CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY
Funded by the European Union
European Research Council
Established by the European Commission

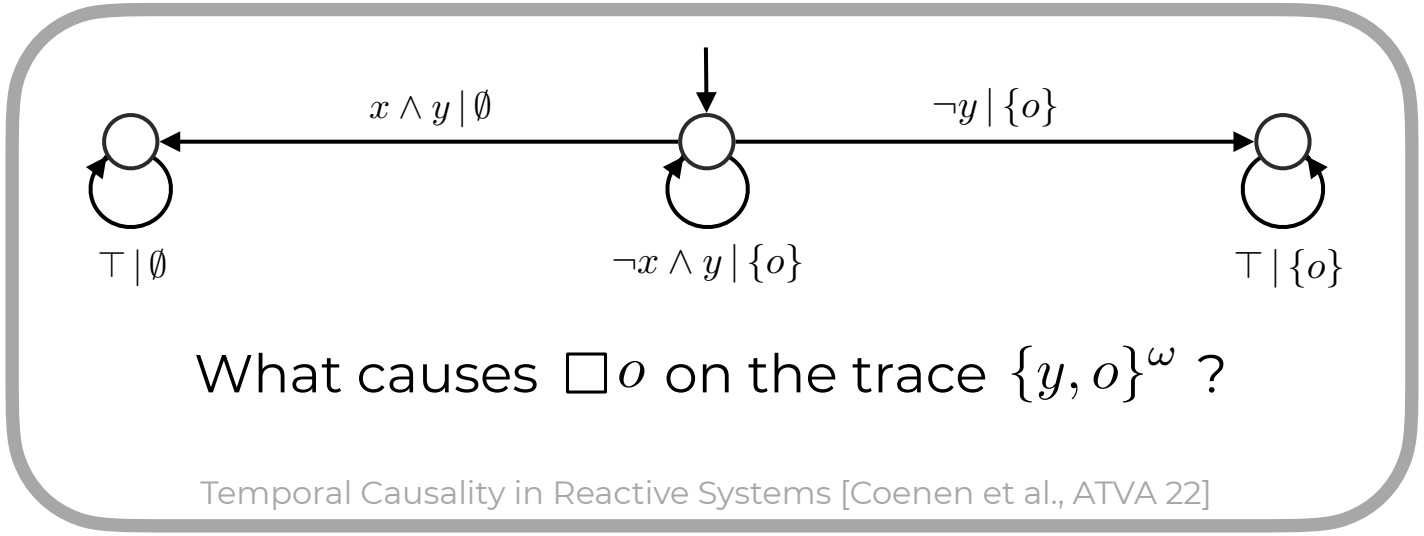
Synthesis of Temporal Causality

Bernd Finkbeiner, Hadar Frenkel, Niklas Metzger, and Julian Siber

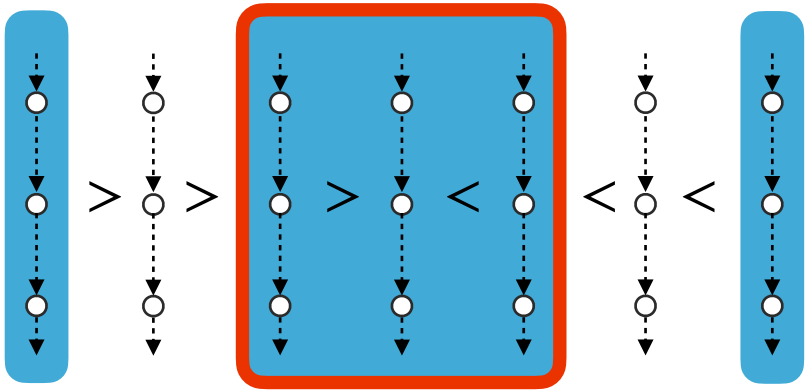
CISPA Helmholtz Center for Information Security

CAV 2024 | 25 July 2024

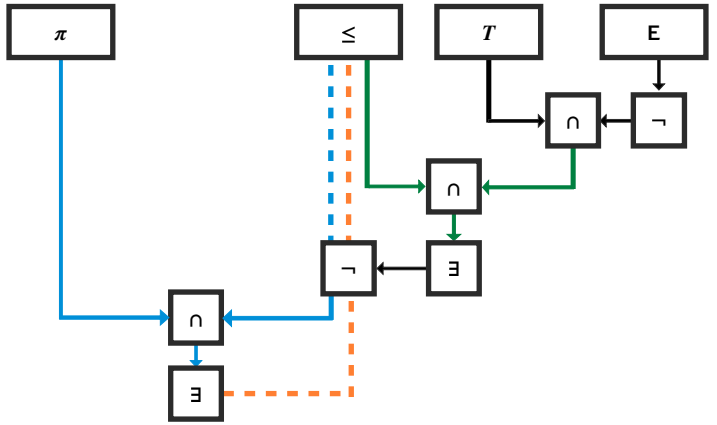
Overview



Temporal Causality in Reactive Systems [Coenen et al., ATVA 22]



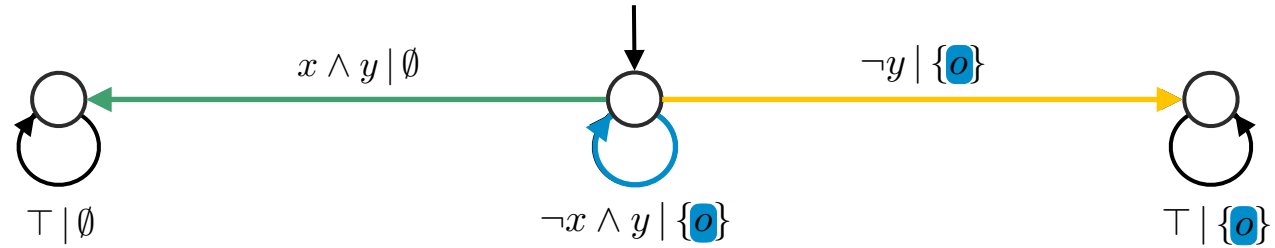
Temporal causes are downward-closed sets



Cause synthesis algorithm for omega-regular effects

Temporal Causality

Causes are temporal properties of the input sequence.



What causes $\square o$ on the trace $\{y, o\}^\omega$?

Temporal Causality in Reactive Systems [Coenen et al., ATVA 22]

SAT: Cause and effect hold on the given trace.

$y \wedge \square \neg x$ ✗ $\quad \square \neg x$ ✓ $\quad \neg y \, \mathcal{R} \, \neg x$ ✗

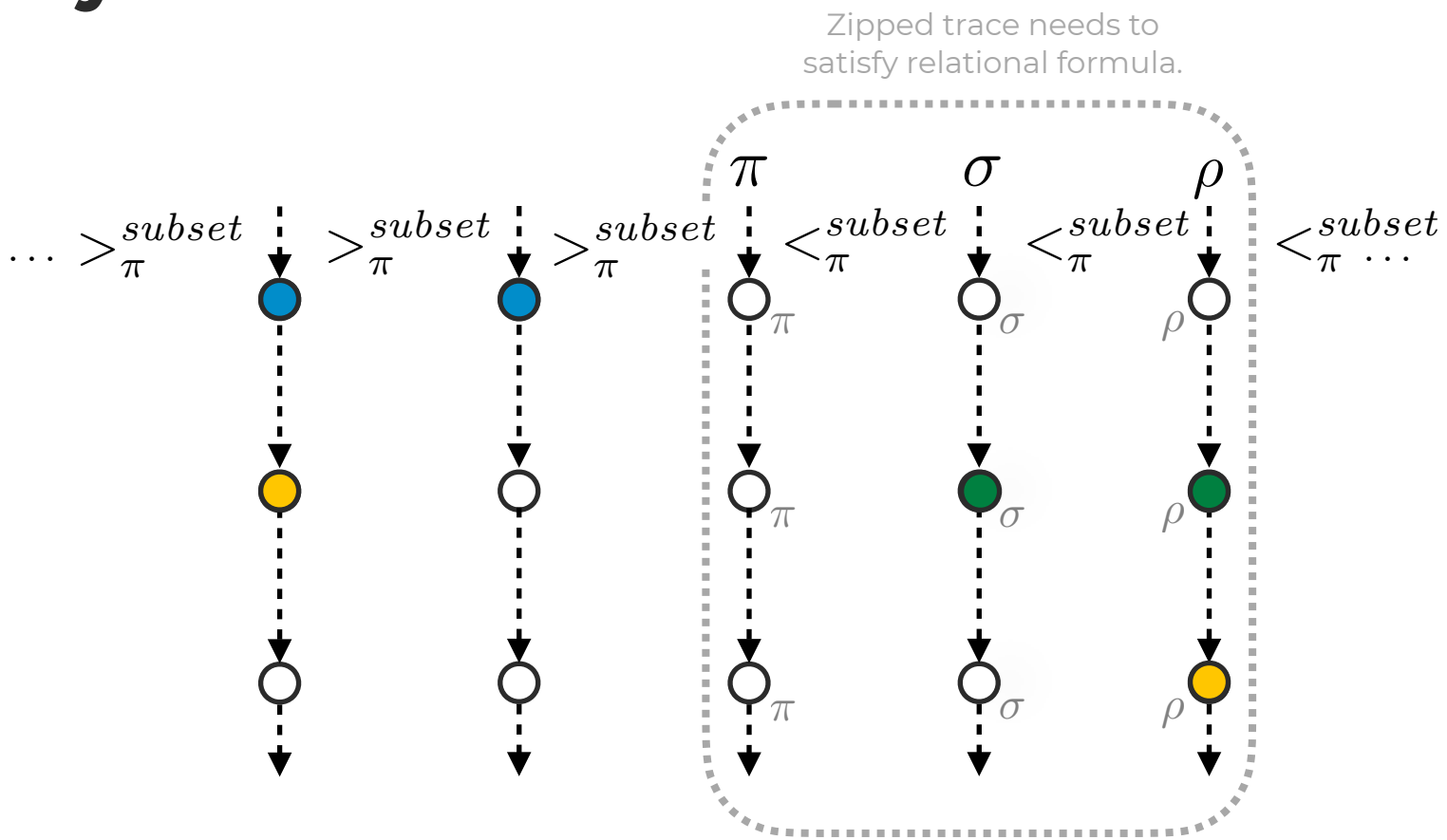
CF: All closest traces that do not satisfy the cause, also do not satisfy the effect.

$\{o\}\{y, o\}^\omega, \ldots \quad \{x, y\}\{y\}^\omega, \ldots$

MIN: No strict subset of the cause satisfies **SAT** and **CF**.

$\mathcal{L}(y \wedge \square \neg x) \subset \mathcal{L}(\square \neg x) \subset \mathcal{L}(\neg y \, \mathcal{R} \, \neg x)$

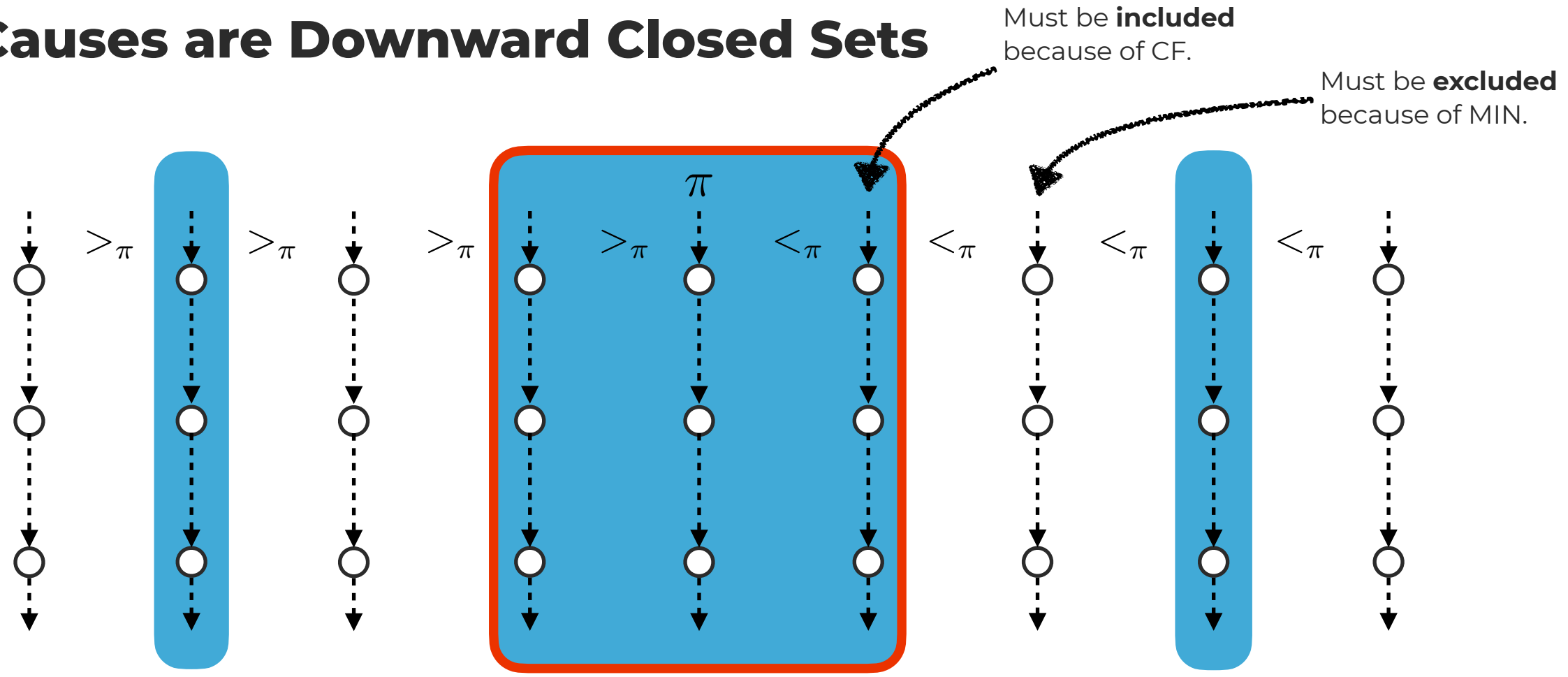
Similarity Relations



The closeness of traces is modeled by a similarity relation.

E.g. $\sigma \leq_\pi^{subset} \rho$ iff $\square \bigwedge_{i \in I} (i_\sigma \not\leftrightarrow i_\pi) \rightarrow (i_\rho \not\leftrightarrow i_\pi)$

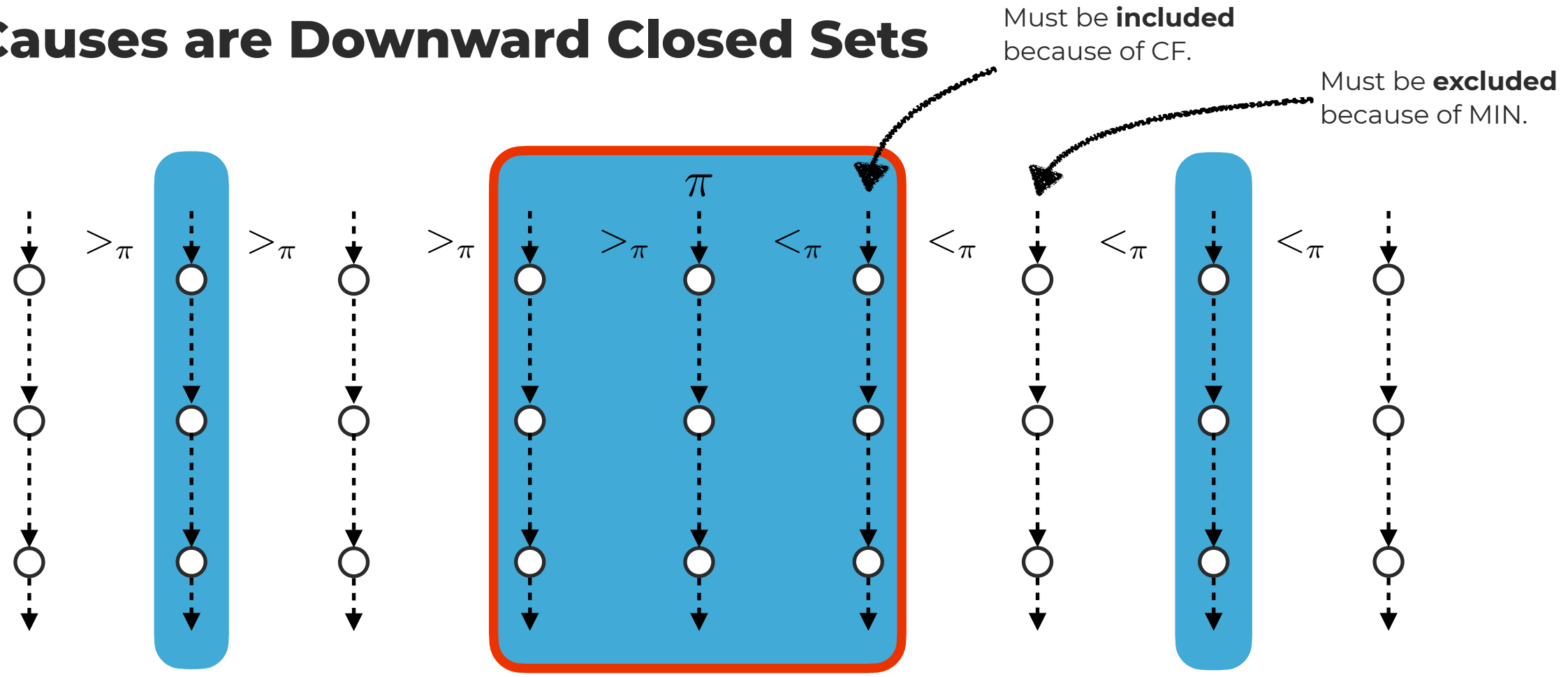
Causes are Downward Closed Sets



$$\text{Cause} = \{\rho \in \text{Effect} \mid \forall \sigma \in T.\, \sigma \leq_\pi \rho \rightarrow \sigma \in \text{Cause}\}$$

= the largest subset of the effect that is downward closed in $(T, \leq_\pi)$

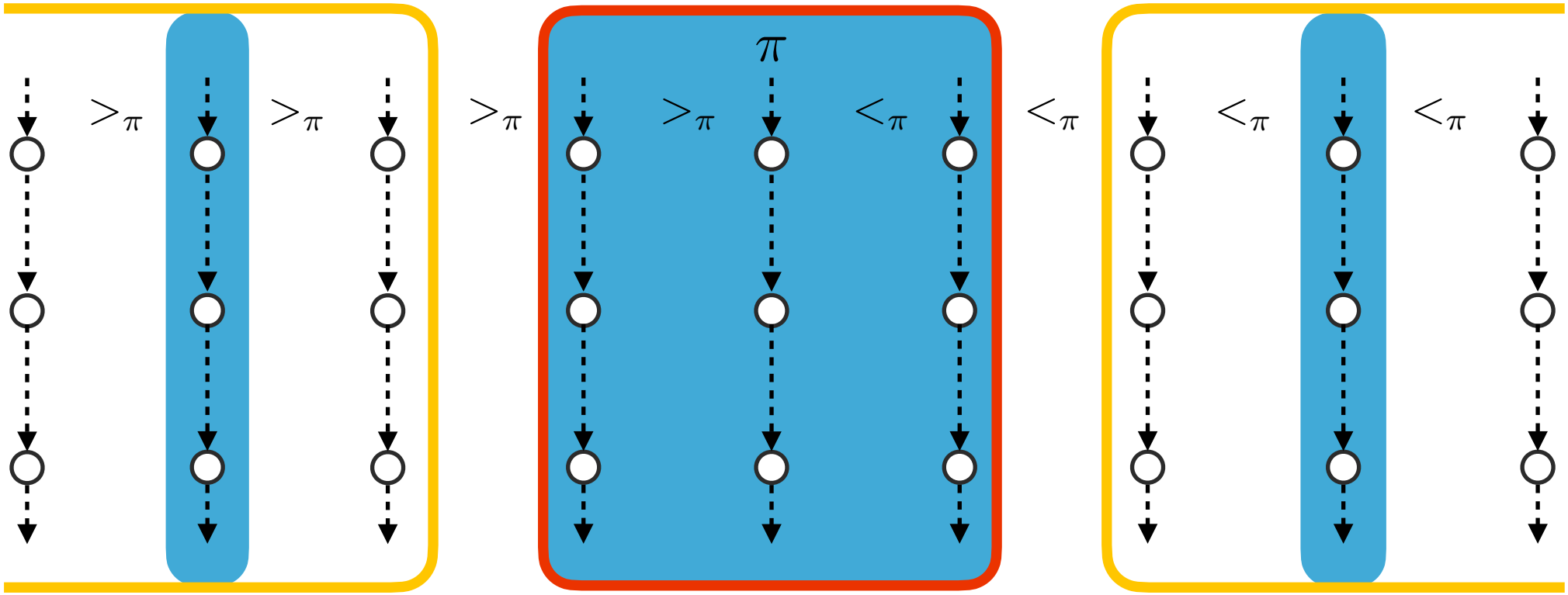
Causes are Downward Closed Sets



Can be rewritten to:

$$\boxed{\text{Cause}} = \{\rho \in T \mid \forall \sigma \in T.\, \sigma \leq_\pi \rho \rightarrow \sigma \in \boxed{\text{Effect}}\}$$

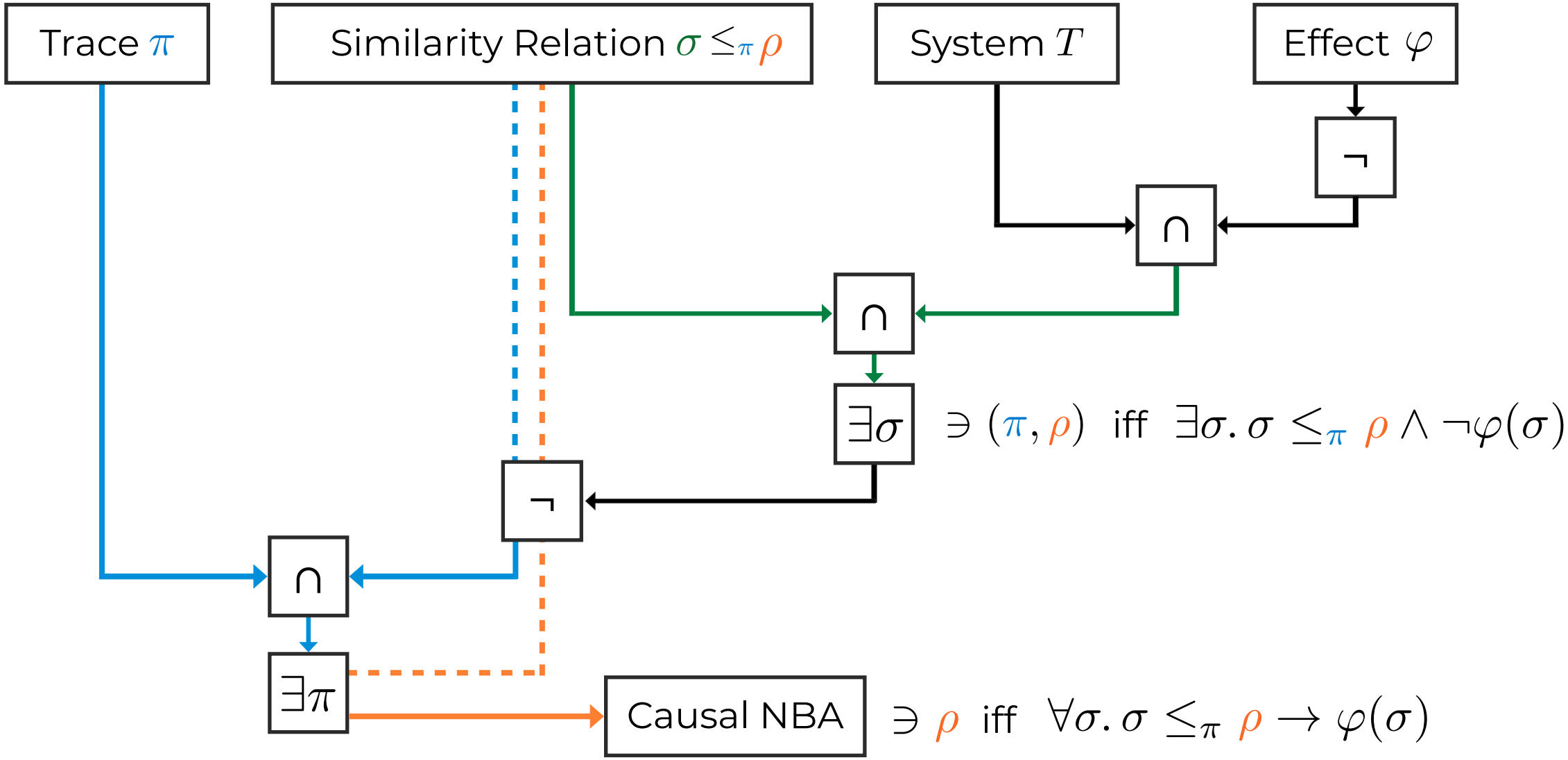
Cause Complements are Upward Closures



$$\overline{\text{Cause}} = \{\rho \in T \mid \exists \sigma \in T.\, \sigma \leq_\pi \rho \wedge \sigma \in \overline{\text{Effect}}\}$$

$$\text{Cause} = \{\rho \in T \mid \forall \sigma \in T.\, \sigma \leq_\pi \rho \rightarrow \sigma \in \text{Effect}\}$$

Automata-based Cause Synthesis



Experiments: Arbiters

CAV Artifact Evaluation ★★★ Reusable

Arbiters grant mutually exclusive access to a resource (e.g., grant g_0).

Instance	$\lvert\mathcal{T}\rvert$	φ_E	$t(\leq^{subset})$	$\lvert\mathcal{A}_C^{subset}\rvert$	φ_C
Full 2	4	$\Diamond g_0$	0.11	2	$\Diamond r_0$
		$\Box\Diamond g_0$	0.11	4	$\Box\Diamond r_0$
Full 3	11	$\Diamond g_0$	0.11	2	$\Diamond r_0$
		$\Box\Diamond g_0$	0.16	24	$\Box\Diamond r_0$
Full 4	46	$\Diamond g_0$	0.11	2	$\Diamond r_0$
		$\Box\Diamond g_0$	33.22	214	$\Box\Diamond r_0$
Spurious 2	2	$\Diamond g_0$	0.11	1	$true$
Spurious 3	3	$\Diamond g_0$	0.11	1	$true$
Spurious 4	4	$\Diamond g_0$	0.11	1	$true$
Unfair 2	2	$\Box\neg g_0$	0.11	1	$\Box r_{prio}$
Unfair 3	4	$\Box\neg g_0$	0.11	1	$\Box r_{prio}$
Unfair 4	6	$\Box\neg g_0$	0.11	1	$\Box r_{prio}$

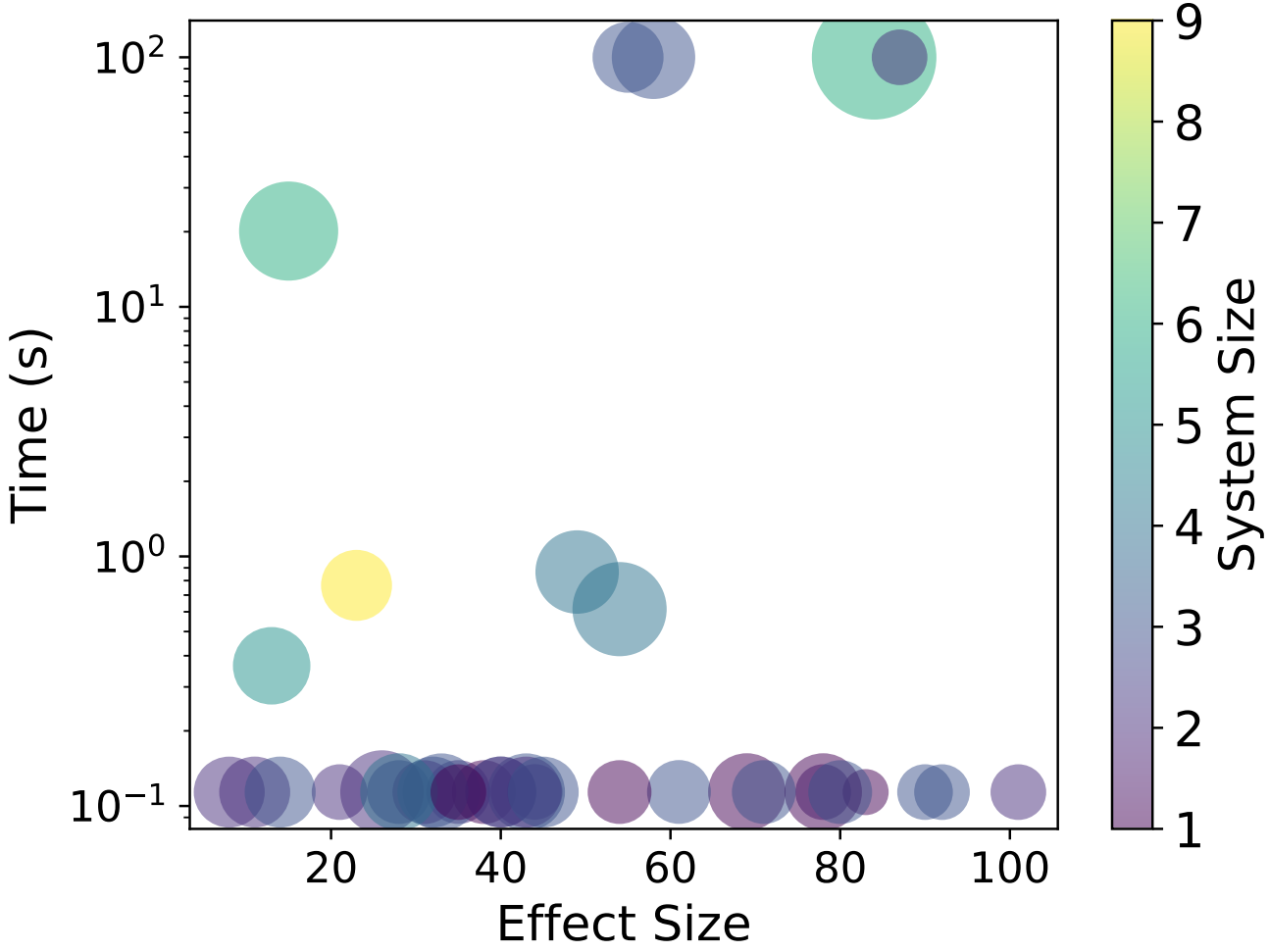
Round-robin strategy

Prioritizes $prio$

We always consider the input sequence where all clients send requests continuously, e.g.:

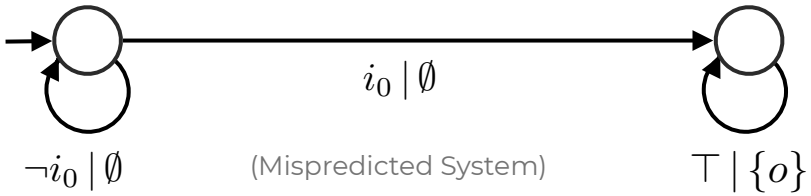
$(\{r_0, r_1, g_0\}\{r_0, r_1, g_1\})^\omega$

Experiments: Neural Synthesis[1]



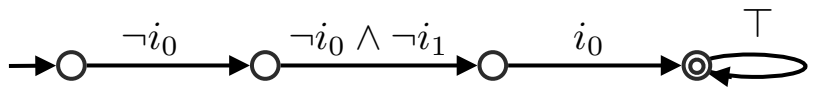
$$\text{Effect} = \neg\big((i_1 \,\mathcal{U}\, i_0) \leftrightarrow (\square\lozenge o)\big)$$

(Negated Specification)



$$\pi = \{i_1\} \cdots \{\} \cdots \{i_0, o\} \cdots (\{i_1, o\})^\omega$$

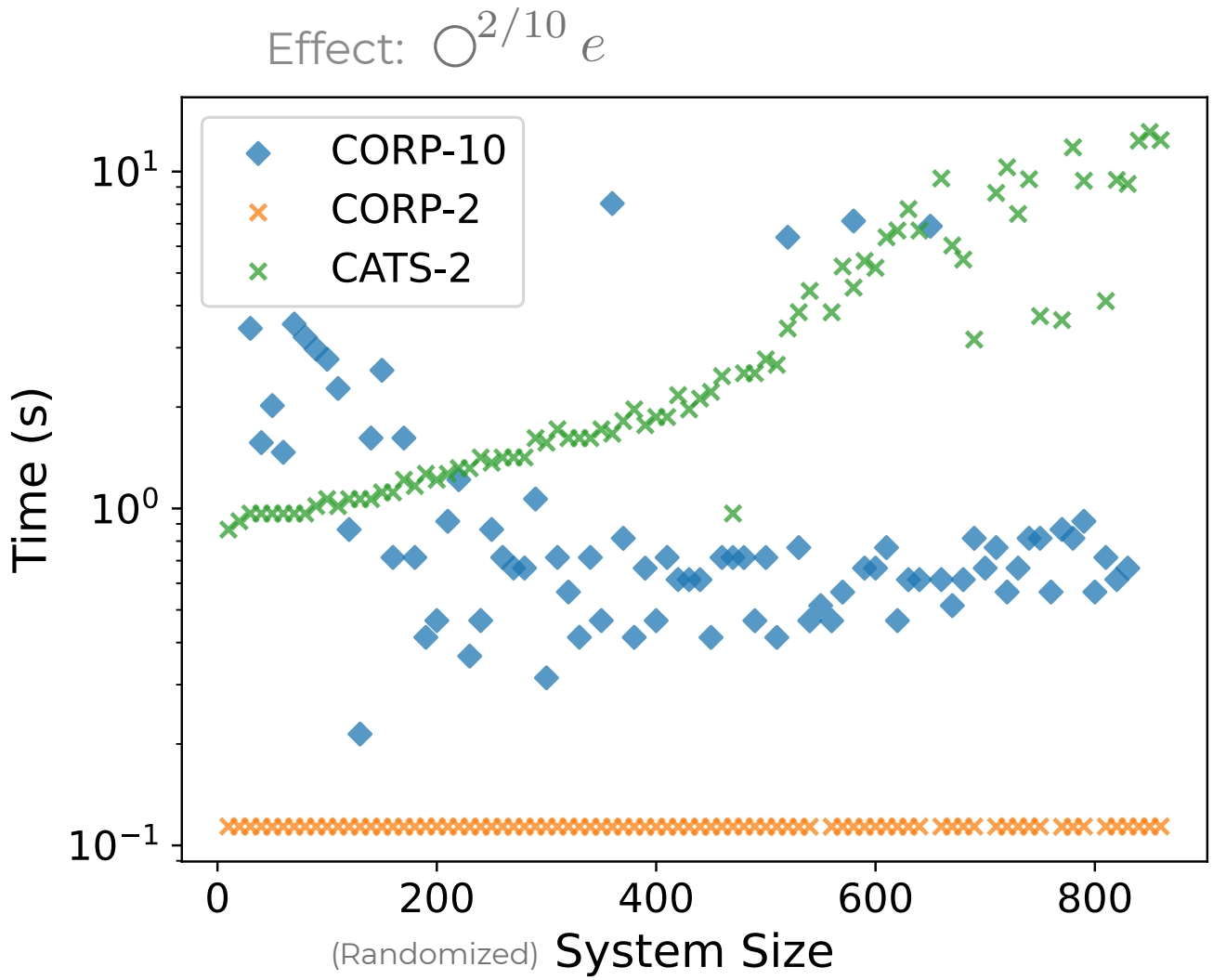
(Counterexample to Specification)



(Causal NBA)

[1] *Neural circuit synthesis from specification patterns*. Schmitt, Hahn, Rabe, and Finkbeiner. NeurIPS 2021.

Experiments: Cause Sketching[1]

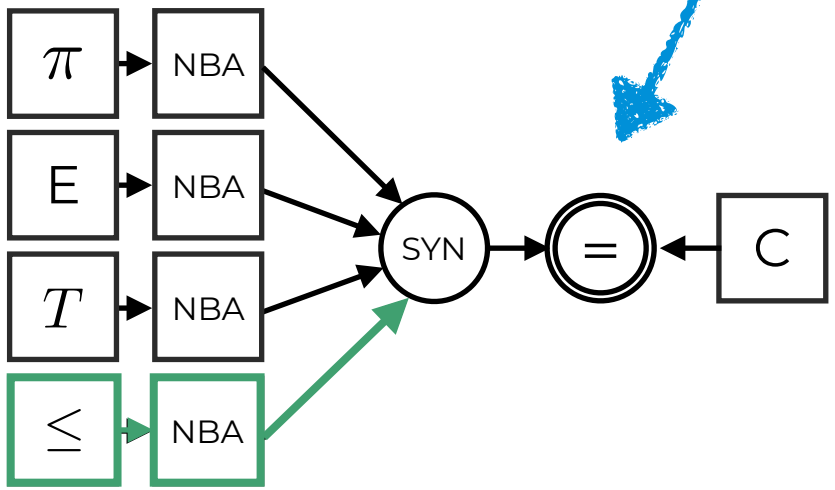


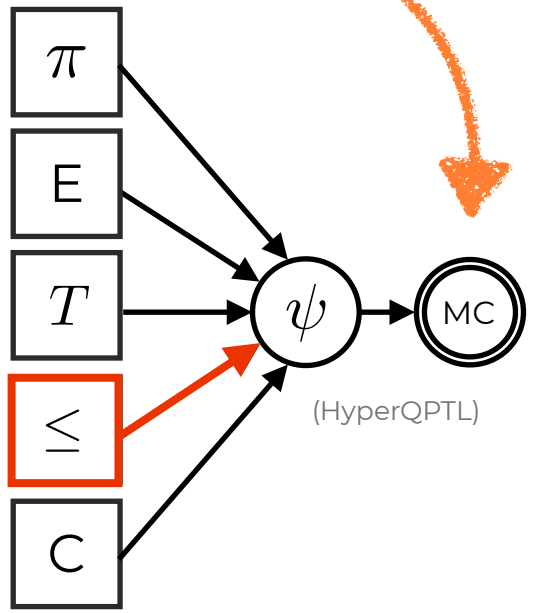
[1] *Checking and Sketching Causes on Temporal Sequences.* Beutner, Finkbeiner, Frenkel, and Siber. ATVA 2023.

Experiments: Cause Checking

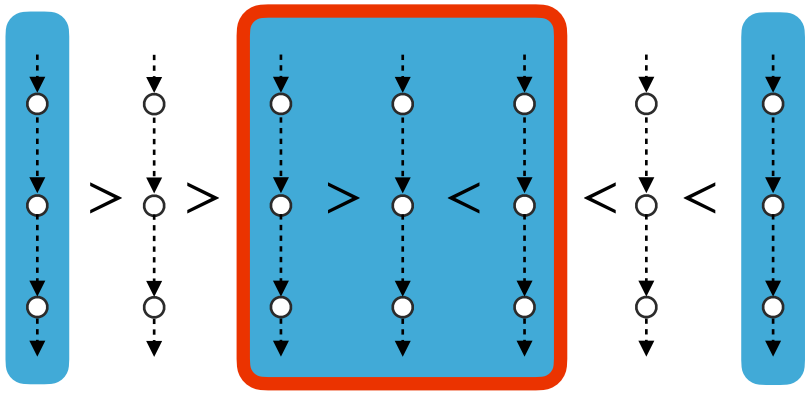
CAV Artifact Evaluation ★ ★ ★ Reusable

Instance	CATS	CORP
Spurious Arbiter	0.72	0.11
Simple Arbiter	0.97	0.11
Arbiter	22.84	0.11
Instance 6, odd	1.81	0.11
Instance 8	0.92	0.11
TP Right	0.87	0.11

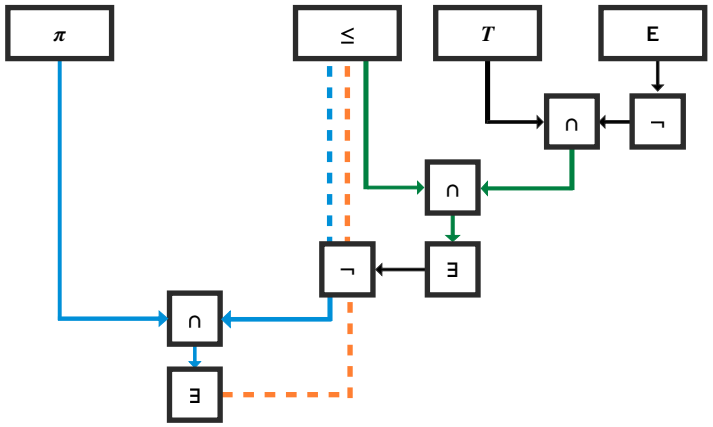




Conclusion



Minimal counterfactual causes are downward-closed sets



Causes can be synthesized under omega-regular assumptions

Are there always closest traces? What about nondeterminism?
See our arXiv version!

CAV Artifact Evaluation ★ ★ ★ Reusable

Want to know what caused the violation on your counterexample?
Check out our artifact!