Learning Structural Causal Models through Deep Generative Models

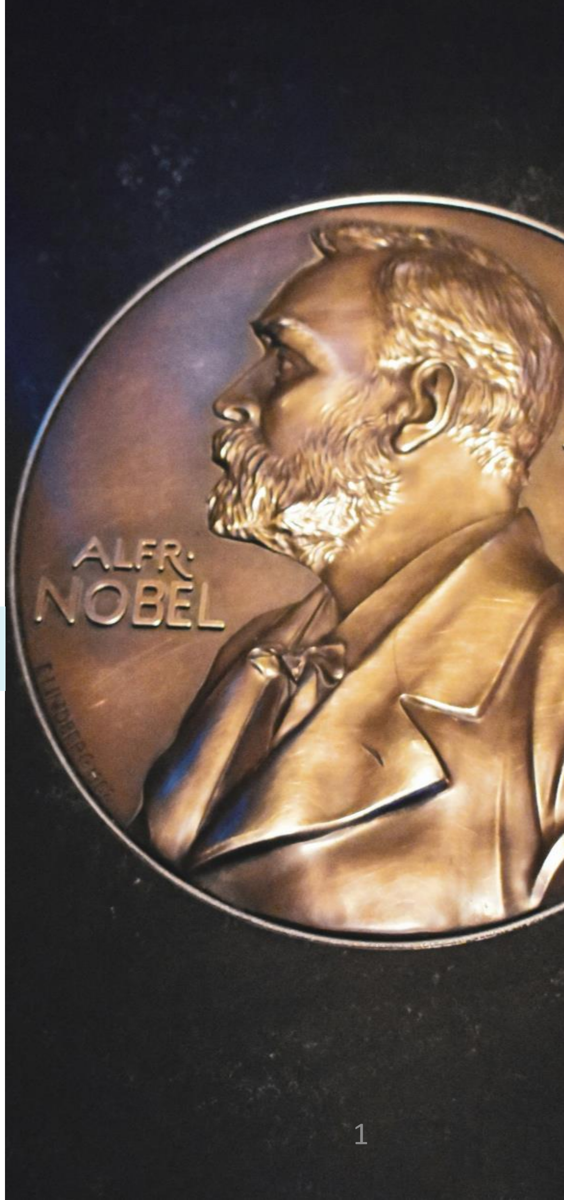
Survey on Methods, Guarantees, and Challenges

Audrey Poinsot, Ph.D. student
audrey.poinsot@ekimetrics.com

Causal Club, University of Pisa, April 2025

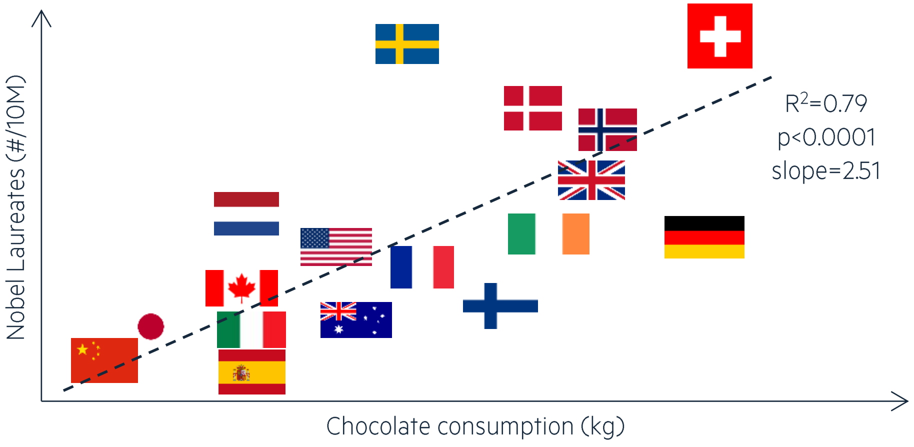
Motivation, correlation is not causation

Does chocolate make you smart?

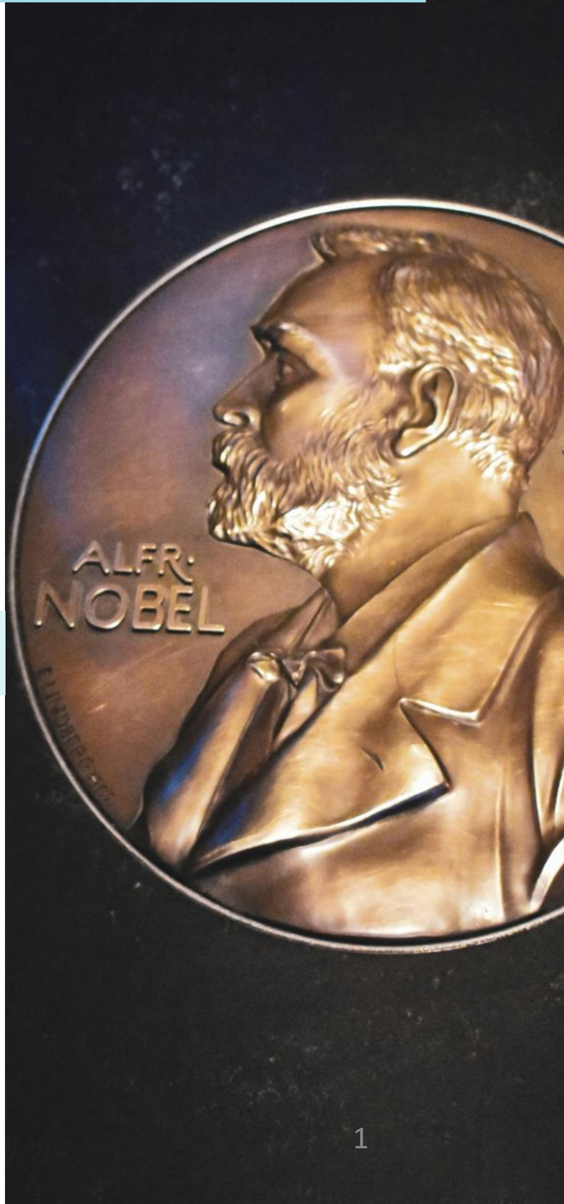


[Messerli, 2012]

Motivation, correlation is not causation

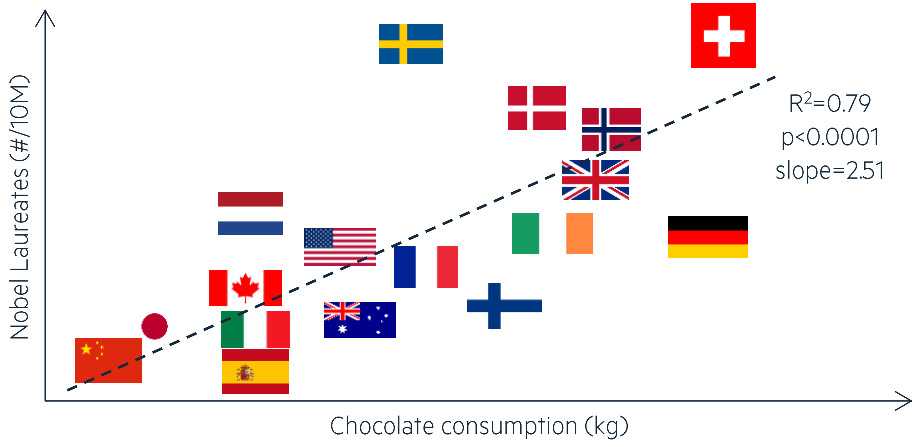


Does chocolate make you smart?

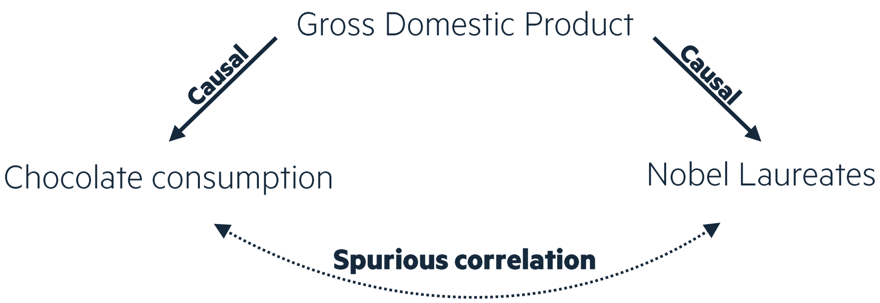


[Messerli, 2012]

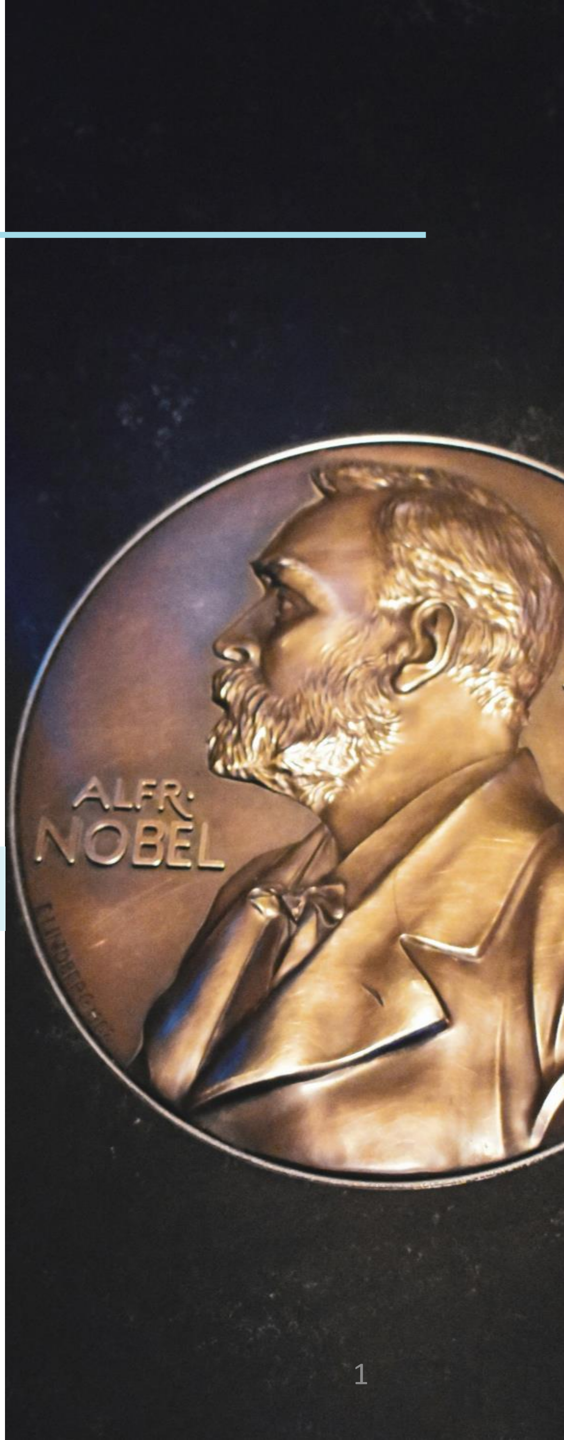
Motivation, correlation is not causation



Does chocolate make you smart?



[Messerli, 2012]

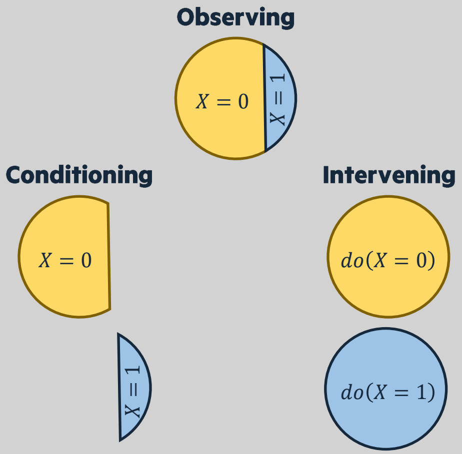


Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

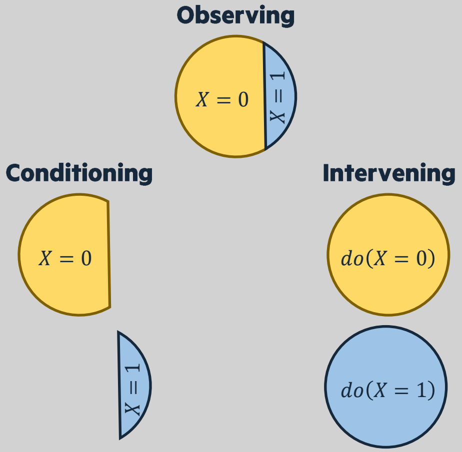


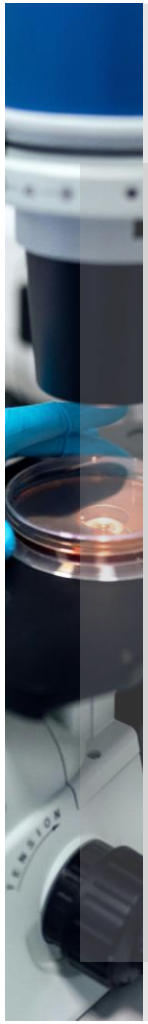
Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

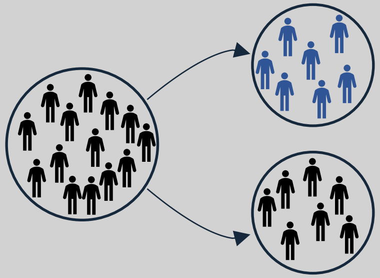




Expensive RCT

Randomized Control Trial (RCT)

Expensive, Unfair, ...

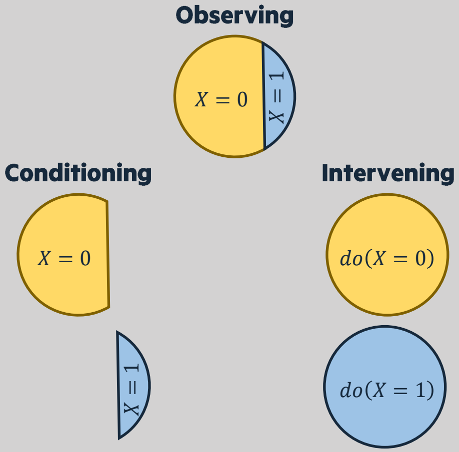


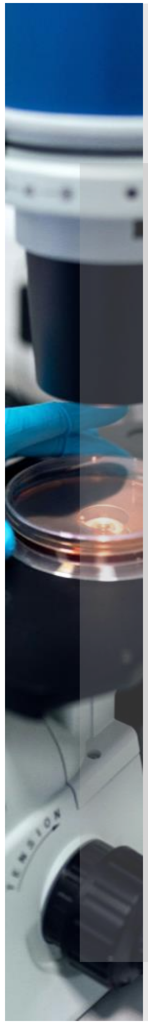
Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

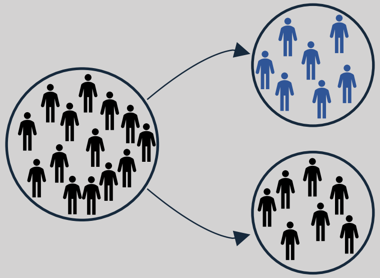




Expensive RCT

Randomized Control Trial (RCT)

Expensive, Unfair, ...



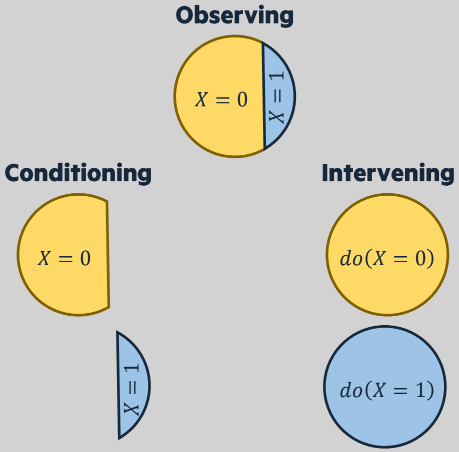
Dealing with **observational data** we need to translate interventions (do) into measurable quantities.

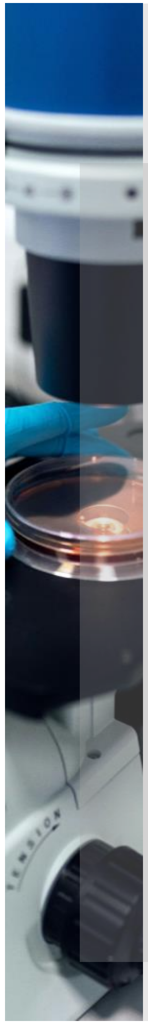
Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

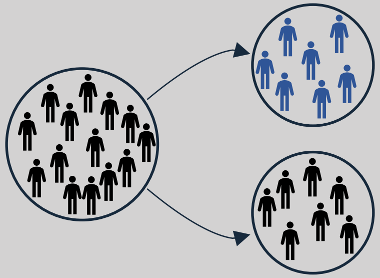




Expensive RCT

Randomized Control Trial (RCT)

Expensive, Unfair, ...



Dealing with **observational data**

we need to translate interventions (do) into measurable quantities.

Do calculus

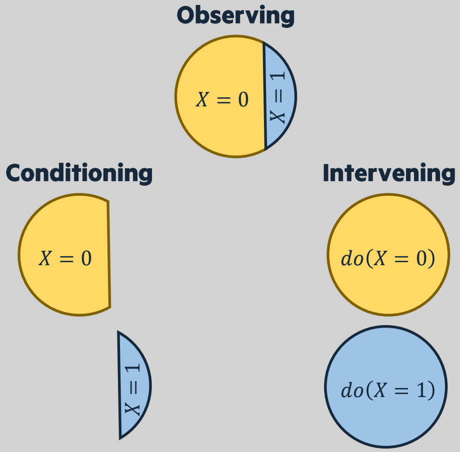
An axiomatic system for **replacing** probability formulas containing the **do operator with** ordinary **conditional probabilities**.

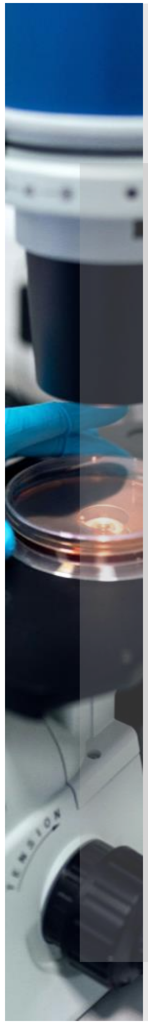
Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

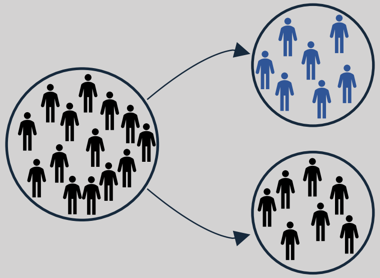




Expensive RCT

Randomized Control Trial (RCT)

Expensive, Unfair, …



Dealing with **observational data** we need to translate interventions (do) into measurable quantities.

Do calculus

An axiomatic system for **replacing** probability formulas containing the **do operator with** ordinary **conditional probabilities**.

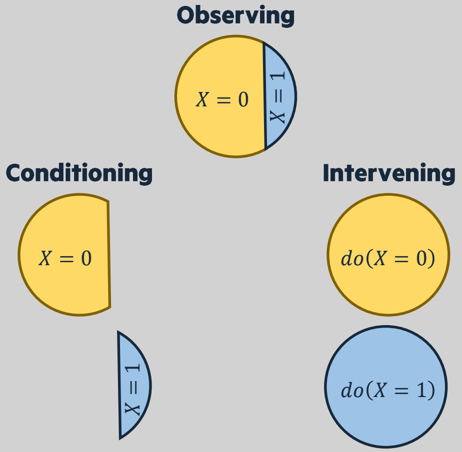
Does such a translation lead to a **unique result**? → With enough assumptions, yes.

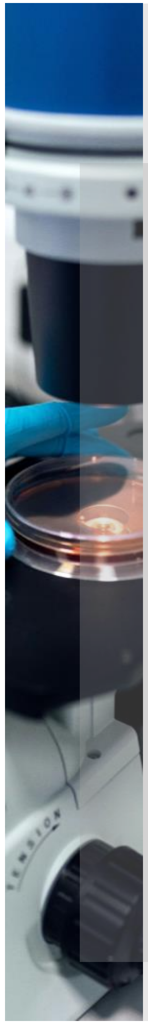
Motivation, Causal Inference to the rescue

Do operator

A **mathematical notation** to denote an **intervention/action**.

$do(X = x)$ sets the variable X to the value x **independently** of any other variable.

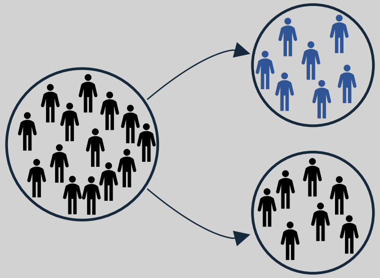




Expensive RCT

Randomized Control Trial (RCT)

Expensive, Unfair, ...



Dealing with **observational data** we need to translate interventions (do) into measurable quantities.

Do calculus

An axiomatic system for **replacing** probability formulas containing the **do operator with** ordinary **conditional probabilities**.

Does such a translation lead to a **unique result**? → With enough assumptions, yes.

Identifiability: A causal query Q is identifiable from a class of models $\mathbf{M}$ (i.e. set of assumptions) if for any pair of models $M_1, M_2 \in \mathbf{M}$, $Q(M_1) = Q(M_2)$.

[Pearl, 2009]

Motivation, towards finer decision-making

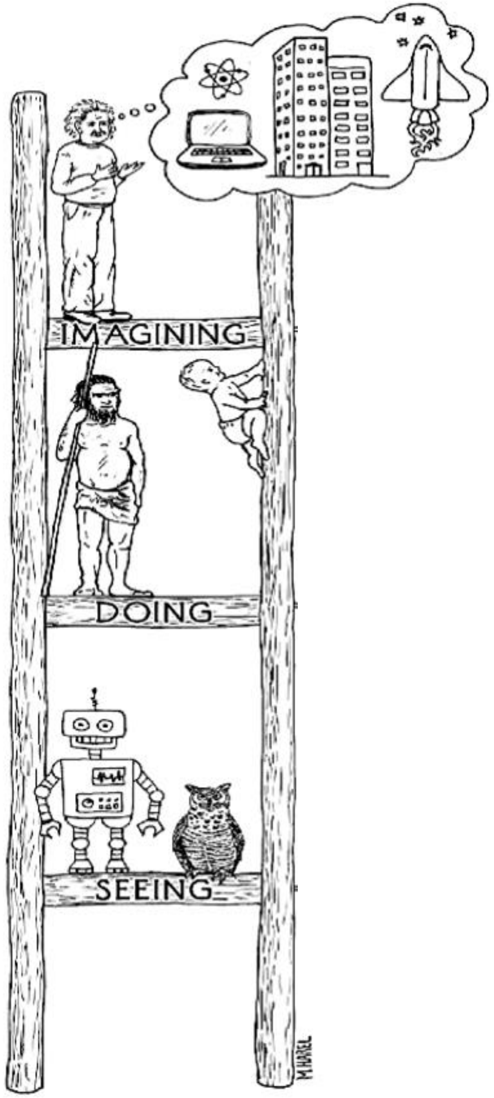
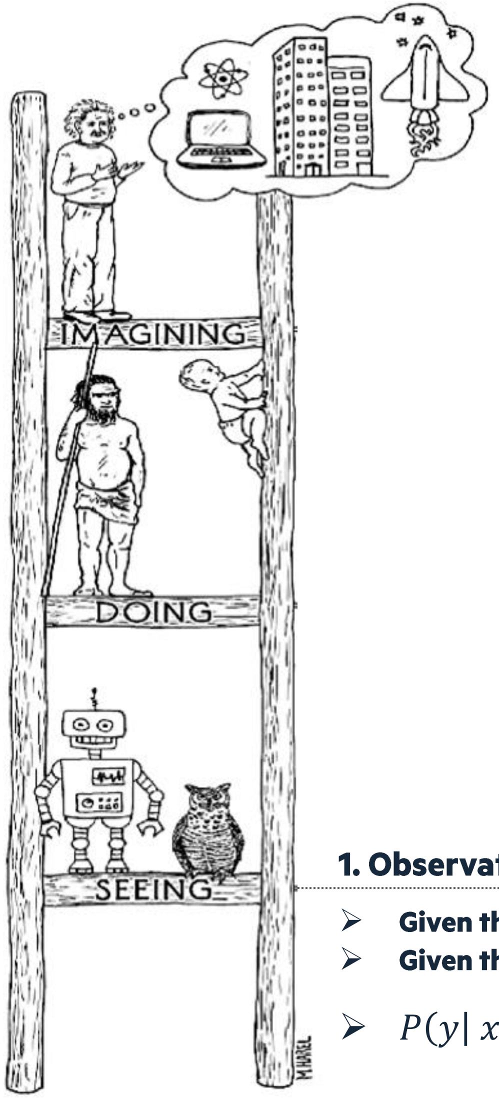


Figure inspired from [Pearl and Mackenzie, 2018]

Motivation, towards finer decision-making



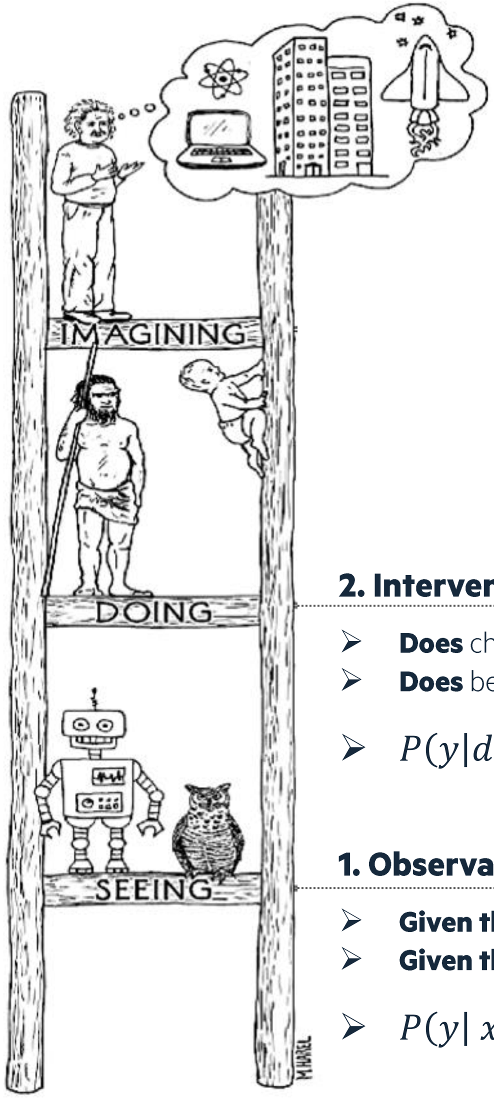
1. Observations – associations (L_1)

- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?

- $P(y|\, x)$

Figure inspired from [Pearl and Mackenzie, 2018]

Motivation, towards finer decision-making



2. Interventions – action-guidance (L_2)

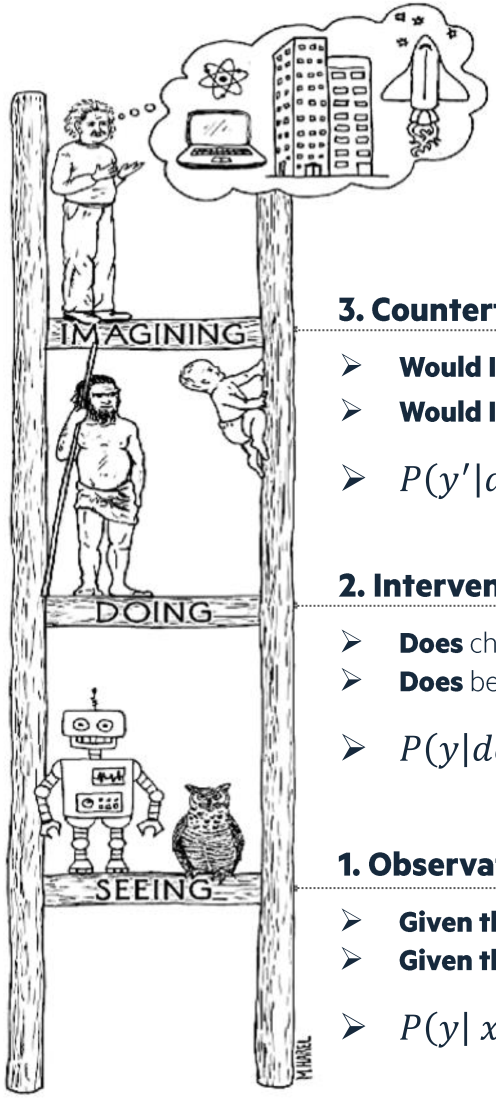
- **Does** chocolate **make you** smarter?
- **Does** being a woman **reduce your** chances of getting a loan?
- $P(y|do(x)) \rightarrow L_2$ assumptions (e.g. causal graph, adjustment sets)

1. Observations – associations (L_1)

- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?
- $P(y|\, x)$

Figure inspired from [Pearl and Mackenzie, 2018]

Motivation, towards finer decision-making



3. Counterfactuals – against existing observations (L_3)

- **Would I** have won the Nobel **if I had** eaten chocolate?
- **Would I** have got the loan **if I had** been a man?
- $P(y'|do(x'), y, x)$ → L_3 assumptions (e.g. exogeneous noise)

2. Interventions – action-guidance (L_2)

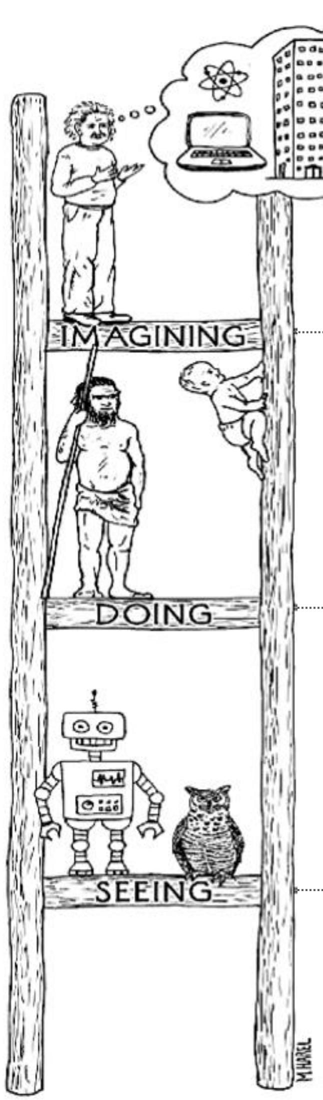
- **Does** chocolate **make you** smarter?
- **Does** being a woman **reduce your** chances of getting a loan?
- $P(y|do(x))$ → L_2 assumptions (e.g. causal graph, adjustment sets)

1. Observations – associations (L_1)

- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?
- $P(y|\, x)$

Figure inspired from [Pearl and Mackenzie, 2018]

Motivation, towards finer decision-making



3. Counterfactuals – against existing observations (L_3)

- **Would I** have won the Nobel **if I had** eaten chocolate?
- **Would I** have got the loan **if I had** been a man?
- $P(y'|do(x'), y, x)$ → L_3 assumptions (e.g. exogeneous noise)

2. Interventions – action-guidance (L_2)

- **Does** chocolate **make you** smarter?
- **Does** being a woman **reduce your** chances of getting a loan?
- $P(y|do(x))$ → L_2 assumptions (e.g. causal graph, adjustment sets)

1. Observations – associations (L_1)

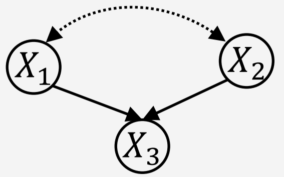
- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?
- $P(y|\, x)$

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$:
$F = \{X_i \; := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

Example

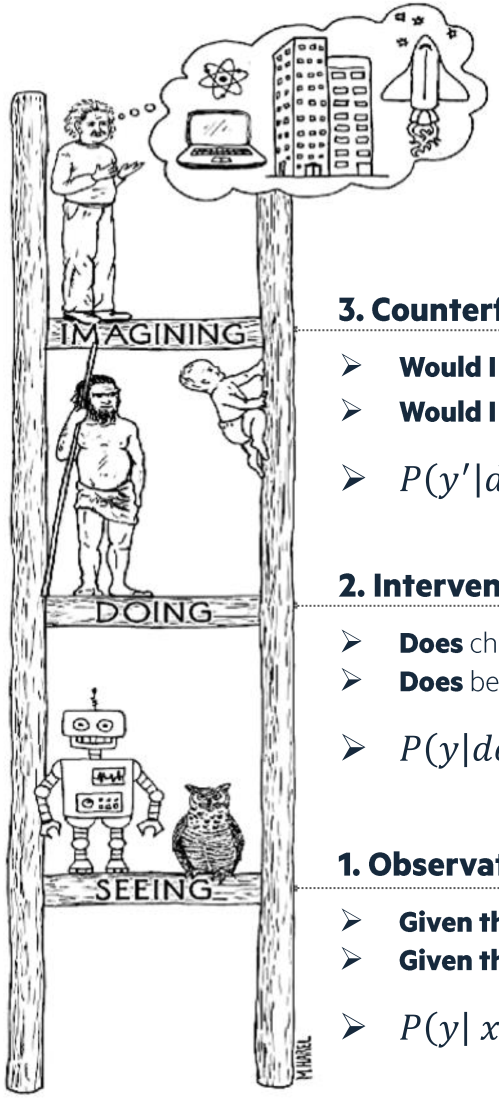


$$X_1 = f_1(U_1)$$
$$X_2 = f_2(U_2)$$
$$X_3 = f_3(X_1, X_2, U_3)$$

with $P(U)$ s.t. $U_3 \perp\!\!\!\perp U_1$ and $U_3 \perp\!\!\!\perp U_2$

Figure inspired from [Pearl and Mackenzie, 2018]

Motivation, towards finer decision-making



3. Counterfactuals – against existing observations (L_3)

- **Would I** have won the Nobel **if I had** eaten chocolate?
- **Would I** have got the loan **if I had** been a man?
- $P(y'|do(x'), y, x)$ → L_3 assumptions (e.g. exogeneous noise)

2. Interventions – action-guidance (L_2)

- **Does** chocolate **make you** smarter?
- **Does** being a woman **reduce your** chances of getting a loan?
- $P(y|do(x))$ → L_2 assumptions (e.g. causal graph, adjustment sets)

1. Observations – associations (L_1)

- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?
- $P(y|\, x)$

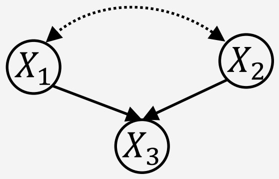
Figure inspired from [Pearl and Mackenzie, 2018]

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$:
$F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

Example



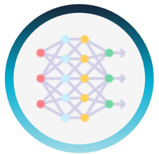
$$X_1 = f_1(U_1)$$
$$X_2 = f_2(U_2)$$
$$X_3 = f_3(X_1, X_2, U_3)$$

with $P(U)$ s.t. $U_3 \perp\!\!\!\perp U_1$ and $U_3 \perp\!\!\!\perp U_2$

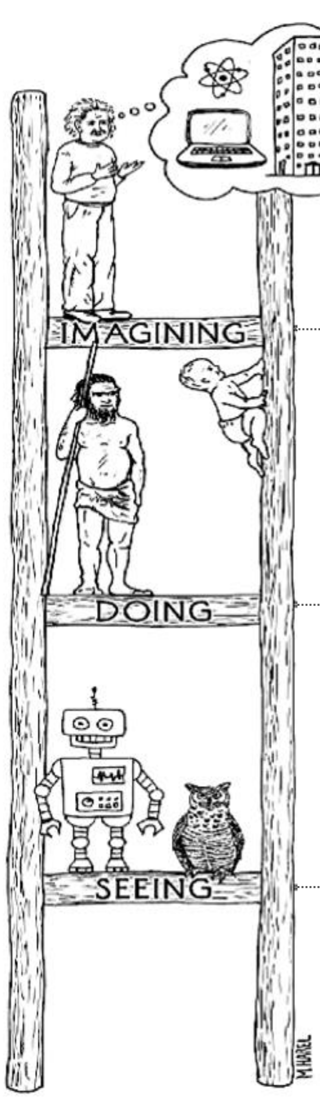
Scope of the survey

Structural Causal Models
Numerous assumptions
Identification concerns

Deep Generative Models
Flexibility, few assumptions
Few guarantees



Motivation, towards finer decision-making



3. Counterfactuals – against existing observations (L_3)

- **Would I** have won the Nobel **if I had** eaten chocolate?
- **Would I** have got the loan **if I had** been a man?
- $P(y'|do(x'), y, x)$ → L_3 assumptions (e.g. exogeneous noise)

2. Interventions – action-guidance (L_2)

- **Does** chocolate **make you** smarter?
- **Does** being a woman **reduce your** chances of getting a loan?
- $P(y|do(x))$ → L_2 assumptions (e.g. causal graph, adjustment sets)

1. Observations – associations (L_1)

- **Given that** I eat chocolate, **how likely am I** to win a Nobel?
- **Given that** I am a woman, **how likely am I** to get a loan?
- $P(y|\, x)$

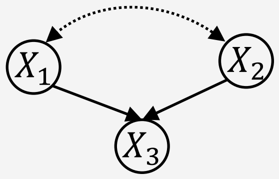
Figure inspired from [Pearl and Mackenzie, 2018]

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$:
$F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

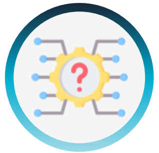
Example



$X_1 = f_1(U_1)$
$X_2 = f_2(U_2)$
$X_3 = f_3(X_1, X_2, U_3)$

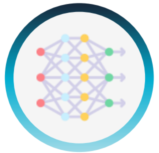
with $P(U)$ s.t. $U_3 \perp\!\!\!\perp U_1$ and $U_3 \perp\!\!\!\perp U_2$

Scope of the survey



Structural Causal Models
Numerous assumptions
Identification concerns

Deep Generative Models
Flexibility, few assumptions
Few guarantees



Learning Structural Causal Models through Deep Generative Models
Existing works, capabilities, and remaining open questions

Literature, the challenge of learning SCMs till DGMs arrived

Motivation: SCMs are convenient tools enabling the modeling of a wide range of causal queries (L_3, multi-treatment, path-specific, …)

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$: $F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

Literature, the challenge of learning SCMs till DGMs arrived

Motivation: SCMs are convenient tools enabling the modeling of a wide range of causal queries (L_3, multi-treatment, path-specific, …)

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$: $F = \{X_i := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

Linear Causal Models: $f_i(PA(X_i), U_i) = \alpha_0 U_i + \sum_{j=1}^{|PA(X_i)|} \alpha_j PA_j(X_i)$ with $\forall j, \alpha_j \in \mathbb{R}$ [Pearl, 2000; Spirtes *et al.*, 2000; Bollen, 1989]

80s …

Literature, the challenge of learning SCMs till DGMs arrived

Motivation: SCMs are convenient tools enabling the modeling of a wide range of causal queries (L_3, multi-treatment, path-specific, …)

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$: $F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

80s …

Linear Causal Models: $f_i(PA(X_i), U_i) = \alpha_0 U_i + \sum_{j=1}^{|PA(X_i)|} \alpha_j PA_j(X_i)$ with $\forall j, \alpha_j \in \mathbb{R}$ [Pearl, 2000; Spirtes *et al.*, 2000; Bollen, 1989]

2008

Additive Noise Models: $f_i(PA(X_i), U_i) = g_i\big(PA(X_i)\big) + U_i$ [18]

2009

Post-Nonlinear Models: $f_i(PA(X_i), U_i) = f_{i,2}(f_{i,1}\big(PA(X_i)\big) + U_i)$ with $f_{i,2}$ an invertible function [20]

Literature, the challenge of learning SCMs till DGMs arrived

Motivation: SCMs are convenient tools enabling the modeling of a wide range of causal queries (L_3, multi-treatment, path-specific, ...)

Definition

A **Structural Causal Model** (SCM) is a tuple $M := (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$: $F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

80s ...

Linear Causal Models: $f_i(PA(X_i), U_i) = \alpha_0 U_i + \sum_{j=1}^{|PA(X_i)|} \alpha_j PA_j(X_i)$ with $\forall j, \alpha_j \in \mathbb{R}$ [Pearl, 2000; Spirtes *et al.*, 2000; Bollen, 1989]

2008

Additive Noise Models: $f_i(PA(X_i), U_i) = g_i\big(PA(X_i)\big) + U_i$ [18]

2009

Post-Nonlinear Models: $f_i(PA(X_i), U_i) = f_{i,2}(f_{i,1}\big(PA(X_i)\big) + U_i)$ with $f_{i,2}$ an invertible function [20]

2018 ... now

Deep Structural Causal Models: $f_i(PA(X_i), U_i) = f_{X_i|PA(X_i)}(U_i)$ Conditional Deep Generative Models (**DGMs**)

- 2018-2019 – **GAN**-based SCMs for L_1 & L_2 tasks like data augmentation [12, 13]
- 2020 – **NF**-based SCMs for L_3 tasks [2, 5] + **DSCM** definition [2]
- 2020-2023 – **Buzz** around DSCMs: VAE-based methods [7, 8, 9] + other GAN-based [11, 14, 15] + other methods [3, 10, 16, 17]

Literature, the challenge of learning SCMs till DGMs arrived

Motivation: SCMs are convenient tools enabling the modeling of a wide range of causal queries (L_3, multi-treatment, path-specific, ...)

Definition

A **Structural Causal Model** (SCM) is a tuple $M \coloneqq (F, P(U))$ where F comprises a set of d structural equations, f_i, one for each endogenous random variable $X_i \in \boldsymbol{X}$: $F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$ with $PA(X_i)$ the parents of X_i and U_i the exogeneous noise

Definition from [Pearl, 2009]

80s ...

Linear Causal Models: $f_i(PA(X_i), U_i) = \alpha_0 U_i + \sum_{j=1}^{|PA(X_i)|} \alpha_j PA_j(X_i)$ with $\forall j, \alpha_j \in \mathbb{R}$ [Pearl, 2000; Spirtes *et al.*, 2000; Bollen, 1989]

2008

Additive Noise Models: $f_i(PA(X_i), U_i) = g_i(PA(X_i)) + U_i$ [18]

2009

Post-Nonlinear Models: $f_i(PA(X_i), U_i) = f_{i,2}(f_{i,1}(PA(X_i)) + U_i)$ with $f_{i,2}$ an invertible function [20]

2018 ... now

Deep Structural Causal Models: $f_i(PA(X_i), U_i) = f_{X_i|PA(X_i)}(U_i)$ Conditional Deep Generative Models (**DGMs**)

- 2018-2019 – **GAN**-based SCMs for L_1 & L_2 tasks like data augmentation [12, 13]
- 2020 – **NF**-based SCMs for L_3 tasks [2, 5] + **DSCM** definition [2]
- 2020-2023 – **Buzz** around DSCMs: VAE-based methods [7, 8, 9] + other GAN-based [11, 14, 15] + other methods [3, 10, 16, 17]

- **What a mess !** We need a review for practitioners and researchers on existing works, their capabilities, and the remaining open questions.

Problem setup, from observations to counterfactuals

The world of SCMs

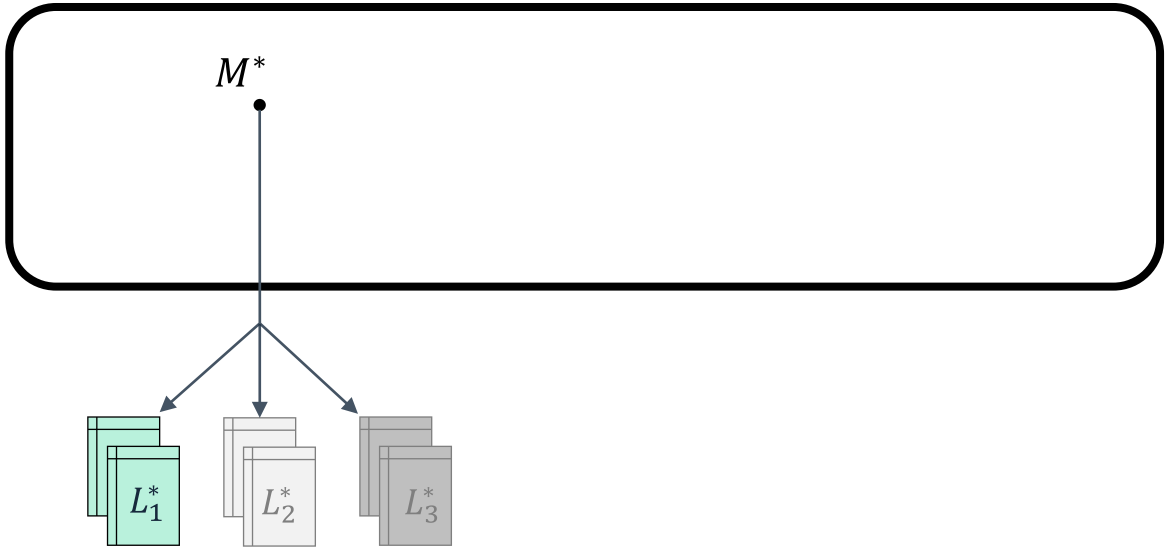


Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals

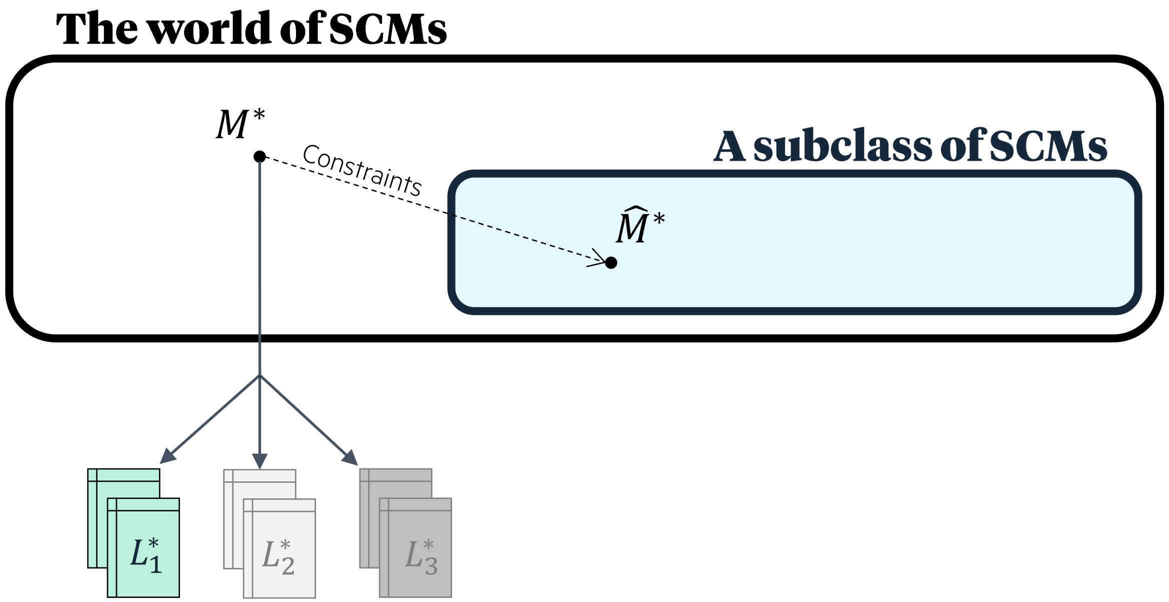


Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals

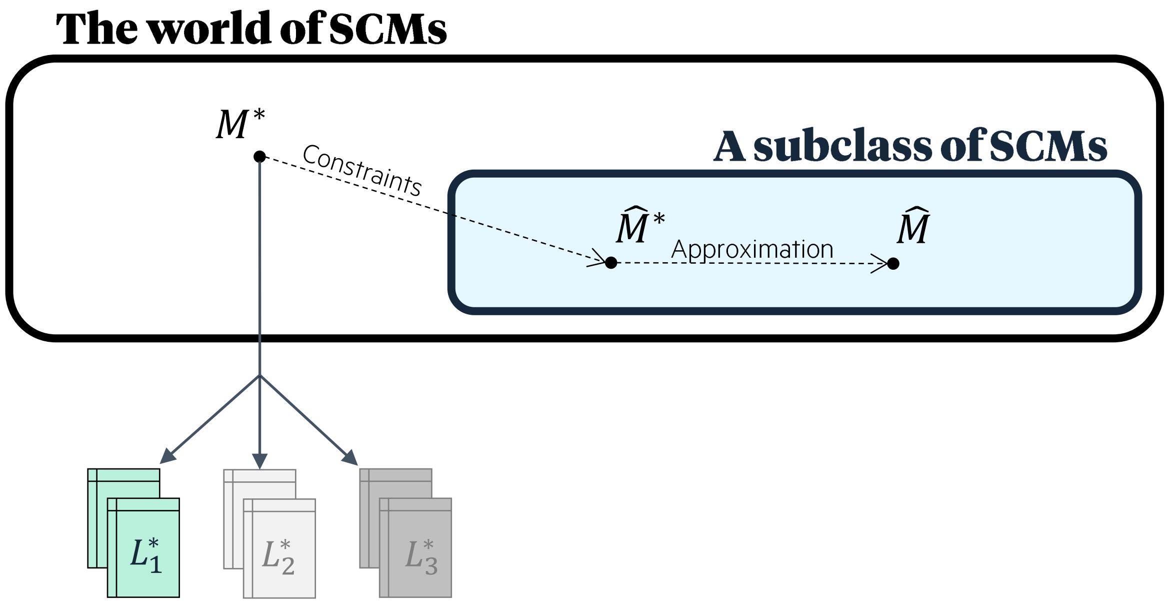


Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals

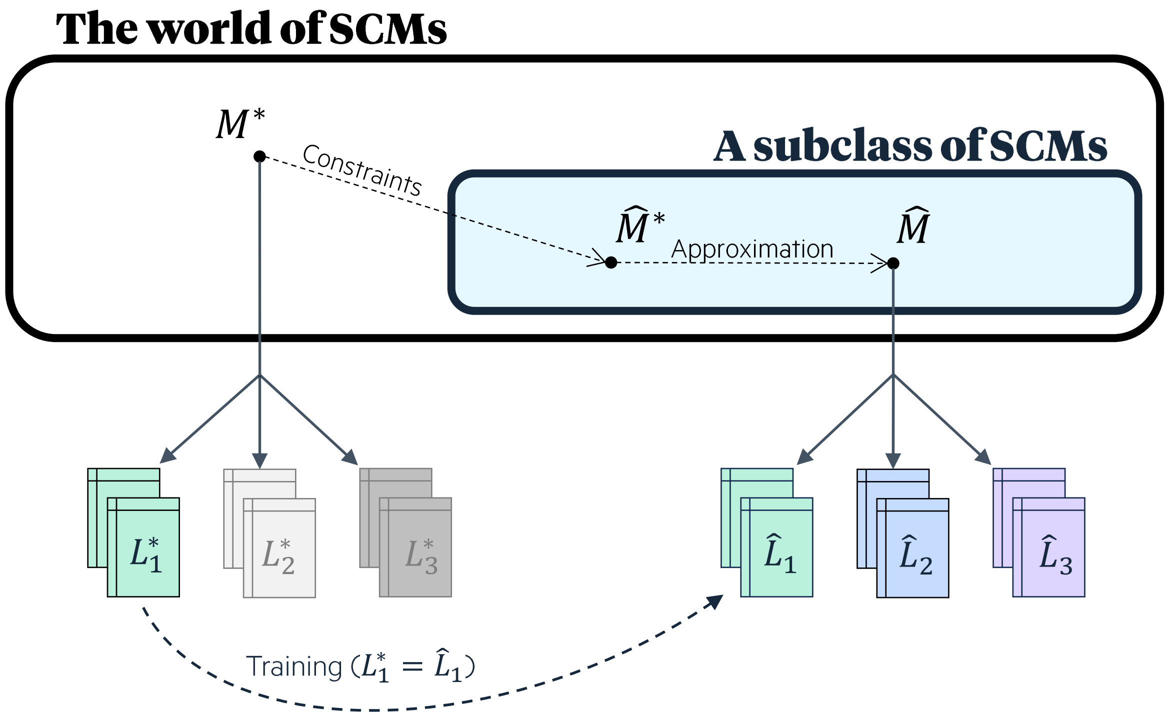
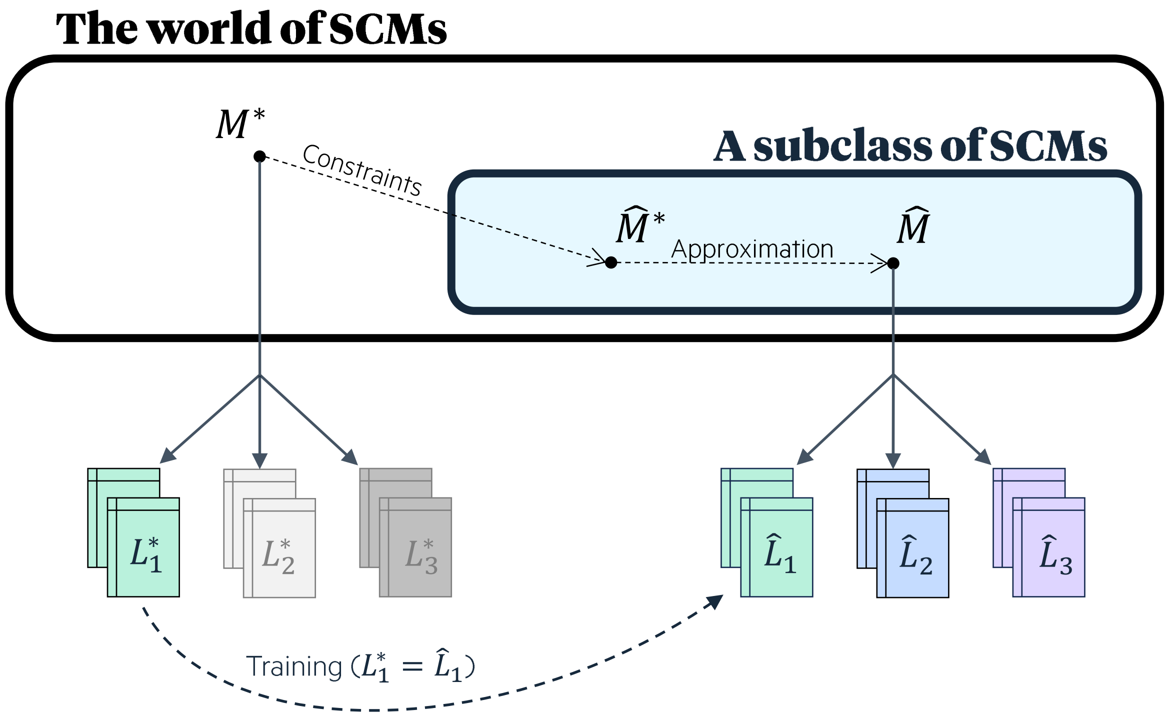


Figure inspired from [Xia *et al.*, 2021]

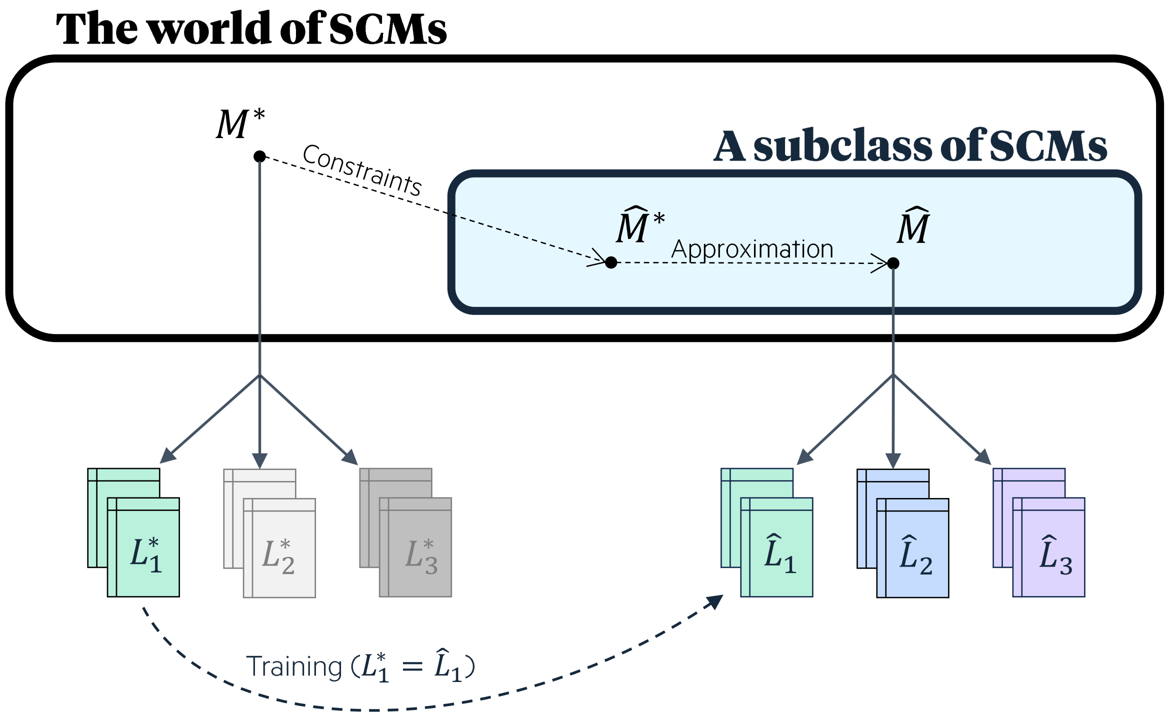
Problem setup, from observations to counterfactuals



Deep Structural Causal Models

Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals



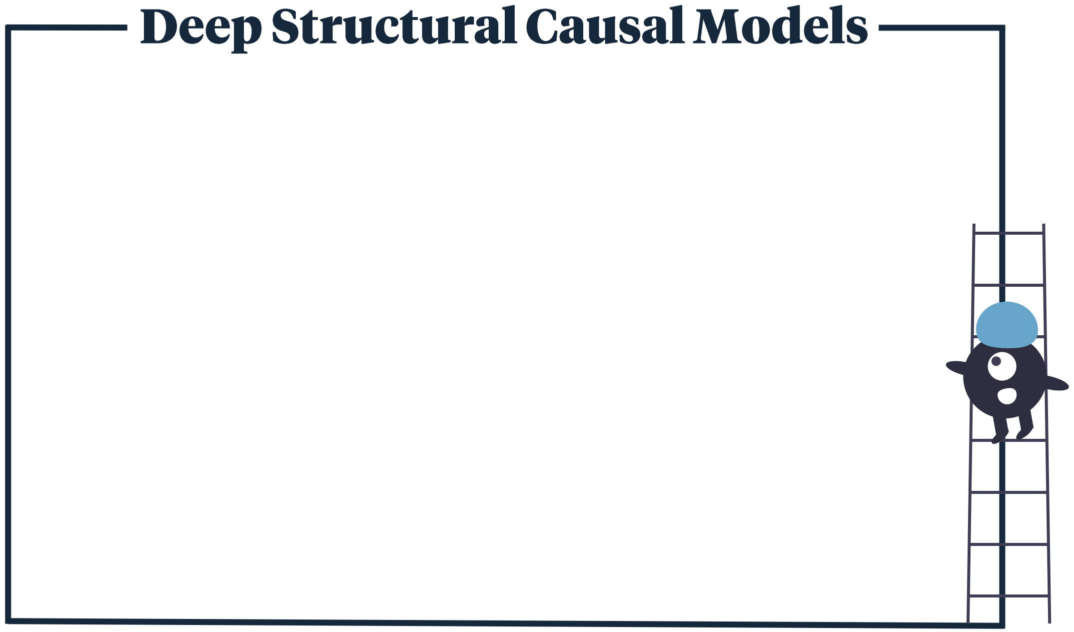
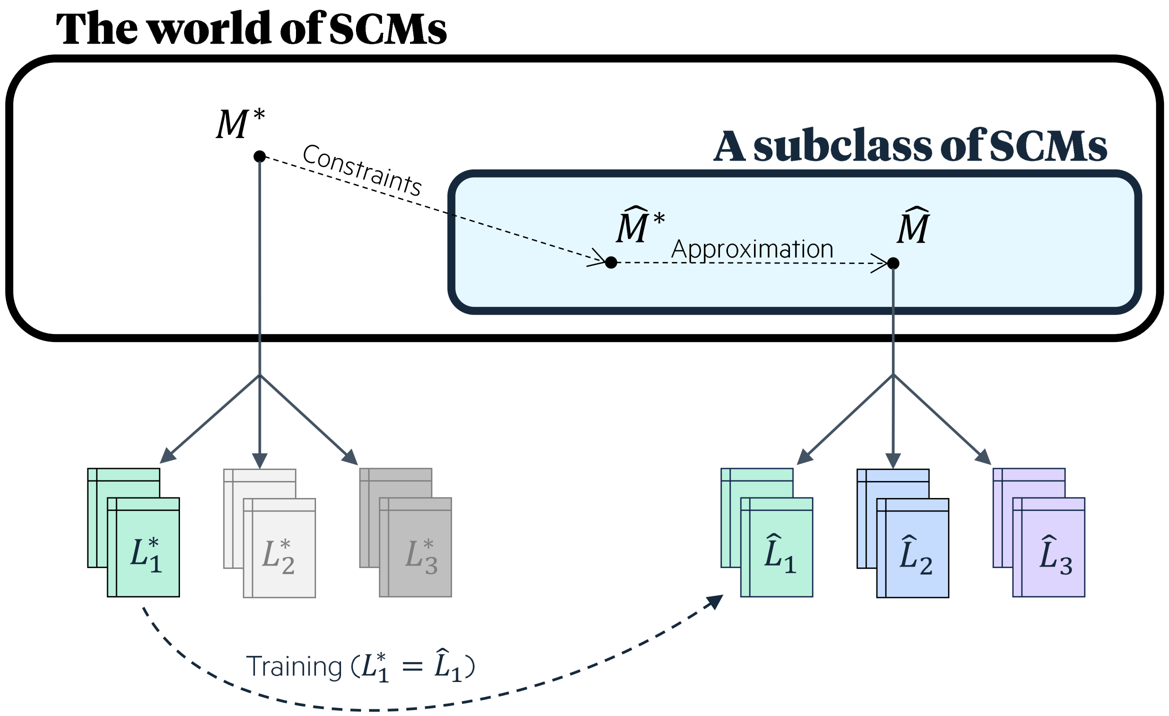


Figure inspired from [Xia *et al.*, 2021]

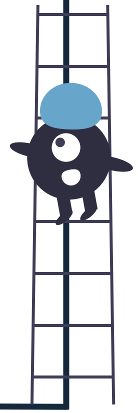
Problem setup, from observations to counterfactuals



Deep Structural Causal Models

Rich literature

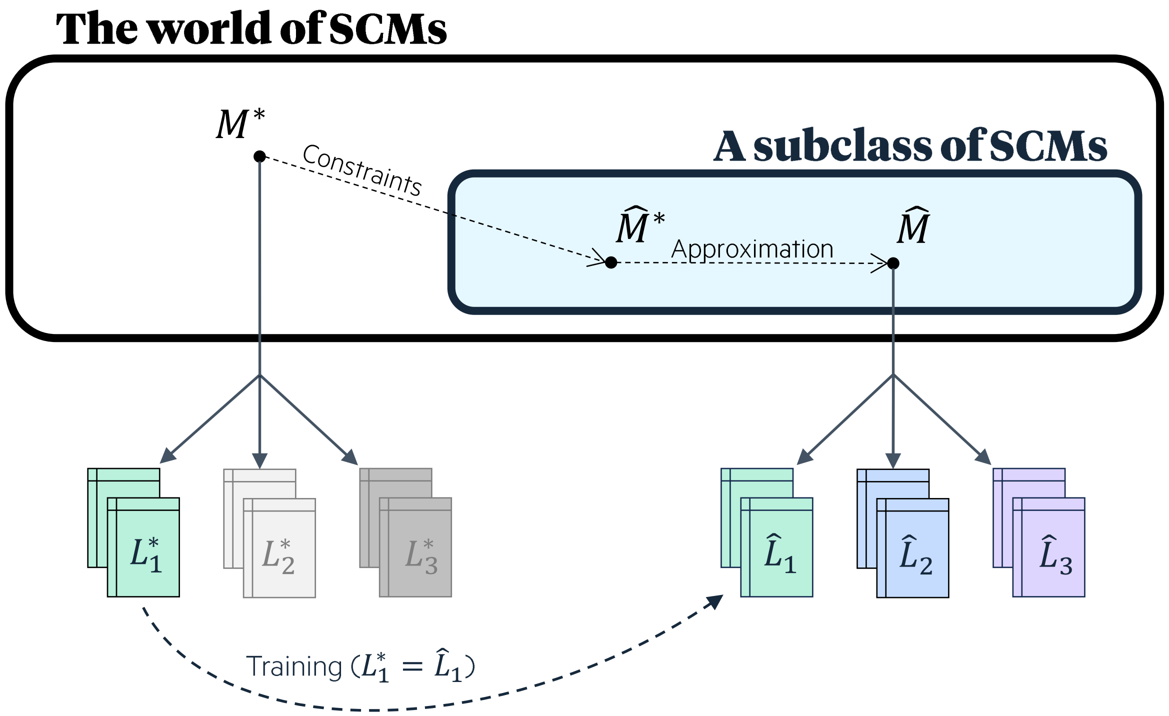
There exist many methods to learn a DSCM and reviews comparing their ML components[1] …



[1] [Zhou *et al.*, 2023; Komanduri *et al.*, 2023; Kaddour *et al.*, 2022]

Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals



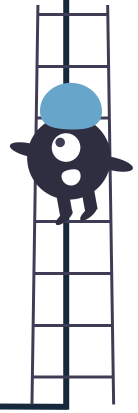
Deep Structural Causal Models

Rich literature

There exist many methods to learn a DSCM and reviews comparing their ML components[1] ...

Lack of comparative analysis

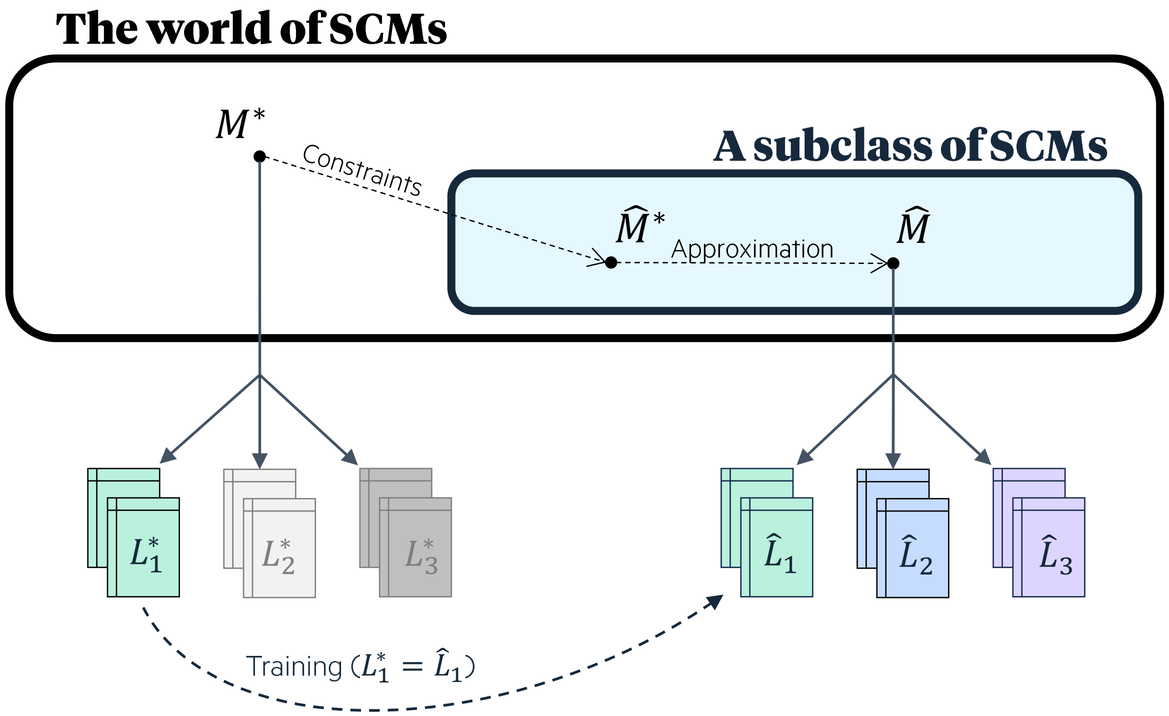
... but nothing in terms of hypotheses, guarantees, and performances.



[1] [Zhou *et al.*, 2023; Komanduri *et al.*, 2023; Kaddour *et al.*, 2022]

Figure inspired from [Xia *et al.*, 2021]

Problem setup, from observations to counterfactuals



Deep Structural Causal Models

Rich literature
There exist many methods to learn a DSCM and reviews comparing their ML components[1] ...

Lack of comparative analysis
... but nothing in terms of hypotheses, guarantees, and performances.

Research Question: Given a known causal structure and observational data, what are the capabilities of existing DSCMs in answering counterfactual questions?
Practical Questions: How can a practitioner choose the most appropriate methods? What are the limitations?

[1] [Zhou *et al.*, 2023; Komanduri *et al.*, 2023; Kaddour *et al.*, 2022]
Figure inspired from [Xia *et al.*, 2021]

DSCM, definition & classification w.r.t DGMs

Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M \coloneqq (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]

DSCM, definition & classification w.r.t DGMs

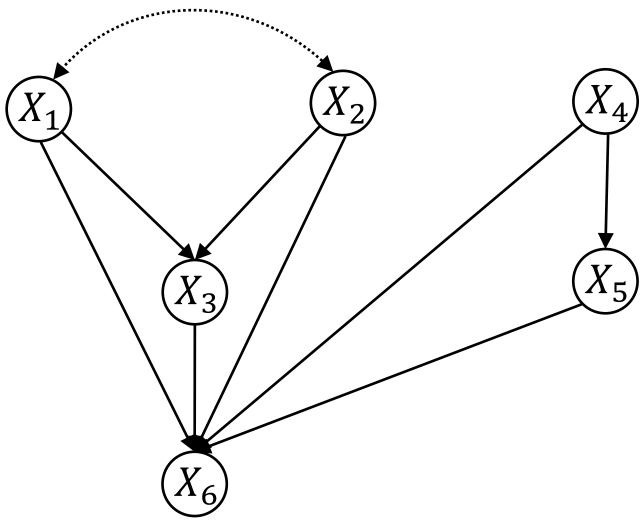
Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M \coloneqq (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]



DSCM, definition & classification w.r.t DGMs

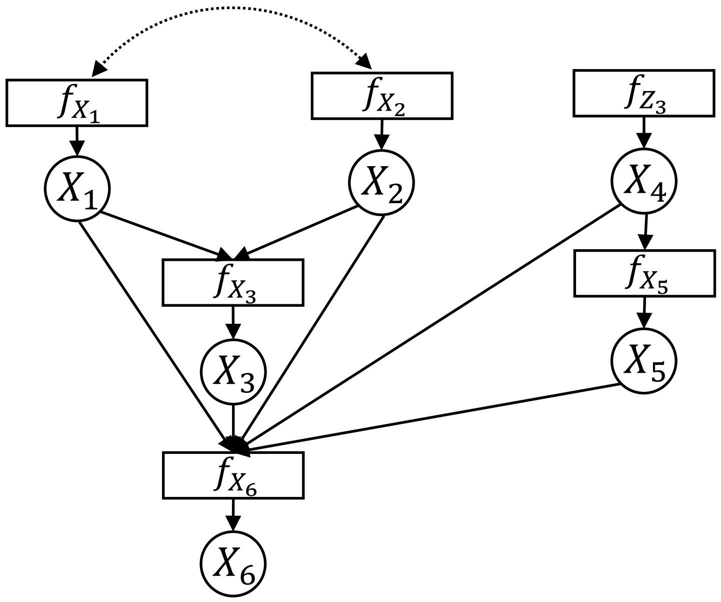
Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M := (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]



DSCM, definition & classification w.r.t DGMs

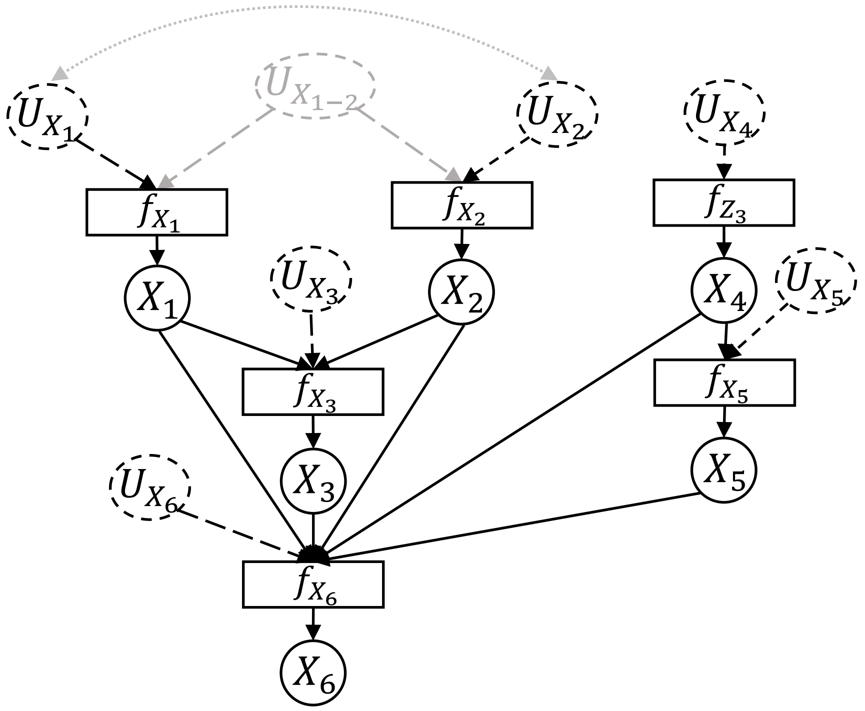
Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M := (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]



DSCM, definition & classification w.r.t DGMs

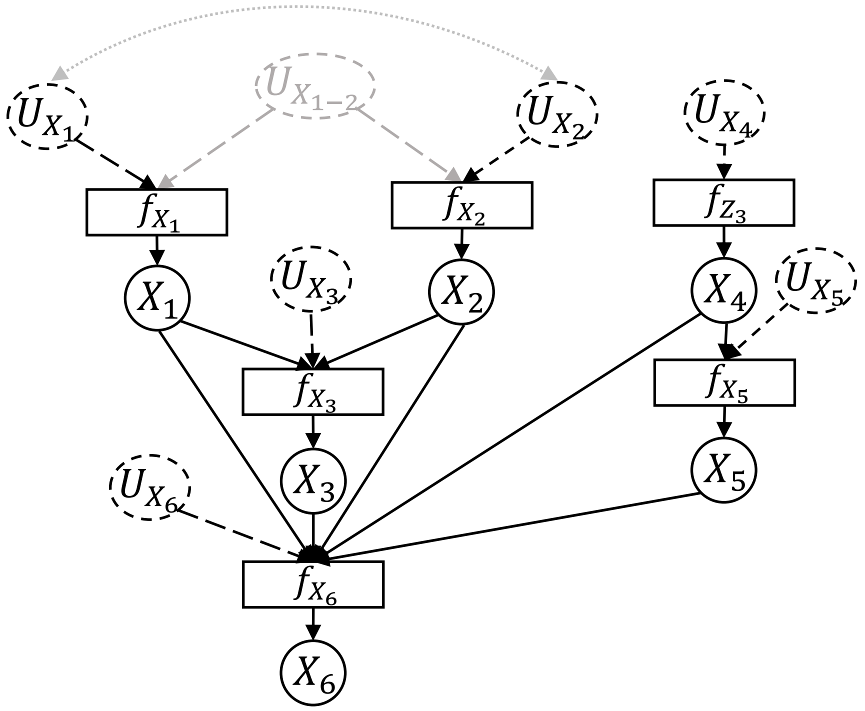
Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M \coloneqq (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]



Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]

DSCM, definition & classification w.r.t DGMs

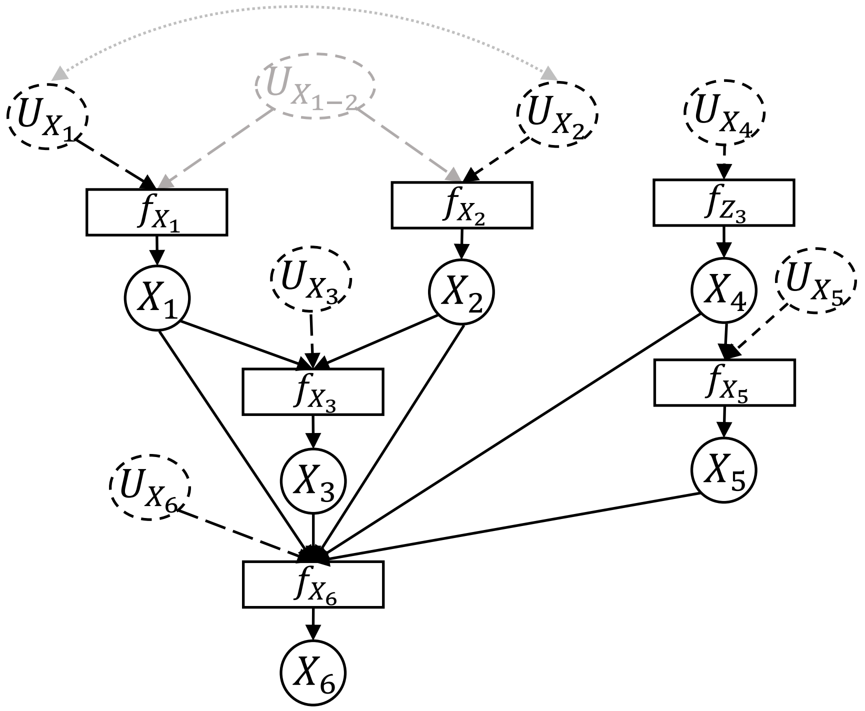
Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M := (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]



Amortized Explicit (AE)

- f_i is learned with an auto-encoder s.t. $g_i(PA(X_i), U_i) = f_i$ and $e_i(X_i, PA(X_i)) = U_i$
- (Variational) (Graph) Auto Encoders, Diffusion Models: **iVGAE** [7], **VACA** [8], **DCM** [9]

Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]

DSCM, definition & classification w.r.t DGMs

Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M := (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i := f_i(PA(X_i), U_i)\}_{i\in[1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski et al., 2020]

Amortized Implicit (AI)

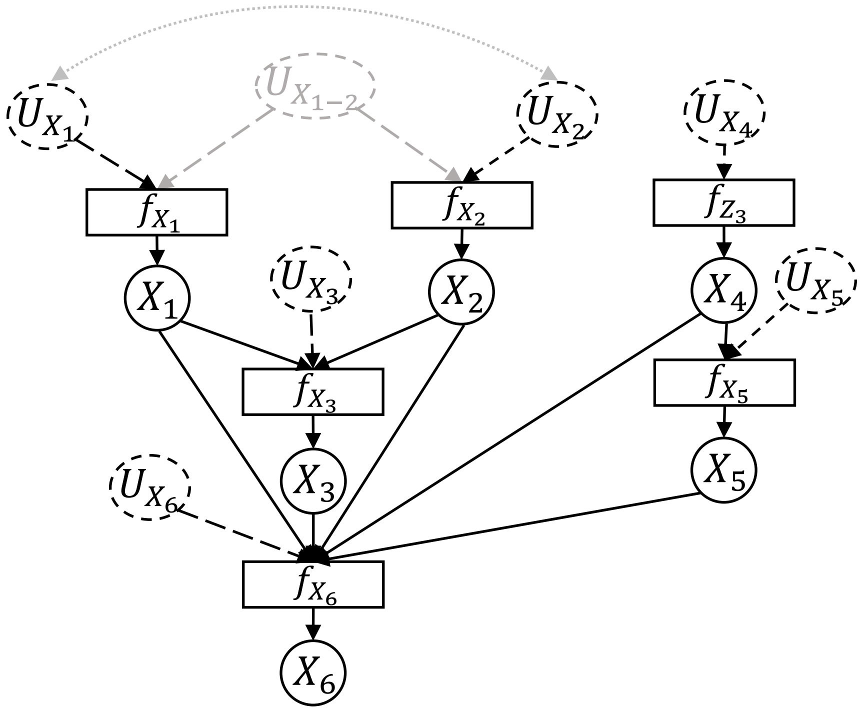
- f_i is a conditional implicit-likelihood model learned with a loss implicitly considering U_i
- Adversarial learning, Loss to fit the distribution: **Causal-TGAN** [11], **CausalGAN** [12], **CFGAN** [13], **DECAF** [14], **WhatIfGAN** [15], **CGN** [16], **DEAR** [17], **GAN-NCM** [3], **MLE-NCM** [3], **SCM-VAE** [10]

Amortized Explicit (AE)

- f_i is learned with an auto-encoder s.t. $g_i(PA(X_i), U_i) = f_i$ and $e_i\big(X_i, PA(X_i)\big) = U_i$
- (Variational) (Graph) Auto Encoders, Diffusion Models: **iVGAE** [7], **VACA** [8], **DCM** [9]

Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]



DSCM, definition & classification w.r.t DGMs

Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M \coloneqq (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski *et al.*, 2020]

Amortized Implicit (AI)

- f_i is a conditional implicit-likelihood model learned with a loss implicitly considering U_i
- Adversarial learning, Loss to fit the distribution: **Causal-TGAN** [11], **CausalGAN** [12], **CFGAN** [13], **DECAF** [14], **WhatIfGAN** [15], **CGN** [16], **DEAR** [17], **GAN-NCM** [3], **MLE-NCM** [3], **SCM-VAE** [10]

Amortized Explicit (AE)

- f_i is learned with an auto-encoder s.t. $g_i(PA(X_i), U_i) = f_i$ and $e_i(X_i, PA(X_i)) = U_i$
- (Variational) (Graph) Auto Encoders, Diffusion Models: **iVGAE** [7], **VACA** [8], **DCM** [9]

Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]

DSCM, definition & classification w.r.t DGMs

Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M \coloneqq (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i \coloneqq f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski *et al.*, 2020]

Amortized Implicit (AI)

- f_i is a conditional implicit-likelihood model learned with a loss implicitly considering U_i
- Adversarial learning, Loss to fit the distribution: **Causal-TGAN** [11], **CausalGAN** [12], **CFGAN** [13], **DECAF** [14], **WhatIfGAN** [15], **CGN** [16], **DEAR** [17], **GAN-NCM** [3], **MLE-NCM** [3], **SCM-VAE** [10]

Amortized Explicit (AE)

- f_i is learned with an auto-encoder s.t. $g_i(PA(X_i), U_i) = f_i$ and $e_i(X_i, PA(X_i)) = U_i$
- (Variational) (Graph) Auto Encoders, Diffusion Models: **iVGAE** [7], **VACA** [8], **DCM** [9]

Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]

Abduction Step

- **Sample Rejection procedure**: U_i s.t. $f_i(PA(X_i), U_i) = X_i$
- **Encoding**: $U_i = e_i(X_i, PA(X_i))$
- **Mechanism Inversion**: $U_i = f_i^{-1}(X_i, PA(X_i))$

DSCM, definition & classification w.r.t DGMs

Definition

A **Deep Structural Causal Model** (DSCM) is an SCM $M := (F, P(U))$ that uses **deep-learning components** to model the structural assignments:

$$F = \{X_i := f_i(PA(X_i), U_i)\}_{i \in [1,d]}$$

with f_i a neural network, $PA(X_i)$ the parents of X_i induced by the known structure and U_i the exogeneous noise

Definition from [Pawlowski *et al.*, 2020]

Amortized Implicit (AI)

- f_i is a conditional implicit-likelihood model learned with a loss implicitly considering U_i
- Adversarial learning, Loss to fit the distribution: **Causal-TGAN** [11], **CausalGAN** [12], **CFGAN** [13], **DECAF** [14], **WhatIfGAN** [15], **CGN** [16], **DEAR** [17], **GAN-NCM** [3], **MLE-NCM** [3], **SCM-VAE** [10]

Amortized Explicit (AE)

- f_i is learned with an auto-encoder s.t. $g_i(PA(X_i), U_i) = f_i$ and $e_i(X_i, PA(X_i)) = U_i$
- (Variational) (Graph) Auto Encoders, Diffusion Models: **iVGAE** [7], **VACA** [8], **DCM** [9]

Invertible Explicit (IE)

- f_i is supposed diffeomorphic w.r.t U_i s.t. $U_i = f_i^{-1}(X_i, PA(X_i))$
- Normalizing Flow: **Causal-NF** [4], **NF-DSCM** [2], **NCF** [5], **CARFEL** [6], **NF-BGM** [1]

Abduction Step

- **Sample Rejection procedure**: U_i s.t. $f_i(PA(X_i), U_i) = X_i$
- **Encoding**: $U_i = e_i(X_i, PA(X_i))$
- **Mechanism Inversion**: $U_i = f_i^{-1}(X_i, PA(X_i))$

Class of DGM	Abduction step: Mechanism Inversion	Encoding	Sample Rejection
Invertible Explicit	✓	✓	✓
Amortised Explicit	✗	✓	✓
Amortised Implicit	✗	✗	✓

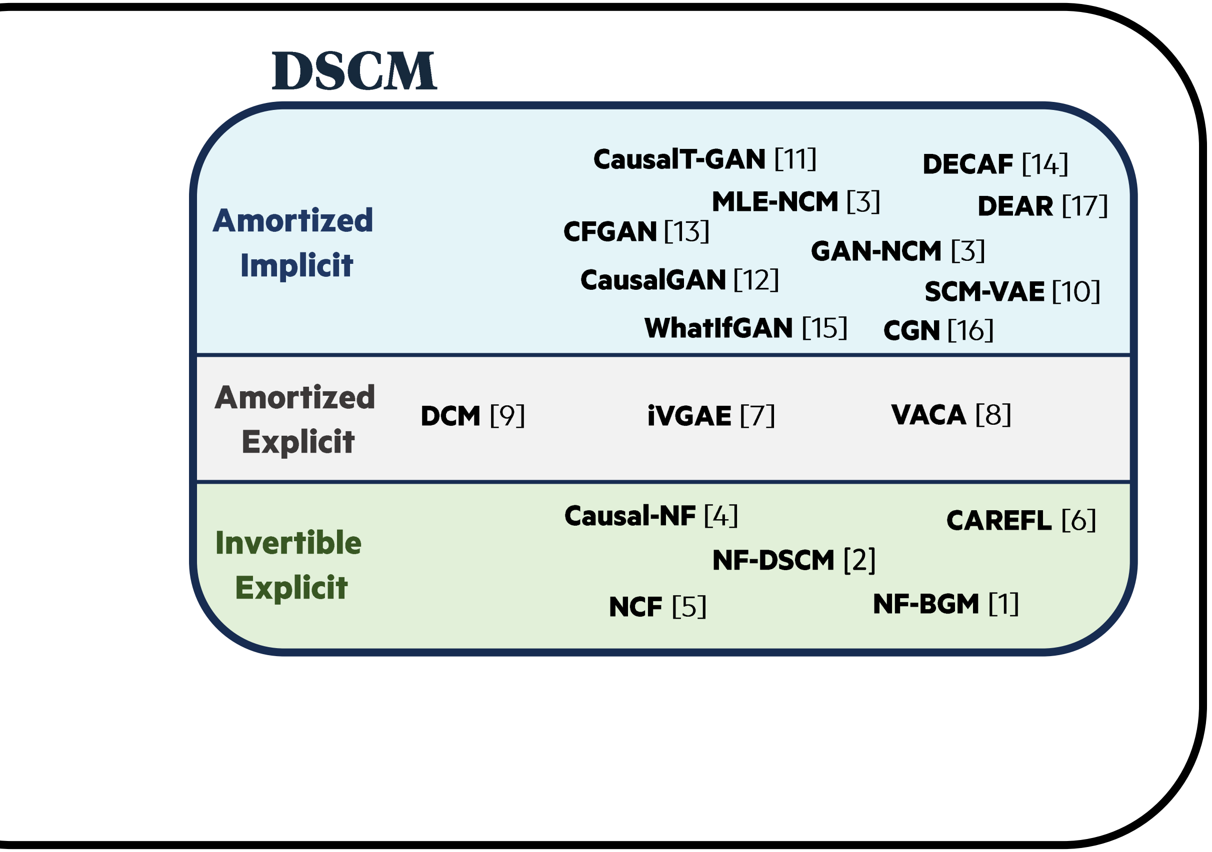
(b) Abduction steps for the classes of DGMs

DSCM, definition & classification w.r.t SCMs

SCM

DSCM, definition & classification w.r.t SCMs

SCM

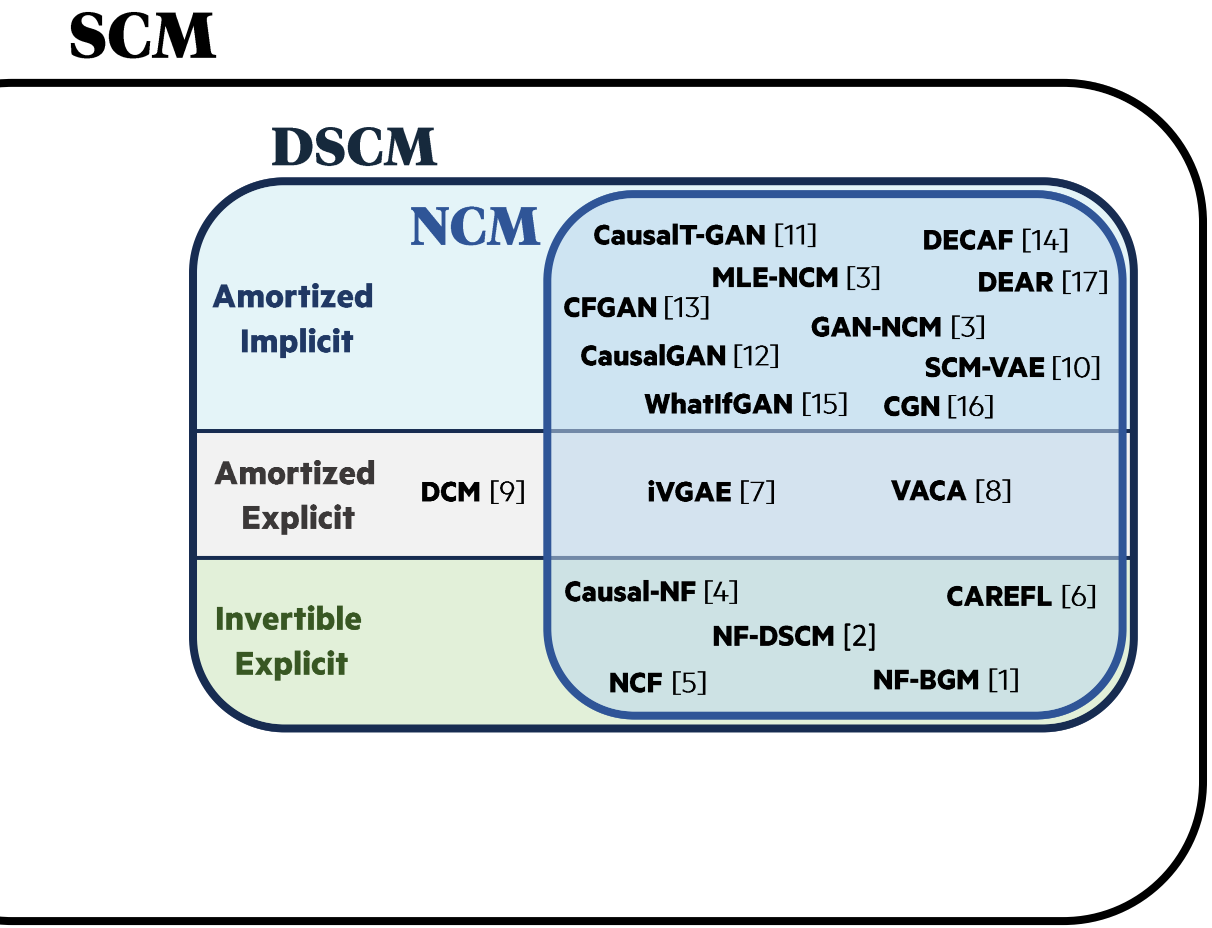


DSCM

[Pawlowski *et al*, 2020]

- Definition: SCM whose causal mechanisms are **deep (conditional) generative models**
- No theoretical guarantees

DSCM, definition & classification w.r.t SCMs



DSCM

[Pawlowski *et al*, 2020]

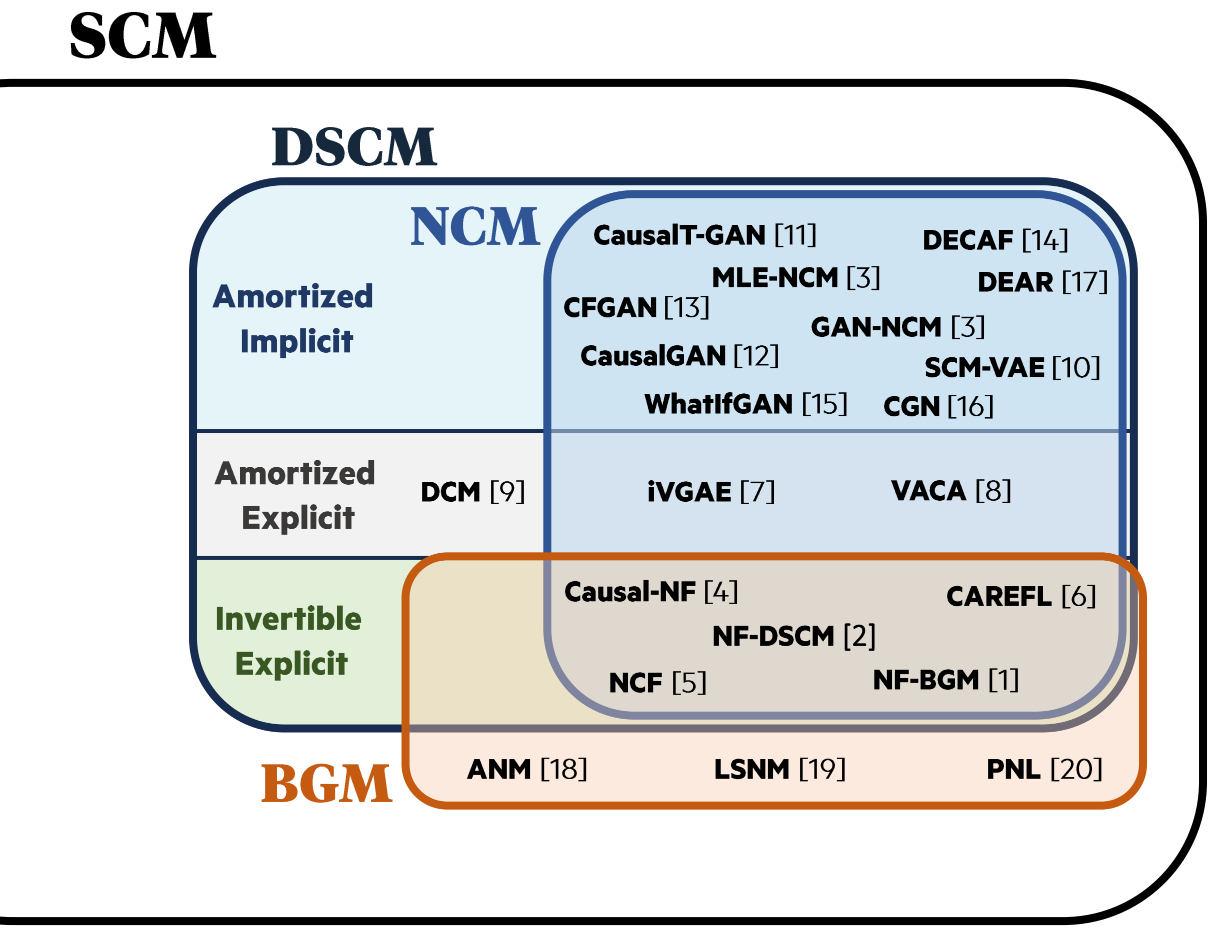
- Definition: SCM whose causal mechanisms are **deep (conditional) generative models**
- No theoretical guarantees

NCM

[Xia *et al*, 2021 & 2023]

- Definition: SCM whose causal mechanisms are **feedforward neural networks**
- Guarantees:
  - **Expressivity**: Given a graph there always exists an NCM L3 consistent with the true SCM
  - L_3**-Identifiability** iif L_3-Identifiability holds for the true SCM

DSCM, definition & classification w.r.t SCMs



DSCM

[Pawlowski *et al*, 2020]

- Definition: SCM whose causal mechanisms are **deep (conditional) generative models**
- No theoretical guarantees

NCM

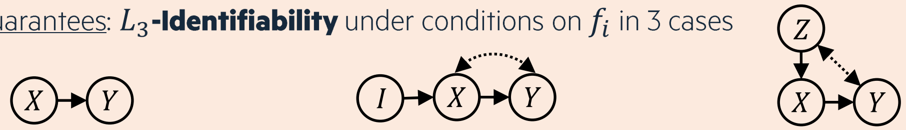
[Xia *et al*, 2021 & 2023]

- Definition: SCM whose causal mechanisms are **feedforward neural networks**
- Guarantees:
  - **Expressivity**: Given a graph there always exists an NCM L3 consistent with the true SCM
  - L_3**-Identifiability** iif L_3-Identifiability holds for the true SCM

BGM

[Nasr-Esfahany *et al*, 2023]

- Definition: SCM whose causal mechanisms are **bijective** w.r.t. the **exogeneous** noises
- Guarantees: L_3**-Identifiability** under conditions on f_i in 3 cases



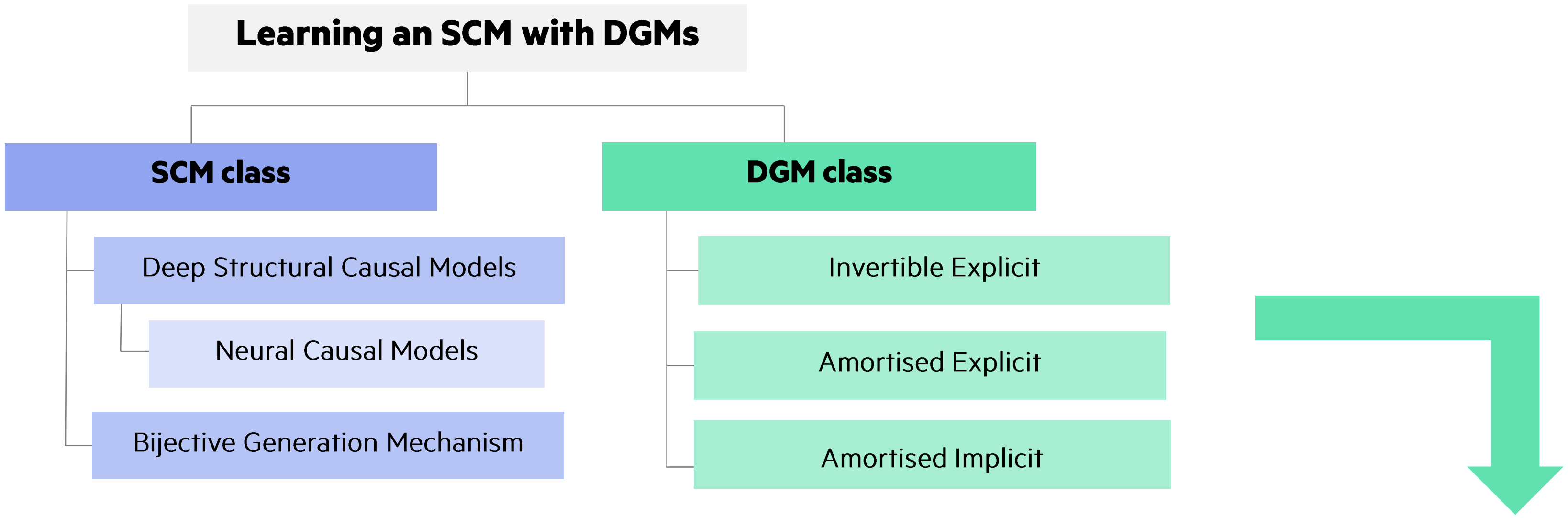
Markovian | Instrumental Variable | Backdoor Criterion

DSCM, classification summary

- **Learning an SCM with DGMs**
  - **SCM class**
    - Deep Structural Causal Models
      - Neural Causal Models
    - Bijective Generation Mechanism
  - **DGM class**
    - Invertible Explicit
    - Amortised Explicit
    - Amortised Implicit

Figures from [Poinsot *et al*, 2024]

DSCM, classification summary

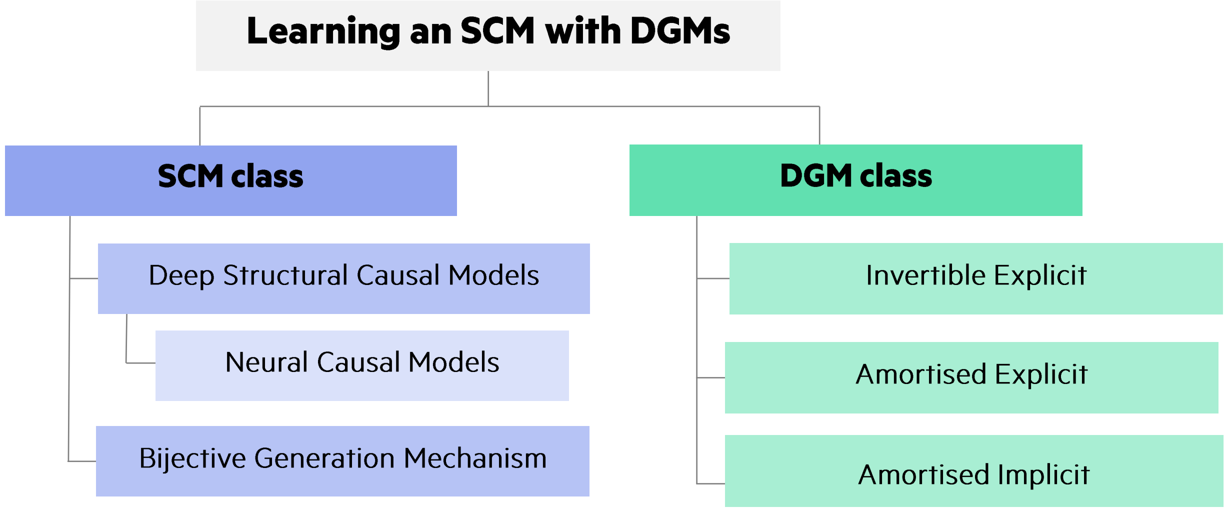


Class of DGM	Abduction step: Mechanism Inversion	Abduction step: Encoding	Abduction step: Sample Rejection
Invertible Explicit	✓	✓	✓
Amortised Explicit	✗	✓	✓
Amortised Implicit	✗	✗	✓

(b) Abduction steps for the classes of DGMs

Figures from [Poinsot *et al*, 2024]

DSCM, classification summary



Class	Mechanism	Identifiability Guarantees
DSCM	Neural network	-
NCM	Feedforward neural network	$\mathcal{L}_3$-id. (resp. $\mathcal{L}_2$) iif $\mathcal{L}_3$-id. (resp. $\mathcal{L}_2$) from the true SCM
BGM	Bijective noise	$\mathcal{L}_3$-id. for three settings* cf. Theorems 5.1, 5.2 and 5.3

*Markovian, Instrumental Variable, and Backdoor Criterion

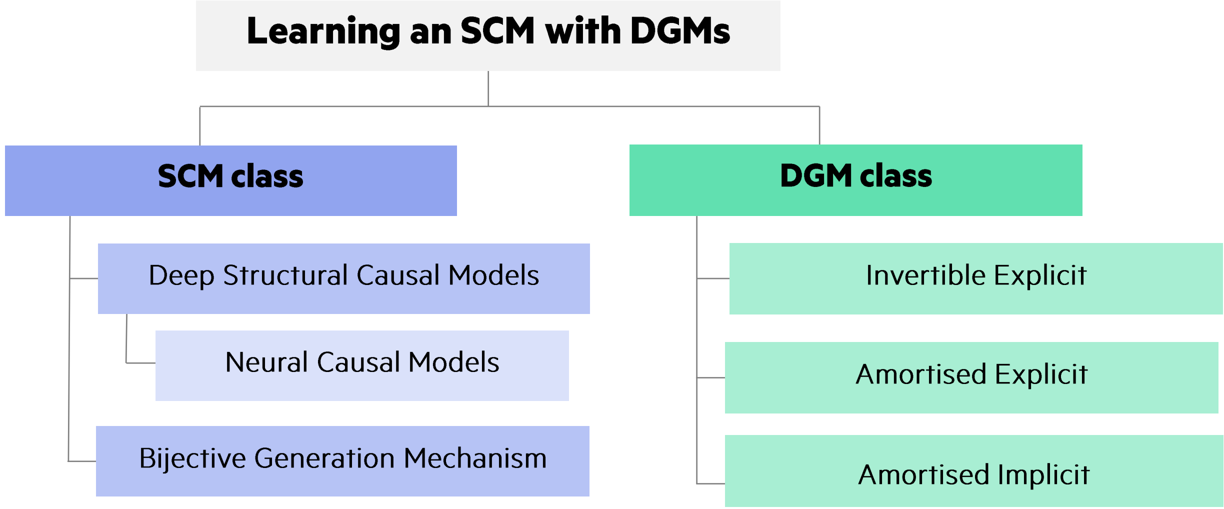
(a) Identifiability guarantees of the classes of SCMs

Class of DGM	Abduction step		
	Mechanism Inversion	Encoding	Sample Rejection
Invertible Explicit	✓	✓	✓
Amortised Explicit	✗	✓	✓
Amortised Implicit	✗	✗	✓

(b) Abduction steps for the classes of DGMs

Figures from [Poinsot *et al*, 2024]

DSCM, classification summary



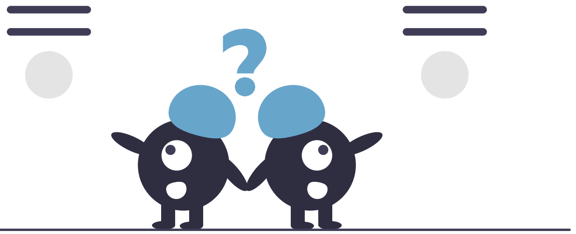
Class	Mechanism	Identifiability Guarantees
DSCM	Neural network	-
NCM	Feedforward neural network	$\mathcal{L}_3$-id. (resp. $\mathcal{L}_2$) iif $\mathcal{L}_3$-id. (resp. $\mathcal{L}_2$) from the true SCM
BGM	Bijective noise	$\mathcal{L}_3$-id. for three settings* cf. Theorems 5.1, 5.2 and 5.3

*Markovian, Instrumental Variable, and Backdoor Criterion

(a) Identifiability guarantees of the classes of SCMs

Class of DGM	Abduction step		
	Mechanism Inversion	Encoding	Sample Rejection
Invertible Explicit	✓	✓	✓
Amortised Explicit	✗	✓	✓
Amortised Implicit	✗	✗	✓

(b) Abduction steps for the classes of DGMs



However, each method has its own characteristics.

"What should I choose?"

Figures from [Poinsot *et al*, 2024]

DSCM, hypotheses & guarantees

Method	Classification: SCM class	Classification: DGM class*	Additional Hypotheses: Causal Structure	Additional Hypotheses: Hidden Confounder	Additional Hypotheses: Data Assumptions	Additional Hypotheses: Available Abduction	Additional Guarantees: Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^\tau$	✓$^\tau$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^\tau$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

- ✓ Hypotheses are linked to the choice of Generative Model
- ✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification: SCM class	Classification: DGM class*	Additional Hypotheses: Causal Structure	Additional Hypotheses: Hidden Confounder	Additional Hypotheses: Data Assumptions	Additional Hypotheses: Available Abduction	Additional Guarantees: Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^{\tau}$	✓$^{\tau}$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^{\tau}$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

- ✓ Hypotheses are linked to the choice of Generative Model
- ✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

➢ **Causal structure**
- Causal order is enough

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification: SCM class	Classification: DGM class*	Additional Hypotheses: Causal Structure	Additional Hypotheses: Hidden Confounder	Additional Hypotheses: Data Assumptions	Additional Hypotheses: Available Abduction	Additional Guarantees: Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, ∅	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^{\tau}$	✓$^{\tau}$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^{\tau}$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

- ✓ Hypotheses are linked to the choice of Generative Model
- ✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

- **Causal structure**
  - Causal order is enough
- **Hidden confounding**
  - BGM and WhatIfGAN consider correlated exogeneous noises
  - NCF and NCM deal with semi-Markovian DAGs

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification		Additional Hypotheses				Additional Guarantees
	SCM class	DGM class*	Causal Structure	Hidden Confounder	Data Assumptions	Available Abduction	Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^{\tau}$	✓$^{\tau}$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^{\tau}$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

- ✓ Hypotheses are linked to the choice of Generative Model
- ✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

- **Causal structure**
  - Causal order is enough
- **Hidden confounding**
  - BGM and WhatIfGAN consider correlated exogeneous noises
  - NCF and NCM deal with semi-Markovian DAGs
- **Data assumptions**
  - SCM-VAE, DEAR encode images intro causally linked attributes
  - NFs → diffeomorphic mechanisms assumptions
  - WhatIfGAN is designed to deal with variables of different dimensions

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification		Additional Hypotheses				Additional Guarantees
	SCM class	DGM class*	Causal Structure	Hidden Confounder	Data Assumptions	Available Abduction	Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	$\checkmark^\sharp$	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	$\checkmark^\sharp$	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG^τ	$\checkmark^\tau$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

$^\sharp$A common cause is represented by an additional exogenous noise, $^\tau$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

✓ Hypotheses are linked to the choice of Generative Model
✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

- **Causal structure**
  - Causal order is enough
- **Hidden confounding**
  - BGM and WhatIfGAN consider correlated exogeneous noises
  - NCF and NCM deal with semi-Markovian DAGs
- **Data assumptions**
  - SCM-VAE, DEAR encode images intro causally linked attributes
  - NFs → diffeomorphic mechanisms assumptions
  - WhatIfGAN is designed to deal with variables of different dimensions
- **Abduction**
  - Only 7 methods implement the abduction step

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification: SCM class	Classification: DGM class*	Additional Hypotheses: Causal Structure	Additional Hypotheses: Hidden Confounder	Additional Hypotheses: Data Assumptions	Additional Hypotheses: Available Abduction	Additional Guarantees: Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^{\tau}$	✓$^{\tau}$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^{\tau}$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

TL;DR

- ✓ Hypotheses are linked to the choice of Generative Model
- ✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

➢ Causal structure
- Causal order is enough

➢ Hidden confounding
- BGM and WhatIfGAN consider correlated exogeneous noises
- NCF and NCM deal with semi-Markovian DAGs

➢ Data assumptions
- SCM-VAE, DEAR encode images intro causally linked attributes
- NFs → diffeomorphic mechanisms assumptions
- WhatIfGAN is designed to deal with variables of different dimensions

➢ Abduction
- Only 7 methods implement the abduction step

➢ Identifiability guarantees
- NCM identifiability result applicable to everyone except DCM
- DCM provides error bounds & L_3-identifiability but under sufficiency
- **NeuralID** algorithm to test for point identification

Table from [Poinsot *et al.*, 2024]

DSCM, hypotheses & guarantees

Method	Classification: SCM class	DGM class*	Additional Hypotheses: Causal Structure	Hidden Confounder	Data Assumptions	Available Abduction	Additional Guarantees: Identifiability, Expressivity, Bounds
NF-BGM [1]	BGM, NCM	IE	DAG	✓	-	Inversion	-
NF-DSCM [2]	BGM, NCM	IE	DAG	✗	f_i diffeomorphic	Inversion	-
GAN-NCM; MLE-NCM [3]	NCM	AI	DAG	✓♯	-	Sample Rejection	-
Causal-NF [4]	BGM, NCM	IE	Ordering	✗	f_i diffeomorphic	Inversion	Model id. up to invertible transformation of U
NCF [5]	BGM, NCM	IE	DAG	✓♯	f_i diffeomorphic	Inversion	-
CARFEL [6]	BGM, NCM	IE	DAG, $\emptyset$	✗	Affine autoregressive flow	Inversion	Model id. in bivariate case
iVGAE [7]	NCM	AE	DAG	✗	-	✗	-
VACA [8]	NCM	AE	DAG	✗	-	Encoding	$\mathcal{L}_2$-expressivity if the decoder is deep enough cf. Prop.2
DCM [9]	-	AE	Ordering	✗	-	Encoding	$\mathcal{L}_3$-id. with error bounds cf. Corollary 1 & 2
SCM-VAE [10]	NCM	AI	DAG	✗	Additive noise on attributes	✗	-
Causal-TGAN [11]	NCM	AI	DAG	✗	-	✗	-
CausalGAN [12]	NCM	AI	DAG	✗	-	✗	-
CFGAN [13]	NCM	AI	DAG	✗	Categ. outcome & sensitive feature	✗	-
DECAF [14]	NCM	AI	DAG	✗	-	✗	-
WhatIfGAN [15]	NCM	AI	DAG	✓	-	✗	-
CGN [16]	NCM	AI	DAG$^{\tau}$	✓$^{\tau}$	Image with attributes	✗	-
DEAR [17]	NCM	AI	Ordering	✗	High-dimensional data with attributes	✗	Data to attribute encoder disentanglement

♯A common cause is represented by an additional exogenous noise, $^{\tau}$Only a confounded trivariate DAG is considered
*Invertible Explicit (IE), Amortised Explicit (AE), and Amortised Implicit (AI)

Table 2: Hypotheses and guarantees of deep structural causal models. The classification (Figure 1) enables one to spot the identifiability results inherited by the SCM class and the compatible abduction step procedures.

Table from [Poinsot *et al.*, 2024]

TL;DR

✓ Hypotheses are linked to the choice of Generative Model
✓ Except NCM & BGM L_3-identifiability results, few to no guarantees

arXiv:2405.05025

➢ Causal structure

- Causal order is enough

➢ Hidden confounding

- BGM and WhatIfGAN consider correlated exogeneous noises
- NCF and NCM deal with semi-Markovian DAGs

➢ Data assumptions

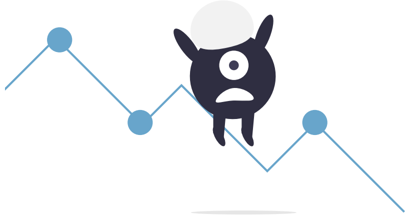
- SCM-VAE, DEAR encode images intro causally linked attributes
- NFs → diffeomorphic mechanisms assumptions
- WhatIfGAN is designed to deal with variables of different dimensions

➢ Abduction

- Only 7 methods implement the abduction step

➢ Identifiability guarantees

- NCM identifiability result applicable to everyone except DCM
- DCM provides error bounds & L_3-identifiability but under sufficiency
- **NeuralID** algorithm to test for point identification



"Theory is a good first filtering. However, I don't want to face a huge drop in performances."

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^{\star}$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^{\sharp}$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^{\sharp}$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^{\tau}$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^{\star}$	✗	-

* Disentanglement, $^{\sharp}$ Fairness debiasing by intervention, $^{\tau}$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

- ✓ High heterogeneity in the evaluation
- ✓ Applied to sensitive use cases

arXiv:2405.05025

Table from [Poinsot *et al.*, 2024]

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^{\star}$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^{\sharp}$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^{\sharp}$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^{\tau}$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^{\star}$	✗	-

* Disentanglement, $^{\sharp}$ Fairness debiasing by intervention, $^{\tau}$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

- ✓ High heterogeneity in the evaluation
- ✓ Applied to sensitive use cases

arXiv:2405.05025

➢ **Empirical evaluation**

- **High heterogeneity**: datasets, PCH of the task, metrics, …
- **Lack of a unified benchmark**
- Datasets are more suited for L_2 tasks
  - fMRI, ColorMNIST, bnlearn → intervention estimation
  - Pendulum, CelebA → disentanglement
  - Morpho-MNIST → counterfactual estimation
- Simulations lack sources of randomness (DAG, noise distrib, …)

Table from [Poinsot *et al.*, 2024]

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^{\star}$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^{\sharp}$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^{\sharp}$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^{\tau}$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^{\star}$	✗	-

* Disentanglement, $^{\sharp}$ Fairness debiasing by intervention, $^{\tau}$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

- ✓ High heterogeneity in the evaluation
- ✓ Applied to sensitive use cases

arXiv:2405.05025

➢ Empirical evaluation

- **High heterogeneity**: datasets, PCH of the task, metrics, …
- **Lack of a unified benchmark**
- Datasets are more suited for L_2 tasks
  - fMRI, ColorMNIST, bnlearn → intervention estimation
  - Pendulum, CelebA → disentanglement
  - Morpho-MNIST → counterfactual estimation
- Simulations lack sources of randomness (DAG, noise distrib, …)

➢ Applications

- **Fairness**
  - Counterfactual fairness evaluation
  - Fair prediction in & pre-processing

Table from [Poinsot *et al.*, 2024]

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^\star$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^\sharp$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^\sharp$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^\tau$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^\star$	✗	-

* Disentanglement, $^\sharp$ Fairness debiasing by intervention, $^\tau$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

✓ High heterogeneity in the evaluation
✓ Applied to sensitive use cases

arXiv:2405.05025

➢ Empirical evaluation

- **High heterogeneity**: datasets, PCH of the task, metrics, …
- **Lack of a unified benchmark**
- Datasets are more suited for L_2 tasks
  - fMRI, ColorMNIST, bnlearn → intervention estimation
  - Pendulum, CelebA → disentanglement
  - Morpho-MNIST → counterfactual estimation
- Simulations lack sources of randomness (DAG, noise distrib, …)

➢ Applications

- **Fairness**
  - Counterfactual fairness evaluation
  - Fair prediction in & pre-processing
- **Explainability**
  - Counterfactual explanations
  - Scientific discovery

Table from [Poinsot *et al.*, 2024]

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^\star$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^\sharp$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^\sharp$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^\tau$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^\star$	✗	-

* Disentanglement, $^\sharp$ Fairness debiasing by intervention, $^\tau$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

- ✓ High heterogeneity in the evaluation
- ✓ Applied to sensitive use cases

arXiv:2405.05025

➢ Empirical evaluation

- **High heterogeneity**: datasets, PCH of the task, metrics, …
- **Lack of a unified benchmark**
- Datasets are more suited for L_2 tasks
  - fMRI, ColorMNIST, bnlearn → intervention estimation
  - Pendulum, CelebA → disentanglement
  - Morpho-MNIST → counterfactual estimation
- Simulations lack sources of randomness (DAG, noise distrib, …)

➢ Applications

- **Fairness**
  - Counterfactual fairness evaluation
  - Fair prediction in & pre-processing
- **Explainability**
  - Counterfactual explanations
  - Scientific discovery
- **Machine Learning robustness/generalization**
  - Out of domain data augmentation
  - Realistic dataset generation

Table from [Poinsot *et al.*, 2024]

DSCM, evaluation & applications

Method	Dataset	PCH	DSCM Comparison	Applications
NF-BGM [1]	Ellips generation simulations	$\mathcal{L}_3$	✗	Video streaming simulations for adaptive bitrate
NF-DSCM [2]	Morpho-MNIST	$\mathcal{L}_3$	✗	Scientific discovery [Yu *et al.*, 2023], cyber-security data generation [Agrawal *et al.*, 2024]
GAN-NCM; MLE-NCM [3]	Simulated SCMs	$\mathcal{L}_3$	✓GAN, MLE NCM	-
Causal-NF [4]	Simulated SCMs	$\mathcal{L}_3$	✓VACA, CARFEL	Counterfactual fairness & fair regularization of classifier
NCF [5]	Salary simulations using a simulated SCM	$\mathcal{L}_3$	✗	Counterfactual fairness and explainability
CARFEL [6]	4-dimentional polynomial simulated SCM, fMRI	$\mathcal{L}_2$ & $\mathcal{L}_3$	✗	-
iVGAE [7]	ASIA	$\mathcal{L}_2$	✗	-
VACA [8]	Simulated SCMs	$\mathcal{L}_3$	✗	Counterfactual fairness
DCM [9]	Simulated SCMs, fMRI	$\mathcal{L}_3$	✓VACA, CARFEL	-
SCM-VAE [10]	Pendulum, CelebA	$\mathcal{L}_2$	✗	-
Causal-TGAN [11]	ASIA, Child, ALARM, Insurance; Adult, Census, News	$\mathcal{L}_1$	✗	In-domain data augmentation
CausalGAN [12]	CelebA	$\mathcal{L}_2^{\star}$	✗	Out-of-domain data augmentation
CFGAN [13]	Adult	$\mathcal{L}_2^{\sharp}$	✗	Fairness debiasing
DECAF [14]	Adult, Credit Approval	$\mathcal{L}_2^{\sharp}$	✓CFGAN	Fairness debiasing
WhatIfGAN [15]	Color-MNIST	$\mathcal{L}_2$	✓NCM	-
CGN [16]	Color-MNIST; ImageNet [Deng *et al.*, 2009]	$\mathcal{L}_2^{\tau}$	✗	Out-of-domain data augmentation
DEAR [17]	Pendulum, CelebA	$\mathcal{L}_2^{\star}$	✗	-

* Disentanglement, $^{\sharp}$ Fairness debiasing by intervention, $^{\tau}$ Invariant classification after intervention

Table 3: Existing evaluations and applications of DSCMs

TL;DR

- ✓ High heterogeneity in the evaluation
- ✓ Applied to sensitive use cases

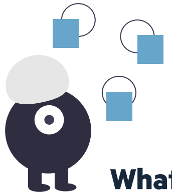
arXiv:2405.05025

➢ Empirical evaluation

- **High heterogeneity**: datasets, PCH of the task, metrics, …
- **Lack of a unified benchmark**
- Datasets are more suited for L_2 tasks
  - fMRI, ColorMNIST, bnlearn → intervention estimation
  - Pendulum, CelebA → disentanglement
  - Morpho-MNIST → counterfactual estimation
- Simulations lack sources of randomness (DAG, noise distrib, …)

➢ Applications

- **Fairness**
  - Counterfactual fairness evaluation
  - Fair prediction in & pre-processing
- **Explainability**
  - Counterfactual explanations
  - Scientific discovery
- **Machine Learning robustness/generalization**
  - Out of domain data augmentation
  - Realistic dataset generation



Well … great.

So, what?

What can or should I do?

Table from [Poinsot *et al.*, 2024]

DSCM, challenges & future directions

For researchers

For practitioners

DSCM, challenges & future directions

For researchers

For practitioners

DSCM, challenges & future directions

For researchers

For practitioners

- **Sensitive applications**
  - The causal graph must be validated by experts beforehand
  - **NeuralID** enables to test point-identification
  - **Sensitivity analysis** is crucial
    - Unobserved confounding
    - Selection bias
  - For decision-makers, important indicators are still missing:
    - **uncertainty quantification**
    - **error measures**

DSCM, challenges & future directions

For researchers

For practitioners

➢ Sensitive applications

- The causal graph must be validated by experts beforehand
- **NeuralID** enables to test point-identification
- **Sensitivity analysis** is crucial
  - Unobserved confounding
  - Selection bias
- For decision-makers, important indicators are still missing:
  - **uncertainty quantification**
  - **error measures**

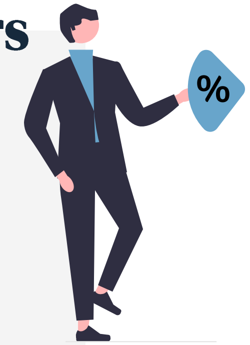
➢ Realistic causal simulations

- Opportunity to simulate causal data close to real ones
  - New way of **benchmarking causal inference methods** on various types of data
  - Causal data augmentation for ML generalization

DSCM, challenges & future directions

For researchers

For practitioners



- **Sensitive applications**
  - Experts validation
  - **NeuralID** for point-identification
  - **Sensitivity analysis** is crucial
  - Uncertainty quantification
  - Error measurement

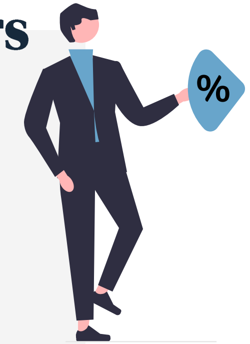
- **Realistic causal simulations**
  - Benchmarking causal inference
  - Causal data augmentation

DSCM, challenges & future directions

For researchers

- **Lack of evaluation**
  - Lack of a proper **benchmark**
    - Simulations to have a ground truth
    - Randomly generating DAGs, noise, mechanisms
  - Lack of a complete **evaluation strategy**
    - Data efficiency
    - Computing time
    - Robustness to selection bias, imbalanced data, imperfect causal knowledge, …
  - Lack of comparison of the **abduction** steps procedures

For practitioners



- **Sensitive applications**
  - Experts validation
  - **NeuralID** for point-identification
  - **Sensitivity analysis** is crucial
  - Uncertainty quantification
  - Error measurement
- **Realistic causal simulations**
  - Benchmarking causal inference
  - Causal data augmentation

DSCM, challenges & future directions

For researchers

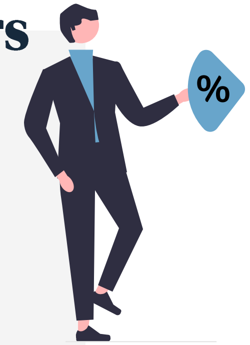
➢ Lack of evaluation

- Lack of a proper **benchmark**
  - Simulations to have a ground truth
  - Randomly generating DAGs, noise, mechanisms
- Lack of a complete **evaluation strategy**
  - Data efficiency
  - Computing time
  - Robustness to selection bias, imbalanced data, imperfect causal knowledge, …
- Lack of comparison of the **abduction** steps procedures

➢ From point to partial identification

- For point identification, **un-testable hypotheses** are taken:
  - Known causal structure
  - No selection bias
  - Unidimentional exogeneous noise
- Some work already done on **partial identification**
  - Discrete SCMs [Zaffalon et al, 2020; Zhang et al., 2022]
  - Non-informative bounds L_3-queries [Melnychuk et al., 2023]
- Whenever partial identification is impossible or too hard to get, **sensitivity analysis** is still possible
  - Unobserved confounding [Schröder at al., 2024, Frauen et al., 2024]

For practitioners



➢ Sensitive applications

- Experts validation
- **NeuralID** for point-identification
- **Sensitivity analysis** is crucial
- Uncertainty quantification
- Error measurement

➢ Realistic causal simulations

- Benchmarking causal inference
- Causal data augmentation

DSCM, challenges & future directions

For researchers

➢ Lack of evaluation

- Lack of a proper **benchmark**
  - Simulations to have a ground truth
  - Randomly generating DAGs, noise, mechanisms
- Lack of a complete **evaluation strategy**
  - Data efficiency
  - Computing time
  - Robustness to selection bias, imbalanced data, imperfect causal knowledge, …
- Lack of comparison of the **abduction** steps procedures

➢ From point to partial identification

- For point identification, **un-testable hypotheses** are taken:
  - Known causal structure
  - No selection bias
  - Unidimentional exogeneous noise
- Some work already done on **partial identification**
  - Discrete SCMs [Zaffalon et al, 2020; Zhang et al., 2022]
  - Non-informative bounds L_3-queries [Melnychuk et al., 2023]
- Whenever partial identification is impossible or too hard to get, **sensitivity analysis** is still possible
  - Unobserved confounding [Schröder at al., 2024, Frauen et al., 2024]

For practitioners

➢ Sensitive applications

- The causal graph must be validated by experts beforehand
- **NeuralID** enables to test point-identification
- **Sensitivity analysis** is crucial
  - Unobserved confounding
  - Selection bias
- For decision-makers, important indicators are still missing:
  - **uncertainty quantification**
  - **error measures**

➢ Realistic causal simulations

- Opportunity to simulate causal data close to real ones
  - New way of **benchmarking causal inference methods** on various types of data
  - Causal data augmentation for ML generalization

Questions

Ekimetrics.

References

[Pearl, 2000] Judea Pearl. Models, Reasoning, and Inference. Cambridge University Press, 2000.

[Pearl, 2009] Judea Pearl. Models, Causal Inference in statistics: an overview. Statistics Surveys, 2009.

[Spirtes *et al.*, 2000] P. Spirtes, C. Glymour, and R. Scheines. Causation, Prediction, and Search. Springer-Verlag, 1993. (2nd ed. MIT Press 2000).

[Bollen, 1989] K. A. Bollen. Structural Equations with Latent Variables. John Wiley & Sons, 1989.

[Pearl and Mackenzie, 2018] Judea Pearl and Dana Mackenzie. The book of why: the new science of cause and effect. Basic books, 2018.

[Xia *et al.*, 2021] Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, and Elias Bareinboim. The causal-neural connection: Expressiveness, learnability, and inference. In Advances in Neural Information Processing Systems, 2021.

[Zhou *et al.*, 2023] Guanglin Zhou, Shaoan Xie, Guangyuan Hao, Shiming Chen, Biwei Huang, Xiwei Xu, Chen Wang, Liming Zhu, Lina Yao, and Kun Zhang. Emerging synergies in causality and deep generative models: A survey. arXiv:2301.12351, 2023.

[Komanduri *et al.*, 2023] Aneesh Komanduri, Xintao Wu, Yongkai Wu, and Feng Chen. From identifiable causal representations to controllable counterfactual generation: A survey on causal generative modeling. arXiv:2310.11011, 2023.

[Kaddour *et al.*, 2022] Jean Kaddour, Aengus Lynch, Qi Liu, Matt J. Kusner, and Ricardo Silva. Causal machine learning: A survey and open problems. arXiv:2206.15475, 2022.

[Pawlowski *et al.*, 2020] Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In Advances in Neural Information Processing Systems, 2020.

[Xia *et al.*, 2023] Kevin Muyuan Xia, Yushu Pan, and Elias Bareinboim. Neural causal models for counterfactual identification and estimation. In 11th International Conference on Learning Representations, 2023.

[Nasr-Esfahany *et al.*, 2023] Arash Nasr-Esfahany, Mohammad Alizadeh, and Devavrat Shah. Counterfactual identifiability of bijective causal models. In 40th International Conference on Machine Learning, 2023.

References

[Poinsot *et al.*, 2024] Audrey Poinsot, Alessandro Leite, Nicolas Chesneau, Michèle Sébag, Marc Schoenauer. Learning Structural Causal Models through Deep Generative Models: Methods, Guarantees, and Challenges. In International Joint Conference on Artificial Intelligence, 2024.

[1] Arash Nasr-Esfahany, Mohammad Alizadeh, and Devavrat Shah. Counterfactual identifiability of bijective causal models. In 40th International Conference on Machine Learning, 2023.

[2] Nick Pawlowski, Daniel Coelho de Castro, and Ben Glocker. Deep structural causal models for tractable counterfactual inference. In Advances in Neural Information Processing Systems, 2020.

[3] Kevin Muyuan Xia, Yushu Pan, and Elias Bareinboim. Neural causal models for counterfactual identification and estimation. In 11th International Conference on Learning Representations, 2023.

[4] Adrian Javaloy, Pablo Sanchez-Martin, and Isabel Valera. Causal normalizing flows: from theory to practice. Advances in Neural Information Processing Systems, 2023.

[5] Alvaro Parafita and Jordi Vitria. Causal inference with deep causal graphs. arXiv:2006.08380, 2020.

[6] Ilyes Khemakhem, Ricardo Monti, Robert Leech, and Aapo Hyvarinen. Causal autoregressive flows. In 24th International Conference on Artificial Intelligence and Statistics, 2021.

[7] Matej Zecevic, Devendra Singh Dhami, Petar Velickovic, and Kristian Kersting. Relating graph neural networks to structural causal models. arXiv:2109.04173, 2021.

[8] Pablo Sanchez-Martin, Miriam Rateike, and Isabel Valera. VACA: designing variational graph autoencoders for causal queries. In AAAI Conference on Artificial Intelligence, 2022.

[9] Patrick Chao, Patrick Blöbaum, and Shiva Prasad Kasiviswanathan. Interventional and counterfactual inference with diffusion models. arxiv:2302.00860, 2023.

[10] Aneesh Komanduri, Yongkai Wu, Wen Huang, Feng Chen, and Xintao Wu. SCM-VAE: Learning identifiable causal representations via structural knowledge. In IEEE International Conference on Big Data, 2022.

[11] Bingyang Wen, Yupeng Cao, Fan Yang, Koduvayur Subbalakshmi, and Rajarathnam Chandramouli. Causal-TGAN: Modeling tabular data using causally-aware GAN. In ICLR Workshop on Deep Generative Models for Highly Structured Data, 2022.

[12] Murat Kocaoglu, Christopher Snyder, Alexandros G. Dimakis, and Sriram Vishwanath. CausalGAN: Learning causal implicit generative models with adversarial training. In International Conference on Learning Representations, 2018.

References

[13] Depeng Xu, Yongkai Wu, Shuhan Yuan, Lu Zhang, and Xintao Wu. Achieving causal fairness through generative adversarial networks. In 28th International Joint Conference on Artificial Intelligence, 2019.

[14] Boris van Breugel, Trent Kyono, Jeroen Berrevoets, and Mihaela van der Schaar. DECAF: Generating fair synthetic data using causally-aware generative networks. In Advances in Neural Information Processing Systems, 2021.

[15] Md Musfiqur Rahman and Murat Kocaoglu. Towards modular learning of deep causal generative models. In ICML Workshop on Structured Probabilistic Inference & Generative Modeling, 2023.

[16] Axel Sauer and Andreas Geiger. Counterfactual generative networks. In International Conference on Learning Representations, 2021.

[17] Xinwei Shen, Furui Liu, Hanze Dong, Qing Lian, Zhitang Chen, and Tong Zhang. Weakly supervised disentangled generative causal representation learning. Journal of Machine Learning Research, 2022.

[18] Patrik Hoyer, Dominik Janzing, Joris Mooij, Jonas Peters and Bernhard Schölkopf. Nonlinear causal discovery with additive noise modes. In Advances in Neural Information Processing Systems, 2008.

[19] Alexander Immer, Christoph Schultheiss, Julia E Vogt, Bernhard Schölkopf, Peter Bühlmann and Alexander Marx. On the Identifiability and Estimation of Causal Location-Scale Noise Models. In International Conference of Machine Learning, 2023.

[20] Kun Zhang and Aapo Hyvärinen. On the Identifiability of the Post-Nonlinear Causal Model. In Conference on Uncertainty in Artificial Intelligence, 2009.

[Zaffalon *et al.*, 2020] Marco Zaffalon, Alessandro Antonucci, and Rafael Cabanas. Structural causal models are (solvable by) credal networks. In 10th International Conference on Probabilistic Graphical Models, 2020.

[Zhang *et al.*, 2022] Junzhe Zhang, Jin Tian, and Elias Bareinboim. Partial counterfactual identification from observational and experimental data. In 39th International Conference on Machine Learning, 2022.

[Schrönder *et al.*, 2024] Maresa Schr¨oder, Dennis Frauen, and Stefan Feuerriegel. Causal fairness under unobserved confounding: a neural sensitivity framework. 12th International Conference on Learning Representations, 2023.

[Frauen *et al.*, 2024] Dennis Frauen, Fergus Imrie, Alicia Curth, Valentyn Melnychuk, Stefan Feuerriegel, and Mihaela van der Schaar. A neural framework for generalized causal sensitivity analysis. In 12th International Conference on Learning Representations, 2024.

[Messerli, 2012] Franz H. Messerli. Chocolate Consumption, Cognitive Function, and Nobel Laureates. The New England Journal of Medicine, 2012.