# Analysis of One-to-One Matching Mechanisms via SAT Solving: Impossibilities for Universal Axioms

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

# Talk Outline

I will try to demonstrate how the AI technique of *SAT solving* can be used for the *axiomatic analysis* of *matching mechanisms*.

- Model: one-to-one matching
- Preservation Theorem for axioms expressed in a formal language
- Approach to proving impossibility theorems via SAT solving
- Application: two impossibility theorems for matching

This talk is based on the following paper:

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

# The Model: One-to-One Matching

Two groups of *agents:* $L_n = \{\ell_1, \ldots, \ell_n\}$ and $R_n = \{r_1, \ldots, r_n\}$.
Each agent *ranks* all the agents on the opposite side of the market.
Need *mechanism* to return one-to-one *matching* given such a *profile*.

Examples: job markets, marriage markets, . . .

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

- *Stability:* no $\ell_i$ and $r_j$ prefer one another over assigned partners
- *Strategyproofness:* best strategy is to truthfully report preferences
- *Fairness:* (for example) no advantage for one side of the market

Gale-Shapley (1962): stable (✓); strategyproof for left side (✓) but not right side (✗) of the market; unfair advantage for left side (✗).

D. Gale and L.S. Shapley. College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69:9–15, 1962.

# Formal Language for Axioms

Would like to have formal language with clear semantics (i.e., a logic) to express axioms, to be able to get results for entire families of axioms.

First-order logic with *sorts*, one for *profiles* and one for agent *indices*, with these basic ingredients:

- $p \triangleright (i, j)$ — in profile $p$, agents $\ell_i$ and $r_j$ will get matched
- $j \succ^{\mathrm{L}}_{p,i} j'$ — in profile $p$, agent $\ell_i$ prefers $r_j$ to $r_{j'}$ (also for $\mathrm{R}$)
- $top^{\mathrm{L}}_{p,i} = j$ — in profile $p$, agent $\ell_i$ most prefers $r_j$ (also for $\mathrm{R}$)
- $p \sim^{\mathrm{L}}_i p'$ — profiles $p$ and $p'$ are $\ell_i$-variants (also for $\mathrm{R}$)
- $p \rightleftarrows p'$ — swapping sides in profile $p$ yields profile $p'$
- $\forall_{\mathrm{P}}$ / $\exists_{\mathrm{P}}$ and $\forall_{\mathrm{N}}$ / $\exists_{\mathrm{N}}$ — quantifiers for variables of two sorts

Recall that axioms describe properties of mechanisms. So *truth* of a *sentence* $\varphi$ in our logic is defined relative to a *mechanism* $\mu$.

## Example

$$\forall_{\mathrm{P}} p . \forall_{\mathrm{P}} p' . \forall_{\mathrm{N}} i . \forall_{\mathrm{N}} j . \forall_{\mathrm{N}} j' \, . \left[ \left( j \succ^{\mathrm{L}}_{p,i} j' \wedge p \sim^{\mathrm{L}}_{i} p' \right) \rightarrow \neg \left( p \triangleright (i, j') \wedge p' \triangleright (i, j) \right) \right]$$

## Another Example

$$\forall_{\mathrm{P}} p . \forall_{\mathrm{N}} i . \forall_{\mathrm{N}} j \,.\, \big[\, (top^{\mathrm{L}}_{p,i} = j \,\wedge\, top^{\mathrm{R}}_{p,j} = i) \;\rightarrow\; (p \triangleright (i,j)) \,\big]$$

# The Preservation Theorem

Call a mechanism *top-stable* if it always matches all mutual favourites. Call an axiom *universal* if it can be written in the form $\forall \vec{x}. \varphi(\vec{x})$.

Similar to (one direction of) the Łoś-Tarski Theorem in model theory (about preservation of first-order $\forall_1$-formulas in substructures):

**Theorem 1** *Let $\mu^+$ be a top-stable mechanism of dimension $n$ that satisfies a given set $\Phi$ of universal axioms. If $n > 1$, then there also exists a top-stable mechanism $\mu$ of dimension $n - 1$ that satisfies $\Phi$.*

Proof idea: Construct larger profile in which extra agents most prefer each other and are least liked by everybody else.

Corollary: Enough to prove impossibility theorems for smallest $n$!

# Proof Detail

Given an $(n-1)$-dimensional profile, construct an $n$-dimensional one, in which top-stability forces the extra agents $\ell_n$ and $r_n$ to be matched:

$$
\begin{array}{llllll}
\ell_1 & : \Box \succ \cdots \succ \Box \succ r_n & \qquad & r_1 & : \Box \succ \cdots \succ \Box \succ \ell_n \\
\ell_2 & : \Box \succ \cdots \succ \Box \succ r_n & \qquad & r_2 & : \Box \succ \cdots \succ \Box \succ \ell_n \\
\vdots & \quad \vdots \qquad\quad \vdots \quad\;\; \vdots & & \vdots & \quad \vdots \qquad\quad \vdots \quad\;\; \vdots \\
\ell_{n-1} & : \Box \succ \cdots \succ \Box \succ r_n & \qquad & r_{n-1} & : \Box \succ \cdots \succ \Box \succ \ell_n \\
\ell_n & : r_n \succ \cdots \succ r_2 \succ r_1 & \qquad & r_n & : \ell_n \succ \cdots \succ \ell_2 \succ \ell_1
\end{array}
$$

## Counterexample

Preservation Theorem might look trivial. *Doesn't this always hold?*

No: some axioms we can satisfy for large but not for small domains.

Suppose we want to design a mechanism under which at least one agent in each group gets assigned to their most preferred partner:

$$\forall_{\mathrm{P}} p. \exists_{\mathrm{N}} i. \forall_{\mathrm{N}} j. [\, (top^{\mathrm{L}}_{p,i} = j) \;\rightarrow\; (p \triangleright (i,j)) \,] \;\wedge$$

$$\forall_{\mathrm{P}} p. \exists_{\mathrm{N}} j. \forall_{\mathrm{N}} i. [\, (top^{\mathrm{R}}_{p,j} = i) \;\rightarrow\; (p \triangleright (i,j)) \,]$$

This is *not* universal! Mechanism exists for $n = 3$ but not for $n = 2$.

# Proving Impossibility Theorems

Suppose we want to prove an impossibility theorem of this form:

"*for $n \geq k$, no matching mechanism satisfies all the axioms in $\Phi$*"

Our Preservation Theorem permits us to proceed as follows:

- Check all axioms in $\Phi$ indeed are universal axioms.
- Check $\Phi$ includes (or implies) top-stability.
- Express all axioms for special case of $n = k$ in *propositional CNF*.
- Using a *SAT solver*, confirm that this CNF is unsatisfiable.
- Using an *MUS extractor*, find a short proof of unsatisfiability.

For example, *stability* for $n = 3$ can be expressed in CNF like this:

$$\bigwedge_{p \in R_3!^3 \times L_3!^3} \; \bigwedge_{i \in \{1,2,3\}} \; \bigwedge_{j \in \{1,2,3\}} \; \bigwedge_{\substack{i' \text{ s.t. } p \text{ has} \\ \ell_i \succ_{r_j} \ell_{i'}}} \; \bigwedge_{\substack{j' \text{ s.t. } p \text{ has} \\ r_j \succ_{\ell_i} r_{j'}}} \left( \neg x_{p \rhd (i,j')} \vee \neg x_{p \rhd (i',j)} \right)$$

Remark: This is a conjunction of *419,904 clauses* (big, yet manageable).

# Application: A Variant of Roth's Theorem

Recall this classic result:

**Theorem 2 (Roth, 1982)** *For $n \geq 2$, no matching mechanism for incomplete preferences is both stable and two-way strategyproof.*

Remark: In our model (with complete preferences) true only for $n \geq 3$.

We can use our approach to prove this stronger variant:

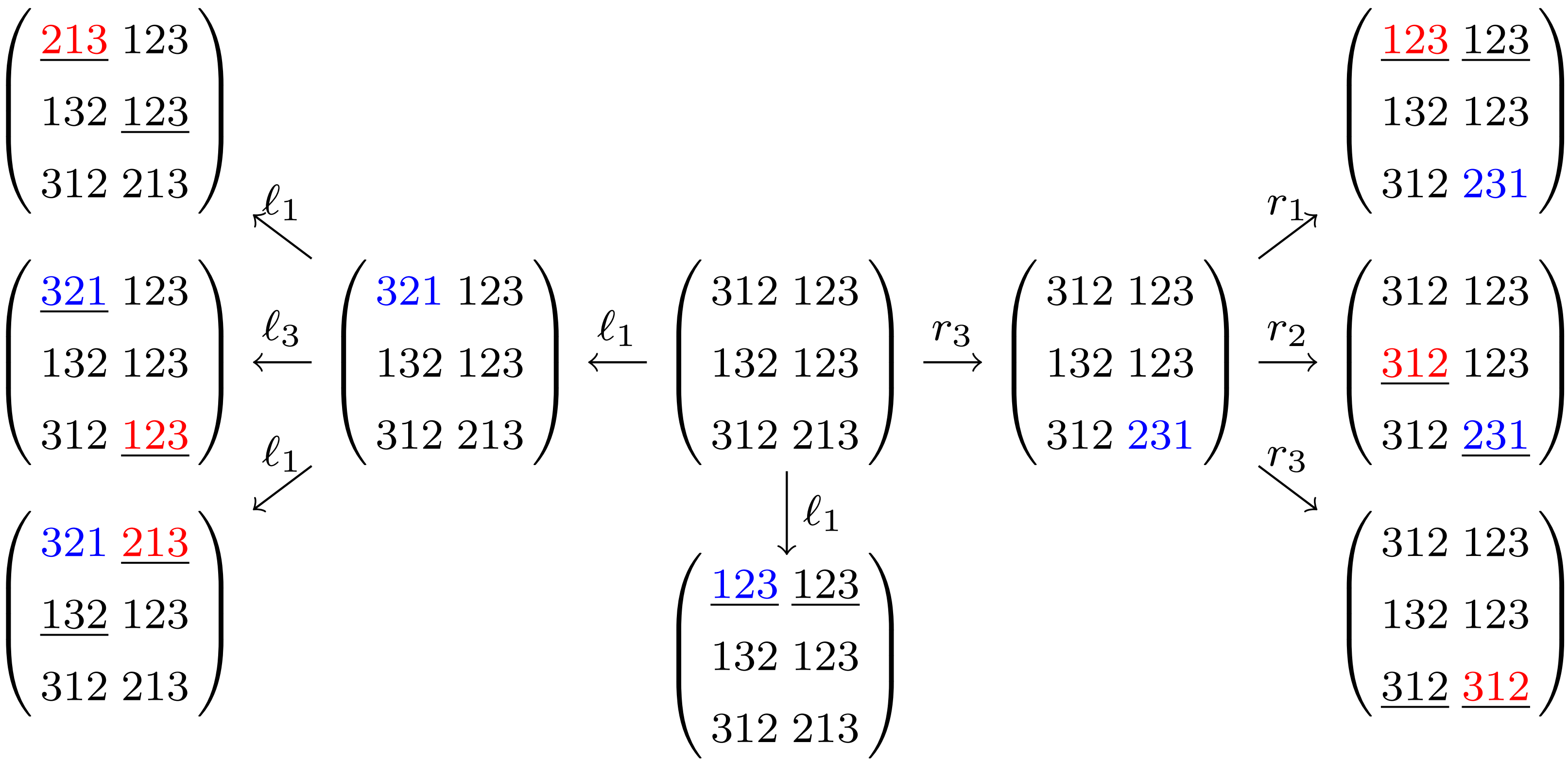
**Theorem 3** *For $n \geq 3$, no matching mechanism is both top-stable and two-way strategyproof (even in our model).*

By the Preservation Theorem, we are done if the claim holds for $n = 3$. SAT solver says it does, and MUS provides human-readable proof ($\hookrightarrow$).

A.E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7:617–628, 1982.

# Proof of Base Case

Found MUS of 23 clauses, referencing 10 profiles. Proof visualisation:




Top-stability forces underlined matches. Colours indicate manipulation opportunities to be ruled out by SP. No matching left for centre profile.

# Application: Stability vs. Gender-Indifference

Call a mechanism *gender-indifferent* if swapping the two sides of the market ("genders") yields the corresponding swap in the outcome:

$$\forall_{\mathrm{P}} p . \forall_{\mathrm{P}} p' . \forall_{\mathrm{N}} i . \forall_{\mathrm{N}} j \,.\, [\, (p \rightleftarrows p') \;\rightarrow\; (\, p \rhd (i,j) \;\rightarrow\; p' \rhd (j,i) \,)\,]$$

Bad news:

**Theorem 4** *For $n \geq 3$, there exists no matching mechanism that is both stable and gender-indifferent.*

Here the MUS extractor finds a particularly simple proof: it identifies a "swap-symmetric" profile for which there exists no admissible outcome (two matchings are ruled out by G-I and the other four by stability).

F. Masarani and S.S. Gokturk. On the Existence of Fair Matching Algorithms. *Theory and Decision*, 26(3):305–322, 1989.

## Last Slide

By the *Preservation Theorem*, for top-stable mechanisms and universal axioms, proving impossibilities can be automated. Specific results:

- Impossible to get *top-stability* and *two-way strategyproofness*.
- Impossible to get *stability* and *gender-indifference*.

[ `tinyurl.com/satmatching` ]