Automated Reasoning for Democracy

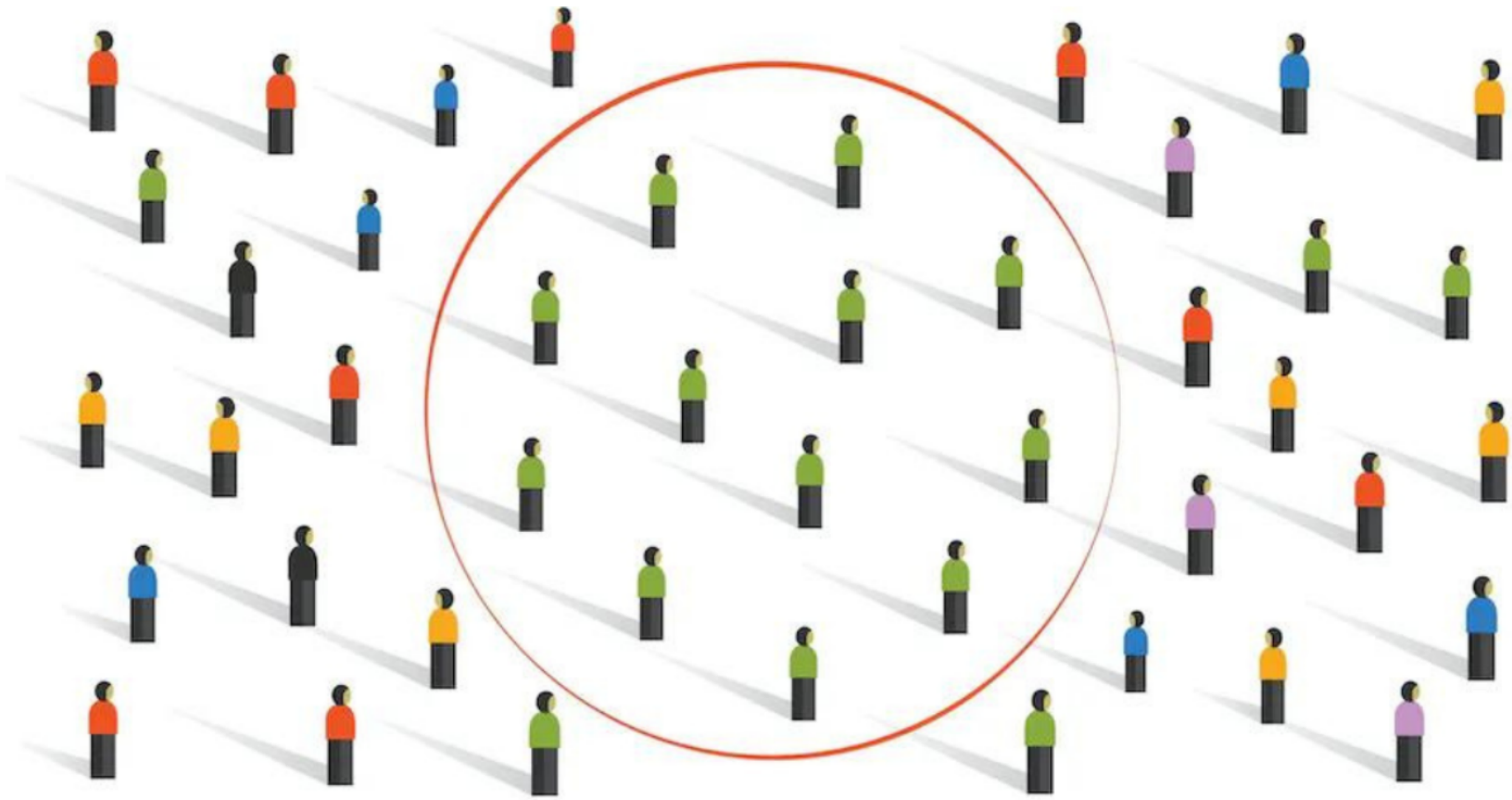
Ulle Endriss
Institute for Logic, Language and Computation
University of Amsterdam

The Washington Post
Democracy Dies in Darkness

Opinion | The next level of AI is approaching. Our democracy isn't ready.

By Danielle Allen
Contributing columnist | + Follow

April 26, 2023 at 6:30 a.m. EDT



Algopopulism: The algorithmic threat to democracy

22nd March 2023 by Editor BizNews

BizNews
Get context. Know more.

AI and Democracy

When talking about democracy, AI has a bad name (for good reasons).

But can we also use algorithmic ideas *to support* democracy?

- Can we support *researchers* investigating democratic mechanisms?
- Can we support *citizens* who are subjected to those mechanisms?

Talk Outline

I'll show how to use *automated reasoning* in support of democracy:

- Case Study 1: helping *researchers* analyse *matching markets*
- Case Study 2: helping *citizens* appreciate *election outcomes*

The Axiomatic Method

When searching for a mechanism to transform individual preferences into democratic decisions, we should start by clarifying our normative requirements ("*axioms*"): *fairness*, *efficiency*, *strategyproofness*, ...

Often impossible to satisfy all axioms. Famous examples:

- **Arrow's Theorem:** *For $m \geqslant 3$ alternatives, no preference aggregation rule is Paretian, independent, and nondicatorial.*
- **Gibbard-Satterthwaite Theorem:** *For $m \geqslant 3$ alternatives, no voting rule is strategyproof, onto, and nondictatorial.*
- **Roth's Theorem:** *For $n \geqslant 2$ agents on each side of the market, no matching mechanism is both stable and strategyproof.*

Such results provide crucial insights but are notoriously hard to prove!

Automated Reasoning

So establishing impossibility theorems is difficult. *Can AI help?* Yes!

Tang and Lin pioneered an exciting approach where we encode axioms as *Boolean formulas* and use a *SAT solver* to prove unsatisfiability.

The approach has been used to find *new proofs* for known results, to discover *new results*, and to *uncover mistakes* in the literature.

P. Tang and F. Lin. Computer-aided Proofs of Arrow's and other Impossibility Theorems. *Artificial Intelligence*, 2009.

Case Study: Fairness in Matching Markets

Scenario: Two groups of n *agents* each. Each agent ranks all the members of the other group. *Find a good matching!*

Applications: job markets, school admissions, kidney transplants

Would like a mechanism with good normative properties (*axioms*):

- *Stability:* never beneficial for two agents to leave the market
- *Fairness:* (for example) no advantage for one side of the market

The classic 1962 algorithm achieves stability, but treats the "left" side of the market better than the "right" side. *Can we do better?*

D. Gale and L. Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 1962.

Encoding

For a fixed number of agents, we can encode axioms in Boolean logic with variables $x_{p \triangleright (i,j)}$ (“*match* i *and* j *in profile* p”). Example:

$$\bigwedge_{p} \bigwedge_{i} \bigwedge_{j} \bigwedge_{i' \prec_j i} \bigwedge_{j' \prec_i j} \left(\neg x_{p \triangleright (i,j')} \vee \neg x_{p \triangleright (i',j)}\right)$$

Exercise: *What is the name of this axiom?*

Remark: For $n = 3$ agents on each side of the market, above formula is a conjunction of $419,904$ *clauses* (big, yet manageable).

Impossibility Theorem

Axiom: call a mechanism *left/right-fair* if swapping the two sides of the market never changes the outcome. Can encode this as well.

Let's run a *SAT solver* on what we prepared:

```
>>> setDimension(3)
>>> cnf = cnfMechanism() + cnfStable() + cnfLeftRight()
>>> solve(cnf)
'UNSATISFIABLE'
```

So we obtain a new impossibility theorem!

Impossibility Theorem: *For $n \geqslant 3$ agents on each side of the market, no matching mechanism is both stable and left/right-fair.*

Discussion: *Does this count? Do we believe in computer proofs?*

U. Endriss. Analysis of One-to-One Matching Mechanisms via SAT Solving: Impossibilities for Universal Axioms. AAAI-2020.

Computer Proofs

We can *proof-read the script* used to generate our formulas just as we would proof-read a paper. And we can use *multiple SAT solvers* and check they agree. So we can have *confidence* in the result.

Missing Pieces

But some pieces are still missing:

- *Does the theorem really generalise to arbitrary $n \geqslant 3$?*

  Clear for our case. But we can do better: *Preservation Theorem* identifies simple conditions on axioms licensing this generalisation.

- *Why does the theorem hold?* This proof does not tell us.

  But SAT technology can help here as well: *MUS extraction*

U. Endriss. Analysis of One-to-One Matching Mechanisms via SAT Solving: Impossibilities for Universal Axioms. AAAI-2020.

Case Study: Explainability in Voting

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

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

axioms $\longrightarrow$ rules $\longrightarrow$ decisions

?

Example

■ $\succ$ ▲ $\succ$ ●

● $\succ$ ▲ $\succ$ ■

■ $\succ$ ▲ $\succ$ ●

Exercise: *Can you think of a voting rule that makes* ■ *win?*

Example

■ ≻ ▲ ≻ ●

● ≻ ▲ ≻ ■

■ ≻ ▲ ≻ ●

Exercise: *Can you think of a voting rule that makes ▲ win?*

Example

■ $\succ$ ▲ $\succ$ ●

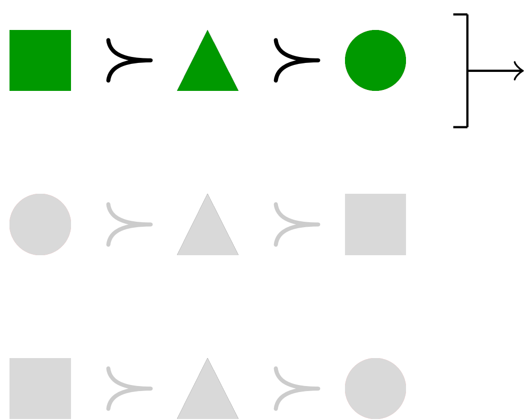
● $\succ$ ▲ $\succ$ ■

■ $\succ$ ▲ $\succ$ ●

What's a good outcome?

Why?

Example



$\{\blacksquare\}$

Clear winner!

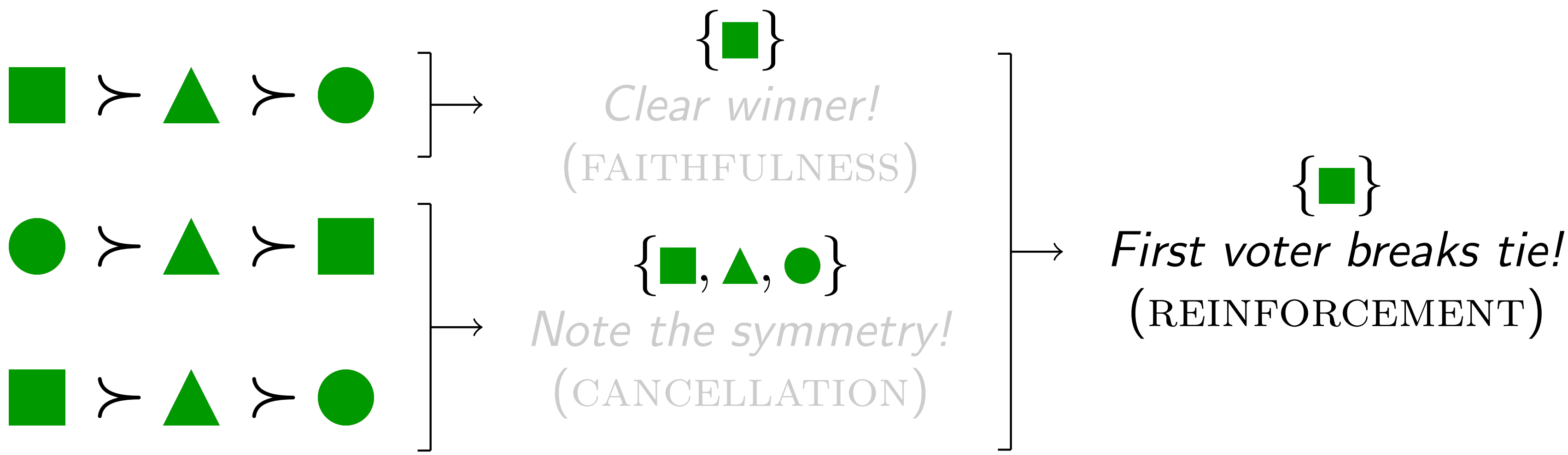
(FAITHFULNESS)

Example

■ $\succ$ ▲ $\succ$ ● ↦ {■}
Clear winner!
(FAITHFULNESS)

● $\succ$ ▲ $\succ$ ■
■ $\succ$ ▲ $\succ$ ● ↦ {■, ▲, ●}
Note the symmetry!
(CANCELLATION)

Example


Clear winner!
(FAITHFULNESS)
Note the symmetry!
(CANCELLATION)
First voter breaks tie!
(REINFORCEMENT)

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 none of $\mathcal{A}^{\mathrm{EX}}$'s proper subsets have the same property)
- *Nontriviality:* at least one voting rule satisfies $\mathcal{A}^{\mathrm{NB}}$

We can operationalise all of this using *SAT-solving* technology!

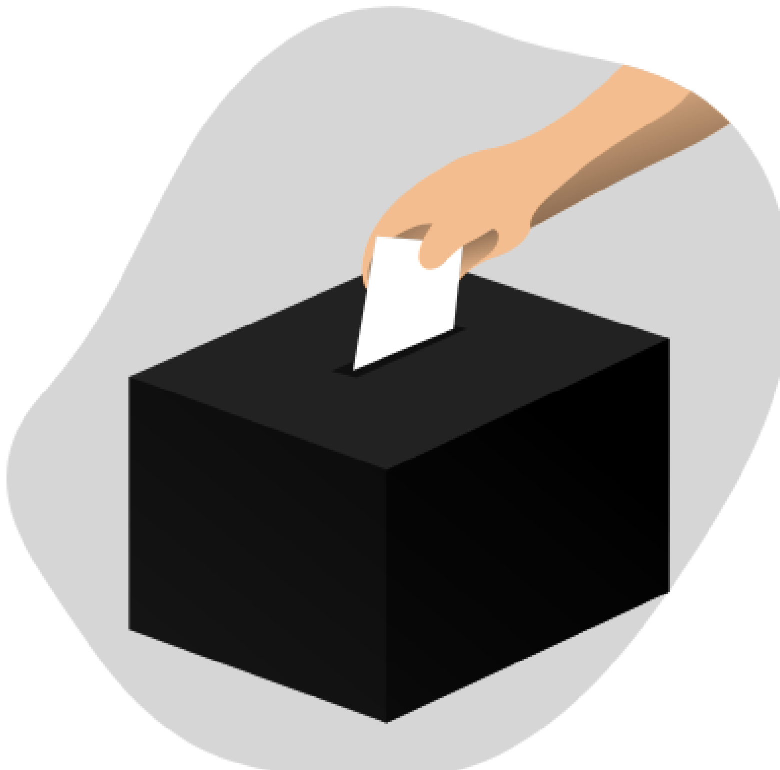
Main idea is to compute *MUS* of all instances of all acceptable axioms, together with formula saying that $X^\star$ is *not* selected in given profile.

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



M.C. Schmidtlein and U. Endriss. Voting by Axioms. AAMAS-2023.

Demo

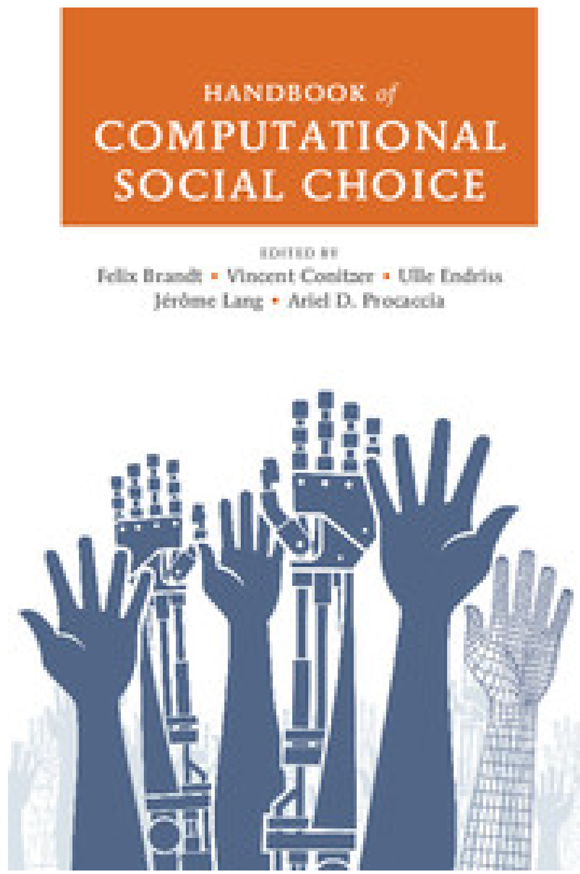
Use this tool to compute an axiomatic justification and a step-by-step explanation for a preference profile and target outcome of your choice:

`bit.ly/xsoc-demo`

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

Computational Social Choice

All of this is part of computational social choice, the study of collective decision making using, amongst others, the tools of computer science.



Last Slide

I illustrated an intriguing approach for using *SAT-solving technology* to support reasoning about *democratic decision making*:

- encode normative requirements as Boolean formulas
- use SAT solver to look for inconsistency / entailment
- use minimally unsatisfiable subset (MUS) as proof / explanation

This approach enables (at least) two exciting applications:

- helping *researchers* to prove (impossibility) theorems
- helping *citizens* to understand normative grounds for decisions

Message: *If done right, AI can have a positive impact on democracy!*

U. Endriss. Automated Reasoning for Social Choice Theory. Hands-on tutorial taught at AAMAS-2023. Slides and code available at bit.ly/tut7aamas.