

# **Causal Inference: Towards Explainable, Generalizable, and Trustworthy AI**

**Adèle H. Ribeiro**

<https://adele.github.io/> | [adele.ribeiro@uni-marburg.de](mailto:adele.ribeiro@uni-marburg.de)

Data Science in Biomedicine  
Faculty of Mathematics and Computer Science  
Philipps-Universität Marburg

**Nordic Probabilistic AI School (ProbAI)**  
**June 16th, 2023**

# Supporting materials:

---

**Prob-AI GitHub (Day 5):** [https://github.com/probabilisticai/probai-2023/tree/main/day\\_5](https://github.com/probabilisticai/probai-2023/tree/main/day_5)

- **Example 1:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_BD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_BD.ipynb)
- **Example 2:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_NonBD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_NonBD.ipynb)

**Please, open the notebooks and install the libraries (it'll take about 15 min!)**

# Recent Breakthroughs in AI

---

- We can design systems that make **predictions** extremely well in high-dimensional settings.
- In particular, there are huge progresses in reinforcement learning, computer vision, and natural language processing.

# Recent Breakthroughs in AI

## NEWSLETTERS

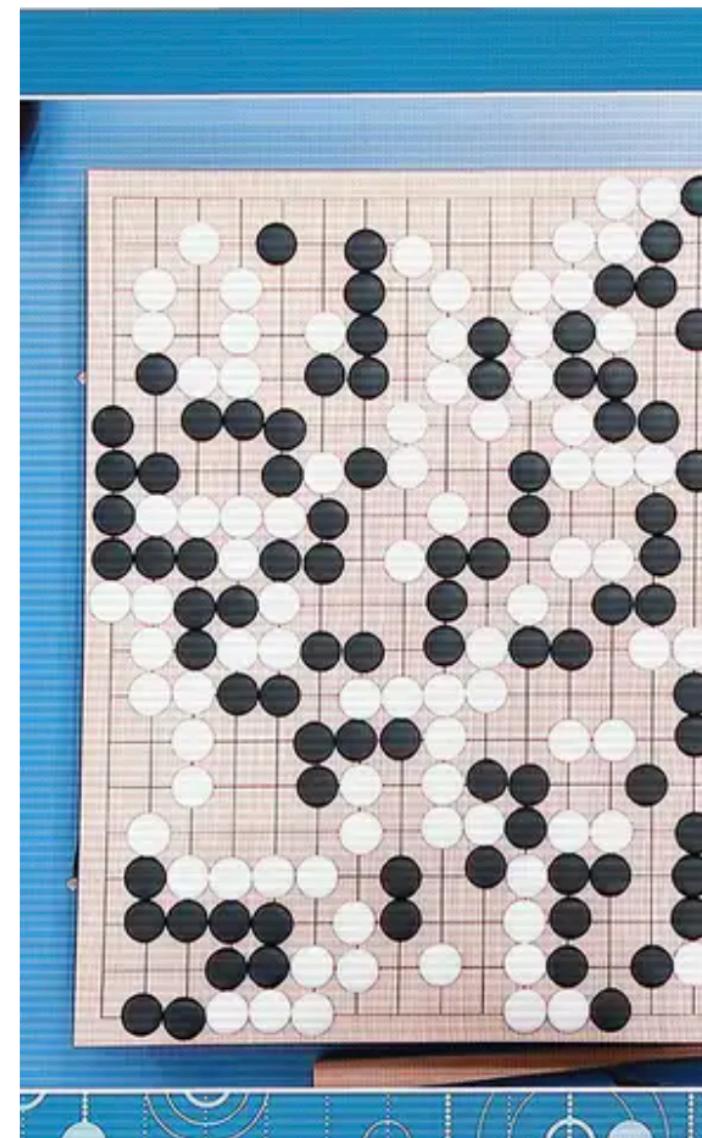
Sign up to read our regular email newsletters

News Podcasts Video **Technology**

## DeepMind's AI player in latest Driverless Cars to San Jose

 TECHNOLOGY 23 May 2017

By Matt Reynolds



Aside from the lidar range-finder unit on its vehicles, including this Lexus hybrid, look reaso

The New York Times

## NEWSLETTERS

Sign up to read our regular email newsletters

News Podcasts Video **Technology** Space Physics Health More ▾ Shop Courses Events 1

## AI system is better than human doctors at

 TECHNOLOGY 1 January 202

By Jessica Hamzelou



Become an A.I. Expert How Chatbots Work Why Chatbots 'Hallucinate' Testing Google Bard

NewScientist

The New York Times

## THE SHIFT

## GPT-4 Is Exciting and Scary

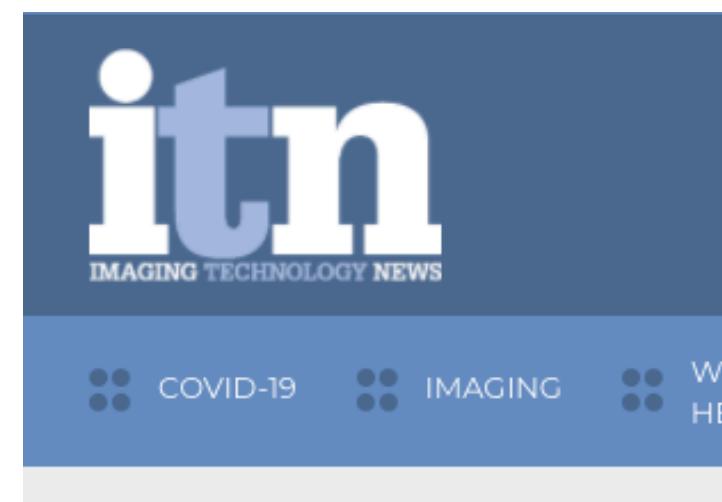
Today, the new language model from OpenAI may not seem all that dangerous. But the worst risks are the ones we cannot anticipate.

Does this mean we are done?

Is there anything missing?



# Current Challenges in AI



nature

Explore content ▾ Journal information ▾ Publish with us ▾ Subscribe

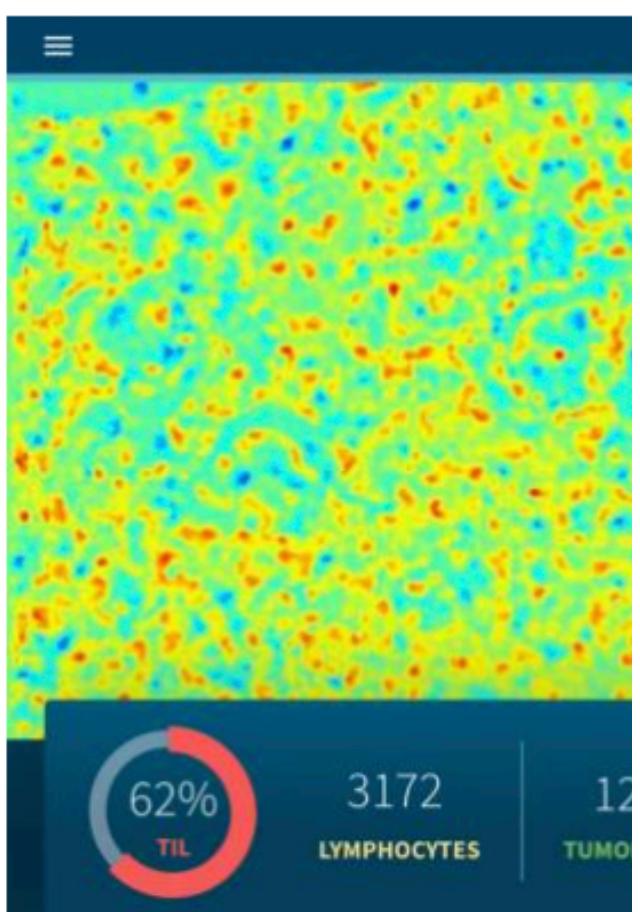
COVID-19 IMAGING WORKING GROUP

nature > outlook > article

NEWS | ARTIFICIAL INTELLIGENCE | MARCH 2023

Making the Role of AI in

Analysis system for the diagnosis of



Without careful implemen

inequality.

nature

Explore content ▾

About the journal ▾

Publish with us ▾

Subscribe

nature > outlook > article

OUTLOOK | 24 February 2023

## A fairer way forward for AI in health care

## Why artificial intelligence needs to understand consequences

A machine with a grasp of cause and effect could learn more like a human, through imagination and regret.

# Why causality is so important?

---

Causality is an essential component in the development of the new generation of Artificial Intelligence methods, allowing important capabilities such as

**Effect identifiability:** can determine the effect of *unrealized* interventions rather than just predicting an outcome (i.e., can distinguish between association and causation).

**Explainability:** provides a better understanding of the underlying mechanisms.

**Fairness:** captures and disentangles any mechanisms of discrimination that may be present, including direct, indirect-mediated, and indirect-confounded.

**Generalizability:** allows the transportability of causal effects across different domains.

**Data Fusion:** provides language and theory to cohesively combine prior knowledge and data from multiple and heterogeneous studies.

# Causal Data Science

---

Goal is to develop language, criteria, and algorithms for:

- **Data-Fusion:** cohesively combining heterogenous datasets,
- **Causal Inference:** inferring the effects of interventions, and
- **Decision-Making:** making robust and generalizable decisions.



## Causal inference and the data-fusion problem

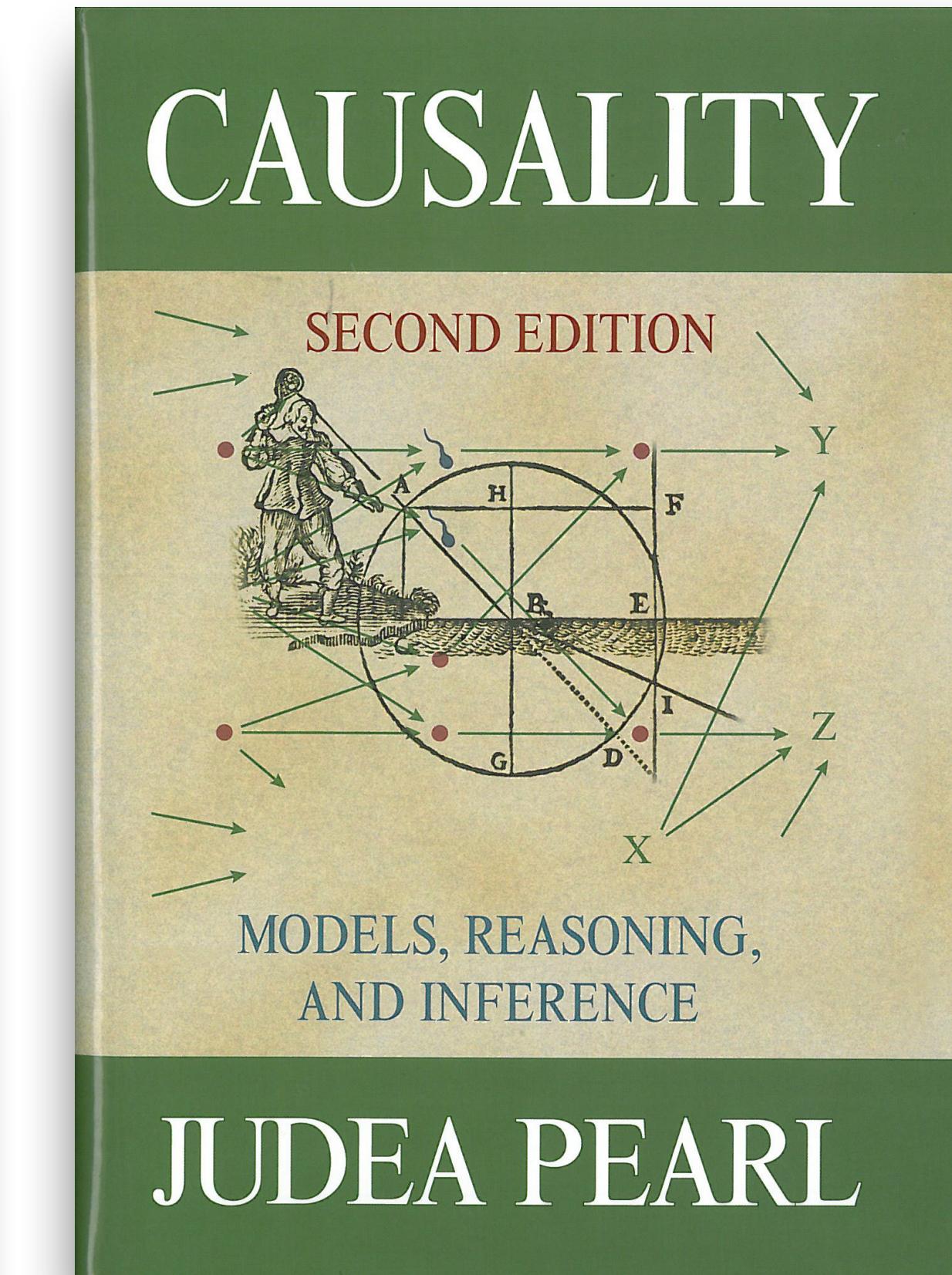
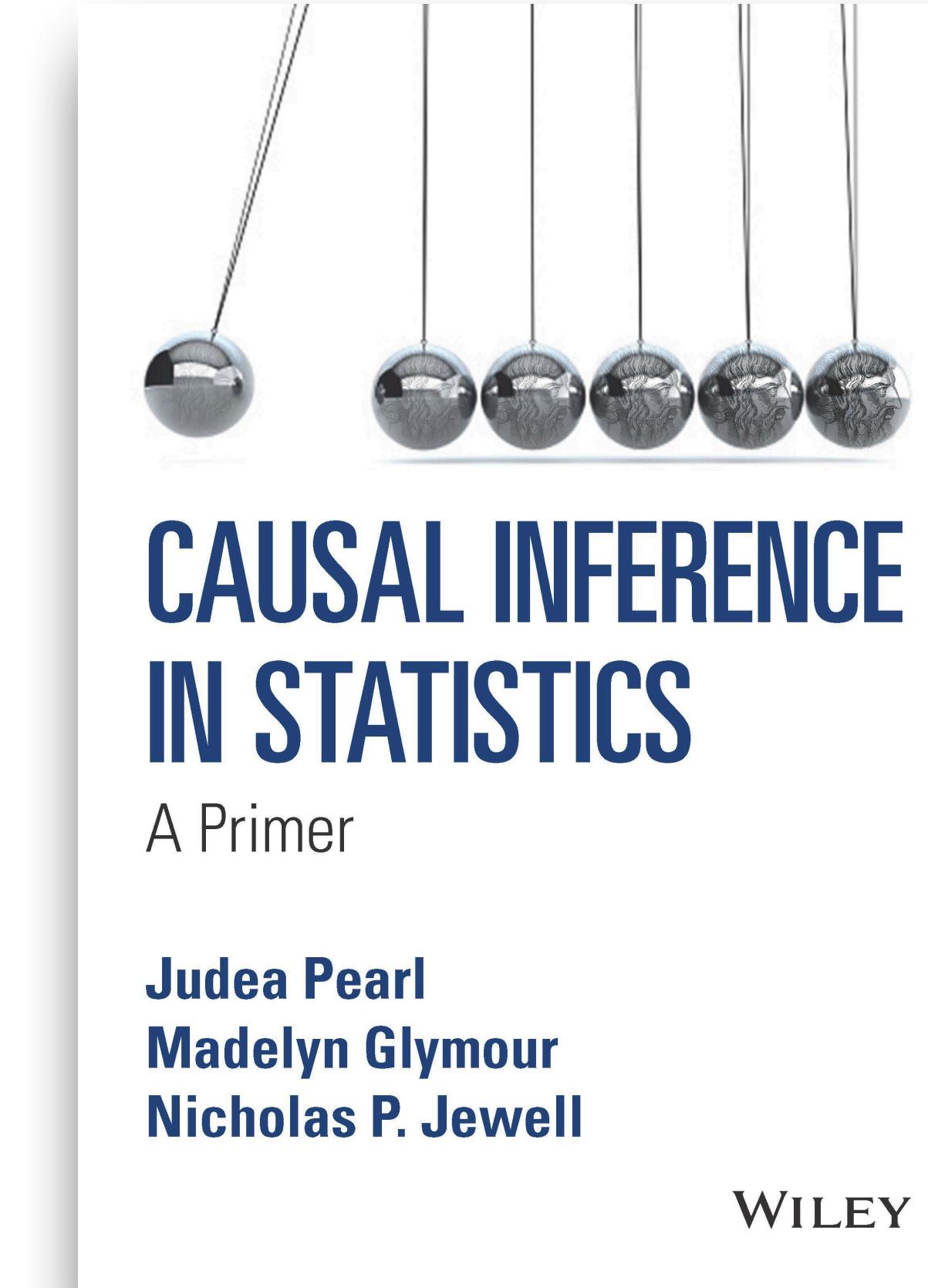
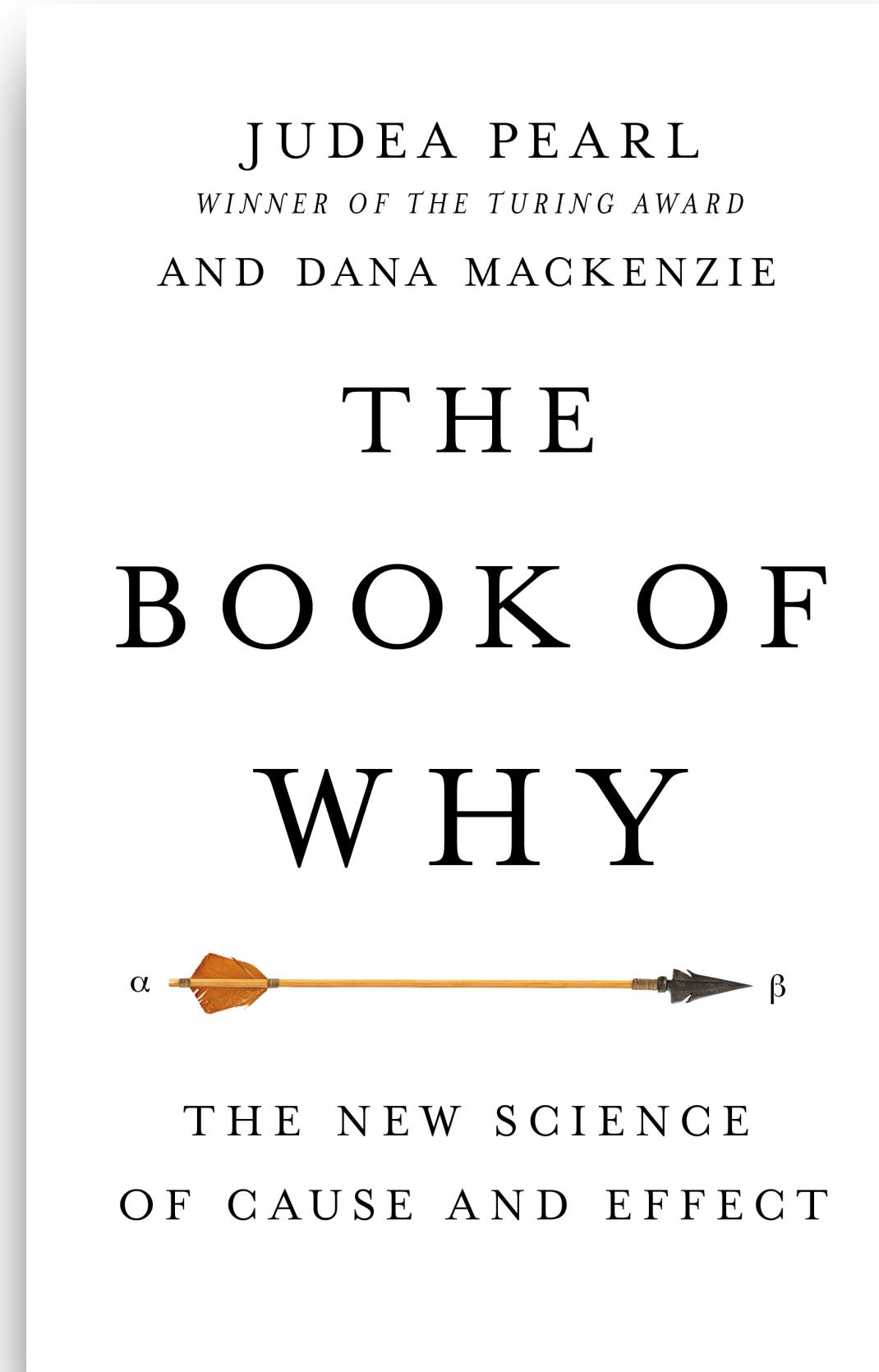
Elias Bareinboim<sup>a,b,1</sup> and Judea Pearl<sup>a</sup>

<sup>a</sup>Department of Computer Science, University of California, Los Angeles, CA 90095; and <sup>b</sup>Department of Computer Science, Purdue University, West Lafayette, IN 47907

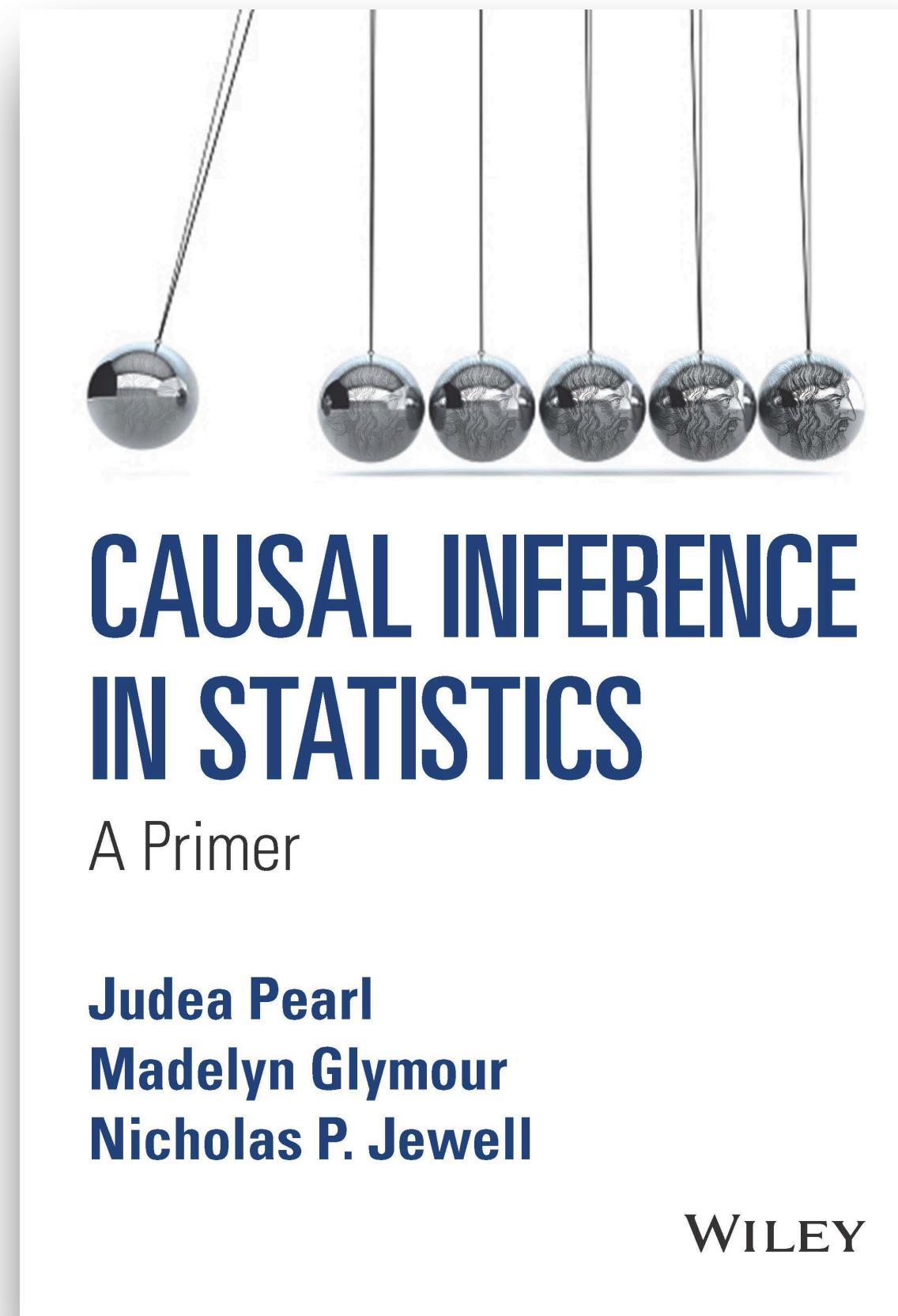
Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved March 15, 2016 (received for review June 29, 2015)

<http://causalfusion.net>

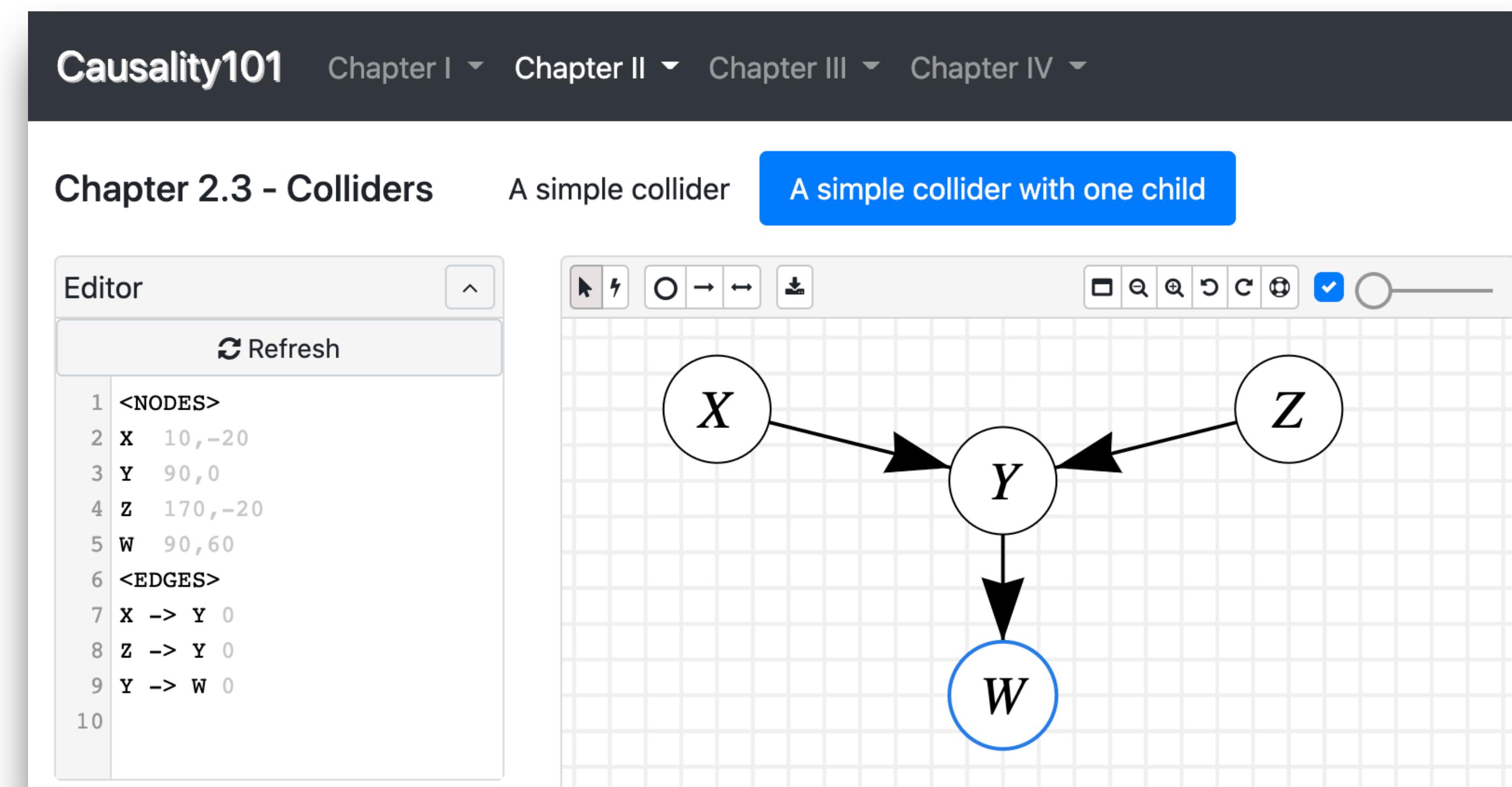
# Causality Theory by Judea Pearl



# Causality Theory by Judea Pearl



<https://causality101.net/>



# **Prediction vs Effect of Interventions**

**Statistical Association vs Causation**

# Predictive Tasks

**Task:** Can I guess how serious/big is the fire by the number of firefighters in action?

**Yes!**

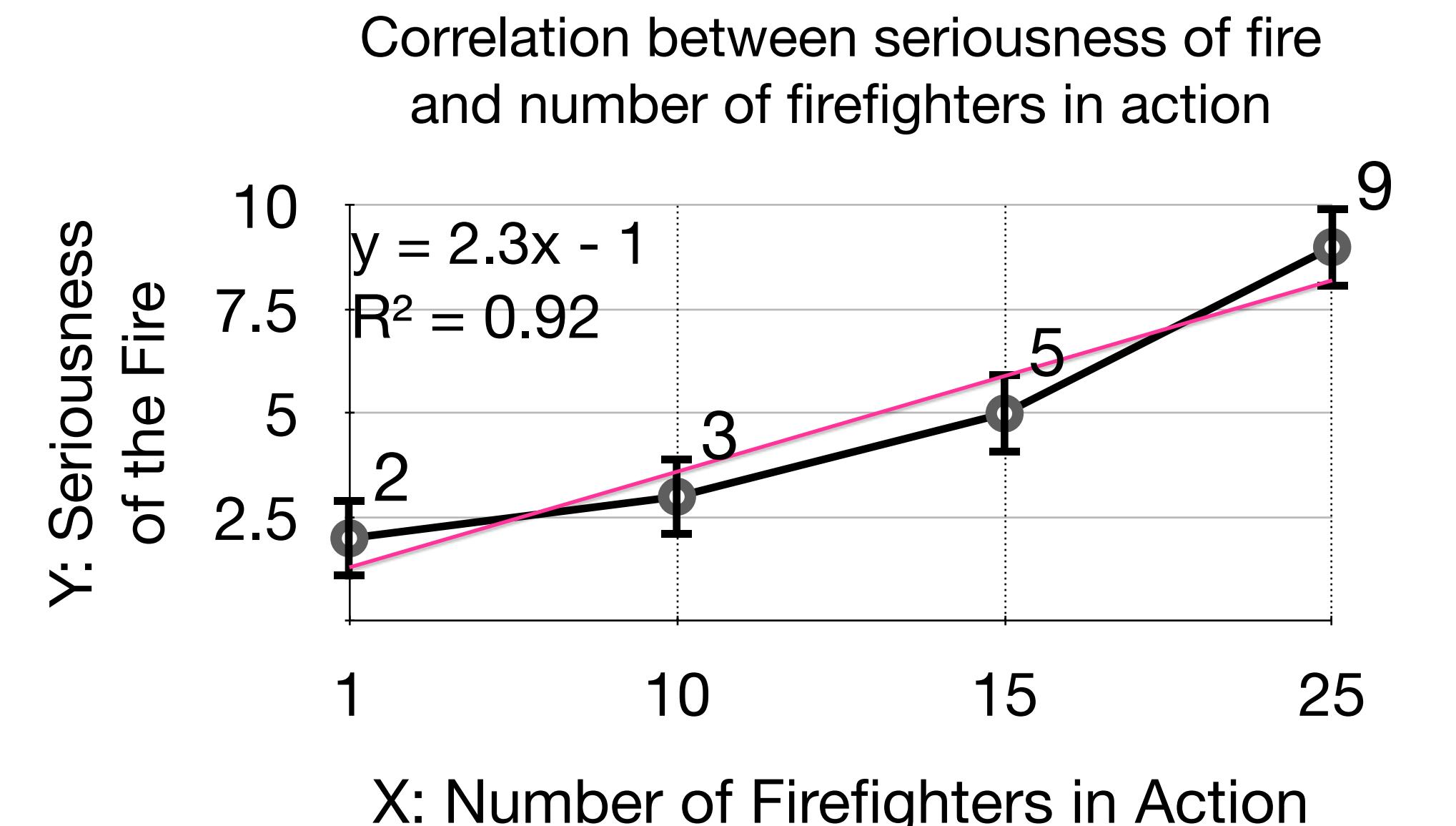
$X$ : Number of firefighters in action  
 $Y$ : Seriousness of fire

$\rho_{XY} \neq 0 \implies X \text{ is a good predictor of } Y$

$$P(Y = y | \textcolor{red}{X} = \textcolor{red}{x}) \neq P(Y = y)$$



**Observational  
Probability Distribution**



**Conclusion:** The seriousness of fire increases with the number of firefighters.

# Prediction $\Rightarrow$ Decision-Making?

---

**Conclusion:** The seriousness of the fire increases with the number of firefighters.

**The fewer firefighters, the weaker the fire.**



**Should we decrease the number of firefighters to reduce the fire?**

# Effect of Interventions

---

$X$ : Number of firefighters in action

$Y$ : Seriousness of fire

$Y$  is not a function of  $X$

In other words,  **$Y$  is not caused by  $X$**

$$\begin{cases} X = f_X(Y, U_X, U_{XY}) \\ Y = f_Y(U_Y, U_{XY}) \end{cases}$$

**Underlying Model**

# Effect of Interventions

---

$X$ : Number of firefighters in action

$Y$ : Seriousness of fire

$$\begin{cases} X = f_X(Y, U_X, U_{XY}) \\ Y = f_Y(U_Y, U_{XY}) \end{cases}$$

**Underlying Model**

$X = x$

$Y$  is not a function of  $X$

In other words,  **$Y$  is not caused by  $X$**

Changing  $X$  won't change the value of  $Y$

$$P(Y = y | \text{do}(X = x)) = P(Y = y)$$

Interventional  
Probability Distribution

The action/intervention on  $X$ ,  $\text{do}(X = x)$   
is independent of  $Y$

**Conclusion:** we cannot change the seriousness of the fire by changing the number of firefighters.

# **Structural Causal Model (SCM)**

**EXPLAINABILITY AND THE DATA GENERATING MODEL**

# Structural Causal Model (SCM)

---

**Definition:** A structural causal model  $\mathcal{M}$  (or, data generating model) is a tuple  $\langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$ , where

- $\mathbf{V} = \{V_1, \dots, V_n\}$ : are endogenous variables
- $\mathbf{U} = \{U_1, \dots, U_m\}$ : are exogenous variables
- $\mathcal{F} = \{f_1, \dots, f_n\}$ : are functions determining  $\mathbf{V}$ , i.e.,  $v_i \leftarrow f_i(pa_i, u_i)$ , where  $Pa_i \subseteq \mathbf{V}$  are endogenous causes (parents) of  $V_i$  and  $U_i \subseteq \mathbf{U}$  are exogenous causes of  $V_i$ .
- $P(\mathbf{U})$  is the probability distribution over  $\mathbf{U}$ .

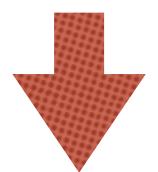
**Assumption:**  $\mathcal{M}$  is recursive, i.e., there are no feedback (cyclic) mechanisms.

# Effect of Interventions in SCMs

---

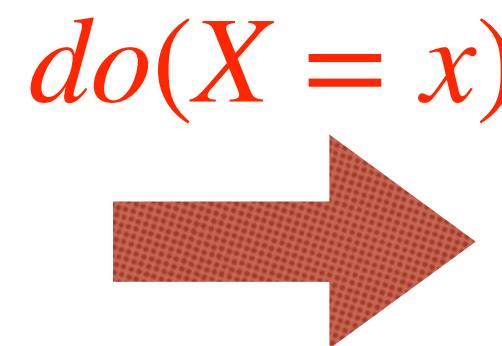
## Pre-Interventional/ Observational SCM

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = f_X(U_X, \textcolor{red}{U}_{XY}) \\ Y = f_Y(\textcolor{red}{X}, U_Y, \textcolor{red}{U}_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



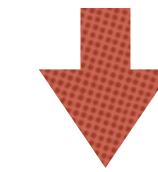
Observational  
Distribution

$$P(\mathbf{V}) \doteq P_{\mathcal{M}}(\mathbf{V})$$



## Post-Interventional / Interventional SCM

$$\mathcal{M}_x = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} \textcolor{red}{X} = x \\ Y = f_Y(x, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



Interventional  
Distribution



$$P(\mathbf{V} | \textcolor{red}{do}(X = x)) \doteq P_{\mathcal{M}_x}(\mathbf{V})$$

Can we ***predict*** better the value of  $Y$  after  
***observing*** que  $X = x$ ?

$$P(Y = y | \textcolor{red}{X} = x) \neq P(Y = y) \implies X \text{ is } \textbf{correlated} \text{ to } Y$$

Can we ***predict*** better the value of  $Y$  after  
***making an intervention***  $do(X = x)$ ?

$$\exists x \text{ s.t. } P_{\mathcal{M}_x}(Y = y) \neq P(Y = y) \implies X \text{ is } \textbf{a cause} \text{ of } Y$$

# Structural Equation Model (SEM)

---

$$\mathcal{M} = \left\{ \begin{array}{l} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{\epsilon_X, \epsilon_Y, \epsilon_Z\} \\ \mathcal{F} = \left\{ \begin{array}{l} Z = \beta_{Z0} + \epsilon_Z \\ X = \beta_{X0} + \beta_{XZ}Z + \epsilon_X \\ Y = \beta_{Y0} + \beta_{YZ}Z + \beta_{YX}X + \epsilon_Y \end{array} \right. \\ \mathbf{U} \sim \mathcal{N}\left(\mathbf{0}, \Sigma = \begin{bmatrix} \sigma_X & 0 & 0 \\ 0 & \sigma_Y & 0 \\ 0 & 0 & \sigma_Z \end{bmatrix}\right) \end{array} \right.$$

- **Linear functions**
- **Normal distribution**
- **Markovianity / Causal Sufficiency:**  
Error terms in  $\mathbf{U}$  are independent of each other (diagonal covariance matrix).

**Full specification of an SCM requires parametric and distributional assumptions.**

**Estimation of such models usually requires strong assumptions (e.g., Markovianity).**

# SCM: Encoder of Functional Knowledge

---

The knowledge required to fully specify an SCM is usually *unavailable* in practice.

Is it possible to identify the effect of interventions from *observational* data without fully specifying the SCM (i.e., in a non-parametric fashion)?



Yes, with structural knowledge encoded as a causal diagram!

# **Encoding Structural Causal Knowledge**

**Acyclic Directed Acyclic Graph (ADMG)**  
**Causal Diagrams**

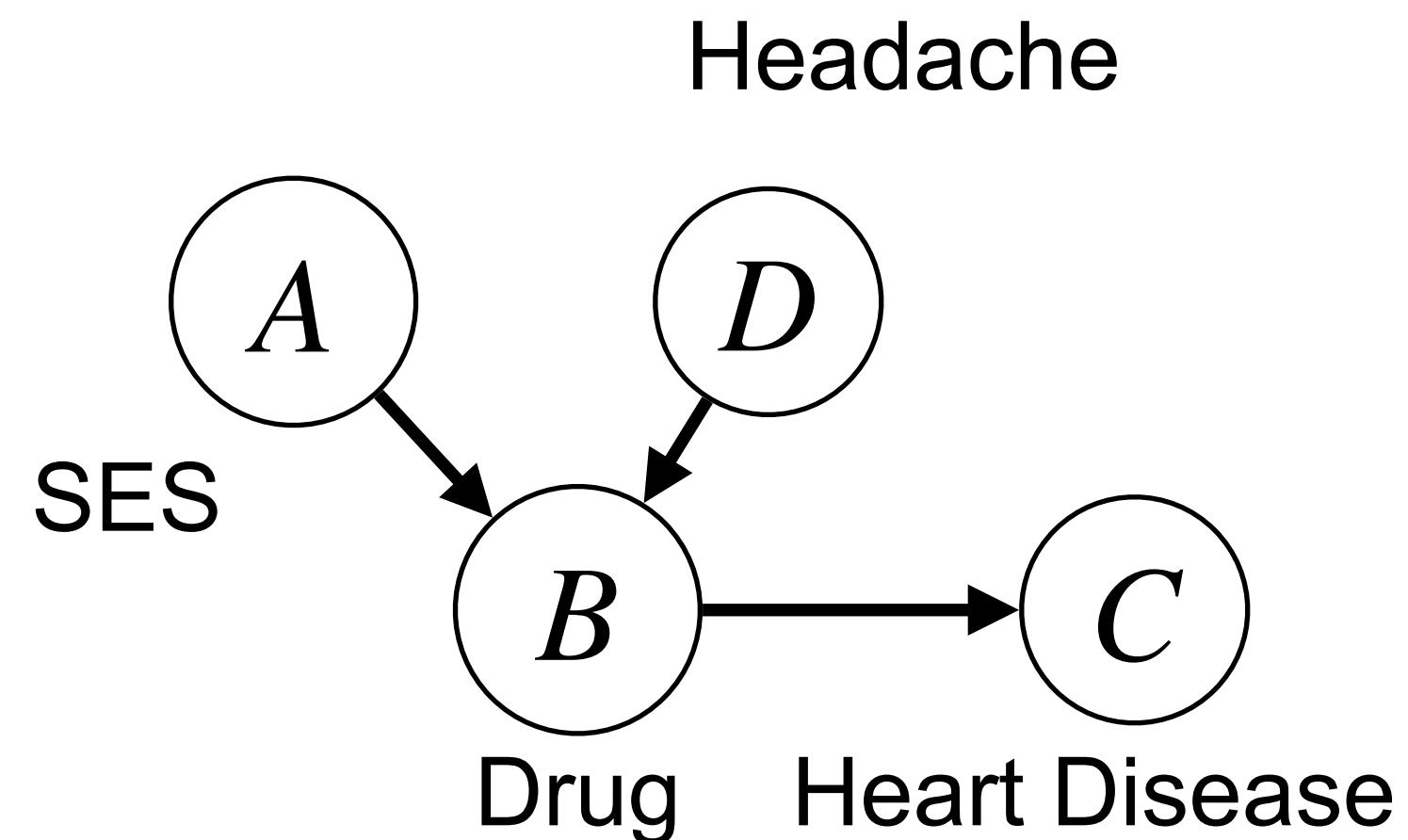
# Causal Diagram: Encoder of Structural Knowledge

Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{A, B, C, D\} \\ \mathbf{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \begin{cases} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram



An SCM  $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$  induces a causal diagram such that, **for every**  $V_i, V_j \in \mathbf{V}$ :

$V_i \rightarrow V_j$ , if  $V_i$  appears as argument of  $f_j \in \mathcal{F}$ .

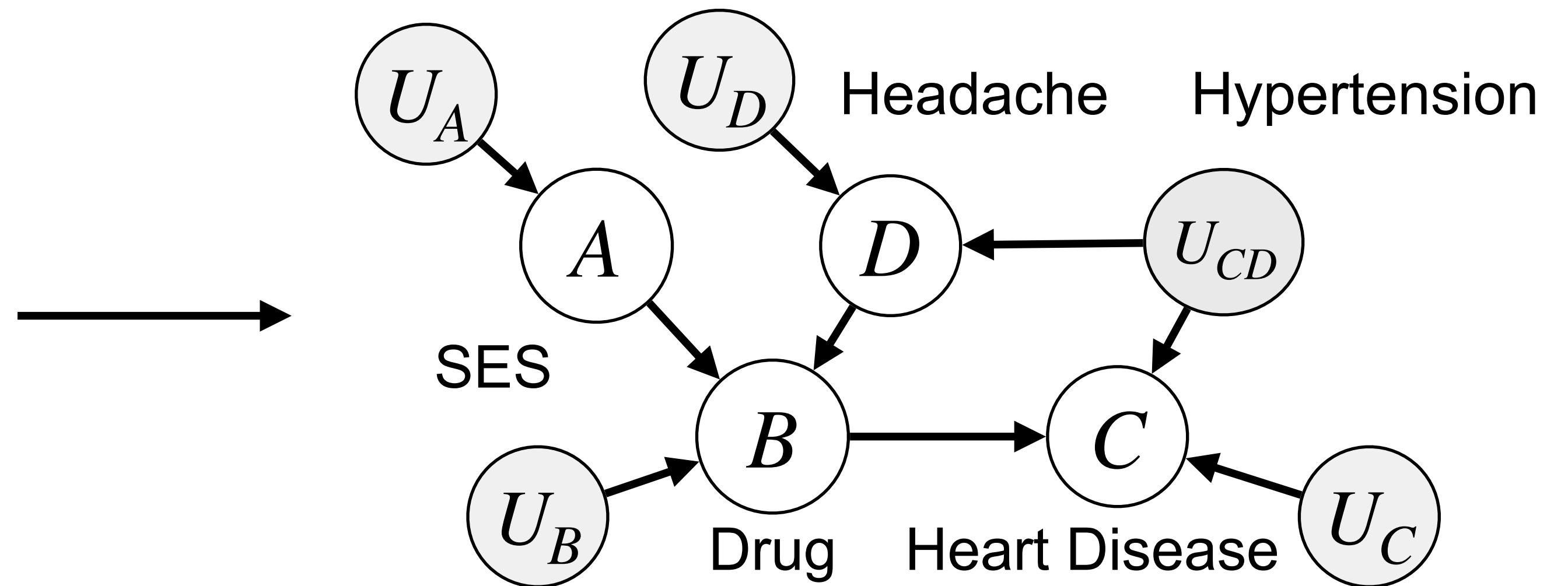
# Causal Diagram: Encoder of Structural Knowledge

Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{A, B, C, D\} \\ \mathbf{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \begin{cases} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram



An SCM  $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$  induces a causal diagram such that, **for every**  $V_i, V_j \in \mathbf{V}$ :

$V_i \rightarrow V_j$ , if  $V_i$  appears as argument of  $f_j \in \mathcal{F}$ .

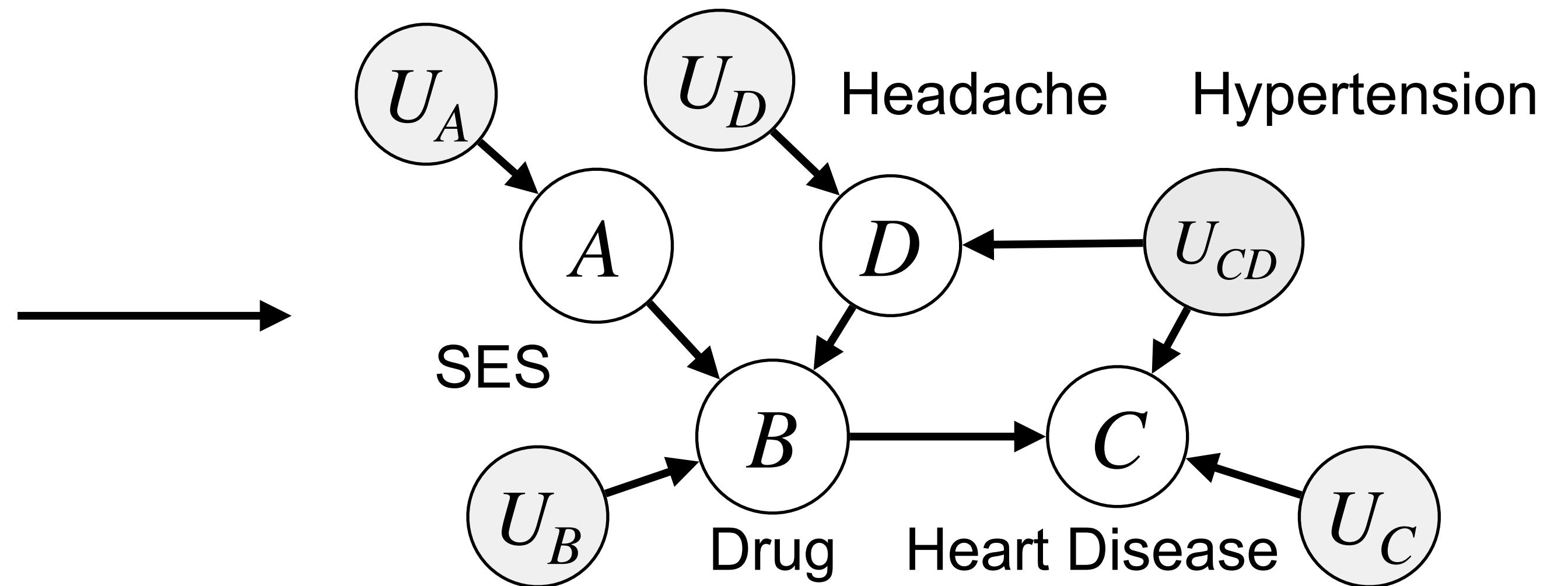
# Causal Diagram: Encoder of Structural Knowledge

Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{A, B, C, D\} \\ \mathbf{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \begin{cases} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram



An SCM  $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$  induces a causal diagram such that, **for every**  $V_i, V_j \in \mathbf{V}$ :

$V_i \rightarrow V_j$ , if  $V_i$  appears as argument of  $f_j \in \mathcal{F}$ .

$V_i \leftrightarrow V_j$  if the corresponding  $U_i, U_j \in \mathbf{U}$  are correlated or  $f_i, f_j$  share some argument  $U \in \mathbf{U}$ .

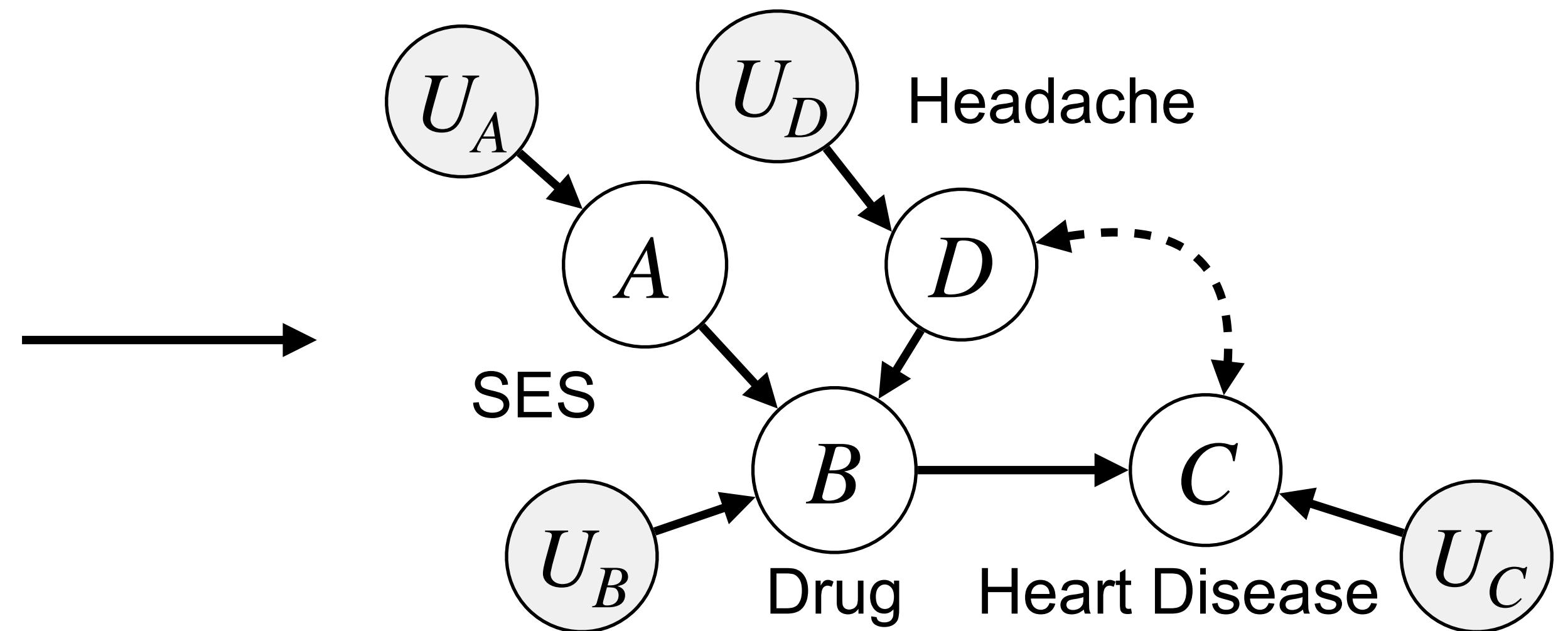
# Causal Diagram: Encoder of Structural Knowledge

Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{A, B, C, D\} \\ \mathbf{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \begin{cases} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

Induced Causal Diagram



An SCM  $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$  induces a causal diagram such that, **for every**  $V_i, V_j \in \mathbf{V}$ :

$V_i \rightarrow V_j$ , if  $V_i$  appears as argument of  $f_j \in \mathcal{F}$ .

$V_i \leftrightarrow V_j$  if the corresponding  $U_i, U_j \in \mathbf{U}$  are correlated or  $f_i, f_j$  share some argument  $U \in \mathbf{U}$ .

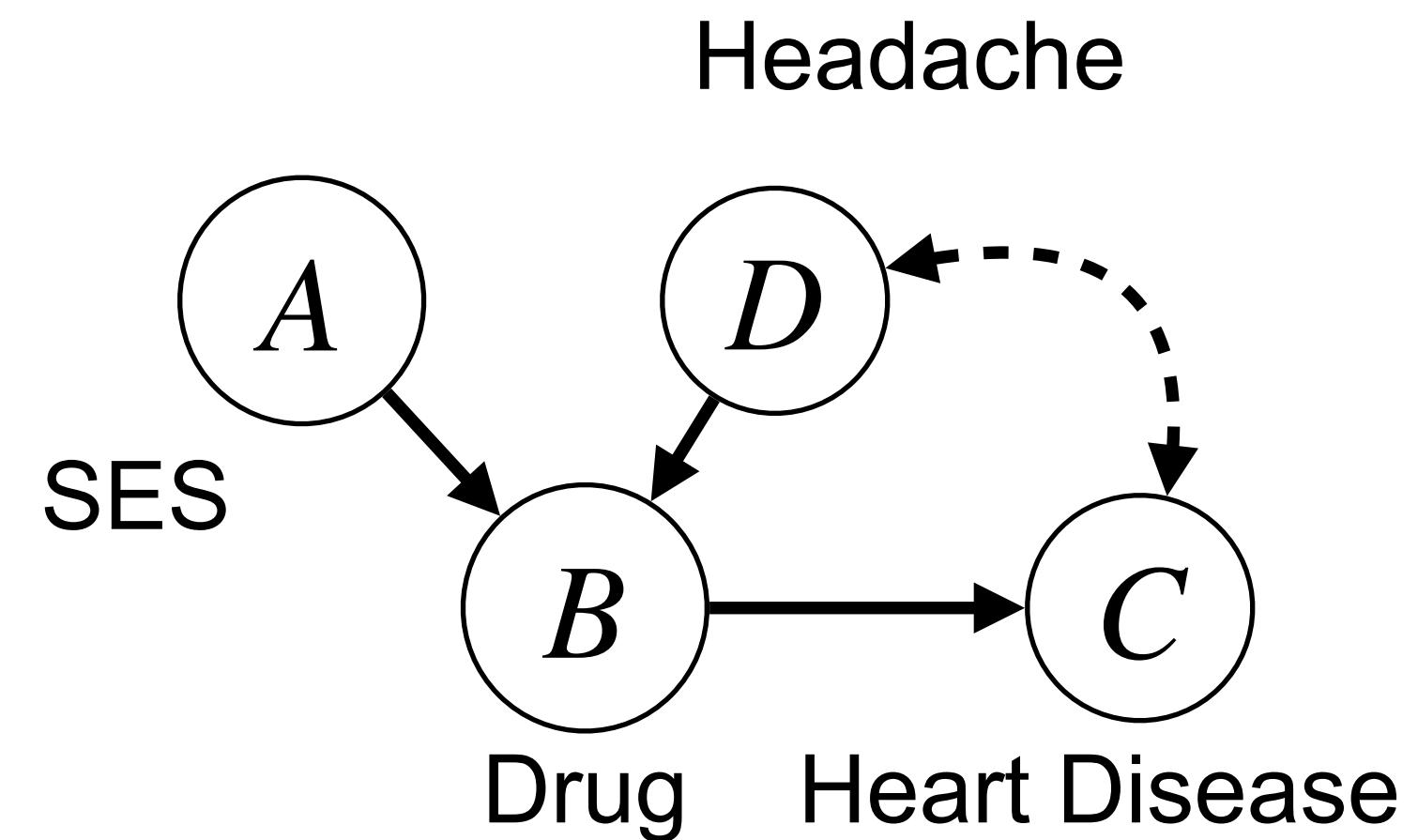
# Causal Diagram: Encoder of Structural Knowledge

Structural Causal Model (SCM)

$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \left\{ \begin{array}{l} \mathbf{V} = \{A, B, C, D\} \\ \mathbf{U} = \{U_A, U_B, U_C, U_D, U_{CD}\} \\ \mathcal{F} = \left\{ \begin{array}{l} A \leftarrow f_A(U_A) \\ B \leftarrow f_B(A, D, U_B) \\ D \leftarrow f_Z(U_D, U_{CD}) \\ C \leftarrow f_X(B, U_C, U_{CD}) \end{array} \right. \\ P(\mathbf{U}) \end{array} \right.$$

Induced Causal Diagram



An SCM  $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$  induces a causal diagram such that, **for every**  $V_i, V_j \in \mathbf{V}$ :

$V_i \rightarrow V_j$ , if  $V_i$  appears as argument of  $f_j \in \mathcal{F}$ .

$V_i \leftrightarrow V_j$  if the corresponding  $U_i, U_j \in \mathbf{U}$  are correlated or  $f_i, f_j$  share some argument  $U \in \mathbf{U}$ .

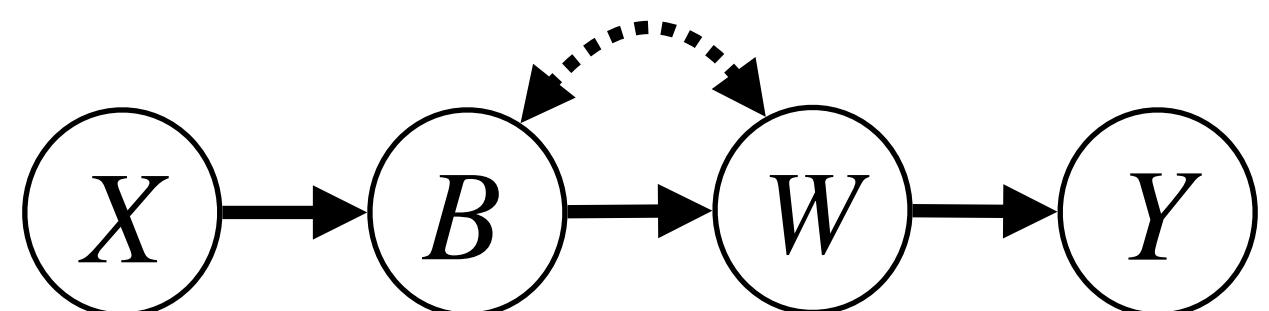
# D-Separation and Implied Conditional Independencies

**Definition (inactive):** A triplet  $\langle V_i, V_m, V_j \rangle$  is said to be **inactive** relative to a set  $Z$  if the middle node  $V_m$ :

1. Is a non-collider and is in  $Z$ ; or
2. Is a collider and neither it nor any of its descendants in  $Z$ .

**Definition (d-separation):** A path  $p$  in a causal diagram  $G$  is said to be **d-separated** (or blocked) by a set of variables  $Z$  if and only if  $p$  contains an inactive triplet in it.

A set  $Z$  d-separates  $X$  and  $Y$  if and only if  $Z$  blocks every path between a node in  $X$  and a node in  $Y$ .



Does  $Z$  d-separates  $X$  and  $Y$  ?

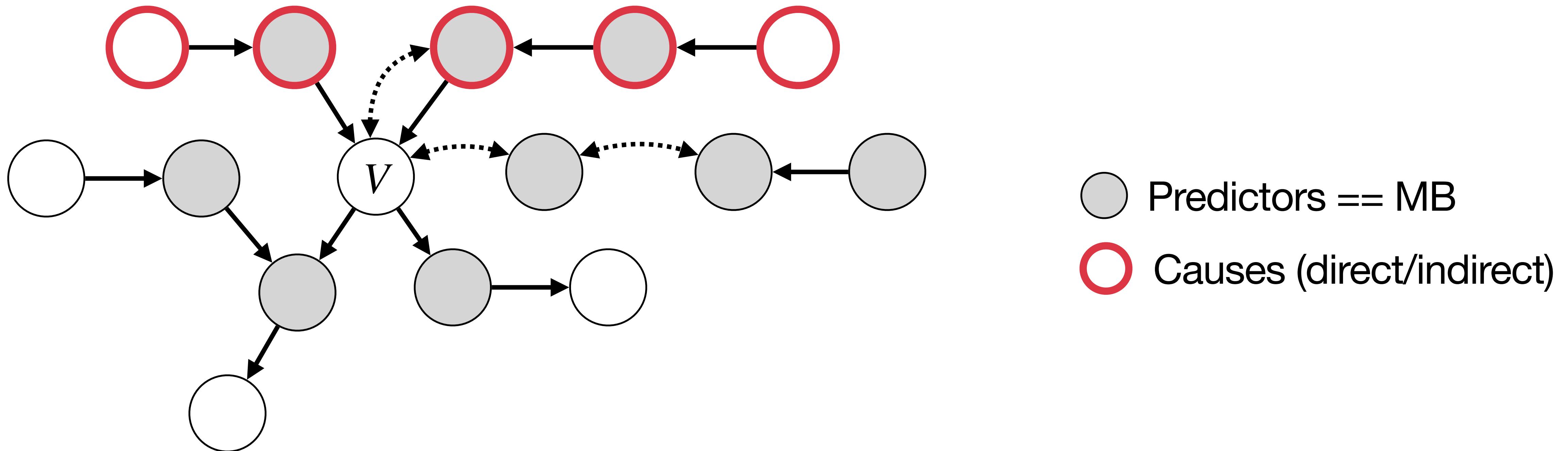
$Z:$    $\{\}$    $\{B\}$    $\{W\}$    $\{B, W\}$

We have that  $(X \perp\!\!\!\perp Y)_G$ ,  $(X \perp\!\!\!\perp Y|B)_G$ , and  $(X \perp\!\!\!\perp Y|W)_G$ , but  $(X \perp\!\!\!\perp Y|B, W)_G$

**Global Markov property:**  $(X \perp\!\!\!\perp Y|Z)_G \Rightarrow (X \perp\!\!\!\perp Y|Z)_P$

D-separations in  $G$  imply conditional independencies in  $P$

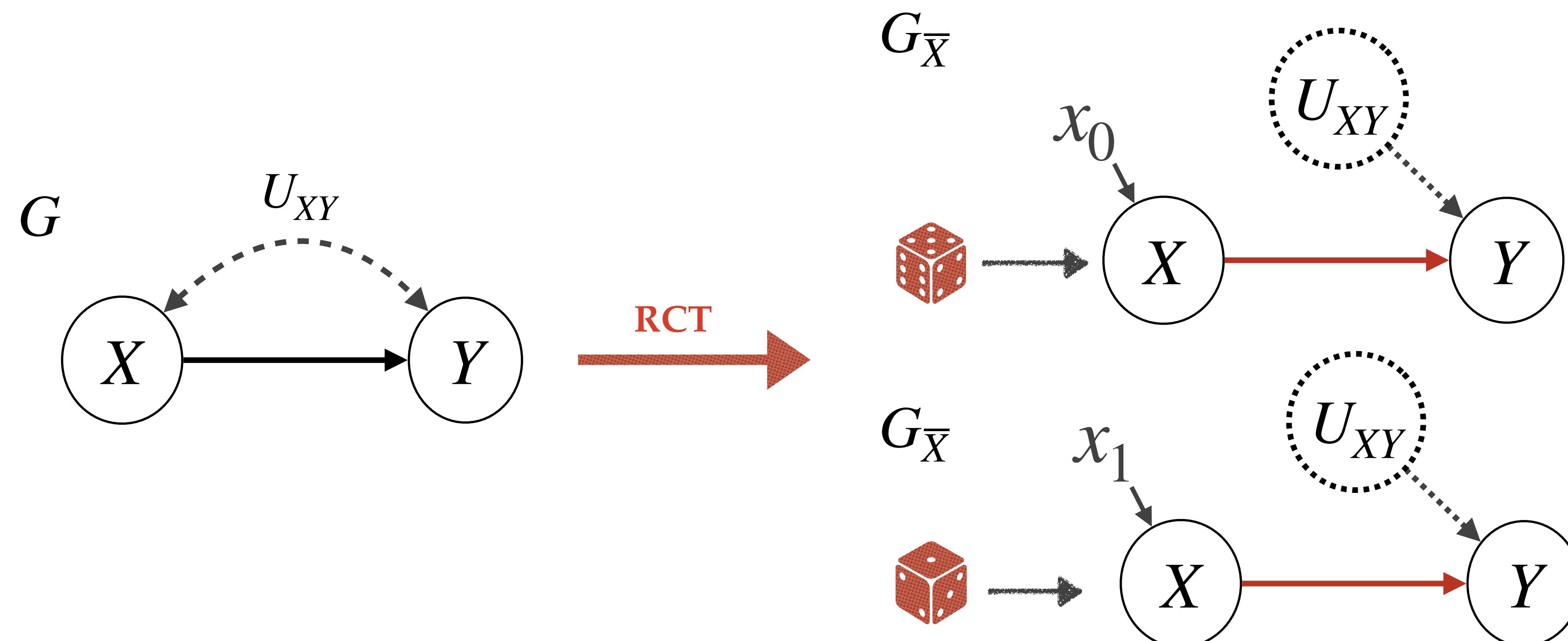
# Graphically Explaining Causes and Predictors



**Markov Blanket (MB) of  $V$ :** the bidirected connected component (district) of  $V$  (excluding  $V$  itself) and the parents of the district of  $V$ , i.e.:  
$$\text{mb}_G(V) = \text{dis}_G(V) \cup \text{Pa}_G(\text{dis}_G(V)) \setminus \{V\}$$

# Randomized Experiments

Randomized Experiments / Control Trials (e.g. RCT) allow the identification of causal effects by leveraging randomization of the treatment assignment.



# Pearl's Inferential Hierarchy

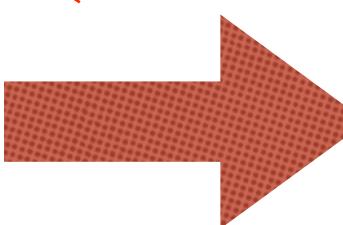
Associational vs Interventional vs Counterfactual

# What is induced by the SCM?

## Observational SCM

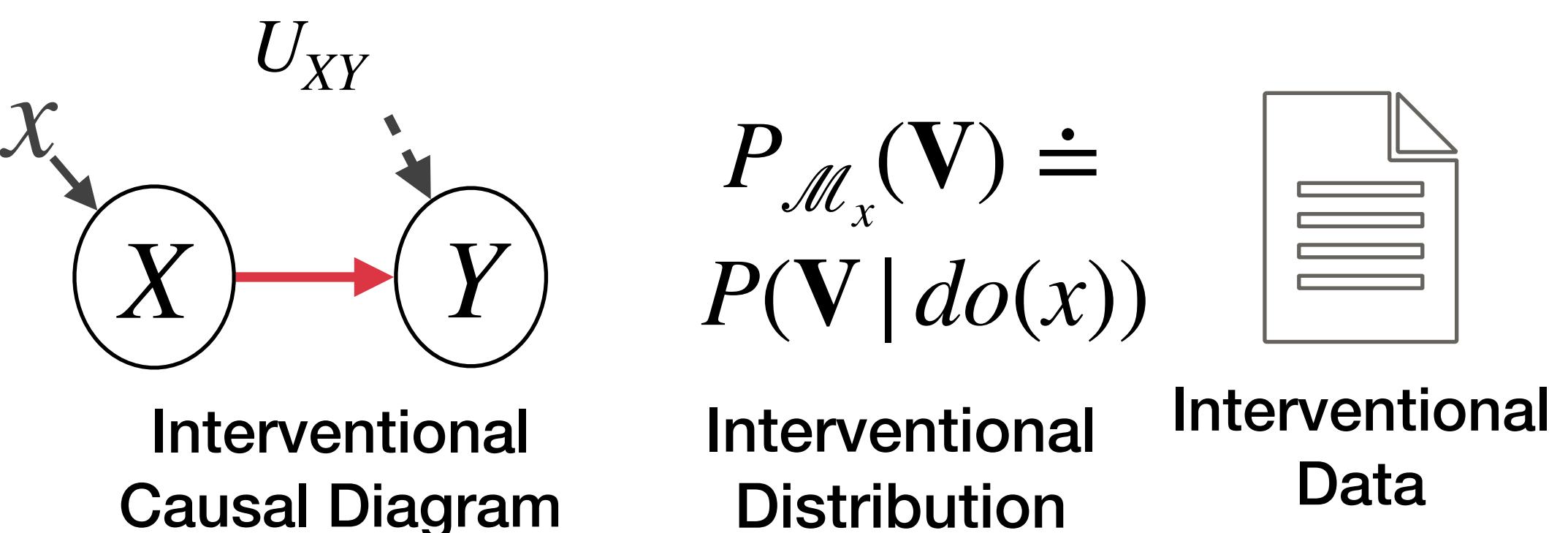
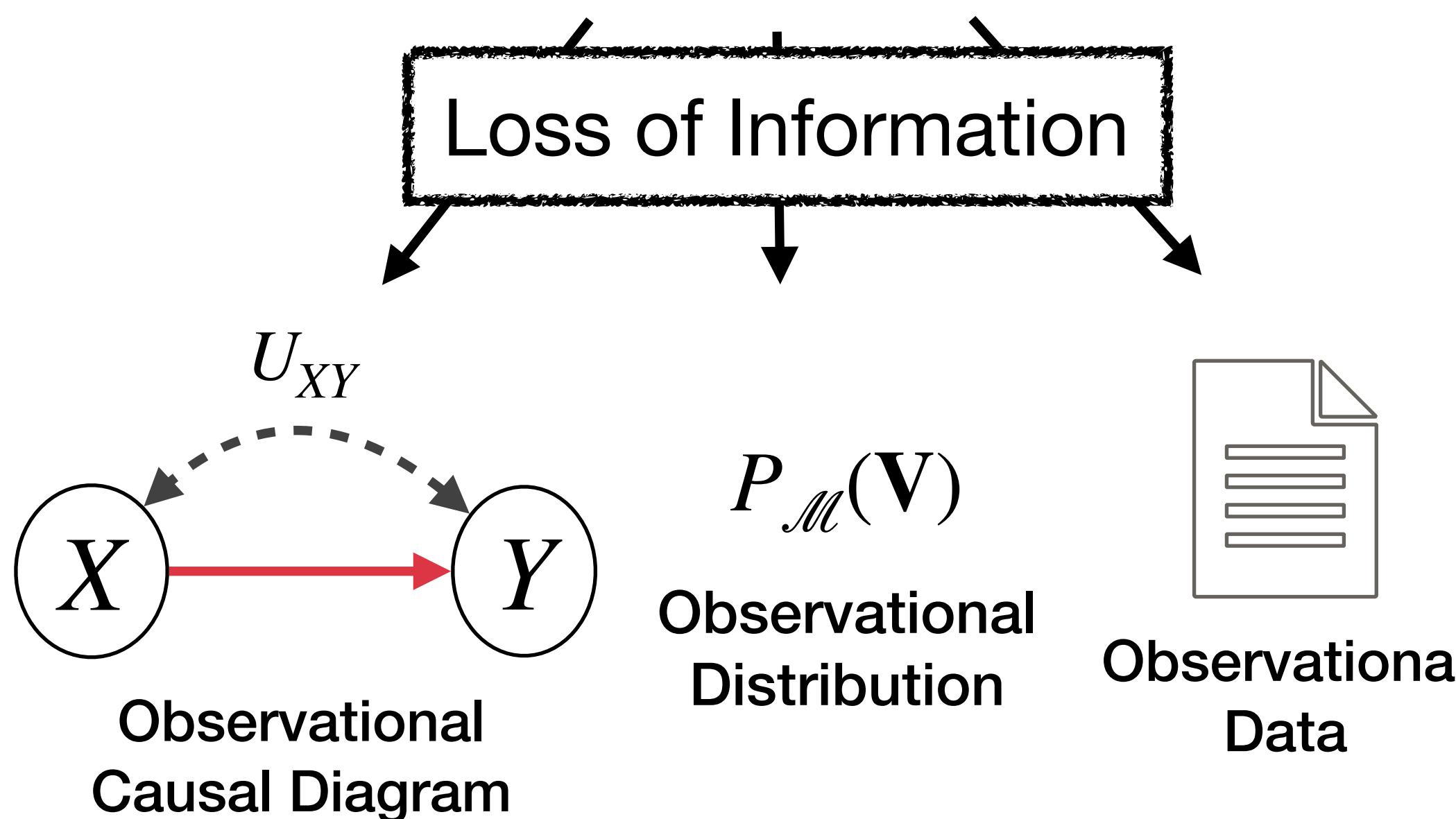
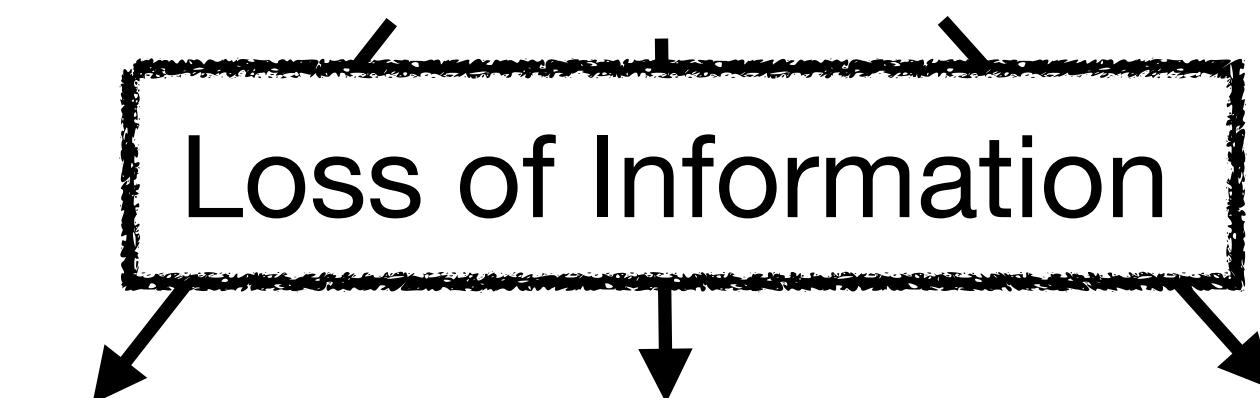
$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = f_X(U_X, U_{XY}) \\ Y = f_Y(X, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$do(X = x)$



## Interventional SCM

$$\mathcal{M}_x = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = x \\ Y = f_Y(x, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



Observational

Structural Causal Model (SCM)

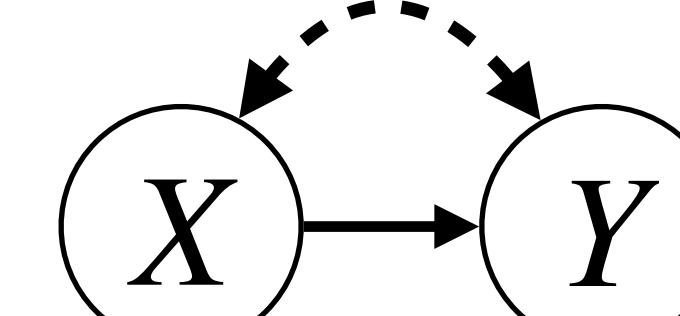
$$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$$

$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \left\{ \begin{array}{l} X \leftarrow f_X(U_X, U_{XY}) \\ Y \leftarrow f_Y(X, U_Y, U_{XY}) \end{array} \right. \\ P(\mathbf{U}) \end{cases}$$

Structural Knowledge

Causal Diagram

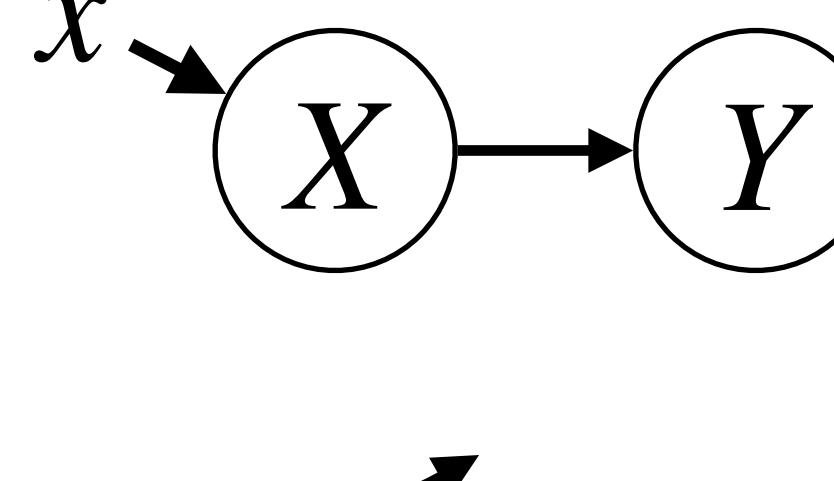
$G$



Interventional

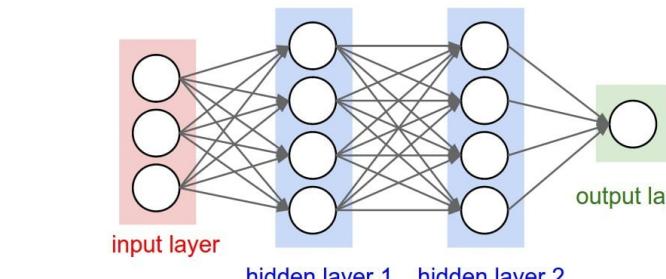
$$\mathcal{M}_x = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \left\{ \begin{array}{l} X \leftarrow x \\ Y \leftarrow f_Y(x, U_Y, U_{XY}) \end{array} \right. \\ P(\mathbf{U}) \end{cases}$$

$G_{\bar{X}}$



Data

X	Z	Y
---	---	---



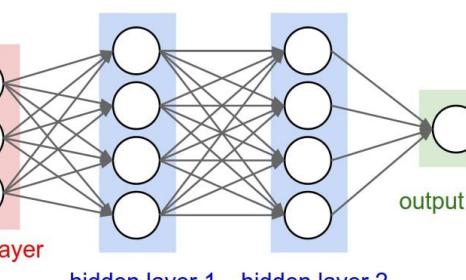
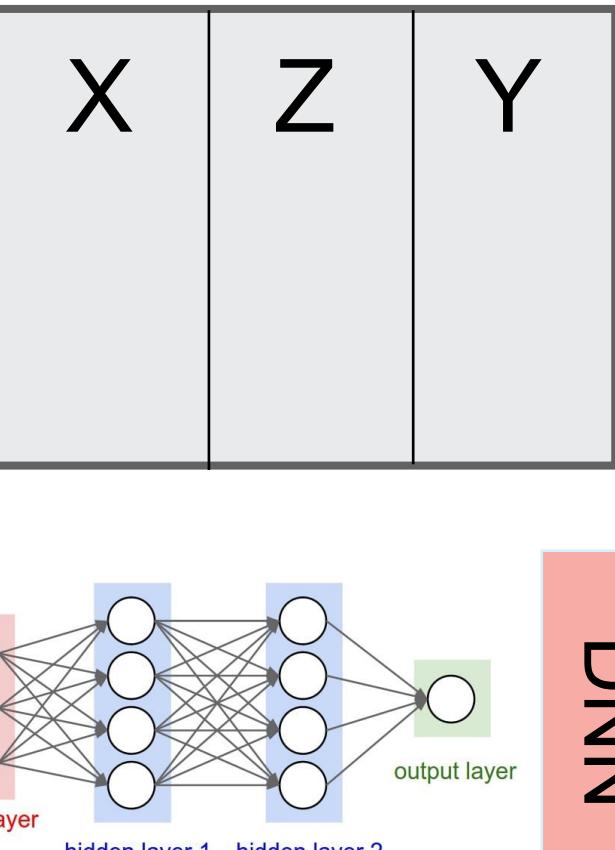
$$\hat{P}(Y|X=x)$$

Seeing

Doing



Observational

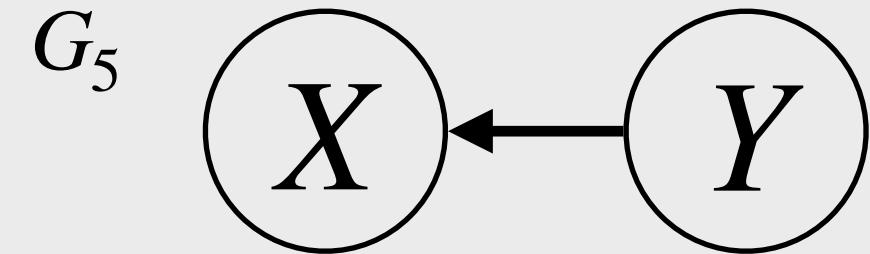
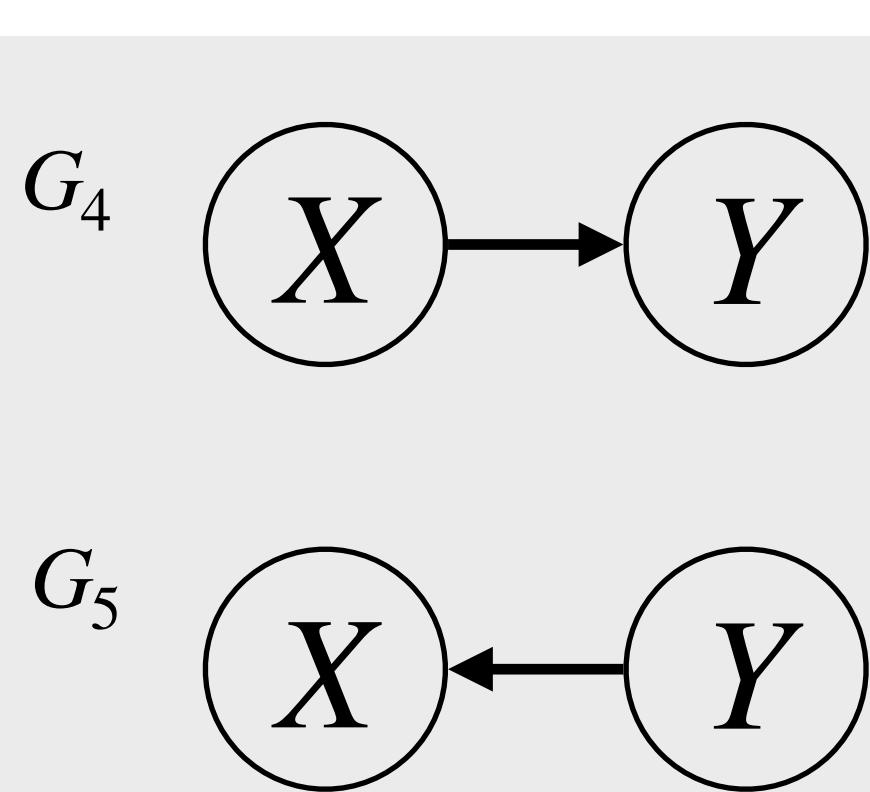
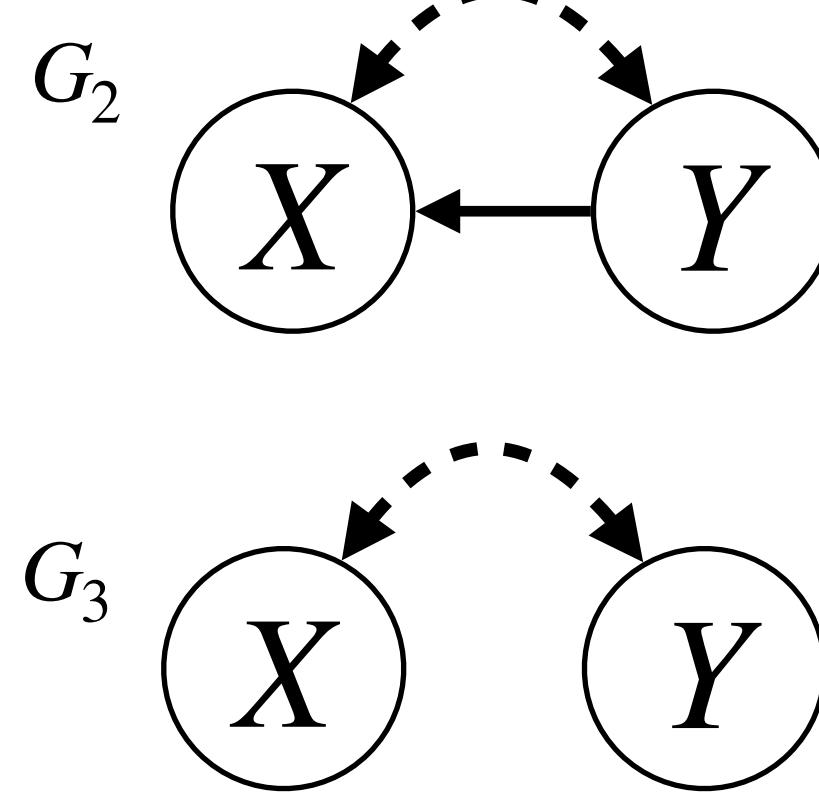
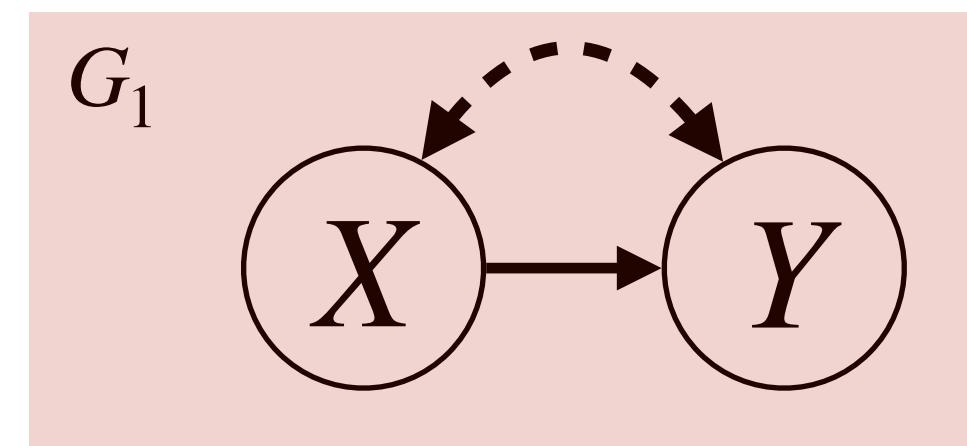


$$P(Y|X=x)$$

Data

Potential Causal Diagrams

Potential SCMs



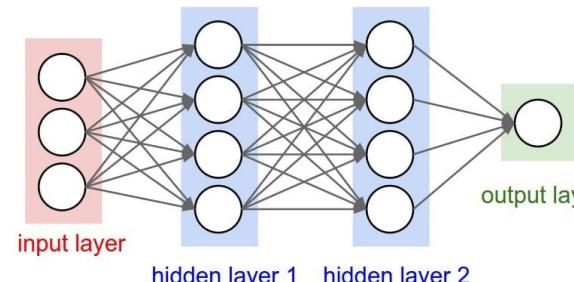
Encoded Knowledge / Assumptions

- $\mathcal{M}_{11} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{11}, P_{11}(\mathbf{u}_1) \rangle$
- $\vdots$
- $\mathcal{M}_{1k_1} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{1k_1}, P_{1k_1}(\mathbf{u}_1) \rangle$
- True Model
- $\mathcal{M}_{21} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{21}, P_{21}(\mathbf{u}_2) \rangle$
- $\vdots$
- $\mathcal{M}_{2k_2} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{2k_2}, P_{2k_2}(\mathbf{u}_2) \rangle$
- $\mathcal{M}_{31} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{31}, P_{31}(\mathbf{u}_3) \rangle$
- $\vdots$
- $\mathcal{M}_{3k_3} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{3k_3}, P_{3k_3}(\mathbf{u}_3) \rangle$
- $\mathcal{M}_{41} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{41}, P_{41}(\mathbf{u}_4) \rangle$
- $\vdots$
- $\mathcal{M}_{4k_4} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{4k_4}, P_{4k_4}(\mathbf{u}_4) \rangle$
- $\mathcal{M}_{51} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{51}, P_{51}(\mathbf{u}_5) \rangle$
- $\vdots$
- $\mathcal{M}_{5k_5} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{5k_5}, P_{5k_5}(\mathbf{u}_5) \rangle$

Parametrization

Observational

X	Z	Y
---	---	---



DNN

$$P(Y|X = r)$$

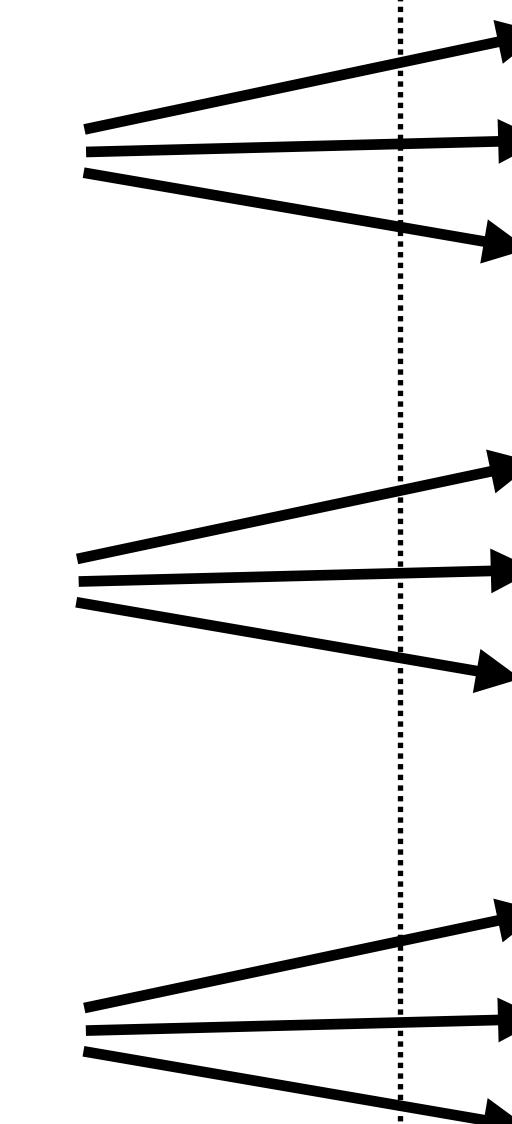
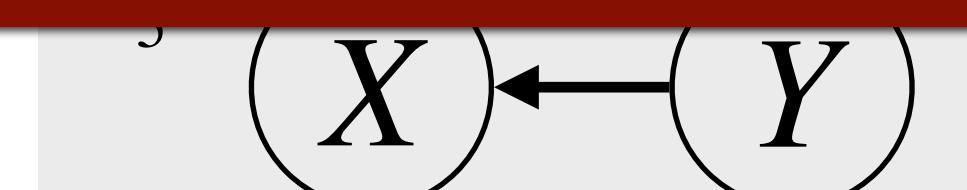
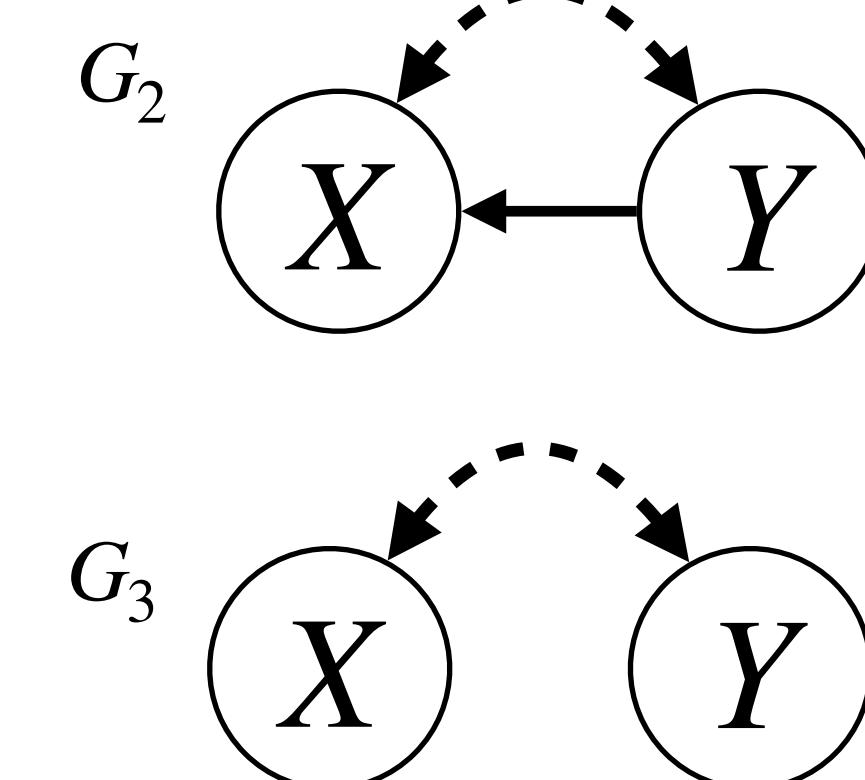
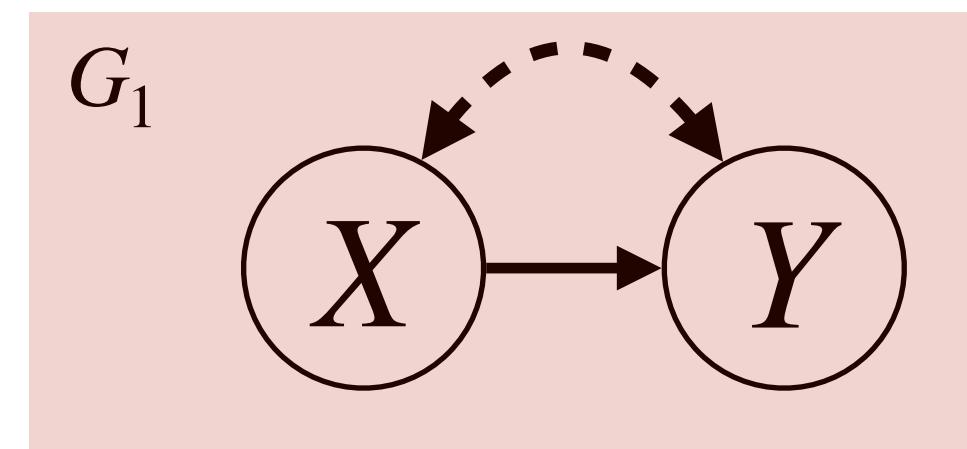
Multiple neural nets fit the data equally well,  
leading to different causal explanations!

Encoded Knowledge / Assumptions

Data

Potential Causal Diagrams

Potential SCMs



$$\mathcal{M}_{11} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{11}, P_{11}(\mathbf{u}_1) \rangle$$

⋮

$$\mathcal{M}_{1k_1} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{1k_1}, P_{1k_1}(\mathbf{u}_1) \rangle$$

True Model

$$\mathcal{M}_{21} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{21}, P_{21}(\mathbf{u}_2) \rangle$$

⋮

$$\mathcal{M}_{2k_2} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{2k_2}, P_{2k_2}(\mathbf{u}_2) \rangle$$

$$\mathcal{M}_{31} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{31}, P_{31}(\mathbf{u}_3) \rangle$$

⋮

$$\mathcal{M}_{3k_3} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{3k_3}, P_{3k_3}(\mathbf{u}_3) \rangle$$

$$\mathcal{M}_{41} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{41}, P_{41}(\mathbf{u}_4) \rangle$$

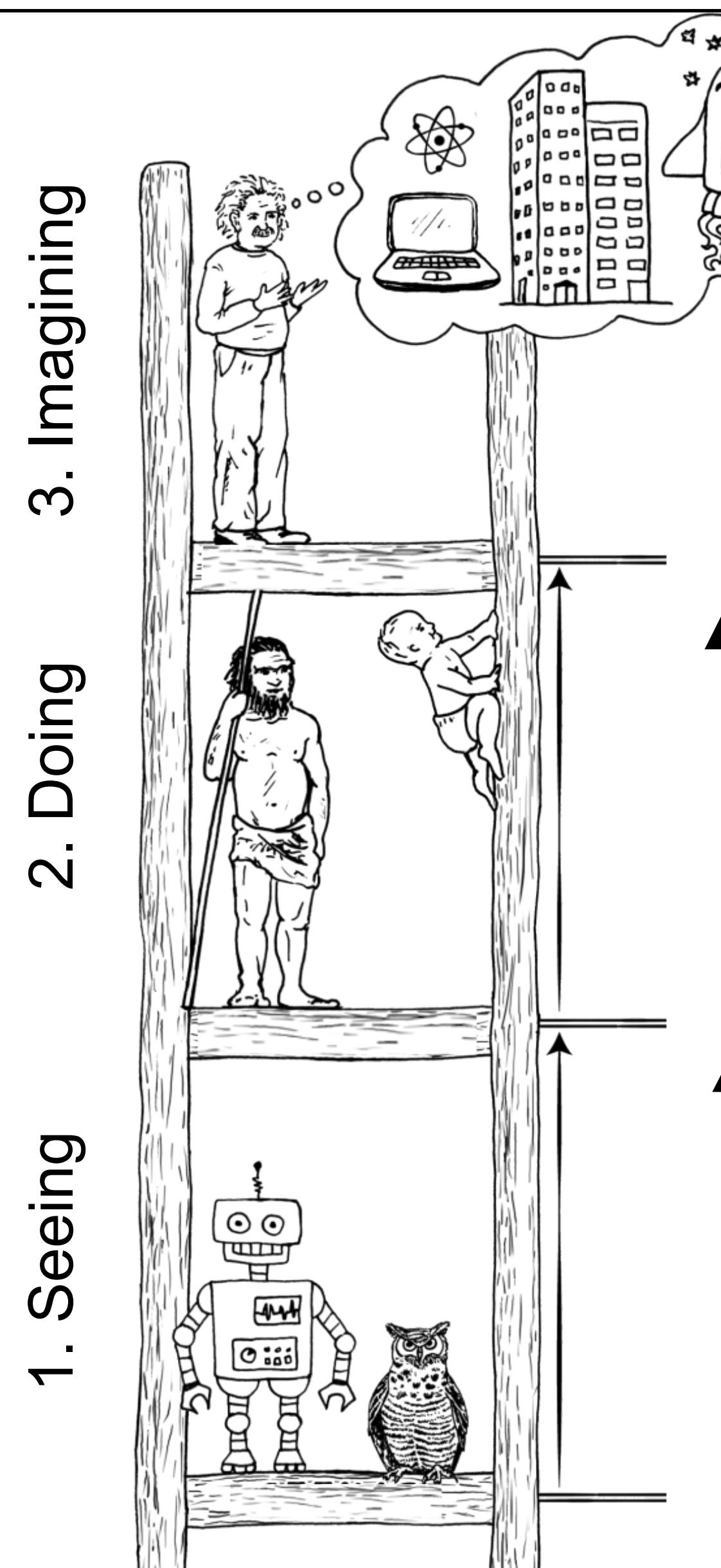
$$\mathcal{M}_{4k_4} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{4k_4}, P_{4k_4}(\mathbf{u}_4) \rangle$$

$$\mathcal{M}_{51} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{51}, P_{51}(\mathbf{u}_5) \rangle$$

$$\mathcal{M}_{5k_5} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{5k_5}, P_{5k_5}(\mathbf{u}_5) \rangle$$

Parametrization

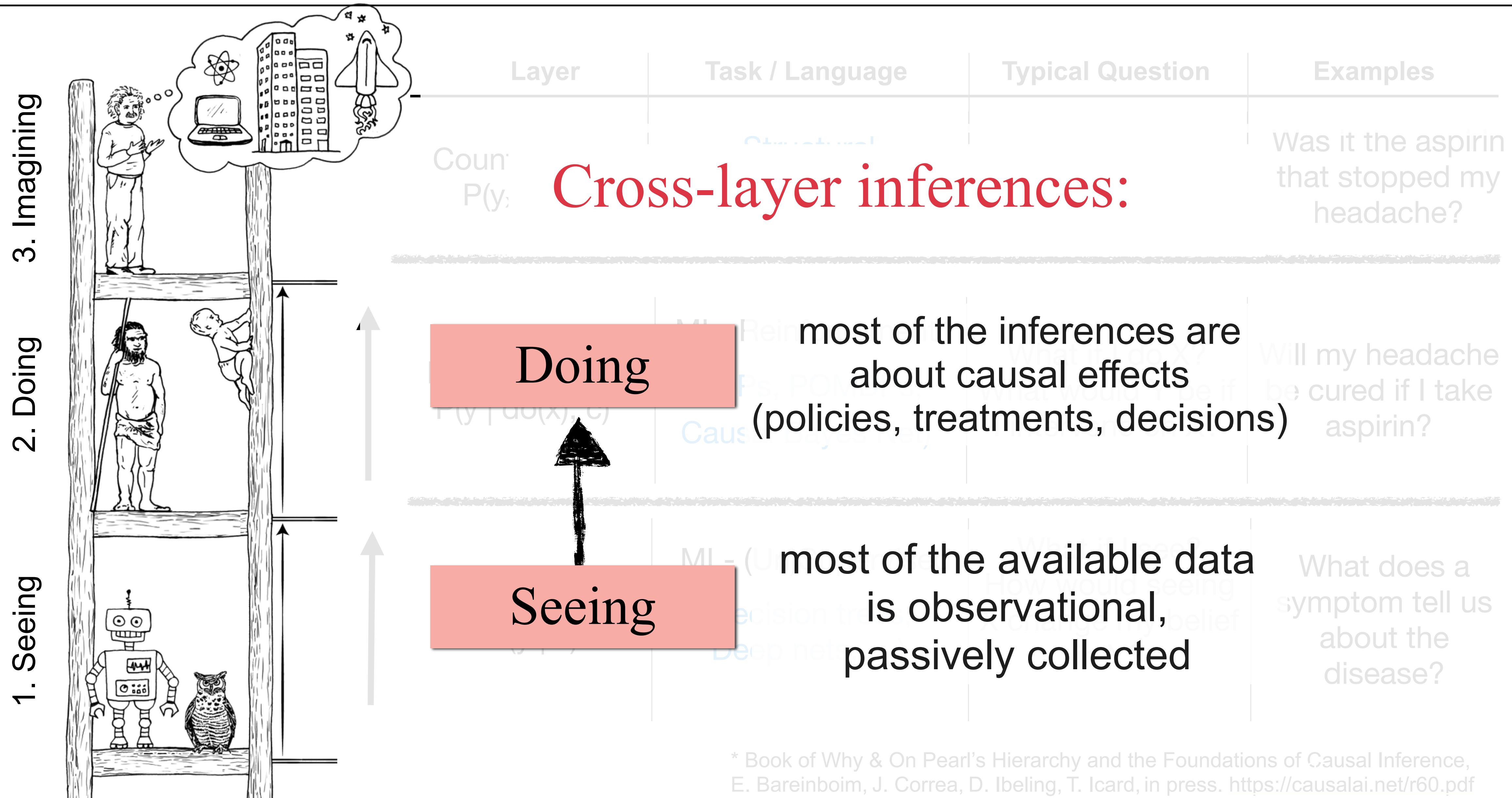
# Ladder of Causation



Layer	Task / Language	Typical Question	Examples
Counterfactual $P(y_x   x', y')$	Structural Causal Model	What if I had acted differently?	Was it the aspirin that stopped my headache?
Interventional $P(y   \text{do}(x), c)$	ML- Reinforcement (Causal Bayes Net)	What if I do X? What would Y be if I intervene on X?	Will my headache be cured if I take aspirin?
Associational $P(y   x)$	ML- (Un)Supervised (Decision trees, Deep nets, ...)	What if I see? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?

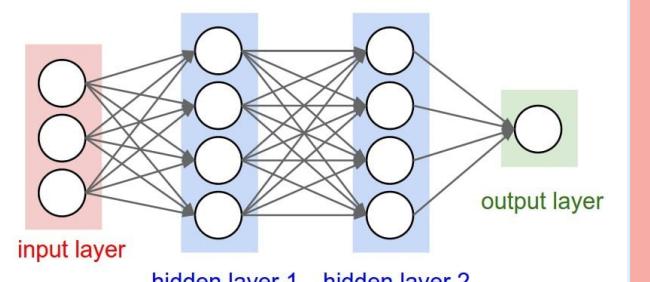
\* Book of Why & On Pearl's Hierarchy and the Foundations of Causal Inference,  
E. Bareinboim, J. Correa, D. Ibeling, T. Icard, in press. <https://causalai.net/r60.pdf>

# Ladder of Causation



## Observational

X	Z	Y
---	---	---

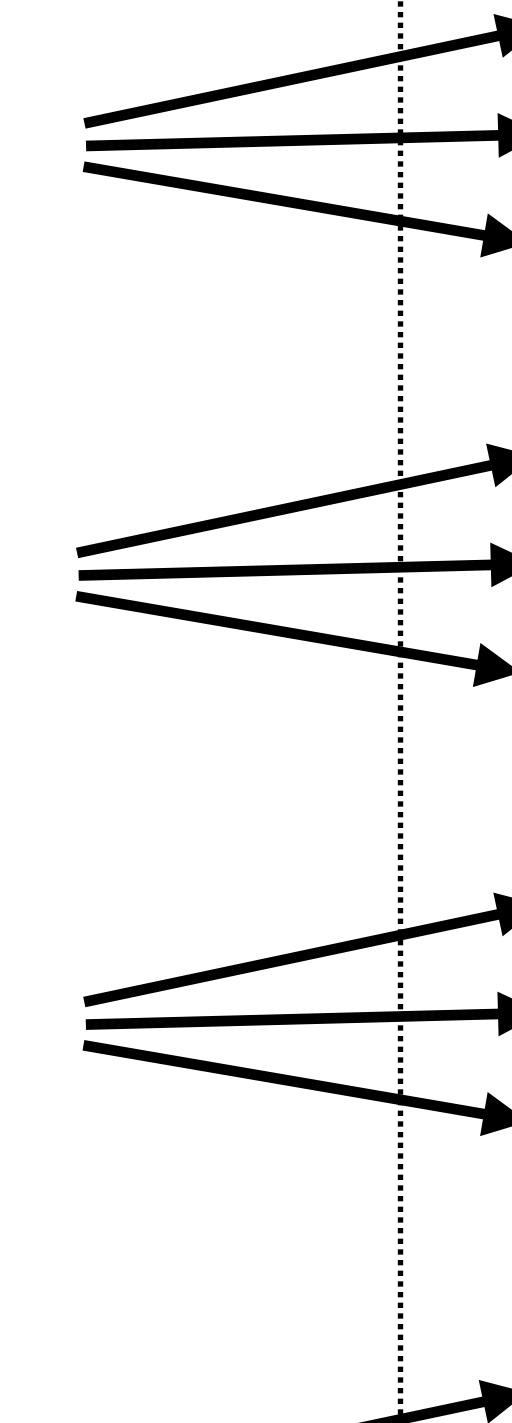
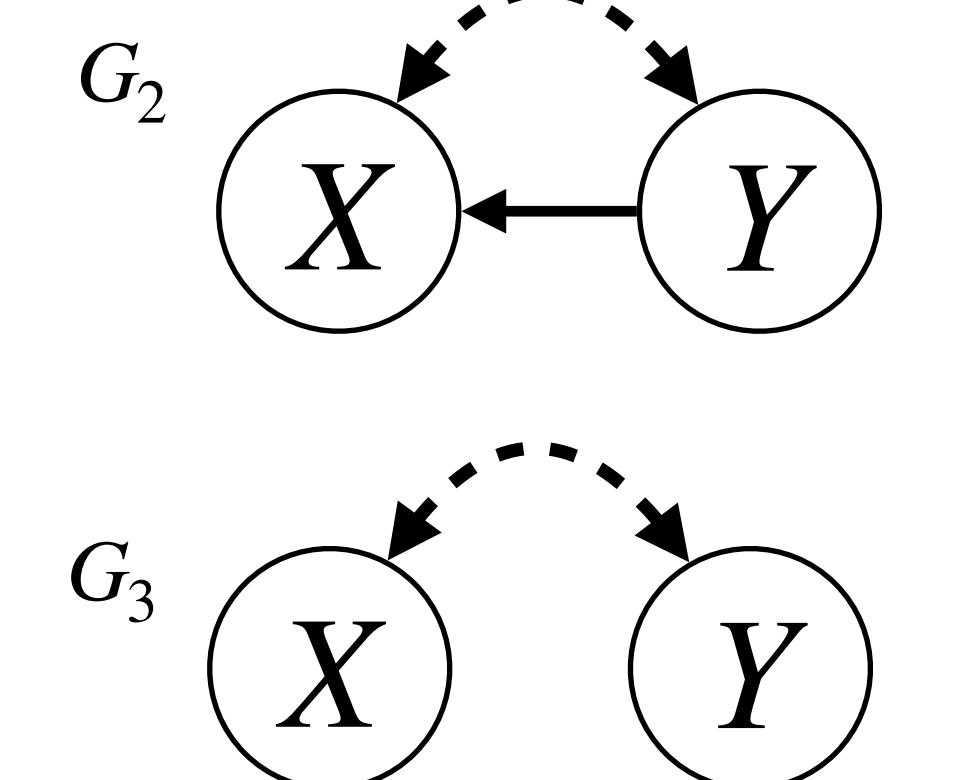
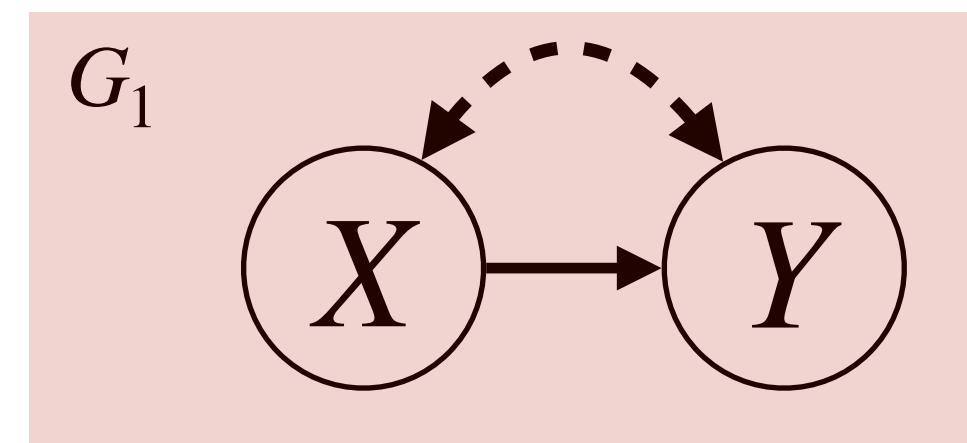


$$P(Y|X=x)$$

Data

Potential Causal Diagrams

Potential SCMs



$$\mathcal{M}_{11} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{11}, P_{11}(\mathbf{u}_1) \rangle$$

⋮

$$\mathcal{M}_{1k_1} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{1k_1}, P_{1k_1}(\mathbf{u}_1) \rangle$$

True Model

$$\mathcal{M}_{21} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{21}, P_{21}(\mathbf{u}_2) \rangle$$

⋮

$$\mathcal{M}_{2k_2} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{2k_2}, P_{2k_2}(\mathbf{u}_2) \rangle$$

$$\mathcal{M}_{31} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{31}, P_{31}(\mathbf{u}_3) \rangle$$

⋮

$$\mathcal{M}_{3k_3} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{3k_3}, P_{3k_3}(\mathbf{u}_3) \rangle$$

$$\mathcal{M}_{41} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{41}, P_{41}(\mathbf{u}_4) \rangle$$

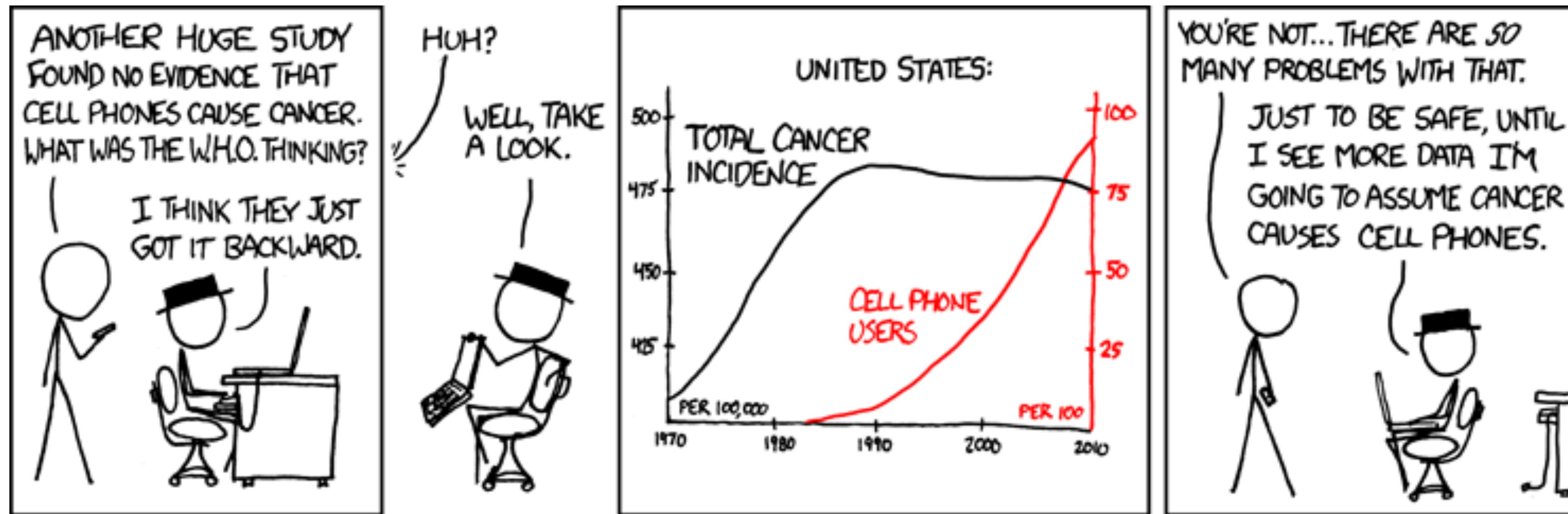
Causal Hierarchy Theorem : to answer questions in layer  $i$ , we need information from layer  $i$  or higher.

$\mathcal{L}_1$

$\mathcal{L}_2$

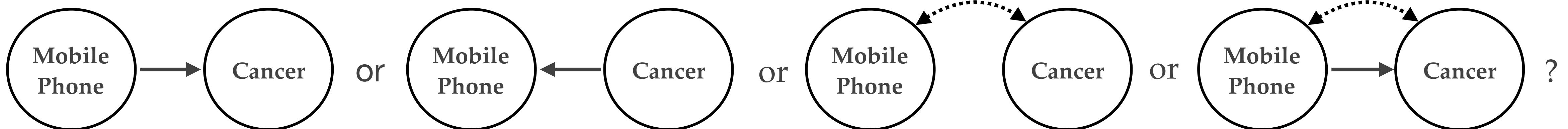
$\mathcal{L}_3$

# Association vs Causation



<https://xkcd.com/925/> - Creative Commons Attribution-NonCommercial 2.5 License.

Will we be able to decide the true relationship just by “seeing” more data?



# Code Examples

---

**Prob-AI GitHub (Day 5):** [https://github.com/probabilisticai/probai-2023/tree/main/day\\_5](https://github.com/probabilisticai/probai-2023/tree/main/day_5)

- **Example 1:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_BD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_BD.ipynb)
- **Example 2:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_NonBD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_NonBD.ipynb)

**Check Items 1 and 2 of both Examples:**

1. Constructing the Structural Causal Model and corresponding Causal Diagram
2. Checking the conditional independence relations implied by the true Causal Diagram

# Questions?



# **Causal Effect Identification**

Graphical Criteria, Do-Calculus, and ID-Algorithm

# Causal Effect

---

The **causal effect** of a (set of) treatment variable(s)  $\mathbf{X}$  on a (set of) outcome variable(s)  $\mathbf{Y}$  is a quantity derived from  $P(\mathbf{Y} | do(\mathbf{X}))$  that tells us how much  $\mathbf{Y}$  changes due to an intervention  $do(\mathbf{X} = \mathbf{x})$ .

## Examples:

- *Average Treatment Effect (ATE)* for discrete treatments:

$$\mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x}')] - \mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x})],$$

where  $\mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x})] = \sum_{\mathbf{y} \in \Omega_{\mathbf{Y}}} \mathbf{y} P(\mathbf{y} | do(\mathbf{x}))$

defined for two treatment levels  $\mathbf{x}'$  and  $\mathbf{x}$  of  $\mathbf{X}$ .

- *Average Treatment Effect (ATE)* for continuous treatments,

$$\frac{\partial \mathbb{E}[Y_i | do(X_j = x_j)]}{\partial x_j}, \text{ for all } Y_i \in \mathbf{Y}, \text{ and } X_j \in \mathbf{X}.$$

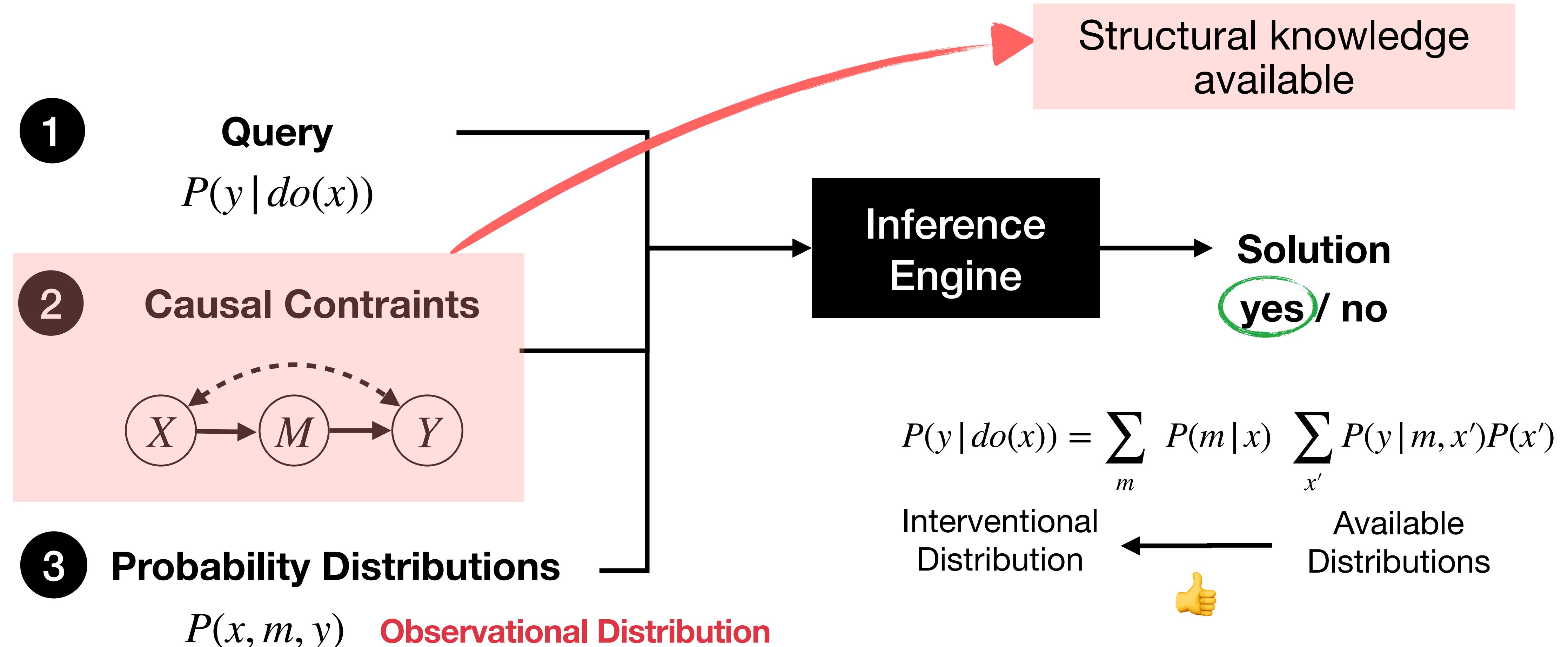
Jacobian of  $\mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x})]$ , where

$$\mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x})] = \int_{\Omega_{\mathbf{Y}}} \mathbf{y} P(\mathbf{y} | do(\mathbf{x})) d\mathbf{y},$$

and  $\Omega_{\mathbf{Y}}$  is the space of all possible values that  $\mathbf{Y}$  might take on

The derivative shows the rate of change of  $\mathbf{Y}$  w.r.t.  $do(\mathbf{X} = \mathbf{x})$

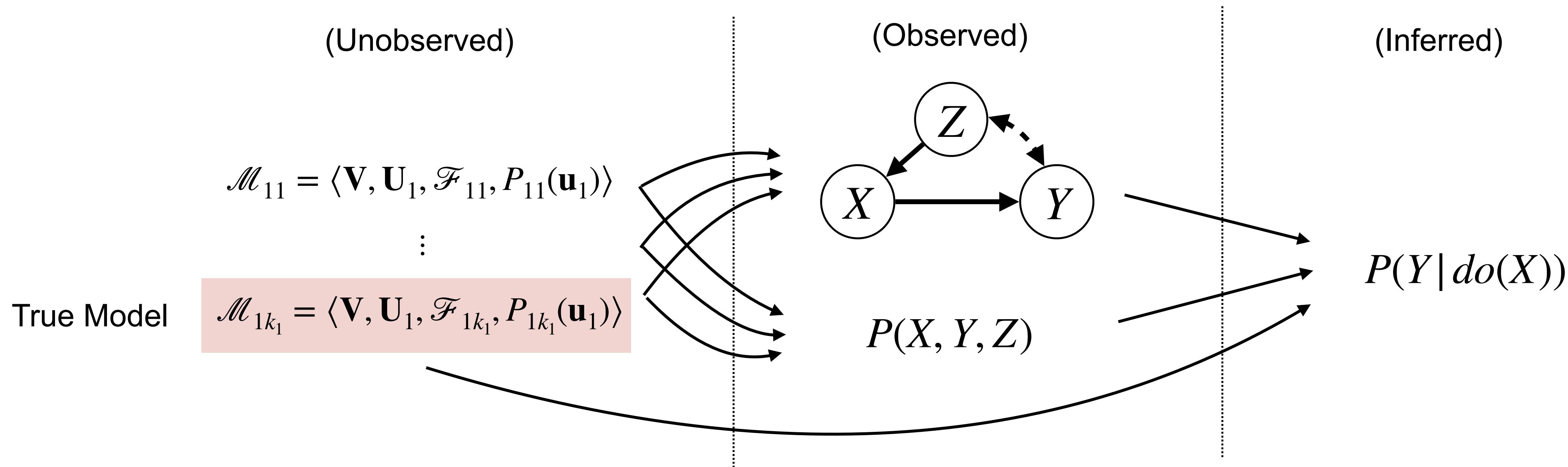
# Classical Causal Effect Identification



- Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pp. 567–573, Menlo Park, CA, 2002. AAAI Press/MIT Press.

# The Effect Identification Problem

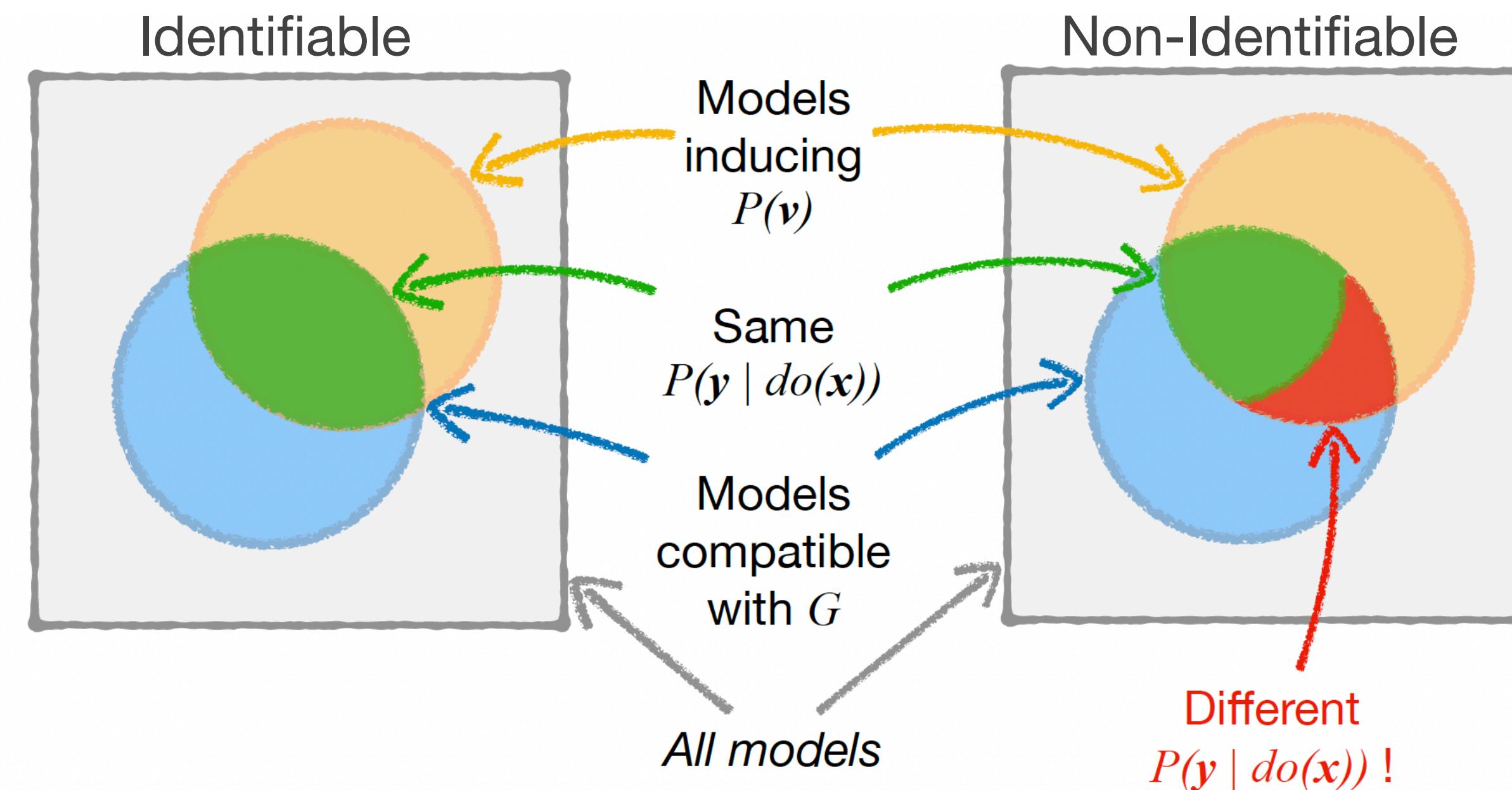
**Causal Effect Identifiability:** The causal effect of a (set of) treatment variable(s)  $\mathbf{X}$  on a (set of) outcome variable(s)  $\mathbf{Y}$  is said to be identifiable from a causal diagram  $G$  and the probability of the observed variables  $P(\mathbf{V})$  if the interventional distribution  $P(\mathbf{Y} | do(\mathbf{X}))$  is *uniquely computable*, i.e., if for every pair of SCMs  $\mathcal{M}_1$  and  $\mathcal{M}_2$  that induce  $G$  and  $P^{\mathcal{M}_1}(\mathbf{V}) = P^{\mathcal{M}_2}(\mathbf{V}) = P(\mathbf{V}) > 0$ ,  $P^{\mathcal{M}_1}(\mathbf{Y} | do(\mathbf{X})) = P^{\mathcal{M}_2}(\mathbf{Y} | do(\mathbf{X})) = P(\mathbf{Y} | do(\mathbf{X}))$ .



In words, causal effect identifiability means that, no matter the form of true SCM, for all models  $\mathcal{M}$  agreeing with  $\langle G, P(\mathbf{V}) \rangle$ , they also agree in  $P(\mathbf{y} | do(\mathbf{x}))$ .

# The Effect Identification Problem

**Causal Effect Identifiability:** The causal effect of a (set of) treatment variable(s)  $\mathbf{X}$  on a (set of) outcome variable(s)  $\mathbf{Y}$  is said to be identifiable from a causal diagram  $G$  and the probability of the observed variables  $P(\mathbf{V})$  if the interventional distribution  $P(\mathbf{Y} | do(\mathbf{X}))$  is *uniquely computable*, i.e., if for every pair of SCMs  $\mathcal{M}_1$  and  $\mathcal{M}_2$  that induce  $G$  and  $P^{\mathcal{M}_1}(\mathbf{V}) = P^{\mathcal{M}_2}(\mathbf{V}) = P(\mathbf{V}) > 0$ ,  $P^{\mathcal{M}_1}(\mathbf{Y} | do(\mathbf{X})) = P^{\mathcal{M}_2}(\mathbf{Y} | do(\mathbf{X})) = P(\mathbf{Y} | do(\mathbf{X}))$ .



In words, causal effect identifiability means that, no matter the form of true SCM, for all models  $\mathcal{M}$  agreeing with  $\langle G, P(\mathbf{V}) \rangle$ , they also agree in  $P(\mathbf{y} | do(\mathbf{x}))$ .

# Tools for Causal Identification

---

1. Markovian Models (No Unmeasured Confounders)
  1. Truncated Factorization / G-computation or G-formula
  2. Adjustment over parents
2. Non-Markovian Models (Under the Presence of Unmeasured Confounders)
  1. Graphical criteria (Backdoor Adjustment, Generalized Adjustment, Front-door Adjustment)
  2. Do-Calculus (a.k.a Causal Calculus)
  3. Identify Algorithm (a.k.a. ID algorithm)

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York. <http://dx.doi.org/10.1017/CBO9780511803161>

Jin Tian. Studies in causal reasoning and learning. PhD thesis, University of California, Los Angeles, 2002.

# Identification in Markovian Models

**Truncated Factorization – Markovian:** Let  $G$  be a causal diagram for the collection  $\mathbf{P}_*$  of all interventional distributions  $P_{\mathbf{x}}(\mathbf{V})$ , for any  $\mathbf{X} \subseteq \mathbf{V}$ . It follows that  $P_{\mathbf{x}}(\mathbf{V})$  factorizes as:

$$P_{\mathbf{x}}(\mathbf{v}) \doteq P(\mathbf{v} \mid do(\mathbf{x})) = \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P_{\mathbf{x}}(v_i \mid pa_i) \Big|_{\mathbf{X}=\mathbf{x}}$$

Follows from  $P_{\mathbf{x}}(\mathbf{v}) \doteq P(\mathbf{v} \mid do(\mathbf{x}))$   
being *Markov* relative to  $G_{\overline{\mathbf{X}}}$

$$= \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P(v_i \mid pa_i) \Big|_{\mathbf{X}=\mathbf{x}}$$

Markovian SCMs have the modularity  
property, i.e.,  $P_{\mathbf{x}}(v_i \mid pa_i) = P(v_i \mid pa_i)$

**Causal Effect of  $\mathbf{X}$  on  $\mathbf{Y}$ :**

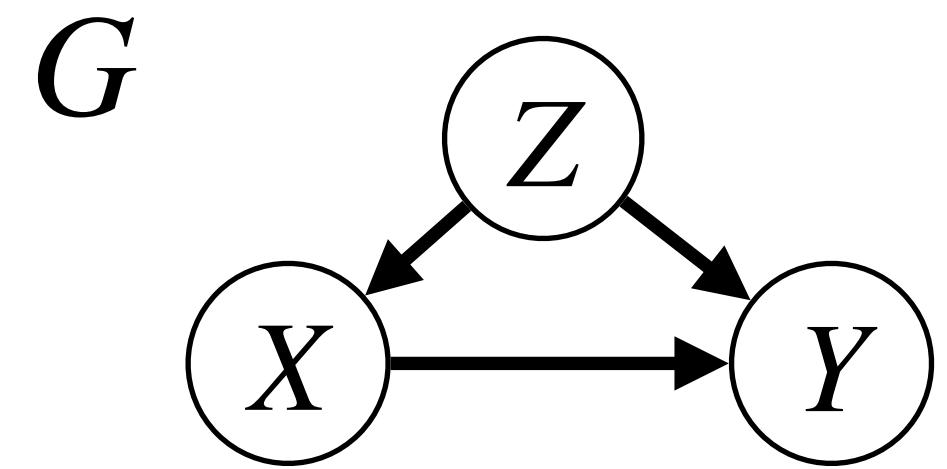
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{V} \setminus (\mathbf{Y} \cup \mathbf{X})} \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P(v_i \mid pa_i) \Big|_{\mathbf{X}=\mathbf{x}}$$

- In Markovian Models, the joint interventional distribution (and hence any causal effect) is always identifiable.
- This factorization is a.k.a “manipulation theorem” (Spirtes et al. 1993) or G-computation (Robins 1986, p. 1423).

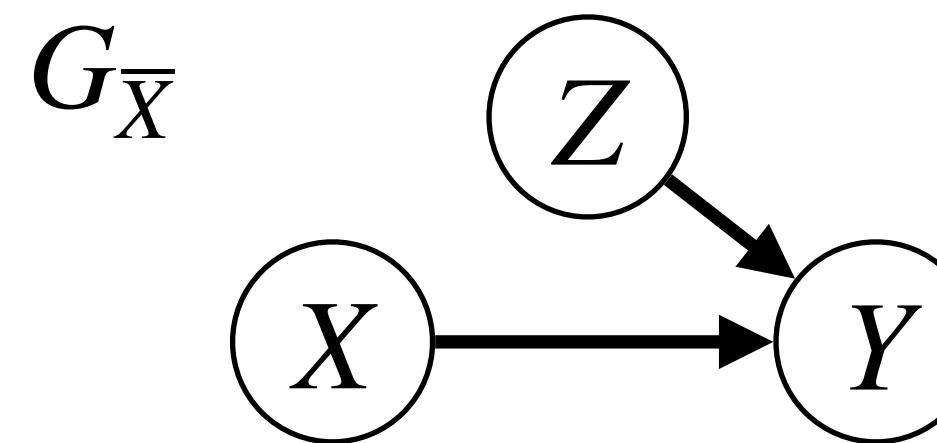
# Example: Identifiable Effect

Causal Effect of  $X$  on  $Y$ :

$$P(y \mid do(x)) = \sum_{V \setminus (Y \cup X)} \prod_{V_i \in V \setminus X} P_x(v_i \mid pa_i) \Big|_{X=x}$$



$do(X = x)$



$$P(x, y, z) = P(z)P(x \mid z)P(y \mid x, z)$$

$$P(y, z \mid do(x)) = P(z)P(y \mid x, z)$$

$$\implies P(y \mid do(x)) = \sum_z P(z)P(y \mid x, z)$$

# Identification in Markovian Models

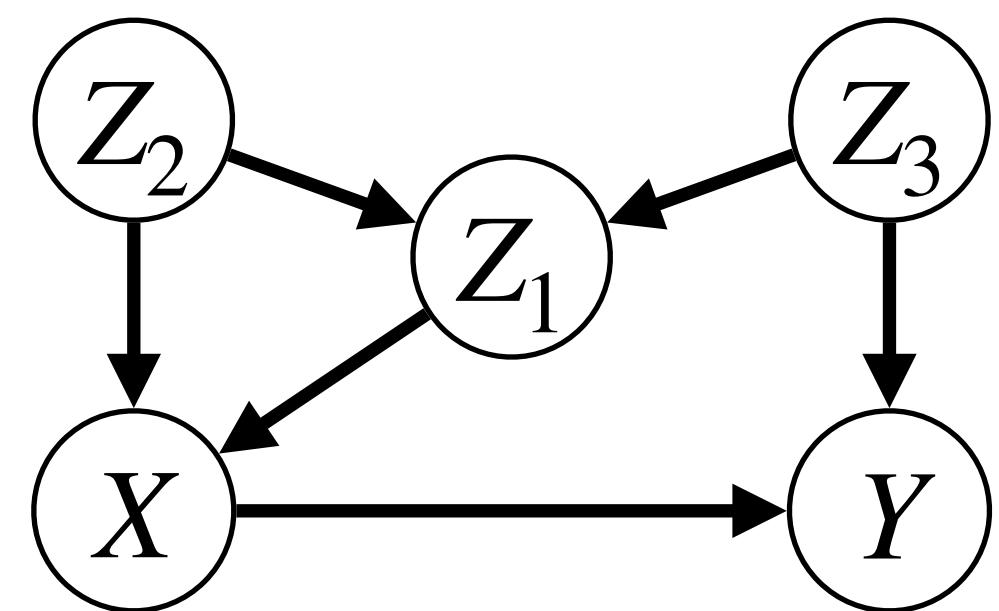
## Via Adjustment over Parents

Let  $G$  be a causal graph with no unmeasured confounders (i.e.,  $\mathcal{M}$  is Markovian).

Then, the effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is given by:

$$P(\mathbf{y} | do(\mathbf{x})) = \sum_{\mathbf{pa}_{\mathbf{x}}} P(y | \mathbf{x}, \mathbf{pa}_{\mathbf{x}}) P(\mathbf{pa}_{\mathbf{x}})$$

Proof follows from the truncated factorization for Markovian models!



$$Pa_x = \{Z_1, Z_2\}$$

$$P(y | do(x)) = \sum_{z_1, z_2} P(y | x, z_1, z_2) P(z_1, z_2)$$

$$\begin{aligned}\mathbf{X} &= \{X\} \\ \mathbf{Y} &= \{Y\} \\ \mathbf{Pa_X} &= \{Z_1, Z_2\}\end{aligned}$$

# Identification in Semi-Markovian Models

---

**Truncated Factorization – Semi-Markovian:** Let  $G$  be the causal diagram for the collection  $\mathbf{P}_*$  of all interventional distributions  $P(\mathbf{V} \mid do(\mathbf{x}))$ , for any  $\mathbf{X} \subseteq \mathbf{V}$ . The interventional distribution  $P(\mathbf{V} \mid do(\mathbf{x}))$  factorizes as followings:

$$P(\mathbf{v} \mid do(\mathbf{x})) = \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V} \setminus \mathbf{X}} P(v_i \mid pa_i, u_i) P(\mathbf{u}) \Big|_{\mathbf{X}=\mathbf{x}}$$

For Semi-Markovian Models, more advanced tools (graphical criteria, calculus, and algorithms) are necessary for evaluating  $P_{\mathbf{x}}(\mathbf{v})$  as an expression that depends only on  $P(\mathbf{V})$ , i.e., a do-free and U-free expression.

# Backdoor Adjustment

Also known as *confounding paths*, or  
*backdoor paths*.

Let  $\mathbf{X}$  be a set of treatment variables and  $\mathbf{Y}$  a set of outcome variables in the causal graph  $G$ .

If there exists a set  $\mathbf{Z}$  such that:

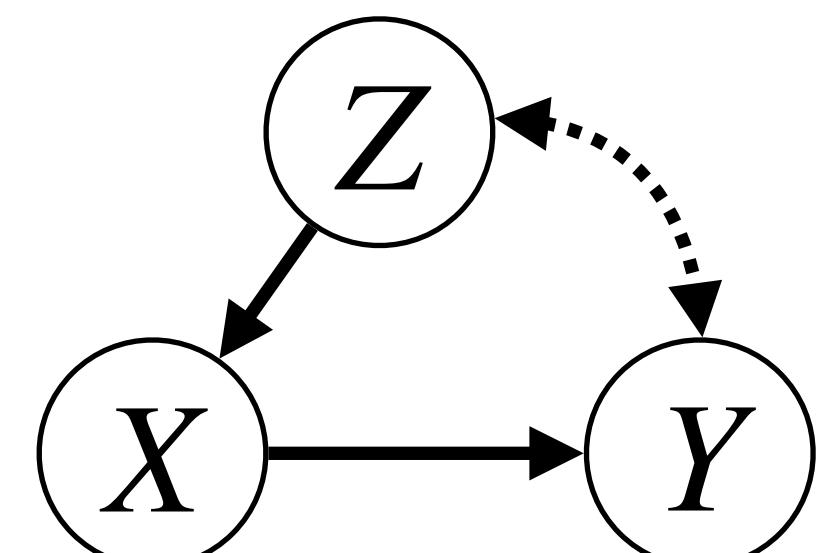
1. for every  $X \in \mathbf{X}$  and  $Y \in \mathbf{Y}$ ,  $\mathbf{Z}$  blocks every path between  $X$  and  $Y$  that has an arrow into  $X$ , and
2. no node in  $\mathbf{Z}$  is a descendant of a variable  $X \in \mathbf{X}$  (all variables in  $\mathbf{Z}$  are pre-treatment)

Then,  $\mathbf{Z}$  satisfies the **backdoor criterion** for  $(\mathbf{X}, \mathbf{Y})$  and, then the effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is given by:

$$P(\mathbf{y} | do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} | \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

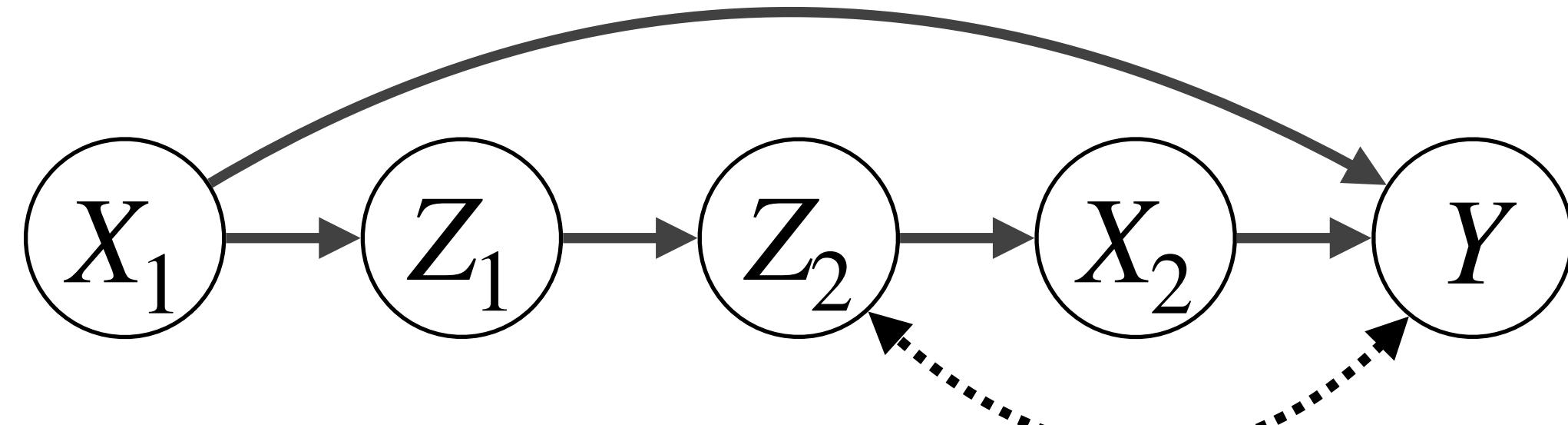
$$\begin{aligned}\mathbf{X} &= \{X\} \\ \mathbf{Y} &= \{Y\} \\ \mathbf{Z} &= \{Z\}\end{aligned}$$

$Z$ , a set of covariates,  
admissible for backdoor  
adjustment



# Is BD complete for covariate adjustment?

---



There are open backdoor paths between  $(\{X_1, X_2\}, \{Y\})$ , e.g..  $\langle X_2, Z_2, Y \rangle$  and  $\langle X_2, Z_2, Z_1, X_1, Y \rangle$ .

However, no set is admissible for **backdoor adjustment** for  $(\{X_1, X_2\}, \{Y\})$ , as both  $Z_1$  and  $Z_2$  are descendants of  $\{X_1, X_2\}$ .

$$\text{However, we have } P(y | do(x_1, x_2)) = \sum_z P(y | x_1, x_2, z_1, z_2) P(z_1, z_2)$$

Backdoor Criterion is **sound (sufficient)** but not **complete (necessary)** for covariate adjustment.

There are descendants of  $\mathbf{X}$  that may be used (and sometimes needed) for adjustment.

**Can we have a sound and complete graphical criterion for covariate adjustment?**

# The Adjustment Criterion

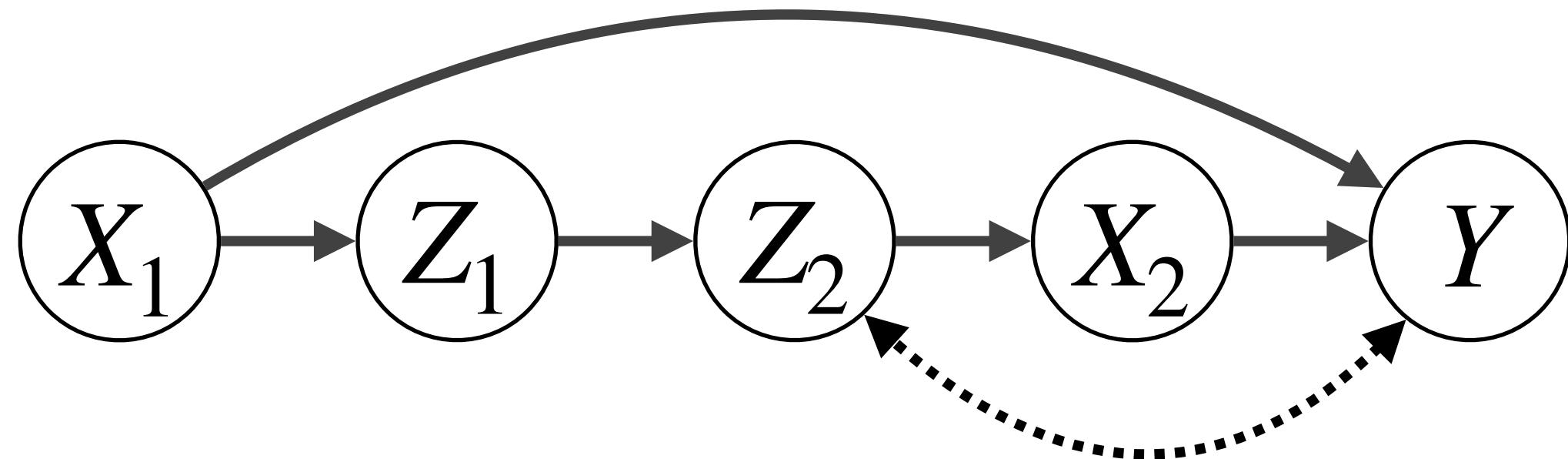
---

## **Sound and Complete Graphical Criterion for Covariate Adjustment:**

- Shpitser. et al (2010) - Shpitser, I., VanderWeele, T., & Robins, J. M. On the validity of covariate adjustment for estimating causal effects. UAI'10: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence - ([Link](#))

# Proper Paths

**Definition (Proper Path):** A path  $p$  between a node in  $\mathbf{X}$  and a node in  $\mathbf{Y}$  is said to be *proper* if only its first node is in  $\mathbf{X}$ .



Causal paths between  $\mathbf{X}$  and  $\mathbf{Y}$ :

- |   |   |                        |
|---|---|------------------------|
| $\langle X_1, Y \rangle$                | } | Proper causal paths    |
| $\langle X_2, Y \rangle$                |   | Non-proper causal path |
| $\langle X_1, Z_1, Z_2, X_2, Y \rangle$ |   |                        |

$$\mathbf{X} = \{X_1, X_2\}$$

$$\mathbf{Y} = \{Y\}$$

Non-causal paths between  $\mathbf{X}$  and  $\mathbf{Y}$ :

- |   |   |                        |
|---|---|------------------------|
| $\langle X_2, Z_2, Y \rangle$           | } | Proper non-causal path |
| $\langle X_2, Z_2, Z_1, X_1, Y \rangle$ |   |                        |

# Adjustment Criterion for ADMGs

Let  $\mathbf{X}$  be a set of treatment variables and  $\mathbf{Y}$  a set of outcome variables in the causal graph  $G$ .

$$W \in Desc(W)$$

If there exists a set  $\mathbf{Z}$  such that:

1. for every  $X \in \mathbf{X}$  and  $Y \in \mathbf{Y}$ ,  $\mathbf{Z}$  blocks every **proper non-causal path between  $X$  and  $Y$** , and
2. no node in  $\mathbf{Z}$  is a descendant in of a variable  $W \notin \mathbf{X}$  which lies on a proper causal path from  $\mathbf{X}$  to  $\mathbf{Y}$

$Z$  cannot contain a variable in a proper causal path or a descendant of such variable.

Then,  $\mathbf{Z}$  satisfies the **adjustment criterion** for  $(\mathbf{X}, \mathbf{Y})$  and, then the effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is given by:

$$P(\mathbf{y} | do(\mathbf{x})) = \sum_{\mathbf{z}} P(\mathbf{y} | \mathbf{x}, \mathbf{z}) P(\mathbf{z})$$

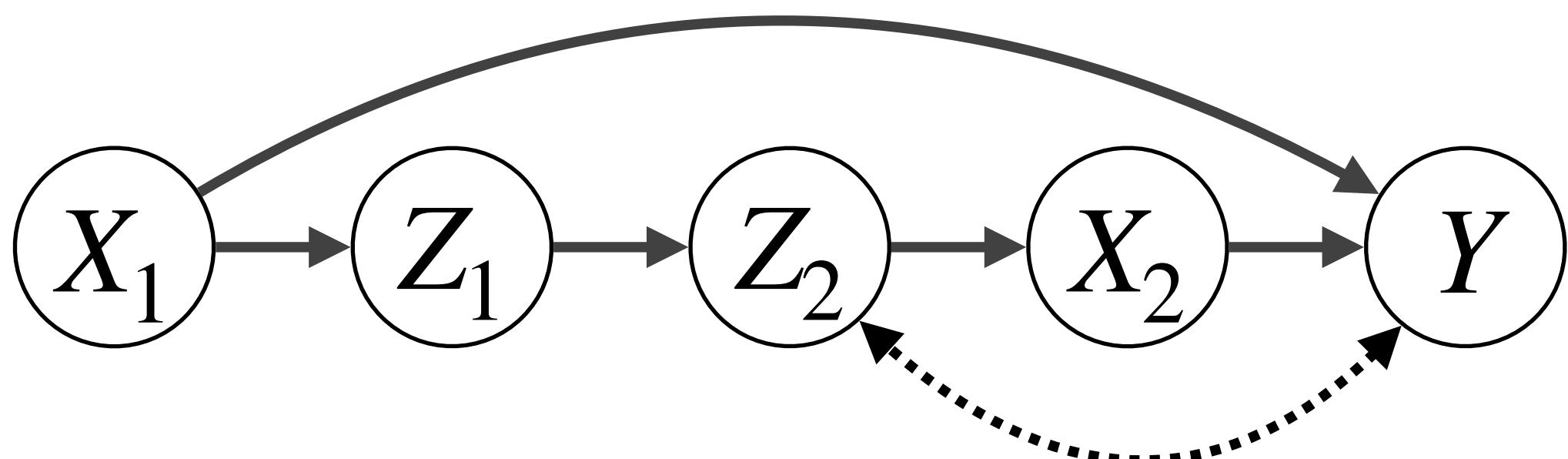
- $\mathbf{Z} = \{\}$  X  
 $\mathbf{Z} = \{Z_2\}$  X  
 $\mathbf{Z} = \{Z_1, Z_2\}$  ✓

Proper non-causal paths between  $\mathbf{X}$  and  $\mathbf{Y}$ :

$$\langle X_2, Z_2, Y \rangle$$

Proper causal paths between  $\mathbf{X}$  and  $\mathbf{Y}$ :

$$\begin{aligned} &\langle X_1, Y \rangle \\ &\langle X_2, Y \rangle \end{aligned}$$



# Estimation via Propensity Scores

**Theorem:** If the set  $\mathbf{Z}$  satisfies the backdoor or generalized adjustment criterion w.r.t. the ordered pair  $(X, Y)$  in the causal graph  $G$ , then the causal effect of  $X$  on  $Y$  is identifiable (**uniquely computable**) and given by:

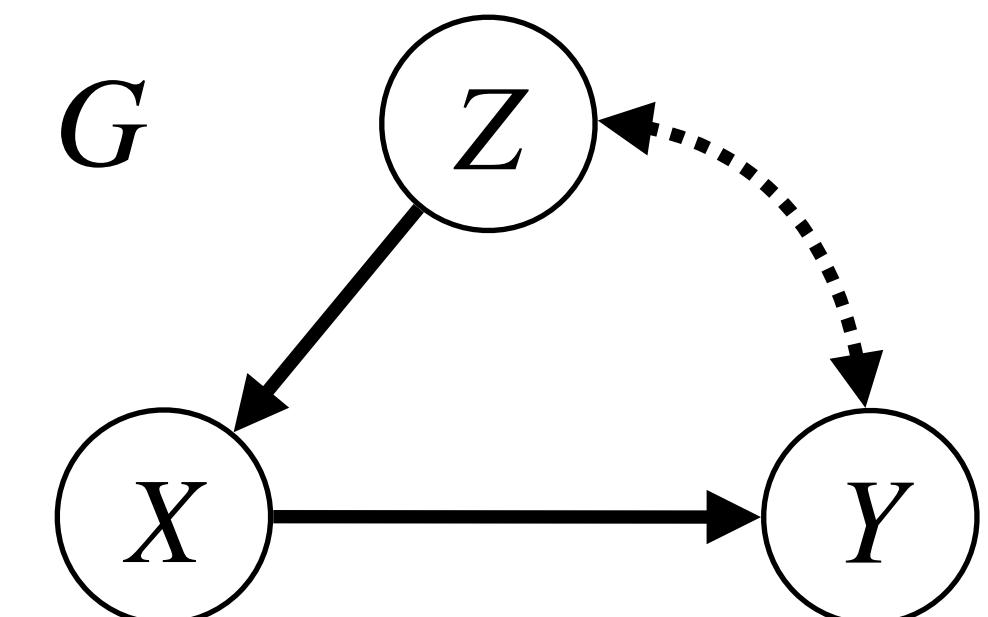
BD Adjustment  $\equiv$   
Conditional Ignorability:

$$Y_x \perp\!\!\!\perp X | \mathbf{Z}$$

Only if  $\mathbf{Z}$  satisfies the BD criterion,  
Inverse Probability Weighting/  
Propensity Score can be used to  
estimate  $P(y | do(x))$ .

$$\begin{aligned} P(y | do(x)) &= \sum_{\mathbf{z}} P(y | x, \mathbf{z})P(\mathbf{z}) \\ &= \sum_{\mathbf{z}} \frac{P(y | x, \mathbf{z})P(x | \mathbf{z})P(\mathbf{z})}{P(x | \mathbf{z})} \\ &= \sum_{\mathbf{z}} \frac{P(y, x, \mathbf{z})}{P(x | \mathbf{z})} \end{aligned}$$

propensity score  
neural nets



More on this later, with  
Fredrik Johansson

# What if covariate adjustment does not work?

## Identification via Front-Door Adjustment

Let  $\mathbf{X}$  be a set of treatment variables and  $\mathbf{Y}$  a set of outcome variables in the causal graph  $G$ .

If there exists a set  $\mathbf{M}$  such that:

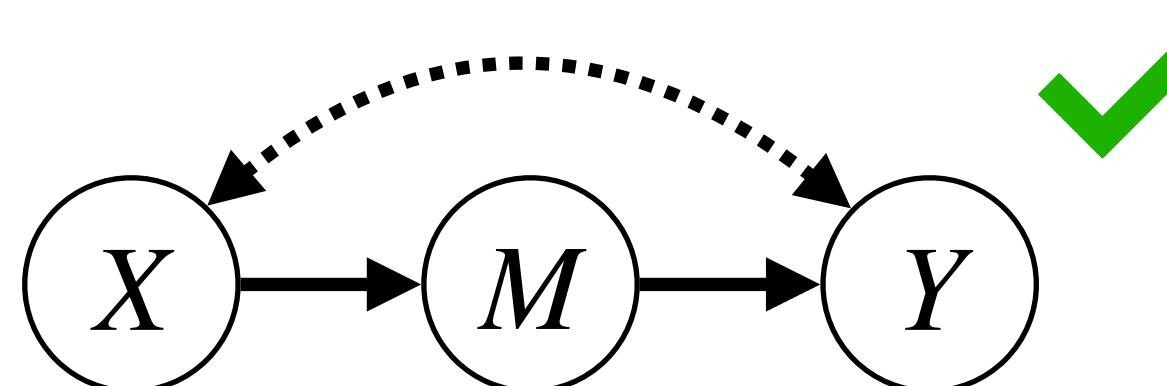
1.  $\mathbf{M}$  intercepts all directed paths from any vertex  $X \in \mathbf{X}$  to any vertex  $Y \in \mathbf{Y}$ ;
2. There is no unblocked back-door path from any vertex  $X \in \mathbf{X}$  to vertex  $M \in \mathbf{M}$ ; and
3. All back-door paths from any vertex  $M \in \mathbf{M}$  to any vertex  $Y \in \mathbf{Y}$  are blocked by  $\mathbf{X}$ .

Then,  $\mathbf{M}$  satisfies the ***front-door criterion*** and, then the effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is given by:

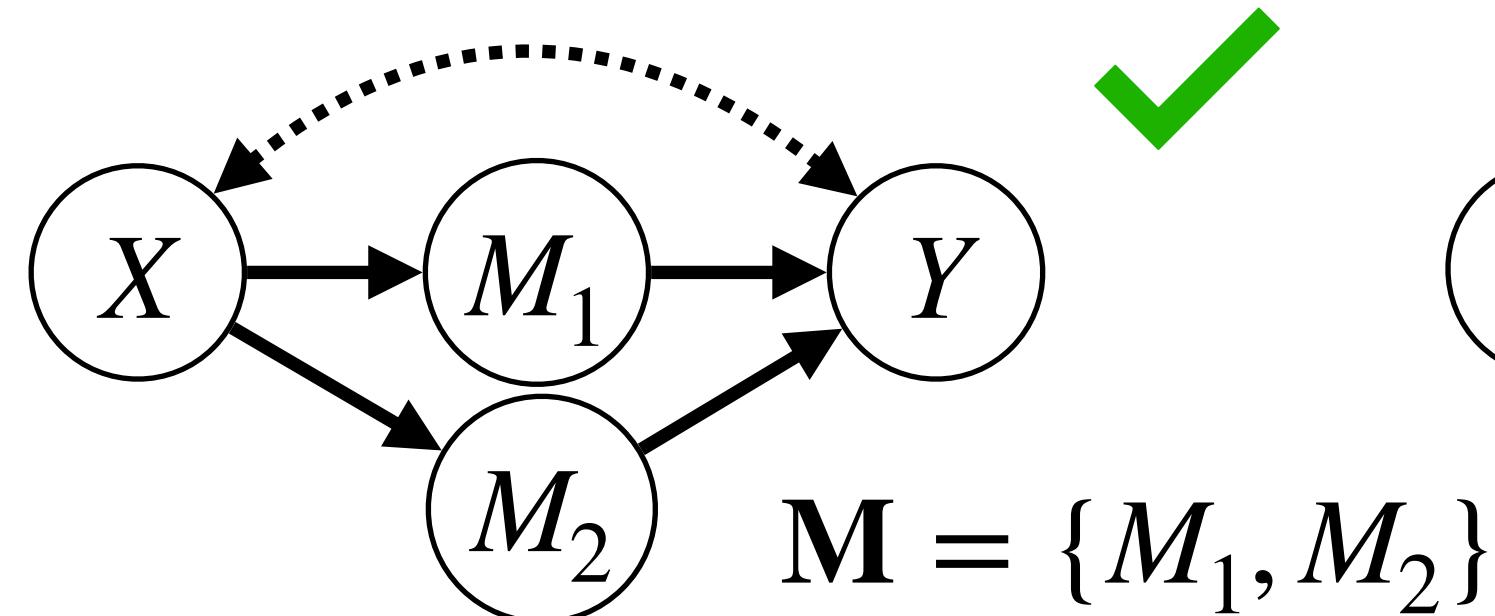
$$P(\mathbf{y} | do(\mathbf{x})) = \sum_{\mathbf{m}} P(\mathbf{m} | \mathbf{x}) \sum_{\mathbf{x}'} P(\mathbf{y} | \mathbf{m}, \mathbf{x}') P(\mathbf{x}')$$

$$\mathbf{X} = \{X\}$$

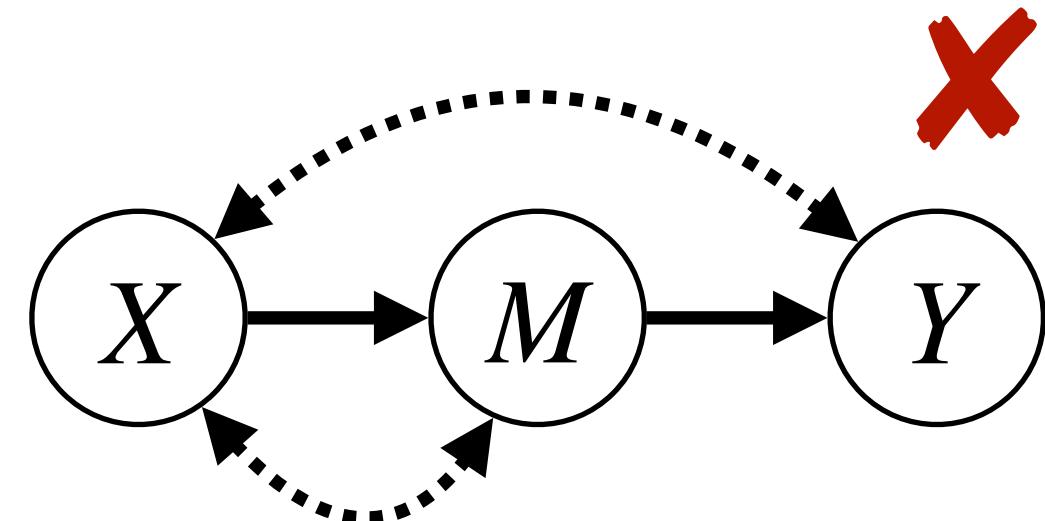
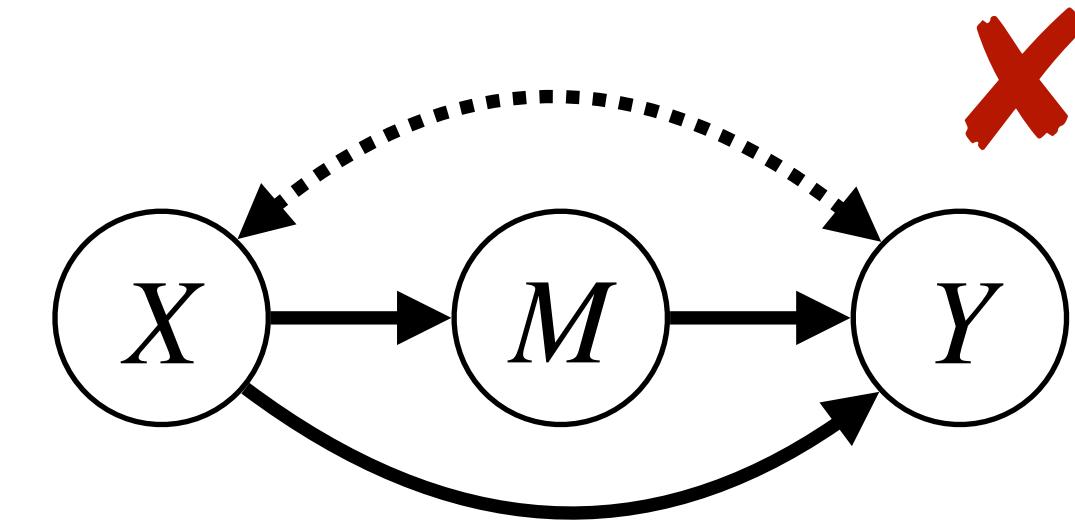
$$\mathbf{Y} = \{Y\}$$



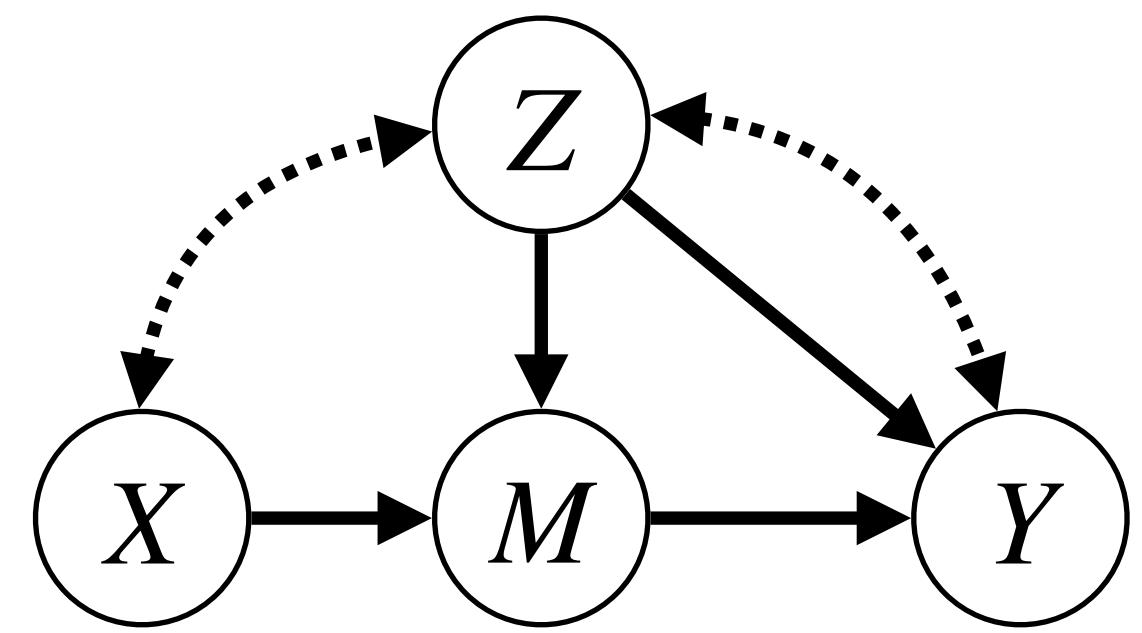
$$\mathbf{M} = \{M\}$$



$$\mathbf{M} = \{M_1, M_2\}$$

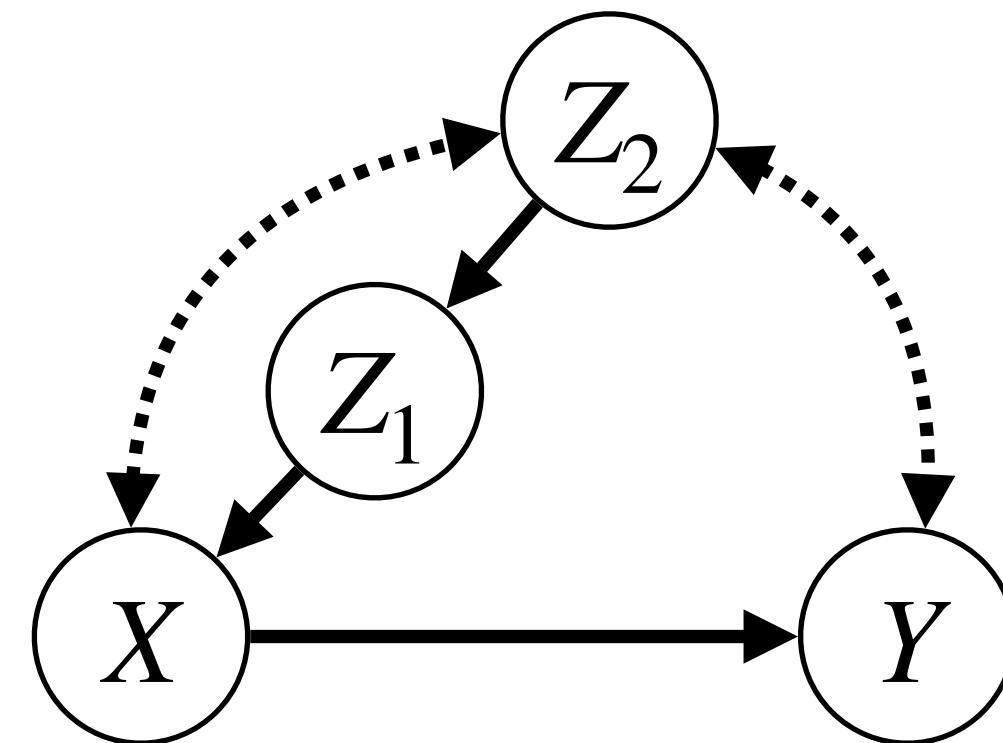


# Many scenarios beyond back-door and front-door!



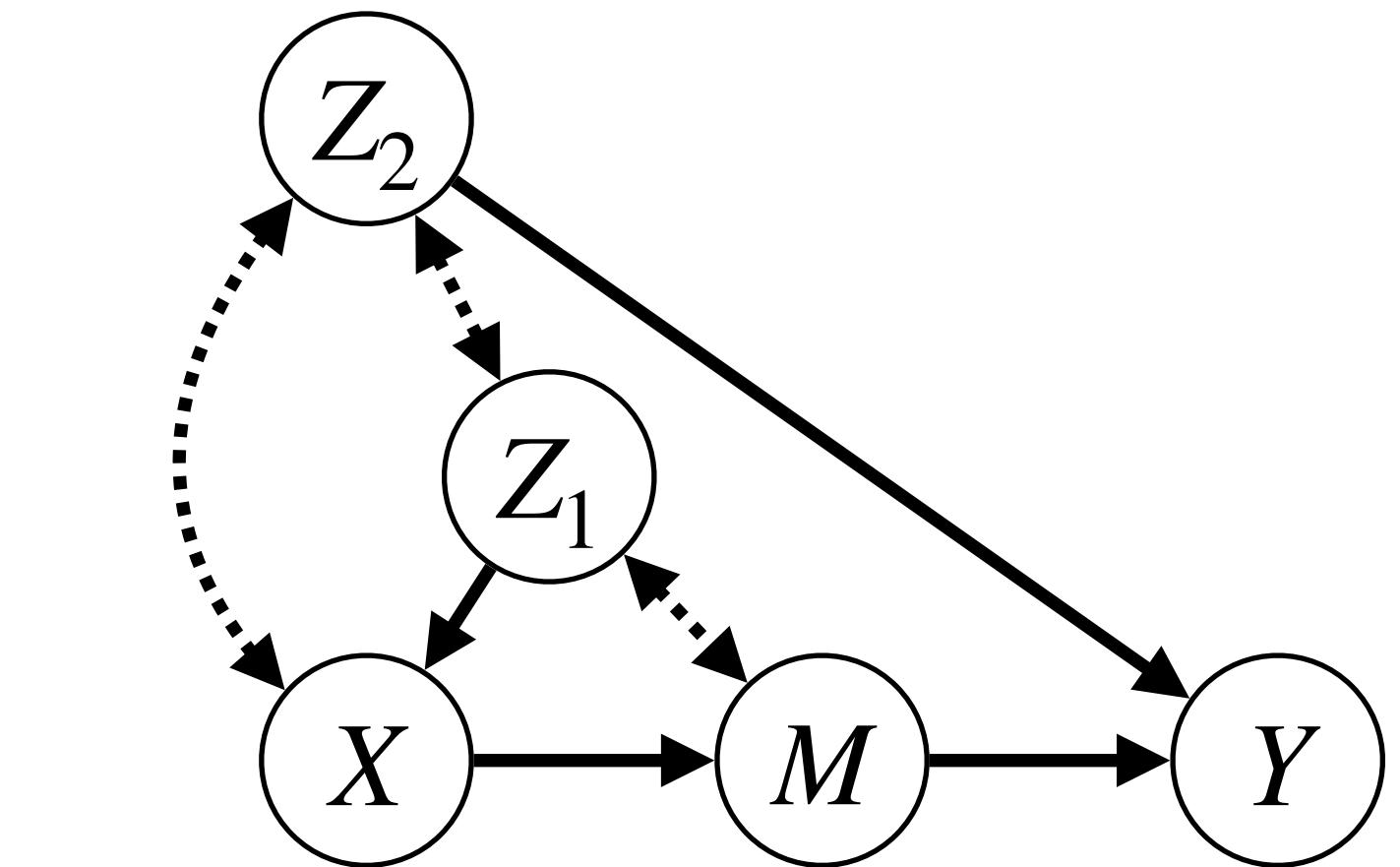
Conditional Front-Door

$$P(y | do(x)) = \sum_{m,z} P(m | x, z) \sum_{x'} P(y | m, x', z) P(x', z)$$



Napkin

$$P(y | do(x)) = \frac{\sum_{z_2} P(x, y | z_1, z_2) P(z_2)}{\sum_{z_2} P(x | z_1, z_2) P(z_2)}$$



Unnamed

$$P(y | do(x)) = \sum_{z_2, z_3} P(y | x, z_1, z_2, z_3) P(z_2) \sum_{z_1} P(z_3 | x, z_1) P(z_1)$$

And many others....

# Do-Calculus (a.k.a. Causal Calculus)

Graphical conditions implying invariances between observational ( $\mathcal{L}_1$ ) and interventional ( $\mathcal{L}_2$ ) distributions

Pearl, 1995

**Theorem:** Let  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$  be any disjoint subjects of variables.

**Rule 1** (Insertion/Deletion of Observations)

$$P(\mathbf{y} | do(\mathbf{w}), \mathbf{x}, \mathbf{z}) = P(\mathbf{y} | do(\mathbf{w}), \mathbf{z}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}, \mathbf{W})_{G_{\overline{\mathbf{W}}}}$$

**Rule 2** (Exchange of Actions and Observations)

$$P(\mathbf{y} | do(\mathbf{w}), do(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} | do(\mathbf{w}), \mathbf{x}, \mathbf{z}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}, \mathbf{W})_{G_{\overline{\mathbf{W}}\underline{\mathbf{X}}}}$$

**Rule 3** (Insertion/Deletion of Actions)

$$P(\mathbf{y} | do(\mathbf{w}), do(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} | do(\mathbf{w}), \mathbf{z}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}, \mathbf{W})_{G_{\overline{\mathbf{W}}, \mathbf{X}(\mathbf{Z})}}$$

$G_{\overline{\mathbf{W}}\underline{\mathbf{X}}}$ : graph  $G$  after removing the incoming arrows into  $\mathbf{W}$  and the outgoing arrows from  $\mathbf{X}$ ;

$\mathbf{X}(\mathbf{Z})$ : set of  $\mathbf{X}$ -nodes that are not ancestors of any  $\mathbf{Z}$ -node in  $G_{\overline{\mathbf{W}}}$ .

# Do-Calculus - Rule 1

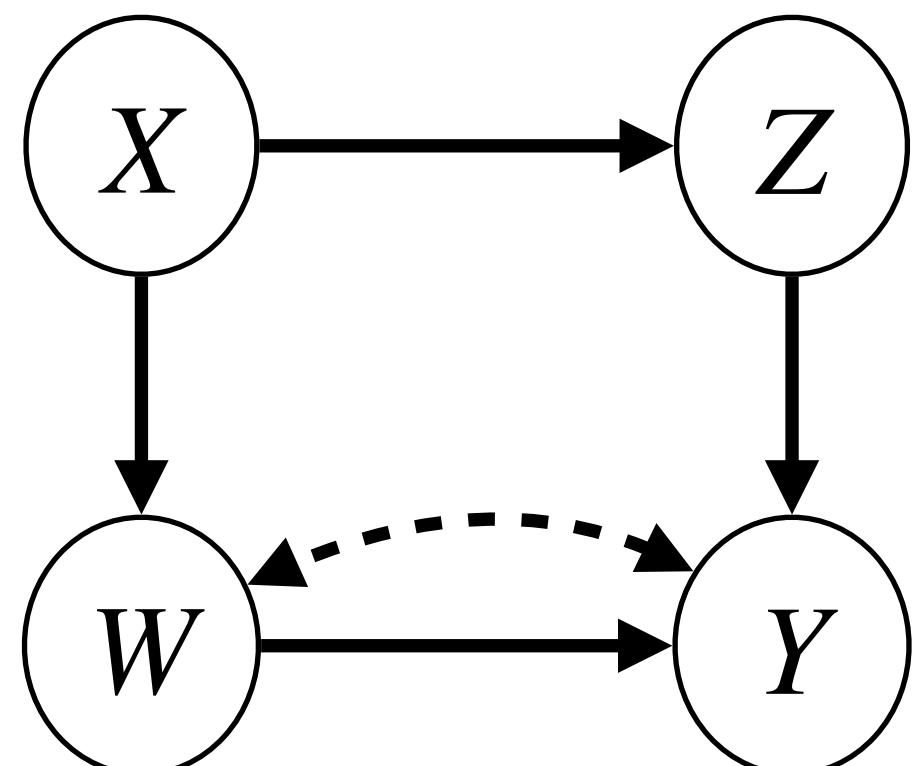
**Theorem:** Let  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$  be any disjoint subjects of variables.

**Rule 1** (Insertion/Deletion of Observations)

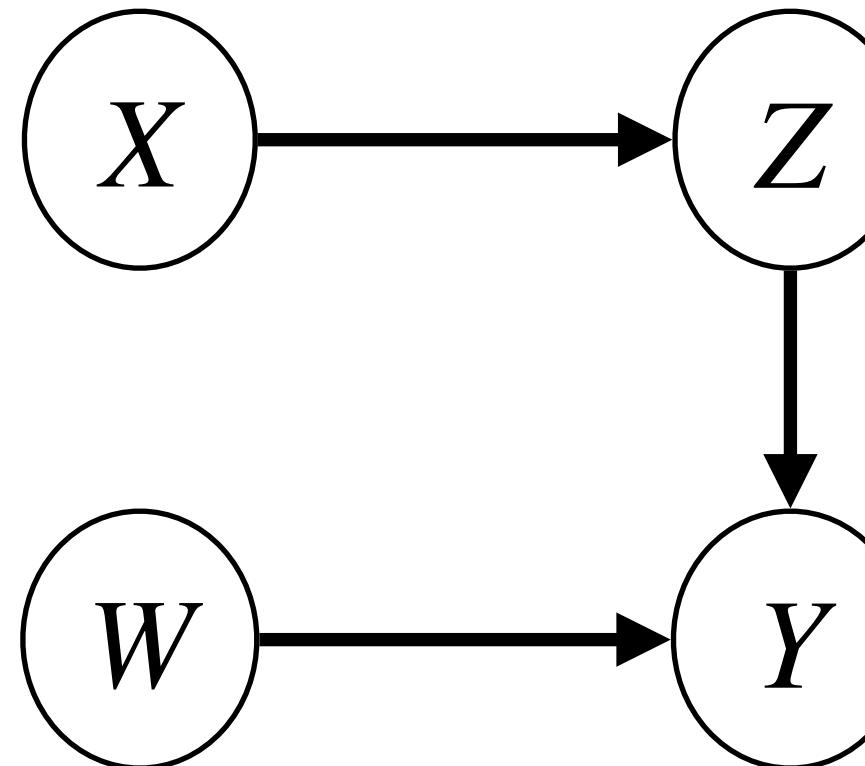
$$P(\mathbf{y} | do(\mathbf{w}), \mathbf{x}, \mathbf{z}) = P(\mathbf{y} | do(\mathbf{w}), \mathbf{z}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}, \mathbf{W})_{G_{\overline{\mathbf{W}}}}$$

$\mathbf{X}$  is conditionally independent  
of  $\mathbf{Y}$  given  $\mathbf{Z} \cup \mathbf{W}$  in the  
interventional model  $G_{\overline{\mathbf{W}}}$

$G$



$G_{\overline{\mathbf{W}}}$



$$(Y \perp\!\!\!\perp X | Z, W)_{G_{\overline{\mathbf{W}}}}$$

$$\implies P(y | do(w), \mathbf{x}, z) = P(y | do(w), z)$$

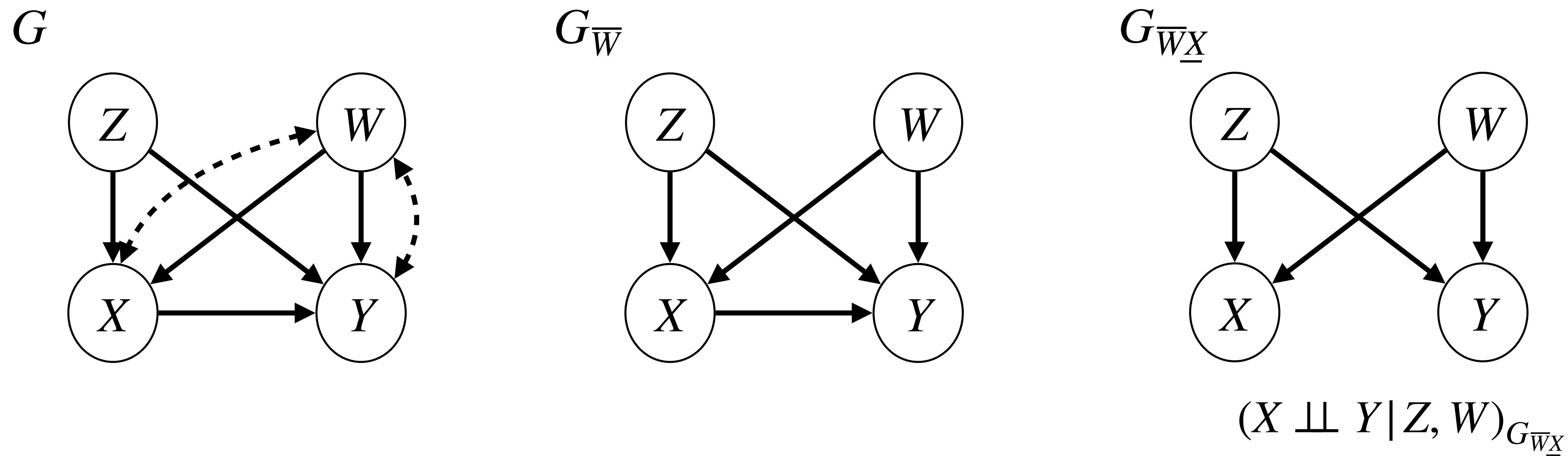
# Do-Calculus - Rule 2

**Theorem:** Let  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$  be any disjoint subjects of variables.

**Rule 2** (Exchange of Actions and Observations)

$$P(\mathbf{y} | do(\mathbf{w}), \textcolor{red}{do}(\mathbf{x}), \mathbf{z}) = P(\mathbf{y} | do(\mathbf{w}), \textcolor{red}{x}, \mathbf{z}), \text{ if } (\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{Z}, \mathbf{W})_{G_{\overline{W}\underline{X}}}$$

$\mathbf{X}$  and  $\mathbf{Y}$  are unconfounded  
given  $\mathbf{Z} \cup \mathbf{W}$  in the  
interventional model  $G_{\overline{W}}$



$$\implies P(y | do(w), \textcolor{red}{do}(x), z) = P(y | do(w), \textcolor{red}{x}, z)$$

# Do-Calculus - Rule 3

**Theorem:** Let  $X, Y, Z, W$  be any disjoint subjects of variables.

**Rule 3 (Insertion/Deletion of Actions)**

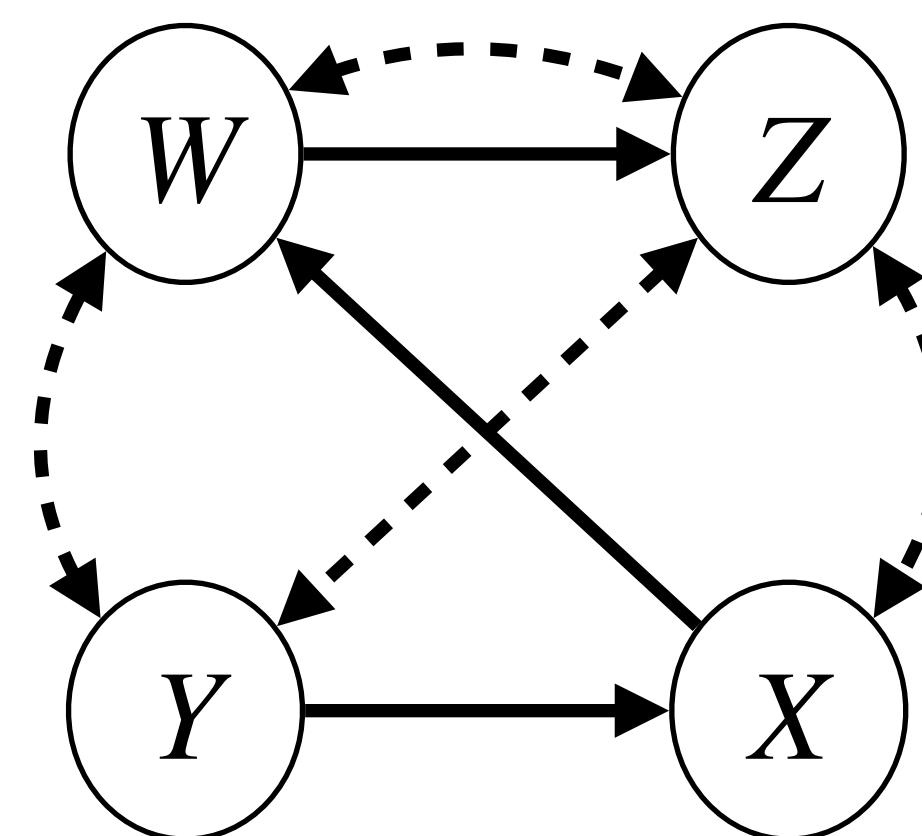
$$P(y | do(w), \textcolor{red}{do(x)}, z) = P(y | do(w), z), \text{ if } (Y \perp\!\!\!\perp X | Z, W)_{G_{\overline{W}, \overline{X}(Z)}}$$

$Y$  is not affected by any action on  $X$  given  $Z \cup W$  in the interventional model  $G_{\overline{W}}$

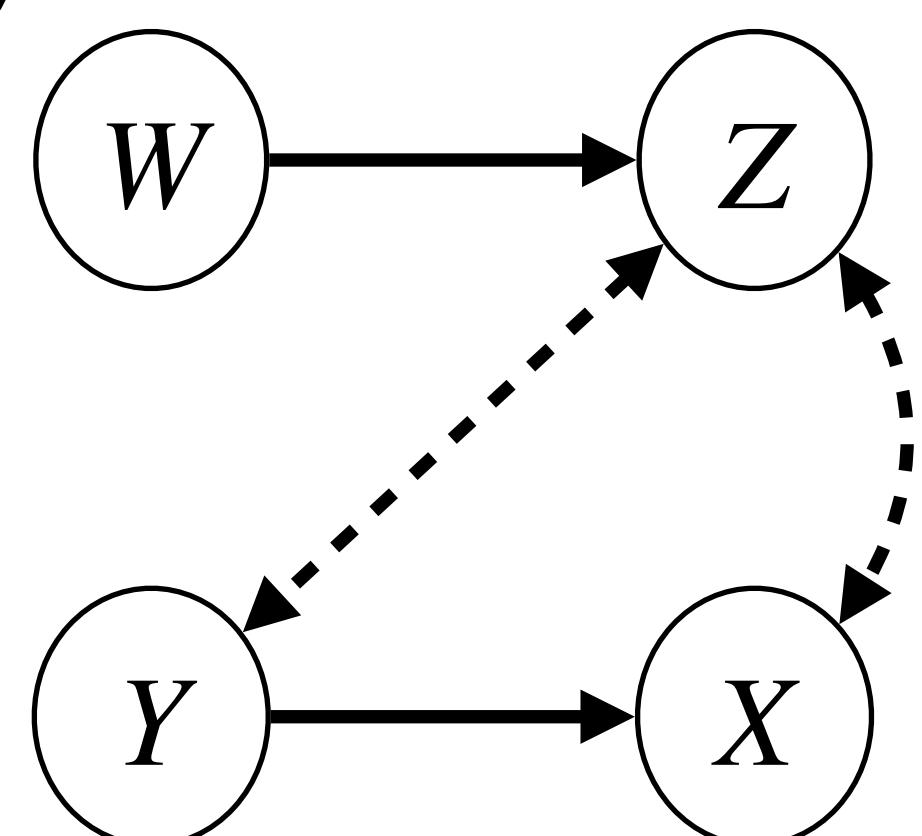
$G_{\overline{W}X}$ : graph  $G$  after removing the incoming arrows into  $W$  and the outgoing arrows from  $X$ ;

$\overline{X}(Z)$ : set of  $X$ -nodes that are not ancestors of any  $Z$ -node in  $G_{\overline{W}}$ .

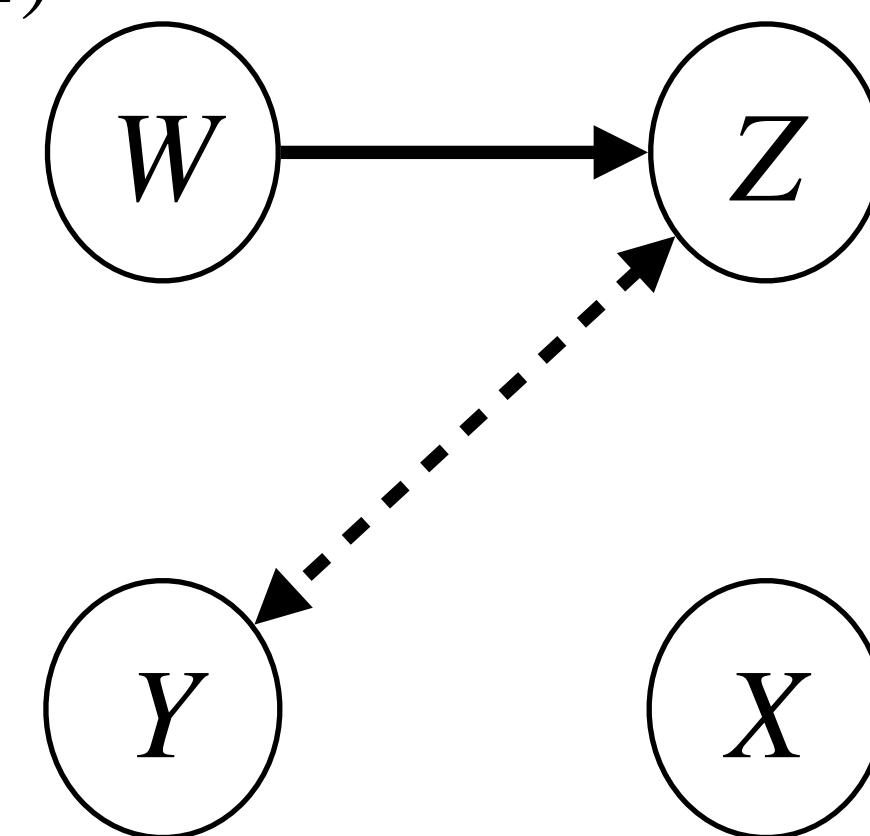
$G$



$G_{\overline{W}}$



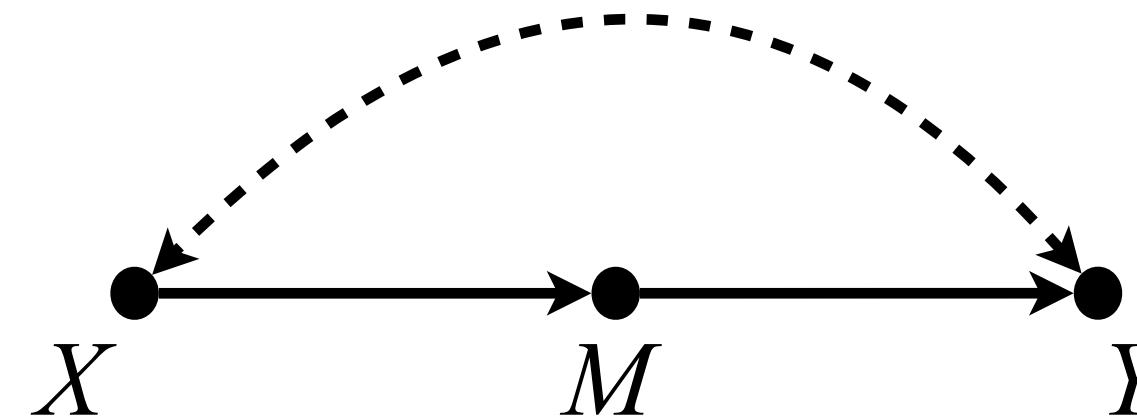
$G_{\overline{W}, \overline{X}(Z)}$



$$(Y \perp\!\!\!\perp X | Z, W)_{G_{\overline{W}, \overline{X}(Z)}}$$

$$\implies P(y | do(w), \textcolor{red}{do(x)}, z) = P(y | do(w), z) \quad 62$$

# Identification in Non-Markovian Models



$$\begin{aligned} P(y \mid do(x)) &= \sum_m P(y \mid do(x), m)P(m \mid do(x)) && \text{Probability Axioms} \\ &= \sum_m P(y \mid do(x), do(m))P(m \mid do(x)) && \text{Rule 2} \\ &= \sum_m P(y \mid do(x), do(m))P(m \mid x) && \text{Rule 2} \\ &= \sum_m P(y \mid do(m))P(m \mid x) && \text{Rule 3} \\ &= \sum_{x'} \sum_m P(y \mid do(m), x')P(x' \mid do(m))P(m \mid x) && \text{Probability Axioms} \\ &= \sum_{x'} \sum_m P(y \mid m, x')P(x' \mid do(m))P(m \mid x) && \text{Rule 2} \\ &= \sum_{x'} \sum_m P(y \mid m, x')P(x' \mid m)P(m \mid x) && \text{Rule 3} \end{aligned}$$



# The Identify (ID) Algorithm

---

---

**Algorithm 1** ID( $\mathbf{x}, \mathbf{y}$ ) given Causal Diagram  $\mathcal{G}$

---

**Input:** two disjoint sets  $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$

**Output:** Expression for  $P_{\mathbf{x}}(\mathbf{y})$  or FAIL

```
1: Let  $\mathbf{D} = \text{An}(\mathbf{Y})_{\mathcal{G}_{\mathbf{V} \setminus \mathbf{x}}}$ 
2: Let the c-components of  $\mathcal{G}_{\mathbf{D}}$  be  $\mathbf{D}_i, i = 1, \dots, k$ 
3:  $P_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{d} \setminus \mathbf{y}} \prod_i \text{IDENTIFY}(\mathbf{D}_i, \mathbf{V}, P)$ 
4: function IDENTIFY( $\mathbf{C}, \mathbf{T}, Q = Q[\mathbf{T}]$ )
5:   if  $\mathbf{C} = \mathbf{T}$  then return  $Q[\mathbf{T}]$ 
    /* Let  $S^B$  denote the c-component of  $\{B\}$  in  $\mathcal{G}_{\mathbf{T}}$  */
6:   if  $\exists B \in \mathbf{T} \setminus \mathbf{C}$  such that  $S^B \cap \text{Ch}(B) = \emptyset$  then
7:     Compute  $Q[\mathbf{T} \setminus \{B\}]$  from  $Q$ ;      ▷ Lemma 1
8:     return IDENTIFY( $\mathbf{C}, \mathbf{T} \setminus \{B\}, Q[\mathbf{T} \setminus \{B\}]$ )
9:   else
10:    throw FAIL
```

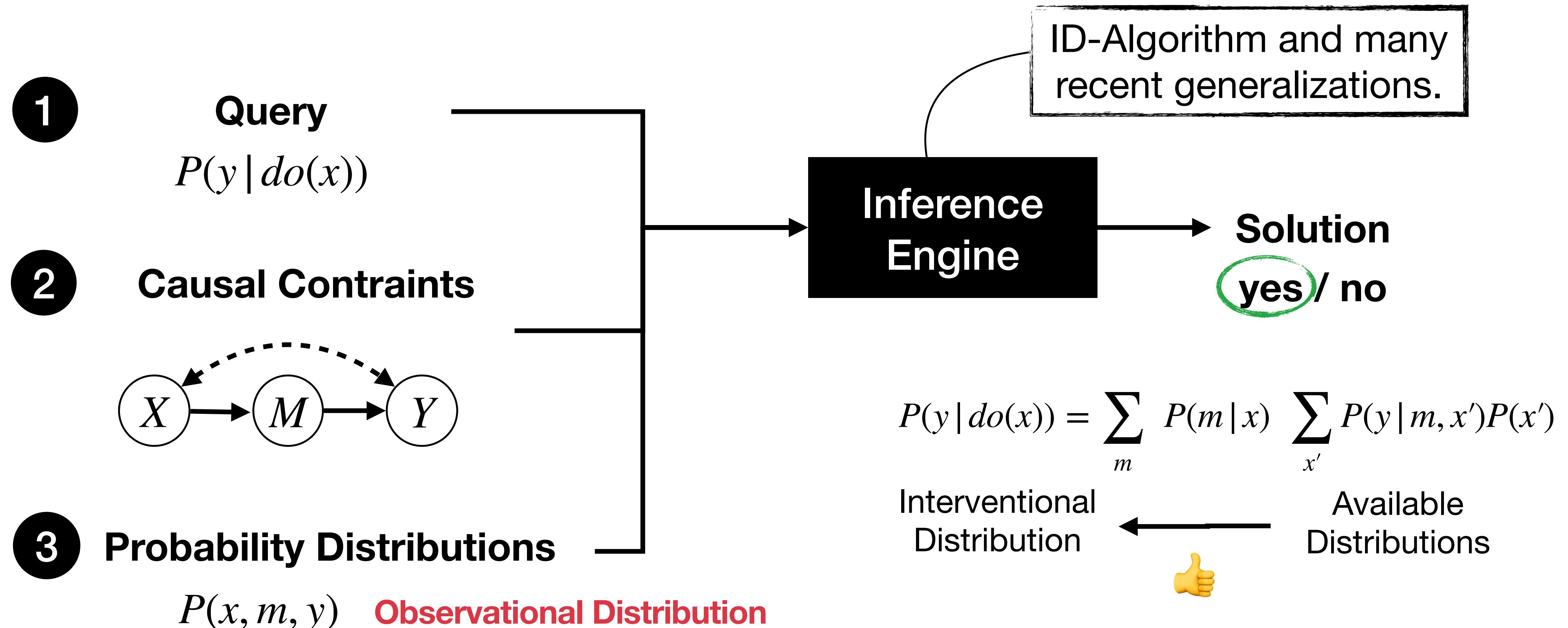
---

**Lemma 1.** Given a causal diagram  $\mathcal{D}$  over  $\mathbf{V}$ ,  $X \in \mathbf{T} \subseteq \mathbf{V}$ , and  $P_{\mathbf{v} \setminus \mathbf{t}}$ , i.e., an expression for  $Q[\mathbf{T}]$ . If  $X$  is not in the same c-component with a child in  $\mathcal{D}_{\mathbf{T}}$ , then  $Q[\mathbf{T} \setminus \{X\}]$  is identifiable and given by

$$Q[\mathbf{T} \setminus \{X\}] = \frac{P_{\mathbf{v} \setminus \mathbf{t}}}{Q[S^X]} \times \sum_x Q[S^X] \quad (2)$$

where  $S^X$  is the c-component of  $X$  in  $\mathcal{D}_{\mathbf{T}}$  and  $Q[S^X]$  is computable from  $P_{\mathbf{v} \setminus \mathbf{t}}$  by [Tian, 2002, Lemma 11].

# Causal Effect Identification



- Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pp. 567–573, Menlo Park, CA, 2002. AAAI Press/MIT Press.

# More on Causal Effect Identification

---

## **Identification from observational and experimental data:**

Lee, S., Correa, J., and Bareinboim, E. (2019). General identifiability with arbitrary surrogate experiments. In *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence*, volume 35, Tel Aviv, Israel. AUAI Press. [Link](#)

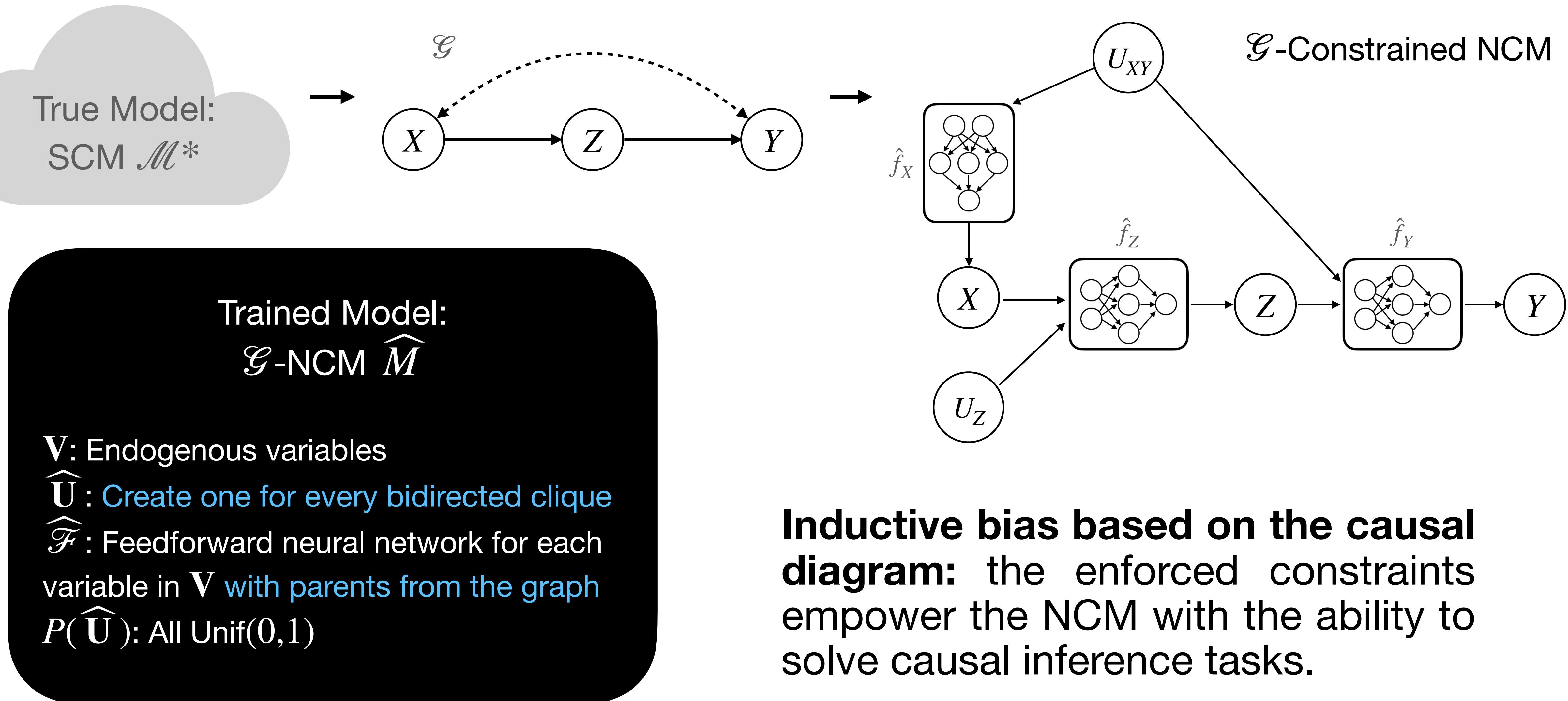
## **Identification of stochastic/soft (and possibly imperfect) interventions:**

Correa, J. and Bareinboim, E. (2020). A calculus for stochastic interventions: Causal effect identification and surrogate experiments. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, NY. AAAI Press. [Link](#)

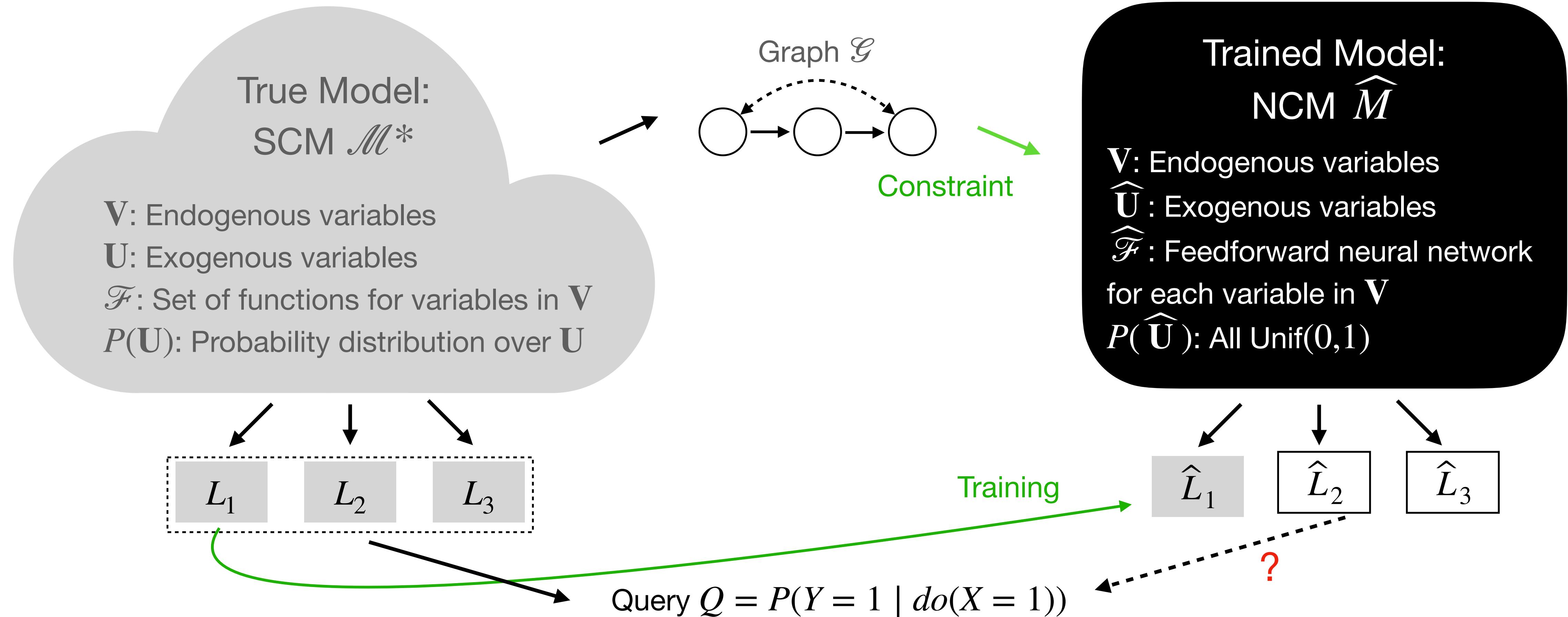
## **Identification and Estimation via Deep Neural Networks:**

Xia, K., Lee, K.-Z., Bengio, Y., and Bareinboim, E. (2021). The causal-neural connection: Expressiveness, learnability, and inference. *Advances in Neural Information Processing Systems*, 34. [Link](#)

# Identification and Estimation via Deep Neural Networks



# Expressiveness of NCMs



**Thm:** For any SCM  $\mathcal{M}^*$ , there exists an NCM  $\hat{M}$  such that  $\hat{M}$  matches  $\mathcal{M}^*$  on all three PCH layers!

This does not imply that the estimated NCM  $\hat{M}$  matches the true SCM  $\mathcal{M}^*$ !

# Solution: A Neural Algorithm for Identification

---

**Algorithm 1:** Identifying/estimating queries with NCMs.

---

**Input :** causal query  $Q = P(\mathbf{y} \mid do(\mathbf{x}))$ ,  $L_1$  data  $P(\mathbf{v})$ , and causal diagram  $\mathcal{G}$

**Output:**  $P^{\mathcal{M}^*}(\mathbf{y} \mid do(\mathbf{x}))$  if identifiable, FAIL otherwise.

```
1  $\widehat{M} \leftarrow \text{NCM}(\mathbf{V}, \mathcal{G})$                                 // from Def. 7
2  $\theta_{\min}^* \leftarrow \arg \min_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
3  $\theta_{\max}^* \leftarrow \arg \max_{\theta} P^{\widehat{M}(\theta)}(\mathbf{y} \mid do(\mathbf{x}))$  s.t.  $L_1(\widehat{M}(\theta)) = P(\mathbf{v})$ 
4 if  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x})) \neq P^{\widehat{M}(\theta_{\max}^*)}(\mathbf{y} \mid do(\mathbf{x}))$  then
5   | return FAIL
6 else
7   | return  $P^{\widehat{M}(\theta_{\min}^*)}(\mathbf{y} \mid do(\mathbf{x}))$  // choose min or max arbitrarily
```

---

Maximize and minimize the induced causal query  $Q$  while maintaining  $L_1$ -consistency (can be done with likelihood estimation).

**Thm.** :  $Q$  is identifiable if and only if they match!

**Corol:** If  $Q$  is identifiable, then we can compute it by performing the mutilation procedure on  $\widehat{M}$ !

The approach is equivalent to established symbolic approaches (Thm. 4), and in identifiable cases, the result is an NCM that can serve as a proxy model for estimating the query (Corol. 2).

# Code Examples

---

**Prob-AI GitHub (Day 5):** [https://github.com/probabilisticai/probai-2023/tree/main/day\\_5](https://github.com/probabilisticai/probai-2023/tree/main/day_5)

- **Example 1:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_BD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_BD.ipynb)
- **Example 2:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_NonBD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_NonBD.ipynb)

**Check Item 3 of both Examples:**

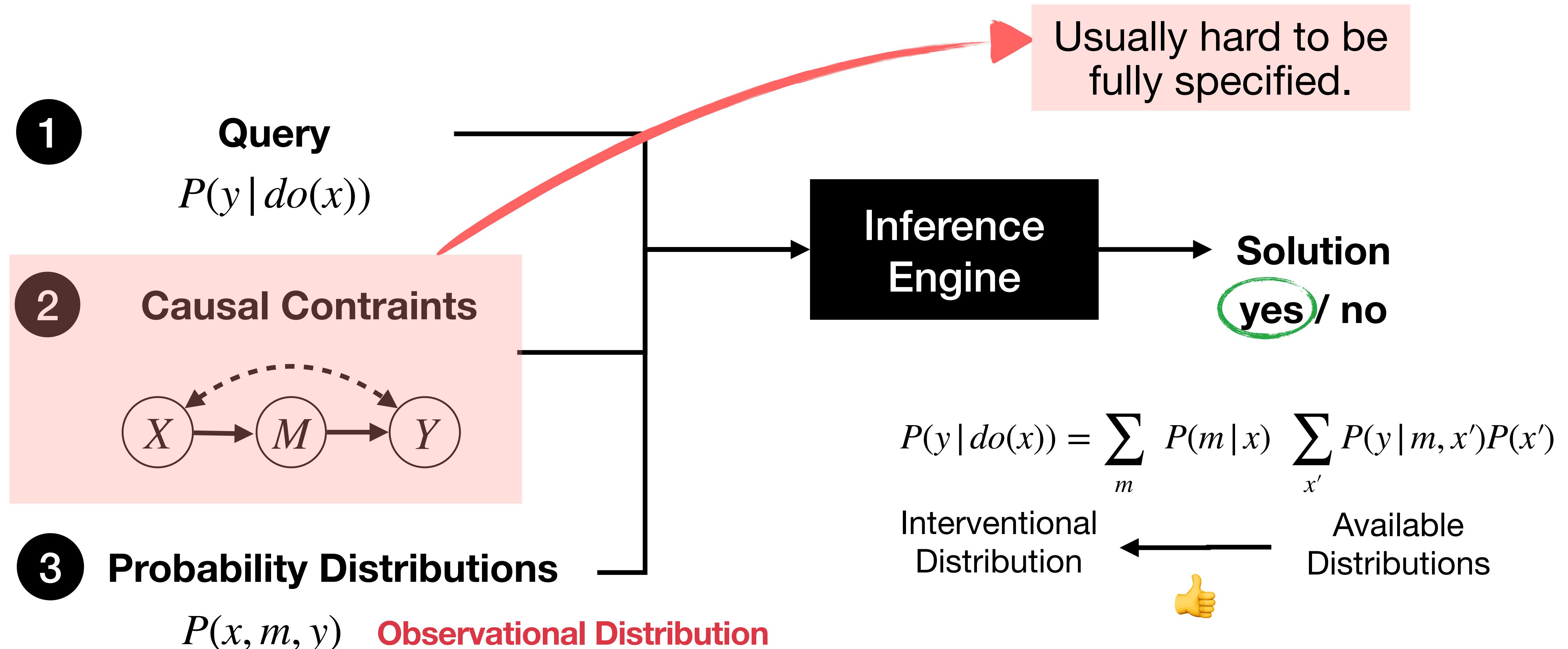
3. Causal Effect Identification from the ADMG
  - A. Identification through Adjustment — using gac function in the pcalg R package
  - B. Identification through ID Algorithm — using the causaleffect R package

# Questions?



**Can we relax some  
causal assumptions?**

# Causal Effect Identification



- Tian, J. and Pearl, J. A General Identification Condition for Causal Effects. In Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002), pp. 567–573, Menlo Park, CA, 2002. AAAI Press/MIT Press.

# Is a Causal Diagram Still Too Much?

---

Causal diagrams are powerful tools that allow for inferences based on weaker knowledge (structural invariances) than the encoded in the true, underlying SCM.

Still, structural knowledge for **every pair of variables** may not be available in many real-world, complex, high-dimensional systems.

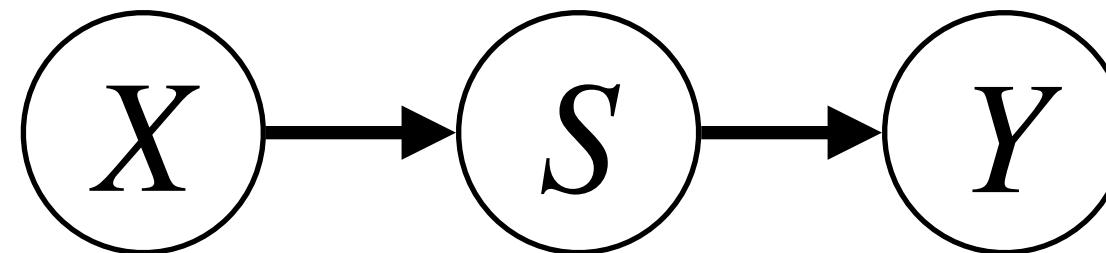
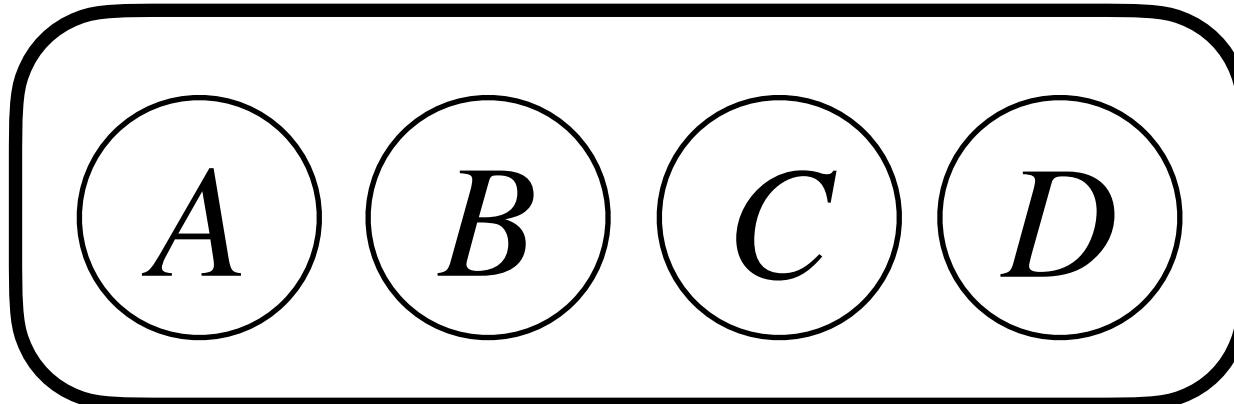
## Question:

Is it possible to relax the assumption of having a fully specified causal diagram and still be able to identify a causal effect?

# Partially Understood Systems

---

- (A) Age
- (B) Blood pressure
- (C) Comorbidities
- (D) Medication history
- (X) Lisinopril
- (S) Sleep Quality
- (Y) Stroke

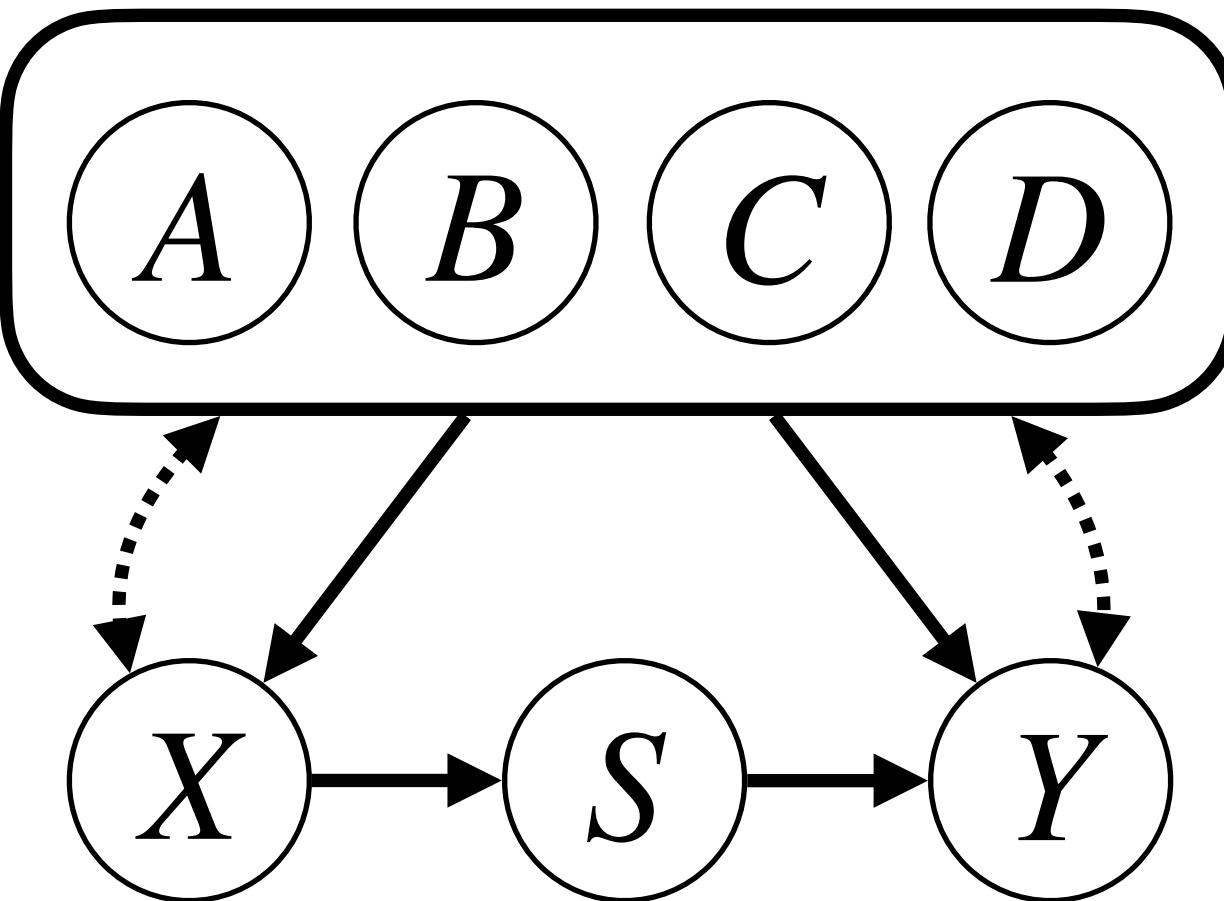


A causal diagram cannot be specified given the existing knowledge!

How can we identify  $P(y | do(x))$  in this case?

# Cluster DAGs (C-DAGs)

- (A) Age
- (B) Blood pressure
- (C) Comorbidities
- (D) Medication history
- (X) Lisinopril
- (S) Sleep Quality
- (Y) Stroke



$\{\{X\}, \{S\}, \{Y\}, \{A, B, C, D\}\}$

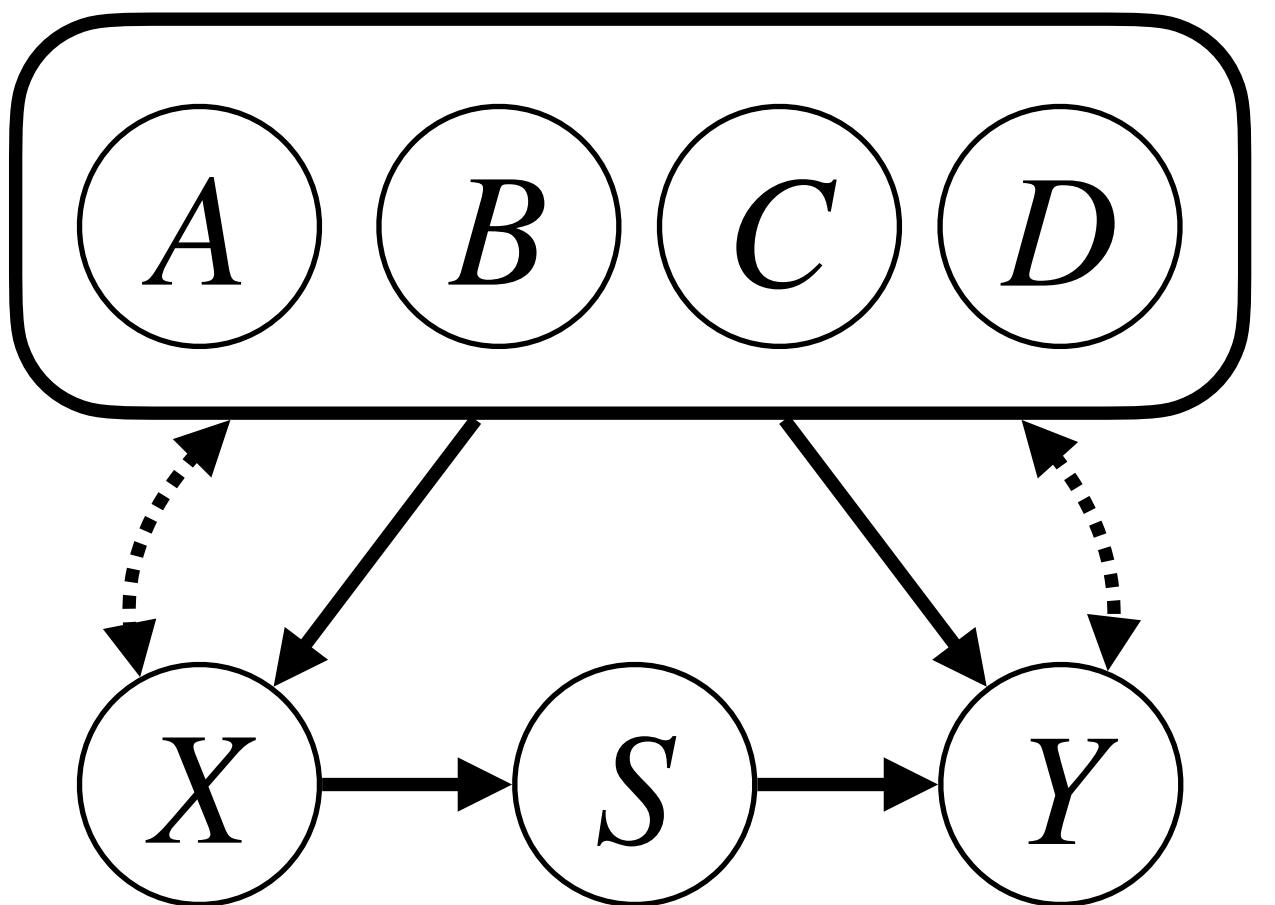
A *cluster DAG*  $G_C$  over a given partition  $\mathbf{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$  of  $\mathbf{V}$  is compatible with a causal diagram  $G$  over  $\mathbf{V}$  if **for every**  $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$ :

- $\mathbf{C}_i \rightarrow \mathbf{C}_j$  if  $\exists V_i \in \mathbf{C}_i$  and  $V_j \in \mathbf{C}_j$  such that  $V_i \rightarrow V_j$
- $\mathbf{C}_i \leftrightarrow \mathbf{C}_j$  if  $\exists V_i \in \mathbf{C}_i$  and  $V_j \in \mathbf{C}_j$  such that  $V_i \leftrightarrow V_j$

and  $G_C$  contains no cycles.

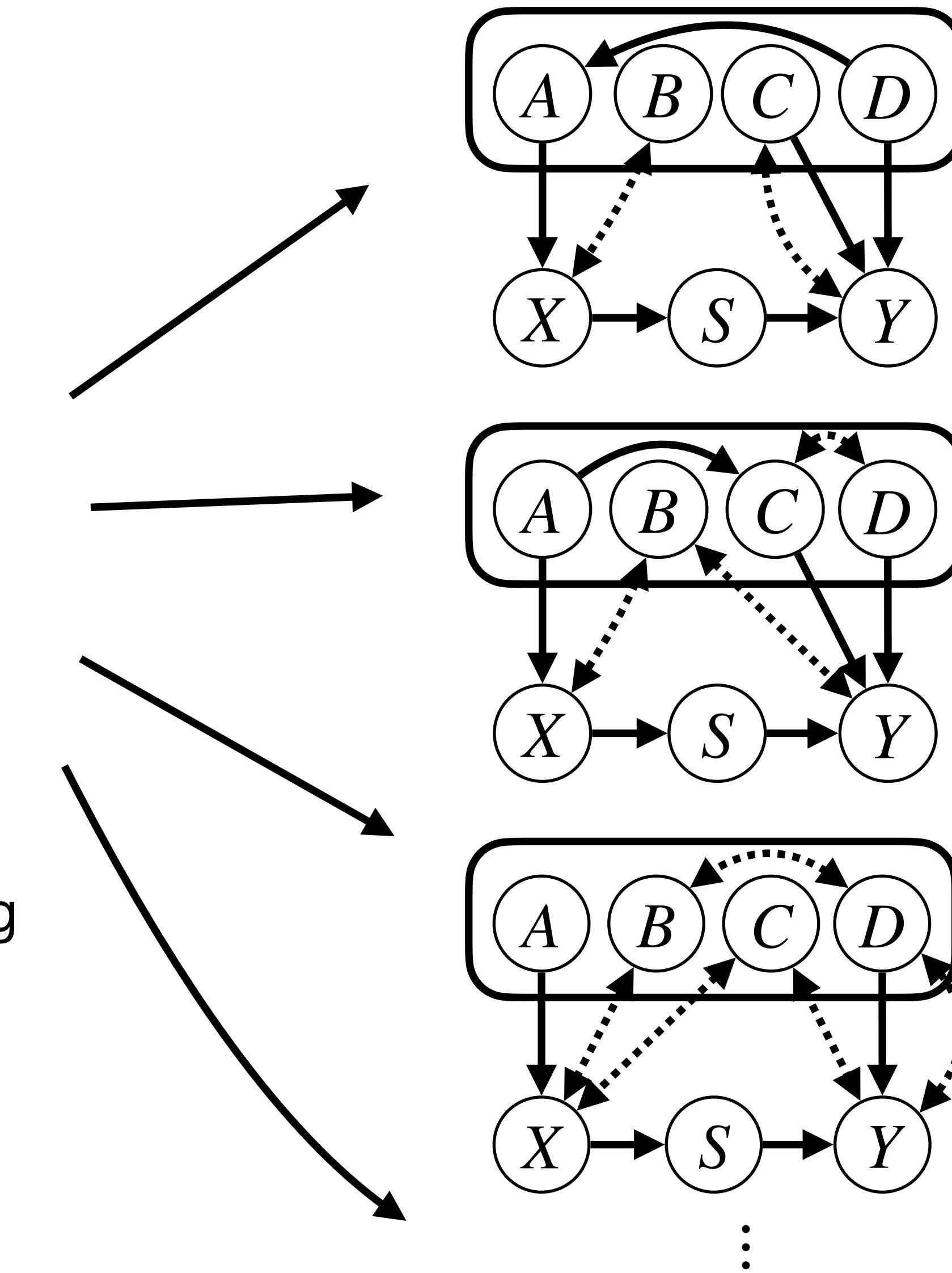
# Partially Understood Systems

Many causal diagrams are compatible with the current knowledge!

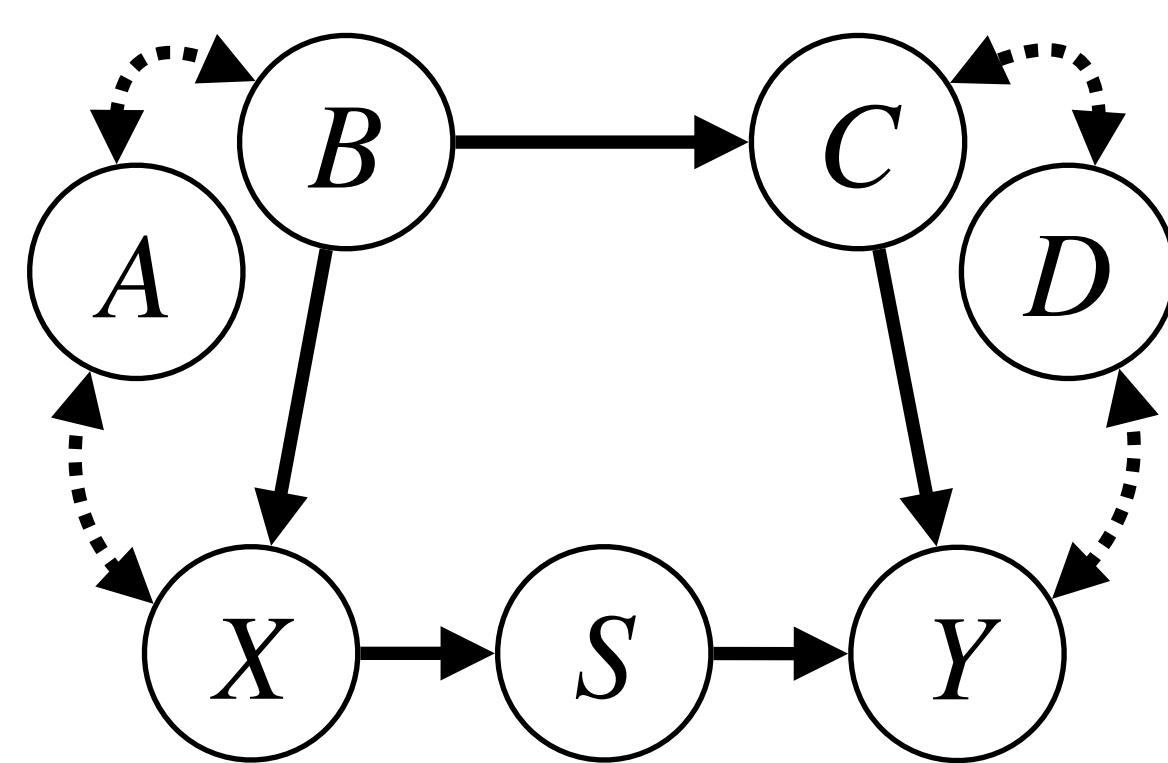


Can be seen as an *equivalence class* of causal diagrams, where any relationships are allowed among the variables within each cluster.

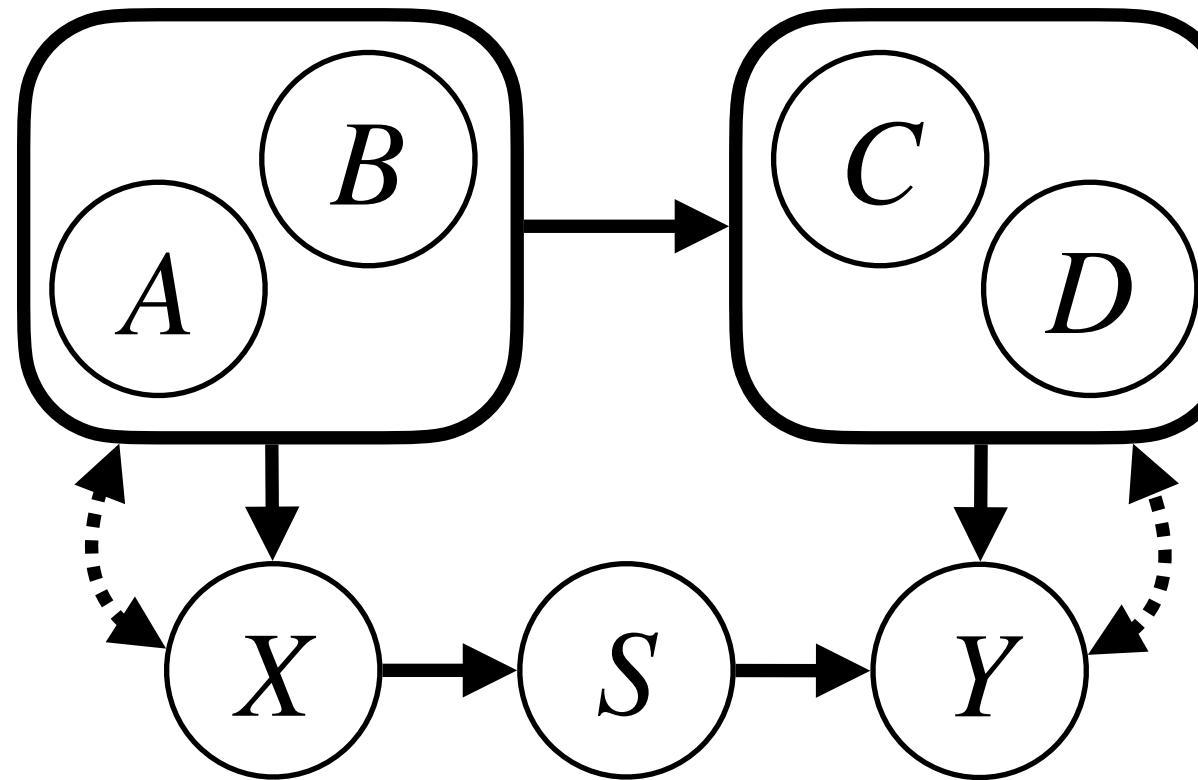
Can we infer causal effects without deciding on any one particular causal diagram?



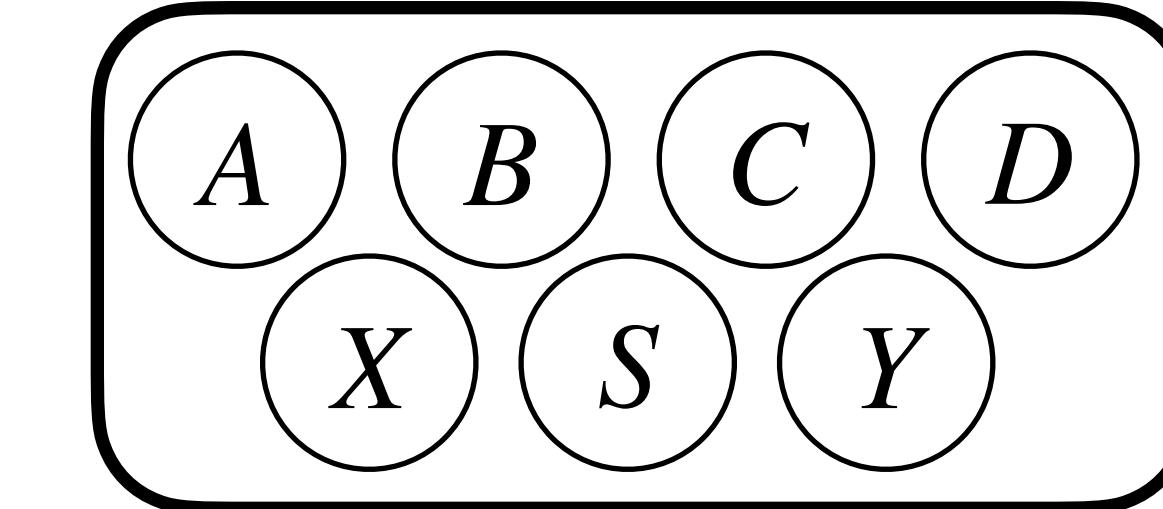
# C-DAG: Flexible Encoder of Model Assumptions



N clusters of size one  
(full knowledge - DAG)



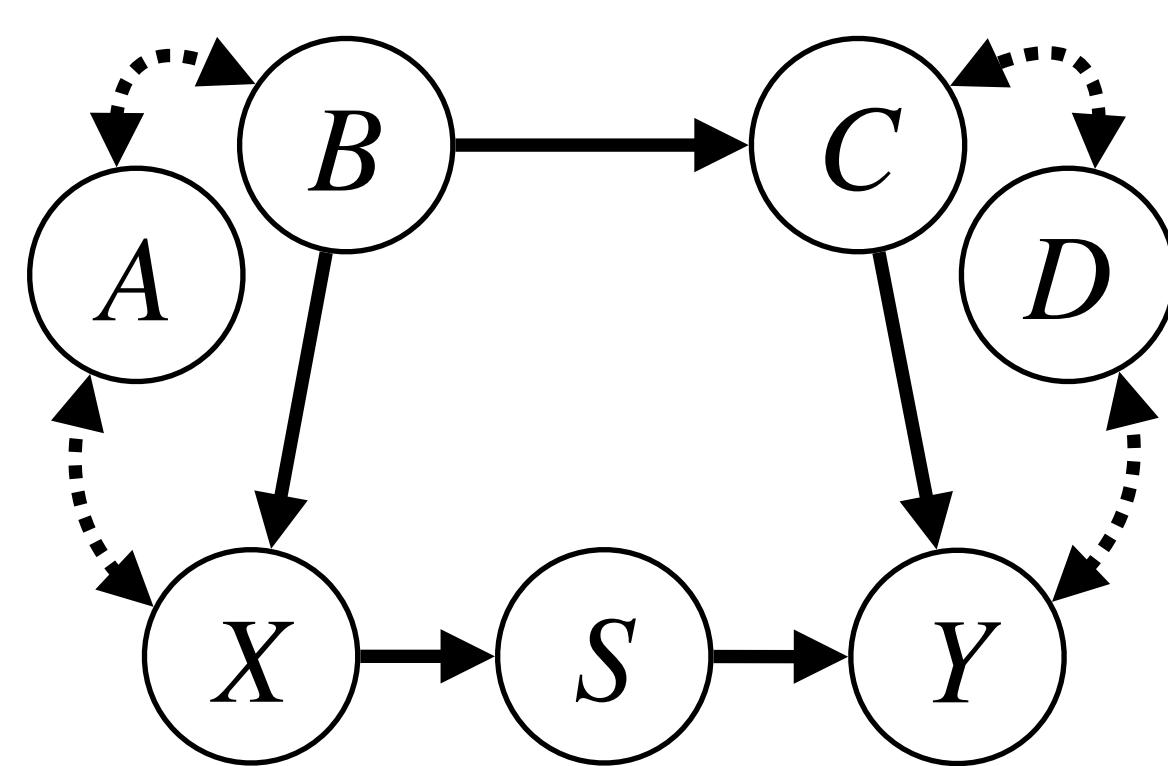
...  
(partial knowledge - C-DAG)



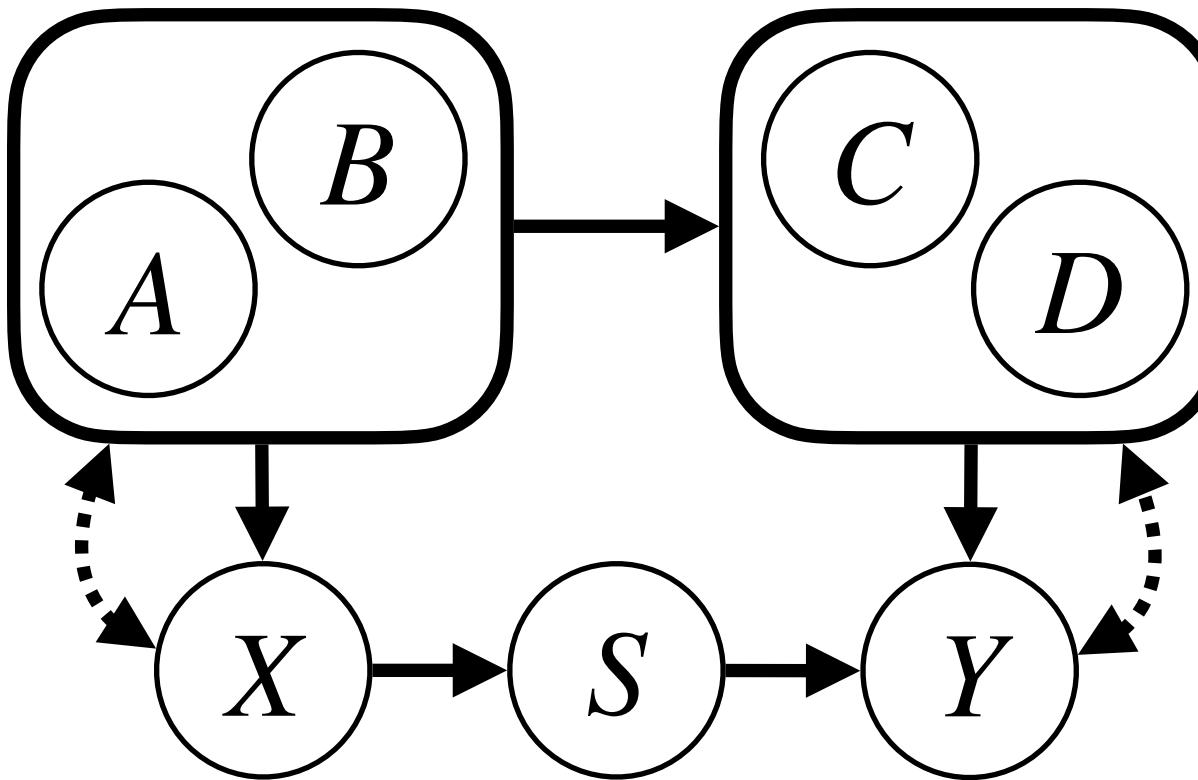
One cluster of size N  
(no knowledge)



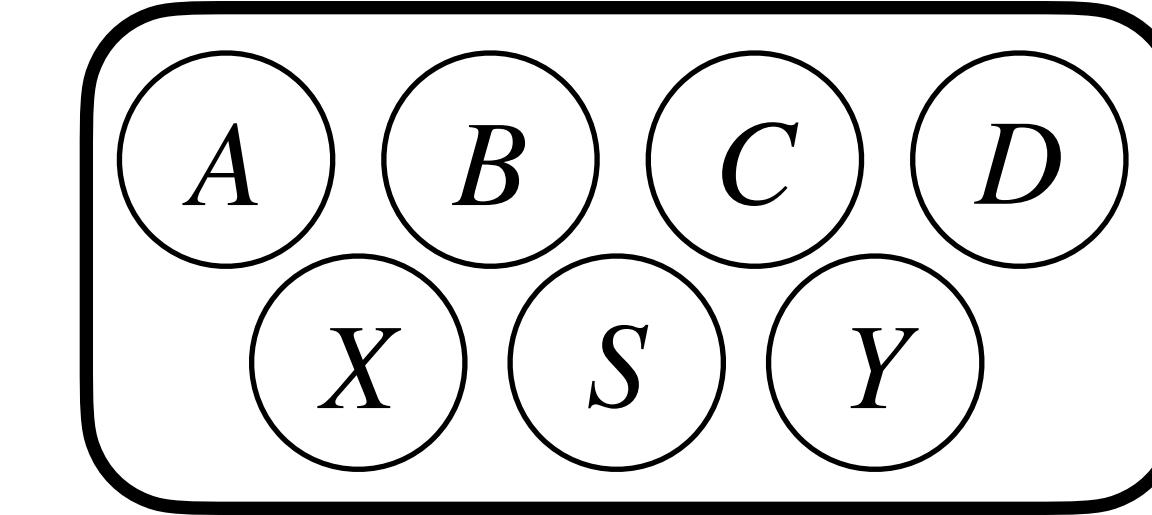
# C-DAG: Flexible Encoder of Model Assumptions



N clusters of size one  
(full knowledge - DAG)



...  
(partial knowledge - C-DAG)

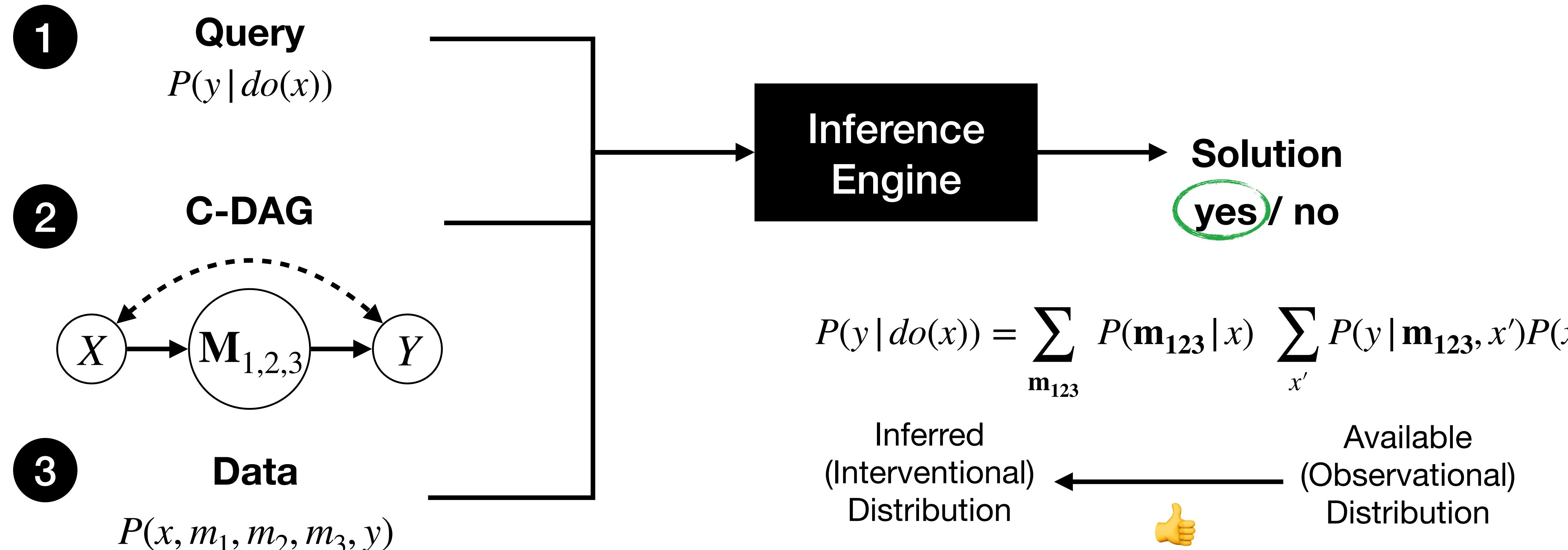


One cluster of size N  
(no knowledge)

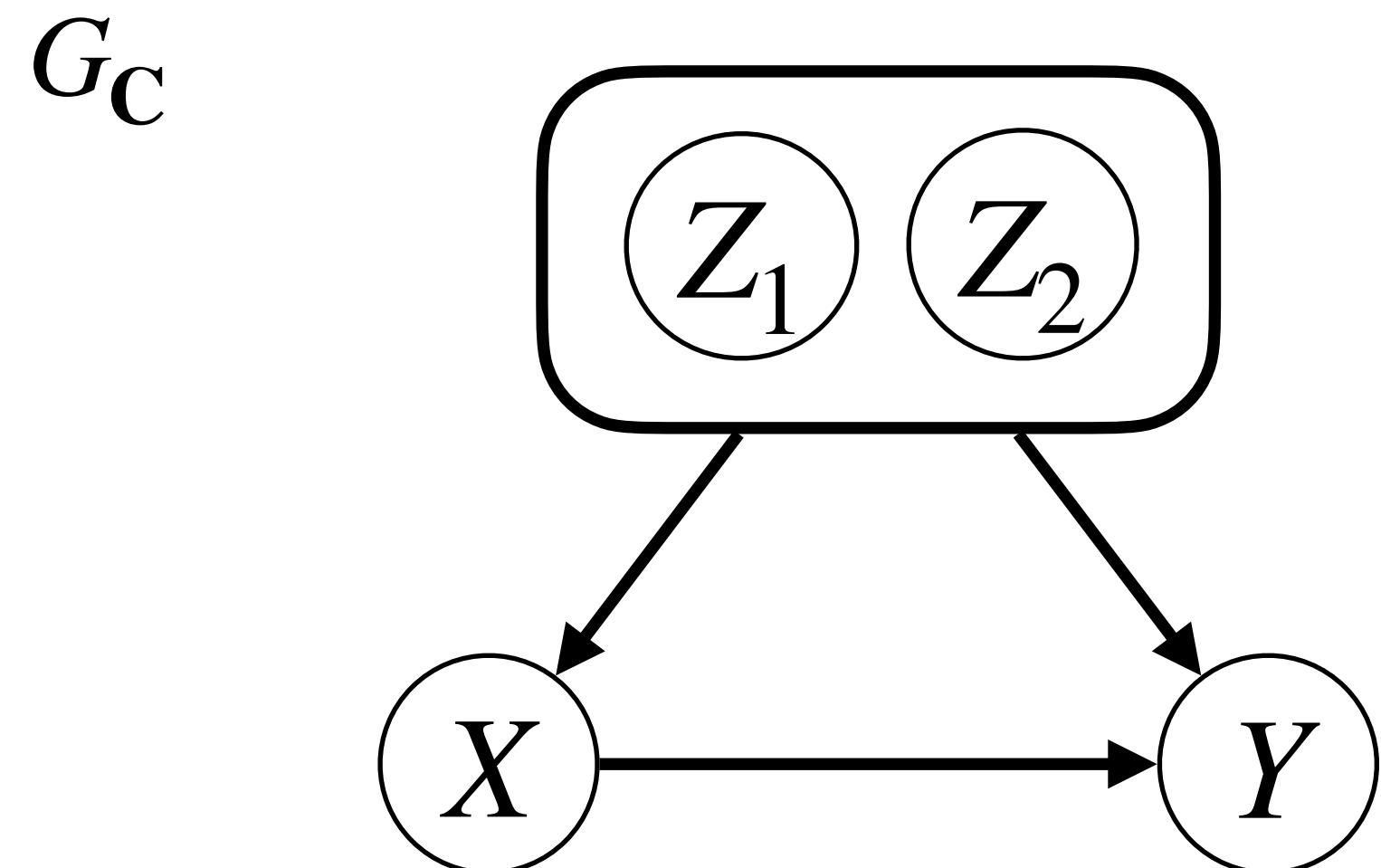
Clusters are manually created by domain experts:

- due to lack of knowledge, consensus, or interest on the internal causal structure;
- to communicate relationships among semantically meaningful entities.

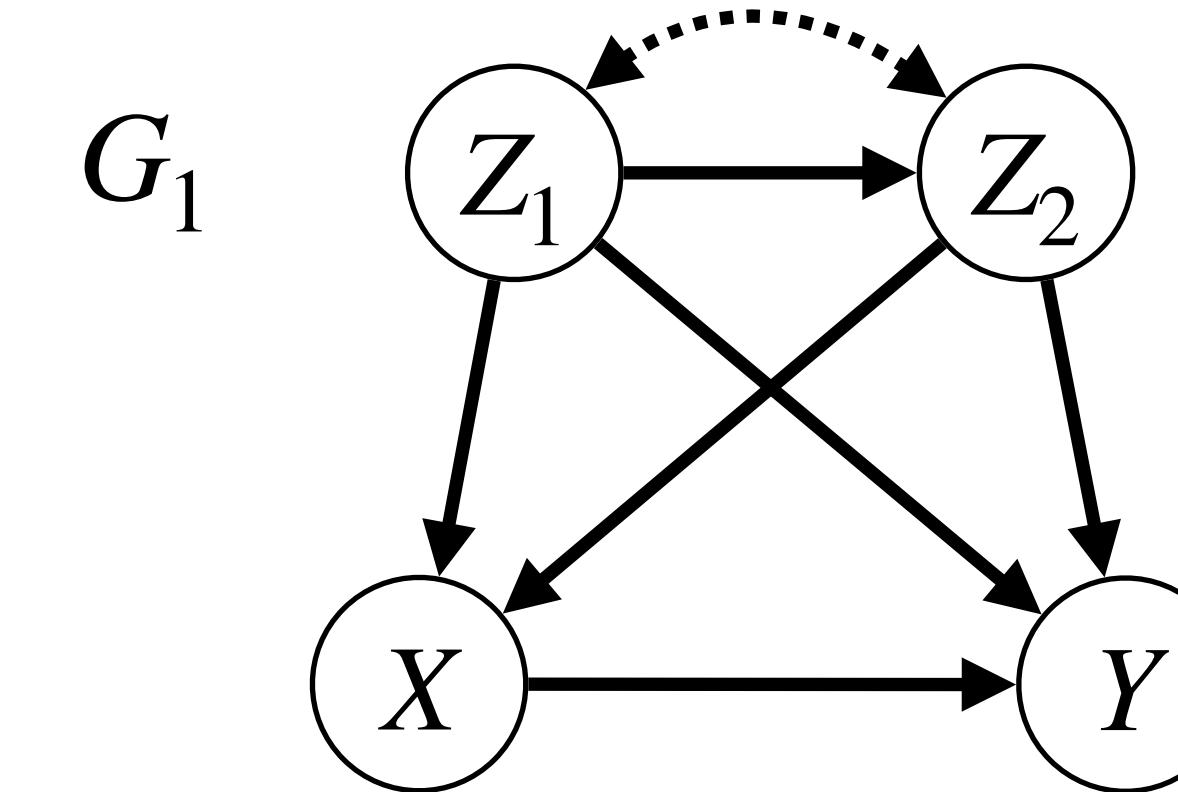
# Identification of Causal Effects from C-DAGs



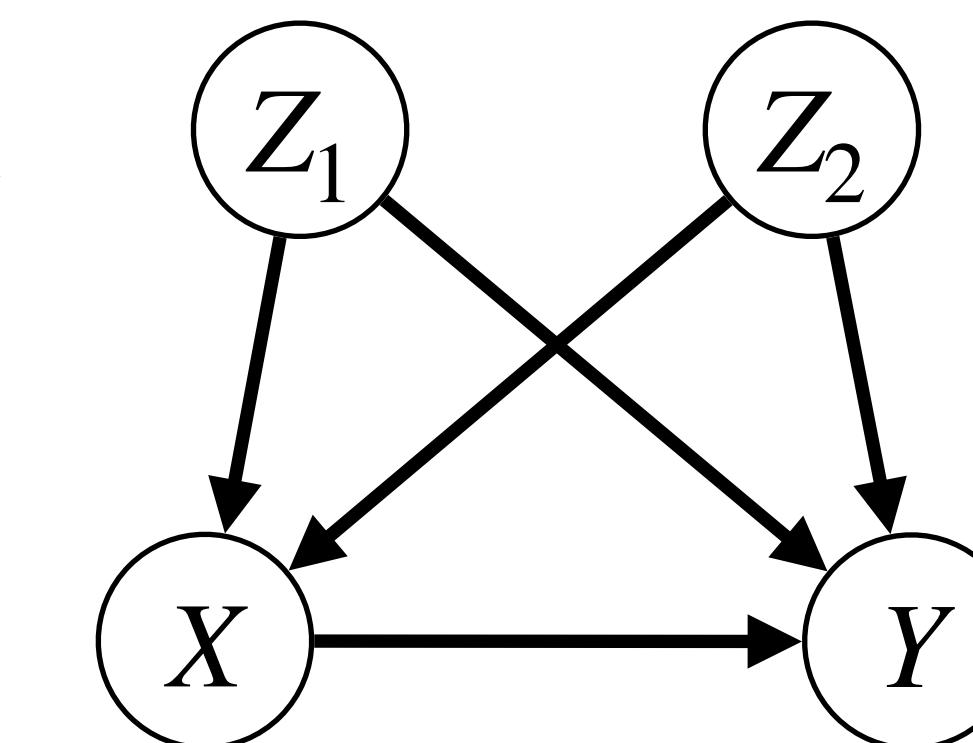
# Effect Identifiability given a C-DAG



$$P(y | do(x)) = \sum_{\mathbf{z}} P(y | x, \mathbf{z}) P(\mathbf{z})$$



$$P(y | do(x)) = \sum_{z_1, z_2} P(y | x, z_1, z_2) P(z_1, z_2)$$

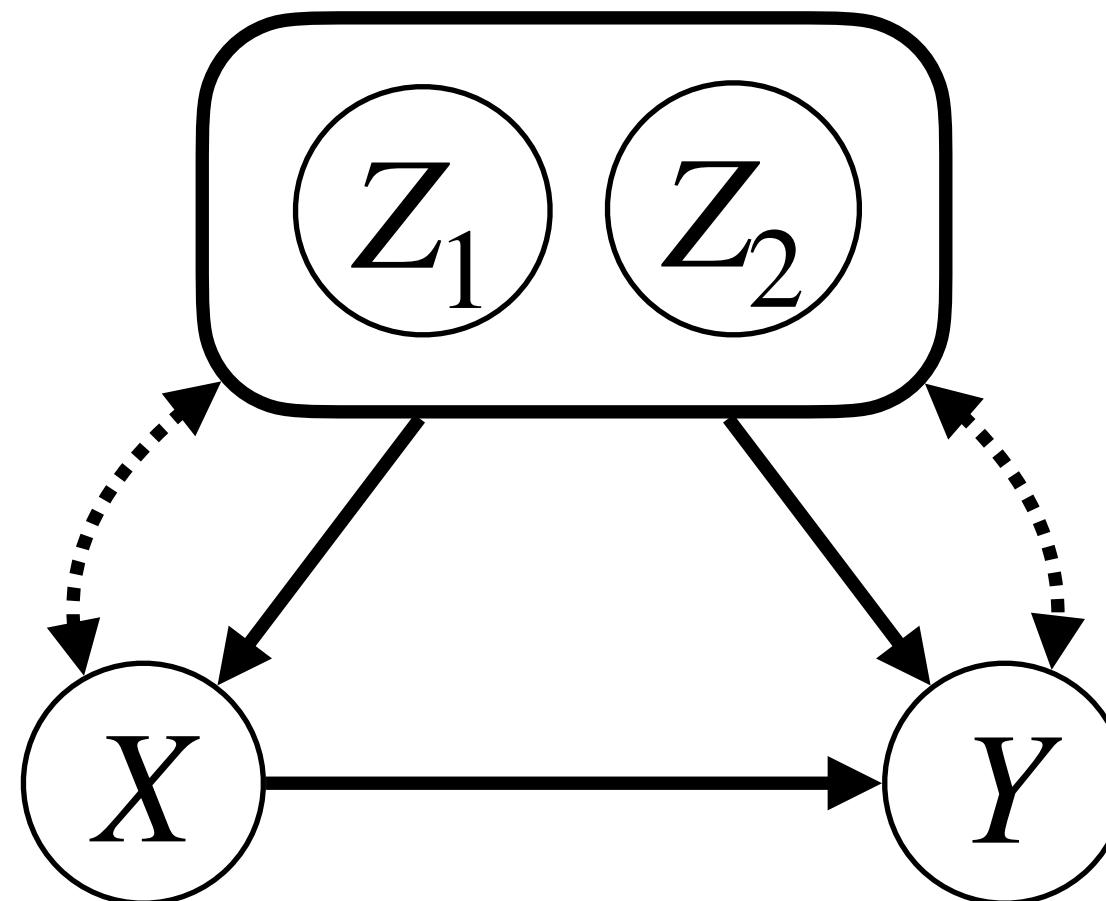


$$P(y | do(x)) = \sum_{z_1, z_2} P(y | x, z_1, z_2) P(z_1, z_2)$$

An effect identifiable in a C-DAG  $G_C$  is identifiable in all compatible causal diagrams  $G$  using the same identification formula!

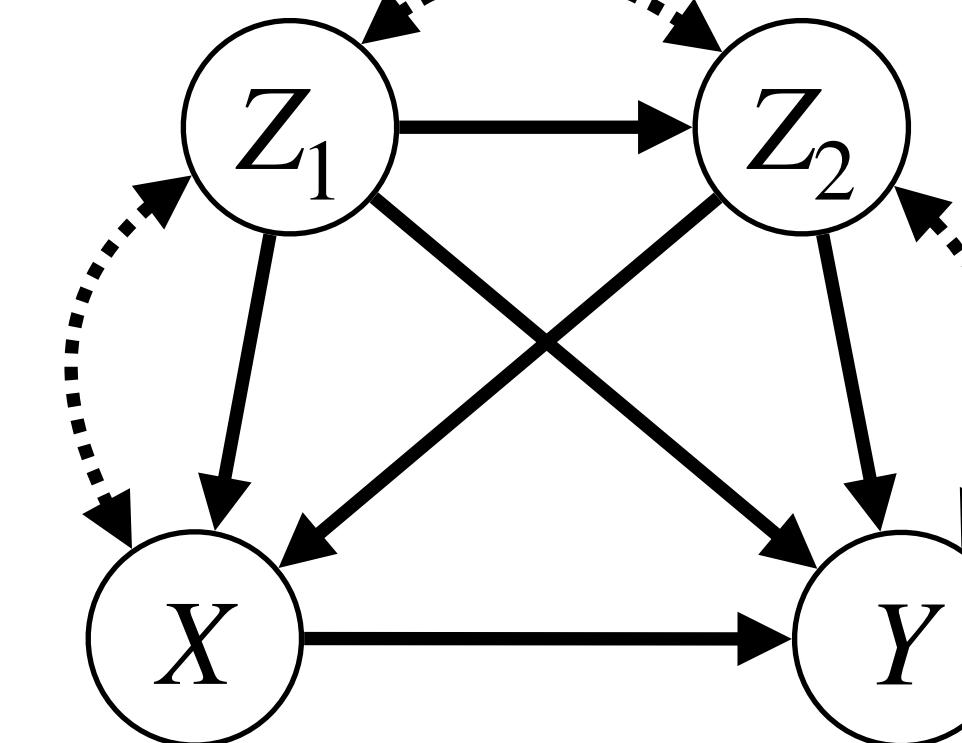
# Effect Non-Identifiability given a C-DAG

$G_C$



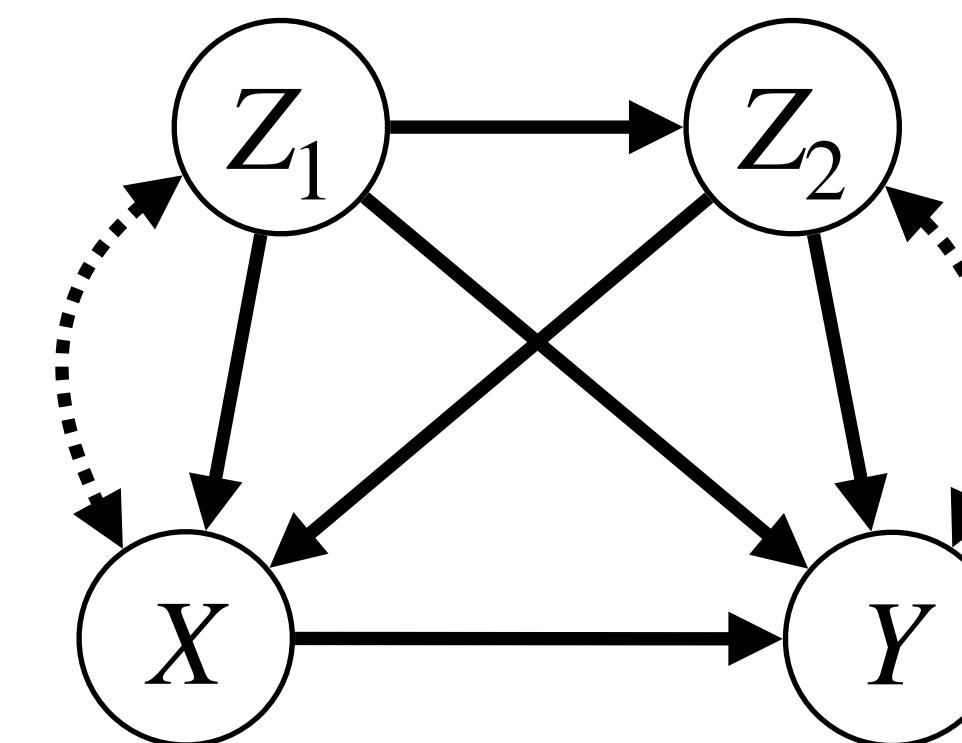
$P(y | do(x))$  is not identifiable

$G_1$



$P(y | do(x))$  is  
not identifiable

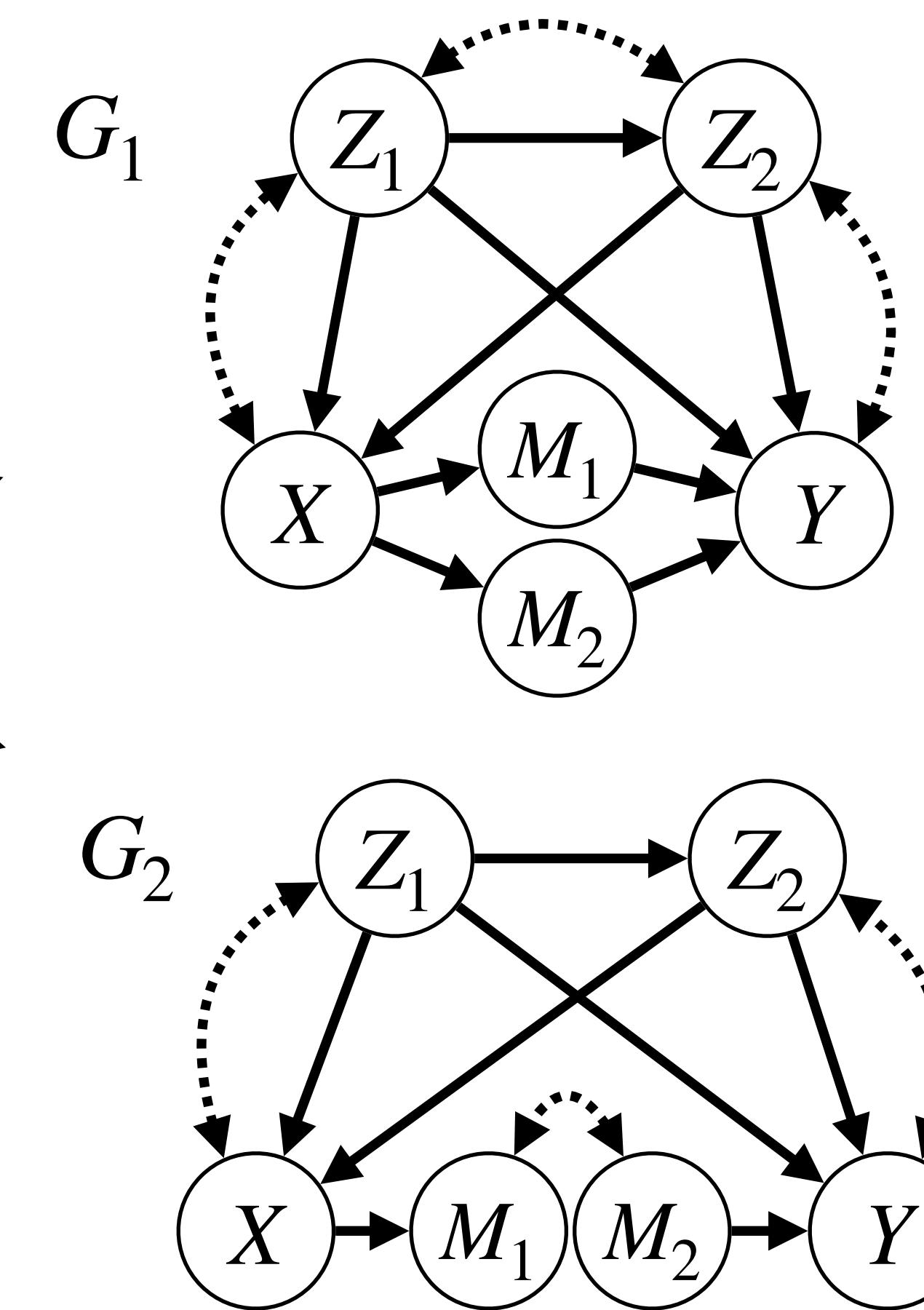
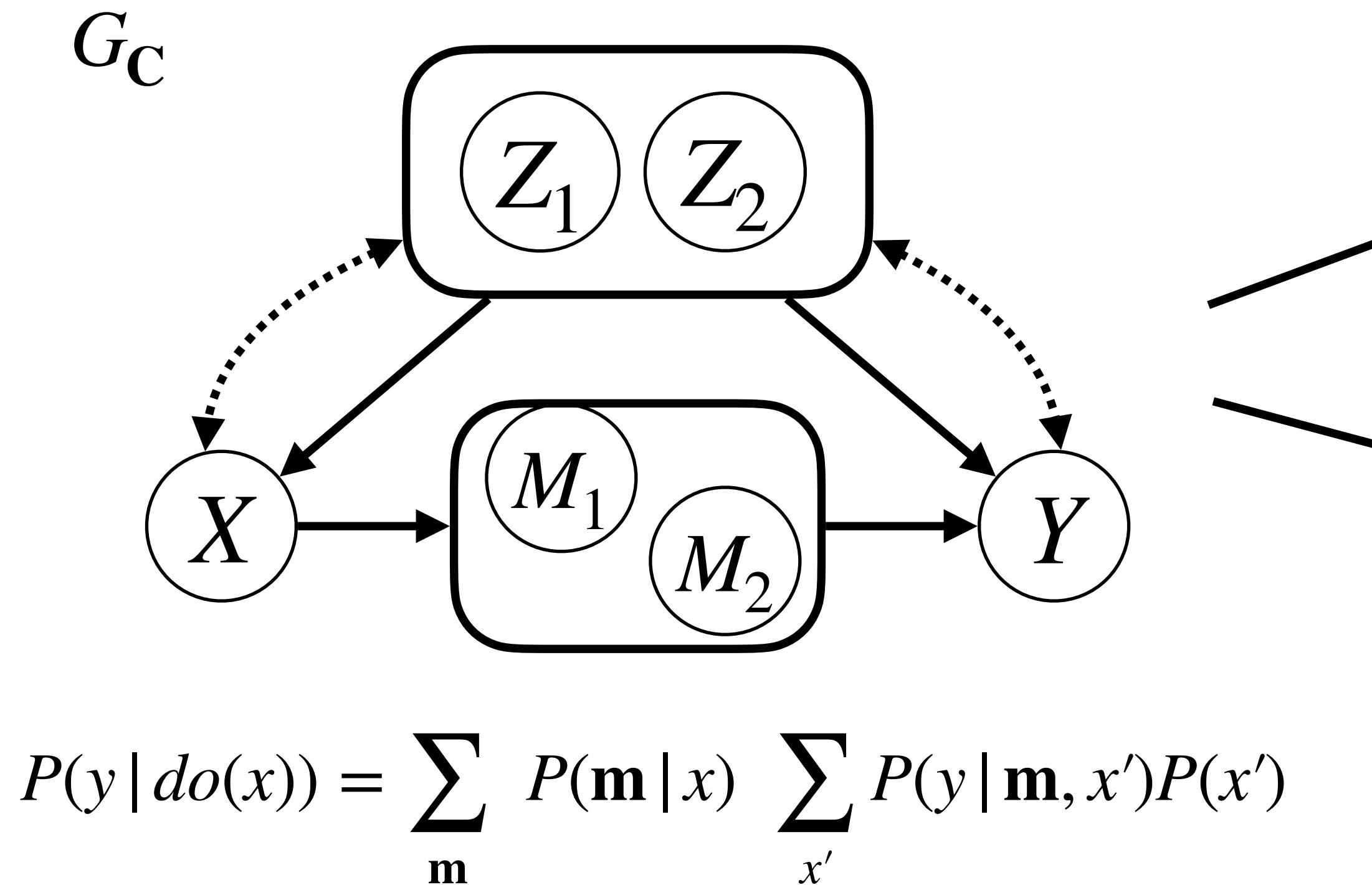
$G_2$



$$P(y | do(x)) = \sum_{z_1, z_2} P(y | x, z_1, z_2)P(z_1, z_2)$$

An effect is not identifiable in a C-DAG  $G_C$  if there exists at least one compatible causal diagrams  $G$  in which the effect is not identifiable.

# Beyond Backdoor Adjustment



$$P(y | do(x)) = \sum_{m_1, m_2} P(m_1, m_2 | x)$$

$$\sum_{x'} P(y | m_1, m_2, x') P(x')$$

$$P(y | do(x)) = \sum_{m_1, m_2} P(m_1, m_2 | x)$$

$$\sum_{x'} P(y | m_1, m_2, x') P(x')$$

Again, an effect identifiable in a C-DAG  $G_C$  is identifiable in all compatible causal diagrams  $G$  using the same identification formula!

# What if no knowledge is available?

---



Can we learn a causal diagram  $\mathcal{G}$  from observational data?

## Causal Discovery:

In non-parametric settings, we can't learn the true causal diagram, but algorithms such as the Fast Causal Inference (FCI) can learn a graphical representation of its *Markov equivalence class*!

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. [Link](#)

# Causal Discovery

---

## Fast Causal Inference (FCI)

A constraint-based causal discovery algorithms that accounts for unobserved confounders

# Causal Discovery

---

**Goal:** Learn a graphical representation of the Markov Equivalence Class from observational data.

**Assumptions:** the observed distribution is the marginal of a distribution  $P$  that satisfies the following conditions for the true causal diagram  $G$  (an **ADMG**):

- 1) **I-Map / Semi-Markov Condition:** for any disjoint subsets  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$ :  
$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z})_G \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z})_P.$$
 $G$  is an ***I-Map of  $P$***   
 $P$  is ***semi-Markov relative*** to  $G$ .
- 2) **Faithfulness Condition:** for any disjoint subsets  $\mathbf{X}$ ,  $\mathbf{Y}$  and  $\mathbf{Z}$ :  
$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z})_P \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z})_G.$$
 $P$  is ***faithful*** to  $G$

**Note:** Estimation of the marginal distribution from limited data requires and **additional assumption**:

- 3) An adequate *conditional independence test* is available.

# Conditional Independence Tests

---

Gaussian errors and independent observations: partial correlation test

Fisher, R.A. (1921). *On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample.*  
R package: <https://cran.r-project.org/web/packages/pcalg/>

Kernel-based non-parametric test:

Zhang, K., Peters, J., Janzing, D., & Schölkopf, B. (2012). *Kernel-based conditional independence test and application in causal discovery.* In: Uncertainty in artificial intelligence. AUAI Press; 2011. p.804–13  
R package: <https://cran.r-project.org/web/packages/CondIndTests>

Continuous (conditional Gaussian) or Discrete (Binary, Ordinal, Multinomial) - Linear Regression

- Tsagris, M., Borboudakis, G., Lagani, V. et al. (2018) Constraint-based causal discovery with mixed data. *Int J Data Sci Anal* 6, 19–30. ([Link](#))
- R package: <https://cran.r-project.org/web/packages/MXM/>

Gaussian errors and correlated observations (family data) :

**Ribeiro A.H.**, Soler J.M.P. (2020). *Learning Genetic and environmental graphical models from family data,* Statistics in Medicine.  
R package: <https://github.com/adele/FamilyBasedPGMs>

# Learning Structural Invariances

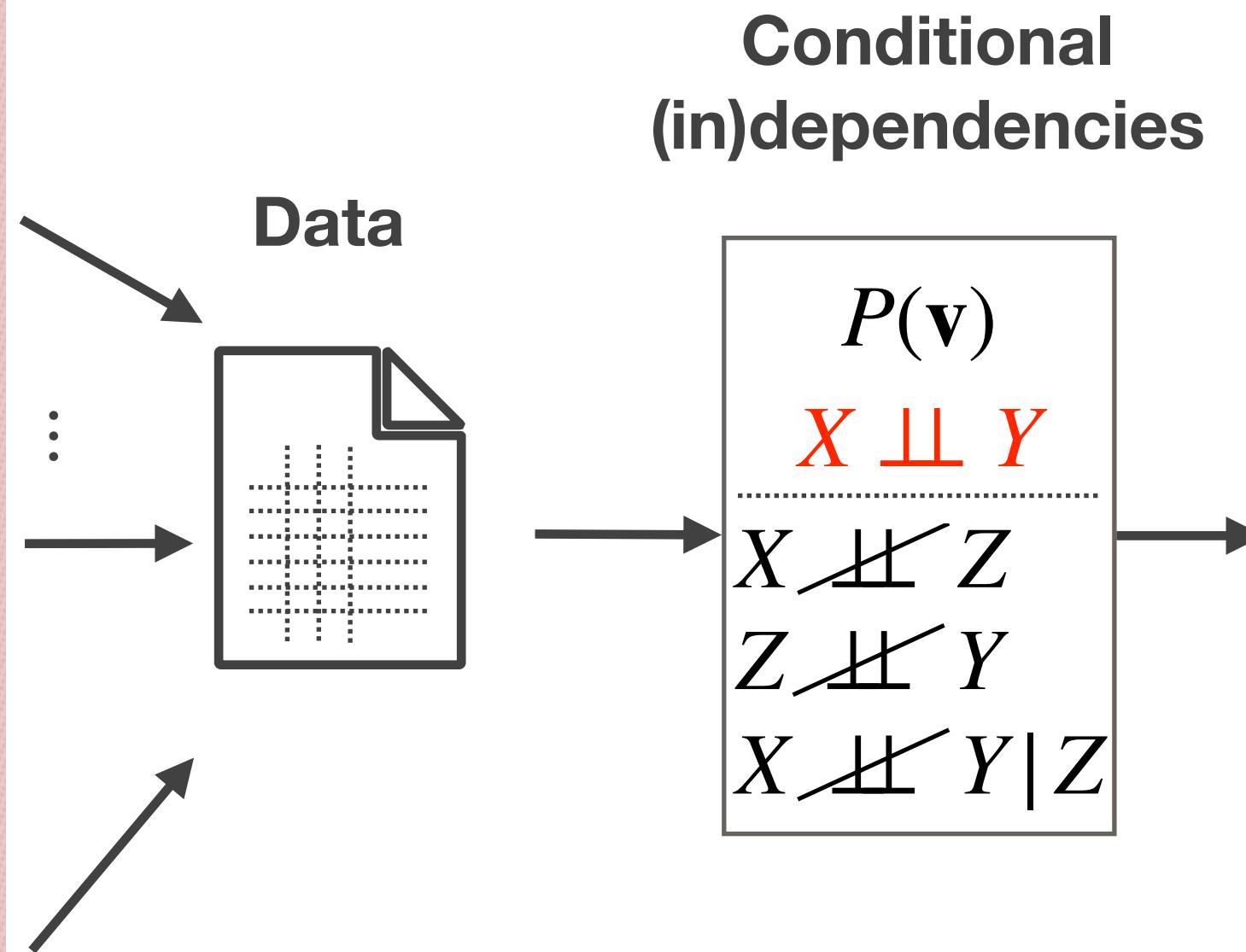
$$\mathcal{M}_1 = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_x, U_y, U_z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_x) \\ Z \leftarrow f_Z(X, Y, U_z) \\ Y \leftarrow f_Y(U_y) \end{cases} \\ P(U) \end{cases}$$

⋮

$$\mathcal{M}_{N-1} = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_{xz}, U_{yz}, U_x, U_y, U_z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{xz}, U_x) \\ Z \leftarrow f_Z(Y, U_{xz}, U_z) \\ Y \leftarrow f_Y(U_y) \end{cases} \\ P(U) \end{cases}$$

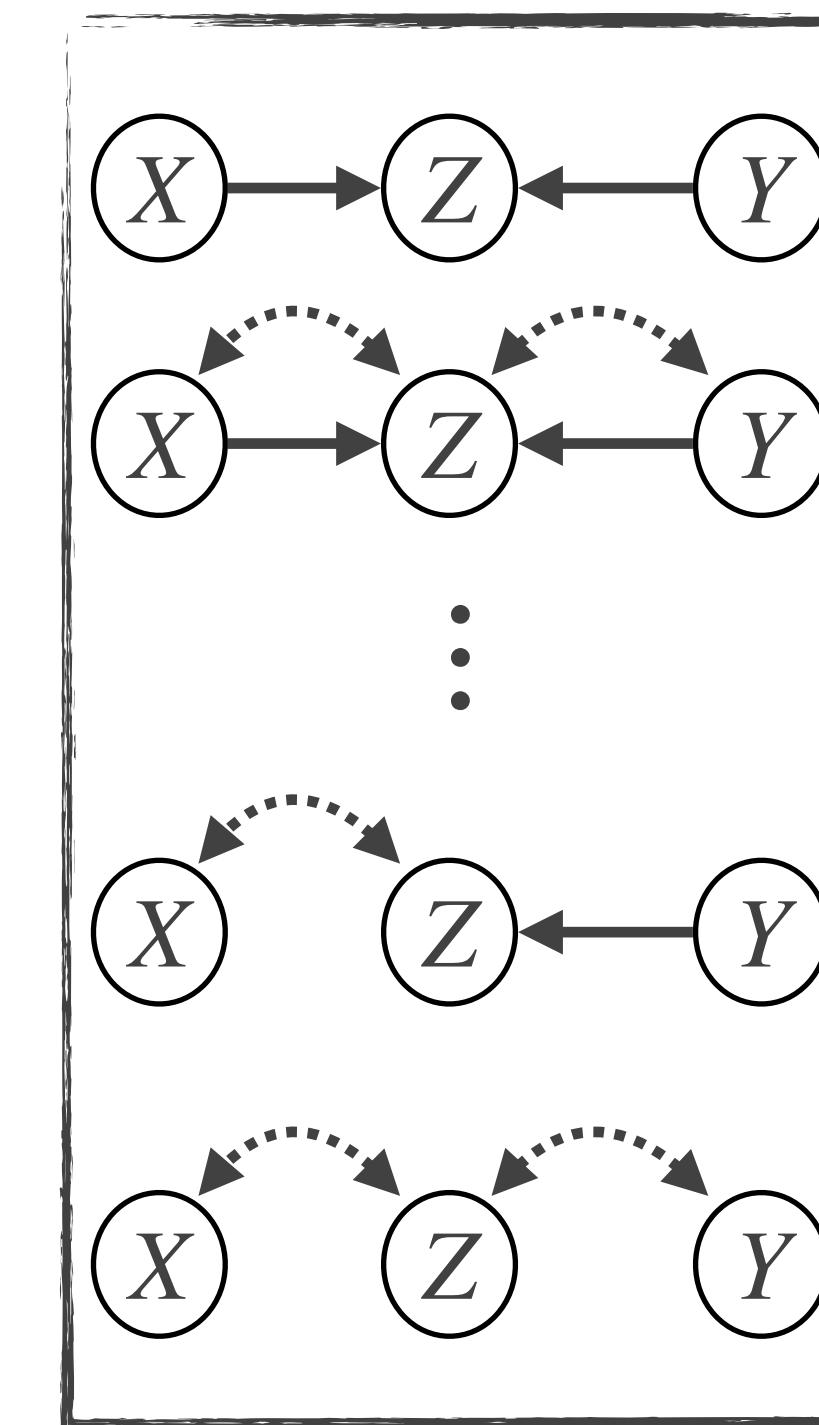
⋮

$$\mathcal{M}_N = \begin{cases} V = \{X, Y, Z\} \\ U = \{U_{xz}, U_{yz}, U_x, U_y, U_z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{xz}, U_x) \\ Z \leftarrow f_Z(U_{xz}, U_{yz}, U_z) \\ Y \leftarrow f_Y(U_{yz}, U_y) \end{cases} \\ P(U) \end{cases}$$

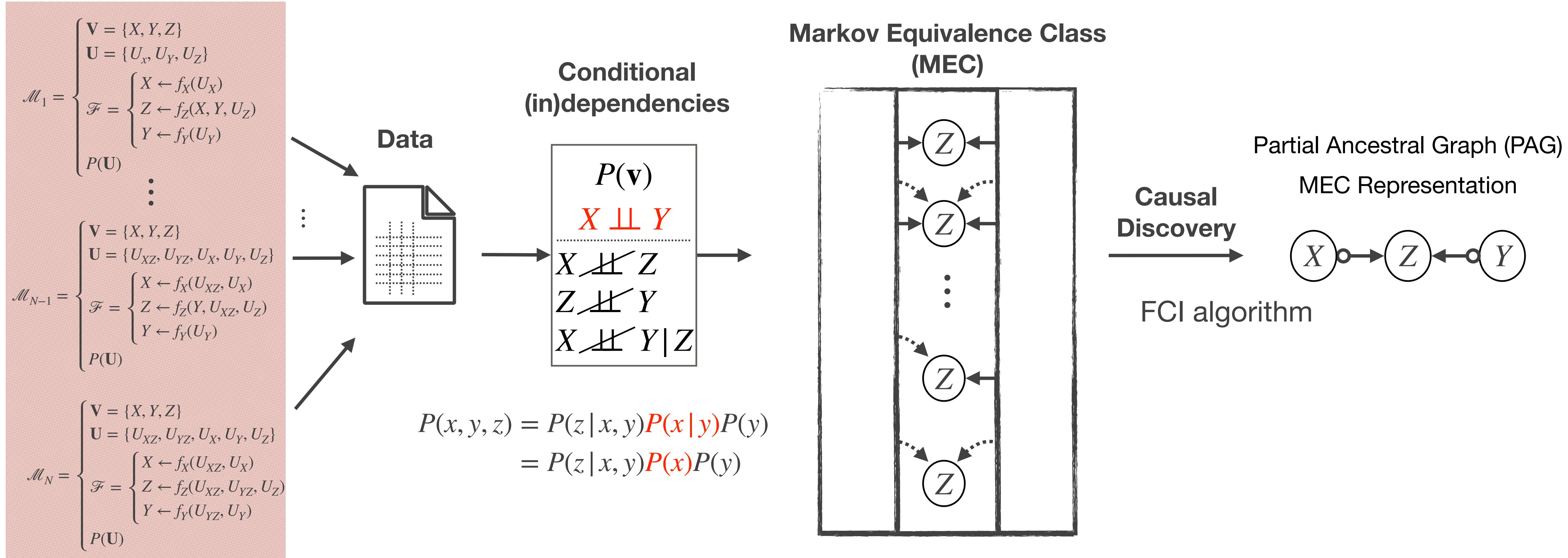


$$P(x, y, z) = P(z|x, y) \color{red}{P(x|y)} P(y) \\ = P(z|x, y) \color{red}{P(x)} P(y)$$

**Markov Equivalence Class (MEC)**

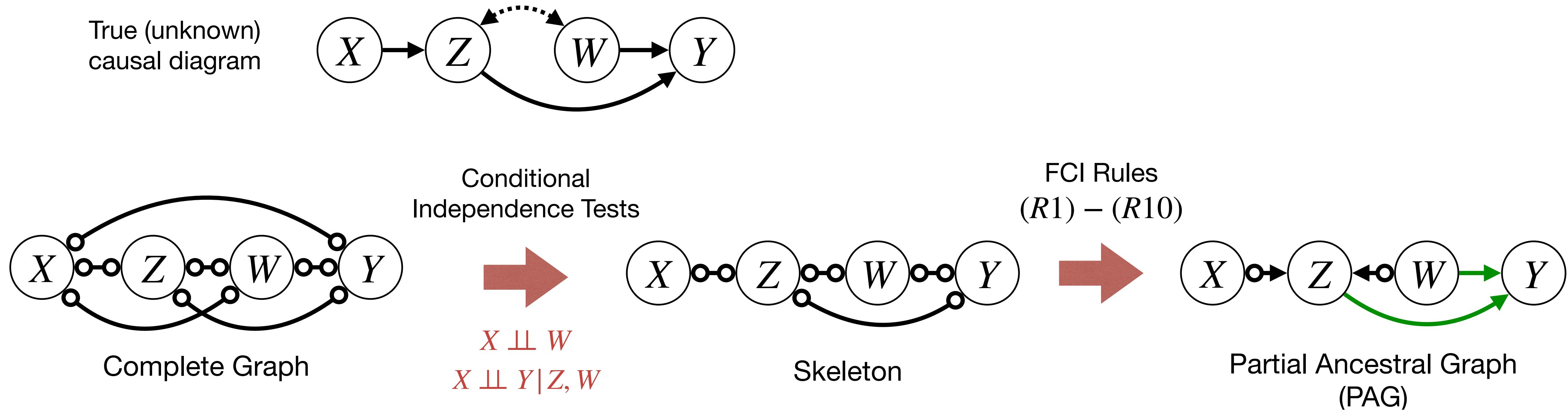


# Learning Structural Invariances



Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. [Link](#)

# Fast Causal Inference (FCI) Algorithm



Arrowhead  $\implies$  non-ancestrality  
Tail  $\implies$  ancestrally  
Circle  $\implies$  non-invariance

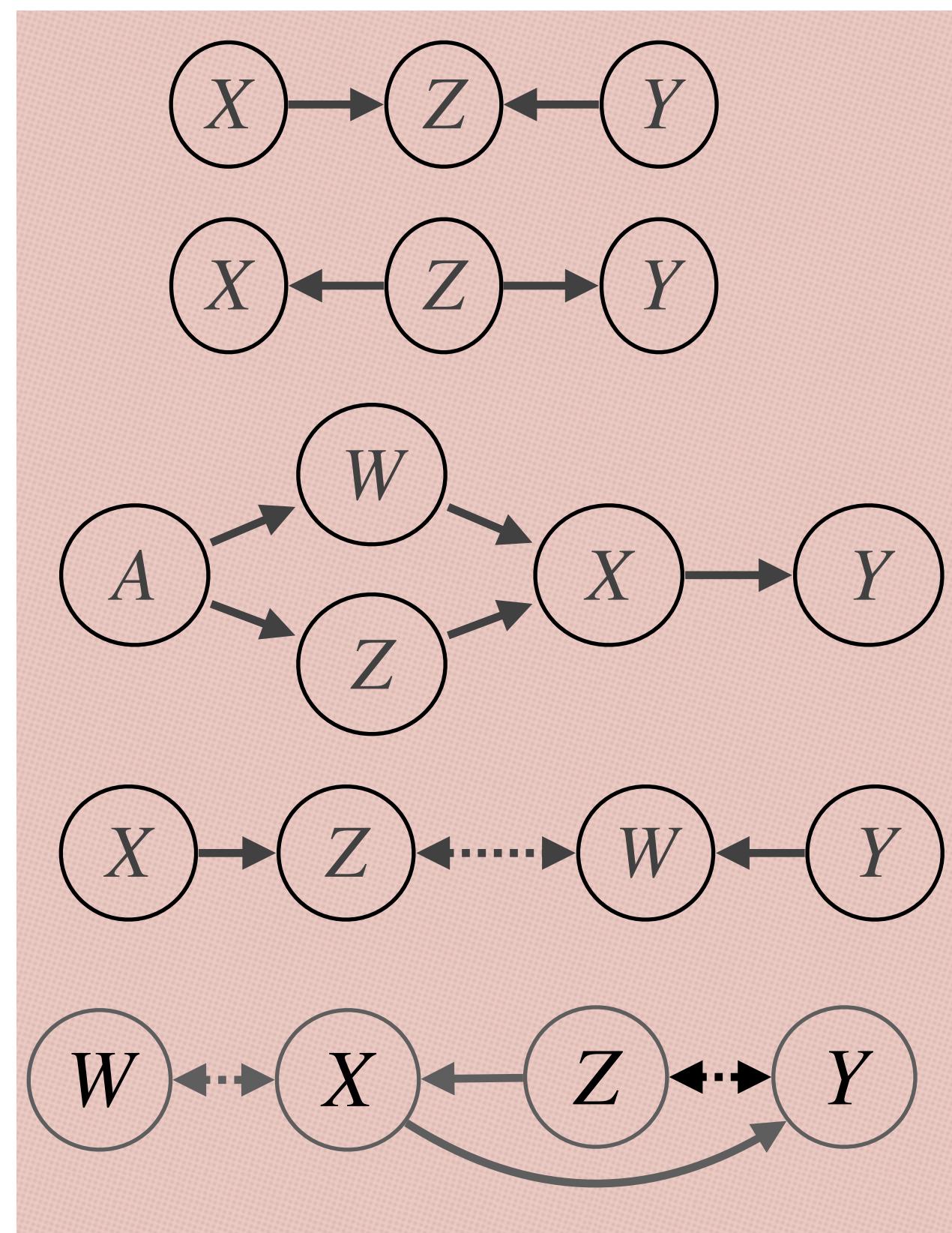
$A \longleftrightarrow B$  – spurious association  
 $A \dashv\! \dashv B$  – selection bias

Z is not an ancestor of X or W.  
Z and W are ancestors (and definite causes) of Y.

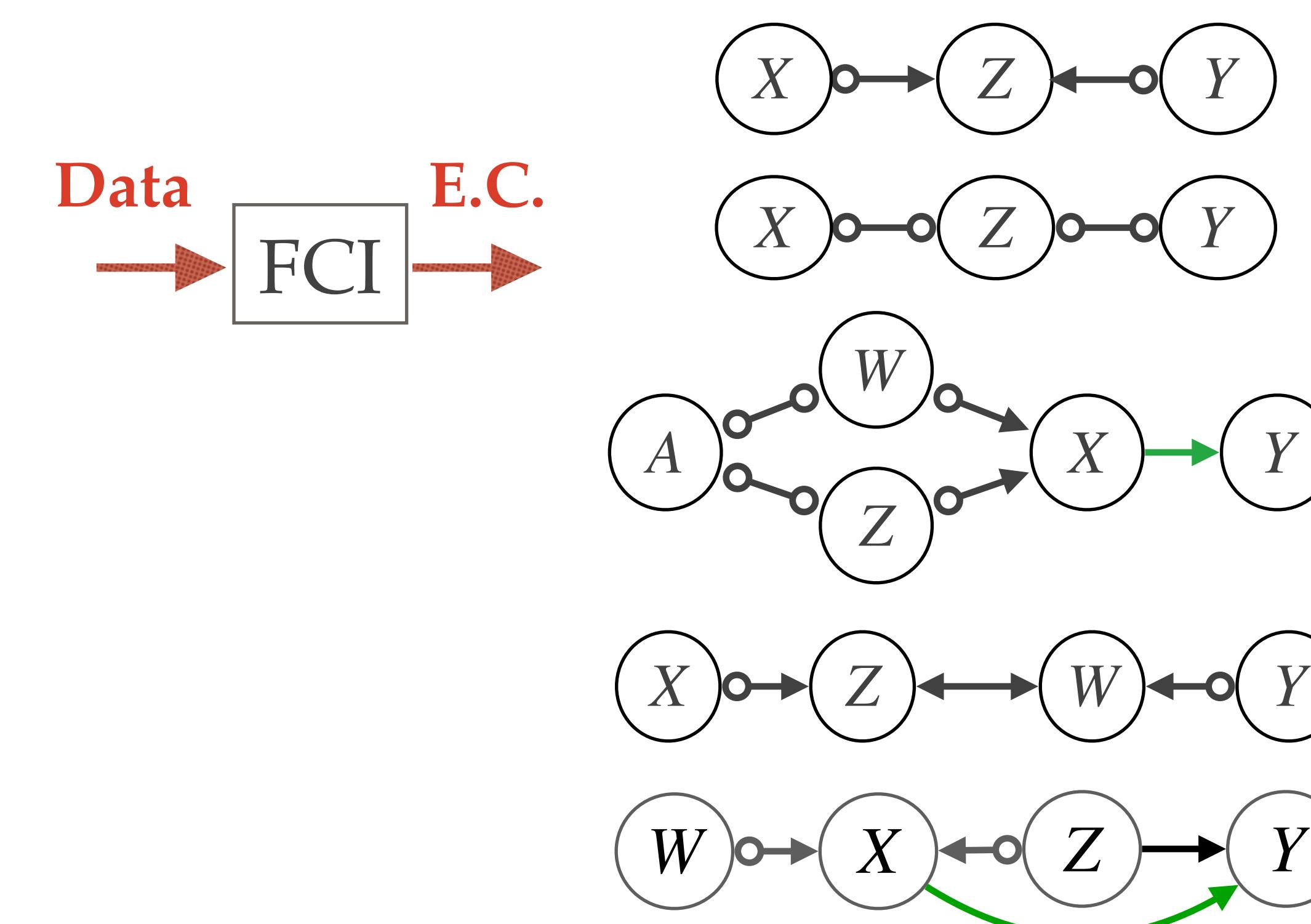
# Causal Structure Learning

Structure Learning Algorithms (e.g. PC/IC, FCI, etc) learn a representation of the equivalence class that contains the true underlying causal diagram:

**Underlying Causal Diagram**



**Partial Ancestral Graph**



# Causal Identification from PAGs

---



Can we identify causal effects from the equivalence class?

## Effect Identification:

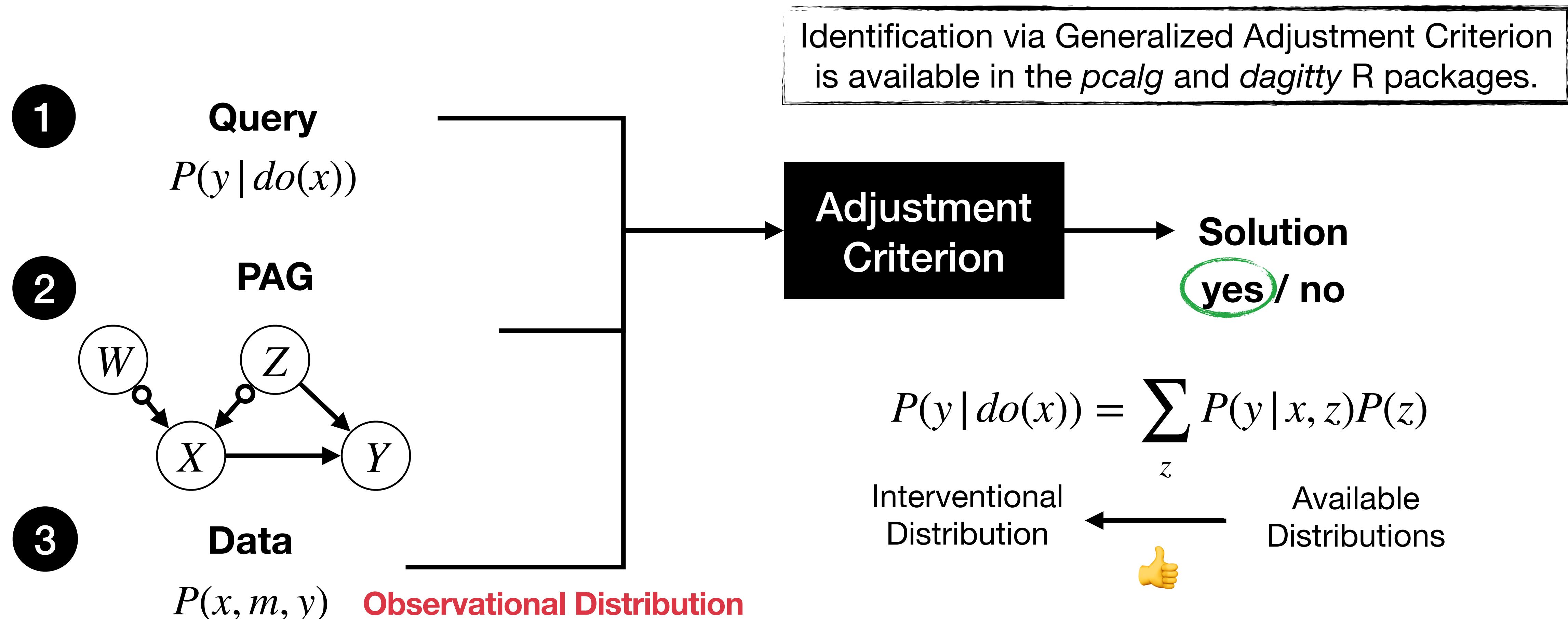
For Covariate Adjustment, we can use the Generalized Adjustment Criterion.

Recently, we proposed complete calculus and algorithms for the identification of marginal and conditional causal effect in PAGs!

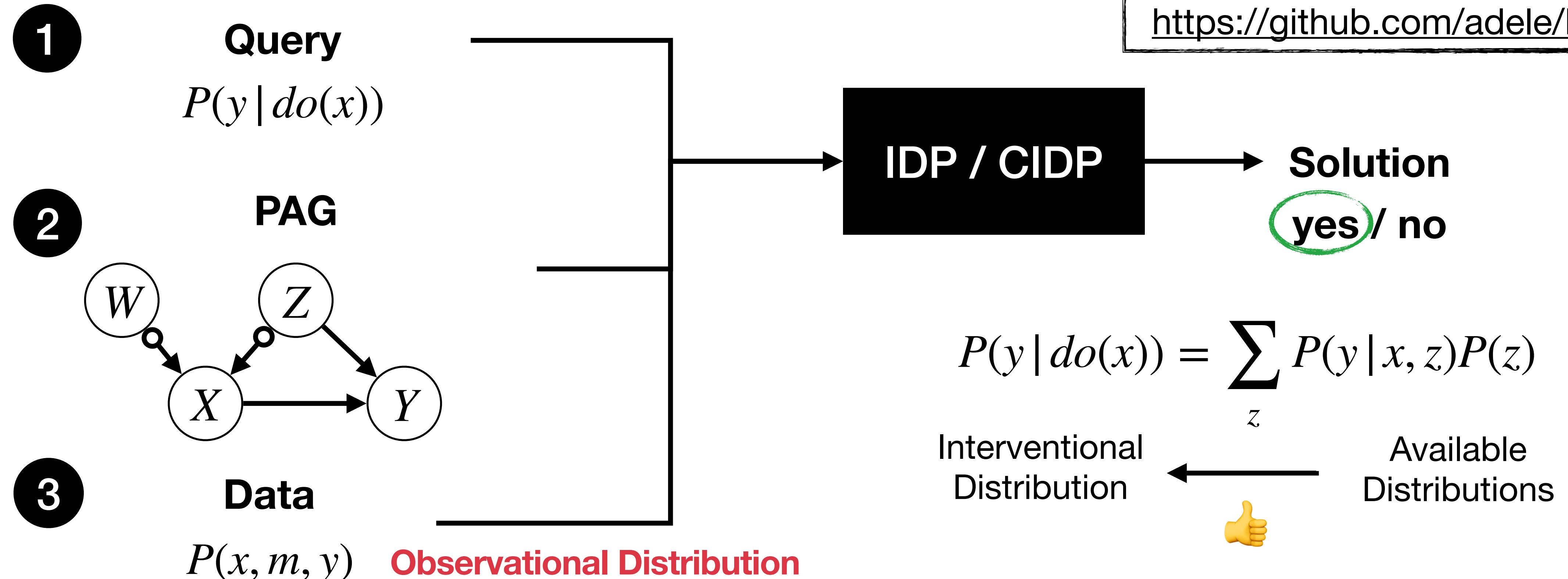
Perkovic, E., Textor, J. C., Kalisch, M., & Maathuis, M. H. (2018). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. Journal of Machine Learning Research 18 (2018) 1-62

Jaber A., **Ribeiro A. H.**, Zhang, J., Bareinboim, E. (2022) Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS. ([Link](#))

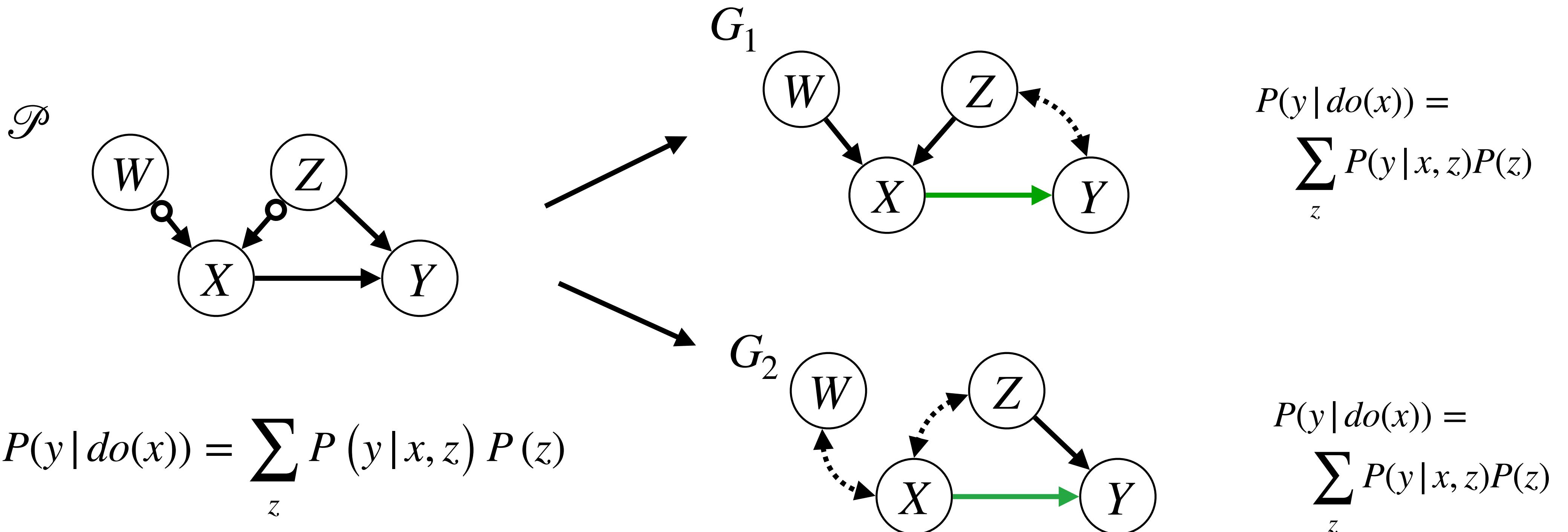
# Adjustment in Markov Equivalence Classes



# General Identification in Markov Equivalence Classes

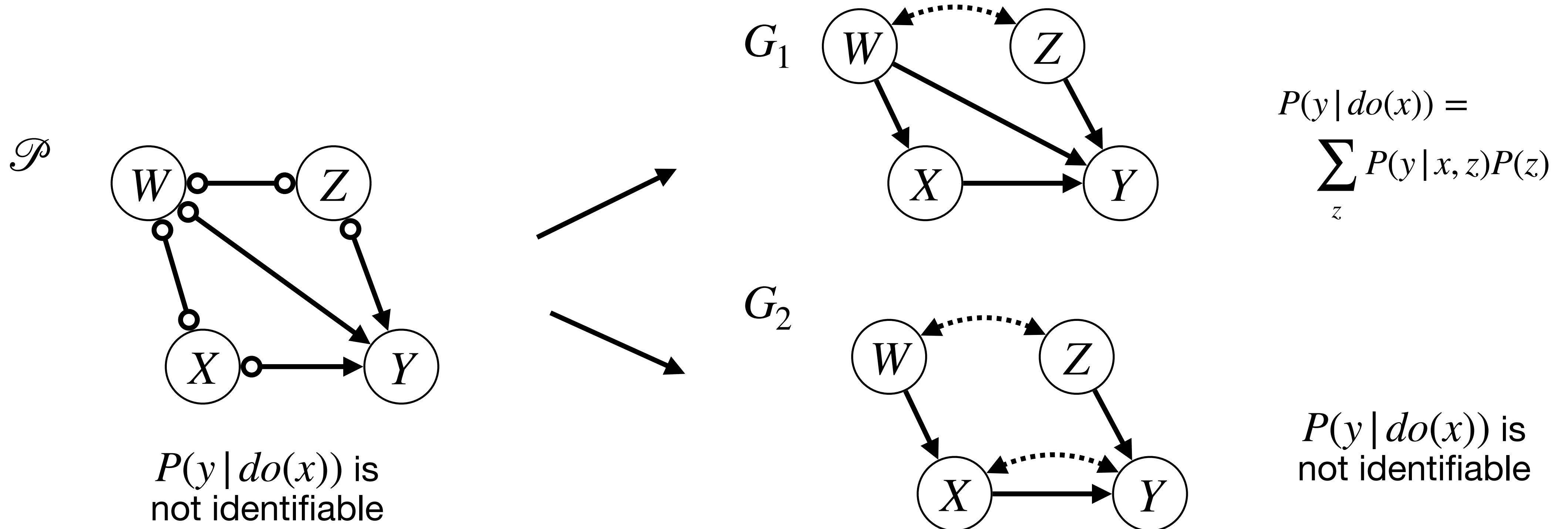


# Effect Identifiability given a PAG



An effect identifiable in a PAG  $\mathcal{P}$  is identifiable in all causal diagrams  $G$  in the Markov Equivalence Class using the same identification formula!

# Effect Non-Identifiability given a PAG



An effect not identifiable in a PAG  $\mathcal{P}$  is not identifiable in at least one causal diagrams  $G$  in the Markov Equivalence Class

# Conclusions

---

Causal inference can help overcome critical challenges in Artificial Intelligence, including robustness, generalizability, explainability, and fairness.

Causal Data Science: principled way of combining data and substantive knowledge about the phenomenon under investigation to generate causal explanations and better decision-making.

Recent developments for causal inference when knowledge is largely unavailable and coarse are expected to help the practice of causal data analysis and meet the growing demand in the Empirical Sciences for sound causal explanations and more robust and generalizable decision-making.

# Code Examples

---

**Prob-AI GitHub (Day 5):** [https://github.com/probabilisticai/probai-2023/tree/main/day\\_5](https://github.com/probabilisticai/probai-2023/tree/main/day_5)

- **Example 1:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_BD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_BD.ipynb)
- **Example 2:** [https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day\\_5/1\\_adele/causal\\_NonBD.ipynb](https://colab.research.google.com/github/probabilisticai/probai-2023/blob/main/day_5/1_adele/causal_NonBD.ipynb)

**Check Items from 6 to 8 of both Examples:**

6. Causal Discovery: Estimating a PAG using **FCI**
7. Checking the conditional independence relations implied by the true Causal Diagram
8. Causal Effect Identification from the PAG

# **Thank you! :)**

Feel free to reach out to me if you have any questions:

[adele.ribeiro@uni-marburg.de](mailto:adele.ribeiro@uni-marburg.de)