IMPERIAL

Optimal Correction Sets for Argumentative Causal Discovery

Fabrizio Russo[1], Giuseppe Mazzotta[2], Carmine Dodaro[2], Francesco Ricca[2] and Francesca Toni[1]

[1]Imperial College London, [2]University of Calabria

LINDA'26

24/07/2026

Outline

1. Causal Discovery
2. Assumption-based Argumentation (ABA)
3. Causal ABA & ASP
4. MUS & MCS
5. Optimal MCS for Causal ASP
6. Experiments
7. Conclusion & Future work

Constraint-based Causal Discovery (Spirtes et al. 1993)

observed data

E	R	O	I
BSc	W	Fin	41
PhD	A	Ed	68
Hi	L	Of	29
BSc	B	Bu	83
MSc	A	Of	52
⋮			

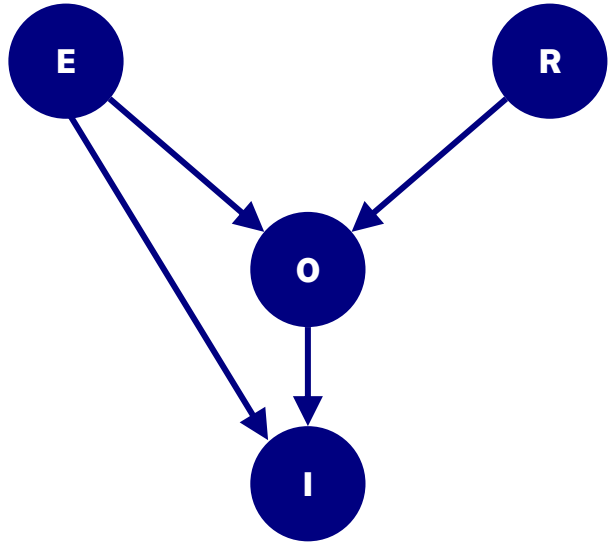
example CI tests

$E \not\perp\!\!\!\perp O$

$E \perp\!\!\!\perp R \mid O$

$R \perp\!\!\!\perp I \mid \{E, O\}$

candidate causal graph



Inconsistencies in CI tests

f_1 $E \not\perp\!\!\!\perp O$

f_2 $E \perp\!\!\!\perp O \mid I$

f_3 $E \not\perp\!\!\!\perp I$

f_4 $O \perp\!\!\!\perp I$

f_1: E and O are dependent → an active E–O path exists

f_2: conditioning on I removes it → I blocks that path

I must be a non-collider on an active E–O route

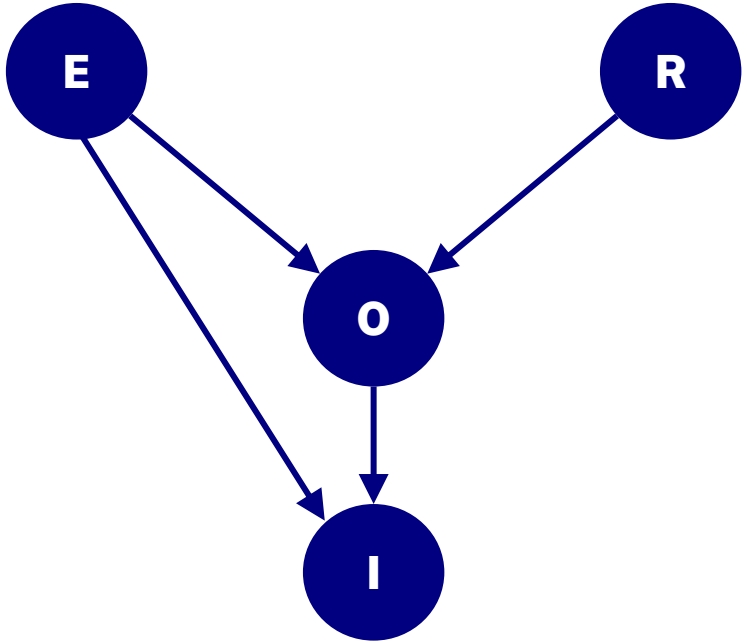
⇒ O and I must be dependent

derived from f_1–f_3: $O \not\perp\!\!\!\perp I$ but f_4 asserts: $O \perp\!\!\!\perp I$

Four individually plausible tests, taken together, describe no DAG

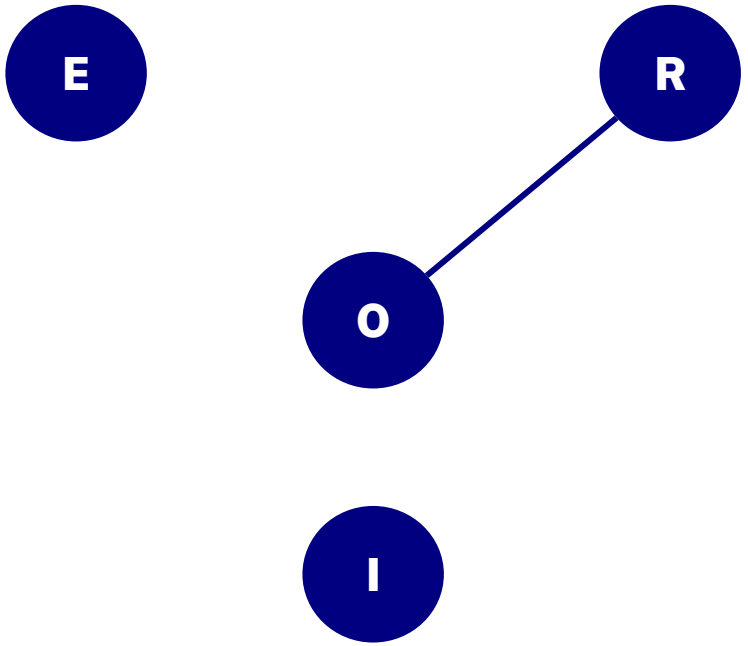
Inconsistency between CI Tests and Graphs

Ground truth



14 tests
5 wrong
jointly inconsistent

Majority-PC output



PC / Majority-PC (Colombo & Maathius, 2014): resolves clashes by majority vote, then orients — the discarded tests leave no trace

Majority-PC returns a graph but never says which tests it stands behind

Causal ABA

(Russo, Rapberger & Toni, 2024)

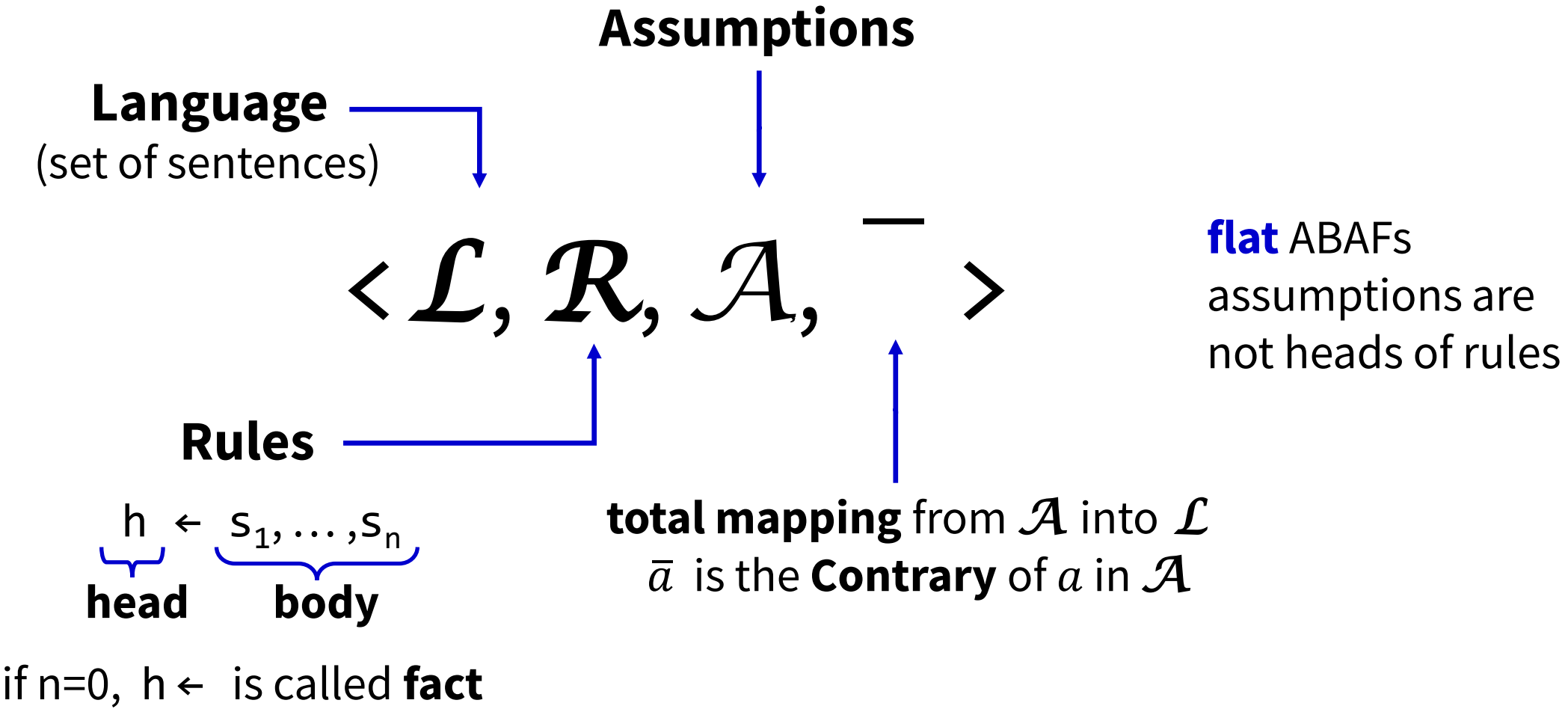
CI tests

⇔

output causal graph (stable extensions)

ABA Framework

(Bondarenko et al. 1997)



flat ABAFs
assumptions are
not heads of rules

ABA Framework

an example – Barber's paradox

person(a) ← person(b) ← person(f) ←
beardless(b) ←

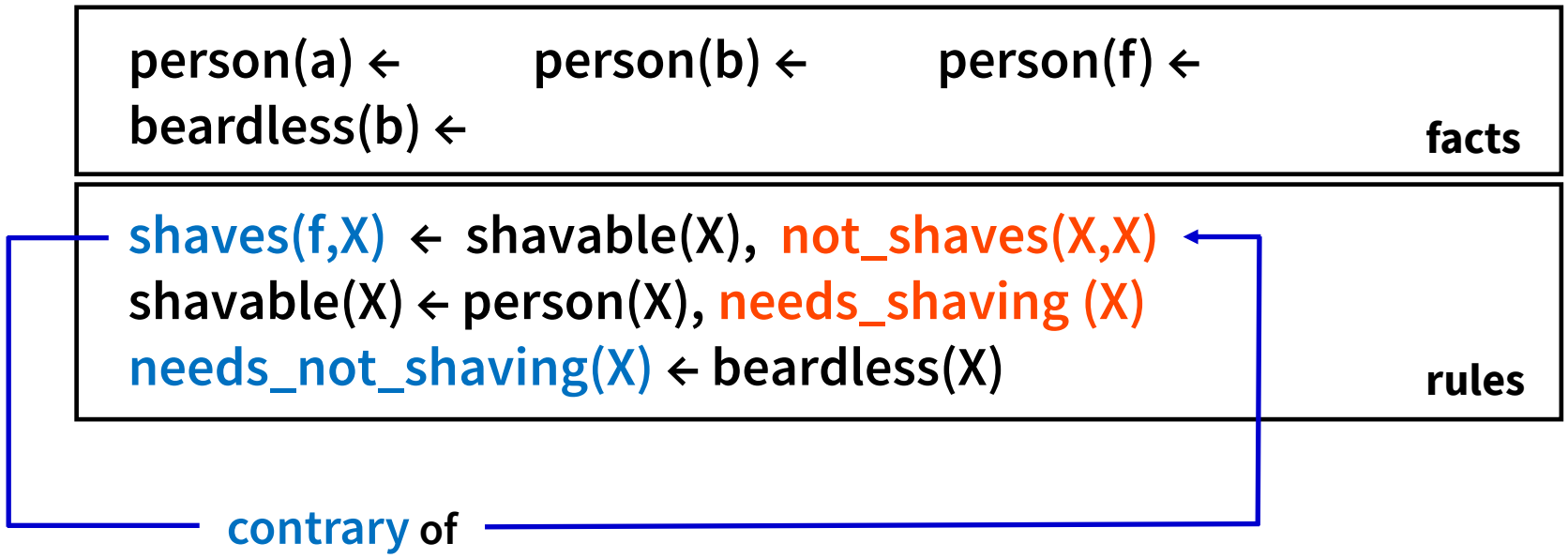
facts

shaves(f,X) ← shavable(X), not_shaves(X,X)
shavable(X) ← person(X), needs_shaving (X)
needs_not_shaving(X) ← beardless(X)

rules

ABA Framework

an example – Barber's paradox



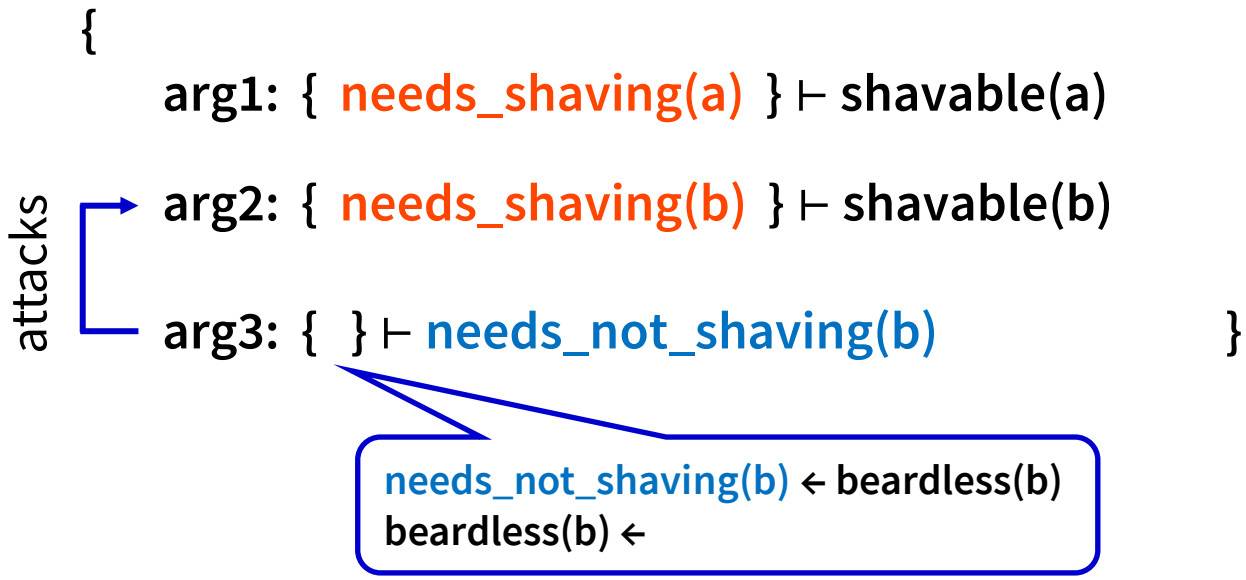
assumptions render rules defeasible:
rules are applicable unless contraries of assumptions can be derived

ABA Framework

Semantics

"**acceptable**" extensions:
sets of **arguments** able to "defend" themselves from "attacks"
as determined by the chosen **semantics σ**

- **Arguments** are **deductions** of claims using **rules** and supported by **assumptions**
- **Attacks** are directed at the assumptions in the support of arguments



ABA Framework

Semantics

Args: set of all arguments of an ABA framework

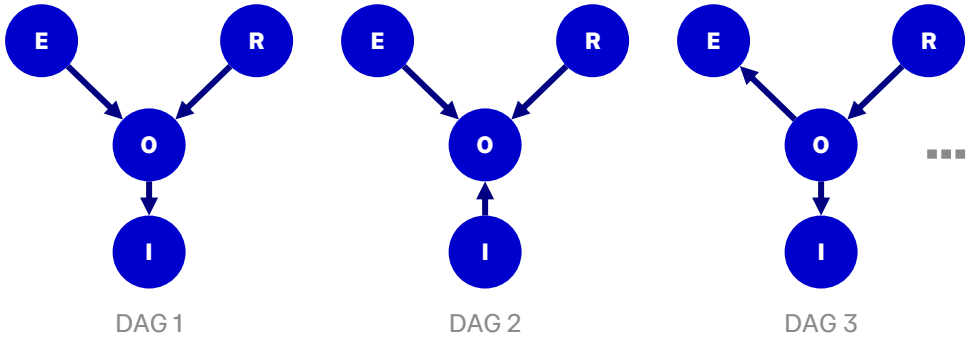
$A, B \subseteq Args$

- A **attacks** B : an argument in A attacks an argument in B
- A is **conflict-free** : A does not attack itself
- A **defends** B : for all $C \subseteq Args$ that attack B, A attacks C

semantics σ	conflict-free extension $A \subseteq Args$
Admissible	it defends itself
Complete	admissible and contains every argument it defends
Grounded	$\subseteq$-minimally complete extension (unique fixed-point)
Preferred	$\subseteq$-maximally admissible extension
Stable	attacks every argument outside of itself

Causal ABA Fragment

stable extensions of the same enforced facts



Assumptions

$$\{arr_{xy} \mid x, y \in \mathbf{V}, x \neq y\} \cup \{noe_{xy} \mid x, y \in \mathbf{V}, x \neq y\}$$

$$\{(x \perp\!\!\!\perp y \mid \mathbf{Z}) \mid \mathbf{Z} \subseteq \mathbf{V}, x, y \in \mathbf{V} \setminus \mathbf{Z}, x \neq y\}.$$

$$\{\mathbf{bp_{p|Z}} \mid \mathbf{p} \text{ is a x-y-path}, \mathbf{Z} \subseteq \mathbf{V} \setminus \{x, y\}\}$$

Rules

- $\overline{a} \leftarrow b,\ a \neq b,\ a, b \in \{arr_{xy}, arr_{yx}, noe_{xy}\},\ x, y \in \mathbf{V} \implies$ **Can only choose one assumption**
- $\overline{arr_{x_i x_{i+1}}} \leftarrow arr_{x_1 x_2}, \ldots, arr_{x_{k-1} x_k}$ for each sequence $x_1 \ldots x_k$ with $x_1 = x_k$, for each $1 \leq i < k \implies$ **Cycles are not allowed**

$dpath_{xy} \leftarrow arr_{xy}$ $\quad$ $dpath_{xz} \leftarrow dpath_{xy}, arr_{yz}$

$e_{xy} \leftarrow arr_{xy}$ $\quad$ $e_{xy} \leftarrow arr_{yx}$ $\quad$ $\overline{noe_{xy}} \leftarrow e_{xy}$

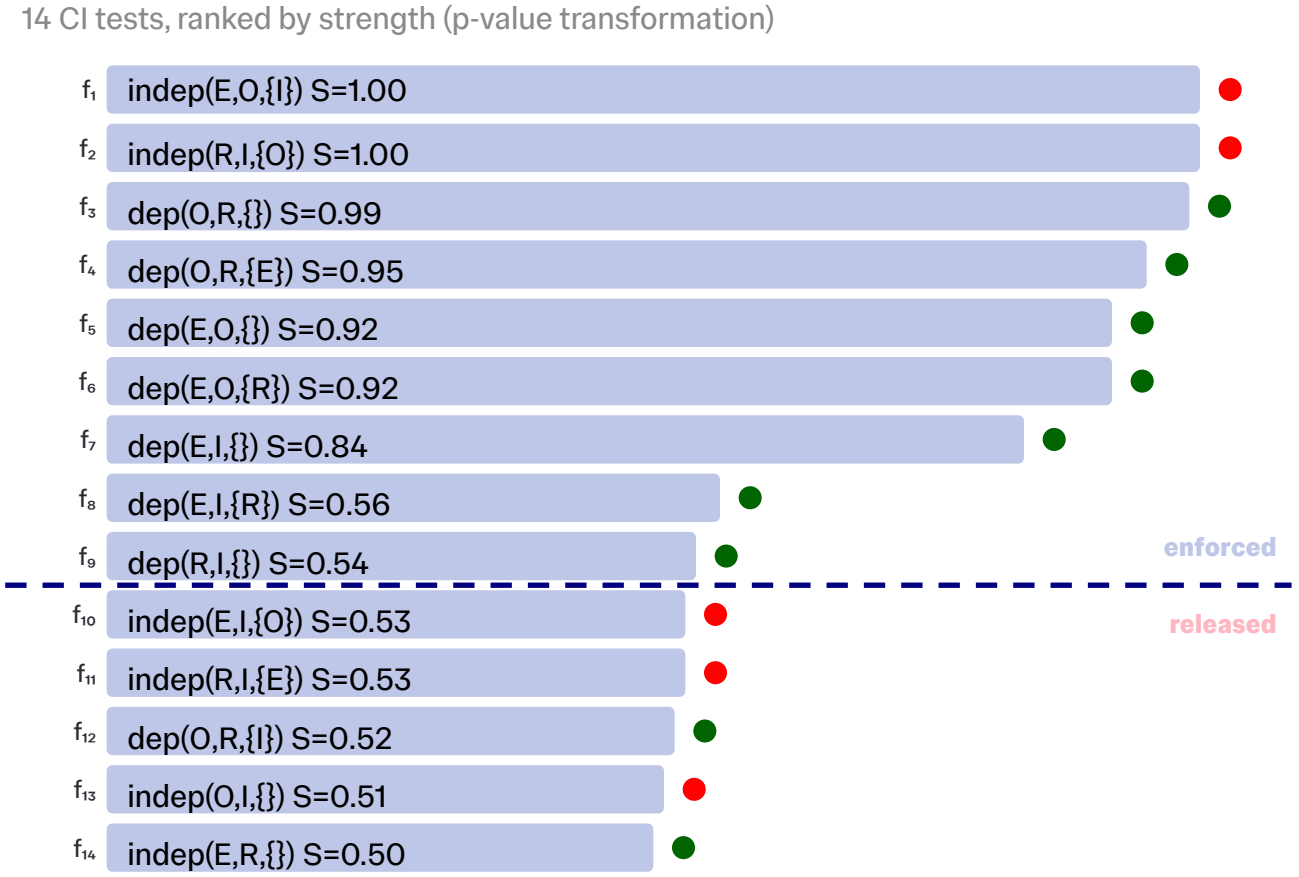
$\overline{x \perp\!\!\!\perp y \mid \mathbf{Z}} \leftarrow p_{\mathbf{Z}}$

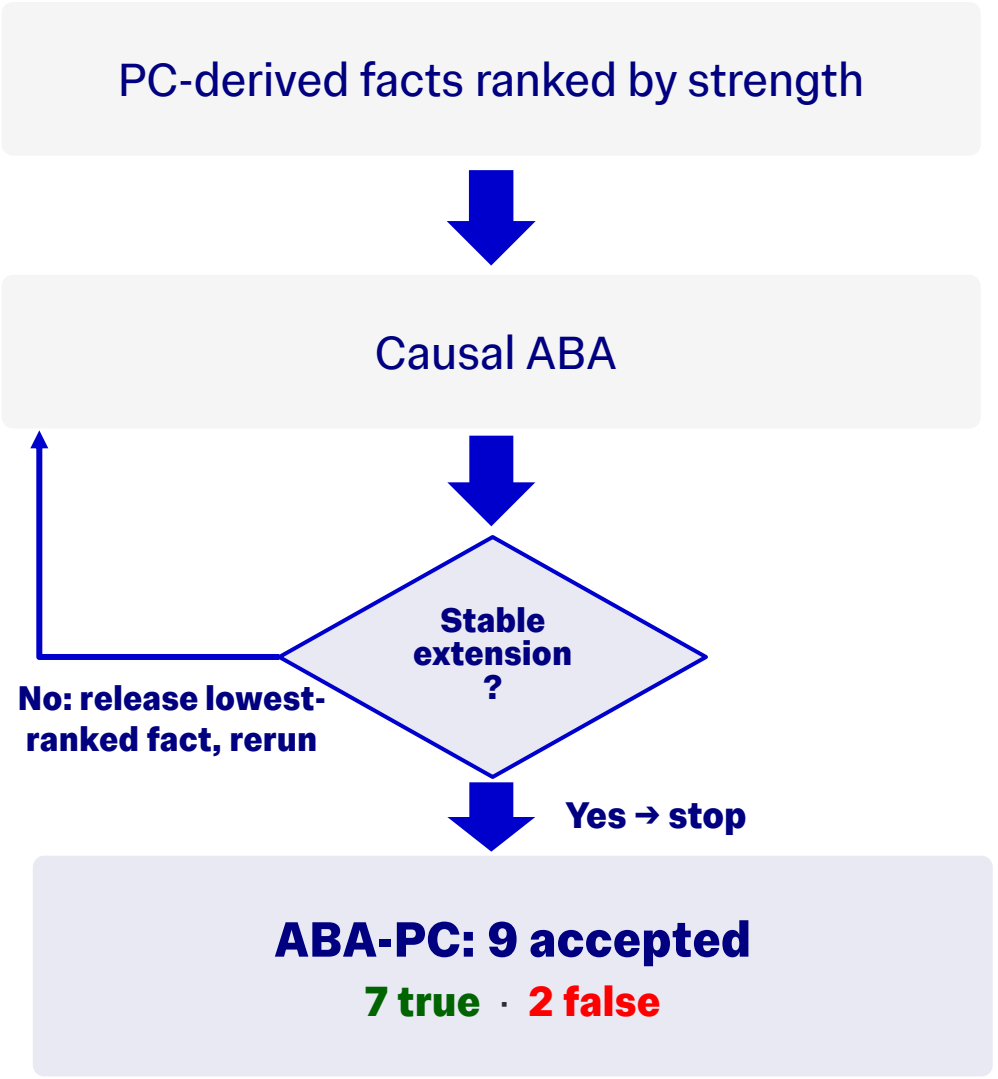
d-Separation (Pearl, 2009)

- $x \perp\!\!\!\perp y \mid \mathbf{Z} \leftarrow bp_{p_1|\mathbf{Z}}, \ldots, bp_{p_k|\mathbf{Z}}$ where $p_1, \ldots, p_k$ denote all paths between x and y;
- $ap_{p|\mathbf{Z}} \leftarrow p$ for each $\mathbf{Z}$-active p

Enforced facts bind every returned graph — and there is usually more than one, or none

ABA-PC (Russo, Rapberger & Toni, 2024)





ABA-PC releases a ranked suffix: transparent, but not the best repair

From Causal ABA to Answer Set Programming

Stable Model for LP ⟺ Stable Extension for ABA (Rapberger, Ulbricht and Toni, 2024)

Base program P

```
a :- o1, not b.
b :- o2, not c.
c :- o3, not a.
:- o4, not d.
```

objectives $O = \{ o_1, o_2, o_3, o_4 \}$ — the claims we may choose to enforce

Enforcing a selection (Alviano et al. 2023)

```
enforce(P,S) = P ∪
      { :- not o.  |  o ∈ S }
```

S is satisfiable iff enforce(P,S) has a stable model.

enforce(P, O) ⇒ no stable model

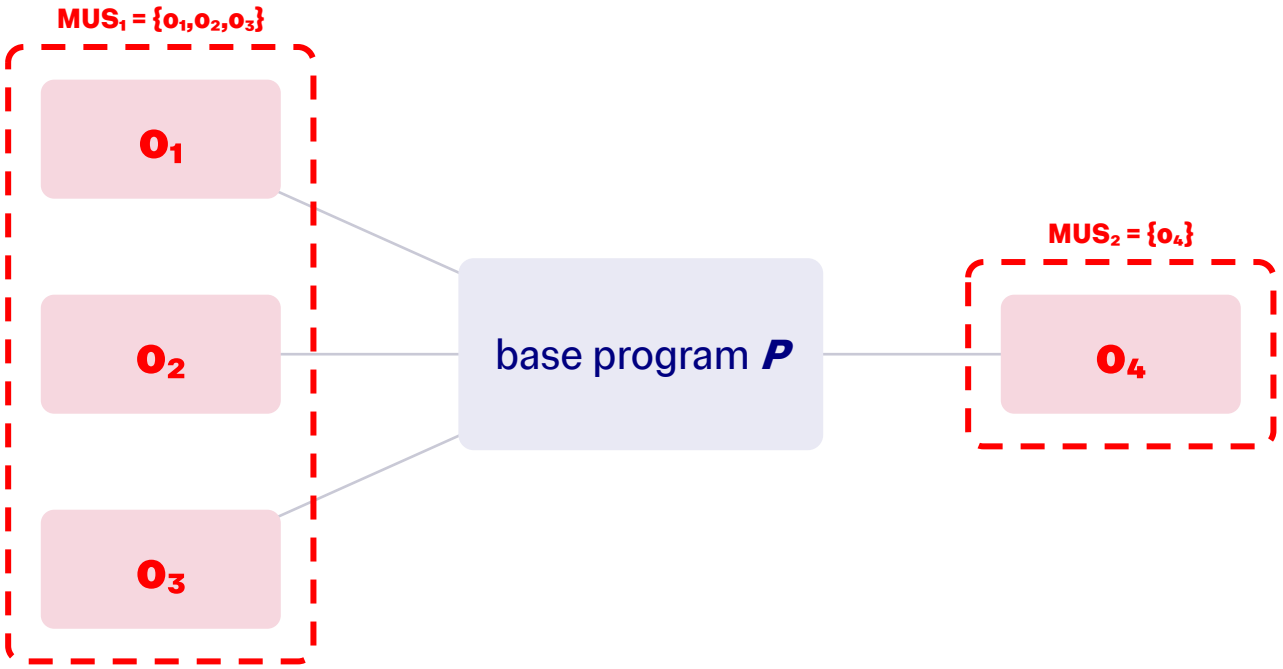
o_1, o_2, o_3 together close an odd negative loop

o_4 requires d, which no rule derives

Objectives are the claims we can switch on; enforcing all of them yields no stable model

Minimal Unsatisfiable Set

(Alviano et al, 2023)



Unsatisfiable set

enforcing all its members makes P incoherent

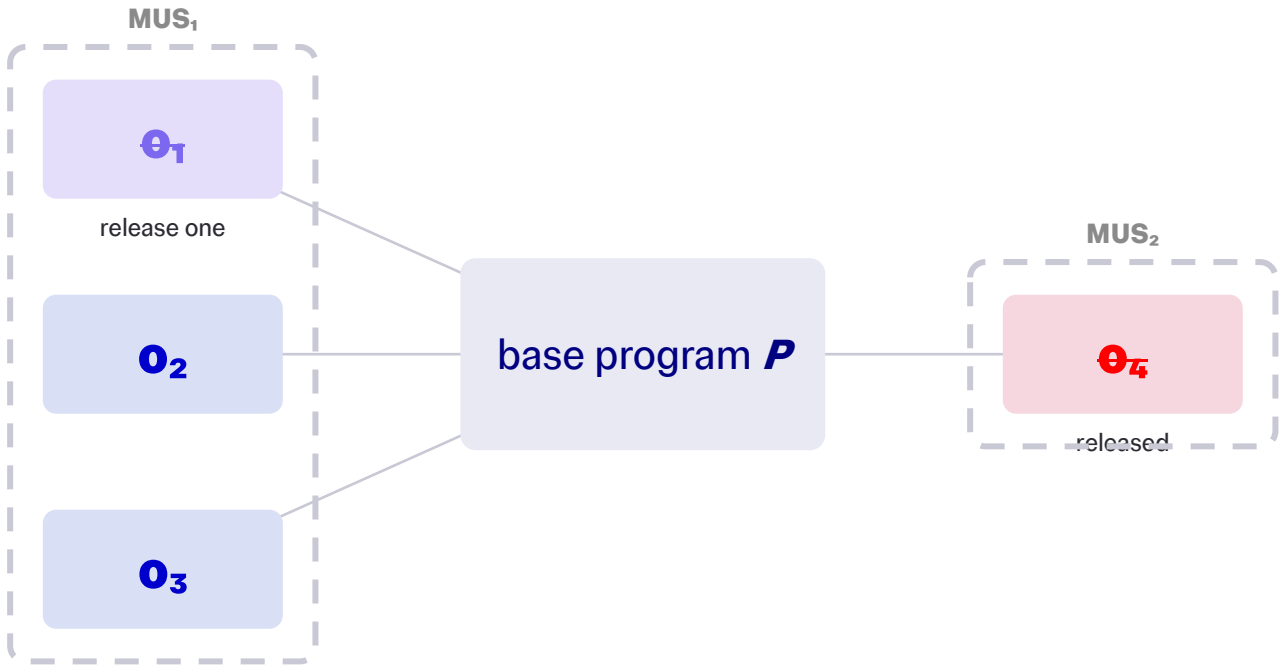
MUS

unsatisfiable, but every strict subset is satisfiable

A MUS is a minimal conflict core — minimal by inclusion, not by size

Minimal Correcting Set

(Alviano et al, 2023)



Correction set

remove it and coherence returns

MCS

no strict subset is still a correction set

$\{o_1,o_4\} \cdot \{o_2,o_4\} \cdot \{o_3,o_4\}$

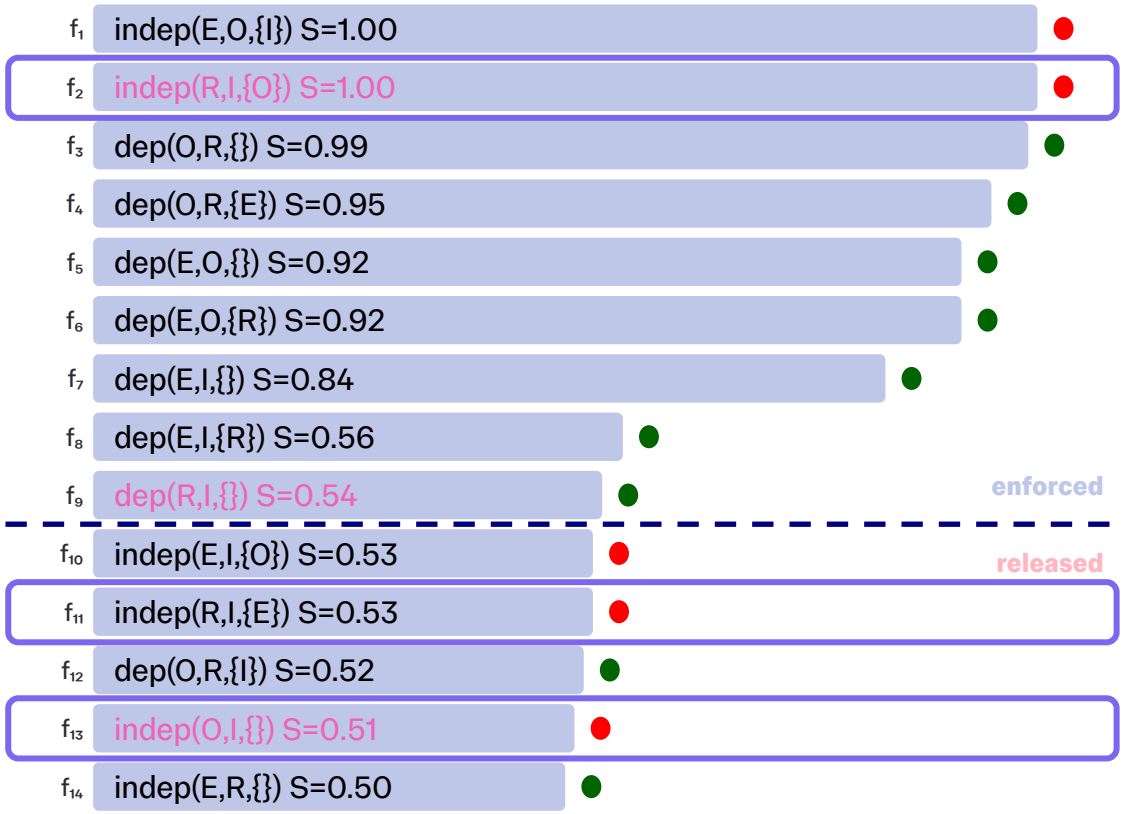
MCSes are the minimal hitting sets of the MUSes.

each repair must drop o_4 (a singleton MUS) and one of $\{o_1,o_2,o_3\}$

Every MCS hits every MUS — MCSes are the minimal hitting sets of the MUSes

MUS and MCS for ABA-PC

14 CI tests, ranked by strength (p-value transformation)



103 MUSes **25** MCSes

a MUS why it is inconsistent

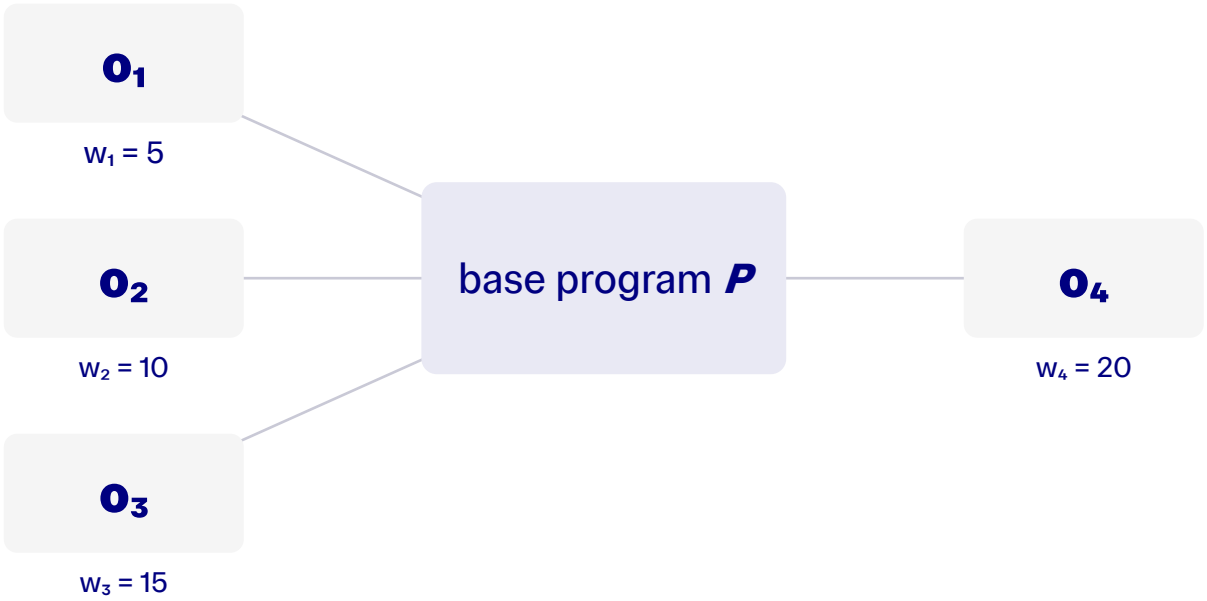
$\{ R \perp\!\!\!\perp I \mid O,\ R \not\perp\!\!\!\perp I,\ O \perp\!\!\!\perp I \}$

an MCS one minimal repair

$\{ R \perp\!\!\!\perp I \mid E,\ O \perp\!\!\!\perp I,\ E \perp\!\!\!\perp O \mid I \}$

A repair may release a strong fact and keep a weak one

Optimal MCS



$C(\{o_1, o_4\}) = 25$ ✓ optimal

$C(\{o_2, o_4\}) = 30$

$C(\{o_3, o_4\}) = 35$

Optimal MCS = a minimal correction set of minimum total release cost

Optimal MCS as the dual of an Optimal Stable Model

the same program P, adorned

```
a :- o1, not b.     b :- o2, not c.
c :- o3, not a.      :- o4, not d.
```

\+ keep or release each objective

```
{ o1; o2; o3; o4 }.
```

\+ pay w_o whenever o is released

```
:~ not o1. [5]       :~ not o2. [10]
:~ not o3. [15]      :~ not o4. [20]
```

Stable model M = {o_2,o_3}; released O \ M = {o_1,o_4}, cost 25.

An optimal stable model minimises exactly the weight of what it releases.

Positive weights make that release set subset-minimal, so it is an optimal MCS.

```
% objectives = the 14 CI facts
{ o(indep(r,i,(e,))) ; ... }.

% weight from the test strength
:~ not o(F), w(F,W). [W]
```

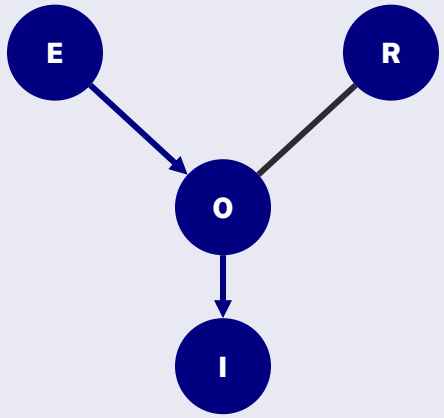
Optimal MCS returned by solver — no MUS or MCS enumeration needed

OptABA-PC

Causal ABA + release an optimal MCS (instead of ABA-PC's weakest-ranked suffix)

ABA-PC

1 of 4 DAGs



9 accepted

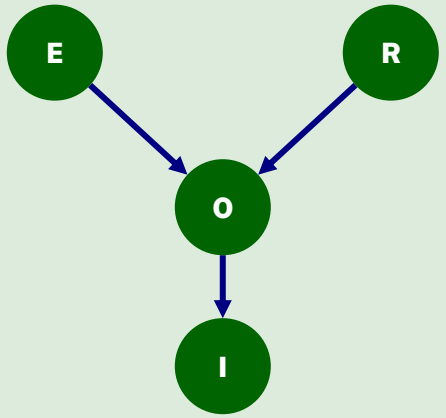
7 true · 2 false

4 compatible DAGs

1 correct orientation

OptABA-PC

unique DAG



10 accepted

9 true · 1 false

4 released facts — all wrong

1 DAG · 3 correct orientations

Global optimisation retains more, releases only wrong facts, and returns one DAG

Results

Dataset (nodes)	True constraints	Fact-classification F1	# CPDAGs	Worst-case SHD	Runtime
Cancer (5)	▲ +10.7%	▲ +5.2%	▼ −88.1%	▼ −66.7%	▲ ×1.65
Survey (5)	▲ +16.0%	▼ −2.3%	▼ −99.4%	▼ −51.7%	▲ ×88.69
Asia (8)	▲ +29.2%	▲ +14.4%	▼ −97.7%	▼ −65.2%	▲ ×5.37
ER (5)	▲ +486.7%	▲ +263.3%	▼ −98.6%	▼ −18.8%	▲ ×5.50
ER (8)	▲ +1455.8%	▲ +655.7%	▼ −99.5%	▼ −31.7%	▲ ×1.69

▲ / ▼ = direction of change, OptABA-PC vs ABA-PC. ER-8 hit the 300 s cap in all runs, figures are best-effort.

Exact repair usually returns fewer and better causal graphs

Inconsistency becomes an inspectable, optimisable object

CI confidence ⟷ release costs

Problem: Does not scale

Future Work:
Optimisation
Cost functions
Explanations

IMPERIAL

Thanks for your attention

scan for code & experiments

github.com/briziorusso/ArgCausalDisco/tree/MCS

fabrizio@imperial.ac.uk