

AUTOMATED JUSTIFICATION OF COLLECTIVE DECISIONS

VIA CONSTRAINT SOLVING

ARTHUR BOIXEL

ILLC, University of Amsterdam

A.BOIXEL@UVA.NL

ULLE ENDRISS

ILLC, University of Amsterdam

U.ENDRISS@UVA.NL

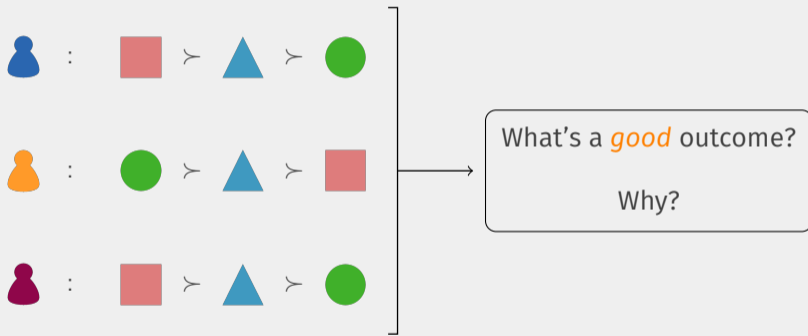


Institute for Logic, Language and Computation



UNIVERSITY OF AMSTERDAM

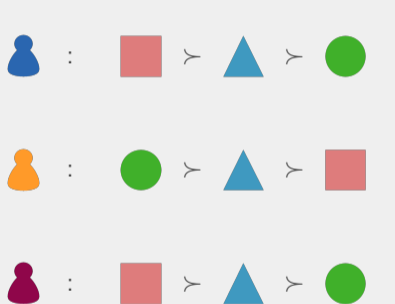
AAMAS - MAY 2020



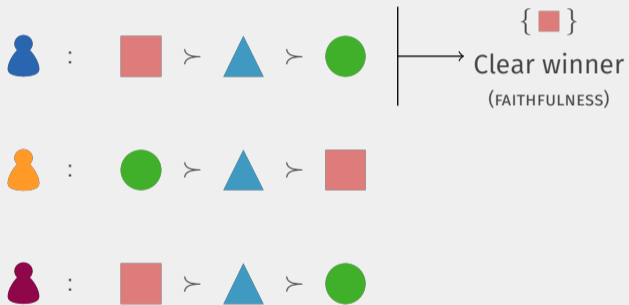
A JUSTIFICATION



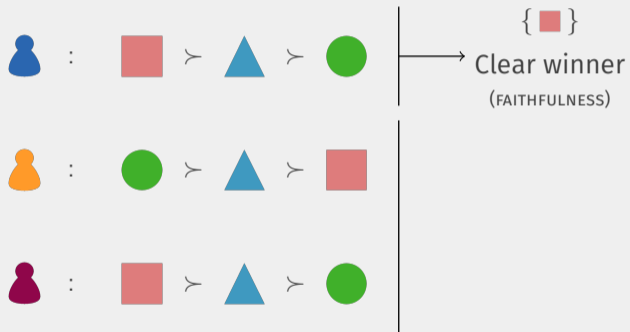
A JUSTIFICATION



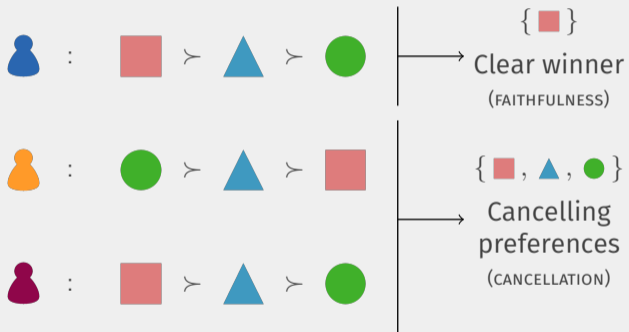
A JUSTIFICATION



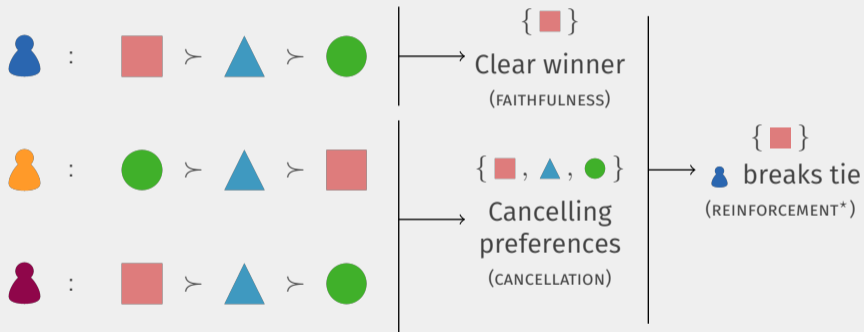
A JUSTIFICATION



A JUSTIFICATION



A JUSTIFICATION





Deliberation Support



Deliberation Support



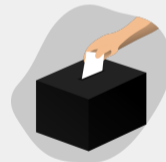
Building Confidence
in Election Results



Deliberation Support



Building Confidence
in Election Results



Justification Generation
as Voting

Agents in	N^*	=	{  ,  ,  }
Alternatives in	X	=	{  ,  ,  }
Preferences in	$\mathcal{L}(X)$	=	{  \succ  \succ  ,  \succ  \succ  , ... }

Voting rules F defined on *profiles* for subelectorates $N \subseteq N^*$:

$$F : \bigcup_{N \in 2^{N^*} \setminus \{\emptyset\}} \mathcal{L}(X)^N \rightarrow 2^X \setminus \{\emptyset\}$$

Interesting voting rules might satisfy some *axioms*:

- *Pareto Principle*: $y \notin F(\succsim_N)$ if $\{i : x \succ_i y\} = N$, for some $x \in X$
 - *Cancellation*: $F(\succsim_N) = X$ if $|\{i : x \succ_i y\}| = \frac{|N|}{2}$ for all $x, y \in X$
 - *Reinforcement**: $F(\succsim_{N \uplus N'}) = F(\succsim_N) \cap F(\succsim_{N'})$, unless empty
-

Interesting voting rules might satisfy some *axioms*:

- *Pareto Principle*: $y \notin F(\succsim_N)$ if $\{i : x \succ_i y\} = N$, for some $x \in X$
- *Cancellation*: $F(\succsim_N) = X$ if $|\{i : x \succ_i y\}| = \frac{|N|}{2}$ for all $x, y \in X$
- *Reinforcement**: $F(\succsim_{N \uplus N'}) = F(\succsim_N) \cap F(\succsim_{N'})$, unless empty

Interpreting Axioms:

- The *interpretation* $\mathbb{I}(A)$ of an axiom A : $\mathbb{I}(A) \subseteq \bigcup_{N \in 2^{N^*} \setminus \{\emptyset\}} \mathcal{L}(X)^N \rightarrow 2^X \setminus \{\emptyset\}$
- An *instance* A' of an axiom A is what you think it is, with $\mathbb{I}(A) = \bigcap_{A' \in \text{Inst}(A)} \mathbb{I}(A')$
 - *Pareto instance*: $\bullet \notin F(\blacksquare \succ_1 \blacktriangle \succ_1 \bullet, \blacktriangle \succ_2 \bullet \succ_2 \blacksquare)$
 - *Cancellation instance*: $F(\blacksquare \succ_1 \blacktriangle \succ_1 \bullet, \bullet \succ_2 \blacktriangle \succ_2 \blacksquare) = \{\blacksquare, \blacktriangle, \bullet\}$

INPUT:

A voting scenario $\langle \succ_{N^*}, X^*, \mathbb{A} \rangle$ with:

- Target Profile \succ_{N^*}
- Target Outcome X^*
- Corpus of Axioms \mathbb{A}

INPUT:

A voting scenario $\langle \succ_{N^*}, X^*, \mathbb{A} \rangle$ with:

- Target Profile \succ_{N^*}
- Target Outcome X^*
- Corpus of Axioms \mathbb{A}

OUTPUT:

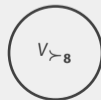
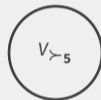
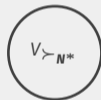
A justification $\langle \mathcal{A}^N, \mathcal{A}^E \rangle$ (**normative basis** \mathcal{A}^N and **explanation** \mathcal{A}^E)

for why $F(\succ_{N^*}) = X^*$ such that:

- (i) **Adequacy:** $\mathcal{A}^N \subseteq \mathbb{A}$
- (ii) **Relevance:** \mathcal{A}^E only contains instances of axioms in \mathcal{A}^N
- (iii) **Explanatoriness:** $F(\succ_{N^*}) = X^*$ for all rules $F \in \mathbb{I}(\mathcal{A}^E)$, and this is not the case for any $\mathcal{A} \subset \mathcal{A}^E$
- (iv) **Nontriviality:** $\mathbb{I}(\mathcal{A}^N) \neq \emptyset$, so *some* rule satisfies all the axioms in \mathcal{A}^N

JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

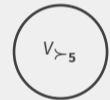
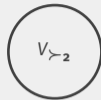
○ Profile



JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

○ Profile

● FAITHFULNESS

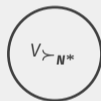


JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

○ Profile

● FAITHFULNESS

● CANCELLATION



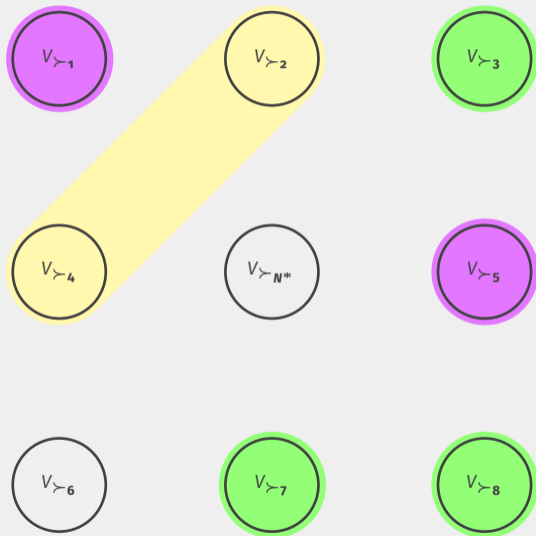
JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

○ Profile

● FAITHFULNESS

● CANCELLATION

● ANONYMITY



JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

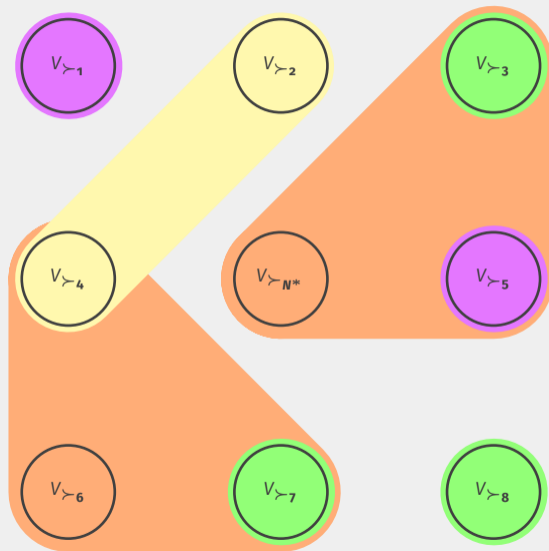
○ Profile

● FAITHFULNESS

● CANCELLATION

● ANONYMITY

● REINFORCEMENT*



JUSTIFICATION PROBLEMS AS CONSTRAINT SATISFACTION PROBLEMS (CSP)

○ Profile

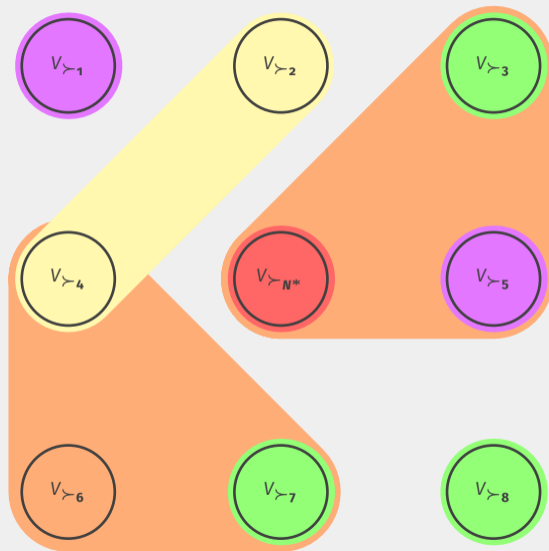
● FAITHFULNESS

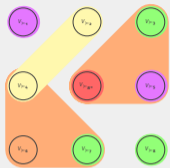
● CANCELLATION

● ANONYMITY

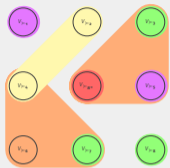
● REINFORCEMENT*

● GOAL: $F(\gamma_{N^*}) \neq X^*$

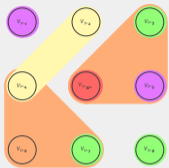




(1) Generate
Constraint Network



(1) Generate
Constraint Network \longrightarrow (2) Is it
Satisfiable?



(1) Generate
Constraint Network



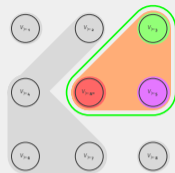
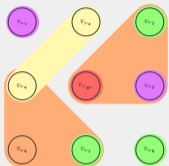
(2) Is it
Satisfiable?

YES



No Justification

AUTOMATED SEARCH FOR JUSTIFICATIONS



(1) Generate
Constraint Network

→ (2) Is it
Satisfiable?

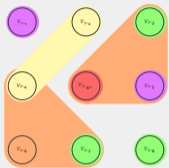
→ NO → (2a) Find MUS with
Goal Constraint

YES

X

No Justification

AUTOMATED SEARCH FOR JUSTIFICATIONS



(1) Generate
Constraint Network

(2) Is it
Satisfiable?

NO

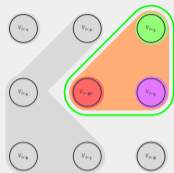
(2a) Find MUS with
Goal Constraint

(2b) Is \mathcal{A}^N
Nontrivial?

YES

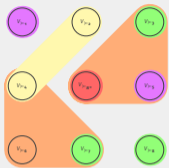
X

No Justification



$$\langle \mathcal{A}^E \approx \text{MUS} \setminus \{\text{GOAL}\}, \\ \mathcal{A}^N = \{\bullet, \bullet, \bullet\} \rangle$$

AUTOMATED SEARCH FOR JUSTIFICATIONS



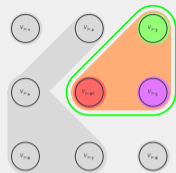
(1) Generate
Constraint Network

(2) Is it
Satisfiable?

YES



No Justification



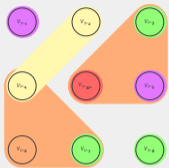
(2a) Find MUS with
Goal Constraint

$$\langle \mathcal{A}^E \approx \text{MUS} \setminus \{\text{GOAL}\}, \mathcal{A}^N = \{\bullet, \bullet, \bullet\} \rangle$$

(2b) Is \mathcal{A}^N
Nontrivial?

NO

AUTOMATED SEARCH FOR JUSTIFICATIONS



(1) Generate
Constraint Network

(2) Is it
Satisfiable?

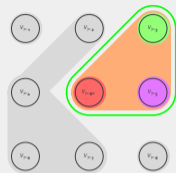
YES



No Justification

NO

(2a) Find MUS with
Goal Constraint



$$\langle \mathcal{A}^E \approx \text{MUS} \setminus \{\text{GOAL}\}, \\ \mathcal{A}^N = \{\bullet, \bullet, \bullet\} \rangle$$

(2b) Is \mathcal{A}^N
Nontrivial?

NO













YES



Justification Found

Corpus of Axioms: $\mathbb{A} = \{\text{FAITHFULNESS, CANCELLATION, REINFORCEMENT}^*\}$

Target Profile: $\gamma_{N^*} =$

	:		γ_1		γ_1	
	:		γ_2		γ_2	
	:		γ_3		γ_3	

Target Outcome: $X^* = \left\{ \left[\text{red square} \right] \right\}$

Goal Constraint: $C_G = V_{p(\gamma_{N^*})} \neq \left\{ \left[\text{red square} \right] \right\}$

TOY EXAMPLE (2)

Target Profile: $\gamma_{N^*} =$

👤 :	■ γ_1	▲ γ_1	●
👤 :	● γ_2	▲ γ_2	■
👤 :	■ γ_3	▲ γ_3	●

MUS: $C_1 : (\text{FAI}) \quad V_{\langle \gamma_{\text{👤}} \rangle} = \left\{ \text{■} \right\}$

$C_2 : (\text{CAN}) \quad V_{\langle \gamma_{\text{👤}}, \gamma_{\text{👤}} \rangle} = \left\{ \text{■}, \text{▲}, \text{●} \right\}$

$C_3 : (\text{REI}) \quad \left[V_{\langle \gamma_{\text{👤}} \rangle} \cap V_{\langle \gamma_{\text{👤}}, \gamma_{\text{👤}} \rangle} \neq \emptyset \right] \rightarrow \left[V_{P(\gamma_{N^*})} = V_{\langle \gamma_{\text{👤}} \rangle} \cap V_{\langle \gamma_{\text{👤}}, \gamma_{\text{👤}} \rangle} \right]$

$C_G : (\text{GC}) \quad V_{P(\gamma_{N^*})} \neq \left\{ \text{■} \right\}$

Target Profile: $\gamma_{N^*} =$

Blue person icon :	Red square	γ_1	Blue triangle	γ_1	Green circle
Orange person icon :	Green circle	γ_2	Blue triangle	γ_2	Red square
Purple person icon :	Red square	γ_3	Blue triangle	γ_3	Green circle

MUS: $c_1 : (\text{FAI}) \quad V_{\langle \gamma_{\text{Blue}} \rangle} = \{ \text{Red square} \}$

$c_2 : (\text{CAN}) \quad V_{\langle \gamma_{\text{Orange}}, \gamma_{\text{Purple}} \rangle} = \{ \text{Red square}, \text{Blue triangle}, \text{Green circle} \}$

$c_3 : (\text{REI}) \quad \left[V_{\langle \gamma_{\text{Blue}} \rangle} \cap V_{\langle \gamma_{\text{Orange}}, \gamma_{\text{Purple}} \rangle} \neq \emptyset \right] \rightarrow \left[V_{p(\gamma_{N^*})} = V_{\langle \gamma_{\text{Blue}} \rangle} \cap V_{\langle \gamma_{\text{Orange}}, \gamma_{\text{Purple}} \rangle} \right]$

$c_G : (\text{GC}) \quad V_{p(\gamma_{N^*})} \neq \{ \text{Red square} \}$

JUSTIFICATION: $\mathcal{A}^E \approx \{c_1, c_2, c_3\}$

$\mathcal{A}^N = \{ \text{FAITHFULNESS, CANCELLATION, REINFORCEMENT}^* \}$

Contribution

- Justification = **Normative Basis** (*axioms*) + **Explanation** (*instances*)
- Generation of Justifications via **Constraint Solving** and **MUS** extraction
- Applications: **Confidence Building** | **Deliberation Support** | **Voting**

Future Work

- How *hard* is it to find a justification?
- How to *best present* a justification?
- What makes for a *good* justification?