Axiomatic Justification of Election Outcomes

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

[joint work with Arthur Boixel]

COMSOC Video Seminar — 3 July 2020

Example

Definition

Scenarios

Algorithm

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

■ $\succ$ ▲ $\succ$ ●

● $\succ$ ▲ $\succ$ ■

■ $\succ$ ▲ $\succ$ ●

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

Example

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

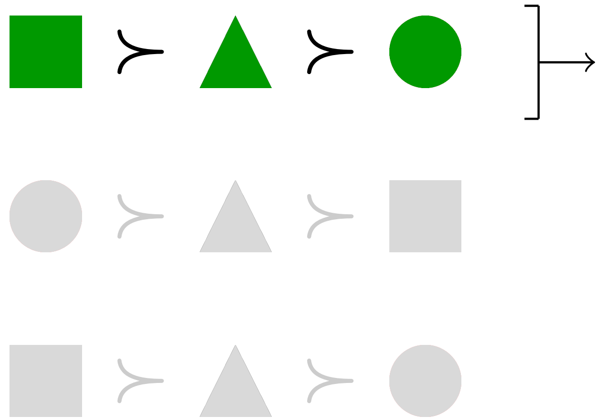
$\bullet \succ \blacktriangle \succ \blacksquare$

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

What's a good outcome?

Why?

Example

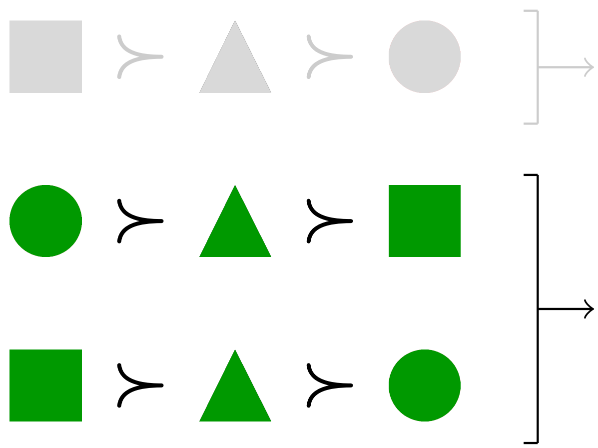


$\{\blacksquare\}$

Clear winner!

(FAITHFULNESS)

Example



{■}

Clear winner!

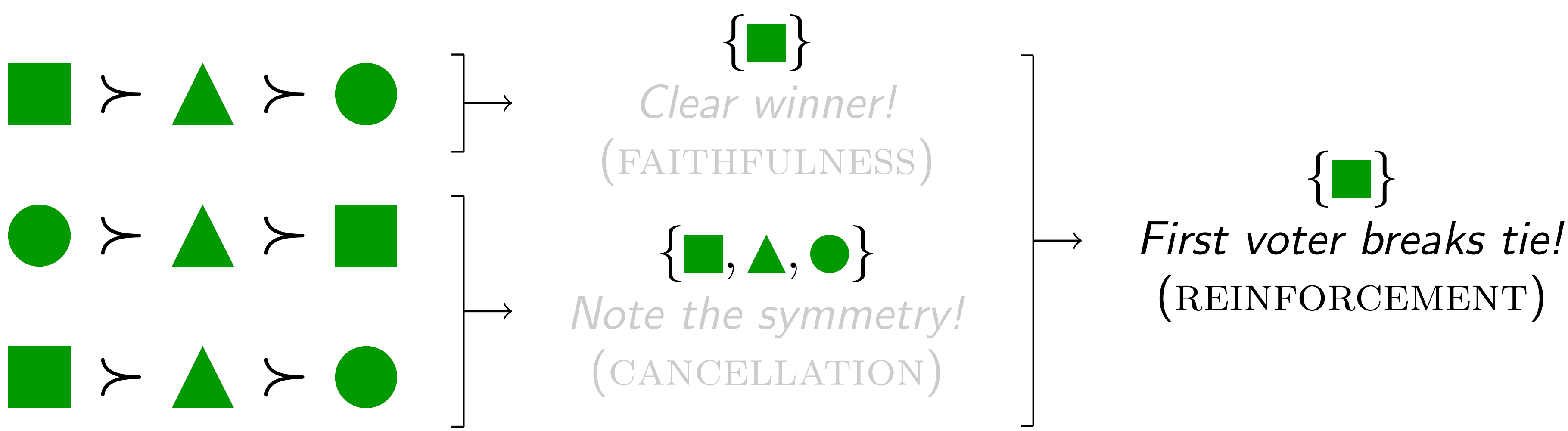
(FAITHFULNESS)

{■, ▲, ●}

Note the symmetry!

(CANCELLATION)

Example



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

The Model

Suppose *agents* in N^* express *preferences* over *alternatives* in X. Consider *voting rules* F defined on *profiles* for subelectorates:

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

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

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

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

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 $(abc^{[2]}, bca^{[5]})$!*", $\ldots \}$

Proposal for a Definition

How can you justify an election outcome $X^* \subseteq X$ for a profile $\succ_{N^*}$ using axioms from a (large!) corpus $\mathbb{A}$?

Justification $=$ *Normative Basis* $+$ *Explanation*

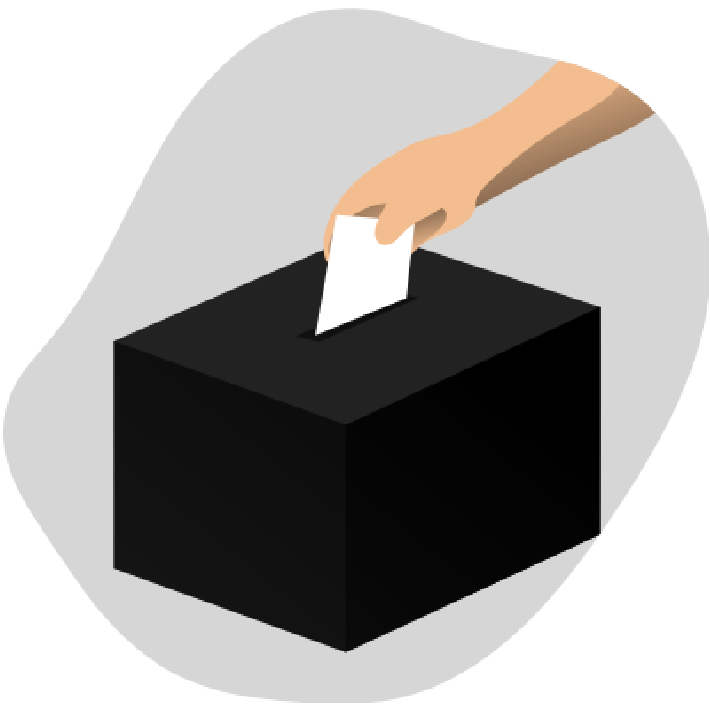
A pair $\langle \mathcal{A}^{\mathrm{NB}}, \mathcal{A}^{\mathrm{EX}} \rangle$ of sets of axioms is a justification if it satisfies:

- *Adequacy:* $\mathcal{A}^{\mathrm{NB}} \subseteq \mathbb{A}$
- *Relevance:* $\mathcal{A}^{\mathrm{EX}}$ is a set of instances of the axioms in $\mathcal{A}^{\mathrm{NB}}$
- *Explanatoriness:* $F(\succ_{N^*}) = X^*$ for all rules $F \in \bigcap_{A' \in \mathcal{A}^{\mathrm{EX}}} \mathbb{I}(A')$ and this is not the case for any proper subset of $\mathcal{A}^{\mathrm{EX}}$
- *Nontriviality:* $\bigcap_{A \in \mathcal{A}^{\mathrm{NB}}} \mathbb{I}(A) \neq \emptyset$ (*some* rule satisfies all axioms)

Scenario 1: Confidence in Election Results

Scenario 2: Deliberation Support

Scenario 3: Justification Generation as Voting



<u>Exercise:</u> *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 $p_{x \in F(\succ_N)}$. Can also use other languages for constraint satisfaction.

Encode *all instances* of axioms in $\mathbb{A}$ together with *goal constraint* expressing $F(\succ_{N^*}) \neq X^*$. 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 axioms in $\mathbb{A}$ with instances in $\mathcal{A}^{\mathrm{EX}}$. Check that $\mathcal{A}^{\mathrm{NB}}$ is *satisfiable* (for nontriviality).

Highly complex! But all computationally intractable tasks directly map to well-studied standard problems in automated reasoning.

Possible Directions for Future Work

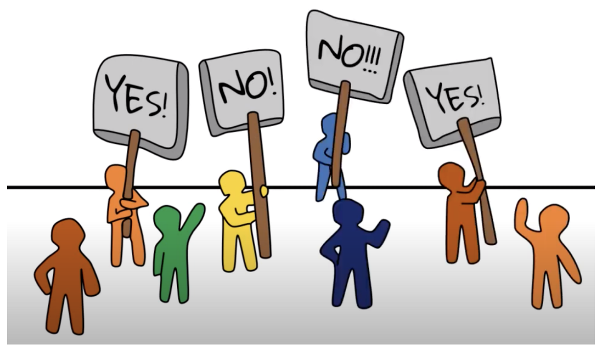
- *Beyond simple voting:* Can you adapt this idea to other models, such as multiwinner voting, matching, or judgment aggregation?
- *Algorithmic angle:* We are using SAT and constraint solving. Can you think of other promising algorithmic approaches?
- *Cognitive angle:* How do you present justifications to people? What makes justifications convincing?
- *Broader research agenda:* How can we use computers to support people in '*arguing about voting rules*'?

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

Last Slide

I proposed a notion of *axiomatic justification* for election outcomes:

- Definition: Justification = Normative Basis + Explanation
- Algorithm: Justification Generation = MUS Generation + SAT
- Scenarios: Confidence Building | Deliberation Support | Voting
- Opportunities: *lots of potential for follow-up research ...*



[http://bit.ly/watch-our-movie]

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