A Syntactic Proof of Arrow's Theorem in a Modal Logic of Social Choice Functions

Cinà, G.; Endriss, U.

Publication date
2015
Document Version
Final published version
Published in
AAMAS '15

A Syntactic Proof of Arrow's Theorem in a Modal Logic of Social Choice Functions

Giovanni Cinà
Institute for Logic, Language and Computation
University of Amsterdam
giovanni.cina88@gmail.com

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam
ulle.endriss@uva.nl

ABSTRACT

We show how to formalise Arrow's Theorem on the impossibility of devising a method for preference aggregation that is both independent of irrelevant alternatives and Pareto efficient by using a modal logic of social choice functions. We also provide a syntactic proof of the theorem in that logic. While prior work has been successful in applying tools from logic and automated reasoning to social choice theory, this is the first human-readable formalisation of the Arrovian framework allowing for a direct derivation of the theorem.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent Systems*; J.4 [**Social and Behavioral Sciences**]: Economics

General Terms

Theory; Economics

Keywords

Social Choice Theory; Logic

1. INTRODUCTION

Social choice theory is the study of mechanisms for collective decision making [24]. This includes voting rules as mechanisms to collectively make political decisions, and consequently social choice theory is chiefly associated with the disciplines of political science and economics. But similar mechanisms can also be used to make decisions in multiagent systems, to coordinate the actions of individual agents, to resolve conflicts between them, and to bundle their information and expertise [6]. Closely related applications of social choice theory in computer science furthermore include recommender systems [20], Internet search engines [2], and crowdsourcing [14].

This widening of the scope of social choice theory has renewed interest in the formal foundations of the field. As we are designing ever more specialised social choice mechanisms for novel types of tasks, better tools to analyze the formal properties of these mechanisms are needed. Specifically, there is now a growing literature on the formal verification of social choice mechanisms by means of logical modelling and the use of techniques from automated reasoning [1, 4, 7, 9, 10, 12, 16, 23, 26, 28]. (We will review some of the contributions to this field in Section 4.)

An obvious yardstick against which to measure different approaches to the formalisation of social choice frameworks is Arrow's Theorem [3], *the* seminal result in the field, which shows that it is impossible to design preference aggregation mechanisms for three or more alternatives that are Pareto efficient and for which the relative ranking of two alternatives is based only on the rankings for the same two alternatives submitted by the individual voters. For instance, recent work has modelled the Arrovian framework in propositional logic [23], first-order logic [12], higher-order logic [16, 28], and a tailor-made modal logic [1]. Some of this work has resulted in methods to prove Arrow's Theorem either automatically [23] or semi-automatically [16, 28], while other work has generated logical formalisations of the theorem that are easily accessible to humans and thus helpful in deepening our understanding of social choice [1, 12]. A shortcoming of the latter contributions, however, is that they have so far not resulted in a full proof of Arrow's Theorem or similar results *within* the chosen logical framework itself. Rather, such work has proceeded by showing that a given logical system is complete w.r.t. an appropriate class of models of social choice theory, thereby proving that a rendering of Arrow's Theorem in the logical language in question must be a theorem of that logic. That is, such work has derived results about a given logic by means of reference to existing "semantic" proofs of Arrow's Theorem. The ultimate goal of such research, however, must be the opposite: to use the logic to derive proofs for Arrow's Theorem and similar results. In this paper, we close this gap by providing a syntactic proof of Arrow's Theorem within a simple tailor-made modal logic.

Our logic of choice is a fragment of the *modal logic of social choice functions* proposed by Troquard et al. [26]. Troquard et al. have used their (full) logic to reason about the strategy-proofness of voting rules (but it has not previously been applied to Arrow's Theorem). This logic can be used to model a (resolute) *social choice function* (SCF), i.e., a function that maps any given profile of preference orders to a single winning alternative. While Arrow originally formulated his theorem for social welfare functions, i.e., functions that map any given profile of preference orders to a single social preference order [3], we will instead work with a standard variant of the theorem for SCF's [24]. Arguably, SCF's (returning a top alternative rather than a full ranking of all

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

alternatives) are relevant to a wider range of applications. In any cases, known techniques to prove either version of the theorem are very similar [9, 24]. We will show how to model Arrow's Theorem for SCF's in our logic and then present a full proof of the theorem from a system of axioms that is shown to be complete for our logic.

The remainder of this paper is organised as follows. Section 2 recalls the definition of SCF's, and then introduces our logic of SCF's and establishes completeness for it. Section 3 shows how to formalise Arrow's Theorem for SCF's in the logic and then presents our proof. Finally, Section 4 discusses related work and Section 5 concludes.

2. LOGIC AND SOCIAL CHOICE

In this section, we recall the formal definition of a SCF and introduce the fragment of the logic put forward by Troquard et al. [26] required to define such a SCF, adapting some of their notation and terminology to our purposes. We then demonstrate that the known completeness theorem for the full logic extends to the fragment that is of interest to us here. Finally, we discuss the limitations of this logic in view of expressing properties of families of SCF's ranging over electorates of varying size, as well as how to overcome these limitations in practice.

2.1 Social Choice Functions

Let $N = \{1, \ldots, n\}$ be a finite set of *agents* (or *individuals*) and let X be a finite set of *alternatives* (or *candidates*). To vote, each agent $i \in N$ expresses her preferences by supplying a linear order $\succcurlyeq_i$ over X, i.e., a binary relation that is reflexive, antisymmetric, complete, and transitive.[1] Let $\mathcal{L}(X)$ denote the set of all such linear orders. We shall also refer to $\succcurlyeq_i$ as the *ballot* provided by agent i, to stress the fact that this is the preference declared by the agent, but not necessarily her true preference. A *profile* is an n-tuple $(\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$ of such ballots, one for each agent.

Definition 1. *A resolute social choice function is a function $F : \mathcal{L}(X)^n \to X$ mapping any given profile of ballots to a single winning alternative.*

Examples for resolute SCF's are well-known voting rules, such as the Borda rule or the plurality rule [24]—when combined with a suitable tie-breaking rule that ensures that there always is just a single winner.

2.2 Language

Troquard et al. [26] have introduced a modal logic, which they call $\Lambda^{\text{scf}}[N, X]$, to reason about resolute SCF's (mapping declared preferences to winners) as well as the agents' truthful preferences. This logic can be used to model strategic behaviour in voting. Here we are not specifically interested in this strategic component, but rather in the purely aggregative aspect of social choice, i.e., in the question of whether a given SCF fairly aggregates individual ballots into a social decision. For the purposes of the present paper, we shall refer to the relevant fragment of the logic of Troquard et al. as $L[N, X]$, the *logic of SCF's* parametrised by N and X. Next, we define the language, i.e., the set of well-formed formulas, of this logic.

This language is built on top of two types of atomic propositions. First, for every $i \in N$ and $x, y \in X$, $p^i_{x \succcurlyeq y}$ is an atomic proposition (with the intuitive meaning that agent i prefers x to y). $\mathit{Pref}[N, X] := \{p^i_{x \succcurlyeq y} \mid i \in N \text{ and } x, y \in X\}$ is the set of all such propositions. Second, by a slight abuse of notation, every alternative $x \in X$ is also an atomic proposition (with the intuitive meaning that x wins). Besides the usual propositional connectives, we have a modal operator $\Diamond_C$ for every coalition of agents $C \subseteq N$ (with the intuitive meaning that C can ensure the truth of a given formula, provided the others do not alter their ballots). Thus:

Definition 2. *The set of well-formed formulas in the language of $L[N, X]$ is defined as follows:*

(*i*) *All atomic propositions of the form $p \in \mathit{Pref}[N, X]$ and $x \in X$ are formulas.*

(*ii*) *If φ and ψ are formulas, then so are $\neg\varphi$ and $\varphi \vee \psi$.*

(*iii*) *For any $C \subseteq N$, if φ is a formula, then so is $\Diamond_C \varphi$.*

(*iv*) *Nothing else is a formula.*

Additional propositional connectives and a dual modal operator are defined in the usual manner: $\varphi \wedge \psi$ is short for $\neg(\neg\varphi \vee \neg\psi)$, $\varphi \to \psi$ is short for for $\neg\varphi \vee \psi$, $\varphi \leftrightarrow \psi$ is short for for $(\varphi \to \psi) \wedge (\psi \to \varphi)$, and $\Box_C \varphi$ is short for $\neg\Diamond_C \neg\varphi$. For $i \in N$, we write $\Diamond_i$ as a shorthand for $\Diamond_{\{i\}}$ and $\Box_i$ as a shorthand for $\Box_{\{i\}}$.

The full logic of Troquard et al. [26] includes an additional pair of modal operators to speak about true preferences.

2.3 Semantics

The semantics of the logic is a standard possible-worlds semantics for modal logics, defined in terms of a set of possible worlds, a family of accessibility relations, and a valuation function [5]. We first give a short high-level description intended for readers familiar with such semantics, and then provide complete formal definitions.

First, the set of possible worlds is the set of all possible profiles—which is fully determined by N and X. The semantics of atomic propositions of the form $p^i_{x \succcurlyeq y}$ will be defined solely in terms of this set of possible worlds: $p^i_{x \succcurlyeq y}$ is true at a given world/profile w, if agent i prefers x to y in w. Only to model the truth of atomic propositions of the form x will we require a valuation function. Valuation functions here are SCF's: x is true at world/profile w if the SCF in question maps profile w to the winning alternative x. Finally, for every coalition $C \subseteq N$, there is an accessibility relation between worlds/profiles: w is connected to w' if they differ only w.r.t. the preferences of agents in C. These accessibility relations will be used to define the semantics of modal formulas of the form $\Diamond_C \varphi$ in the usual manner.

Definition 3. *A model is a triple $M = \langle N, X, F \rangle$, consisting of a finite set of agents N with $n = |N|$, a finite set of alternatives X, and a SCF $F : \mathcal{L}(X)^n \to X$.*

For fixed sets N and X, we sometimes write M_F for the model $M = \langle N, X, F \rangle$ based on the SCF F. From now on we shall use the terms *'world'* and *'profile'* interchangeably. We are now ready to define what it means for formula φ to be true at world $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$ in a given model M.

[1] The strict part $\succ_i$ of $\succcurlyeq_i$ is a strict linear order, a relation that is irreflexive, complete, and transitive. While most work in voting theory tends to take such strict linear orders as primitive, we instead follow Troquard et al. [26] and work with non-strict linear orders. Ultimately, both approaches are equivalent: $\succcurlyeq_i$ uniquely determines $\succ_i$, and *vice versa*.

DEFINITION 4. *Let $M = \langle N, X, F\rangle$ be a model. We write $M, w \models \varphi$ to express that the formula φ is true at the world $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$ in M. The satisfaction relation $\models$ is defined inductively:*

- $M, w \models p^i_{x \succcurlyeq y}$ *iff* $x \succcurlyeq_i y$
- $M, w \models x$ *iff* $F(\succcurlyeq_1, \ldots, \succcurlyeq_n) = x$
- $M, w \models \neg\varphi$ *if* $M, w \not\models \varphi$
- $M, w \models \varphi \vee \psi$ *iff* $M, w \models \varphi$ *or* $M, w \models \psi$
- $M, w \models \Diamond_C \varphi$ *iff* $M, w' \models \varphi$ *for some world* $w' = (\succcurlyeq'_1, \ldots, \succcurlyeq'_n) \in \mathcal{L}(X)^n$ *with* $\succcurlyeq_i = \succcurlyeq'_i$ *for all* $i \in N \setminus C$.

That is, $\Diamond_C \varphi$ is true at w, if the agents in C can make φ true by changing their own ballots (assuming none of the other agents change as well). Thus, $\Box_C \varphi$ is true at w if φ holds at every world that is reachable from w by only the agents in C changing their ballots.

In some sense, the truth of every formula of the form $p^i_{x \succcurlyeq y}$ is under the control of agent i. Because of this feature, the logic is classified as a *logic of propositional control*. The motivation underlying such logics is essentially game-theoretic: every individual is conceived as having "control" over a set of atomic propositions. The choice of a particular truth value for these atomic propositions can be seen as an action of the individual, and therefore a valuation of all the atomic propositions of this sort corresponds to a strategy profile. For more details and motivations on logics of propositional control we refer to the work of van der Hoek and Wooldridge [27], Gerbrandy [11], and Troquard et al. [26], amongst others.

Let φ be a formula in the language based on N and X. Then φ is called *satisfiable*, if there exist a SCF F and a world $w \in \mathcal{L}(X)^n$ such that $M_F, w \models \varphi$. It is called *true in the model* M, denoted $M \models \varphi$, if $M, w \models \varphi$ for every world $w \in \mathcal{L}(X)^n$. Finally, it is called *valid*, denoted $\models \varphi$, if $M \models \varphi$ for every model M based on N and X.

The logic of Troquard et al. [26] is known to be decidable and this result immediately extends to the fragment of their logic discussed here:

PROPOSITION 1. *Determining whether a formula in the language of $L[N, X]$ is valid is a decidable problem.*

PROOF. Since N and X are fixed, we can enumerate all models and check for each of them whether our formula is true at every world in the model. □

2.4 Axiomatisation and Completeness

Next, we review the axiomatisation due to Troquard et al. [26], restricted to the fragment $L[N, X]$ discussed here, and then adapt their completeness result to this fragment. The first group of axioms ensure that the propositions of the form $p^i_{x \succcurlyeq y}$ really encode linear orders.

(1) $p^i_{x \succcurlyeq x}$ (reflexivity)

(2) $p^i_{x \succcurlyeq y} \leftrightarrow \neg p^i_{y \succcurlyeq x}$ for $x \neq y$ (antisym./completeness)

(3) $p^i_{x \succcurlyeq y} \wedge p^i_{y \succcurlyeq z} \rightarrow p^i_{x \succcurlyeq z}$ (transitivity)

Before we continue with the axiomatisation, let us first introduce a couple of additional language constructs to refer to ballots and profiles within the logical language. Consider a profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n) \in \mathcal{L}(X)^n$. For a given agent $i \in N$, let $x_1, x_2, \ldots, x_m$ be a permutation of the elements of X such that $x_1 \succcurlyeq_i x_2 \succcurlyeq_i \cdots \succcurlyeq_i x_m$. Then $\mathit{ballot}_i(w)$ is defined as the following formula:

$$\mathit{ballot}_i(w) \quad := \quad p^i_{x_1 \succcurlyeq x_2} \wedge p^i_{x_2 \succcurlyeq x_3} \wedge \cdots \wedge p^i_{x_{m-1} \succcurlyeq x_m}$$

Thus, $\mathit{ballot}_i(w)$ is true at world w' *iff* w and w' agree as far as the ballot of agent i is concerned. Note that $\mathit{ballot}_i(w)$ is a purely syntactic representation of a semantic notion (namely, agent i's preference order $\succcurlyeq_i$). Similarly, we define $\mathit{profile}(w)$ as the following formula:

$$\mathit{profile}(w) \quad := \quad \mathit{ballot}_1(w) \wedge \mathit{ballot}_2(w) \wedge \cdots \wedge \mathit{ballot}_n(w)$$

Hence, the formula $\mathit{profile}(w)$ is true at world w, and only there. This shows that *nominals*, i.e., formulas uniquely identifying worlds [5], are definable within this logic at no extra cost. Finally, for any two alternatives $x, y \in X$, we define $\mathit{profile}(w)(x, y)$ as the formula fixing the relative ordering of x and y for all agents as in profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$:

$$\mathit{profile}(w)(x, y) \quad := \quad \bigwedge_{i \in N} \{p^i_{x \succcurlyeq y} \mid x \succcurlyeq_i y\} \wedge \bigwedge_{i \in N} \{p^i_{y \succcurlyeq x} \mid y \succcurlyeq_i x\}$$

This formula will be used to express the fact that two profiles 'agree' on the preferences concerning the alternatives x and y. We now state the remaining axioms:

(4) all propositional tautologies

(5) $\Box_i(\varphi \rightarrow \psi) \rightarrow (\Box_i \varphi \rightarrow \Box_i \psi)$ (K(i))

(6) $\Box_i \varphi \rightarrow \varphi$ (T(i))

(7) $\varphi \rightarrow \Box_i \Diamond_i \varphi$ (B(i))

(8) $\Diamond_i \Box_j \varphi \leftrightarrow \Box_j \Diamond_i \varphi$ (confluence)

(9) $\Box_{C_1} \Box_{C_2} \varphi \leftrightarrow \Box_{C_1 \cup C_2} \varphi$ (union)

(10) $\Box_\emptyset \varphi \leftrightarrow \varphi$ (empty coalition)

(11) $(\Diamond_i p \wedge \Diamond_i \neg p) \rightarrow (\Box_j p \vee \Box_j \neg p)$, where $i \neq j$ (exclusive)

(12) $\Diamond_i \mathit{ballot}_i(w)$ (ballot)

(13) $\Diamond_{C_1} \delta_1 \wedge \Diamond_{C_2} \delta_2 \rightarrow \Diamond_{C_1 \cup C_2}(\delta_1 \wedge \delta_2)$ (cooperation)

(14) $\bigvee_{x \in X}(x \wedge \bigwedge_{y \in X \setminus \{x\}} \neg y)$ (resolute)

(15) $(\mathit{profile}(w) \wedge \varphi) \rightarrow \Box_N(\mathit{profile}(w) \rightarrow \varphi)$ (functional)

Here φ and ψ range over arbitrary formulas, x over atomic propositions in X, i and j over agents, C_1 and C_2 over coalitions, and w over profiles. In axiom (11), p is ranging only over atomic propositions in the set $\mathit{Pref}[N, X]$, and in axiom (13) δ_1 and δ_2 do not contain any common atoms.

Axioms (4)–(8) describe well-known properties of normal modal logics [5]. Axiom (9) describes the capability of a coalition to enforce a certain formula in terms of the capabilities of its sub-coalitions. Axiom (10) states that the empty coalition cannot enforce any formula. Axiom (11) enforces a division among the atomic propositions of the shape $p^i_{x \succcurlyeq y}$: if an atom is controlled by an agent i, then other agents cannot change its value. Axiom (12) ensures that every agent can express every possible preference. Due to axiom (13), if two formulas δ_1 and δ_2 do not contain a common atom and two coalitions C_1 and C_2 can each enforce one of the formulas, then the joint coalition can enforce the conjunction $\delta_1 \wedge \delta_1$. Axiom (14) expresses that any outcome associated with a profile must be a single winning alternative. Thus,

this axioms encodes the resoluteness of the SCF in question. Finally, axiom (15) ensures that every profile is associated with a single outcome, i.e., it encodes the fact that the SCF being modelled must be a function.

The inference rules of the logic are *modus ponens* and *necessitation* w.r.t. all modalities of the form $\Box_i$ [5]:

- (MP) from $\varphi \to \psi$ and φ, infer ψ
- (Nec_i) from $\vdash \varphi$, infer $\vdash \Box_i \varphi$

We write $\vdash \varphi$ to mean that a well-formed formula φ in the language parametrised by N and X is a *theorem* of the logic $L[N, X]$, in the sense that it can be derived from axioms (1)–(15), together with the above inference rules.[2] The theorems coincide with the valid formulas:

THEOREM 2. *The logic $L[N, X]$ is sound and complete w.r.t. the class of models of SCF's.*

PROOF (SKETCH). Since our logic is a fragment of $\Lambda^{\text{scf}}[N, X]$, the soundness result due to Troquard et al. [26] applies directly. The same is not true for completeness. However, as we shall outline next, *mutatis mutandis*, the proof of Troquard et al. [26] for the richer logic can be adapted to our fragment.

First, we show the existence of an isomorphism between the models of Definition 3 and particular Kripke models. The latter structures are tuples $\langle W, (R_C)_{C \subseteq N} \rangle$ where W is the set of profiles and $R_C \subseteq W \times W$ are relations defined as

$$w R_C w' \quad \text{iff} \quad w \restriction N \setminus C = w' \restriction N \setminus C,$$

where $w \restriction N \setminus C$ is the profile w restricted to only the individuals outside of C. Intuitively, $w R_C w'$ holds if all the agents outside of C express the same preferences in w and w'.

Second, given a consistent formula φ, we build a maximally consistent set Γ_φ containing it using the usual Lindenbaum construction. Define $\mathit{Cluster}(\Gamma_\varphi)$ to be the set of maximally consistent sets that describe the same SCF:

$$\begin{aligned} \mathit{Cluster}(\Gamma_\varphi) \quad := \quad & \{\Gamma \mid \forall w \in \mathcal{L}(X)^n, \forall x \in X : \\ & \Diamond_N(\mathit{profile}(w) \wedge x) \in \Gamma \quad \textit{iff} \\ & \Diamond_N(\mathit{profile}(w) \wedge x) \in \Gamma_\varphi\} \end{aligned}$$

Finally, we consider the submodel of the canonical model generated by $\mathit{Cluster}(\Gamma_\varphi)$. Let us call this submodel M_φ. It remains to check that:

- the Truth Lemma holds for M_φ,
- there is a bijection between profiles and states of M_φ,
- M_φ is one of the aforementioned particular Kripke models corresponding to the models of our logic.

The first item is shown in the customary way, while the other items are proven exploiting the axioms. □

2.5 Representing Families of SCF's

To complete the outline of the expressive capabilities of $L[N, X]$, we illustrate how it is possible to encode a SCF as a formula. Given a SCF F, its representation will be:

$$\rho^F \quad = \quad \bigwedge \{\mathit{profile}(w) \to x \mid w \in \mathcal{L}(X)^n \text{ and } F(w) = x\}$$

That is, ρ^F is simply the conjunction, over all profiles w, of implications between a formula describing w and a formula identifying the winning alternative for profile w under F. In other words, we need to have the full graph of the function, that is, the full set of input-output pairs, to be able to encode F in the language. This is indeed possible, because, strictly speaking, ρ^F represents the function only for a fixed number of alternatives and a fixed number of agents. Moreover, since we are able to encode any set of input-output pairs, we can represent any SCF in the language.

Unfortunately, for the very same reason, ρ^F cannot be taken as a proper representative of a SCF, because it only tells us what the output of the function is in a very limited case: when the alternatives are exactly those in X and when the agents are exactly those in N. In practice, however, we are interested in *families* of SCF's. If, say, F is the Borda rule and X and N both have cardinality 3, then ρ^F will only express the workings of the Borda rule for 3 alternatives and 3 agents. A full representation of the Borda rule (which formally is a family of SCF's in the sense of Definition 1), however, should contain the information necessary to compute the output from *any* given profile. It should be a conjunction of all the formulas ρ^F for all possible choices of X and N. But even assuming that we had all such sets of pairs, there are countably many ρ's of this kind, and our logical language does not contain countable conjunctions. Given that the language is not powerful enough to encode an algorithmic specification, there is no hope that our logic, or a similar logic, will do better than using ρ^F in representing SCF's. Indeed, this restriction to specific sets of alternatives and agents is a recognised limitation of most existing logic-based approaches to modelling frameworks of social choice [9].

Interestingly, however, this problem affects the representations of the properties of SCF's only partially. Since most of the properties do not directly refer to the specific number of alternatives and agents, we can formulate the properties leaving X and N as parameters. The same can be done when proving the relative dependencies between properties. This means that, to prove that property P_1 entails P_2, we prove that, for fixed choices of X and N, there is a proof in the logic from the formula encoding P_1 to the formula encoding P_2 (both these formulas are instantiated to X and N themselves). This is the approach we shall take here.

3. ARROW'S THEOREM

First published in 1951, Arrow's Theorem is widely regarded as *the* seminal contributions to social choice theory [3]. The original theorem concerns *social welfare functions*, i.e., functions mapping profiles of (weak) preference orders (permitting indifference between alternatives) to single collective preference orders. The version we present here is adapted for preference orders that do not permit indifferences between alternatives and to SCF's (which return a single winning alternative rather than a collective order). We refer to Taylor [24] for an extensive discussion of this variant of the theorem. From a mathematical point of view, both variants are essentially equivalent and can be proven using the same methods [9, 24]. We focus on linear orders (not permitting indifferences), because most standard voting rules impose this requirement on ballots [24]. We furthermore focus on SCF's, because the problem of choosing a single best alternative is more pervasive in applications than that of choosing a full ranking over alternatives.

In this section, we first recall Arrow's Theorem and the

[2]The $\vdash \varphi$ appearing in the second rule indicates that the rule can only be applied to theorems.

properties it involves. We then express these properties in our logic and prove a number of correspondence results establishing the correctness of this encoding. Finally, we demonstrate how to construct a full proof of Arrow's Theorem within the axiomatic system we have seen to be complete for our logic (cf. Theorem 2).

3.1 Properties of SCF's

Arrow showed that, rather surprisingly, any aggregator for three or more alternatives that is Pareto efficient and that satisfies the property of independence of irrelevant alternatives must be dictatorial. We start by recalling the theorem's main ingredients: independence of irrelevant alternatives, Pareto efficiency, and dictatorships.

Denote with $N^w_{x \succcurlyeq y} := \{i \in N \mid x \succcurlyeq_i y\}$ the set of agents that prefer x over y in profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$; and denote with top^w_i that alternative $x \in X$ for which $x \succcurlyeq_i y$ for all other alternatives $y \in X$ in profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$.

Independence of irrelevant alternatives, henceforth IIA, expresses the intuitively desirable property of a SCF F that, for every two profiles and for every two alternatives x and y, if the outcome of F in the first profile is x and the two profiles are identical as far as the preferences of the agents over x and y are concerned, then the outcome of F in the second profile should not be y. This is formalised as follows:

Definition 5. *A SCF F satisfies IIA if, for every pair of profiles $w, w' \in \mathcal{L}(X)^n$ and every pair of distinct alternatives $x, y \in X$ with $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$, $F(w) = x$ implies $F(w') \neq y$.*

Pareto efficiency expresses the desideratum that, if all the agents rank an alternative x above alternative y, then y certainly should not win. This is formalised as follows:

Definition 6. *A SCF F is Pareto efficient if, for every profile $w \in \mathcal{L}(X)^n$ and every pair of distinct alternatives $x, y \in X$ with $N^w_{x \succcurlyeq y} = N$, we obtain $F(w) \neq y$.*

Finally, dictatorships are SCF's for which one individual, the dictator, can enforce their top alternative as the outcome:

Definition 7. *A SCF F is a dictatorship if there exists an agent $i \in N$ (the dictator) such that, for every profile $w \in \mathcal{L}(X)^n$, we obtain $F(w) = top^w_i$.*

We are now ready to state Arrow's Theorem itself:

Theorem 3 (Arrow). *Any SCF for $\geqslant 3$ alternatives that satisfies IIA and the Pareto condition is a dictatorship.*

3.2 Correspondences

Given that the language of $L[N, X]$ is parametrised by the set of individuals and the set of alternatives, Arrow's Theorem itself cannot be stated or proven in this logic. To prove it, we have to make a meta-argument, using a proof schema, to show that, for each choice of N and X, it is possible to prove a version of Arrow's Theorem in the logic instantiated to those two parameters. The same *proviso* holds for the properties of SCF's featuring in the theorem: rather than being formulas in the logic, they are schemas of the representations of the properties in the logic.

We exploit freely the finiteness of the language. This means that we will use big conjunctions and disjunctions to quantify over individuals or alternatives, and we will assume that we have a finite number of formulas of the form $\mathit{profile}(w)$ and $\mathit{profile}(w)(x, y)$. Here is how the aforementioned properties are coded in the logical language:

$$\begin{aligned} \mathit{IIA} &:= \bigwedge_{w \in \mathcal{L}(X)^n} \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} \\ &\quad [\Diamond_N(\mathit{profile}(w) \wedge x) \rightarrow (\mathit{profile}(w)(x, y) \rightarrow \neg y)] \\ P &:= \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} \left[\left(\bigwedge_{i \in N} p^i_{x \succcurlyeq y} \right) \rightarrow \neg y \right] \\ D &:= \bigvee_{i \in N} \bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} \left(p^i_{x \succcurlyeq y} \rightarrow \neg y \right) \end{aligned}$$

Notice that, in the presence of axiom (14), encoding resoluteness, the disjunction in the formula D is actually an exclusive one, i.e., not only must there be some dictator, but there must be exactly one dictator.[3]

Before proceeding to the proof of Arrow's Theorem in our logic, we must show that the above formulas indeed correspond to the standard definitions of properties of SCF's introduced earlier (in Definitions 5–7).

Lemma 4. *For every SCF F, $M_F \vDash \mathit{IIA}$ if and only if F satisfies independence of irrelevant alternatives.*

Proof. From right to left, assume F satisfies IIA. We want to prove every conjunct of the formula *IIA*. So take any generic world w' such that $M_F, w' \vDash \Diamond_N(\mathit{profile}(w) \wedge x)$. We want to show that $M_F, w' \vDash (\mathit{profile}(w)(x, y) \rightarrow \neg y)$. So suppose $M_F, w' \vDash \mathit{profile}(w)(x, y)$, which entails $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$. By the semantics of $\Diamond_N$, there is a world w'' such that $M_F, w'' \vDash \mathit{profile}(w) \wedge x$, which entails $N^w_{x \succcurlyeq y} = N^{w''}_{x \succcurlyeq y}$. Thus, also $N^{w'}_{x \succcurlyeq y} = N^{w''}_{x \succcurlyeq y}$. From $M_F, w'' \vDash x$ we can infer $F(w'') = x$. Now we can apply IIA to w'' and w' and obtain $F(w') = x$ and thus $F(w') \neq y$. Again by the semantics, this is tantamount to $M_F, w' \vDash \neg y$.

From left to right, assume $M_F \vDash \mathit{IIA}$. Take any two profiles w, w' and two alternatives x, y with $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$. Now assume $F(w) = x$. We thus have $M_F, w \vDash \mathit{profile}(w) \wedge x$ and, by the semantics of $\Diamond_N$, also $M_F, w' \vDash \Diamond_N(\mathit{profile}(w) \wedge x)$. Using modus ponens and formula *IIA*, we get $M_F, w' \vDash (\mathit{profile}(w)(x, y) \rightarrow \neg y)$. But we assumed $N^w_{x \succcurlyeq y} = N^{w'}_{x \succcurlyeq y}$, hence $M_F, w' \vDash \mathit{profile}(w)(x, y)$ and thus $M_F, w' \vDash \neg y$, which by the semantics entails $F(w') \neq y$. □

Lemma 5. *For every SCF F, $M_F \vDash P$ if and only if F is Pareto efficient.*

Proof. Straightforward. □

Lemma 6. *For every SCF F, $M_F \vDash D$ if and only if F is a dictatorship.*

Proof. From right to left, suppose F is a dictatorship, and call the dictator i. Take any world $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$. We want to show that the disjunct corresponding to i is true at w. Thus, for any two distinct alternative x, y we want to show that $p^i_{x \succcurlyeq y} \rightarrow \neg y$ is true at w. First, if $x \succcurlyeq_i y$, then $top^w_i \neq y$ and thus, due to F being a dictatorship of i, we have $F(w) \neq y$. By the semantics, this entails $M_F, w \vDash \neg y$

[3]The reader can prove this using the Universal Domain Lemma from the next section, formula D, and axiom (14). The gist of the proof is: take a profile were two dictators disagree and show that it leads to a contradiction.

and thus $M_F, w \vDash p^i_{x \succcurlyeq y} \rightarrow \neg y$. Second, if $x \not\succcurlyeq_i y$, then $M_F, w \nvDash p^i_{x \succcurlyeq y}$, and the implication holds vacuously.

From left to right, suppose $M_F \vDash D$. Then one of the disjuncts must be valid, say for agent i. Suppose $x = top^w_i$ under profile w. Then $M_F, w \vDash \bigwedge_{y \in X \setminus \{x\}} p^i_{x \succcurlyeq y}$. Since (the disjunct referring to i in) the condition D is true at w, we obtain $M_F, w \vDash \bigwedge_{y \in X \setminus \{x\}} \neg y$. By resoluteness, this entails $M_F, w \vDash x$ and thus $F(w) = x$. □

3.3 Coding the Proof of Arrow's Theorem

We now proceed to code a proof of Arrow's Theorem in our logic. We will use a familiar technique, based on the concept of *decisive coalitions*, to guide our search for a proof [9, 22]. What is novel about our approach is that we show that this technique can be fully embedded into a formal derivation of the axiomatic system for $L[N, X]$ presented earlier.

We first need to introduce some additional concepts. We will call a coalition $C \subseteq N$ *decisive* over a pair of alternatives $(x, y) \in X^2$ if the members of C preferring x to y is a sufficient condition for preventing y from winning. We use the following formula to encode decisiveness of C over (x, y):

$$C\,dec(x, y) \quad := \quad \left(\bigwedge_{i \in C} p^i_{x \succcurlyeq y} \right) \rightarrow \neg y$$

If C is decisive on every pair, we will simply write $C\,dec$. Along the same lines, we define a *weakly decisive* coalition C for (x, y) as a coalition that can bar y from winning if *exactly* the agents in C prefer x to y. We use the following formula to encode weak decisiveness of C over (x, y):

$$C\,wdec(x, y) \quad := \quad \left(\bigwedge_{i \in C} p^i_{x \succcurlyeq y} \wedge \bigwedge_{i \notin C} p^i_{y \succcurlyeq x} \right) \rightarrow \neg y$$

The reader can easily check that these syntactic notions match the semantic ones; for example, in the case of decisiveness we have that $C\,dec(x, y)$ is true in the model M_F *iff* the coalition C is decisive over that pair of alternatives for the corresponding SCF F. Finally, observe that D is equivalent to $\bigvee_{i \in N} \{i\}dec$, i.e., a SCF is dictatorial *iff* there is an individual that is decisive on every pair. Likewise, P is equivalent to $N\,dec$, i.e., to saying that the grand coalition N is decisive on every pair.

Our next lemma states that all the possible profiles are also possible worlds in the semantics. This fact, which is implicit in our earlier exposition of Arrow's Theorem, is called the *universal domain* condition in Arrow's original work [3].

LEMMA 7 (UNIVERSAL DOMAIN). *For every possible profile $w \in \mathcal{L}(X)^n$, we have $\vdash \Diamond_N profile(w)$.*

PROOF. Take any profile w. Then $ballot_1(w)$ encodes the preferences of the first agent. We have, by axiom (12), that $\Diamond_1 ballot_1(w)$, and similarly for the second agent we get $\Diamond_2 ballot_2(w)$. Because $ballot_1(w)$ and $ballot_2(w)$ contain different atoms (the former only atoms with superscript 1, the latter only atoms with superscript 2), we can apply axiom (13) and obtain $\Diamond_{\{1,2\}}(ballot_1(w) \wedge ballot_2(w))$. We can repeat this reasoning for all the finitely many agents in N to prove $\Diamond_N profile(w)$. □

We now turn to the two main lemmas in the proof. We offer an outline on the main steps of the proof, from which a complete formal derivation can be recovered. Semantically speaking, the first of these two lemmas shows that, under certain conditions, a coalition being weakly decisive over a specific pair of alternatives implies that the same coalition is (not only weakly) decisive over all pairs.

LEMMA 8. *Consider a language parametrised by X such that $|X| \geqslant 3$. Then for any coalition $C \subseteq N$ and any two distinct alternatives $x, y \in X$, we have that:*

$$\vdash P \wedge IIA \wedge C\,wdec(x, y) \rightarrow C\,dec$$

PROOF. Suppose x, y, x' and y' are distinct alternatives (the other cases are analogous). In order to prove $C\,dec$ we need to prove each of the conjuncts in the following formula:

$$\bigwedge_{x \in X} \bigwedge_{y \in X \setminus \{x\}} \left[\left(\bigwedge_{i \in C} p^i_{x \succcurlyeq y} \right) \rightarrow \neg y \right]$$

Denote by C' one of the possible subsets of $N \setminus C$ preferring x' over y'. Now consider the following derivation:

(1) $(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \rightarrow [(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge \bigvee_{C' \subseteq N \setminus C}((\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'}))]$
By finiteness of agents and alternatives and the theorems $p^i_{x' \succcurlyeq y'} \vee p^i_{y' \succcurlyeq x'}$ for all $i \in N$ we can, rearranging conjunctions and disjunctions, prove the second line of the formula; the implication follows.

(2) $(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \rightarrow \bigvee_{C' \subseteq N \setminus C} [(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'})]$
by distributivity from (1)

(3) This part of the proof contains the derivation of the following formula, for every $C' \subseteq N \setminus C$:
$P \wedge IIA \wedge C\,wdec(x, y) \rightarrow [(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'}) \rightarrow \neg y']$
We will present the derivation for any such C' below.

(4) $P \wedge IIA \wedge C\,wdec(x, y) \rightarrow \bigvee_{C' \subseteq N \setminus C} [(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'}) \rightarrow \neg y']$
by propositional reasoning from all the instances of (3)

(5) $P \wedge IIA \wedge C\,wdec(x, y) \rightarrow [(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \rightarrow \neg y']$
by implication concatenation from (2) and (4)

We still need to show (all the finitely many instances of) step (3). We prove each of them in the following way. Consider a specific profile $w = (\succcurlyeq_1, \ldots, \succcurlyeq_n)$ for which we can rearrange the conjuncts in the formula $profile(w)$ as follows:

$$\begin{aligned} profile(w) \quad = \quad & (\bigwedge_{i \in C} p^i_{x \succcurlyeq y}) \wedge (\bigwedge_{i \in N} (p^i_{x' \succcurlyeq x} \wedge p^i_{y \succcurlyeq y'})) \wedge \\ & (\bigwedge_{i \in C \cup C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C} p^i_{y \succcurlyeq x}) \wedge (\bigwedge_{i \notin C \cup C'} p^i_{y' \succcurlyeq x'}) \wedge \alpha \end{aligned}$$

Here α is the formula expressing the fact that all the other alternatives (if any) are ranked by all agents below x, y, x', y'. We are now ready to present a derivation for a specific C':

(a) For any $z \in X \setminus \{x, y, x', y'\}$:
$P \wedge profile(w) \rightarrow \neg x \wedge \neg y' \wedge \neg z$
from formula P, the second part of $profile(w)$, and α

(b) $C\,wdec(x,y) \wedge profile(w) \rightarrow \neg y$
by definition of $C\,wdec(x,y)$

(c) $P \wedge C\,wdec(x,y) \rightarrow (profile(w) \rightarrow x')$
by axiom (14), encoding resoluteness, with (a) and (b)

(d) $\Diamond_N profile(w)$
by the Universal Domain Lemma

(e) $P \wedge C\,wdec(x,y) \rightarrow \Diamond_N(profile(w) \wedge x')$
by standard modal reasoning from (c) and (d)

(f) $P \wedge IIA \wedge C\,wdec(x,y) \rightarrow \Diamond_N(profile(w) \wedge x')$
by propositional reasoning from (e)

(g) $P \wedge IIA \wedge C\,wdec(x,y) \rightarrow [(profile(w)(x',y') \rightarrow \neg y')]$
from (f) and formula IIA w.r.t. x' and y'

But $profile(w)(x',y')$ consists of the following conjuncts:

$$(\bigwedge_{i \in C} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \in C'} p^i_{x' \succcurlyeq y'}) \wedge (\bigwedge_{i \notin C' \cup C} p^i_{y' \succcurlyeq x'})$$

Hence, we may infer that this latter formula entails $\neg y'$. This shows step (3) and concludes the proof. □

The next lemma establishes a syntactic counterpart of what is known as the *Contraction Lemma* in the literature [22]. It says that, under certain conditions, for any way of splitting a decisive coalition of two or more agents into two sub-coalitions, one of those sub-coalitions must also be decisive.

LEMMA 9 (CONTRACTION LEMMA). *Consider a language parametrised by X such that $|X| \geqslant 3$. Then for any coalition $C \subseteq N$ with and any two coalitions C_1 and C_2 that form a partition of C, we have that:*

$$\vdash P \wedge IIA \wedge C\,dec \rightarrow (C_1\,dec \vee C_2\,dec)$$

PROOF. Consider C, C_1 and C_2 as in the statement of the lemma (i.e., $C = C_1 \cup C_2$ and $C_1 \cap C_2 = \emptyset$) and let x, y, z be three distinct alternatives. Now consider any profile w for which $profile(w)$ has the following form:

$$\begin{aligned} profile(w) \;=\; & (\bigwedge_{i \notin C_2} p^i_{x \succcurlyeq y}) \wedge (\bigwedge_{i \in C_1} p^i_{x \succcurlyeq z}) \wedge (\bigwedge_{i \in C_1 \cup C_2} p^i_{y \succcurlyeq z}) \\ & (\bigwedge_{i \in C_2} p^i_{y \succcurlyeq x}) \wedge (\bigwedge_{i \notin C_1} p^i_{z \succcurlyeq x}) \wedge (\bigwedge_{i \notin C_1 \cup C_2} p^i_{z \succcurlyeq y}) \wedge \alpha \end{aligned}$$

Here α encodes the fact that all other alternatives (if any) are ranked by all agents below x, y, z.

Now assume P, IIA, and $C\,dec$. We want to prove $(C_1\,dec \vee C_2\,dec)$. By $C\,dec$ and propositional reasoning, we have that $profile(w) \rightarrow \neg z$ is the case. As all other alternatives are ruled out by P and α, axiom (14), encoding resoluteness, enforces that x or y must be the outcome. Hence, the formula $(profile(w) \rightarrow x) \vee (profile(w) \rightarrow y)$ must be the case. As an aside, we note that we know (again from resoluteness) that this disjunction must be exclusive.

By the Universal Domain Lemma, we have $\Diamond_N profile(w)$, and thus, using standard modal reasoning, we obtain $\Diamond_N(profile(w) \wedge x) \vee \Diamond_N(profile(w) \wedge y)$. Now propositional reasoning together with IIA, first w.r.t. the pair (x,z) and then w.r.t. the pair (y,x), allows us to derive the formula $(profile(w)(x,z) \rightarrow \neg z) \vee (profile(w)(y,x) \rightarrow \neg x)$.

Recall that in $profile(w)$ the agents in C_1 are the only ones supporting x over z. Hence, $(profile(w)(x,z) \rightarrow \neg z)$ means that C_1 is weakly decisive for the pair (x,z). Likewise, the agents in C_2 are the only ones supporting y over x; thus $(profile(w)(y,x) \rightarrow \neg x)$ means that C_2 is weakly decisive for the pair (y,x). In this fashion we obtain that the formula $C_1\,wdec(x,z) \vee C_2\,wdec(y,x)$ must be the case.

We can now use Lemma 8 to derive $C_1\,dec \vee C_2\,dec$. We have thus shown that $P \wedge IIA \wedge C\,dec \rightarrow (C_1\,dec \vee C_2\,dec)$ must be a theorem of the logic. Note that the disjunction is still exclusive. □

We can now state and prove our main result, a syntactic counterpart of Arrow's Theorem:

THEOREM 10. *Consider a language parametrised by X such that $|X| \geqslant 3$. Then we have:*

$$\vdash P \wedge IIA \rightarrow D$$

PROOF. As mentioned earlier, P is equivalent to $N\,dec$. Exploiting the formula IIA, we can apply the Contraction Lemma and prove that one of two disjoint subsets of N is decisive. Repeating the process finitely many times (we have finitely many agents), we can show that one of the singletons that form N is decisive. But this is tantamount to saying that there exist a decisive agent, i.e., a dictator, so the formula D can be derived as claimed. □

Note that throughout the proof we have made implicit use of the condition $|X| \geqslant 3$ when assuming the availability of three distinct alternatives (in fact, in the proof of Lemma 8 we have only gone through the most interesting case, requiring at least four alternatives).

As we already mentioned, the proof provided here is not, strictly speaking, a full syntactic proof of Arrow's Theorem *within* the logic, because the language is parametric in the set of agents N and the set of alternatives X. Nevertheless, apart from the *proviso* on the number of alternatives stated in Theorem 10, our proof is independent of the choice of N and X; that is to say, this proof can be used as a *template* to prove the appropriate instance of Arrow's Theorem in *any* logic $L[N,X]$ for N and X such that $|X| \geqslant 3$.

Due to Theorem 2 establishing completeness of the logic and Lemmas 4–6 establishing the correctness of our representation of the Arrovian conditions within the logic, Theorem 10 is equivalent to the usual, semantic, rendering of Arrow's Theorem for SCF's stated as Theorem 3. Thus, our purely syntactic proof constitutes an independent proof of the theorem. This shows that the logic $L[N,X]$ is a useful tool for reasoning about nontrivial concepts in social choice.

4. RELATED WORK

The idea of using logic, and formal methods more generally, to subject social procedures, such as voting rules, to the same kind of formal analysis routinely applied to algorithms and software systems can be traced back to, at least, the work of Parikh [17, 18]. The two main arguments motivating this kind of enterprise are obvious and well known: formal analysis will deepen our understanding of social procedures; and formal analysis can increase our confidence in the correctness of social procedures. Pauly [19] has suggested a third argument that is specific to the use of logic in social choice theory: the expressive power of a logical language required to express a choice-theoretic property (such

a IIA) is a relevant criterion in judging the interestingness of a characterisation result making use of such a property. A fourth argument fueling this line of research is that it has the potential to uncover entirely new characterisation and impossibility results [7, 10, 23]—results that are of independent interest to economists [8].

Successful applications of logic and automated reasoning to social choice theory have included the automated verification of the correctness of practical algorithms for implementing voting rules [4] and the automated search for new impossibility theorems in the domain of ranking sets of objects [10]. However, most work to date has focussed on the Arrovian framework of preference aggregation and the challenges of representing Arrow's Theorem in a variety of logical frameworks [1, 12], of verifying the correctness of existing proofs for the theorem [16, 28], and of finding new such proofs [23]. Indeed, Arrow's Theorem is arguably the best yardstick against which to measure new formal methods for reasoning about problems of social choice. Our own work also falls into this category. The work of Lange et al. [13] on the use of automated reasoning in different areas of economic theory, such as auctions and cooperative games, demonstrates that the basic concepts and techniques developed for the seemingly narrow domain of Arrovian preference aggregation can have a ripple-on effect on the use of formal methods in economics more widely.

Regarding Arrow's Theorem, starting at the top as far as the expressive power of the logical systems employed is concerned, Nipkow [16] and Wiedijk [28] have shown how to verify existing proofs for the theorem in higher-order logic proof assistants. Grandi and Endriss [12] have shown that classical first-order logic is sufficiently expressive to model all aspects of Arrow's Theorem, with the sole exception being the requirement that the set of agents be finite (the theorem is not valid for infinite electorates; cf. the use of induction in the proof of Theorem 10). In particular, modelling IIA does not require second-order quantification. At the most extreme end of the spectrum, Tang and Lin [23] have shown that the theorem can even be embedded into classical propositional logic, albeit only for a fixed set of agents and a fixed set of alternatives. This embedding itself ceases to be useful for deepening our understanding of social choice (as it involves thousands of clauses, even for the simplest case of $|N| = 2$ and $|X| = 3$). Instead, the great significance of the work of Tang and Lin derives from the fact that they have been able to provide a fully automated proof of the theorem based on this embedding. The work of Ågotnes et al. [1], like our own work, is orthogonal to these other contributions, in that they design a new tailor-made logic for social choice theory, rather than encoding those concepts into already existing logics. Note that Troquard et al. [26], the originators of the logic $\Lambda^{\mathrm{scf}}[N, X]$ we have used here, have themselves not attempted to model Arrow's Theorem.

To date, the approaches to modelling Arrow's Theorem in logical frameworks that are human-readable, namely the contributions of Ågotnes et al. [1] and of Grandi and Endriss [12], have not yet yielded a complete proof of the theorem *within* that same logical framework, although Ågotnes et al. [1] do succeed in providing a syntactic proof of a relevant lemma. The most satisfactory attempt in this respect is that of Perkov [21], who has outlined a natural deduction proof of Arrow's Theorem using the language of Ågotnes et al. [1], albeit for a calculus that currently is not known to be complete. Our work provides a complete formalisation of the theorem and its premises, in the form of a clear recipe for constructing a derivation of the theorem from the axioms of the logic, in a sound and complete calculus that is easily readable.

A recent survey on logic and social choice theory [9] has identified three critical points in existing work on logics for modelling concepts in social choice: (1) whether the approach does not require us to fix the sets of agents and alternatives upfront, (2) whether the universal domain assumption can be expressed in an elegant manner, and (3) whether the approach facilitates automation. Regarding point (1), as discussed in Section 2.5, our logic is indeed subject to the common limitation of requiring us to fix the cardinalities of N and X before even the notion of a well-formed formula can be defined, but we have also demonstrated that in practice this limitation can be overcome by working with schemas parametrised by N and X. Point (2) is convincingly taken care of by Lemma 7, the Universal Domain Lemma. Point (3), finally, is not directly addressed here, but we believe that our proof shows that automation, certainly automated verification of our proof in terms of the axiomatisation given, is clearly possible in principle. Further evidence for the claim that the automation of reasoning tasks for the modal logic of SCF's used here is feasible and promising is given by Troquard [25], who has initiated a study of algorithms for model checking for the full logic $\Lambda^{\mathrm{scf}}[N, X]$, including a prototype implementation.

5. CONCLUSION

We have shown how to obtain a syntactic proof of Arrow's Theorem within a simple modal logic for speaking about basic concepts of preference aggregation. The logic in question is a fragment of a logic introduced by Troquard et al. [26], which we have shown to be complete by adapting their original completeness proof. While prior work has been successful in applying tools from logic and automated reasoning to social choice theory, this is the first human-readable formalisation of the framework of preference aggregation that allows for a direct derivation of Arrow's Theorem.

Because of the central role of Arrow's Theorem not only in social choice theory at large, but also in the emerging literature on logics for social choice, where it has served as a yardstick for assessing the suitability of a variety of approaches to logical modelling, we believe the closure of this gap constitutes a useful step towards the longterm aim of the field. This aim is to offer tangible computer-aided support for reasoning about methods for collective decision making, be it in the context of political decision making, economic interaction, or multiagent systems.

Our results suggest two important directions for future work. First, it certainly is possible, at least in principle, to encode most of the commonly studied desiderata for voting rules in the logic considered here. To what extent this is also practically feasible, and to what extent this might allow us to verify a given voting rule's satisfaction of a given desideratum, or to what extent this might allow us to re-prove other classical results in social choice theory, such as May's Theorem on the characterisation of the simple majority rule [15], are intriguing open questions. Second, our demonstration of the usefulness of modal logics of social choice underlines the importance of further developing the reasoning machinery for such logic, including optimised implementations.

REFERENCES

[1] T. Ågotnes, W. van der Hoek, and M. Wooldridge. On the logic of preference and judgment aggregation. *Autonomous Agents and Multiagent Systems*, 22(1):4–30, 2011.

[2] A. Altman and M. Tennenholtz. Axiomatic foundations for ranking systems. *Journal of Artificial Intelligence Research*, 31:473–495, 2008.

[3] K. J. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, 2nd edition, 1963. First edition published in 1951.

[4] B. Beckert, R. Goré, C. Schürmann, T. Bormer, and J. Wang. Verifying voting schemes. *Journal of Information Security and Applications*, 19(2):115–129, 2014.

[5] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

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

[7] 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)*, 2014.

[8] S. Chatterjee and A. Sen. Automated reasoning in social choice theory: Some remarks. *Mathematics in Computer Science*, 8(1):5–10, 2014.

[9] 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.

[10] C. Geist and U. Endriss. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research*, 40:143–174, 2011.

[11] J. Gerbrandy. Logics of propositional control. In *Proc. 5th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, 2006.

[12] U. Grandi and U. Endriss. First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic*, 42(4):595–618, 2013.

[13] C. Lange, C. Rowat, and M. Kerber. The ForMaRE Project: Formal mathematical reasoning in economics. In *Intelligent Computer Mathematics*, pages 330–334. Springer-Verlag, 2013.

[14] A. Mao, A. D. Procaccia, and Y. Chen. Better human computation through principled voting. In *Proc. 27th AAAI Conference on Artificial Intelligence (AAAI-2013)*, 2013.

[15] K. O. May. A set of independent necessary and sufficient conditions for simple majority decisions. *Econometrica*, 20(4):680–684, 1952.

[16] T. Nipkow. Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 43(3):289–304, 2009.

[17] R. Parikh. The logic of games and its applications. In *Topics in the Theory of Computation*, volume 24 of *Annals of Discrete Mathematics*. North-Holland, 1985.

[18] R. Parikh. Social software. *Synthese*, 132(3):187–211, 2002.

[19] M. Pauly. On the role of language in social choice theory. *Synthese*, 163(2):227–243, 2008.

[20] D. M. Pennock, E. Horvitz, and C. L. Giles. Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *Proc. 17th National Conference on Artificial Intelligence (AAAI-2000)*, 2000.

[21] T. Perkov. Natural deduction for a fragment of modal logic of social choice. Presented at ESSLLI-2014 Workshop on Information Dynamics in Artificial Societies, 2014.

[22] A. K. Sen. Social choice theory. In K. J. Arrow and M. D. Intriligator, editors, *Handbook of Mathematical Economics*, volume 3. North-Holland, 1986.

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

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

[25] N. Troquard. Logics of social choice and perspectives on their software implementation. Presented at Dagstuhl Seminar 11101 on Reasoning about Interaction: From Game Theory to Logic and Back, 2011.

[26] N. Troquard, W. van der Hoek, and M. Wooldridge. Reasoning about social choice functions. *Journal of Philosophical Logic*, 40(4):473–498, 2011.

[27] W. van der Hoek and M. Wooldridge. On the logic of cooperation and propositional control. *Artificial Intelligence*, 164(1):81–119, 2005.

[28] F. Wiedijk. Arrow's Impossibility Theorem. *Formalized Mathematics*, 15(4):171–174, 2007.