beats in each linear order, minus the sum of the numbers of alternatives it is beaten by in each linear order. Under the *Borda rule* f_B the alternatives with the highest beta score win.

Remark 2. Observe that Borda scores and beta scores define the same rule. Indeed, let n be the number of voters and recall that m is the number of alternatives. The beta score, for a given voter, is $b-(m-1-b)=2b-(m-1)$, where b is the Borda score of that same voter. Thus, the total beta score of an alternative is twice its total Borda score minus $n(m-1)$.

In this section we want to use our logic to justify a given outcome of Borda. That is, starting from any profile $\boldsymbol{R}^*$, we want to be able to give a proof, grounded in $\mathcal{L}$-axioms that are as appealing as possible, for the claim that the only "reasonable" winners must be the ones Borda picks (provided the reader of the argument finds these instantiations of axioms indeed reasonable). We will thus, first, present an $\mathcal{L}$-axiomatisation of Borda and, second, provide an algorithm that, given any $\boldsymbol{R}^*$, builds a proof for $[\boldsymbol{R}^* \mapsto f_B(\boldsymbol{R}^*)]$.

3.1 Borda $\mathcal{L}$-Axiomatisation

To present the $\mathcal{L}$-axiomatisation that we will use to argue in favour of Borda, we require a few definitions. Fix an arbitrary linear order $\succ$ on $\mathcal{A}$. (We will use the alphabetic ordering in our illustrative examples.)

Definition 5. The *elementary profile* $\boldsymbol{R}_e^A$, $\emptyset \subset A \subseteq \mathcal{A}$, has two voters and is defined as follows. Let $k = \succ|_A$ be the restriction of $\succ$ on A and let $\ell = \succ|_{\mathcal{A}\setminus A}$. The first voter has the linear order defined by k then ℓ; the second has $\overline{k}$ then $\overline{\ell}$.

Example 3. The elementary profile $\boldsymbol{R}_e^{\{a,b\}}$ corresponding to $A=\{a,b\}$, with $\mathcal{A}=\{a,b,c,d\}$, is composed of the linear orders (a,b,c,d) and (b,a,d,c). △

Let us call a bijection S on $\mathcal{A}$ an *m-cycle* if $(\mathcal{A}, S)$ is a strongly connected graph, thus, if S represents a cycle that visits each alternative in $\mathcal{A}$ exactly once. It is formally defined as a set of pairs of alternatives, but we will denote such a cycle using a tuple of alternatives, where the first and last alternatives are equal, and all other alternatives appear exactly once. For example, $\langle a,c,b,d,a\rangle$ denotes the m-cycle $\{(a,c),(c,b),(b,d),(d,a)\}$ in $\{a,b,c,d\}$. This cycle can also be represented as $\langle b,d,a,c,b\rangle$. We say that a cycle in $\mathcal{A}$ *generates* $m=|\mathcal{A}|$ linear orders on $\mathcal{A}$, in the natural way. For example, $\langle a,c,b,a\rangle$ generates (a,c,b), (c,b,a), and (b,a,c). We write linear orders with regular parentheses $(\cdots)$ to distinguish them from cycles $\langle\cdots\rangle$. Conversely, observe that a linear order involving all alternatives in $\mathcal{A}$ is generated by exactly one m-cycle.

Definition 6. The *cyclic profile* $\boldsymbol{R}_c^S$, with S an m-cycle, is the profile composed of all m linear orders generated by S.

Example 4. The cyclic profile $\boldsymbol{R}_c^{\langle a,b,c,d,a\rangle}$ corresponding to $S=\langle a,b,c,d,a\rangle$ with $\mathcal{A}=\{a,b,c,d\}$ has the preference orders (a,b,c,d), (b,c,d,a), (c,d,a,b) and (d,a,b,c). △

A *delta vector* δ is a mapping from $\succ$ to the rationals: such a vector has $\binom{m}{2}$ coordinates, each mapping a pair of alternatives to a rational number. For every pair of alternatives $(a,b) \in \succ$, define $\delta_{ba} = -\delta_{ab}$ (slightly abusing notation). The set of delta vectors, denoted by $\boxed{\delta}$, together with addition and multiplication by a rational defined in the natural way, is a vector space.

Definition 7. For any profile $\boldsymbol{R}$, the *delta vector* $\delta^{\boldsymbol{R}}$ maps every $(a,b) \in \succ$ to the signed number of victories of a against b, i.e., $\delta_{ab}^{\boldsymbol{R}}$ is the number of voters who prefer a to b minus the number of voters who prefer b to a.

Thus, $\delta^{\boldsymbol{R}}$ represents the *weighted majority graph* of $\boldsymbol{R}$.

We say that two profiles $\boldsymbol{R}$ and $\boldsymbol{R}'$ *cancel* when $\delta^{\overline{\boldsymbol{R}}} = \delta^{\boldsymbol{R}'}$, thus when $\overline{\boldsymbol{R}}$ and $\boldsymbol{R}'$ have the same weighted majority graph, or equivalently, observing that $\delta^{\overline{\boldsymbol{R}}} = -\delta^{\boldsymbol{R}}$, when $\delta^{\boldsymbol{R}\oplus\boldsymbol{R}'} = \mathbf{0}$, where $\mathbf{0}$ is the zero vector.

Below is the $\mathcal{L}$-axiomatisation that we use for the Borda rule. The fact that it is a *correct* axiomatisation of the Borda rule will follow from Theorem 2 below. As discussed below, in Section 3.4, our axiomatisation is very similar but not identical to the axiomatisation given by Young [21].

Elem For any elementary profile $\boldsymbol{R}_e^A$, the only reasonable set of winners is A: for all $\emptyset \subset A \subseteq \mathcal{A}$, $[\boldsymbol{R}_e^A \mapsto A]$.

Cycl For any cyclic profile $\boldsymbol{R}_c^S$, the only reasonable set of winners is $\mathcal{A}$: for all m-cycles S, $[\boldsymbol{R}_c^S \mapsto \mathcal{A}]$.

Canc If all pairs of alternatives (a,b) are such that a is preferred to b as many times as b is to a, then the set of winners must be $\mathcal{A}$: for all $\boldsymbol{R}$ such that $\delta_{ab}^{\boldsymbol{R}} = 0$ for all $(a,b) \in \succ$, $[\boldsymbol{R} \mapsto \mathcal{A}]$.

Reinf Reinforcement, as defined earlier (cf. Section 2.2).

Reinf-sub Subtracting a profile with a full winner-set does not change the outcome. For all $\boldsymbol{R}, \boldsymbol{R}', \emptyset \subset A \subseteq \mathcal{A}$: $([\boldsymbol{R}\oplus\boldsymbol{R}' \mapsto A] \wedge [\boldsymbol{R}' \mapsto \mathcal{A}]) \rightarrow [\boldsymbol{R} \mapsto A]$.

Simp If a profile consists of a repetition of the same sub-profile, then the sub-profile must have the same winners (i.e., we can simplify): for all $\boldsymbol{R}$, $2 \le k \in \mathbb{N}$, $\emptyset \subset A \subseteq \mathcal{A}$, $[k\boldsymbol{R} \mapsto A] \rightarrow [\boldsymbol{R} \mapsto A]$.

We denote our $\mathcal{L}$-axiomatisation by J_B, the set of all six sets of formulæ just defined.

Remark 3. Observe that Simp and Reinf-sub logically follow from Reinf, i.e., they are in fact not required for the characterisation itself. We introduce them nevertheless, as explained in Remark 1, because they can shorten proofs, and because we assume they will appear sufficiently intuitive to the reader of such proofs to be used without requiring separate justification themselves.

3.2 An Example

Consider the set of alternatives $\mathcal{A}=\{a,b,c,d\}$ and a profile $\boldsymbol{R}^*$ composed of the two preference orders (a,b,d,c) and (c,b,a,d). Observe that Borda selects $\{a,b\}$ as winners for this profile. We will now build a proof grounded in J_B of the claim $[\boldsymbol{R}^* \mapsto \{a,b\}]$.

The proof consists of two parts. First (steps 1–8 in this example), we define a profile $\boldsymbol{R}'$ that is the sum of several profiles for which the winners are uncontroversial, either because of Elem or because of Cycl, and use this to argue that our Borda winners should win for $\boldsymbol{R}'$. For our example, let $\boldsymbol{R}_E = \boldsymbol{R}_e^{\{a,b\}} \oplus 2\boldsymbol{R}_e^{\{a,b,c\}}$, $\boldsymbol{R}_C = \boldsymbol{R}_c^{\langle a,d,c,b,a\rangle} \oplus \boldsymbol{R}_c^{\langle a,b,d,c,a\rangle}$, and $\boldsymbol{R}' = \boldsymbol{R}_E \oplus \boldsymbol{R}_C$. Second (steps 9–16 in this example), we argue that $\boldsymbol{R}'$ must have the same winners as $\boldsymbol{R}^*$. This works, because we also chose $\boldsymbol{R}'$ in such a way that it has the same weighted majority graph as some multiple of $\boldsymbol{R}^*$. Indeed, step 12 uses the fact that $\overline{4\boldsymbol{R}^*}$ and $\boldsymbol{R}'$ cancel (this can be verified manually by counting the number of wins for each pair of alternatives). Step 9 is valid, as any profile cancels with its inverse.

1. $[\boldsymbol{R}_e^{\{a,b\}} \mapsto \{a,b\}]$ (Elem)
2. $[\boldsymbol{R}_e^{\{a,b,c\}} \mapsto \{a,b,c\}]$ (Elem)
3. $[\boldsymbol{R}_c^{\langle a,d,c,b,a\rangle} \mapsto \mathcal{A}]$ (Cycl)
4. $[\boldsymbol{R}_c^{\langle a,b,d,c,a\rangle} \mapsto \mathcal{A}]$ (Cycl)
5. $((1) \wedge (2)) \rightarrow [\boldsymbol{R}_E \mapsto \{a,b\}]$ (Reinf)
6. $((3) \wedge (4)) \rightarrow [\boldsymbol{R}_C \mapsto \mathcal{A}]$ (Reinf)
7. $([\boldsymbol{R}_E \mapsto \{a,b\}] \wedge [\boldsymbol{R}_C \mapsto \mathcal{A}]) \rightarrow [\boldsymbol{R}' \mapsto \{a,b\}]$ (Reinf)
8. $[\boldsymbol{R}' \mapsto \{a,b\}]$ (PR from 5–7)
9. $[4\boldsymbol{R}^* \oplus \overline{4\boldsymbol{R}^*} \mapsto \mathcal{A}]$ (Canc)
10. $([4\boldsymbol{R}^* \oplus \overline{4\boldsymbol{R}^*} \mapsto \mathcal{A}] \wedge [\boldsymbol{R}' \mapsto \{a,b\}]) \rightarrow [4\boldsymbol{R}^* \oplus \overline{4\boldsymbol{R}^*} \oplus \boldsymbol{R}' \mapsto \{a,b\}]$ (Reinf)
11. $[4\boldsymbol{R}^* \oplus \overline{4\boldsymbol{R}^*} \oplus \boldsymbol{R}' \mapsto \{a,b\}]$ (PR from 8–10)
12. $[\overline{4\boldsymbol{R}^*} \oplus \boldsymbol{R}' \mapsto \mathcal{A}]$ (Canc)
13. $([4\boldsymbol{R}^* \oplus \overline{4\boldsymbol{R}^*} \oplus \boldsymbol{R}' \mapsto \{a,b\}] \wedge [\overline{4\boldsymbol{R}^*} \oplus \boldsymbol{R}' \mapsto \mathcal{A}]) \rightarrow [4\boldsymbol{R}^* \mapsto \{a,b\}]$ (Reinf-sub)
14. $[4\boldsymbol{R}^* \mapsto \{a,b\}]$ (PR from 11–13)
15. $[4\boldsymbol{R}^* \mapsto \{a,b\}] \rightarrow [\boldsymbol{R}^* \mapsto \{a,b\}]$ (Simp)
16. $[\boldsymbol{R}^* \mapsto \{a,b\}]$ (PR from 14 & 15)

Simplifications are possible. For instance, step 8 could be presented to a user as following directly from steps 1–4, together with Reinf and basic propositional reasoning.

3.3 The General Algorithm

We now define an algorithm, *Borda-expl*, which, given any profile $\boldsymbol{R}^*$, builds a proof grounded in J_B of the claim $[\boldsymbol{R}^* \mapsto f_B(\boldsymbol{R}^*)]$, i.e., a justification for the Borda outcome. Our proofs all have the same structure as in the example above; only the concrete profiles used along the way differ. We show how to compute a natural number r, a profile $\boldsymbol{R}_E$ that is the sum of several elementary profiles, and a profile $\boldsymbol{R}_C$ that is the sum of several cyclic profiles such that, for $\boldsymbol{R}' = r\boldsymbol{R}_E \oplus \boldsymbol{R}_C$, (i) the winners for $\boldsymbol{R}'$ are $f_B(\boldsymbol{R}^*)$, and (ii) $rm\boldsymbol{R}^*$ and $\boldsymbol{R}'$ have the same weighted majority graph.

First, let us define $\boldsymbol{R}_E$. Define a beta vector as a vector mapping alternatives from $\mathcal{A}$ to rationals, with the condition that it sums to zero. The set of beta vectors, denoted by $\boxed{\beta}$, together with addition and multiplication by a rational defined in the natural way, is a vector space. We write $\beta^{\boldsymbol{R}} = \langle \beta_a^{\boldsymbol{R}}, a \in \mathcal{A}\rangle$ for the beta vector corresponding to a profile $\boldsymbol{R}$, where $\beta_a^{\boldsymbol{R}}$ denotes the beta score of a in $\boldsymbol{R}$. Name alternatives $a_1, a_2, \dots, a_m$ by decreasing beta score in $\boldsymbol{R}^*$, thus $\beta_{a_1}^{\boldsymbol{R}^*} \geq \beta_{a_2}^{\boldsymbol{R}^*} \geq \dots \geq \beta_{a_m}^{\boldsymbol{R}^*}$. Define $\boldsymbol{R}_E = \bigoplus_{i=1}^{m-1} \frac{\beta_{a_i}^{\boldsymbol{R}^*} - \beta_{a_{i+1}}^{\boldsymbol{R}^*}}{2} \boldsymbol{R}_e^{\{a_1,\dots,a_i\}}$.

Remark 4. This definition of $\boldsymbol{R}_E$ is legal as the coefficients are natural numbers: $(\beta_{a_i}^{\boldsymbol{R}^*} - \beta_{a_{i+1}}^{\boldsymbol{R}^*})$ is even because, depending on m, either all beta scores are even, or all are odd (as may be seen by revisiting Remark 2).

We can now show that $\boldsymbol{R}_E$ has the same beta scores as $m\boldsymbol{R}^*$, which is equivalent to a result due to Young [21].

Lemma 1 (Young, 1974). *With the above definitions, for each $a \in \mathcal{A}$: $\beta_a^{\boldsymbol{R}_E} = \beta_a^{m\boldsymbol{R}^*}$.*

Proof. First observe that $\forall \emptyset \subset A \subseteq \mathcal{A}$, $\beta_a^{\boldsymbol{R}_e^A}$ equals $2(m - |A|)$ if $a \in A$ and $-2|A|$ if $a \notin A$. Write β_a instead of $\beta_a^{\boldsymbol{R}^*}$. Now $\beta_{a_i}^{\boldsymbol{R}_E} = \sum_{j=1}^{i-1} \frac{\beta_{a_j} - \beta_{a_{j+1}}}{2}(-2j) + \sum_{j=i}^{m-1} \frac{\beta_{a_j} - \beta_{a_{j+1}}}{2} 2(m-j) = \sum_{j=1}^{m-1}(\beta_{a_j} - \beta_{a_{j+1}})(-j) + \sum_{j=i}^{m-1}(\beta_{a_j} - \beta_{a_{j+1}})m = (\sum_{j=1}^{m-1} -\beta_{a_j}) + (m-1)\beta_{a_m} + m\beta_{a_i} - m\beta_{a_m} = (\sum_{j=1}^{m} -\beta_{a_j}) + m\beta_{a_i}$. The claim now follows from $\sum_{a\in\mathcal{A}} \beta_a = 0$. □

Thus, $\boldsymbol{R}_E$ and $m\boldsymbol{R}^*$ have the same Borda winners.[2] We now have to define $\boldsymbol{R}'$. Recall that we want $\boldsymbol{R}'$ and $rm\boldsymbol{R}^*$ to have equal weighted majority graphs for some r. Thus, our objective is to obtain $\delta^{\boldsymbol{R}'} = \delta^{rm\boldsymbol{R}^*}$. Assume we can find a set of m-cycles $\mathcal{S}$ and rationals $\langle q_S, S \in \mathcal{S}\rangle$ that solve the linear system of equations $\delta^{m\boldsymbol{R}^*} = \delta^{\boldsymbol{R}_E} + \sum_{S\in\mathcal{S}} q_S \delta^{\boldsymbol{R}_c^S}$. (We will shortly define $\mathcal{S}$ and prove that this system always admits a solution.) Because $\delta^{\boldsymbol{R}_c^S} = -\delta^{\boldsymbol{R}_c^{-S}}$, where $-S$ denotes the inverse cycle of S, we can then choose coefficients q_S that are all non-negative. Then, it remains only to define r as the smallest strictly positive integer such that $\{ rq_S, S \in \mathcal{S}\}$ are all natural numbers, and to define $\boldsymbol{R}_C = \bigoplus_S rq_S \boldsymbol{R}_c^S$. Indeed, $\boldsymbol{R}' = r\boldsymbol{R}_E \oplus \boldsymbol{R}_C$ then has the same winners as $\boldsymbol{R}_E$ (hence, the same as $\boldsymbol{R}^*$), and $\delta^{rm\boldsymbol{R}^*} = \delta^{\boldsymbol{R}'}$. (The comment made in Footnote 2 applies here as well.)

It remains to show that the system above can be solved. Indeed, even considering all $(m-1)!$ m-cycles as the set $\mathcal{S}$, the claim that such a system can always be solved requires a proof. Furthermore, our proof permits to use a quadratic—instead of a factorial—number of unknowns, as we define precisely which cycles $\mathcal{S}$ must contain.

Lemma 2. *There exists a set of m-cycles $\mathcal{S}$ such that, for any two profiles $\boldsymbol{R}^*, \boldsymbol{R}_E$ satisfying $\beta^{m\boldsymbol{R}^*} = \beta^{\boldsymbol{R}_E}$, there exist rationals q_S with $\sum_{S\in\mathcal{S}} q_S \delta^{\boldsymbol{R}_c^S} = \delta^{m\boldsymbol{R}^*} - \delta^{\boldsymbol{R}_E}$.*

In order to prove Lemma 2, we will first define a beta transformation $\hat{\beta}$, a linear transformation from $\boxed{\delta}$ to $\boxed{\beta}$. We will then show that $\delta^{m\boldsymbol{R}^*} - \delta^{\boldsymbol{R}_E}$ belongs to its kernel $\mathcal{K}(\hat{\beta})$ (Lemma 3). Next, we will define $\mathcal{S}$ and show that we can find rationals q_S with $\sum_{S\in\mathcal{S}} q_S \delta^{\boldsymbol{R}_c^S} = \delta$, for any delta vector $\delta \in \mathcal{K}(\hat{\beta})$. Equivalently, defining $\rho = \left\{ \delta^{\boldsymbol{R}_c^S}, S \in \mathcal{S} \right\}$, we will show that ρ spans $\mathcal{K}(\hat{\beta})$ (Lemma 4). In order to do this, observing that $\rho \subseteq \mathcal{K}(\hat{\beta})$, we will show that $\dim \mathcal{K}(\hat{\beta}) = \dim \rho$, or equivalently, that ρ has $\dim \mathcal{K}(\hat{\beta})$ independent vectors.

Define the beta transformation $\hat{\beta}(\delta)$ of a delta vector δ as the following beta vector: $\hat{\beta}(\delta)_a = \sum_{b\in\mathcal{A}\setminus\{a\}} \delta_{ab}$. Let $\mathcal{K}(\hat{\beta})$ denote the kernel of the beta transformation, the vector space of vectors δ such that $\hat{\beta}(\delta) = \mathbf{0}$. Because $\hat{\beta}$ is a surjective linear transformation from $\boxed{\delta}$ of dimension $\binom{m}{2}$ to $\boxed{\beta}$ of dimension $m-1$, $\dim \mathcal{K}(\hat{\beta}) = \binom{m}{2} - (m-1)$.

Lemma 3 (Young, 1974). *For any two profiles $\boldsymbol{R}^*, \boldsymbol{R}_E$ satisfying $\beta^{m\boldsymbol{R}^*} = \beta^{\boldsymbol{R}_E}$, $\delta^{m\boldsymbol{R}^*} - \delta^{\boldsymbol{R}_E} \in \mathcal{K}(\hat{\beta})$.*

Proof. Because $\hat{\beta}$ is linear and for any $\boldsymbol{R}$, $\beta^{\boldsymbol{R}} = \hat{\beta}(\delta^{\boldsymbol{R}})$, we get $\hat{\beta}(\delta^{m\boldsymbol{R}^*} - \delta^{\boldsymbol{R}_E}) = \beta^{m\boldsymbol{R}^*} - \beta^{\boldsymbol{R}_E}$. By the hypothesis of this lemma, this equals zero. Thus, $\delta^{m\boldsymbol{R}^*} - \delta^{\boldsymbol{R}_E} \in \mathcal{K}(\hat{\beta})$. □

[2] If $\boldsymbol{R}_E$ or $\boldsymbol{R}_C$ are null profiles, then the Borda winners for these profiles are undefined. We describe here our algorithm assuming these profiles to be non-null. The reader will easily find out which modifications are required in those simpler cases.

$\mathcal{S}$	S^{ab}	S^{ac}	S^{bc}
	$\langle a,$	$\langle b,$	$\langle b,$
	$b,$	$a,$	$c,$
	$c,$	$c,$	$a,$
	$d,$	$d,$	$d,$
	$a\rangle$	$b\rangle$	$b\rangle$

T	ab	ac	ad	bc	bd	cd
ab	1	0	−1	0	1	0
ac	0	1	−1	0	0	1
bc	0	0	0	1	−1	1

M	$\delta^{\boldsymbol{R}_c^{S^{ab}}}$	$\delta^{\boldsymbol{R}_c^{S^{ac}}}$	$\delta^{\boldsymbol{R}_c^{S^{bc}}}$
ab	2	−2	0
ac	0	2	−2
ad	−2	0	2
bc	2	0	2
bd	0	−2	−2
cd	2	2	0

TM	k^{ab}	k^{ac}	k^{bc}
ab	4	−4	−4
ac	4	4	−4
bc	4	4	4

Figure 4: Illustrations for Lemma 2, with $\mathcal{A} = \{\,a, b, c, d\,\}$. The set $\mathcal{S}$ (up, right) defines the delta vectors in ρ forming the columns of the matrix M (middle, right), which get transformed by T (middle, left), obtaining $T\delta^{\boldsymbol{R}_c^{S^{tu}}} = k^{tu}$.

Lemma 4. *There exists a set of m-cycles $\mathcal{S}$ such that $\rho = \left\{\,\delta^{\boldsymbol{R}_c^{S}}, S \in \mathcal{S}\,\right\}$ spans $\mathcal{K}(\hat{\beta})$.*

Proof. Let z denote the least alternative in $\succ$. For $(t, u) \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$, define S^{tu} as the m-cycle constituted by all alternatives that are in between t and u in $\succ$ (in the order they come in $\succ$), followed by t, followed by u, followed by all alternatives that come after u in $\succ$ except z (in the order they come in $\succ$), followed by all alternatives that come before t (in the reverse order of the order they come in $\succ$), followed by z. Let $\mathcal{S}$ be the set of m-cycles $\{S^{tu}, (t, u) \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}\}$ (Figure 4, top right).

Let ρ denote the set of vectors $\left\{\,\delta^{\boldsymbol{R}_c^{S}}, S \in \mathcal{S}\,\right\}$ (Figure 4, middle right). Let us show that ρ spans $\mathcal{K}(\hat{\beta})$. We leave it to the reader to check that $\rho \subseteq \mathcal{K}(\hat{\beta})$. It remains to show that ρ is constituted of $\binom{m}{2} - (m-1) = \binom{m-1}{2}$ independent vectors. We do so by transforming each of the $\binom{m-1}{2}$ vectors in ρ using linear combinations of their coordinates. Their independence will then be visible from their simple format. This is equivalent to defining a linear transform T, defining a matrix M constituted by the vectors $\delta \in \rho$ as column vectors, and showing that TM is non-singular.

Let us transform $\delta^{tu} = \delta^{\boldsymbol{R}_c^{S^{tu}}} \in \rho$ into a new vector k^{tu}, of size $\binom{m-1}{2}$, with coordinates indexed using the pairs in ${\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$. For each pair $(v, w) \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$, define $k^{tu}_{vw} = -\delta^{tu}_{vz} + \delta^{tu}_{vw} + \delta^{tu}_{wz}$. Equivalently, define a $\binom{m-1}{2} \times \binom{m}{2}$ matrix T whose line T_{vw}, $vw \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$, is defined as $T^{vz}_{vw} = -1$, $T^{vw}_{vw} = T^{wz}_{vw} = 1$, the rest of the line being zero (Figure 4, middle left). This yields $T_{vw}\delta^{tu} = k^{tu}_{vw}$.

$$\boldsymbol{R}_c^{\langle a,b,c,y,d,e,x,a\rangle} = \begin{array}{ccccccc} a & b & c & y & d & e & x \\ b & c & y & d & e & x & a \\ c & y & d & e & x & a & b \\ y & d & e & x & a & b & c \\ d & e & x & a & b & c & y \\ e & x & a & b & c & y & d \\ x & a & b & c & y & d & e \end{array}.$$

Figure 5: Illustration of the computation of the delta score corresponding to an m-cycle, here with $d = 4$ the distance between x and y in the cycle.

Observe that $\delta^{tu}_{xy} = m - 2d_{xy}$ for any $xy \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$, where d_{xy} is the number of alternatives in between x and y in the cycle S^{tu}, counting y but not x. To see this, recall δ^{tu}_{xy} is the signed number of victories of x against y in the profile $\boldsymbol{R}_c^{S^{tu}}$, composed of all preference orders obtained by starting the cycle S^{tu} at different positions. Assume the distance between x and y is d in S^{tu}. See Figure 5. Consider first the preference order starting with the $(d-1)$ alternatives in between x and y, then y, then the remaining $m-d$ alternatives in $\mathcal{A}$ ending with x. In this preference order, y is better than x, and this will be the case for the preference orders of the first d voters where we gradually shift (cyclically) y towards the first position in the ranking. The next shift will make y beaten by x, and it will remain so for the rest of the voters. Thus, x has won $m-d$ times and y has won d times.

We obtain $k^{tu}_{vw} = 3m - 2(d_{zv} + d_{vw} + d_{wz})$. Now only two cases need to be distinguished, thanks to the order of the alternatives in the cycle S^{tu} compared to the ordering $\succ$. Consider as an example ${\succ} = (a, b, c, t, d, e, f, u, g, h, z)$, $S^{tu} = \langle d, e, f, t, u, g, h, c, b, a, z, d\rangle$. (The reasoning which follows is general though.) Because (t, u) and (v, w) are taken in ${\succ}|_{\mathcal{A}\setminus\{\,z\,\}}$, we know that in $\succ$, t comes before u and v comes before w, and all are different from z. Consider any $v \in \mathcal{A} \setminus \{\,z\,\}$. Assume $t \succ v$. Then, w being in between v and z in $\succ$ implies that w is in between v and z in S^{tu}. Similarly, w is in between v and z in S^{tu} whenever $(t = v \wedge u = w)$ or $(t = v \wedge u \succ w)$. It is easy to check that in all other cases, v is in between w and z in S^{tu}. Finally, observe that $d_{zv} + d_{vw} + d_{wz}$ equals m when w is in between v and z in S^{tu} and equals $2m$ otherwise.

We conclude that $k^{tu}_{vw} = m$ whenever $t \succ v \vee (t = v \wedge u = w) \vee (t = v \wedge u \succ w)$, and $k^{tu}_{vw} = -m$ in the remaining cases. Hence, TM (Figure 4, bottom) has its entries on the diagonal and below equal to m and the rest equal to $-m$, which shows it is nonsingular, or equivalently, that $\{k^{tu}, tu \in {\succ}|_{\mathcal{A}\setminus\{\,z\,\}}\}$ is a set of $\binom{m-1}{2}$ independent vectors. □

Once suitable values for r, $\boldsymbol{R}_E$ and $\boldsymbol{R}_C$ have been found, it suffices to present a proof following the structure presented in Section 3.2 (or simple modifications thereof, in case some of the profiles found are null profiles). This terminates the proof of correctness of our Borda-expl algorithm.

Theorem 2 (Borda justification). *For any given profile,* Borda-expl *computes a proof justifying the outcome of the Borda rule in terms of the $\mathcal{L}$-axiomatisation J_B.*

Proof. This is a consequence of Lemma 2, which itself follows from Lemma 3 and Lemma 4. □

3.4 Comparison with Young's Axiomatisation

This instantiation of our framework is based on the axiomatisation of the Borda rule given by Young [21]. Young used the axioms of *neutrality*, *faithfulness*, *cancellation*, and *reinforcement*. We chose a slightly different set of axioms to make the argument more concrete and the proofs built by our algorithm shorter. In particular, we included elementary and cyclic profiles in the axioms themselves (rather than make them follow from the axioms of Young).

Young's work also inspired our approach to proving correctness of our algorithm. Lemmas 1 and 3, the idea of using the spaces $\boxed{\delta}$, $\boxed{\beta}$ and the transformation $\widehat{\beta}$ are due to him. The novelty in our approach is that our vectors can be defined from a restricted set of cyclic profiles, whereas Young uses a more general construction. Thus, the results specific to the space of cyclic profiles (the construction and exploitation of $\mathcal{S}$ as done in Lemma 4) are new.

4. BEYOND OUTCOME JUSTIFICATION

In this section we briefly explore additional opportunities for putting our general framework to use and sketch how it may be applied to argue about voting rules in other ways that simply justifying a given outcome.

4.1 Types of Arguments

Proofs of claims may be used in various ways to argue in favour of one voting rule or to attack another rule. There are clear links with argumentation theory [3], which could be further developed to arrive at a fully fledged framework for arguing about voting rules. Here we only define a few categories of arguments we can create in our framework. In the context of a voting rule f, a proof for a claim $[\boldsymbol{R} \overset{\in}{\mapsto} \alpha]$, saying that in profile $\boldsymbol{R}$ the set of winners should be selected from α, can constitute different types of arguments:

- a *partial justification* for f when $f(\boldsymbol{R}) \in \alpha$;
- a *full justification* for f on $\boldsymbol{R}$ when $\alpha = \{ f(\boldsymbol{R}) \}$;
- an *attack* against f when $f(\boldsymbol{R}) \notin \alpha$.

An argument may belong to more than one of these categories, e.g., it may simultaneously be a justification for some rule and an attack against some other rules.

An argument can also attack an $\mathcal{L}$-axiomatisation instead of a specific voting rule. A system using an $\mathcal{L}$-axiomatisation J could establish that J' is incompatible with J (meaning that voting rules conforming to J necessarily give different results in some cases from rules conforming to J') and, assuming that the user will favour J over J' when realising that they are incompatible, could thus argue by simply giving an example illustrating the incompatibility. It is then up to that system to choose its example as wisely as possible. Formally, an attack against J' by J consists of two proofs, one of $[\boldsymbol{R} \overset{\in}{\mapsto} \alpha]$ grounded in J and one of $[\boldsymbol{R} \overset{\in}{\mapsto} \alpha']$ grounded in J', for some profile $\boldsymbol{R}$ and some sets α and α' with $\alpha \cap \alpha' = \emptyset$. An attack against J' is also an attack against any rule f' conforming to J'.

4.2 Attacking and Defending Borda

As an illustration, we present here, first, an argument that could be given against Borda, namely, that it does not satisfy the Condorcet property. We then also show how to defend Borda against this argument by producing a counter-argument to the Condorcet argument.

Consider $J_C = \{ \textsc{Cond} \}$, including only the $\mathcal{L}$-axiom saying that, if there is a Condorcet winner, it must be returned as the sole winner. Now take any profile with a Condorcet winner where Borda does not select that Condorcet winner. For example, take $\mathcal{A} = \{ a, b, c \}$ and $\boldsymbol{R}$ defined as follows:

$$\boldsymbol{R} = \begin{matrix} b & b & a & a & a \\ c & c & b & b & b \\ a & a & c & c & c \end{matrix} .$$

Although a is the Condorcet winner, Borda shamelessly selects $\{ b \}$. Thus, an attack against Borda can be built by putting forward the claim $[\boldsymbol{R} \mapsto \{ a \}]$ and its (trivial) proof grounded in J_C, whilst observing that this contradicts Borda's choice.

As a defence, a system arguing for Borda may give a justification for choosing $\{ b \}$ using its own $\mathcal{L}$-axiomatisation, by giving an argument for $[\boldsymbol{R} \mapsto \{ b \}]$ grounded in J_B as computed by Borda-expl. But this is unlikely to be convincing: such an attack rather calls for a more specific response. The system could also counter-attack by saying that we do not want to follow Condorcet in general, by using Fishburn's argument. Define J'_B as the set of $\mathcal{L}$-axioms for Borda described above, together with FvsC, the Fishburn-versus-Condorcet $\mathcal{L}$-axiom (see Section 2.2). An attack against J_C can now be produced by giving a proof grounded in J'_B for $[\boldsymbol{R}_F \mapsto \{ a \}]$, together with a proof grounded in J_C for $[\boldsymbol{R}_F \mapsto \{ w \}]$. This shows the incompatibility between these two $\mathcal{L}$-axiomatisations.

5. CONCLUSION AND RELATED WORK

We have developed a general logic-based framework for representing arguments in favour of or against specific election outcomes. While these arguments can be based on general axioms familiar from social choice theory, when actually used, they apply to concrete instances of elections, thereby making them understandable to non-experts. We have also devised a practical algorithm for generating the arguments required to justify the election outcome selected by the Borda rule, for any given profile of preferences.

Related work has aimed at explaining or justifying recommendations [4, 10, 11, 12] or outcomes of elections [15, 16]. However, these approaches are all based on specific ways of justifying decisions and propose no general framework capable of integrating different kinds of arguments, including in particular counter-arguments against their own claims.

Our work is also related to existing work on logic and automated reasoning for social choice theory [5, 7, 17], aimed at automatically deriving theorems in social choice theory. However, to date work in that literature has not attempted to tackle the problem of justifying election outcomes.

Acknowledgments

Part of this work was carried out while the first author was employed at the ILLC, University of Amsterdam, and then at the Université de technologie de Compiègne, France. It was supported by the Dutch Government through an NWO Free Competition Grant (No. 612.001.015), and funded in the framework of the Labex MS2T supported by the French Government through the programme "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02), respectively.

References

[1] K. J. Arrow, A. K. Sen, and K. Suzumura, editors. *Handbook of Social Choice and Welfare*, volume 1. North-Holland, 2002.

[2] Y. Bertot and L. Théry. A generic approach to building user interfaces for theorem provers. *Journal of Symbolic Computation*, 25(2):161–194, 1998.

[3] P. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[4] D. Bouyssou and M. Pirlot. Conjoint measurement tools for MCDM. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 73–132. Springer-Verlag, 2005.

[5] F. Brandt and C. Geist. Finding strategyproof social choice functions via SAT solving. In *Proc. 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2014)*, pages 1193–1200, 2014.

[6] F. Brandt, V. Conitzer, and U. Endriss. Computational social choice. In G. Weiss, editor, *Multiagent Systems*, pages 213–283. MIT Press, 2013.

[7] U. Endriss. Logic and social choice theory. In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*, volume 2, pages 333–377. College Publications, 2011.

[8] P. C. Fishburn. Paradoxes of voting. *The American Political Science Review*, 68(2):537–546, 1974.

[9] W. Gaertner. *A Primer in Social Choice Theory*. LSE Perspectives in Economic Analysis. Oxford University Press, 2006.

[10] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, 2nd edition, 1993.

[11] C. Labreuche. A general framework for explaining the results of a multi-attribute preference model. *Artificial Intelligence*, 175(7–8):1410–1448, 2011.

[12] C. Labreuche, N. Maudet, V. Mousseau, and W. Ouerdane. Explanation of the robust additive preference model by even swap sequences. In *Proc. 6th Multidisciplinary Workshop on Advances in Preference Handling*, pages 21–26, 2012.

[13] I. McLean and A. B. Urken. *Classics of Social Choice*. University of Michigan Press, 1995.

[14] A. Ranta. Translating between language and logic: What is easy and what is difficult. In N. Bjørner and V. Sofronie-Stokkermans, editors, *Proc. 23rd International Conference on Automated Deduction (CADE-2011)*, pages 5–25. Springer-Verlag, 2011.

[15] D. G. Saari. Explaining all three-alternative voting outcomes. *Journal of Economic Theory*, 87(2): 313–355, 1999.

[16] D. G. Saari. *Decisions and Elections: Explaining the Unexpected*. Cambridge University Press, 2001.

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

[18] A. D. Taylor. *Social Choice and the Mathematics of Manipulation*. Cambridge University Press, 2005.

[19] D. van Dalen. *Logic and Structure*. Springer-Verlag, 5th edition, 2013.

[20] M. Wenzel. Isar: A generic interpretative approach to readable formal proof documents. In *Proc. 12th International Conference on Theorem Proving in Higher Order Logics (TPHOLs-1999)*, pages 167–183. Springer-Verlag, 1999.

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