Axiomatic Engineering

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

[20th European Meeting on Game Theory
Maastricht, 18 June 2025]

Outline

The axiomatic method is a major cornerstone of economic theory.

How can algorithmic thinking inform how we use the axiomatic method?

(1) classification — (2) automation — (3) explainability

The Axiomatic Method in Economic Theory

Context: Examples will come from voting theory and matching under preferences, where solution concepts map profiles of preferences into collective decisions.

How we use the axiomatic method:

- identify relevant normative requirements: *fairness*, *strategyproofness*, ...
- formalise those requirements in the form of axioms
- explore the logical consequences of those axioms

Results might include separation, characterisation, or impossibility results.

Credo: Find natural and logically weak axioms with surprising consequences!

Formal Representation of Axioms

We tend to describe axioms using a combination of English and Mathematics:

> *"Condition P: If every individual prefers any alternative x to another alternative y, then society must prefer x to y." — Amartya Sen, 1970*

Often appropriate. *But could (or <u>should</u>) we be more formal than that?*

Explicit Representation

Think of axiom as the set of mechanisms that satisfy it (extensional semantics):

$$\mathbb{I}(A) \quad = \quad \{\, F \mid \textit{mechanism } F \textit{ satisfies axiom } A \,\}$$

Discussion: *set-inclusion* — *intersection* — *cardinality*

Example: Classification of Axioms

In voting theory, Fishburn (in 1973) introduced the notion of intraprofile axiom, albeit without providing a formal definition. We can now give such a definition:

- Set of outcomes for profile P that would be consistent with axiom A:

$$A(P) \;=\; \{\, F(P) \mid F \in \mathbb{I}(A) \,\}$$

- Axiom A is an intraprofile axiom iff this is true:

$$\mathbb{I}(A) \;=\; \bigcap_{\textit{profile } P} \{\, F \mid F(P) \in A(P) \,\}$$

Details worked out in the MSc thesis of my student Marie Schmidtlein (UvA, 2022).

Logical Representation

Another form of representation is to encode axioms into mathematical logic:

propositional logic — modal logic — predicate logic

Axioms now become:

comparable in view of the expressive power required to encode them

computer-readable objects we can pass on to an algorithm

U. Endriss. Logic and Social Choice Theory. In *Logic and Philosophy Today*, 2011.

Example: Encoding in Propositional Logic

If the set of profile/outcome pairs is finite, then propositional (boolean) logic can express anything we might want. Just create propositional variables like this:

$x_{p \,\triangleright\, c}$ <u>true</u> if in the pth profile the cth candidate wins

$x_{p \,\triangleright (i,j)}$ <u>true</u> if in the pth profile the ith left and jth right agent are matched

Now axioms become formulas of propositional logic. Example from matching:

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

<u>Exercise:</u> *What is the name of this axiom?*

Impossibility Theorems

Often impossible to satisfy all axioms we care about. Famous examples:

- **Gibbard-Satterthwaite Theorem:** *For elections with $m \geqslant 3$ alternatives, no resolute voting rule is strategyproof, nonimposed, and nondictatorial.*
- **Roth's Theorem:** *For matching scenarios with $n \geqslant 2$ agents on each side of the market, no matching mechanism is both stable and strategyproof.*

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

Automated Theorem Proving

Insight: A given combination of axioms is impossible to satisfy together iff the corresponding conjunction of propositional formulas is unsatisfiable.

This suggests an approach to automating the search for impossibility results:

(1) For fixed parameters (say, 2 voters and 3 alternatives for voting; $3+3$ agents for matching), encode the axioms of interest in propositional logic.

(2) Use a SAT-solver to check the conjunction of our formulas for unsatisfiability. This conjunction might be big (millions of clauses), but this often works well. Use additional tools to extract, shorten, and understand the proof trace.

(3) Use conventional methods to generalise to arbitrary parameters.

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

U. Endriss. *Tutorial on Automated Reasoning for Social Choice Theory*, 2024. [bit.ly/satsct]

Some Results

Examples from my own work (others have done similar work):

- **Approval-Based Committee Voting** (with Kluiving et al., 2020)
  Generalising an impossibility due to Dominik Peters regarding proportionality, strategyproofness, and efficiency to the case of irresolute voting rules.

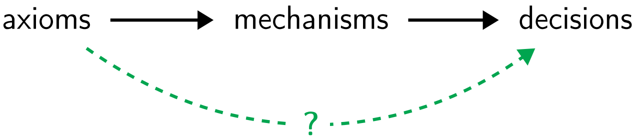
- **One-to-One Matching** (2020)
  General Preservation Theorem, yielding strengthening of Roth's Theorem and impossibility for stability and "fairness" uncovering a mistake in the literature.

- **Ranking Sets of Objects** (with Geist, 2011)
  Found all 84 impossibility theorems in a space of 20 axioms for scenarios with $n \geqslant k$ *objects (for* $k \in \{2, \ldots, 8\}$*), both interesting and trivial, including the Kannai-Peleg impossibility and one uncovering a mistake in the literature.*

Explainability

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

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



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* ■ *win?*

Example

■ $\succ$ ▲ $\succ$ ●

● $\succ$ ▲ $\succ$ ■

■ $\succ$ ▲ $\succ$ ●

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

Example

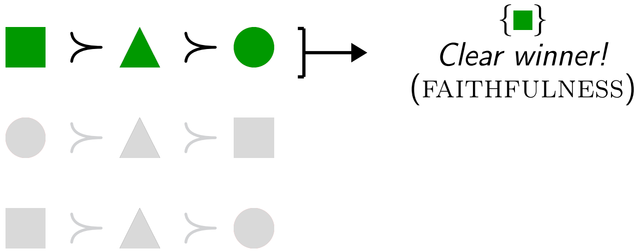
■ $\succ$ ▲ $\succ$ ●

● $\succ$ ▲ $\succ$ ■

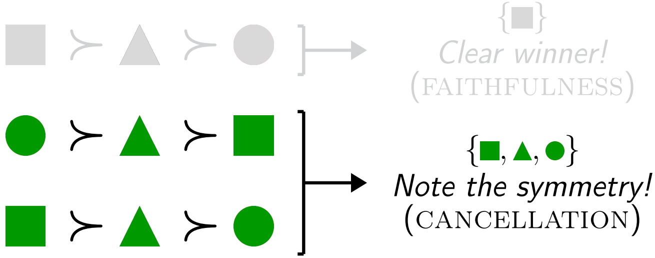
■ $\succ$ ▲ $\succ$ ●

What's a good outcome?
Why?

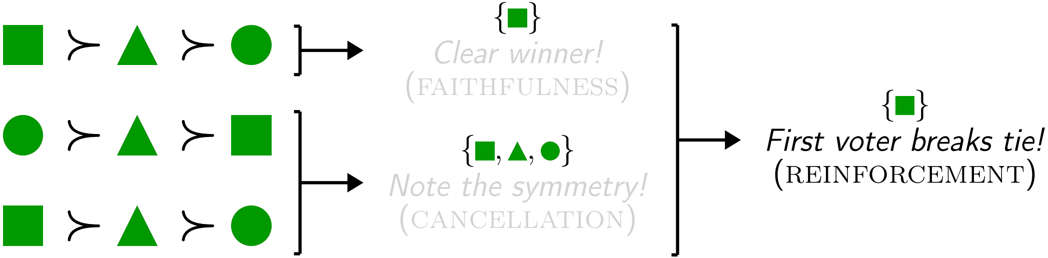
Example



Example


Clear winner!
(FAITHFULNESS)
Note the symmetry!
(CANCELLATION)

Example



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

Axiomatic Justification of Outcomes

Given a corpus of acceptable axioms, a profile, and a target outcome, we can try to compute a justification of the outcome in terms of some of those axioms:

(1) express all axiom instances as propositional formulas
(2) express that the target outcome should not be chosen as a further formula
(3) any minimally unsatisfiable subset now becomes explanation for our choice
(4) ensure nontriviality by checking the corresponding set of axioms is satisfiable

Such raw explanations can then be turned into human-readable explanations.

A. Boixel, U. Endriss, and O. Nardi. Displaying Justifications for Collective Decisions. *International Joint Conference on Artificial Intelligence*, 2022.

Last Slide

Message: Treating axioms as objects we can represent formally and reason about algorithmically can greatly enrich the axiomatic method in economic theory!

slides available at bit.ly/endriss-sing20

We saw examples for three research directions:

- classification of types of axioms
- automation of proof search for axiomatic results
- explainability of outcomes in terms of axioms