

Voting Theory in the Lean Theorem Prover

Wesley H. Holliday¹[0000–0001–6054–9052], Chase Norman¹[0000–0001–8954–3770],
and Eric Pacuit²[0000–0002–0751–9011]

¹ University of California, Berkeley
{wesholliday,c.}@berkeley.edu

² University of Maryland, College Park
epacuit@umd.edu

Abstract. There is a long tradition of fruitful interaction between logic and social choice theory. In recent years, much of this interaction has focused on computer-aided methods such as SAT solving and interactive theorem proving. In this paper, we report on the development of a framework for formalizing voting theory in the Lean theorem prover, which we have applied to verify properties of a recently proposed voting method. While previous applications of interactive theorem proving to social choice (using Isabelle/HOL and Mizar) have focused on the verification of impossibility theorems, we aim to cover a variety of results ranging from impossibility theorems to the verification of properties of specific voting methods (e.g., Condorcet consistency, independence of clones, etc.). In order to formalize voting theoretic axioms concerning adding or removing candidates and voters, we work in a variable-election setting whose formalization makes use of dependent types in Lean.

Keywords: logic and social choice theory · voting theory · interactive theorem proving · Lean theorem prover

1 Introduction

There is a long tradition of fruitful interaction between logic and social choice theory. Both Kenneth Arrow [3, p. 154] and Amartya Sen [34, p. 108] have noted the influence of mathematical logic on their thinking about the foundations of social choice theory. Early work using logical methods in social choice theory includes Murakami’s [26] application of results about three-valued logic to the analysis of voting rules, Rubinstein’s [32] proof of the equivalence between multi-profile and single-profile approaches to social choice, and Parikh’s [29] development of a logic of games to study social procedures. There is now a rich literature developing logical systems that can formalize results in social choice theory (see, e.g., [30,1,27,35,38,16,20,12,28,22]).

In recent years, much of the research on logic and social choice has focused on computer-aided methods such as SAT solving and interactive theorem proving [19]. The first applications of interactive theorem proving used Isabelle/HOL [27] and Mizar [39] to formalize different proofs of Arrow’s Impossibility Theorem [18]. More recently, [9] and [15] used Isabelle to verify impossibility theorems from

[10] and [8], respectively. These projects demonstrate, as Nipkow [27] notes, that “social choice theory turns out to be perfectly suitable for mechanical theorem proving.” In this paper, we provide further evidence of this by developing a framework for formalizing voting theory using an interactive theorem prover.

But why formalize? One obvious benefit of such a project is the verification of the correctness of mathematical claims in voting theory. Several published claims, including Arrow’s [2] original statement of his impossibility theorem (for more than 3 candidates), Baigent’s [5] variation involving “weak IIA” (in the case of 3 candidates), and Routley’s [31] claimed generalization of Arrow’s theorem to infinite populations, were disproved by counterexamples (see [7], [11], and [6]). Second, formalization allows us to carefully track which assumptions—e.g., about voter preferences, cardinalities, choice of primitive concepts, etc.—are needed for which results, leading to generalizations and perhaps even new avenues for research. Third, formalization may eventually facilitate automated search of the corpus of proved results for use by researchers in proving new results.

For our formalization project we chose to use the Lean theorem prover [14], a framework that supports both interactive and automated theorem proving. Lean’s kernel is based on dependent type theory and implements a version of the calculus of inductive constructions [13] and Martin-Löf type theory [25]. There is an extensive and actively maintained library of mathematical results formalized in Lean (see https://leanprover-community.github.io/mathlib_docs/). In addition, Lean is the system chosen for the Formal Abstracts project initiated by Thomas Hales (<https://formalabstracts.github.io>).

Our aim was to use Lean to verify results about axioms for voting methods (e.g., Condorcet consistency, independence of clones, etc.). In order to formalize axioms concerning adding or removing candidates and voters, we work in a variable-election setting whose formalization makes use of dependent types, as explained in Section 2. In Section 3, we discuss our formal verification of results from [23] about a recently proposed voting method, Split Cycle (defined in Example 3 below), illustrating the usefulness of our framework. We conclude in Section 4 with directions for further work. All of the code for our project is available at <https://github.com/chasenorman/Formalized-Voting>.

2 Framework

In this section, we define the basic objects of voting theory: profiles, social choice correspondences, etc. We first give standard set-theoretic definitions and then their type-theoretic counterparts in Lean syntax. After defining these objects, we discuss our formalization of standard axioms used to evaluate voting procedures.

2.1 Profiles

For our set-theoretic definitions, we fix infinite sets \mathcal{V} and \mathcal{X} of voters and candidates, respectively. Given $X \subseteq \mathcal{X}$, let $\mathcal{B}(X)$ be the set of all binary relations on X . Instead of thinking of a binary relation as a set of ordered pairs, here it is more

convenient to think of a binary relation on X as a function $S : X \times X \rightarrow \{0, 1\}$. In fact, to better match our Lean formalization, we “curry” all functions with multiple arguments, transforming them into functions with single arguments that output functions. Thus, we regard a binary relation on X as a function $S : X \rightarrow \{0, 1\}^X$, where $\{0, 1\}^X$ is the set of functions from X to $\{0, 1\}$. For any $x \in X$, $S(x) : X \rightarrow \{0, 1\}$, and $S(x)(y) = 1$ means that the binary relation S holds of (x, y) . In what follows, we write ‘ xSy ’ instead of $S(x)(y) = 1$.

Definition 1. For $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$, a (V, X) -*profile* is a map $\mathbf{Q} : V \rightarrow \mathcal{B}(X)$. We write ‘ \mathbf{Q}_i ’ for the relation $\mathbf{Q}(i)$. Given a (V, X) -profile \mathbf{Q} , let $V(\mathbf{Q})$ be V and $X(\mathbf{Q})$ be X . We then define a function **Prof** that assigns to each pair (V, X) of $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$ the set $\text{Prof}(V, X)$ of all (V, X) -profiles. Finally, define $\text{PROF} = \bigcup_{V \subseteq \mathcal{V}, X \subseteq \mathcal{X}} \text{Prof}(V, X)$.

Depending on the application, one can interpret $x\mathbf{Q}_iy$ to mean either (i) that voter i strictly prefers x to y or (ii) that voter i strictly prefers x to y or is indifferent between x and y . Under interpretation (i), we use ‘ \mathbf{P} ’ for a profile; under interpretation (ii), we use ‘ \mathbf{R} ’ for a profile.³ A profile \mathbf{Q} is said to be *asymmetric* (*transitive*, etc.) if for every $i \in V$, \mathbf{Q}_i is asymmetric (*transitive*, etc.). Of course, asymmetric profiles only make sense under interpretation (i), whereas under interpretation (ii), profiles should be reflexive.

To translate Definition 1 into Lean, we first think of V and X as types, rather than sets, and then represent the function **Prof** from Definition 1 as follows:⁴

```
def Prof : Type → Type → Type :=
  λ (V X : Type), V → X → X → Prop
```

Here **Prop** is the type of propositions, which in the definition plays the role of $\{0, 1\}$ in the treatment of binary relations mentioned above. The definition states that **Prof** is a function that given two types, V and X , outputs the type $V \rightarrow X \rightarrow X \rightarrow \text{Prop}$. Because $X \rightarrow X \rightarrow \text{Prop}$ is the type of binary relations on X , an inhabitant of the type $V \rightarrow X \rightarrow X \rightarrow \text{Prop}$ can be viewed as a (V, X) -profile. Thus, we may think of **Prof** V X as the type of (V, X) -profiles.

One of the most important kinds of information to read off from a profile is whether one candidate is majority preferred to another.

Definition 2. Given a profile \mathbf{P} and $x, y \in X(\mathbf{P})$, we say that x is *majority preferred to y in \mathbf{P}* if more voters rank x above y than rank y above x .

In Lean, we formalize Definition 2 as follows:

³ Approach (ii) is more general, since it allows one to distinguish between voter i being *indifferent* between x and y , defined as $x\mathbf{R}_iy$ and $y\mathbf{R}_ix$, vs. x and y being *noncomparable* for i , defined as *neither* $x\mathbf{R}_iy$ *nor* $y\mathbf{R}_ix$. When the distinction between voter indifference and noncomparability is not needed, approach (i) can be simpler.

⁴ When writing type expressions, arrows associate to the right, so, e.g., the expression ‘ $V \rightarrow X \rightarrow X \rightarrow \text{Prop}$ ’ stands for $V \rightarrow (X \rightarrow (X \rightarrow \text{Prop}))$.

```

def majority_preferred {V X : Type} :
  Prof V X → X → X → Prop := λ P x y,
  cardinal.mk {v : V // P v x y} > cardinal.mk {v : V // P v y x}

```

Here ‘{V X : Type}’ indicates that V and X are implicit arguments⁵ to the function `majority_preferred` of type `Type`. Then `majority_preferred` takes in explicit arguments of a (V, X) -profile and two candidates and returns the proposition stating that the cardinality of the set of voters who prefer x to y is greater than the cardinality of the set of voters who prefer y to x. Here the ‘//’ notation indicates that we are identifying the subtype of voters with a certain property, and `cardinal.mk` gives us the cardinality of the subtype.

We are often concerned not only with whether one candidate is majority preferred to another but also, if so, what is the margin of majority preference.

Definition 3. Given a profile \mathbf{P} and $x, y \in X(\mathbf{P})$, the *margin of x over y in P*, denoted $\text{Margin}_{\mathbf{P}}(x, y)$, is $|\{i \in V(\mathbf{P}) \mid x\mathbf{P}_i y\}| - |\{i \in V(\mathbf{P}) \mid y\mathbf{P}_i x\}|$.

In Lean, Definition 3 becomes:

```

def margin {V X : Type} [fintype V] : Prof V X → X → X → ℤ
:= λ P x y, ↑(finset.univ.filter (λ v, P v x y)).card -
  ↑(finset.univ.filter (λ v, P v y x)).card

```

Here ‘[fintype V]’ can be understood as an implicit assumption that V is finite,⁶ which we make so that we can perform the subtraction in the definition of `margin`. The `margin` function takes in explicit arguments of a (V, X) -profile and two candidates and returns the margin of the first over the second; in particular, ‘`finset.univ.filter (λ v, P v x y)`’ is syntax for constructing the set $\{v \in V(\mathbf{P}) \mid x\mathbf{P}_v y\}$, `.card` takes the cardinality of the set (a natural number), and `↑` shifts the type from natural number to integer (so we can subtract).

As usual, we can regard the $\text{Margin}_{\mathbf{P}}$ function as an $|X(\mathbf{P})| \times |X(\mathbf{P})|$ matrix. Since $\text{Margin}_{\mathbf{P}}(x, y) = -\text{Margin}_{\mathbf{P}}(y, x)$, the matrix is skew-symmetric. Treating an integer-valued square matrix as a function from a set X to functions from X to \mathbb{Z} , the property of skew-symmetry takes in such a function and outputs the proposition stating that the skew-symmetry equation holds for all pairs:

```

def skew_symmetric {X : Type} : (X → X → ℤ) → Prop :=
  λ M, ∀ x y, M x y = - M y x.

```

Verifying that $\text{Margin}_{\mathbf{P}}$ is skew-symmetric is trivial using Lean’s automation:

```

lemma margin_skew_symmetric {V X : Type} (P : Prof V X)
  [fintype V] : skew_symmetric (margin P) :=
begin
  unfold margin,
  obviously,
end

```

⁵ See Section 3.3 of the Lean documentation on implicit arguments.

⁶ See the Lean community page on Sets and set-like objects.

The `unfold` tactic writes `margin P` in terms of the definition of `margin` above, allowing the `obviously` tactic to fill in the details of the proof of skew-symmetry.

Returning to properties of profiles, one of the most important to consider is whether a profile has a so-called Condorcet winner or even a majority winner.

Definition 4. Given a profile \mathbf{P} and $x \in X(\mathbf{P})$, x is a *Condorcet winner* in \mathbf{P} if for all $y \in X(\mathbf{P})$ with $y \neq x$, x is majority preferred to y in \mathbf{P} . We say that x is a *majority winner* in \mathbf{P} if the number of voters who rank x (and only x) in first place is greater than the number of voters who do not rank x in first place.

In Lean, Definition 4 becomes:

```
def condorcet_winner {V X : Type} (P : Prof V X) (x : X) :
  Prop := ∀ y ≠ x, majority_preferred P x y

def majority_winner {V X : Type} (P : Prof V X) (x : X) :
  Prop := cardinal.mk {v : V // ∀ y ≠ x, P v x y} > cardinal.mk
  {v : V // ∃ y ≠ x, P v y x}
```

As an example of a more involved proof than the one above showing that the margin matrix is skew-symmetric, we present a proof in Lean that a majority winner is also a Condorcet winner. For this we use several basic theorems provided by Mathlib, including one formalizing the fact that a subtype of a type has cardinality less than or equal to that of the type:⁷

```
theorem cardinal.mk_subtype_mono {α : Type u} {φ ψ : α → Prop}
  (h : ∀ x, φ x → ψ x) :
  cardinal.mk {x // φ x} ≤ cardinal.mk {x // ψ x}
```

We explain the following Lean proof in detail below:

```
lemma condorcet_of_majority_winner {V X : Type} (P : Prof V X)
  [fintype V] (x : X) :
  majority_winner P x → condorcet_winner P x :=
begin
1.  intros majority z z_ne_x,
2.  have imp1 : ∀ v, (∀ y ≠ x, P v x y) → P v x z := by finish,
3.  refine lt_of_lt_of_le _ (cardinal.mk_subtype_mono imp1),
4.  have imp2 : ∀ v, P v z x → (∃ y ≠ x, P v y x) := by finish,
5.  apply lt_of_le_of_lt (cardinal.mk_subtype_mono imp2),
6.  exact majority,
end
```

Since the logical form of what we want to prove, `majority_winner P x → condorcet_winner P x`, is an implication, we use `intros` on line 1 to introduce a name `majority` for a proof of `majority_winner P x`. Then since the consequent, `condorcet_winner P x`, is a universal claim, $\forall y \neq x, \text{majority_preferred}$

⁷ We have changed variable names and replaced ‘#’ with ‘`cardinal.mk`’.

$P \ x \ y$, we introduce a name z for an arbitrary candidate and a name z_ne_x for a proof of $z \neq x$. Our goal is now to prove `majority_preferred` $P \ x \ z$.

The first key move on line 2 is to prove that everyone who ranks x first ranks x above z , which Lean does automatically using the `finish` tactic. Since `imp1` is a proof of proposition of the form $(\forall v, \varphi \ v \rightarrow \psi \ v)$, we can apply the Mathlib theorem `cardinal.mk_subtype_mono` to get a proof `cardinal.mk_subtype_mono imp1` that the number of voters who rank x first is less than or equal to the number of voters who rank x above z .

On line 3, we use a Mathlib theorem, `lt_of_lt_of_le`, which states that $n < m \rightarrow m \leq k \rightarrow n < k$ (recall that implication associates to the right). Take n to be the number of voters who rank z above x , m to be the number who rank x first, and k to be the number who rank x above z . Thus, our goal is to prove $n < k$, and above we proved $m \leq k$. Now $m \leq k$ is not the antecedent of $n < m \rightarrow m \leq k \rightarrow n < k$, but Lean’s `refine` tactic allows us to insert a placeholder `_` for the antecedent, so our goal then becomes proving $n < m$.

To prove $n < m$, the key move on line 4 is to prove that everyone who ranks z above x does not rank x first, which Lean does automatically using the `finish` tactic. Then we can apply `cardinal.mk_subtype_mono` to obtain a proof `cardinal.mk_subtype_mono imp2` that the number n of voters who rank z above x is less than or equal to the number—call it m' —of voters who do not rank x first. Thus, we have a proof of $n \leq m'$, so we can apply the implication $n \leq m' \rightarrow m' < m \rightarrow n < m$ provided by the Mathlib theorem `lt_of_le_of_lt` to obtain a proof of $m' < m \rightarrow n < m$. Then since `majority` is exactly a proof of the antecedent of $m' < m \rightarrow n < m$, we obtain a proof of our goal $n < m$.

2.2 Functions on profiles

Next we define two kinds of functions that take profiles as inputs. The first, called a *social choice correspondence* (SCC), assign to a given profile a set of candidates, who are considered tied for winning the election. It is common to consider “domain restrictions” on the set of profiles for which the SCC is defined. Thus, one may define an SCC as a function F on some set \mathcal{D} of profiles such that for all $\mathbf{Q} \in \mathcal{D}$, we have $\emptyset \neq F(\mathbf{Q}) \subseteq X(\mathbf{Q})$. However, for our formalization purposes, it is more convenient to use the following equivalent approach.

Definition 5. For $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$, a *social choice correspondence for* (V, X) , or (V, X) -SCC, is a function $F : \text{Prof}(V, X) \rightarrow \wp(X)$. We abuse terminology and call the set $\{\mathbf{Q} \in \text{Prof}(V, X) \mid F(\mathbf{Q}) \neq \emptyset\}$ the *domain* of F . We say that F satisfies *universal domain* if its domain is $\text{Prof}(V, X)$.

Let `SCC` be a function that assigns to each pair (V, X) of $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$ the set of all (V, X) -SCCs.

We represent the function `SCC` in Lean as follows, where `set X` is the type of subsets of X :⁸

⁸ When writing type expressions, function application binds more strongly than arrow, so ‘`Prof V X → set X`’ stands for $(\text{Prof } V \ X) \rightarrow \text{set } X$.

```
def SCC := λ (V X : Type), Prof V X → set X
```

The definition states that `SCC` is a function that given two types, `V` and `X`, outputs the type `Prof V X → set X`, which is the type of (V, X) -SCCs.

We formalize universal domain as follows:

```
def universal_domain_SCC {V X : Type} (F : SCC V X) : Prop :=
  ∀ P : Prof V X, F P ≠ ∅
```

Example 1. For any V and X , consider the Condorcet SCC for (V, X) defined by:

$$\text{Cond}_{(V,X)}(\mathbf{P}) = \begin{cases} \{x\} & \text{if } x \text{ is a Condorcet winner in } \mathbf{P} \\ X(\mathbf{P}) & \text{otherwise} \end{cases}.$$

The definition states that given a (V, X) -profile \mathbf{P} , if there is a Condorcet winner—in which case it is unique—then output the set containing the Condorcet winner, and otherwise output all candidates in X .

We represent this (V, X) -SCC in Lean as follows:

```
def condorcet_SCC {V X : Type} : SCC V X := λ P,
  {x : X | condorcet_winner P x ∨ ¬ ∃ y, condorcet_winner P y}
```

Most voting methods (e.g., Plurality, Borda, Instant Runoff) are defined not only for a fixed set of voters and candidates but for any set of voters and candidates, which motivates the following definition.

Definition 6. A *variable-election social choice correspondence* (VSCC) is a function F that assigns to each pair (V, X) of a $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$ a (V, X) -SCC. We abuse terminology and call the set $\{\mathbf{Q} \in \text{PROF} \mid F(V(\mathbf{Q}), X(\mathbf{Q}))(\mathbf{Q}) \neq \emptyset\}$ the *domain* of F . We say that F satisfies (*finite*) *universal domain* if the domain of F includes $\{\mathbf{P} \in \text{PROF} \mid V(\mathbf{P}) \text{ and } X(\mathbf{P}) \text{ nonempty and finite}\}$.⁹

An equivalent but perhaps more intuitive approach would define a VSCC to be a function on `PROF` (rather than $\wp(\mathcal{V}) \times \wp(\mathcal{X})$) such that for each $\mathbf{Q} \in \text{PROF}$, we have $F(\mathbf{Q}) \subseteq X(\mathbf{Q})$;¹⁰ abusing terminology, we could then call the set $\{\mathbf{Q} \in \text{PROF} \mid F(\mathbf{Q}) \neq \emptyset\}$ the *domain* of the VSCC. However, we have presented Definition 6 above because it nicely connects with our formalization in Lean.

In Lean, we define the type of VSCCs as a *dependent function type*:

```
def VSCC : Type 1 := Π (V X : Type), SCC V X
```

Given $\alpha : \text{Type}$ and $\beta : \alpha \rightarrow \text{Type}$, the type $\Pi y z : \alpha, \beta y z$ is the type of functions f such that for each $a b : \alpha$, we have that $f a b$ is an element of $\beta a b$. In the definition of `VSCC` above, α is `Type` and β is `SCC`. Thus, the definition states that an element of the type `VSCC` is a function that for any types V and X returns a function of the type `SCC V X`, i.e., a (V, X) -SCC.

We formalize (*finite*) universal domain as follows:

⁹ Of course, one could also consider the stronger condition that the domain of F contains all profiles even with infinite sets of voters and/or candidates.

¹⁰ This is the definition of a *voting method* used in [23] with the additional stipulations that $F(\mathbf{Q}) \neq \emptyset$ and that $V(\mathbf{Q})$ and $X(\mathbf{Q})$ are nonempty and finite.

```
def finite_universal_domain_VSCC (F : VSCC) : Prop :=
  ∀ V X [inhabited V] [inhabited X] [fintype V] [fintype X],
  universal_domain_SCC (F V X)
```

Example 2. We define the Condorcet VSCC as follows, taking advantage of our definition for any V and X of the Condorcet (V, X) -SCC in Example 1:

```
def condorcet_VSCC : VSCC := λ V X, condorcet_SCC
```

The second type of function we consider assigns to a given profile a binary relation on the set of candidates in the profile.

Definition 7. For $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$, a *collective choice rule for (V, X)* , or (V, X) -CCR, is a function $f : \text{Prof}(V, X) \rightarrow \mathcal{B}(X)$. Let CCR be a function that assigns to each pair (V, X) of $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$ the set of all (V, X) -CCRs.

Depending on the application, one can interpret the binary relation $f(\mathbf{Q})$ in one of two ways: $xf(\mathbf{Q})y$ can mean (a) x is strictly preferred to y socially or (b) x is strictly preferred to or tied with y socially.¹¹ Once again, there is also the issue of “domain restrictions.” Under approach (a), we can mark that the CCR is “undefined” on a profile \mathbf{Q} by setting $f(\mathbf{Q}) = X(\mathbf{Q}) \times X(\mathbf{Q})$. Then we can abuse terminology and call $\{\mathbf{Q} \in \text{Prof}(V, X) \mid f(\mathbf{Q}) \neq X(\mathbf{Q}) \times X(\mathbf{Q})\}$ the domain of f . Under approach (b), we can mark that the CCR is “undefined” on \mathbf{Q} by setting $f(\mathbf{Q}) = \emptyset$. Then we can abuse terminology and call $\{\mathbf{Q} \in \text{Prof}(V, X) \mid f(\mathbf{Q}) \neq \emptyset\}$ the domain of f . A CCR f is said to be *asymmetric* (resp. *transitive*, etc.), if for all \mathbf{Q} in the domain of f , $f(\mathbf{Q})$ is asymmetric (transitive, etc.). Of course, asymmetric CCRs only make sense under interpretation (a) above, whereas under interpretation (b), CCRs should be reflexive.

In Lean, our representation of the function CCR is similar to that of SCC:

```
def CCR := λ (V X : Type), Prof V X → X → X → Prop
```

Example 3. As an example of a CCR, we consider the Split Cycle CCR studied in [24]. The output of the Split Cycle CCR is an asymmetric relation understood as a relation of “defeat” between candidates. A candidate x defeats a candidate y in \mathbf{P} just in case the margin of x over y is (i) positive and (ii) greater than the weakest margin in each majority cycle containing x and y . To formalize this definition, we first need a definition of a cycle in a binary relation:

```
def cycle {X : Type} := λ (R : X → X → Prop) (c : list X),
  ∃ (e : c ≠ list.nil), list.chain R (c.last e) c
```

Here the function `cycle` takes in a binary relation R and a list c of elements of X and outputs the proposition stating that (i) there is a proof e that c is not

¹¹ As in Footnote 3, approach (b) is more general, since it allows one to distinguish between “social indifference” and “social noncomparability” (for examples of theorems in social choice in which this distinction matters, see [21]). When notions of social indifference and noncomparability are not needed, approach (a) can be simpler.

the empty list, and (ii) c is a cycle in R . To express (ii), we use the construction `list.chain R a c`, where R is a binary relation, a is an element of X , and c is a list of elements of X , which means that a is R -related to the first element of c and that every element in the list c is related to the next element in c . Thus, if we take a as the last element of c , this implies that c is a cycle. Applying `c.last` to the proof e that c is not the empty list outputs the last element of c .

Now we are ready to define the Split Cycle (V, X) -CCR in Lean:

```
def split_cycle_CCR {V X : Type} : CCR V X :=
  λ (P : Prof V X) (x y : X), ∀ [n: fintype V],
    0 < @margin V X n P x y ∧
    ¬ (∃ (c : list X), x ∈ c ∧ y ∈ c ∧
      cycle (λ a b, @margin V X n P x y ≤ @margin V X n P a b) c)
```

Recall that the `margin` function takes as implicit arguments the set V of voters, the set X of candidates, and a proof that V is finite. The `@` symbol is used when explicitly supplying these implicit arguments. Thus, the definition states that given a profile P and two candidates x and y , the binary relation outputted by `split_cycle_CCR` holds of x, y if for any proof n that V is finite, the margin of x over y in P (supplying the `margin` function with V, X , and n) is greater than 0 and there is no list c of elements containing x and y such that c is a majority cycle for which the margin of x over y is less than or equal to every margin in the cycle, i.e., c is a cycle in the binary relation R that holds of a, b just in case the margin of x over y is less than or equal to the margin of a over b .

Once again, we can consider functions that are not restricted to a fixed set of voters and candidates.

Definition 8. A *variable-election collective choice rule* (VCCR) is a function that assigns to each pair (V, X) of a $V \subseteq \mathcal{V}$ and $X \subseteq \mathcal{X}$ a (V, X) -CCR.

An equivalent but perhaps more intuitive definition takes a VCCR to be a function f on PROF (instead of $\mathcal{V} \times \mathcal{X}$) such that for all $\mathbf{Q} \in \text{PROF}$, $f(\mathbf{Q})$ is a binary relation on $X(\mathbf{Q})$.¹² However, we have presented Definition 8 above because it nicely connects with our formalization in Lean, which as before defines the type of VCCRs to be a dependent function type:

```
def VCCR := Π (V X : Type), CCR V X
```

The definition states that an element of the type `VCCR` is a function that for any types V and X returns a function of the type `CCR V X`, i.e., a (V, X) -CCR.

Example 4. We define the Split Cycle VCCR as follows, taking advantage of our definition for any V and X of the Split Cycle (V, X) -SCC in Example 3:

```
def split_cycle_VCCR : VCCR := λ V X, split_cycle_CCR
```

¹² This is the definition of a VCCR used in [24] with the additional stipulation that $V(\mathbf{Q})$ and $X(\mathbf{Q})$ are nonempty and finite.

Any VCCR, regarded as outputting for a given profile (for a given V, X) a relation of strict social preference or “defeat,” can be transformed into a VSCC by assigning to a given profile the set of candidates who are not defeated.¹³

Definition 9. Given an asymmetric VCCR f , we define the *maximal-element induced* VSCC f_M such that for any $V \subseteq \mathcal{V}$, $X \subseteq \mathcal{X}$, and (V, X) -profile \mathbf{P} ,

$$f_M(V, X)(\mathbf{P}) = \{x \in X(\mathbf{P}) \mid \forall y \in X(\mathbf{P}), (y, x) \notin f(V, X)(\mathbf{P})\}.$$

In Lean, we formalize Definition 9 as follows:

```
def max_el_VSCC : VCCR → VSCC := λ f V X P,
  {x : X | ∀ y : X, ¬ f V X P y x}
```

Example 5. The Split Cycle voting method [23] is the maximal-element induced VSCC from the Split Cycle VCCR defined in Example 4:

```
def split_cycle : VSCC := max_el_VSCC split_cycle_VCCR
```

As is well known, any acyclic VCCR (i.e., VCCR that assigns an acyclic CCR to each V, X) induces a VSCC satisfying (finite) universal domain:

```
def acyclic {X : Type} : (X → X → Prop) → Prop :=
  λ Q, ∀ (c : list X), ¬ cycle Q c

theorem max_el_VSCC_universal_domain (f : VCCR)
  (a : ∀ V X [inhabited V] [inhabited X] [fintype V] [fintype X]
    (P : Prof V X), acyclic (F V X P)) :
  finite_universal_domain_VSCC (max_el_VSCC f) := ...
```

The proof can be found in our online repository.

2.3 Voting axioms

After formalizing the basic objects of voting theory, we formalized a number of standard axioms by which voting procedures are evaluated:

Domain axioms:

- *universal domain* (resp. *finite universal domain*): all profiles (resp. finite profiles) are in the domain of the VSCC.

Intra-profile axioms:

- *Condorcet criterion*: if there is a Condorcet winner in a profile, that candidate is the unique winner.
- *Condorcet loser criterion*: if there is a Condorcet loser in a profile—a candidate who loses to every other candidate in a head-to-head majority comparison—that candidate does not win.

¹³ An alternative approach, also easily formalizable, assigns to a given profile the set of candidates who are weakly socially preferred to all other candidates.

- *Pareto*: if all voters rank candidate x above candidate y in a profile, then y does not win.

Inter-profile axioms:

- *monotonicity*: if x wins in a profile \mathbf{P} , and \mathbf{P}' is obtained from \mathbf{P} by voters moving x up in their rankings, then x still wins in \mathbf{P}' .
- *reversal symmetry*: if x is the unique winner in a profile \mathbf{P} , then x is not a winner in the profile \mathbf{P}^r obtained from \mathbf{P} by reversing all voters' rankings.

Variable-candidate inter-profile axioms:

- *independence of clones*: see Section 3.3.
- *strong stability for winners*: if x wins in a profile \mathbf{P} , and \mathbf{P}' is obtained from \mathbf{P} by adding a new candidate y who does not beat x in a head-to-head majority comparison, then x is still a winner in \mathbf{P}' .

Variable-voter inter-profile axioms:

- *positive involvement*: if x wins in a profile \mathbf{P} , and \mathbf{P}' is obtained from \mathbf{P} by adding a new voter who ranks x as their unique first choice, then x still wins in \mathbf{P}' .
- *negative involvement*: if x does not win in a profile \mathbf{P} , and \mathbf{P}' is obtained from \mathbf{P} by adding a new voter who ranks x as their unique last choice, then x still does not win in \mathbf{P}' .

Several of these axioms involved formalizing auxiliary relations or operations on profiles. For example, monotonicity requires the ternary relation of one profile being related to another by a *simple lift* of a candidate x , meaning that x may go up in voters' rankings but the rest of their rankings remains the same:

```
def simple_lift {V X : Type} : Prof V X → Prof V X → X → Prop
:= λ P' P x, (∀ (a ≠ x) (b ≠ x) i, P i a b ↔ P' i a b) ∧
  ∀ a i, ((P i x a → P' i x a) ∧ (P' i a x → P i a x))
```

The variable-candidate and variable-voter axioms involve moving from one profile to another with a different set of candidates or voters. In fact, it is most convenient to formalize these axioms using the operations of *removing* a candidate or voter from a profile, moving from a profile of type $\text{Prof } V \ X$ to a profile of type $\text{Prof } V \ \{x : X // x \neq b\}$, in the case of removing a candidate b , or of type $\text{Prof } \{v : V // v \neq i\} \ X$, in the case of removing a voter i . See Section 3.3 for the definition of the *minus_candidate* operation.

We stress that the axiom definitions we formalized are the standard ones used throughout the literature, not ones tailored to any particular VSCC, so others can immediately use our formalizations to verify results for any VSCCs.

3 Theorems

As a proof of concept of verifying theorems using our formal framework, we verified most of the results concerning the Split Cycle voting method in [23]. In particular, we verified the equivalence of two definitions of Split Cycle (one quantifying over all cycles containing x and y , the other quantifying over only paths from y to x) and that Split Cycle satisfies all of the axioms listed in Section 2.3. However, we emphasize that Split Cycle was only chosen as a test case. Our formalization of the basic objects and axioms of voting theory can be used to verify properties of any other voting method.

Moreover, verifying the properties of Split Cycle required formalizing a number of general-purpose lemmas that will be needed—and can now be used—to verify the properties of other voting methods. Indeed, one of the benefits of a formalization project such as ours is to extract those general-purpose lemmas, which may be buried in proofs for a particular voting method, so that future formalization efforts can move more quickly to formalizing sophisticated results.

3.1 Graph-theoretic background

Before formalizing voting-theoretic proofs, we had to build up basic infrastructure for reasoning about cycles, walks, and paths in graphs, such as rotating and reversing cycles and converting walks to paths, which was not available in Mathlib. For example, to convert walks to paths, we defined an inductive type¹⁴:

```
noncomputable def to_path {X : Type} : list X → list X
| [] := []
| (u :: p) := let p' := to_path p in
  if u ∈ p' then (p'.drop (p'.index_of u)) else (u :: p')
```

Hence `to_path` maps the empty list to itself, and given a list $u :: p$ constructed by adding u to the front of the list p , if u is an element of `to_path p`, we output the result of dropping from `to_path p` all elements before u in the list, and otherwise we add u to the front of `to_path p`. A significant part of the formalization effort was proving needed properties of `to_path` and other operations on lists.

3.2 Reasoning about margins

In addition to developing graph-theoretic infrastructure, we formalized a number of general-purpose lemmas for reasoning about majority margins between candidates and the relation between changes in profiles and changes in margins. These lemmas are applicable to all voting methods for which the selection of winners is invariant between profiles with the same majority margins matrices—so-called C2 voting methods [17] (e.g., Minimax, Ranked Pairs, Beat Path, Split Cycle,

¹⁴ We can eliminate ‘noncomputable’ if we assume that equality for X is decidable, adding ‘[decidable_eq X]’ as an implicit argument.

and even Borda). To take one example, when reasoning about monotonicity (recall Section 2.3) for C2 methods, a key fact is that a simple lift of a candidate x cannot increase the margin of any candidate over x :

```
lemma margin_lt_margin_of_lift {V X : Type} (P P' : Prof V X)
  [fintype V] (y x : X) :
  simple_lift P' P x → margin P' y x ≤ margin P y x := ...
```

Other general-purpose lemmas about margins that we formalized concern how margins change when adding a voter who ranks x first—for verifying the axiom of positive involvement—or last—for verifying the axiom of negative involvement—or how clones of a candidate all bear the same margins to a given non-clone—for verifying the axiom of independence of clones. These are basic lemmas that anyone wishing to verify the relevant properties of C2 voting methods will need.

For some of the most elementary facts, Lean’s automation was adequate to complete proofs without our help. For example, the fact that reversing all voters rankings reverses all margins—used in verifying the axiom of reversal symmetry—was provable using Lean’s `obviously` tactic:

```
def reverse_profile {V X : Type} : Prof V X → Prof V X :=
  λ P v x y, P v y x

lemma margin_reverse_eq {V X : Type} [fintype V] (P : Prof V X)
  (a b : X) : margin (reverse_profile P) b a = margin P a b :=
begin
  obviously,
end
```

Similarly, Lean’s `obviously` tactic immediately proves that removing a candidate does not change the margins between remaining candidates. By contrast, proofs of lemmas like `margin_lt_margin_of_lift` required more human input.

3.3 Example: independence of clones

Our most involved formalization was of the proof from [23] that Split Cycle satisfies Tideman’s [37] axiom of independence of clones. A set C of two or more candidates is a set of *clones* in a profile \mathbf{P} if no voter ranks any candidates outside of C in between two candidates from C . Given a particular candidate c , we say that a nonempty set D of candidates (not containing c) is a set of *clones of c* if $D \cup \{c\}$ is a set of clones. In Lean, we formalize this as follows:

```
def clones {V X : Type} (P : Prof V X) (c : X)
  (D : set {x : X // x ≠ c}) : Prop :=
D.nonempty ∧ (∀ (c' ∈ D) (x : {x : X // x ≠ c}) (i : V),
  x ∉ D → ((P i c x ↔ P i c' x) ∧ (P i x c ↔ P i x c')))
```

Independence of clones for VSCCs states that (i) removing a clone from a profile should not change which non-clones win and (ii) removing a clone from a profile should not change whether at least one clone is among the winners (though which clone wins is allowed to change upon removing a clone). To formalize this, we need a way of removing a candidate from a profile, accomplished as follows:

```
def minus_candidate {V X : Type} (P : Prof V X) (b : X) :
  Prof V {x : X // x ≠ b} := λ v x y, P v x y
```

Thus, `minus_candidate` takes in a profile P for V and X , as well as a candidate b from X , and outputs the profile for V and $\{x : X // x \neq b\}$ that agrees with P on how every voter ranks the candidates other than b . Using `minus_candidate`, we formalize condition (i) of independence of clones as follows:

```
def non_clone_choice_ind_clones {V X : Type} (P : Prof V X)
  (c : X) (D : set {x : X // x ≠ c}) : VSCC → Prop := λ F,
  clones P c D → (∀ a : {x : X // x ≠ c}, a ∉ D →
    (a.val ∈ (F V X P) ↔ a ∈ (F V {x : X // x ≠ c}
      (minus_candidate P c))))
```

Since $\{x : X // x \neq c\}$ is a subtype of X , a consists of an element of X , called $a.val$, together with a proof that $a.val \neq c$. Since $F V X P$ is a set of elements of X , we must write ' $a.val \in (F V X P)$ ' instead of ' $a \in (F V X P)$ '. Finally, we can state that Split Cycle satisfies part (i) of independence of clones as follows:

```
theorem non_clone_choice_ind_clones_split_cycle {V X : Type}
  [fintype V] (P : Prof V X) (c : X) (D : set {x : X // x ≠ c}) :
  non_clone_choice_ind_clones P c D split_cycle := ...
```

The formalization of part (ii) of independence of clones is similar. The proof that Split Cycle satisfies independence of clones involves manipulating paths in the majority graph of a profile—in particular, replacing all clones in a path by a distinguished clone and then eliminating repetitions of candidates in the resulting sequence using the `to_path` operation.

4 Conclusion

Our goal was to set up a general framework for formally verifying results in voting theory. One of the benefits of such a formalization project, beyond certifying the correctness of results, is that it forces formalizers to think about how the fundamental notions of the field ought to be formulated (see, e.g., our definitions of VSCCs and VCCRs). We expect such benefits to accrue for formalization projects in other areas of mathematical social science or even natural science.

What did we learn from the verification stage of our project? As usual in formalization, we caught some omitted assumptions (e.g., of nonemptiness) in definitions needed to prove results about the Split Cycle voting method in a draft of [23], prompting corrections. A more striking lesson of formalizing these

results is how little depends on assumptions about properties of voter preferences. While it was initially assumed in [23] that voter preference relations are linear orders,¹⁵ the full strength of this assumption turned out not to be used in any proofs we formalized. In fact, most results work with no assumptions about voter preferences at all (except the default asymmetry of strict preference). The only exception was the Pareto principle, whose proof used the acyclicity of voter preferences. It would be fascinating to see exactly what properties of voter preferences are needed in formalized proofs of properties of other voting methods.

It would also be desirable to abstract away from the definition of margin in Definition 3 to define voting methods and prove theorems in terms of an abstract relation $(a, b) \succ_{\mathbf{P}} (c, d)$ expressing that the strength of majority preference for a over b in profile \mathbf{P} is stronger than the strength of majority preference for c over d in \mathbf{P} . One definition of $(a, b) \succ_{\mathbf{P}} (c, d)$ is that $\text{Margin}_{\mathbf{P}}(a, b) > \text{Margin}_{\mathbf{P}}(c, d)$. But if voter preference relations are non-linear, there are alternative, inequivalent definitions of $(a, b) \succ_{\mathbf{P}} (c, d)$, such as the *winning votes* definition: $|\{i \in V(\mathbf{P}) \mid a \mathbf{P}_i b\}| \geq |\{i \in V(\mathbf{P}) \mid c \mathbf{P}_i d\}|$. Schulze [33] considers other definitions besides margins and winning votes, and rather than settling for any one of them, lays down general axioms on the relation of relative strength of majority preference needed for proofs to go through. This is a natural next step in our project, which formalization facilitates: identify which of the properties of the $\succ_{\mathbf{P}}$ relation are actually used in our formalized proofs and then assume only those properties.

Another natural next step would be to extend the verification of voting axioms beyond voting methods, such as Split Cycle, that are based on head-to-head majority comparisons to voting methods, such as Instant Runoff, that are based on iterative elimination procedures. Instant Runoff can be defined *recursively*: the Instant Runoff winners in a profile \mathbf{P} are the Instant Runoff winners in the profile \mathbf{P}_{-Y} obtained from \mathbf{P} by removing the set Y of candidates who receive the fewest first-place votes in \mathbf{P} , if $Y \subsetneq X(\mathbf{P})$; otherwise all candidates win.¹⁶ As Lean offer natural ways of defining functions recursively and writing proofs by induction [4, § 8], we expect Lean to be well suited to verifying properties of recursively definable voting methods such as Instant Runoff.

With its axiomatic approach and discrete mathematical character, voting theory is especially amenable to formal verification. Moreover, given the importance of democratic decision making in society, we find it desirable to formally verify that democratic decision procedures have the desirable properties claimed for them. We have done so for one recently proposed voting method, but we would like to see this done for all methods proposed for use in democratic elections.

¹⁵ This assumption was made for its simplifying consequence of rendering the majority margin between two candidates the canonical measure of the strength of majority preference between candidates. Cf. the following paragraph in the main text.

¹⁶ While this definition of Instant Runoff, based on [36, p. 7], eliminates all candidates who receive the fewest first-place votes at a given stage, an alternative “parallel universe” definition states that x is an Instant Runoff winner in a profile \mathbf{P} if either x is the only candidate in \mathbf{P} or there is some candidate y who receives the fewest first-place votes in \mathbf{P} , and x is an Instant Runoff winner in \mathbf{P}_{-y} .

Acknowledgement

We thank the Lean Zulip chat community, the Berkeley Lean Seminar, and Jeremy Avigad for advice about using Lean and the two anonymous referees for helpful feedback on our paper.

References

1. Agotnes, T., van der Hoek, W., Wooldridge, M.: On the logic of preference and judgment aggregation. *Autonomous Agents and Multi-Agent Systems* **22**(1), 4–30 (2009)
2. Arrow, K.J.: *Social Choice and Individual Values*. John Wiley & Sons, Inc., New York, 1st edn. (1951)
3. Arrow, K.J.: Origins of the impossibility theorem. In: Maskin, E., Sen, A. (eds.) *The Arrow Impossibility Theorem*, pp. 143–148. Columbia University Press, New York (2014)
4. Avigad, J., de Moura, L., Kong, S.: Theorem Proving in Lean. Release 3.23.0 edn. (2021), https://leanprover.github.io/theorem_proving_in_lean/theorem_proving_in_lean.pdf
5. Baigent, N.: Twitching weak dictators. *Journal of Economics* **47**(4), 407–411 (1987)
6. Blau, J.H.: Review of the article “Repairing proofs of arrow’s general impossibility theorem and enlarging the scope of the theorem” by R. Routley. *Mathematical Reviews* **0545435** (1979), <https://mathscinet.ams.org/mathscinet-getitem?mr=545435>
7. Blau, J.H.: The existence of social welfare functions. *Econometrica* **25**(2), 302–313 (1957)
8. Brandl, F., Brandt, F., Eberl, M., Geist, C.: Proving the incompatibility of efficiency and strategyproofness via SMT solving. *Journal of the ACM* **65**(2), 6:1–6:28 (2018)
9. Brandt, F., Eberl, M., Saile, C., Stricker, C.: The incompatibility of fishburn-strategyproofness and pareto-efficiency. *Archive of Formal Proofs* (2018), https://isa-afp.org/entries/Fishburn_Impossibility.html
10. Brandt, F., Saile, C., Stricker, C.: Voting with ties: Strong impossibilities via SAT solving. In: Dastani, M., Sukthankar, G., André, E., Koenig, S. (eds.) *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018)*. pp. 1285–1293. International Foundation for Autonomous Agents and Multiagent Systems (2018)
11. Campbell, D.E., Kelly, J.S.: Weak independence and veto power. *Economics Letters* **66**, 183–189 (2000)
12. Ciná, G., Endriss, U.: Proving classical theorems of social choice theory in modal logic. *Autonomous Agents and Multi-Agent Systems* **30**(5), 963–989 (2016)
13. Coquand, T., Huet, G.: The calculus of constructions. *Information and Computation* **76**(2-3), 95–120 (1988)
14. de Moura, L., Kong, S., Avigad, J., van Doorn, F., von Raumer, J.: The lean theorem prover. In: *25th International Conference on Automated Deduction (CADE-25)* (2015)
15. Eberl, M.: Verifying randomised social choice. In: *International Symposium on Frontiers of Combining Systems, FroCoS 2019*. pp. 240–256 (2019)
16. Endriss, U.: Logic and social choice theory. In: Gupta, A., van Benthem, J. (eds.) *Logic and Philosophy Today*, pp. 333–377. College Publications, London (2011)

17. Fishburn, P.C.: Condorcet social choice functions. *SIAM Journal on Applied Mathematics* **33**(3), 469–489 (1977)
18. Geanakoplos, J.: Three brief proofs of Arrow’s theorem. *Economic Theory* **26**(1), 211–215 (2005)
19. Geist, C., Peters, D.: Computer-aided methods for social choice theory. In: Endriss, U. (ed.) *Trends in Computational Social Choice*. pp. 249–267. *AI Access* (2017)
20. Grandi, U., Endriss, U.: First-order logic formalisation of impossibility theorems in preference aggregation. *Journal of Philosophical Logic* **42**(4), 595–618 (2013)
21. Holliday, W.H., Kelley, M.: A note on Murakami’s theorems and incomplete social choice without the Pareto principle. *Social Choice and Welfare* **55**, 243–253 (2020)
22. Holliday, W.H., Pacuit, E.: Arrow’s decisive coalitions. *Social Choice and Welfare* **54**, 463–505 (2020)
23. Holliday, W.H., Pacuit, E.: Split Cycle: A new Condorcet consistent voting method independent of clones and immune to spoilers (2020), [arXiv:2004.02350](https://arxiv.org/abs/2004.02350)
24. Holliday, W.H., Pacuit, E.: Axioms for defeat in democratic elections. *Journal of Theoretical Politics* (Forthcoming), [arXiv:2008.08451](https://arxiv.org/abs/2008.08451)
25. Marginal-Löf, P.: *Intuitionistic type theory*. Bibliopolis, Napoli (1984)
26. Murakami, Y.: *Logic and Social Choice*. Dover Publications, New York (1968)
27. Nipkow, T.: Social choice theory in HOL. *Journal of Automated Reasoning* **43**(3), 289–304 (2009)
28. Pacuit, E., Yang, F.: Dependence and independence in social choice: Arrow’s theorem. In: Abramsky, S., Kontinen, J., Väänänen, J., Vollmer, H. (eds.) *Dependence Logic*, pp. 235–260. Springer (2016)
29. Parikh, R.: The logic of games and its applications. In: Karpinski, M., van Leeuwen, J. (eds.) *Topics in the Theory of Computation*, North-Holland Mathematics Studies, vol. 102, pp. 111–139. North-Holland (1985)
30. Pauly, M.: On the role of language in social choice theory. *Social Choice and Welfare* **163**, 227–243 (2008)
31. Routley, R.: Repairing proofs of arrow’s general impossibility theorem and enlarging the scope of the theorem. *Notre Dame Journal of Formal Logic* **20**(4), 879–890 (1979)
32. Rubinstein, A.: The single profile analogues to multi profile theorems: Mathematical logic’s approach. *International Economic Review* **25**(3), 719–730 (1984)
33. Schulze, M.: A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social Choice and Welfare* **36**, 267–303 (2011)
34. Sen, A.: *Collective Choice and Social Welfare: An Expanded Edition*. Harvard University Press, Cambridge, Mass. (2017)
35. Tang, P., Lin, F.: Discovering theorems in game theory: Two-person games with unique pure nash equilibrium payoffs. *Artificial Intelligence* **175**, 2010–2020 (2011)
36. Taylor, A.D., Pacelli, A.M.: *Mathematics and Politics: Strategy, Voting, Power, and Proof*. Springer, New York, 2nd edn. (2008)
37. Tideman, T.N.: Independence of clones as a criterion for voting rules. *Social Choice and Welfare* **4**, 185–206 (1987)
38. Troquard, N., van der Hoek, W., Wooldridge, M.: Reasoning about social choice functions. *Journal of Philosophical Logic* **40**(4), 473–498 (2011)
39. Wiedijk, F.: Arrow’s impossibility theorem. *Formalized Mathematics* **15**(4), 171–174 (2007)