Abstraction between Structural Causal Models: A Review of Definitions and Properties

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

University of Warwick
August 5th, 2022

Abstraction

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

Thermodynamics example:

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

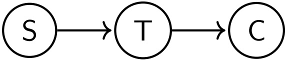
High-level / Abstracted model:
Macroscopic description P, T, V.

A LoA may be *inaccessible*, *inadequate* or *computationally expensive*, so we may want to *shift* among LoAs.

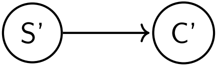
Can we move among LoAs reliably?

Abstraction for SCMs

Assume two *SCMs* at different *levels of abstractions*:



$$\mathcal{M} = \langle \mathcal{X}, \mathcal{E}, \mathcal{F}, \mathcal{P} \rangle$$



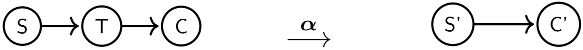
$$\mathcal{M}' = \langle \mathcal{X}', \mathcal{E}', \mathcal{F}', \mathcal{P}' \rangle$$

How do we relate two SCMs at different LoA?

1. How do we define a mapping $\boldsymbol{\alpha} : \mathcal{M} \rightarrow \mathcal{M}'$?
2. (How do we guarantee consistency between models?)

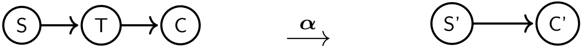
How do we define a mapping $\boldsymbol{\alpha} : \mathcal{M} \to \mathcal{M}'$?

A SCM carries a lot of data:



(Variables)	$\mathcal{X} = \{S, T, C\}$	$\mathcal{X}' = \{S', C'\}$
(Outcomes)	$\mathcal{M}[S] = M[T] = \mathcal{M}[C] = \{0, 1\}$	$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$
(Joint Outcomes)	$\mathcal{M}[\mathcal{X}] = \{0, 1\}^3$	$\mathcal{M}'[\mathcal{X}'] = \{0, 1\}^3$
(DAG)	$\mathcal{G}_{\mathcal{M}}$	$\mathcal{G}_{\mathcal{M}'}$
(Joint PMF)	$P_{\mathcal{M}}(S, T, C)$	$P_{\mathcal{M}'}(S', C')$

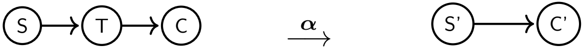
τ-ω transformation [5, 1]



(Variables)	$\mathcal{X} = \{S, T, C\}$		$\mathcal{X}' = \{S', C'\}$
(Outcomes)	$\mathcal{M}[S] = M[T] = \mathcal{M}[C] = \{0, 1\}$		$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$
(Joint Outcomes)	$\mathcal{M}[\mathcal{X}] = \{0, 1\}^3$	$\xrightarrow{\tau}$	$\mathcal{M}'[\mathcal{X}'] = \{0, 1\}^3$
(DAG)	$\mathcal{G}_{\mathcal{M}}$		$\mathcal{G}_{\mathcal{M}'}$
(Joint PMF)	$P_{\mathcal{M}}(S, T, C)$	$\overset{\tau_\#}{\dashrightarrow}$	$P_{\mathcal{M}'}(S', C')$

Second proposal

(R, a, α) *abstraction* [4, 3]



(Variables)	$\mathcal{X} = \{S, T, C\}$	$\overset{a}{\dashrightarrow}$	$\mathcal{X}' = \{S', C'\}$
(Outcomes)	$\mathcal{M}[S] = M[T] = \mathcal{M}[C] = \{0, 1\}$	$\overset{\alpha_{X'}}{\twoheadrightarrow}$	$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$
(Joint Outcomes)	$\mathcal{M}[\mathcal{X}] = \{0, 1\}^3$		$\mathcal{M}'[\mathcal{X}'] = \{0, 1\}^3$
(DAG)	$\mathcal{G}_{\mathcal{M}}$		$\mathcal{G}_{\mathcal{M}'}$
(Joint PMF)	$P_{\mathcal{M}}(S, T, C)$		$P_{\mathcal{M}'}(S', C')$

Problem definition

How do we define a mapping $\alpha : \mathcal{M} \rightarrow \mathcal{M}'$*?*

Large degree of freedom, several answers: [5, 1, 4, 3, 2].

Different answers imply:

- Different *understandings of abstraction*
- Different *constraints*
- Different *applications*

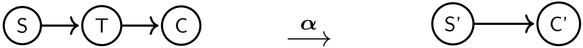
Framework

We want a framework to align and compare definitions

1. We distinguish *layers* at which we could define the mapping α
2. We consider what *formal properties* we can enforce on the mapping α
3. We evaluate what *forms of abstraction* are allowed

We can tailor definitions of abstraction around our understanding/need.

Layers of an abstraction



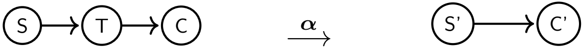
Structural layer:

(DAG)	$\mathcal{G}_{\mathcal{M}}$	$\mathcal{G}_{\mathcal{M}'}$
(Variables)	$\mathcal{X} = \{S, T, C\}$	$\mathcal{X}' = \{S', C'\}$

Distributional layer:

(Outcomes)	$\mathcal{M}[S] = M[T] = \mathcal{M}[C] = \{0, 1\}$	$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0, 1\}$
(Joint Outcomes)	$\mathcal{M}[\mathcal{X}] = \{0, 1\}^3$	$\mathcal{M}'[\mathcal{X}'] = \{0, 1\}^3$
(Joint PMF)	$P_{\mathcal{M}}(S, T, C)$	$P_{\mathcal{M}'}(S', C')$

Structural properties (nodes)

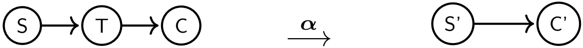


Structural layer:

(DAG)	$\mathcal{G}_{\mathcal{M}}$		$\mathcal{G}_{\mathcal{M}'}$
(Variables)	$\mathcal{X} = \{S, T, C\}$	$\xrightarrow{f}$	$\mathcal{X}' = \{S', C'\}$

Map f can be:

- *functional*: must every node in $\mathcal{M}$ be accounted for?
  (Yes [2]; no [4])
- *surjectivity*: is every node in $\mathcal{M}'$ explained by the micromodel?
  (Yes [4, 1]; no [2])
- *injectivity*: is coarsening of nodes to be prevented?



Structural layer:

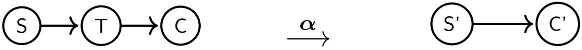
(DAG)	$\mathcal{G}_{\mathcal{M}}$	$\xrightarrow{F}$	$\mathcal{G}_{\mathcal{M}'}$
(Variables)	$\mathcal{X} = \{S, T, C\}$		$\mathcal{X}' = \{S', C'\}$

Map f can be:

- *functoriality*: must every edge and its directionality in $\mathcal{M}$ be preserved?

  (Yes [2]; no [4])
- *fulness*: is every edge in $\mathcal{M}'$ explained by the micromodel?
- *faithfulness*: is coarsening of edges to be prevented?

Distributional properties



Distributional layer:

(Outcomes)	$\mathcal{M}[S] = M[T] = \mathcal{M}[C] = \{0,1\}$	$\xrightarrow{f}$	$\mathcal{M}'[S'] = \mathcal{M}'[C'] = \{0,1\}$
(Joint Outcomes)	$\mathcal{M}[\mathcal{X}] = \{0,1\}^3$	$\xrightarrow{f}$	$\mathcal{M}'[\mathcal{X}'] = \{0,1\}^3$
(Joint PMF)	$P_{\mathcal{M}}(S,T,C)$		$P_{\mathcal{M}'}(S',C')$

Map f can be:

- *functional*: must every outcome in $\mathcal{M}$ be accounted for?
  (Yes [4])
- *surjectivity*: is every outcome in $\mathcal{M}'$ explained by the micromodel?
  (Yes [4]; no [5])
- *injectivity*: is coarsening of outcomes to be prevented?

Discussion

- We can distinguish aspects of abstraction at different layers (*coarsening*)
- We can reason about desirable properties and forms of abstraction

- We could discuss other formal properties (*bijectivity, macro-to-micro, stochasticity*)
- We need to discuss consistency properties

Thanks!

Thank you for listening!

References I

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

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

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

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

[5] 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.