ACTUAL CAUSALITY AND COUNTERFACTUAL REASONING

MOHAMMAD MOUSAVI
SCHOOL OF INFORMATICS
UNIVERSITY OF LEICESTER

UNIVERSITY OF LEICESTER

CAUSALITY: BACKGROUND

A Railway Crossing Hazard

Safety goal:

- “It shall always be the case that there is never a car and a train in crossing at the same time"

What is a Cause?

[Lewis 1973] "Causation". Journal of Philosophy (1973)

- possible world semantics for counterfactuals
  - **c** is causal for **e** (in a model **m**), if were **c** not to occur, then **e** would not occur either

What is a Cause?

[Lewis 1973] "Causation". Journal of Philosophy (1973)

- possible world semantics for counterfactuals
  - **c** is causal for **e** (in a model **m**), if were **c** not to occur, then **e** would not occur either

[Halpern, Pearl 2005] "Causes and explanations: A structural-model approach. Part I: Causes". The British Journal for the Philosophy of Science (2005)

What is a Cause?

[Lewis 1973] "Causation". Journal of Philosophy (1973)

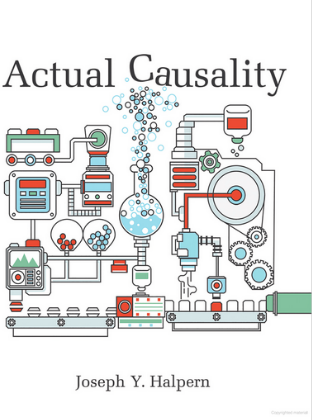
- possible world semantics for counterfactuals
  - **c** is causal for **e** (in a model **m**), if were **c** not to occur, then **e** would not occur either

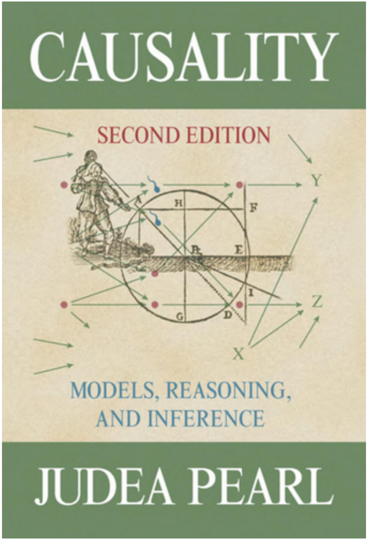
[Halpern, Pearl 2005] "Causes and explanations: A structural-model approach. Part I: Causes". The British Journal for the Philosophy of Science (2005)

[Leitner-Fischer, Leue 2013] "Causality Checking for Complex System Models". VMCAI (2013)

- adaptation of [Halpern, Pearl 2005] to concurrent computations and reachability properties
- considers ordering and non-occurrence of events as potential causal factors

Textbooks


Actual Causality
Joseph Y. Halpern


CAUSALITY
SECOND EDITION
MODELS, REASONING, AND INFERENCE
JUDEA PEARL

Our Order of Business

1. Formalising a notion of causality for reactive systems
2. Studying its compositionality
3. Discussing the extension of causality for cyber-physical- and autonomous systems

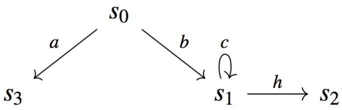
UNIVERSITY OF LEICESTER

CAUSALITY FOR REACTIVE SYSTEMS

Labelled Transition Systems

Labelled Transition Systems (LTS's)

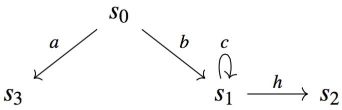
1. transitions: $s_0 \xrightarrow{b} s_1$



Labelled Transition Systems

Labelled Transition Systems (LTS's)

1. transitions: $s_0 \xrightarrow{b} s_1$
2. trace: $s_0 \overset{bcch}{\twoheadrightarrow} s_2$, ε – trace

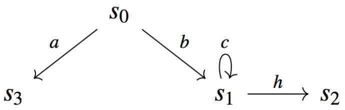


Labelled Transition Systems

Labelled Transition Systems (LTS's)

1. transitions: $s_0 \xrightarrow{b} s_1$
2. trace: $s_0 \overset{bcch}{\twoheadrightarrow} s_2$, ε − trace
3. computations, *e.g.*,

   $traces(\pi) = \{$
4. $s_0 \overset{\epsilon}{\twoheadrightarrow} s_0$,
5. $s_0 \xrightarrow{b} s_1 \xrightarrow{h} s_2$,
6. $s_0 \xrightarrow{b} s_1 \xrightarrow{c} s_1 \xrightarrow{h} s_2$,
7. ...
8. $s_0 \xrightarrow{b} s_1 \xrightarrow{c} \dots \xrightarrow{c} s_1 \xrightarrow{h} s_2 \}$



Labelled Transition Systems

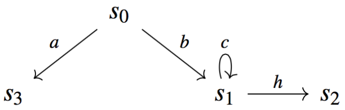
Labelled Transition Systems (LTS's)

1. transitions: $s_0 \xrightarrow{b} s_1$
2. trace: $s_0 \overset{bcch}{\twoheadrightarrow} s_2$, ε – trace
3. computations, *e.g.*,

   $\mathit{traces}(\pi) = \{$
4. $s_0 \overset{\epsilon}{\twoheadrightarrow} s_0$,
5. $s_0 \xrightarrow{b} s_1 \xrightarrow{h} s_2$,
6. $s_0 \xrightarrow{b} s_1 \xrightarrow{c} s_1 \xrightarrow{h} s_2$,
7. $\ldots$
8. $s_0 \xrightarrow{b} s_1 \xrightarrow{c} \ldots \xrightarrow{c} s_1 \xrightarrow{h} s_2$ $\}$

   $\pi = (s_0, b, [\varepsilon, c, cc, \ldots]), (s_1, h, [\varepsilon, \varepsilon, \varepsilon, \ldots]), s_2$

   $(s_0, b, [\varepsilon, c]), s_1 \in \mathit{sub}(\pi)$



Hennessy-Milner Logic

Hennessy-Milner Logic (HML). Syntax & Semantics.

$$\phi, \psi ::= \top \mid \neg\phi \mid \phi \wedge \psi \mid \langle a \rangle \phi \qquad (a \in A).$$

$s \models \top$ for all $s \in \mathbb{S}$
$s \models \neg\phi$ whenever s does not satisfy ϕ; also written as $s \not\models \phi$
$s \models \phi \wedge \psi$ if and only if $s \models \phi$ and $s \models \psi$
$s \models \langle a \rangle \phi$ if and only if $s \xrightarrow{a} s'$ for some $s' \in \mathbb{S}$ such that $s' \models \phi$

Causality for LTS's – AC1

Consider an LTS T and an HML property ϕ in T.
$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

1. **Positive causality, AC1**

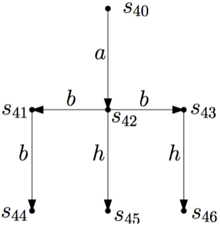
   The causal trace leads to the effect:
   $s_0 \xrightarrow{l_0} \ldots s_n \xrightarrow{l_n} s_{n+1} \wedge s_{n+1} \models \phi$

Causality for LTS's – AC1

Consider an LTS T and an HML property ϕ in T.
$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

1. **Positive causality, AC1**

   The causal trace leads to the effect:
   $s_0 \xrightarrow{l_0} \ldots s_n \xrightarrow{l_n} s_{n+1} \wedge s_{n+1} \vDash \phi$



$\phi = \langle h \rangle \top$
$\pi = (s_{40}, a, \mathcal{D}_{40}), s_{42}$

Causality for LTS's – AC2(a)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

2. **Counter-factual, AC2(a)**

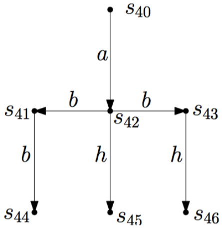
   The effect does not hold trivially:
   $\exists \chi \in A^*, s' \in \mathbb{S} : s_0 \overset{\chi}{\twoheadrightarrow} s' \wedge s' \models \neg\phi$

Causality for LTS's – AC2(a)

$\pi = (s_0, l_0, \mathcal{D}_0), \dots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

2. **Counter-factual, AC2(a)**

   The effect does not hold trivially:
   $\exists \chi \in A^*, s' \in \mathbb{S} : s_0 \xrightarrow{\chi}\!\!\!\!\rightarrow s' \wedge s' \vDash \neg\phi$



$\phi = \langle h \rangle \top$
e.g., $\chi = abb$, $\chi = ah$

Causality of non-occurrence

What if the car leaves (*Cl*) the crossing before the train enters the crossing?

- *Cl* is causal by its non-occurrence...

Causality for LTS's – AC2(b)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in \mathit{Causes}(\phi, T)$ iff:

3. **Causality of occurrence, AC2(b)** Interleaving "other actions" with the causal trace keeps the effect:

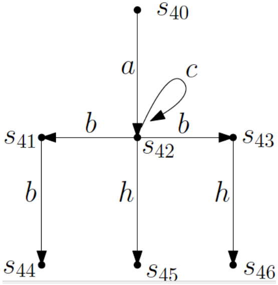
   $\forall \chi' = l_0 \chi_0 \ldots l_n \chi_n \in (A^* \setminus \mathit{traces}(\pi)) \cup \{l_0 \ldots l_n\}$,

   $s_0 \xrightarrow{\chi'}\!\!\!\!\rightarrow s' \Rightarrow s' \models \phi$

Causality for LTS's – AC2(b)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

3. **Causality of occurrence, AC2(b)** Interleaving "other actions" with the causal trace keeps the effect:
   $\forall \chi' = l_0\chi_0 \ldots l_n\chi_n \in (A^* \setminus traces(\pi)) \cup \{l_0 \ldots l_n\},$
   $s_0 \xrightarrow{\chi'}\!\!\!\!\!\rightarrow s' \Rightarrow s' \models \phi$



$\phi = \langle h \rangle \top$
$\pi = (s_{40}, a, [h, bb, bh]), s_{42}$

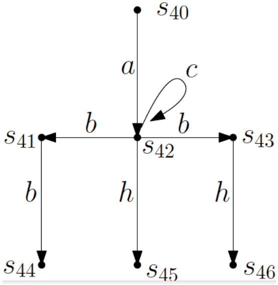
Causality for LTS's – AC2(b)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

3. **Causality of occurrence, AC2(b)** Interleaving "other actions" with the causal trace keeps the effect:

   $\forall \chi' = l_0\chi_0 \ldots l_n\chi_n \in (A^* \setminus traces(\pi)) \cup \{l_0 \ldots l_n\},$

   $s_0 \xrightarrow{\chi'}\!\!\!\!\rightarrow s' \Rightarrow s' \vDash \phi$



$\phi = \langle h \rangle \top$

$\pi = (s_{40}, a, [h, bb, bh]), s_{42}$ but not $\pi = (s_{40}, a, [c, \ldots]), s_{42}$

Causality for LTS's – AC2(c)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in \mathit{Causes}(\phi, T)$ iff:

4. **Causality of non-occurrence, AC2(c)** Interleaving "preventive actions" will remove the effect:

   $\forall \chi' \in (\mathit{traces}(\pi) \setminus \{l_0 \ldots l_n\}), s' \in \mathbb{S}:$

   $s_0 \xrightarrow{\chi'}\!\!\!\!\rightarrow s' \Rightarrow s' \models \neg\phi$

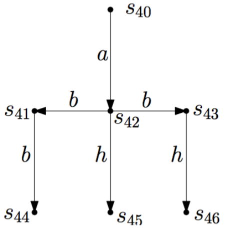
Causality for LTS's – AC2(c)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

4. **Causality of non-occurrence, AC2(c)** Interleaving "preventive actions" will remove the effect:

   $\forall \chi' \in (traces(\pi) \setminus \{l_0 \ldots l_n\}), s' \in \mathbb{S}$:

   $s_0 \xrightarrow{\chi'}\!\!\!\!\rightarrow s' \Rightarrow s' \models \neg\phi$



$\phi = \langle h \rangle \top$

$\pi = (s_{40}, a, [h, bb, bh]), s_{42}$

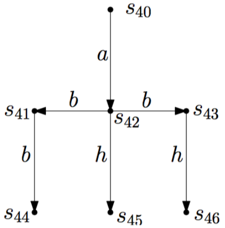
Causality for LTS's – AC2(c)

$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in Causes(\phi, T)$ iff:

4. **Causality of non-occurrence, AC2(c)** Interleaving "preventive actions" will remove the effect:

   $\forall \chi' \in (traces(\pi) \setminus \{l_0 \ldots l_n\}), s' \in \mathbb{S}:$

   $s_0 \xrightarrow{\chi'}\!\!\!\!\rightarrow s' \Rightarrow s' \models \neg\phi$



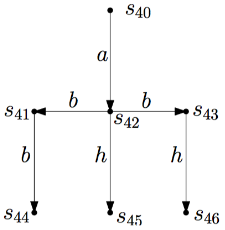
$\phi = \langle h \rangle \top$

$\pi = (s_{40}, a, [h, bb, bh]), s_{42}$ but not $\pi = (s_{40}, a, [h]), s_{42}$

Causality for LTS's – AC3

Consider an LTS T and an HML property ϕ in T.
$\pi = (s_0, l_0, \mathcal{D}_0), \ldots, (s_n, l_n, \mathcal{D}_n), s_{n+1} \in \mathit{Causes}(\phi, T)$ iff:

5. **Minimality, AC3**
   $\forall \pi' \in \mathit{sub}(\pi) : \pi'$ does not satisfy AC1–AC2(c)



$\phi = \langle h \rangle \top$
$\pi = (s_{40}, a, [h, bb, bh]), s_{42}$ satisfies AC1–AC2(c)
$\mu = (s_{40}, a, [\varepsilon, \varepsilon]), (s_{42}, b, [h, b]), s_{43}$ violates AC3 as $\pi \in \mathit{sub}(\mu)$

UNIVERSITY OF LEICESTER

DECOMPOSING CAUSALITY

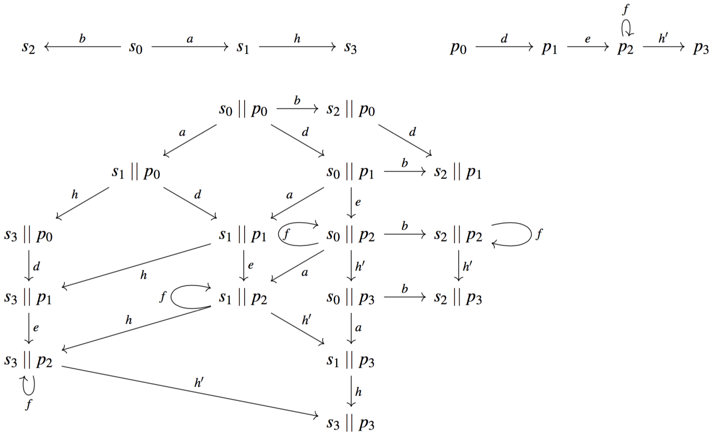
Composing LTS's

$$\frac{s \xrightarrow{a} s'}{s \mid\mid p \xrightarrow{a} s' \mid\mid p} \qquad \frac{p \xrightarrow{a} p'}{s \mid\mid p \xrightarrow{a} s \mid\mid p'}$$

$$\frac{s \xrightarrow{a} s'}{s + p \xrightarrow{a} s'} \qquad \frac{p \xrightarrow{a} p'}{s + p \xrightarrow{a} p'}$$

(De-)Composing Causality

From causality in $s_0 \mid\mid p_0$ to causality in s_0 and/or p_0?

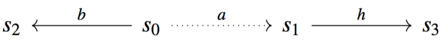


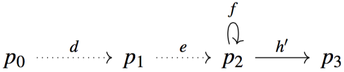
Causal Projection

Consider an LTS T and an HML property ϕ in T.

$T \downarrow \phi$ (or $s_0 \downarrow \phi$): causal projection of T w.r.t. ϕ

- *e.g.*, $s_0 \downarrow \langle h \rangle \top$ and $p_0 \downarrow \langle h' \rangle \top$:



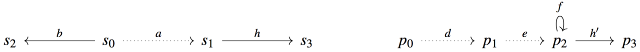


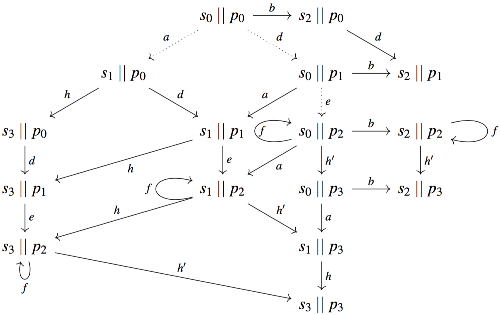
(De-)Composing Disjunction

Consider LTS's $T = (\mathbb{S}, s_0, A, \rightarrow)$ and $T' = (\mathbb{S}', s_0', B, \rightarrow')$ such that $A \cap B = \emptyset$. Assume two HML formulae ϕ and ψ over A and B, respectively. The following holds:

$$T \,||\, T' \downarrow (\phi \vee \psi) \;\simeq\; T \downarrow \phi + T' \downarrow \psi.$$

Example: $\langle h \rangle \top \vee \langle h' \rangle \top$



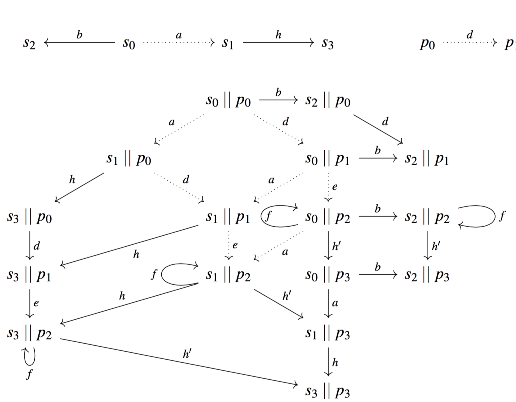


(De-)Composing Conjunction

Consider LTS's $T = (\mathbb{S}, s_0, A, \rightarrow)$ and $T' = (\mathbb{S}', s_0', B, \rightarrow')$ such that $A \cap B = \emptyset$. Assume two HML formulae ϕ and ψ over A and B, respectively. The following holds:

$$T \,||\, T' \downarrow (\phi \wedge \psi) \; = \; (T \downarrow \phi) \,||\, (T' \downarrow \psi).$$

Example: $\langle h \rangle \top \wedge \langle h' \rangle \top$



Conclusions & Future Work

Our contributions:

- defined causality for LTS's & HML (safety properties)
- established first compositionality results for non-communicating LTS's
- implemented in a model-checker (mCRL2)

Future work:

- extension to communicating LTS's (in the style of CCS)
- extension to liveness properties (in the modal μ-calculus)

[Caltais, Mousavi, and Singh, Causal Reasoning for Safety in HML, Fundamenta Informaticae, 2020]

UNIVERSITY OF LEICESTER

THANK YOU VERY MUCH!

QUESTIONS?

mm789@le.ac.uk