Computational Social Choice 2023

Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

[http://www.illc.uva.nl/~ulle/teaching/comsoc/2023/]

Plan for Today

How do you explain why a given collective decision is the right one?

The axiomatic method seems relevant, given that axioms can motivate voting rules, which in turn produce decisions when applied to profiles.

$$\begin{tikzcd} \text{axioms} \arrow[r] \arrow[rr, bend right, dashed, "?"'] & \text{rules} \arrow[r] & \text{decisions} \end{tikzcd}$$

Today we want to explore this idea in some detail. It will turn out to be another potential application of SAT solvers in social choice.

Notational Remark

Today (only) we will use X rather than A for the set of alternatives, as we need A as a variable ranging over axioms.

Explainability in Social Choice

Can the axiomatic method help us *explain* why a given outcome for a given profile of preferences might be *the right outcome*?

Yes, to a certain extent:

- let's say, given profile $r^\star$, we want to explain the election of $X^\star$
- suppose $F(r^\star) = X^\star$ for some voting rule F
- suppose F is characterised by the set of axioms $\mathcal{A}$
- suppose we consider the axioms in $\mathcal{A}$ to be normatively appealing
- then we might say that we have an argument for electing $X^\star$ in $r^\star$

But there are a number of problems here:

- *few characterisation results*, some with *unattractive axioms*
- some appealing axioms also feature in *impossibility results*
- we hardly can expect our audience to *understand* the results used
- overkill: we just care about $r^\star$, *not all profiles*

Exercise: *Any ideas for how to think about explainability instead?*

Example

$\blacksquare \succ \blacktriangle \succ \bullet$

$\bullet \succ \blacktriangle \succ \blacksquare$

$\blacksquare \succ \blacktriangle \succ \bullet$

Exercise: *Can you think of a voting rule that makes* $\blacksquare$ *win?*

Example

$\blacksquare \succ \blacktriangle \succ \bullet$

$\bullet \succ \blacktriangle \succ \blacksquare$

$\blacksquare \succ \blacktriangle \succ \bullet$

Exercise: *Can you think of a voting rule that makes* $\blacktriangle$ *win?*

Example

$\blacksquare \succ \blacktriangle \succ \bullet$

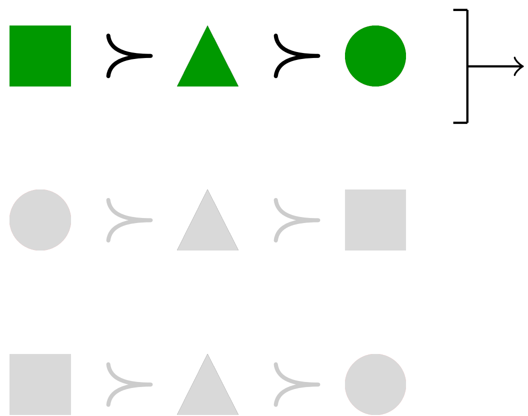
$\bullet \succ \blacktriangle \succ \blacksquare$

$\blacksquare \succ \blacktriangle \succ \bullet$

What's a good outcome?

Why?

Example



$\{\blacksquare\}$

Clear winner!

(FAITHFULNESS)

Example

$\blacksquare \succ \blacktriangle \succ \bullet$ $\mapsto$ $\{\blacksquare\}$ *Clear winner!* (FAITHFULNESS)

$\bullet \succ \blacktriangle \succ \blacksquare$

$\blacksquare \succ \blacktriangle \succ \bullet$

$\mapsto$ $\{\blacksquare, \blacktriangle, \bullet\}$ *Note the symmetry!* (CANCELLATION)

Example

$\blacksquare \succ \blacktriangle \succ \bullet$ $\longrightarrow$ $\{\blacksquare\}$

Clear winner!

(FAITHFULNESS)

$\bullet \succ \blacktriangle \succ \blacksquare$

$\blacksquare \succ \blacktriangle \succ \bullet$ $\longrightarrow$ $\{\blacksquare, \blacktriangle, \bullet\}$

Note the symmetry!

(CANCELLATION)

$\longrightarrow$ $\{\blacksquare\}$

First voter breaks tie!

(REINFORCEMENT)

The Model

We will work wit a *variable-electorate model* (with a finite universe) to be able to formally deal with axioms such as reinforcement.

Suppose some of the *voters* in a finite universe $N^\star$ express *preferences* over the *alternatives* in a set X.

We consider *voting rules* defined on all *profiles* for subelectorates:

$$F : \mathcal{L}(X)^{N \subseteq N^\star} \to 2^X \setminus \{\emptyset\}$$

Axioms: Interpretation and Instances

Attractive rules might satisfy *axioms* such as *neutrality*, *Pareto*, . . .

The *interpretation* of an axiom A is just a set of voting rules:

$$\mathbb{I}(A) \quad \subseteq \quad \mathcal{L}(X)^{N \subseteq N^\star} \to 2^X \setminus \{\emptyset\}$$

Example: $\mathbb{I}(\textsc{neu}) = \{\, \textsc{borda}, \textsc{copeland}, \ldots, F_{4711}, \ldots \}$

An *instance* A' of axiom A (for a specific profile, etc.) is what you think it is, and itself an axiom, with $\mathbb{I}(A) = \bigcap_{A' \in \mathrm{Inst}(A)} \mathbb{I}(A')$.

Example: $\mathrm{Inst}(\textsc{par}) = \{$ "*don't elect* c *in* $(abc^{[2]}, bca^{[5]})$*!*", . . . $\}$

Justification = Normative Basis + Explanation

How do you justify selecting outcome $X^\star$ for a given preference profile?

Find axiom set $\mathcal{A}^{\mathrm{NB}}$ (*normative basis*) and set of axiom instances $\mathcal{A}^{\mathrm{EX}}$ (*explanation*) regarding specific scenarios meeting these conditions:

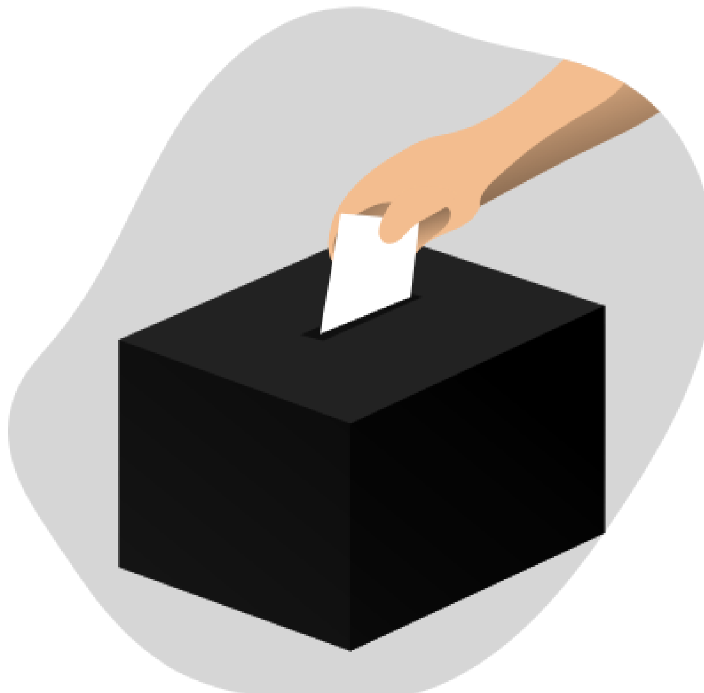
- *Adequacy:* axioms in $\mathcal{A}^{\mathrm{NB}}$ are acceptable to the user
- *Relevance:* $\mathcal{A}^{\mathrm{EX}}$ only includes instances of axioms in $\mathcal{A}^{\mathrm{NB}}$
- *Explanatoriness:* every voting rule satisfying $\mathcal{A}^{\mathrm{EX}}$ returns $X^\star$ (and $\mathcal{A}^{\mathrm{EX}}$ is tight: none of its proper subsets have the same property)
- *Nontriviality:* at least one voting rule satisfies $\mathcal{A}^{\mathrm{NB}}$

A. Boixel and U. Endriss. Automated Justification of Collective Decisions via Constraint Solving. AAMAS-2020.

Scenario 1: Confidence in Election Results

Scenario 2: Deliberation Support

Scenario 3: Justification Generation as Voting



Exercise: *What is the name of this well-known voting rule?*

$$F_{\{\mathrm{CON}\} \gg \{\mathrm{NEU}, \mathrm{REI}, \mathrm{FAI}, \mathrm{CAN}\}}$$

Computing Justifications

We can encode axiom instances in propositional logic with variables $p_{r,x}$ to say alternative x is amongst the winners in profile r.

Encode *instances* of adequate axioms and *goal constraint* $F(r^\star) \neq X^\star$. Then use a *SAT solver* to check whether this set is *satisfiable:*

- If *yes*, no justification exists.
- If *no*, a justification $\langle \mathcal{A}^{\mathrm{NB}}, \mathcal{A}^{\mathrm{EX}} \rangle$ exists if these steps succeed:
  - Find an MUS (*minimal unsatisfiable subset*) that includes the goal constraint. Let $\mathcal{A}^{\mathrm{EX}}$ be MUS $\setminus$ {goal constraint}.
  - Let $\mathcal{A}^{\mathrm{NB}}$ be the set of adequate axioms with instances in $\mathcal{A}^{\mathrm{EX}}$. Check that $\mathcal{A}^{\mathrm{NB}}$ is *satisfiable* (for nontriviality).

Algorithmic Refinement

Highly complex! Luckily, all intractable subtasks map to a well-studied problems in automated reasoning (*SAT solving* and *MUS extraction*).

The main algorithmic challenge then is to generate the solver input. Generating all axiom instances is too much, so we need heuristics to identify *relevant* axiom instances.

An approach that works well is to do *breadth-first search* in the graph induced by profiles (nodes) and axiom instances (edges).

O. Nardi, A. Boixel, and U. Endriss. A Graph-Based Algorithm for the Automated Justification of Collective Decisions. AAMAS-2022.

Structured Explanations

For now, an *explanation* is a minimal set of axiom instances that forces the outcome we want. But *how* it does so is not (yet) captured.

Ultimately, we want to get a *structured explanation* that encodes an easily understandable proof for this claim of $\mathcal{A}^{\mathrm{EX}}$ forcing $X^\star$.

One idea is to use a *tableaux-style calculus* for reasoning about voting rules to construct such structured explanations.

The calculus manipulates statements of the form $\langle r, \mathcal{O} \rangle$, where r is a profile and $\mathcal{O}$ is the range of outcomes still considered possible for r. We use axioms to narrow down these ranges until we find $\langle r^\star, \{X^\star\} \rangle$.

This representation is reasonably close a *natural-language explanation*.

A. Boixel, U. Endriss, and R. de Haan. A Calculus for Computing Structured Justifications for Election Outcomes. AAAI-2022.

Demo

For small preference profiles, you can try it out for yourself:

https://demo.illc.uva.nl/justify/

Remark: For this demo the axiom of *anonymity* is always included, so we can express profiles more compactly (number of voters per ballot).

Exercise: *Generate a justification for our original example!*

$$\blacksquare \succ \blacktriangle \succ \bullet$$

$$\bullet \succ \blacktriangle \succ \blacksquare$$

$$\blacksquare \succ \blacktriangle \succ \bullet$$

A. Boixel, U. Endriss, and O. Nardi. Displaying Justifications for Collective Decisions. IJCAI-2022 (Demo Track).

Good Explanations

What makes for a good/convincing/understandable explanation?

We don't really know (yet). This will require careful *empirical studies*.

The Bigger Picture

Consult my paper with Olivier Cailloux (2016), the position paper by Procaccia (2019), and the survey by Suryanarayana et al. (2022) for a broader discussion of explainability in multiagent decision making.

O. Cailloux and U. Endriss. Arguing about Voting Rules. AAMAS-2016.

A.D. Procaccia. Axioms Should Explain Solutions. In J.-F. Laslier et al. (eds), *The Future of Economic Design*. Springer, 2019.

S.A. Suryanarayana, D. Sarne, and S. Kraus. Explainability in Mechanism Design: Recent Advances and the Road Ahead. EUMAS-2022.

Summary

To approach the idea of *explainability in social choice*, we investigated the notion of *axiomatic justification* for election outcomes:

- Scenarios: Confidence Building | Deliberation Support | Voting
- Definition: Justification = Normative Basis + Explanation
- Algorithm: Graph Search + MUS Generation + SAT Solving
- Structured Explanations via Tableaux-style Calculus for Voting

What next? More on the use of logical modelling in social choice.