# Abstraction of Structural Causal Models

Fabio Massimo Zennaro
fabio.zennaro@warwick.ac.uk

University of Warwick
October 31st, 2022

# Outline

$$\begin{tikzcd}
 & \textit{1. Background} \arrow[dl] \arrow[dr] & \\
\textit{2. Transformation Approach} \arrow[dr] & & \textit{3. Abstraction Approach} \arrow[dl] \\
 & \textit{4. Research directions} &
\end{tikzcd}$$

# 1. Background

# Problem definition

Systems may be represented at different **levels of abstraction** (LoA).

*Thermodynamics example:*

*Low-level / Base model:*
Microscopic description $\mathbf{p}, \dot{\mathbf{p}}$.

*High-level / Abstracted model:*
Macroscopic description $P, T, V$.

LoA may be inaccessible, so we may want to *shift* among LoAs.

1. We need a *mapping* between LoAs.
2. We want the mapping to be *consistent*.

# Problem definition

SCMs are becoming more popular for encoding causal models.

*Lung cancer scenario example:*

$$S \longrightarrow T \longrightarrow C \qquad\qquad S' \longrightarrow C'$$

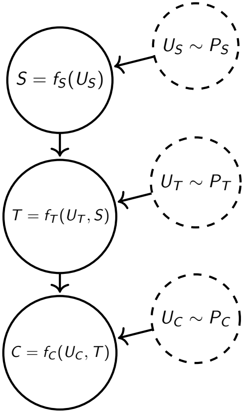
1. How do we find a *mapping*?
2. How do we define and guarantee some form of *consistency*?

This could allow us to shift between LoAs of SCMs, taking advantage of data and computational resources.

# SCMs

We express a causal model as a **structural causal model** $\mathcal{M}$ [4, 5]:

- $\mathcal{X}$: set of *endogenous nodes* $(S, T, C)$ representing variables of interest
- $\mathcal{E}$: Set of *exogenous nodes* $(U_S, U_T, U_C)$ representing stochastic factors
- $\mathcal{F}$: Set of *structural functions* $(f_S, f_T, f_C)$ describing the dynamics of each variable
- $\mathcal{P}$: Set of *distributions* $(P_S, P_T, P_C)$ describing the random factors



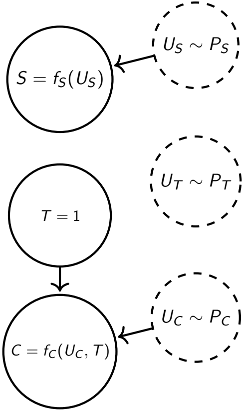

Every SCM $\mathcal{M}$ implies a (joint) **distribution** $P_{\mathcal{M}}$: $P_{\mathcal{M}}(S, T, C)$

# Interventions

We can perform **interventions** on a causal model [4, 5]:

$do(T = 1)$

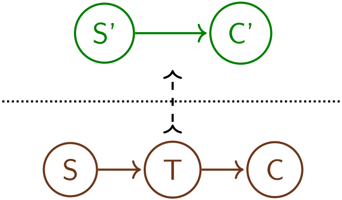
1. Remove incoming edges in the intervened node
2. Set the value of the intervened node




An intervention $\iota_1$ effectively defines a new **intervened model** $\mathcal{M}_{\iota_1}$ such that $P_{\mathcal{M}}(S, T, C) \neq P_{\mathcal{M}_{\iota_1}}(S, T, C)$

# Two approaches

*Lung cancer scenario example:*




$$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$$

$$\mathcal{M}[S] = \mathcal{M}[T] = \mathcal{M}[C] = \{0, 1\}$$

- The **transformation** approach [8]
- The **abstraction** approach [7]

# 2. Transformation approach [8]

# The *transformation* approach: mapping

Given two SCMs $\mathcal{M}$ and $\mathcal{M}'$, let us consider the **transformation**:

$$\tau : \prod_i \mathcal{M}[X_i] \rightarrow \prod_j \mathcal{M}'[X_j]$$

$\tau$ : domain of the variables of $\mathcal{M}$ $\rightarrow$ domain of the variables of $\mathcal{M}'$.
$\tau$ : an output/configuration of $\mathcal{M}$ $\mapsto$ an output/configuration of $\mathcal{M}'$.

This implies a (pushforwarded) distribution on $\mathcal{M}'$:

$$\begin{tikzcd}
\prod_i \mathcal{M}[X_i] \arrow[r, "\tau"] \arrow[dd, dash, dashed] & \prod_j \mathcal{M}'[X_j] \arrow[d, dash, dashed] \\
 & P_{\mathcal{M}'} \\
P_{\mathcal{M}} \arrow[r, "\tau"'] & \tau(P_{\mathcal{M}})
\end{tikzcd}$$

If $\tau_{\#}(P_{\mathcal{M}}) = P_{\mathcal{M}'}$ we have *observational consistency*.

# The *transformation* approach: consistency

Let us consider a mapping between interventions:

$$\omega : \mathcal{I} \to \mathcal{I}'$$

$$\omega : \text{an intervention on } \mathcal{M} \mapsto \text{an intervention on } \mathcal{M}'.$$

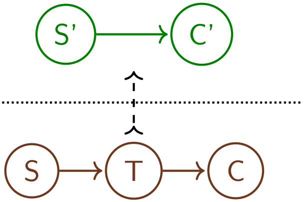
A transformation is an *exact transformation* if there exist a surjective order-preserving $\omega$ such that:

$$\begin{tikzcd}
P_{\mathcal{M}} \arrow[r, "\tau"] \arrow[dd, "\iota"'] & \tau(P_{\mathcal{M}}) = P_{\mathcal{M}'} \arrow[d, "\omega(\iota)"] \\
& P_{\mathcal{M}_{\omega(\iota)}} \\
P_{\mathcal{M}_\iota} \arrow[r, "\tau"'] & \tau(P_{\mathcal{M}_\iota})
\end{tikzcd}$$

where $\tau(P_{\mathcal{M}_\iota}) = P_{\mathcal{M}_{\omega(\iota)}}$, $\forall \iota \in \mathcal{I}$. We have *interventional consistency*.

# The *transformation* approach: example

*Lung cancer scenario example:*




$\tau : \mathcal{M}[S] \times \mathcal{M}[T] \times \mathcal{M}[C] \to \mathcal{M}'[S'] \times \mathcal{M}'[C']$

$\tau : (s, t, c) \mapsto (s, c)$

Set of interventions: $\mathcal{I} = \{\emptyset, do(S = 0)\}$

$$\omega : \begin{cases} \emptyset \mapsto \emptyset \\ do(S = 0) \mapsto do(S' = 0) \end{cases}$$

*Consistency condition*:

$$\begin{tikzcd} P_{\mathcal{M}}(S, T, C) \arrow[r, "\tau"] \arrow[d, "\iota"'] & P_{\mathcal{M}'}(S', C') \arrow[d, "\omega(\iota)"] \\ P_{\mathcal{M}}(T, C | do(S = 0)) \arrow[r, "/" marking, "\tau"'] & P_{\mathcal{M}'}(C' | do(S' = 0)) \end{tikzcd}$$

# The *transformation* approach: summary

Given:

- A low-level model $\mathcal{M}$ with a set of interventions of interest $\mathcal{I}$;
- A high-level model $\mathcal{M}'$;
- A surjective order-preserving $\omega : \mathcal{I} \rightarrow \mathcal{I}'$

an **exact transformation** $\tau$ guarantees that if I:

- work (intervene) at low-level and then switch (abstract) to high-level,
- or, switch first to high-level and then work there,

I will observe the same statistical behavior in the two models.

# 3. Abstraction approach [7]

# The *abstraction* approach: mapping

Let $\mathcal{M}$ and $\mathcal{M}'$ be two finite SCMs with finite domains. An **abstraction** is a tuple

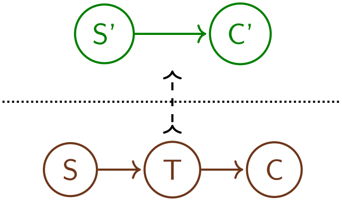
$$(R, a, \alpha)$$

where

- $R \subseteq \mathcal{X}_{\mathcal{M}}$ is a subset of *relevant nodes* among the endogenous nodes of $\mathcal{M}$.
- $a : R \rightarrow \mathcal{X}_{\mathcal{M}'}$ is a *surjective function* mapping a low-level node in $\mathcal{M}$ to a high-level node in $\mathcal{M}'$.
- $\alpha$ is a *collection of surjective functions*, one for each high-level node $X'$, defined as $\alpha_{X'} : \mathcal{M}[a^{-1}(X')] \rightarrow \mathcal{M}'[X']$.
  $\alpha'_X$ maps an output of the low-level nodes sent onto $X'$ by $a$ onto an output of $X'$.

# The *abstraction* approach: example (I)

*Lung cancer scenario example:*




$$R = \{S, C\} \subseteq \mathcal{X}_{\mathcal{M}}$$

$$a : R \rightarrow \mathcal{X}_{\mathcal{M}'}$$

$$a : \begin{cases} S \mapsto S' \\ C \mapsto C' \end{cases}$$

$$\alpha : \begin{cases} \alpha_{S'} : \{0,1\} \rightarrow \{0,1\} \\ \alpha_{S'} : s \mapsto s \\ \alpha_{C'} : \{0,1\} \rightarrow \{0,1\} \\ \alpha_{C'} : c \mapsto c \end{cases}$$

# The *abstraction* approach: consistency

We have *(interventional) consistency* if the following diagram commutes for all the disjoint subsets $X', Y' \in \mathcal{X}_{\mathcal{M}'}$ for every value in $\mathcal{M}[a^{-1}(X')]$:

$$\begin{tikzcd}
\mathcal{M}[a^{-1}(X')] \arrow[r, "{\mathcal{M}[\phi_{a^{-1}(Y')}]}"] \arrow[d, "\alpha_{X'}"'] & \mathcal{M}[a^{-1}(Y')] \arrow[d, "\alpha_{Y'}"] \\
\mathcal{M}'[X'] \arrow[r, "{\mathcal{M}'[\phi_{Y'}]}"'] & \mathcal{M}'[Y']
\end{tikzcd}$$

that is, we get an identity:

$$\alpha_{Y'} \circ \mathcal{M}[\phi_{a^{-1}(Y')}] = \mathcal{M}'[\phi_{Y'}] \circ \alpha_{X'}$$

# The *abstraction* approach: abstraction error

If the diagram does not commute for $X', Y' \in \mathcal{X}_{\mathcal{M}'}$:

$$\begin{tikzcd}
\mathcal{M}[a^{-1}(X')] \arrow[r, "{\mathcal{M}[\phi_{a^{-1}(Y')}]}"] \arrow[d, "\alpha_{X'}"'] & \mathcal{M}[a^{-1}(Y')] \arrow[d, "\alpha_{Y'}"] \\
\mathcal{M}'[X'] \arrow[r, "{\mathcal{M}'[\phi_{Y'}]}"'] & \mathcal{M}'[Y']
\end{tikzcd}$$

I can compute the *abstraction error* for $X', Y'$:

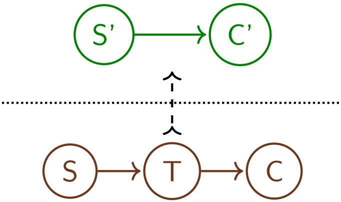
$$E_\alpha(X', Y') = D_{JSD}(\alpha_{Y'} \circ \mathcal{M}[\phi_{a^{-1}(Y')}], \mathcal{M}'[\phi_{Y'}] \circ \alpha_{X'})$$

I can compute the *overall abstraction error* as the worst-case:

$$e(\alpha) = \sup_{X', Y' \in \mathcal{X}_{\mathcal{M}'}} E_\alpha(X', Y')$$

# The *abstraction* approach: example (II)

*Lung cancer scenario example:*




Assuming no commutativity

$$\begin{tikzcd}
\mathcal{M}[S] \arrow[r, "\mathcal{M}[\phi_{\tilde{C}}]"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\
\mathcal{M}'[S'] \arrow[r, "\mathcal{M}'[\phi_{C'}]"'] & \mathcal{M}'[C']
\end{tikzcd}$$

I can compute *abstraction error*:

$E_\alpha(S', C') = D_{JSD}(\alpha_{C'} \circ \mathcal{M}[\phi_{\tilde{C}}], \mathcal{M}'[\phi_{C'}] \circ \alpha_{S'})$

Since there are not other subsets this is also the *overall abstraction error*:

$e_\alpha = E_\alpha(S', C')$

# The *abstraction* approach: summary

Given:

- A low-level model $\mathcal{M}$;
- A high-level model $\mathcal{M}'$;
- An abstraction $(R, a, \alpha)$

a **zero-error abstraction** guarantees that, under intervention, if I:

- work (mechanism) at low-level and then switch (abstract) to high-level,
- or, switch first to high-level and then work there,

I will observe the same statistical behavior in the two models.

# 4. Research directions

# A quick comparison of the approaches [9]

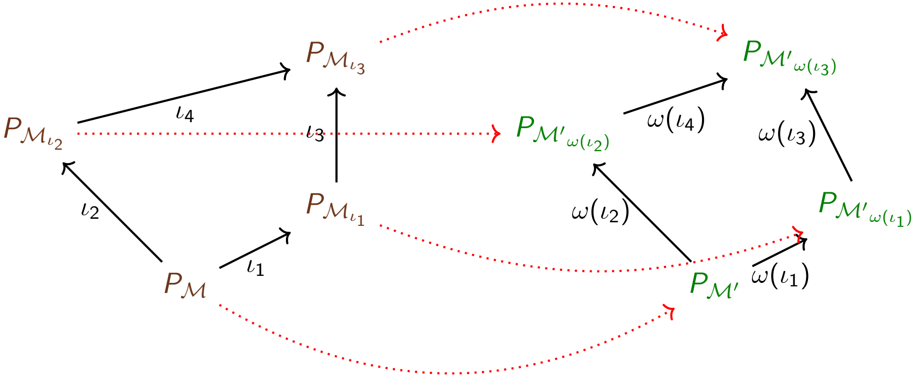
## Transformation approach

- Given: $\mathcal{M}, \mathcal{M}', \mathcal{I}, \omega$,
- a transformation is $\tau$.
- Consistency wrt *intervention-transformation*.
- Concerned with *distributional* information only (*structural* mediated through interventions).
- Works with *continuous models*.
- Consistency wrt to a limited *set of interventions*.

## Abstraction approach

- Given: $\mathcal{M}, \mathcal{M}'$,
- an abstraction is $(R, a, \alpha)$.
- Consistency wrt to *(intervened) mechanism-abstraction*.
- Concerned with *structural* and *distributional* information.
- Works with *finite models*.
- Consistency wrt to *all interventions* (in a finite set).

# Learning transformations




- A *single map* ($\tau$) across multiple distributions.
- *Transport problem?*

# Learning abstractions

$$
\begin{tikzcd}
\mathcal{M}[S] \arrow[r, "\mathcal{M}[\phi_{\tilde{C}}]"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\
\mathcal{M}'[S'] \arrow[r, "\mathcal{M}'[\phi_{\tilde{C}'}]"'] & \mathcal{M}'[C']
\end{tikzcd}
\quad
\begin{tikzcd}
\mathcal{M}[S] \arrow[r, "\mathcal{M}[\phi_{T}]"] \arrow[d, "\alpha_{S'}"'] & \mathcal{M}[T] \arrow[d, "\alpha_{T'}"] \\
\mathcal{M}'[S'] \arrow[r, "\mathcal{M}'[\phi_{T'}]"'] & \mathcal{M}'[T']
\end{tikzcd}
\quad
\begin{tikzcd}
\mathcal{M}[T] \arrow[r, "\mathcal{M}[\phi_{C}]"] \arrow[d, "\alpha_{T'}"'] & \mathcal{M}[C] \arrow[d, "\alpha_{C'}"] \\
\mathcal{M}'[T'] \arrow[r, "\mathcal{M}'[\phi_{C'}]"'] & \mathcal{M}'[C']
\end{tikzcd}
$$

- *Multiple maps* between intervened variables
- *Combinatorial problem*

# Other directions

Other topics:

- Compositionality [8, 7, 6]
- Counterfactual consistency
- Locality
- Other formalizations [2, 1, 3]
- Optimal criteria for learning [10]
- Transferring knowledge between models

# Thanks!

Thank you for listening!

# References I

[1] Sander Beckers, Frederick Eberhardt, and Joseph Y Halpern. Approximate causal abstractions. In *Uncertainty in Artificial Intelligence*, pages 606–615. PMLR, 2020.

[2] Sander Beckers and Joseph Y Halpern. Abstracting causal models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2678–2685, 2019.

[3] Jun Otsuka and Hayato Saigo. On the equivalence of causal models: A category-theoretic approach. *arXiv preprint arXiv:2201.06981*, 2022.

[4] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[5] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: Foundations and learning algorithms*. MIT Press, 2017.

# References II

[6] Eigil F Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. *arXiv preprint arXiv:2103.15758*, 2021.

[7] Eigil Fjeldgren Rischel. The category theory of causal models. 2020.

[8] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. In *33rd Conference on Uncertainty in Artificial Intelligence (UAI 2017)*, pages 808–817. Curran Associates, Inc., 2017.

[9] Fabio Massimo Zennaro. Abstraction between structural causal models: A review of definitions and properties. In *UAI 2022 Workshop on Causal Representation Learning*, 2022.

[10] Fabio Massimo Zennaro, Paolo Turrini, and Theo Damoulas. Towards computing an optimal abstraction for structural causal models. In *UAI 2022 Workshop on Causal Representation Learning*, 2022.