Structural Causal Models, Abstraction, and Learning

Fabio Massimo Zennaro

University of Bergen

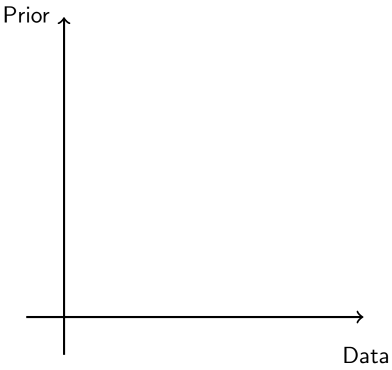
October 7, 2024

1. Structural Causal Modelling

Modelling

Assume we want to model a *system*.

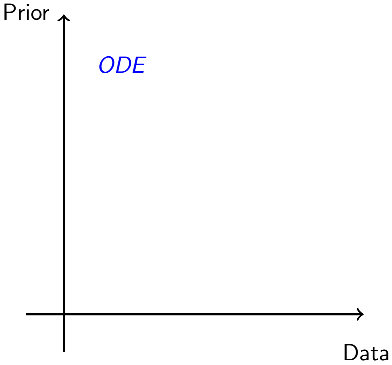
Different types of model will negotiate a trade-off between priors and data:



Modelling

Assume we want to model a *system*.

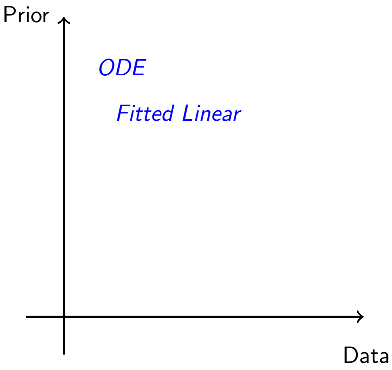
Different types of model will negotiate a trade-off between priors and data:



Modelling

Assume we want to model a *system*.

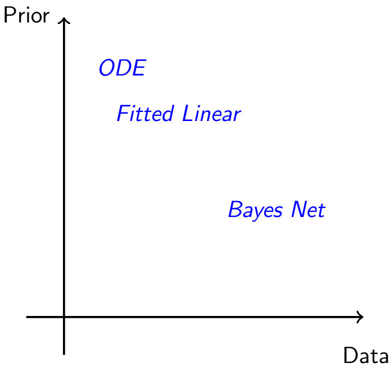
Different types of model will negotiate a trade-off between priors and data:



Modelling

Assume we want to model a *system*.

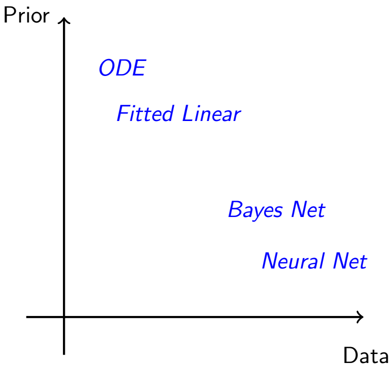
Different types of model will negotiate a trade-off between priors and data:



Modelling

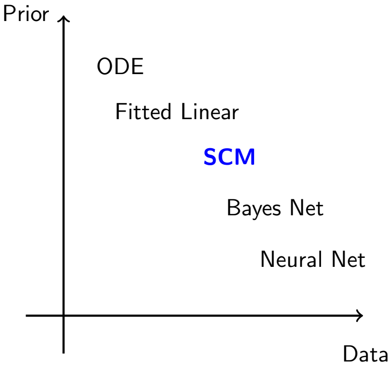
Assume we want to model a *system*.

Different types of model will negotiate a trade-off between priors and data:



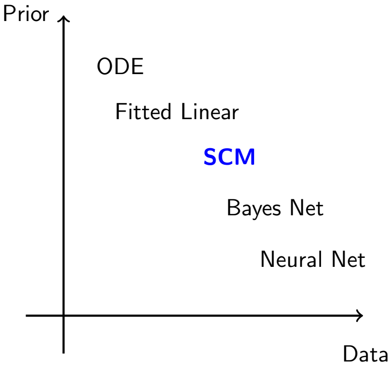
Structural Causal Modeling

Structural causal models rely on a strong prior given by *causality* [14, 15].



Structural Causal Modeling

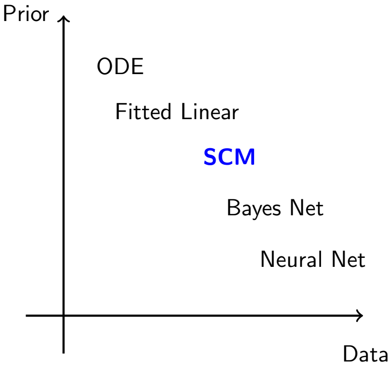
Structural causal models rely on a strong prior given by *causality* [14, 15].



- It discriminates *correlations* and *causes*.

Structural Causal Modeling

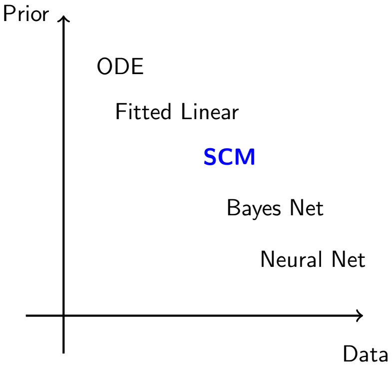
Structural causal models rely on a strong prior given by *causality* [14, 15].



- It discriminates *correlations* and *causes*.
- It allows for reasoning about *interventions*.

Structural Causal Modeling

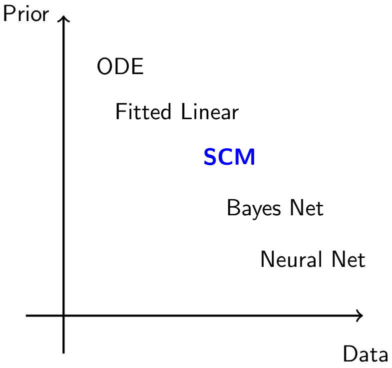
Structural causal models rely on a strong prior given by *causality* [14, 15].



- It discriminates *correlations* and *causes*.
- It allows for reasoning about *interventions*.
- It allows for reasoning about *counterfactuals*.

Structural Causal Modeling

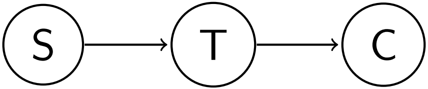
Structural causal models rely on a strong prior given by *causality* [14, 15].



- It discriminates *correlations* and *causes*.
- It allows for reasoning about *interventions*.
- It allows for reasoning about *counterfactuals*.
- It implies a *causality ladder* of reasoning.

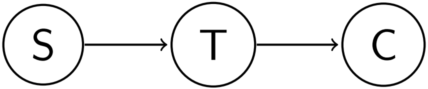
A Motivating Example

SCMs represent **causal systems**.



A Motivating Example

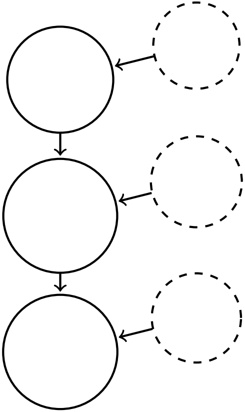
SCMs represent **causal systems**.



SCMs integrates a *graphical model* and *probabilities distributions*.

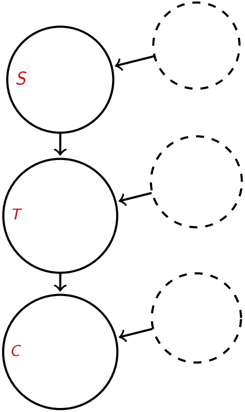
Structural Causal Models (SCMs) - Definition

We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:



Structural Causal Models (SCMs) - Definition

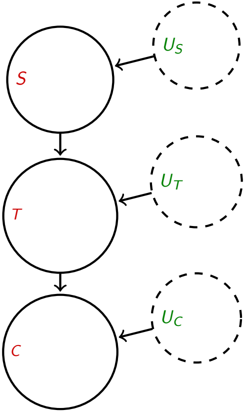
We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:



- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest

Structural Causal Models (SCMs) - Definition

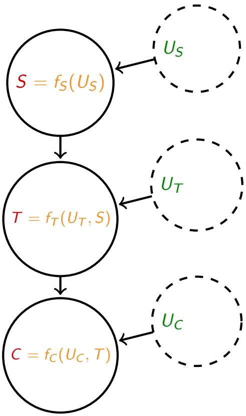
We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:



- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{U}$: Set of *exogenous nodes* (U_S, U_T, U_C) representing stochastic factors

Structural Causal Models (SCMs) - Definition

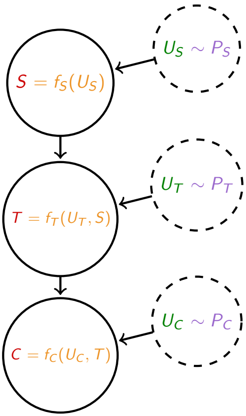
We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:



- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{U}$: 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

Structural Causal Models (SCMs) - Definition

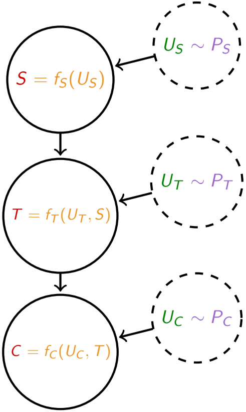
We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:



- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{U}$: 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

Structural Causal Models (SCMs) - Definition

We express a **SCM** as $\mathcal{M} = \langle \mathcal{X}, \mathcal{U}, \mathcal{F}, \mathcal{P} \rangle$ [14, 15]:

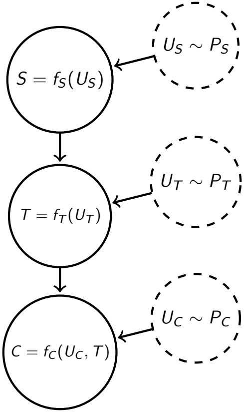


- $\mathcal{X}$: set of *endogenous nodes* (S, T, C) representing variables of interest
- $\mathcal{U}$: Set of *exogenous nodes* (U_S, U_T, U_C) representing stochastic factors
- $\mathcal{F}$: Set of *structural functions* (f_S, f_T, f_C) describing the dynamics of each variable
- $\mathcal{P}$: Set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

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

Structural Causal Models (SCMs) - Interventions

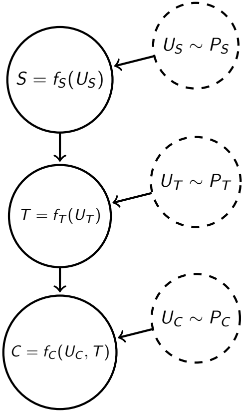
We can perform **interventions** on a causal model [14, 15]:



1.
2.

Structural Causal Models (SCMs) - Interventions

We can perform **interventions** on a causal model [14, 15]:

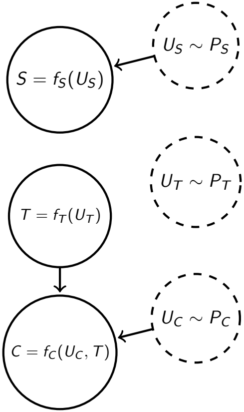


$do(T = 1)$

1.
2.

Structural Causal Models (SCMs) - Interventions

We can perform **interventions** on a causal model [14, 15]:

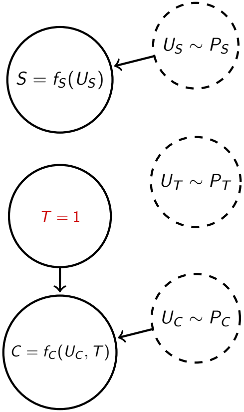


$do(T = 1)$

1. Remove incoming edges in the intervened node
2.

Structural Causal Models (SCMs) - Interventions

We can perform **interventions** on a causal model [14, 15]:



$do(T = 1)$

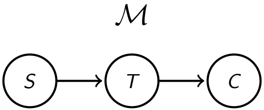
1. Remove incoming edges in the intervened node
2. Set the value of the intervened node

Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.



Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

$\mathcal{M}$

$$S \longrightarrow T \longrightarrow C$$

$\mathcal{M}_\iota$

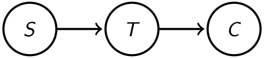
$$S \qquad 1 \longrightarrow C$$

$P_{\mathcal{M}}$

Structural Causal Models (SCMs) - Distributions

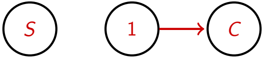
An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

$\mathcal{M}$



$P_{\mathcal{M}}$

$\mathcal{M}_\iota$

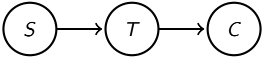


$P_{\mathcal{M}_\iota}$

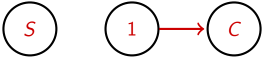
Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

$\mathcal{M}$



$\mathcal{M}_\iota$



$P_{\mathcal{M}}$

$P_{\mathcal{M}_\iota}$

$P_{\mathcal{M}}(C|S)$

Structural Causal Models (SCMs) - Distributions

An *intervention* ι defines a new **intervened model** $\mathcal{M}_\iota$ with new distributions.

$\mathcal{M}$	$\mathcal{M}_\iota$
$S \longrightarrow T \longrightarrow C$	$S \qquad 1 \longrightarrow C$
$P_{\mathcal{M}}$	$P_{\mathcal{M}_\iota}$
$P_{\mathcal{M}}(C\|S)$	$P_{\mathcal{M}_\iota}(C\|S) = P_{\mathcal{M}}(C\|S, do(T = 1))$

2. Causal Abstraction

Levels of Abstraction

Systems may be represented at different **levels of abstraction** (LoA) [7].

Levels of Abstraction

Systems may be represented at different **levels of abstraction** (LoA) [7].

Thermodynamics example:

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

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

Levels of Abstraction

Systems may be represented at different **levels of abstraction** (LoA) [7].

Thermodynamics example:

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

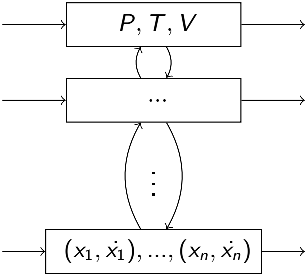
High-level / Abstracted model:
Macroscopic description P, T, V.

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*.

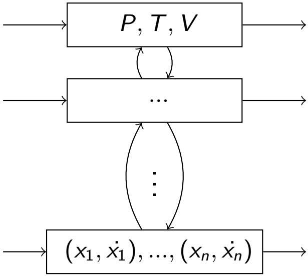
Abstraction

Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



Abstraction

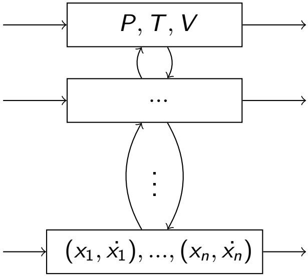
Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



- It combines models from *different sources*.

Abstraction

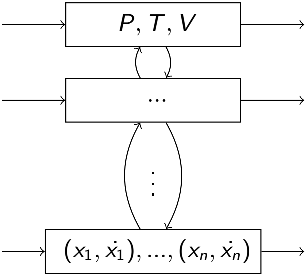
Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



- It combines models from *different sources*.
- It aggregates information from *different resolutions*.

Abstraction

Abstraction (aka, *multi-level modelling* or *multi-resolution modelling*) aims at relating these levels.



- It combines models from *different sources*.
- It aggregates information from *different resolutions*.
- It allows for *computation with minimal effort*.

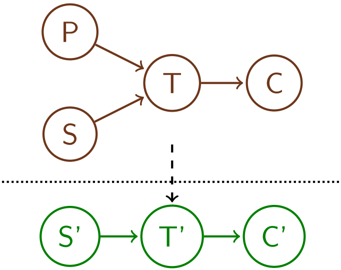
A Motivating Example

Lung cancer scenario example:

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

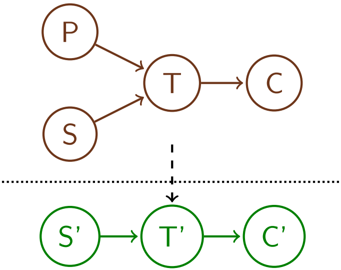
A Motivating Example

Lung cancer scenario example:



A Motivating Example

Lung cancer scenario example:

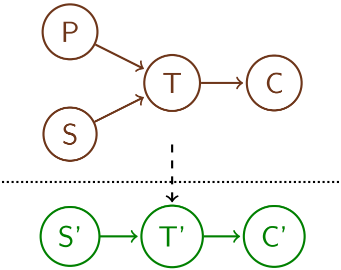


All vars in $0, 1$

All vars in $0, 1$

A Motivating Example

Lung cancer scenario example:



All vars in $0, 1$

All vars in $0, 1$

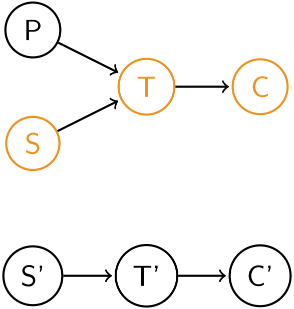
- The *transformation* approach [18, 2]
- The α-**abstraction** approach [17, 16]
- The Φ-*abstraction* approach [12, 13]

Causal Abstractions (CAs) - Definition

An α-**abstraction** $\langle R, a, \alpha_i \rangle$ [17, 16] is defined as:

Causal Abstractions (CAs) - Definition

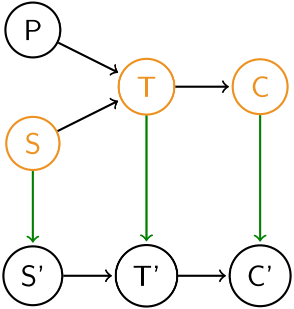
An α-**abstraction** $\langle R, a, \alpha_i \rangle$ [17, 16] is defined as:



- R: a set of *relevant variables*;

Causal Abstractions (CAs) - Definition

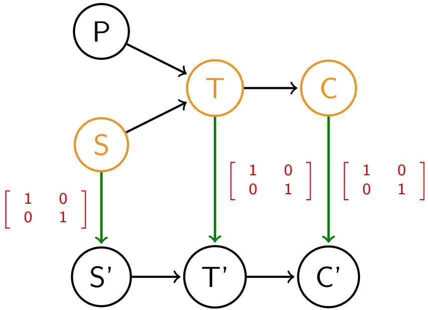
An **α-abstraction** $\langle R, a, \alpha_i \rangle$ [17, 16] is defined as:



- R: a set of *relevant variables*;
- a: a surjective function between *variables*;

Causal Abstractions (CAs) - Definition

An α-**abstraction** $\langle R, a, \alpha_i \rangle$ [17, 16] is defined as:



- R: a set of *relevant variables*;
- a: a surjective function between *variables*;
- α_i: a collection of surjective functions between *outcomes*.

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

.

$$S' \xrightarrow[P_{\mathcal{M}'_{\iota'}}(T'|do(S'))]{\nu} T'$$

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

$$\begin{tikzcd}
S \arrow[d, "\alpha_{S'}"'] & \\
S' \arrow[r, "\nu", "P_{\mathcal{M}'_{\iota'}}(T'|do(S'))"'] & T'
\end{tikzcd}$$

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

$$
\begin{tikzcd}
S \arrow[r, "\mu", "P_{\mathcal{M}_\iota}(T|do(S))"'] \arrow[d, "\alpha_{S'}"'] & T \\
S' \arrow[r, "\nu", "P_{\mathcal{M'}_{\iota'}}(T'|do(S'))"'] & T'
\end{tikzcd}
$$

.

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

$$\begin{tikzcd} S \arrow[r, "\mu", "P_{\mathcal{M}_\iota}(T|do(S))"'] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu", "P_{\mathcal{M'}_{\iota'}}(T'|do(S'))"'] & T' \end{tikzcd}$$

.

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

$$
\begin{tikzcd}
S \arrow[r, "\mu", "P_{\mathcal{M}_\iota}(T|do(S))"'] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\
S' \arrow[r, "\nu", "P_{\mathcal{M}'_{\iota'}}(T'|do(S'))"'] & T'
\end{tikzcd}
$$

- Ideally, mechanisms and abstractions *commute*.

Causal Abstractions (CAs) - Consistency

We want an abstraction to guarantee *interventional consistency*.

$$\begin{tikzcd}[column sep=huge, row sep=huge]
S \arrow[r, "\mu", "P_{\mathcal{M}_\iota}(T|do(S))"'] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\
S' \arrow[r, "\nu", "P_{\mathcal{M'}_{\iota'}}(T'|do(S'))"'] & T'
\end{tikzcd}$$

- Ideally, mechanisms and abstractions *commute*.
- Otherwise, we compute an abstraction error as the *worst-case discrepancy* over all possible interventions:

$$E_{\boldsymbol{\alpha}}(S', T') = \max_{\iota} D(\alpha_{T'} \cdot \mu, \nu \cdot \alpha_{S'})$$

Causal Abstractions (CAs) - Abstraction Error

An abstraction implies multiple *causal mechanism diagrams*:

Causal Abstractions (CAs) - Abstraction Error

An abstraction implies multiple *causal mechanism diagrams*:

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd}$$

Causal Abstractions (CAs) - Abstraction Error

An abstraction implies multiple *causal mechanism diagrams*:

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

Causal Abstractions (CAs) - Abstraction Error

An abstraction implies multiple *causal mechanism diagrams*:

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

Causal Abstractions (CAs) - Abstraction Error

An abstraction implies multiple *causal mechanism diagrams*:

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

A **(global) abstraction error** [17] $e(\boldsymbol{\alpha})$ is the maximum abstraction error over all diagrams.

$$e(\boldsymbol{\alpha}) = \sup_{\mathbf{X}', \mathbf{Y}' \subseteq \mathcal{X}'} E_{\boldsymbol{\alpha}}(\mathbf{X}', \mathbf{Y}')$$

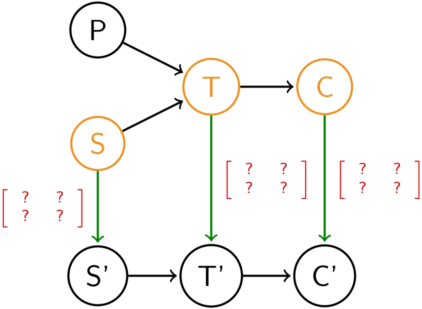
3. Abstraction Learning

Joint work of FMZ, M. Drávucz, G. Apachitei, W.D. Widanage and T. Damoulas

Problem statement [22]

Given a partially define *abstraction* $\boldsymbol{\alpha}$ in terms of $\langle R, a \rangle$ can I learn α_i as:

$$\min_{\boldsymbol{\alpha}} \; e(\boldsymbol{\alpha})$$



Challenges [22]

(i) *Multiple related problems*

$$\alpha_{S'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{T'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{C'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

Challenges [22]

(i) *Multiple related problems*

(ii) *Combinatorial optimization*

$$\alpha_{S'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{T'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{C'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

Challenges [22]

(i) *Multiple related problems*

(ii) *Combinatorial optimization*

(iii) *Surjectivity constraints*

$$\alpha_{S'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{T'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{C'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

Challenges [22]

$$\alpha_{S'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{T'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}, \alpha_{C'} = \begin{bmatrix} ? & ? \\ ? & ? \end{bmatrix}$$

(i) *Multiple related problems*

(ii) *Combinatorial optimization*

(iii) *Surjectivity constraints*

$$\begin{tikzcd} S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\ S' \arrow[r, "\nu"] & T' \end{tikzcd} \qquad \begin{tikzcd} T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\ T' \arrow[r, "\nu'"] & C' \end{tikzcd}$$

$$\begin{tikzcd} S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\ S' \arrow[r, "\nu' \circ \nu"] & C' \end{tikzcd}$$

Baselines: parallel or sequential approaches.

Relaxation and parametrization [22]

We address *(ii) combinatorial optimization* by *relaxing* and *parametrizing* all α_i.

$$\min_{\boldsymbol{\alpha}(\mathbf{W})} e(\boldsymbol{\alpha}(\mathbf{W}))$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in \mathbb{R}^{2\times 2}$$

$$\begin{bmatrix} 0.7 & 1.2 \\ -0.2 & 3.3 \end{bmatrix}$$

Relaxation and parametrization [22]

We address *(ii) combinatorial optimization* by *relaxing* and *parametrizing* all α_i.

$$\min_{\boldsymbol{\alpha}(\mathbf{W})} \; e(\boldsymbol{\alpha}(\mathbf{W}))$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in \mathbb{R}^{2\times 2}$$

$$\begin{bmatrix} 0.7 & 1.2 \\ -0.2 & 3.3 \end{bmatrix}$$

We add *tempering* $t(W) = \frac{e^{\frac{W_{ij}}{T}}}{\sum_i e^{\frac{W_{ij}}{T}}}$ along the matrix columns to binarize them.

$$\mathcal{L}_1 : \min_{\boldsymbol{\alpha}(\mathbf{W})} \; e(\boldsymbol{\alpha}(t(\mathbf{W})))$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in [0,1]^{2\times 2}$$

$$t\left(\begin{bmatrix} 0.7 & 1.2 \\ -0.2 & 3.3 \end{bmatrix}\right) = \begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix}$$

Enforcing surjectivity [22]

We address *(iii) surjective constraints* through a *penalty function*:

$$\mathcal{L}_2 : \min_{\mathbf{W}} \sum_{\mathbf{W}} \sum_{i} \left(1 - \max_{j} t(W)_{ij}\right)$$

$$\alpha_{S'}, \alpha_{T'}, \alpha_{C'} \in [0,1]^{2\times 2}$$

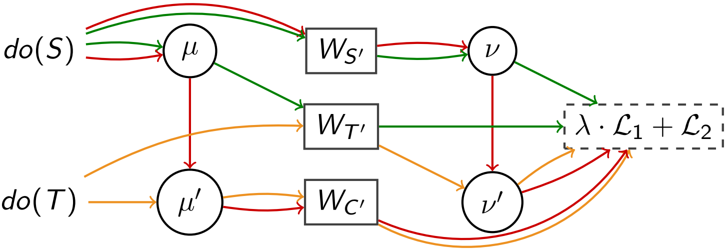
$$\begin{bmatrix} 0.99 & 0.02 \\ 0.01 & 0.98 \end{bmatrix} \overset{\mathcal{L}_2}{\rightsquigarrow}$$

$$(1-0.99)+(1-0.98)$$

Solution by gradient descent [22]

We address *(i) multiple related problems* by *jointly* solving all the problems via *gradient descent*:

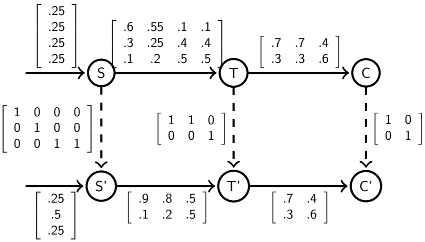
$$
\begin{tikzcd}
S \arrow[r, "\mu"] \arrow[d, "\alpha_{S'}"'] & T \arrow[d, "\alpha_{T'}"] \\
S' \arrow[r, "\nu"] & T'
\end{tikzcd}
\qquad
\begin{tikzcd}
T \arrow[r, "\mu'"] \arrow[d, "\alpha_{T'}"'] & C \arrow[d, "\alpha_{C'}"] \\
T' \arrow[r, "\nu'"] & C'
\end{tikzcd}
\qquad
\begin{tikzcd}
S \arrow[r, "\mu' \circ \mu"] \arrow[d, "\alpha_{S'}"'] & C \arrow[d, "\alpha_{C'}"] \\
S' \arrow[r, "\nu' \circ \nu"] & C'
\end{tikzcd}
$$

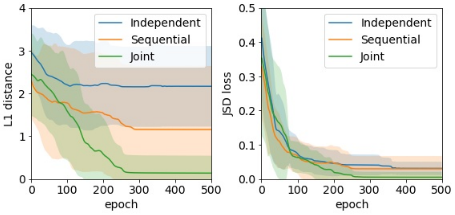


Synthetic Experiments [22]

We evaluated our learning method:

- On multiple synthetic models;
- Against *independent* and *sequential* approach;
- Monitoring *loss functions*, *L1-dist from ground truth*, *wall-clock time*.





Real-World Experiments [22]

We want to model the stage of **coating** in lithium-ion battery manufacturing:

$$\text{Mass Loading} = \text{f(input)}$$

Experiments are costly, so we want to integrate data[1] collected by two groups running similar (but not identical) experiments:

LRCS (France)	**WMG** (UK)
Collection of few statistics in each a few stages of battery manufacturing [**?**].	Collection of detailed space- and time-dependent measurements during coating.

[1] https://chemistry-europe.onlinelibrary.wiley.com/doi/full/10.1002/batt.201900135
https://github.com/mattdravucz/jointly-learning-causal-abstraction/

Real-World Experiments [22]

We evaluated our learning method:

- Performing abstraction of data from base to abstracted (WMG $\rightarrow$ LRCS);
- Evaluating change in performance using aggregated data when predicting *out-of-sample* (k).

	Training set	**Test Set**	**MSE**
(a)	LRCS[$CG \neq k$]	LRCS[$CG = k$]	1.86 ± 1.75
(b)	LRCS[$CG \neq k$] + WMG	LRCS[$CG = k$]	0.22 ± 0.26
(c)	LRCS[$CG \neq k$] + WMG[$CG \neq k$]	LRCS[$CG = k$] + WMG[$CG = k$]	1.22 ± 0.95

4. Current Developments

Current Developments

Causality and *abstraction* may both play important role in modelling.

Current Developments

Causality and *abstraction* may both play important role in modelling.

Large space for conceptual and practical development of **causal abstraction frameworks**:

Current Developments

Causality and *abstraction* may both play important role in modelling.

Large space for conceptual and practical development of **causal abstraction frameworks**:

1. *Foundations* of the framemorks
   - Category theory [13]
   - Measure theory [3]
   - Review [20]

Current Developments

Causality and *abstraction* may both play important role in modelling.

Large space for conceptual and practical development of **causal abstraction frameworks**:

1. *Foundations* of the framemorks
   - Category theory [13]
   - Measure theory [3]
   - Review [20]
2. *Characterization* of these frameworks
   - Measures of abstraction [23]
   - Abstraction with soft interventions [10]
   - Cluster DAGs and do-calculus [1]
   - Causal bandits and abstraction [21]

Current Developments

3. *Algorithmic and empirical* development
   - Learning with optimal transport [6]
   - Learning linear abstraction [11]
   - Target learning [9]
   - Neural models [19]

Current Developments

3. *Algorithmic and empirical* development
   - Learning with optimal transport [6]
   - Learning linear abstraction [11]
   - Target learning [9]
   - Neural models [19]
4. *Applications* of causal abstraction
   - Surrogate for agent-based models [5]
   - Explainable AI [8]
   - Visual coarsening [4]

Current Developments

3. *Algorithmic and empirical* development
   - Learning with optimal transport [6]
   - Learning linear abstraction [11]
   - Target learning [9]
   - Neural models [19]
4. *Applications* of causal abstraction
   - Surrogate for agent-based models [5]
   - Explainable AI [8]
   - Visual coarsening [4]

And connections to *causal representation learning*, *reinforcement learning*...

Current Developments

3. *Algorithmic and empirical* development
   - Learning with optimal transport [6]
   - Learning linear abstraction [11]
   - Target learning [9]
   - Neural models [19]
4. *Applications* of causal abstraction
   - Surrogate for agent-based models [5]
   - Explainable AI [8]
   - Visual coarsening [4]

And connections to *causal representation learning*, *reinforcement learning*...

More about causal abstraction:

https://github.com/FMZennaro/CausalAbstraction/

Thanks!

Thank you for listening!

References I

[1] G Anand, Swarnava Ghosh, Liwei Zhang, Angesh Anupam, Colin L Freeman, Christoph Ortner, Markus Eisenbach, and James R Kermode. Exploiting machine learning in multiscale modelling of materials. *Journal of The Institution of Engineers (India): Series D*, pages 1–11, 2022.

[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] Simon Buchholz, Junhyung Park, and Bernhard Schölkopf. Products, abstractions and inclusions of causal spaces. *arXiv preprint arXiv:2406.00388*, 2024.

[4] Krzysztof Chalupka, Pietro Perona, and Frederick Eberhardt. Visual causal feature learning. *arXiv preprint arXiv:1412.2309*, 2014.

References II

[5] Joel Dyer, Nicholas Bishop, Yorgos Felekis, Fabio Massimo Zennaro, Anisoara Calinescu, Theodoros Damoulas, and Michael Wooldridge. Interventionally consistent surrogates for agent-based simulators. *arXiv preprint arXiv:2312.11158*, 2023.

[6] Yorgos Felekis, Fabio Massimo Zennaro, Nicola Branchini, and Theodoros Damoulas. Causal optimal transport of abstractions. *arXiv preprint arXiv:2312.08107*, 2023.

[7] Luciano Floridi. The method of levels of abstraction. *Minds and machines*, 18:303–329, 2008.

[8] Atticus Geiger, Hanson Lu, Thomas Icard, and Christopher Potts. Causal abstractions of neural networks. *Advances in Neural Information Processing Systems*, 34:9574–9586, 2021.

[9] Armin Kekić, Bernhard Schölkopf, and Michel Besserve. Targeted reduction of causal models. *arXiv preprint arXiv:2311.18639*, 2023.

References III

[10] Riccardo Massidda, Atticus Geiger, Thomas Icard, and Davide Bacciu. Causal abstraction with soft interventions. *arXiv preprint arXiv:2211.12270*, 2022.

[11] Riccardo Massidda, Sara Magliacane, and Davide Bacciu. Learning causal abstractions of linear structural causal models. *arXiv preprint arXiv:2406.00394*, 2024.

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

[13] Jun Otsuka and Hayato Saigo. The process theory of causality: an overview. 2022.

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

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

References IV

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

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

[18] 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.

[19] Kevin Xia and Elias Bareinboim. Neural causal abstractions. *arXiv preprint arXiv:2401.02602*, 2024.

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

References V

[21] Fabio Massimo Zennaro, Nicholas George Bishop, Joel Dyer, Yorgos Felekis, Ani Calinescu, Michael J Wooldridge, and Theodoros Damoulas. Causally abstracted multi-armed bandits. In *The 40th Conference on Uncertainty in Artificial Intelligence*, 2024.

[22] Fabio Massimo Zennaro, Máté Drávucz, Geanina Apachitei, W. Dhammika Widanage, and Theodoros Damoulas. Jointly learning consistent causal abstractions over multiple interventional distributions. In *2nd Conference on Causal Learning and Reasoning*, 2023.

[23] Fabio Massimo Zennaro, Paolo Turrini, and Theo Damoulas. Quantifying consistency and information loss for causal abstraction learning. In *Proceedings of the Thrity-Second International Conference on International Joint Conferences on Artificial Intelligence*, 2023.