

# Introduction to Causal Inference

Adèle H. Ribeiro

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

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

**Nordic Probabilistic AI School**  
**June 17-21, 2024**  
**Copenhagen, Denmark**

# Predictive vs Explainable, Trustworthy AI

## NEWSLETTERS

Sign up to read our regular email newsletters

# NewScientist

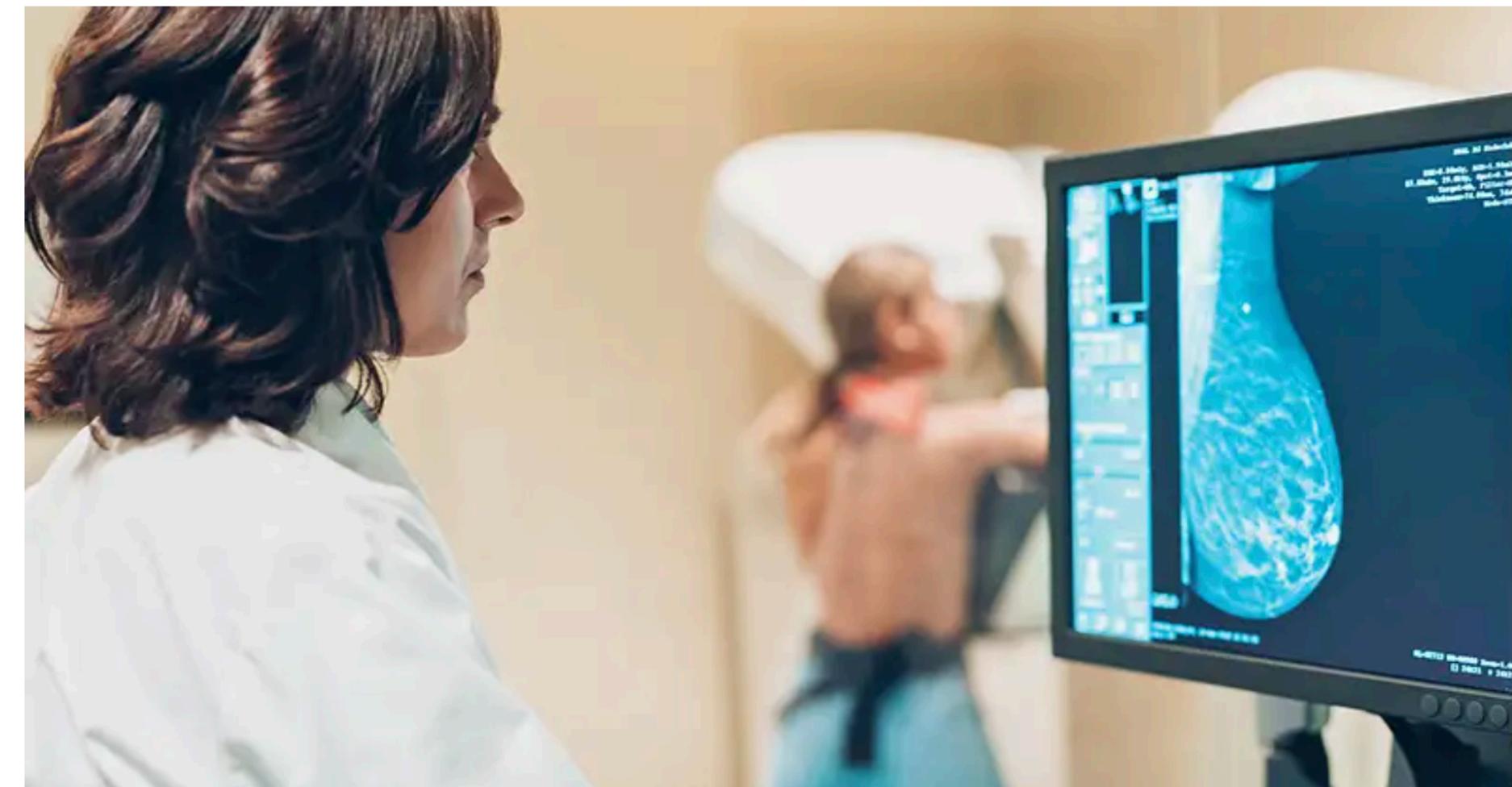
News Podcasts Video **Technology** Space Physics Health More ▾ Shop Courses Events 1

## AI system is better than human doctors at predicting breast cancer



TECHNOLOGY 1 January 2020

By [Jessica Hamzelou](#)



**itn**  
IMAGING TECHNOLOGY NEWS

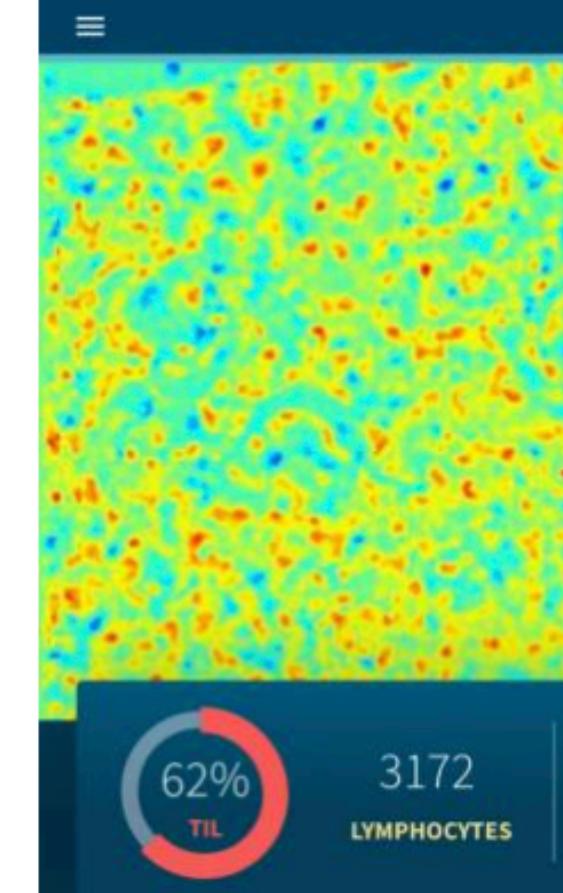
Covid-19    IMAGING    WOMEN'S    INFORMATION    RADIATION

**nature**

NEWS | ARTIFICIAL INTELLIGENCE | Explore content ▾ Journal information ▾ Publish with us ▾ Subscribe

Making the Role of AI nature > outlook > article

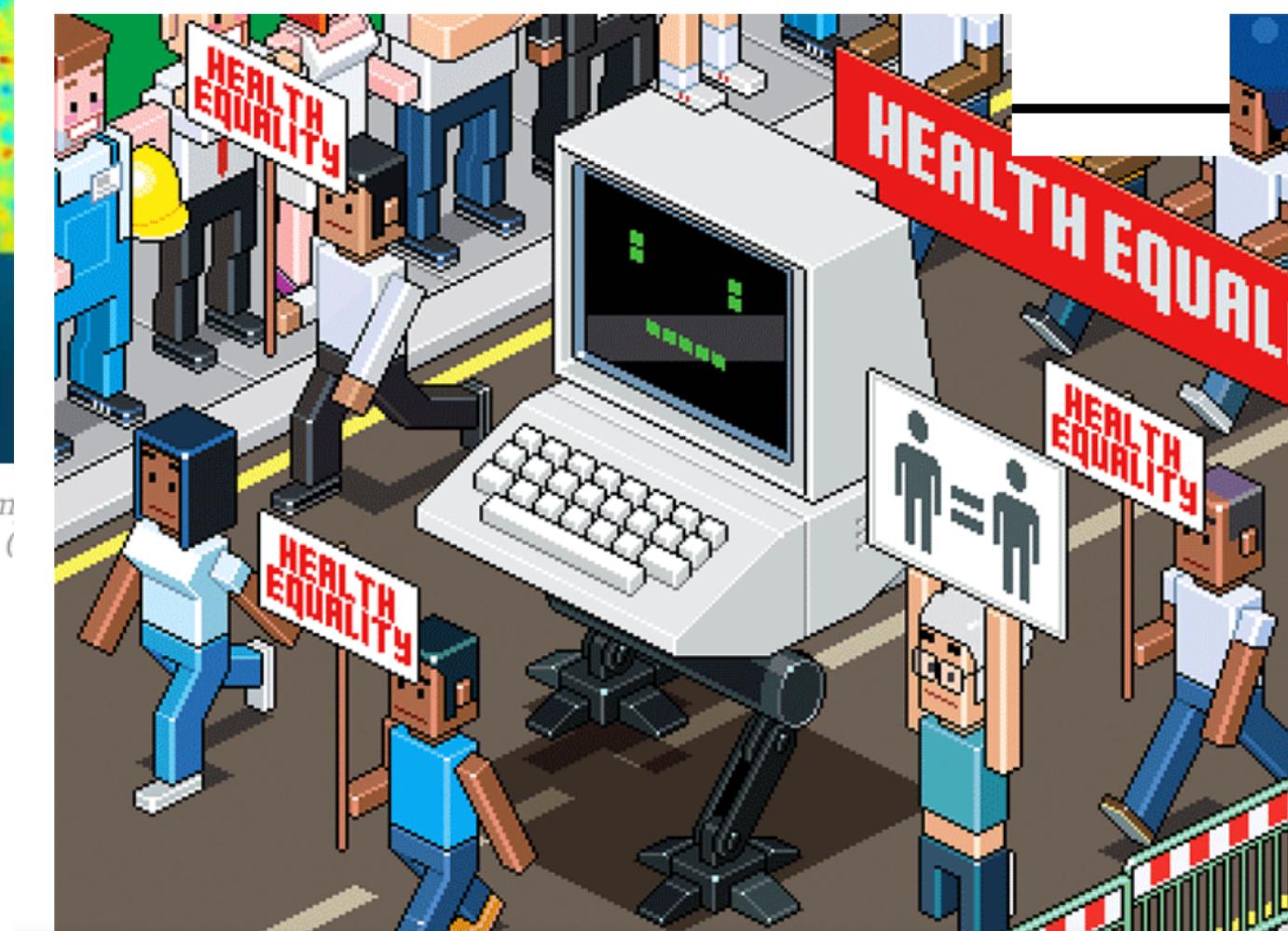
Analysis system for the diagnosis

  
Detection of tumor-infiltrating lymphocytes generate a heatmap showing TILs (of Klauschen/Charité)

A fairer way forward for AI in health care

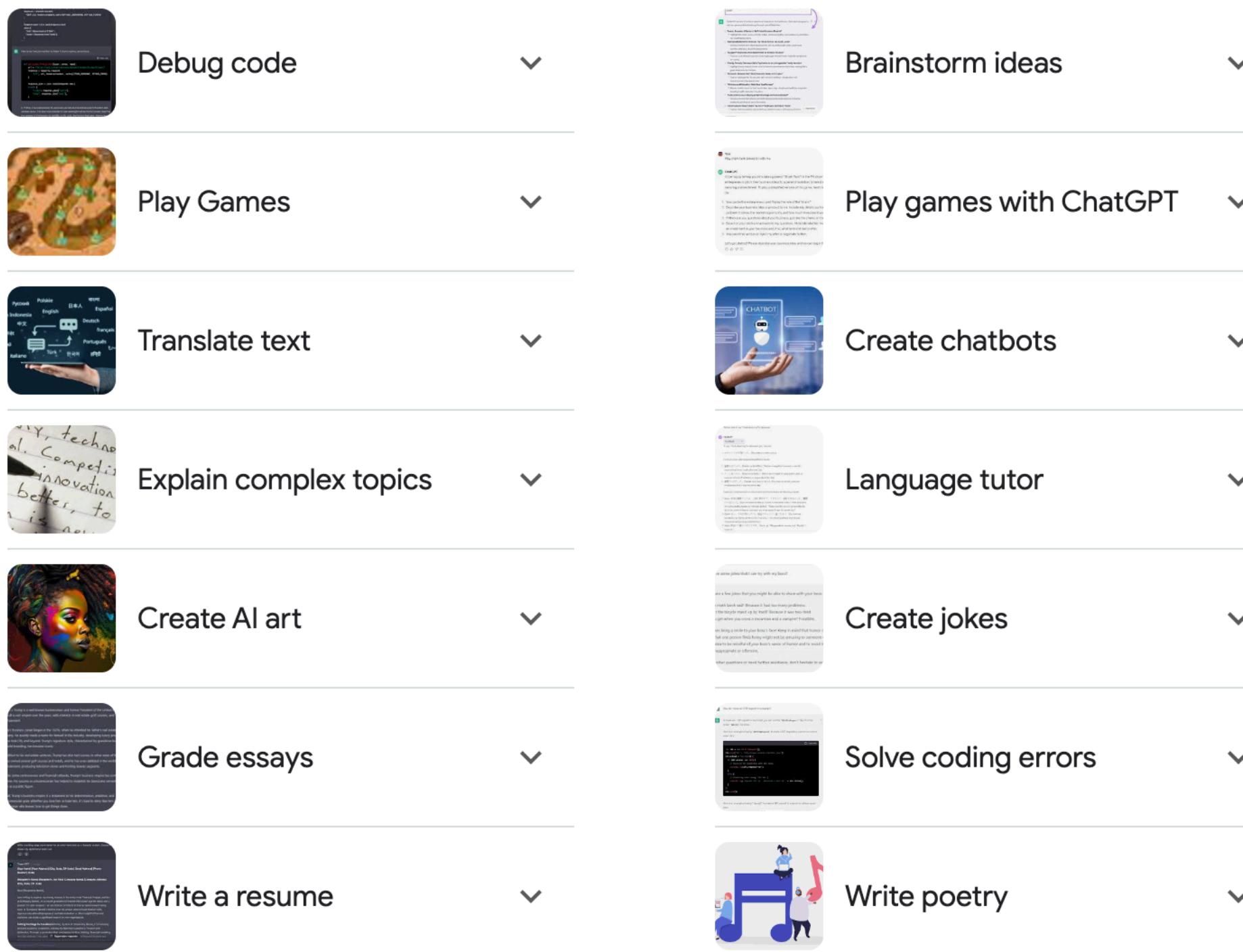
Without careful implementation, artificial intelligence could widen health-care inequality.

Linda Nordling



# Predictive vs Explainable, Trustworthy AI

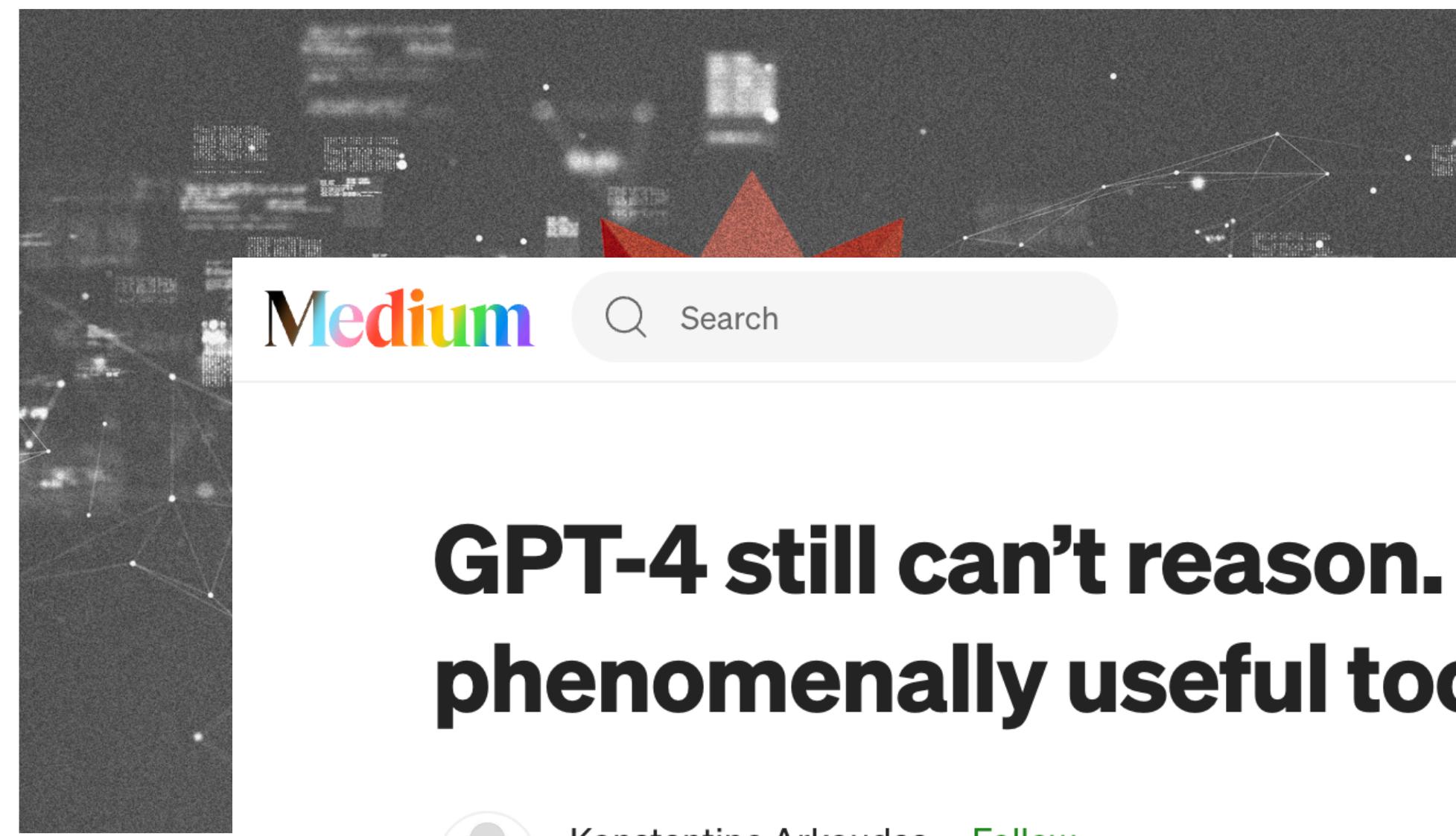
## Chat GPT - Impressive Abilities:



06-08-2024 | TECH

## This classic answer engine still outsmarts AI chatbots

For questions involving hard data and math calculations, 15-year-old WolframAlpha is a fast, accurate alternative to inaccurate AI chatbots.



## GPT-4 still can't reason. But it's a phenomenally useful tool anyway.

Konstantine Arkoudas · Follow

12 min read · Apr 11, 2024

# Causality: A Missing Link to Reasoning in AI

---

The ability to understand cause-and-effect relationships is crucial for deeper understanding and decision-making processes.

**nature**

Explore content ▾

About the journal ▾

Publish with us ▾

Subscribe

---

[nature](#) > [outlook](#) > article

OUTLOOK | 24 February 2023

## 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.

# **Prediction vs Effect of Interventions**

**Statistical Association vs Causation**

# Predictive Tasks

**Task:** Can I guess how severe is a fire by **observing** the number of firefighters?

**Yes!**

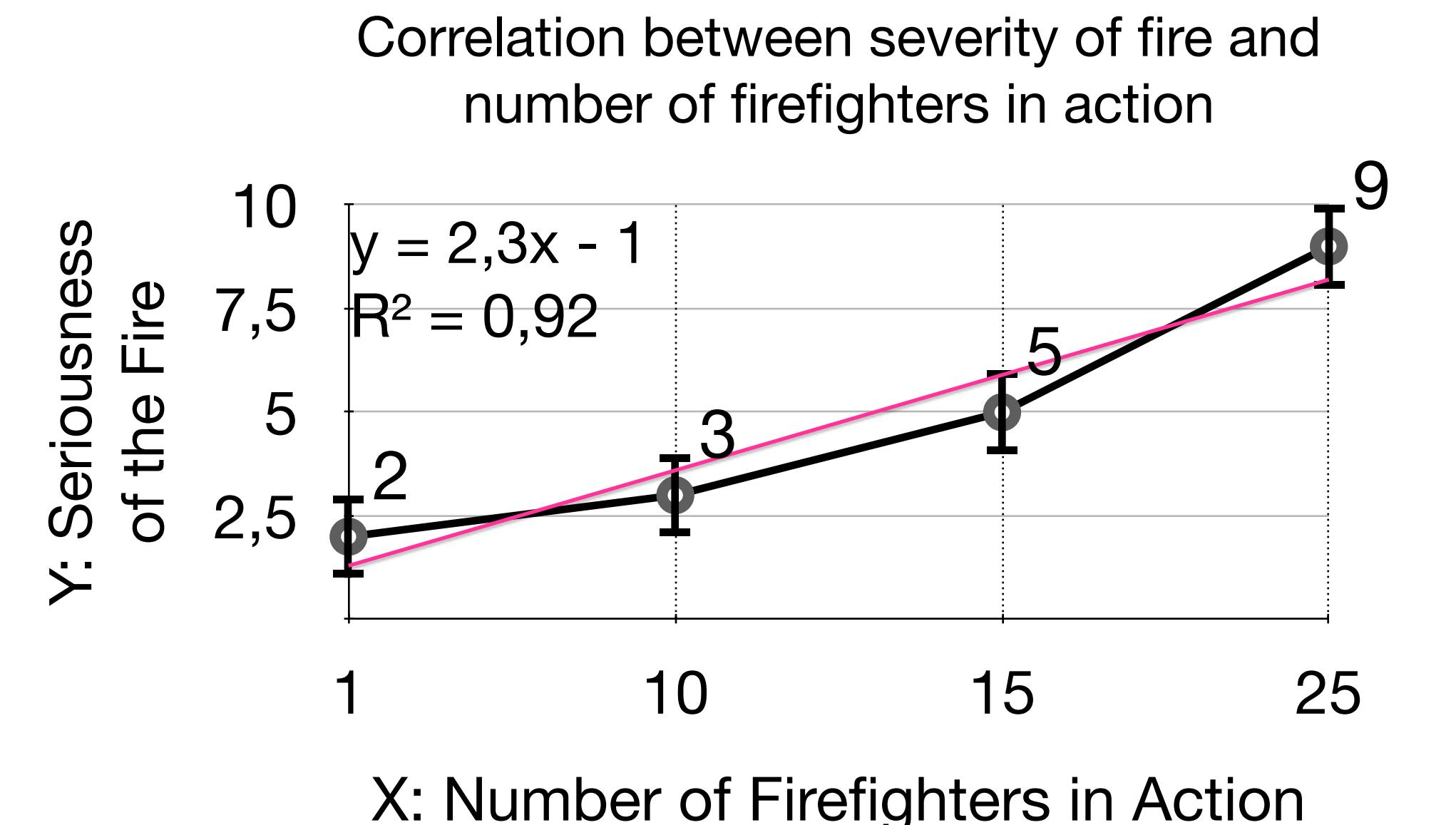
$X$ : Number of firefighters in action  
 $Y$ : Severity of the (initial) fire

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

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



**Observational  
Probability Distribution**



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

# Prediction $\Rightarrow$ Decision-Making?

---

**Conclusion:** The severity 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?**

# Causal Effect $\neq$ Statistical Association

---

The **causal effect** of  $X$  on  $Y$  is a quantity that tells us how much  $Y$  changes after an intervention  $do(X = x)$ , e.g.,  $E[Y | do(X = x)]$ .



If a different number of firefighters were dispatched,  $do(X = x)$ , would this change the expected severity of the initial fire  $Y$  ?

## Causal Effect $\equiv$ Effect of an Intervention

This quantity is derived from the probability of  $Y$  obtained after making the intervention  $do(X = x)$ , which is denoted by  $P(Y | do(X = x))$  and is commonly referred as ***Interventional Distribution***.

# Causal Effect $\equiv$ Effect of an Intervention

---

This question can also be answered if knowledge about the underlying reality is available!

$X$ : Number of firefighters in action

$Y$ : (Initial) Severity of the fire

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

**Underlying  
Structural Causal Model (SCM)**

$Y$  is not a function of  $X$

In other words,  **$X$  is not a cause of  $Y$**

In this case,  $\forall x, E[Y | do(X = x)] = E[Y]$ .

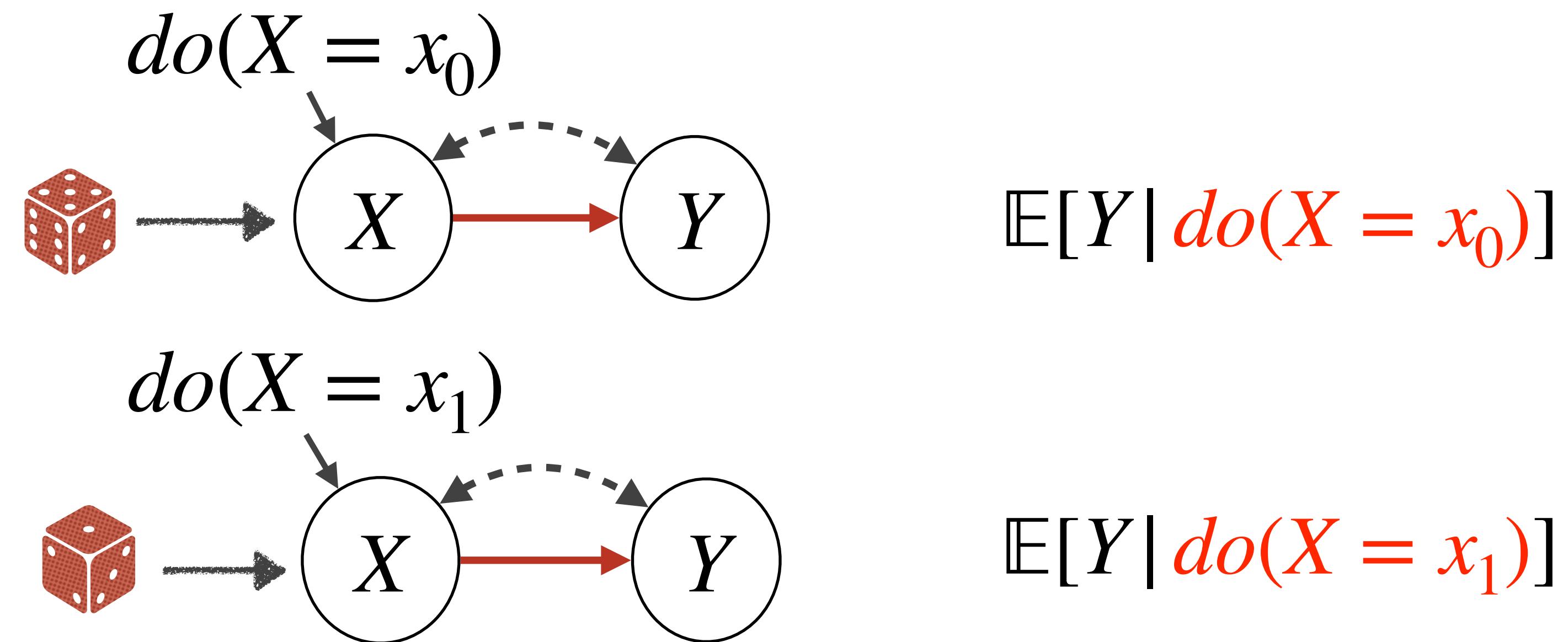
Changing the number of firefighters does not change the (initial) severity of the fire.

The action/intervention on  $X$ ,  $do(X = x)$  is independent of  $Y$ ,  
i.e.,  $P(Y = y | do(X = x)) = P(Y = y)$

# Randomized Experiments

One way to access the interventional distribution  $P(Y | \text{do}(X = x))$  is through a *perfectly realized* Randomized Experiments / Control Trials (e.g. RCT):

Randomization of the  
 $X$ 's assignment

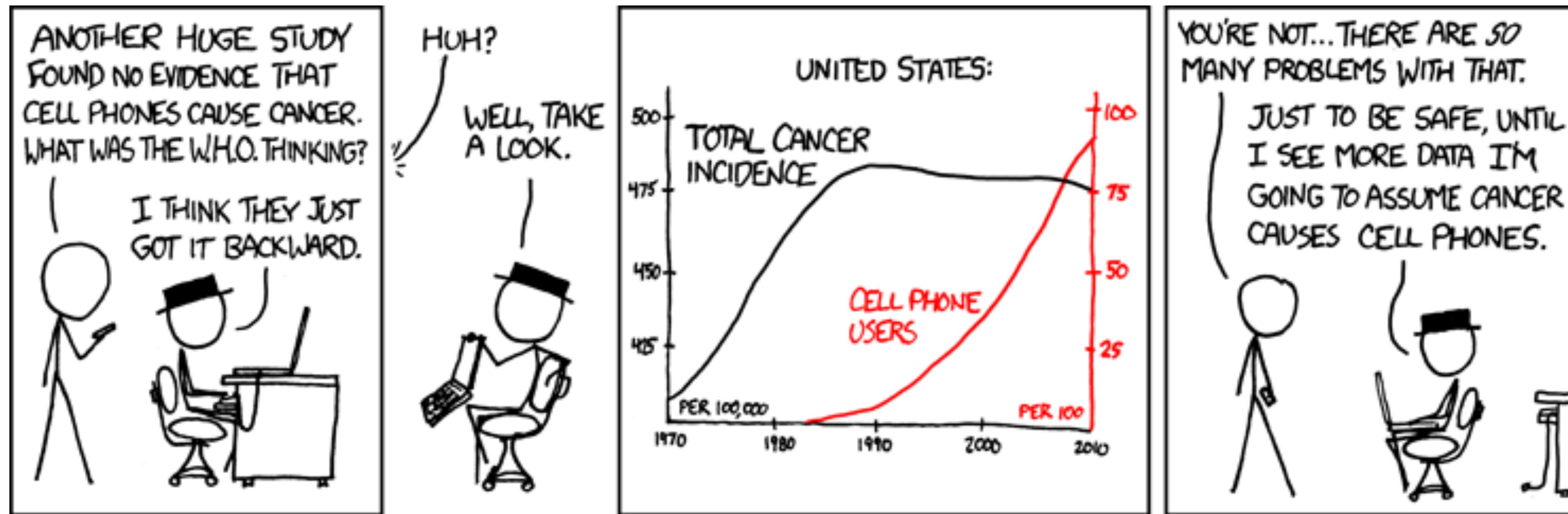


**Average Causal Effect:**  $\mathbb{E}[Y | \text{do}(X = x_0)] - \mathbb{E}[Y | \text{do}(X = x_1)]$

What is the causal effect of the number of firefighters  $X$  on the severity of the initial fire  $Y$  ?

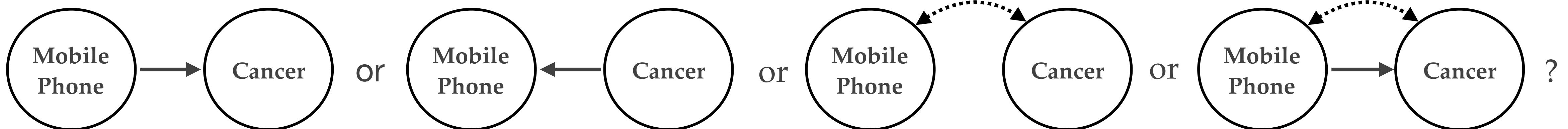
**Can we identify causal effects without  
conducting a randomized experiment?**

# Association vs Causation



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

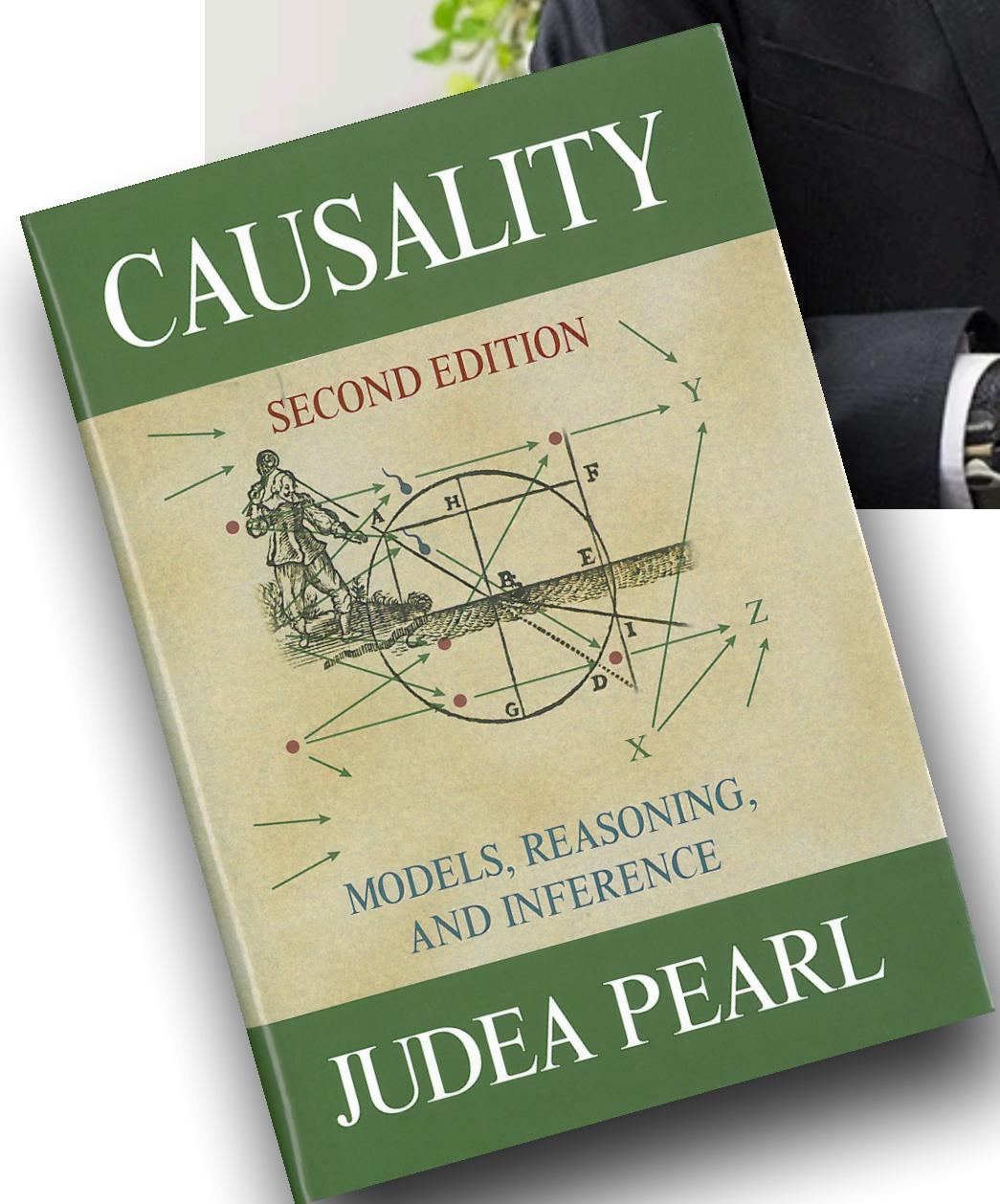
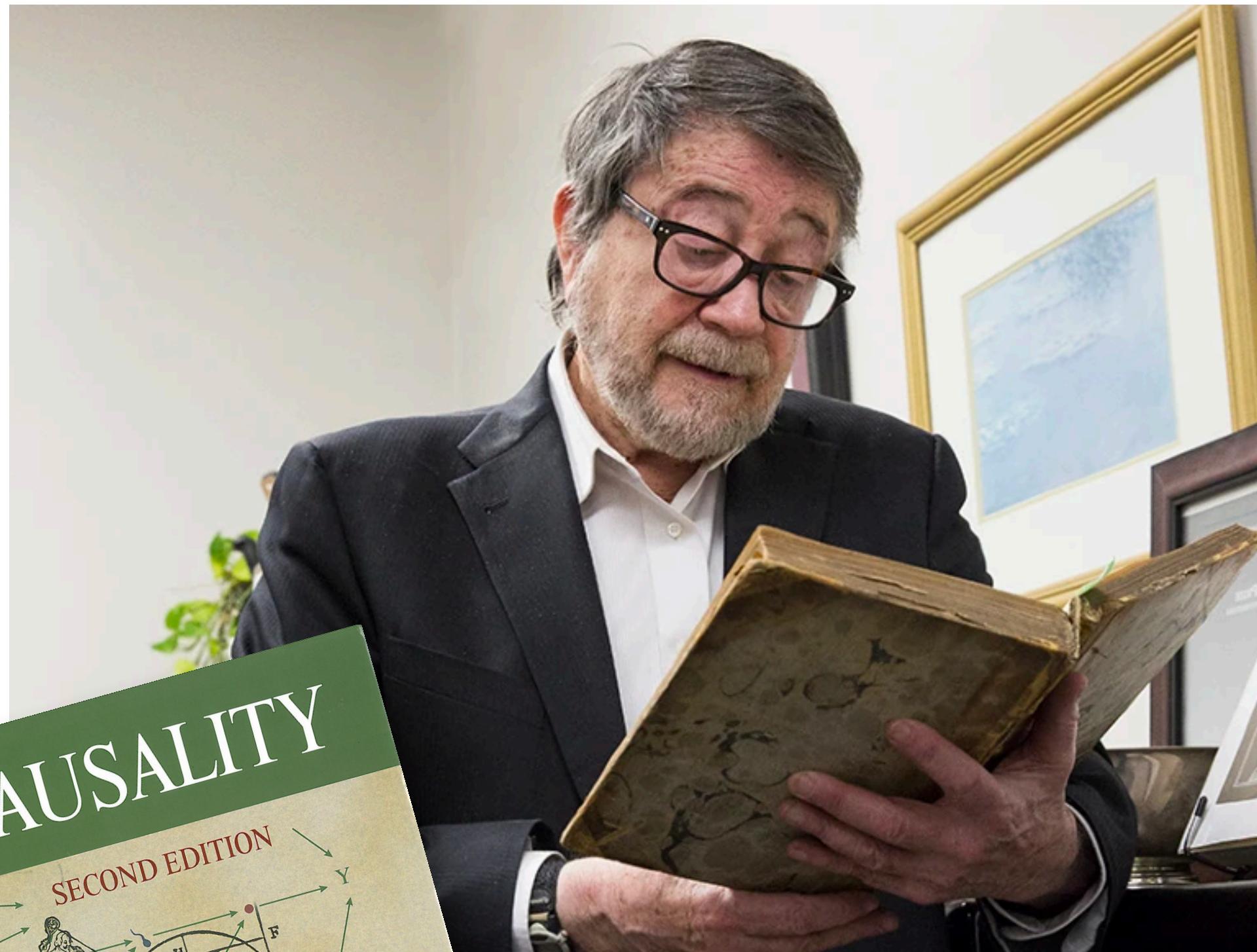
Will we be able to decide the true relationship just by “seeing” more data?



# The Mathematical Framework of Causal Data Science

# Judea Pearl – Causality

---



Director of the Cognitive Systems Laboratory at the University of California, Los Angeles.

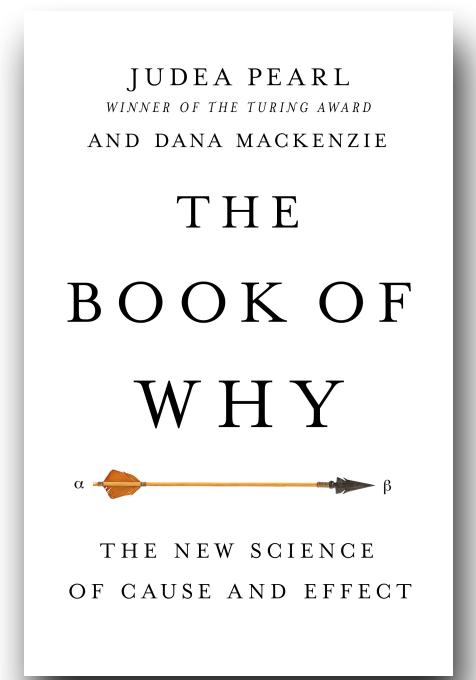
In 2011, he won the A. M. Turing Award (the highest distinction in computer science and a \$250,000 prize)

“for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.”

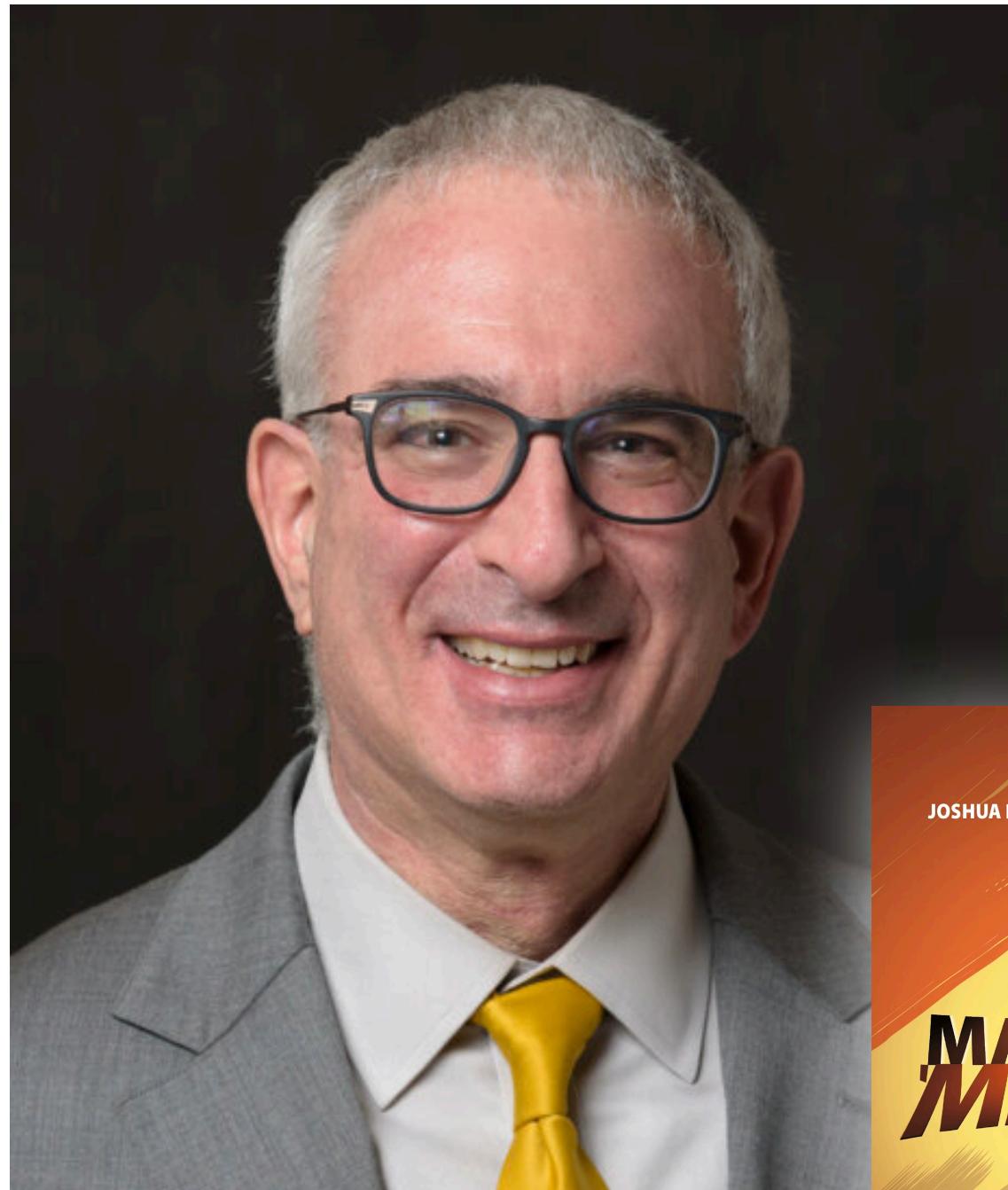
— Association for Computing Machinery (ACM)

“Deep learning has instead given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality.”

— The Book of Why: The New Science of Cause and Effect



# Guido W. Imbens, Joshua D. Angrist & Donald B. Rubin

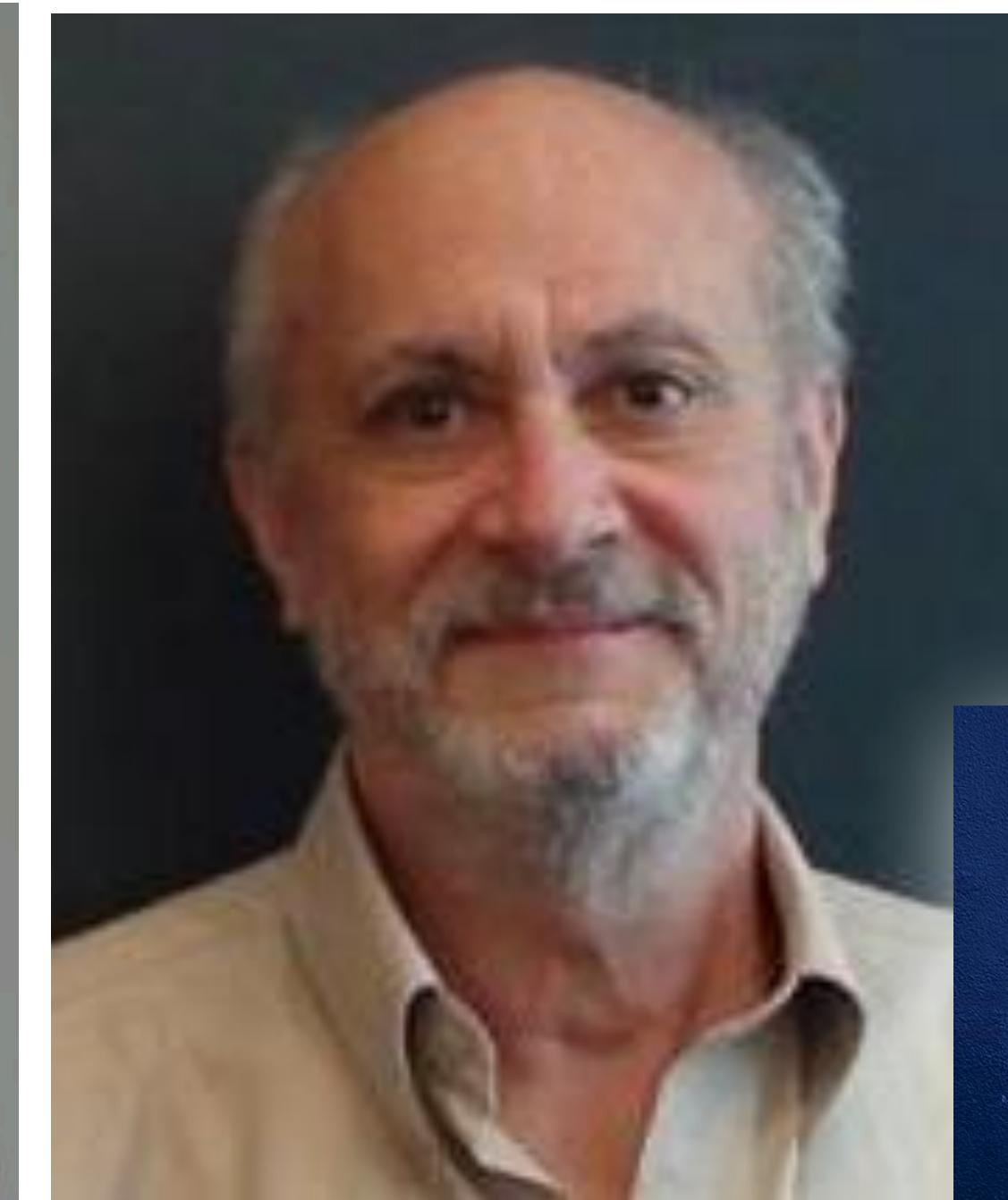


Joshua D. Angrist

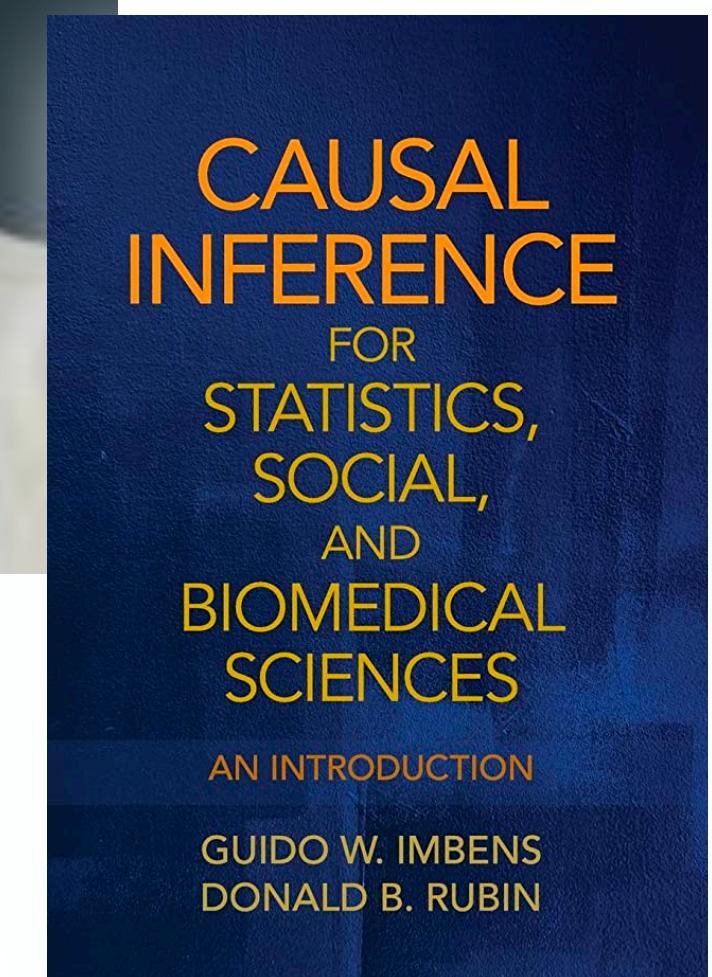
Professor of  
Economics at MIT



Guido W. Imbens  
Professor of Applied  
Econometrics at  
Stanford University



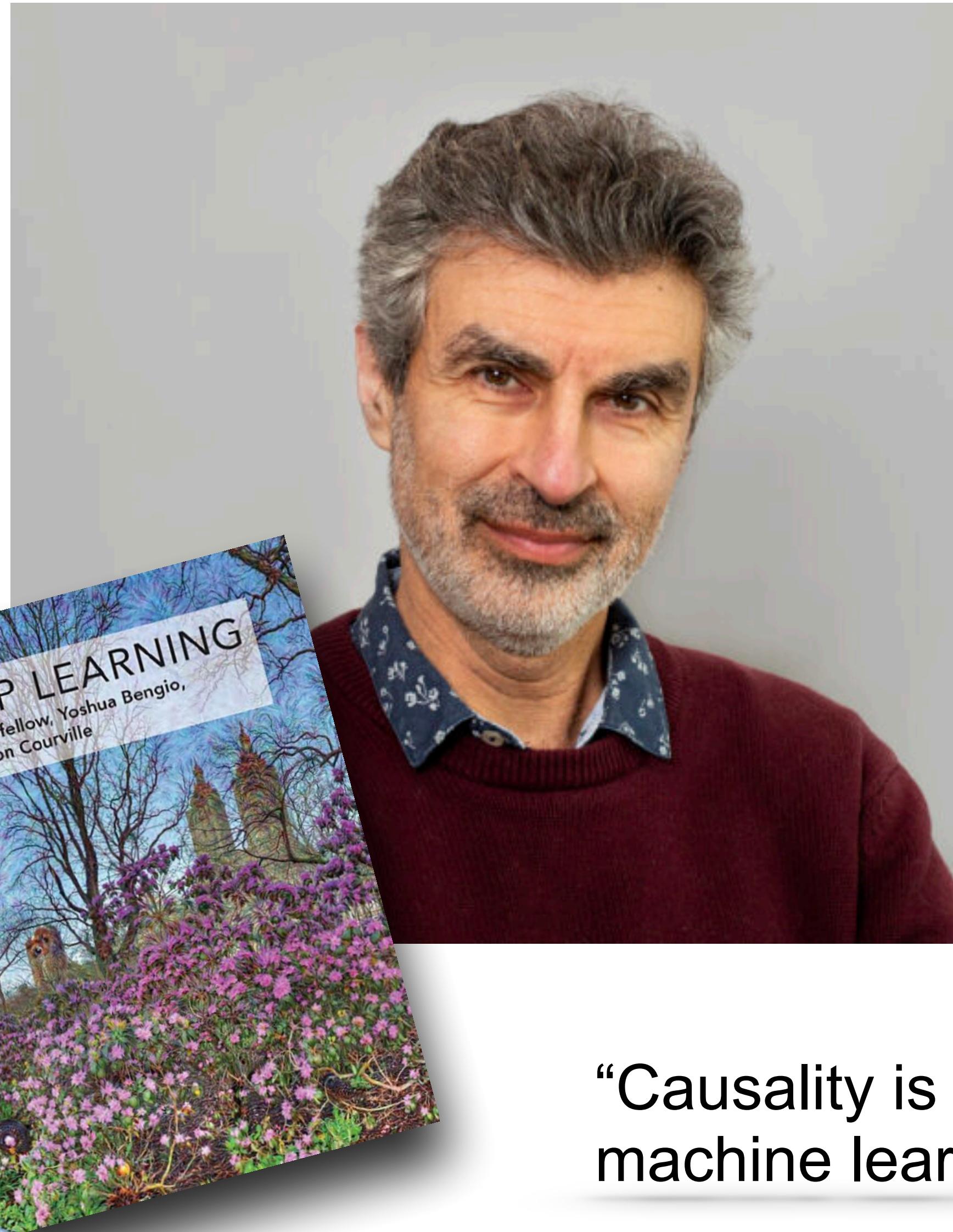
Donald B. Rubin  
Professor of  
Statistics at  
Harvard University



In 2021, Angrist & Imbens won the Nobel Prize in Economics  
“for their methodological contributions to the analysis of causal relationships”

# Yoshua Bengio – Deep Learning

---



Professor at the University of Montreal, and the Founder and Scientific Director of Mila – Quebec AI Institute

In 2018, he won the A. M. Turing Award, with Geoffrey Hinton, and Yann LeCun

“for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.”

— Association for Computing Machinery (ACM)

“Causality is very important for the next steps of progress of machine learning,” — interview with *IEEE Spectrum*.

# Why causality is so important?

---

Causality allows important capabilities such as

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

- **Causal Effect Identification and Estimation**

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

- **Causal Discovery**

**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

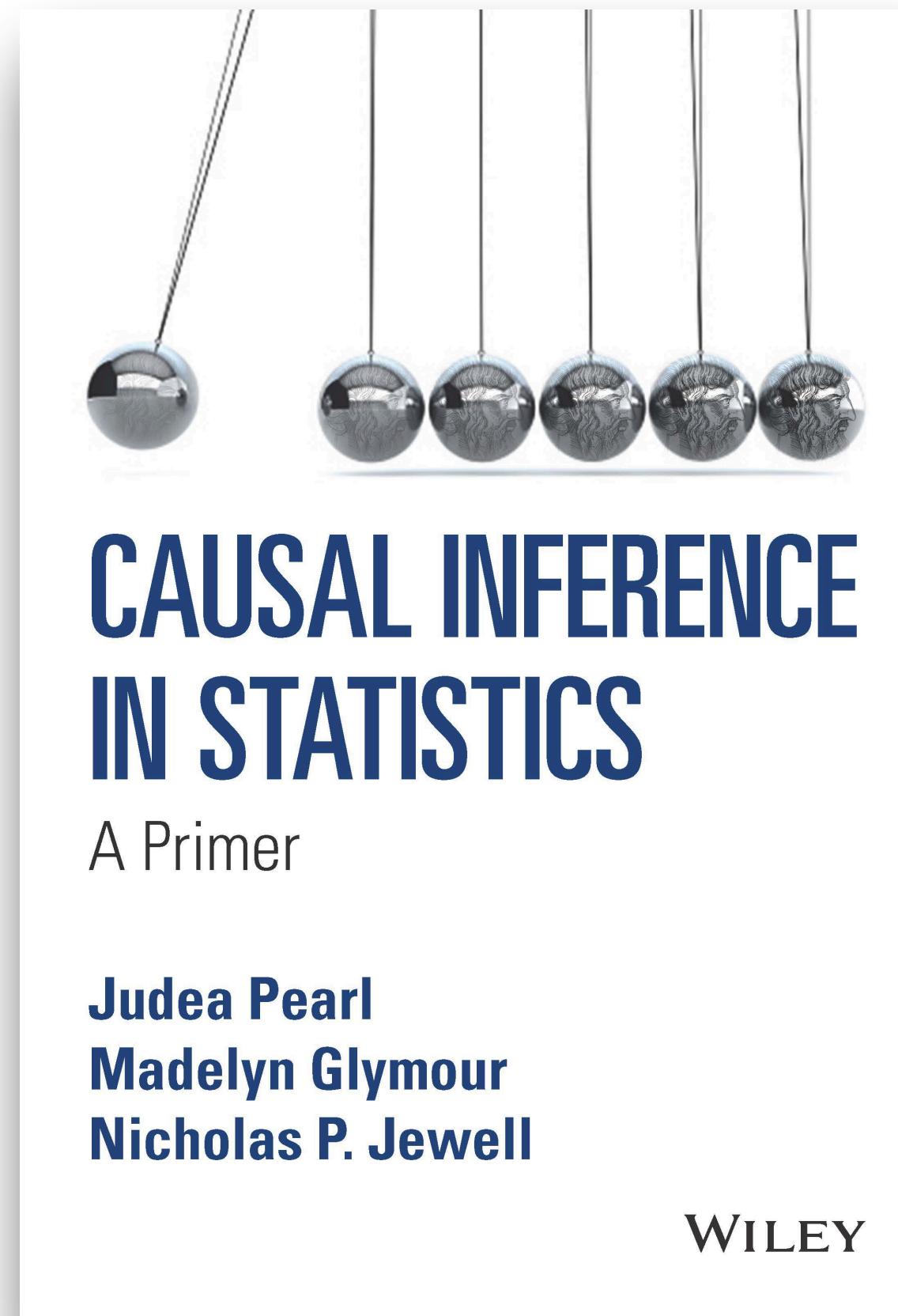
Elias Bareinboim<sup>a,b,1</sup> and Judea Pearl<sup>a</sup>

<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

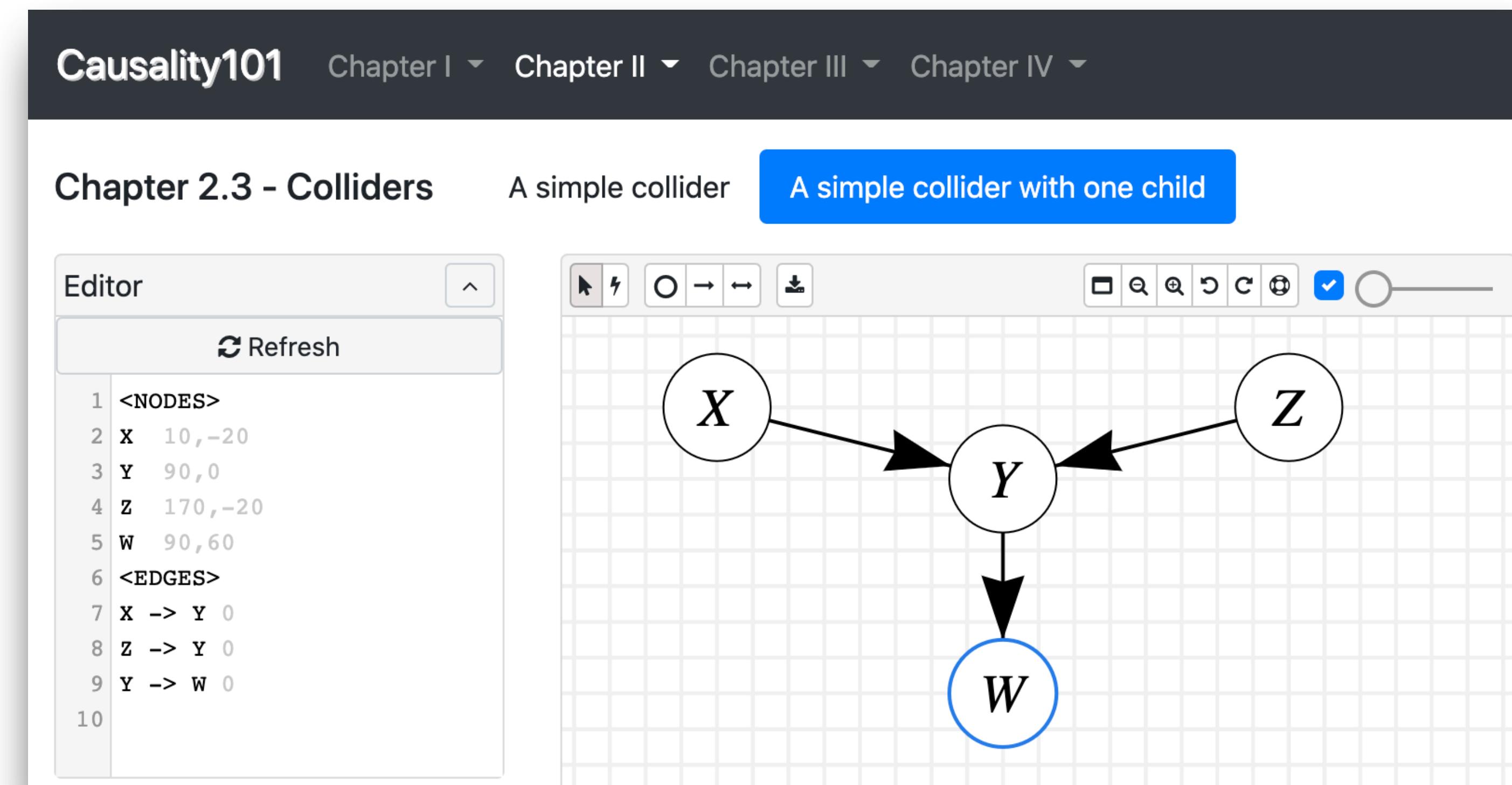
Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved March 15, 2016 (received for review June 29, 2015)

<http://causalfusion.net>

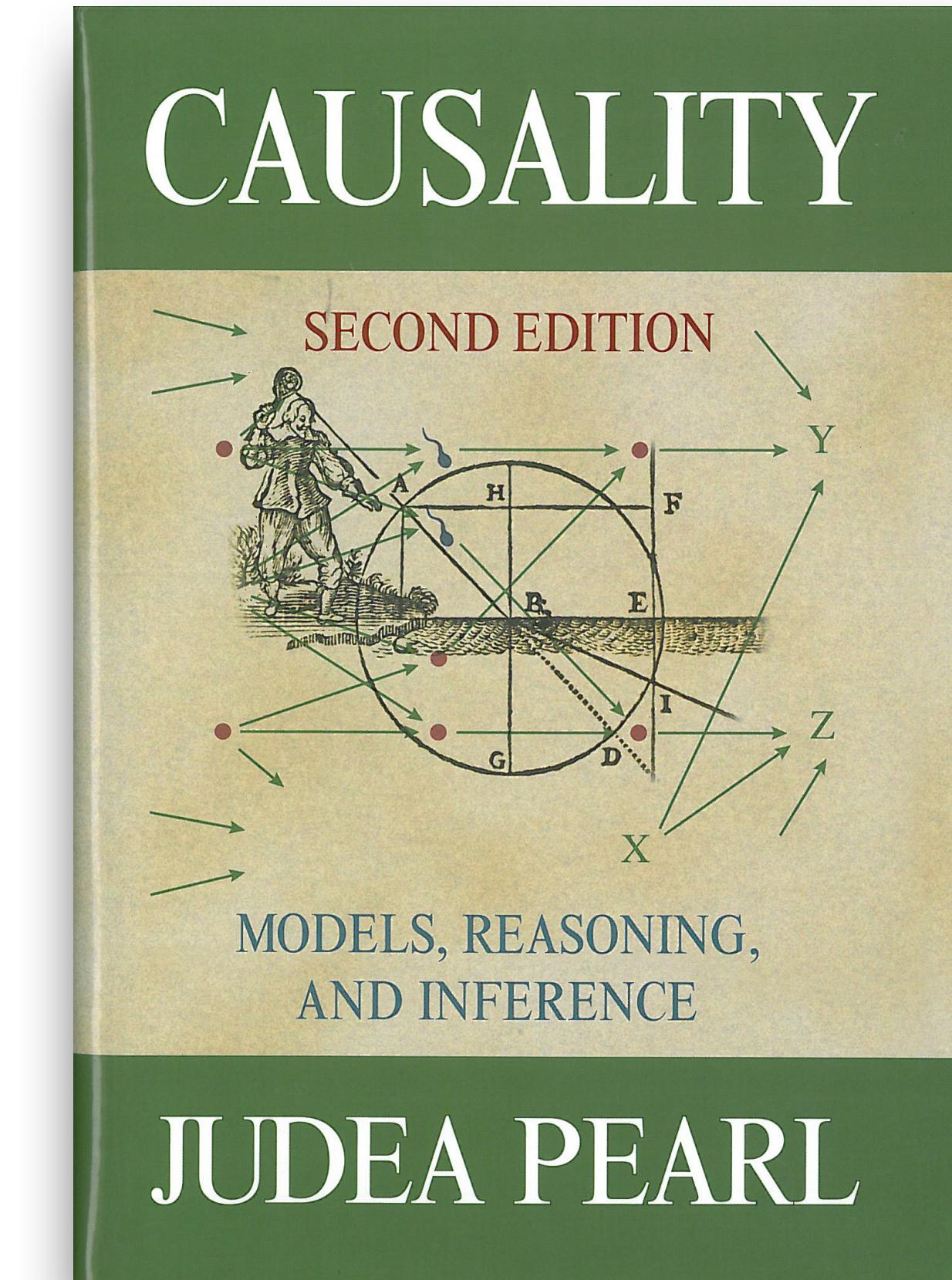
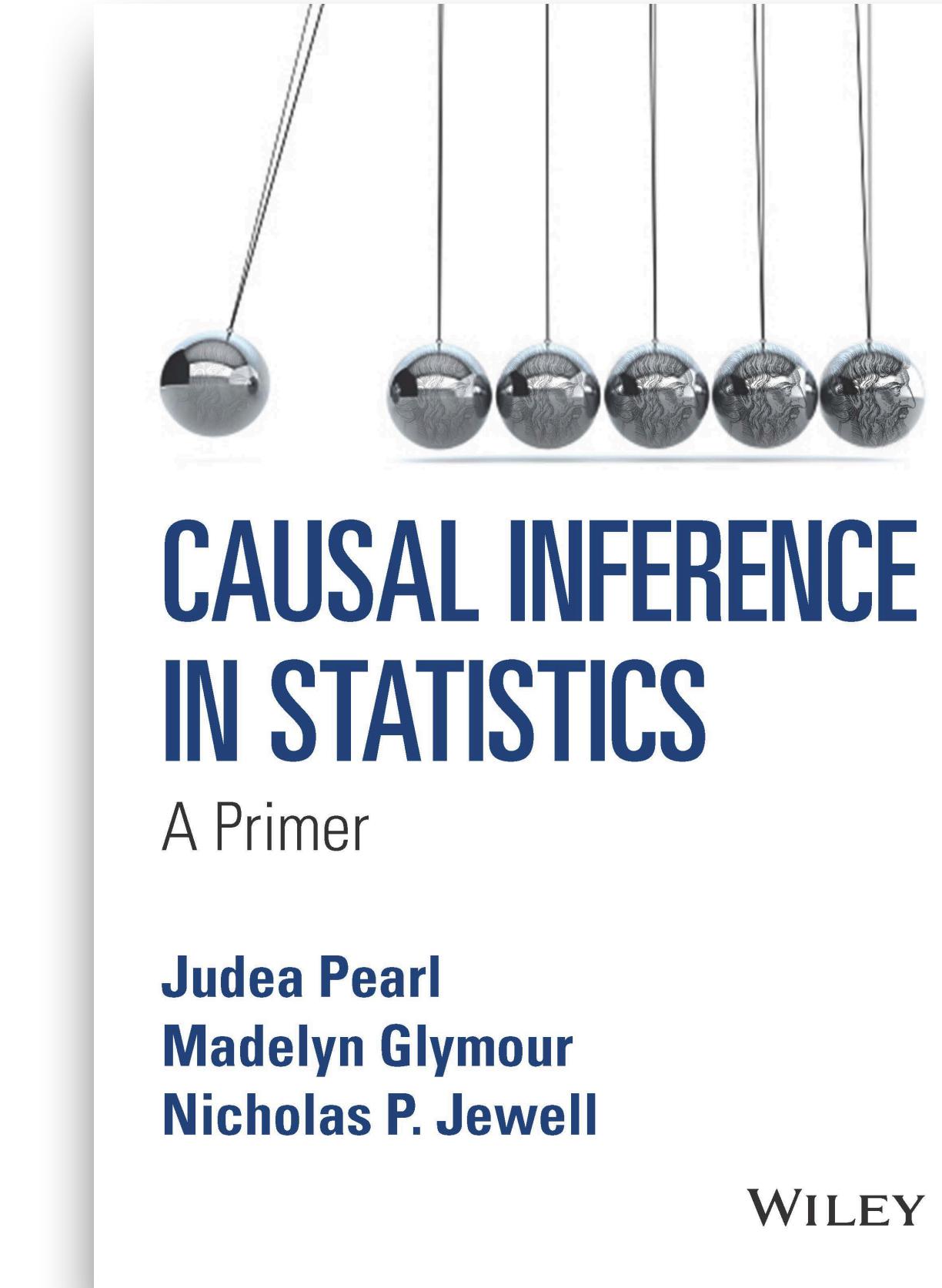
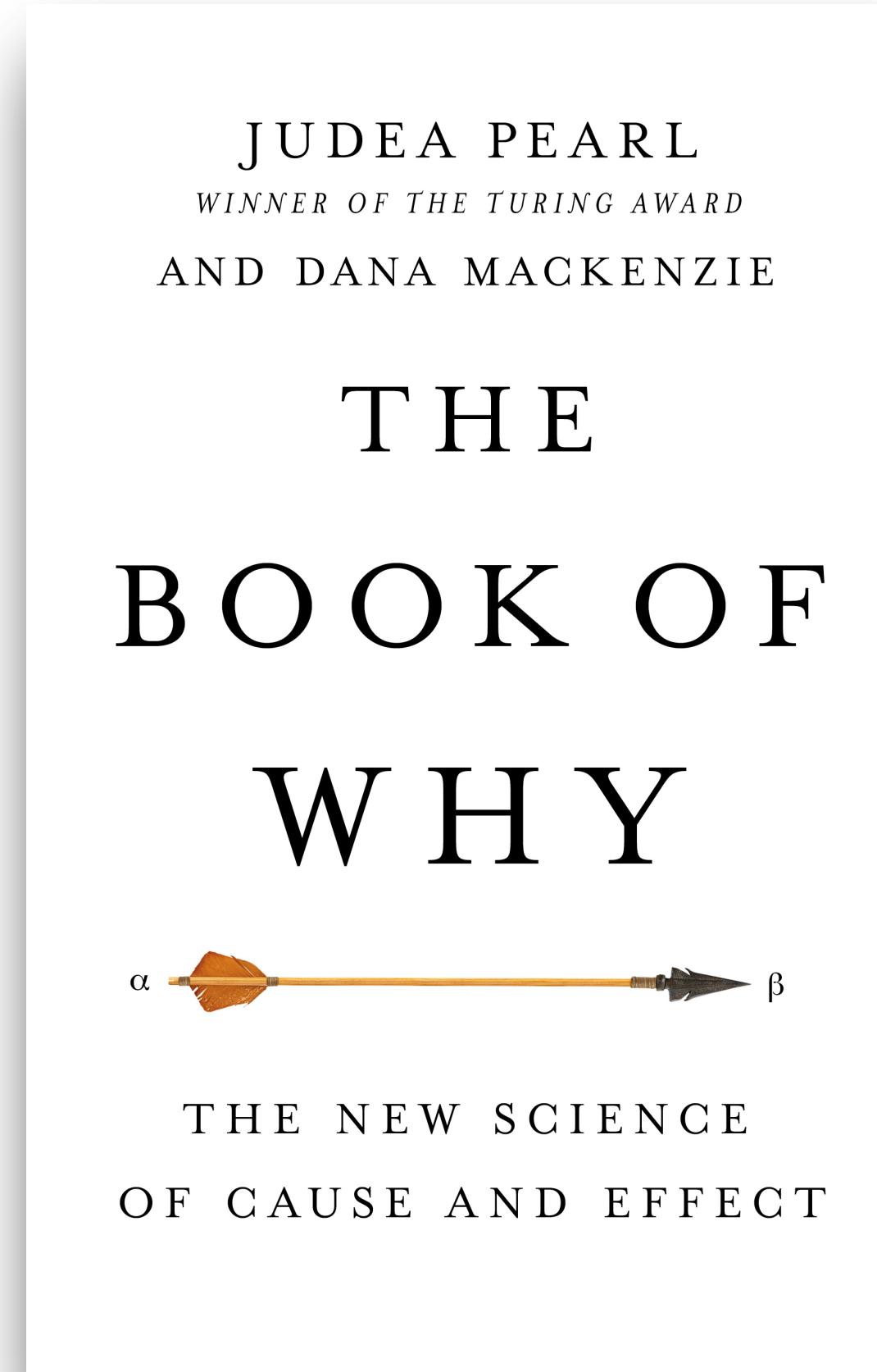
# Causality Theory by Judea Pearl



<https://causality101.net/>



# Causality Theory by Judea Pearl



# **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.

# 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).

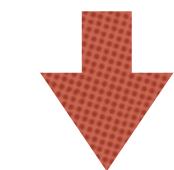
**Full specification of an SCM requires parametric and distributional assumptions.**

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

# Statistical Association vs Causation

## Pre-Interventional/ Observational SCM

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



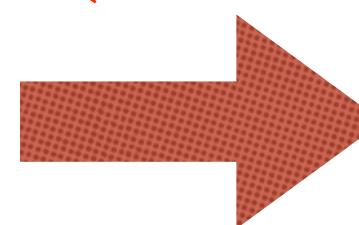
Observational  
Distribution

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

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

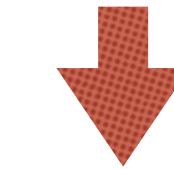
$$P(Y = y | X = x) \neq P(Y = y) \implies X \text{ is } \text{correlated} \text{ to } Y$$

$do(X = x)$



## Post-Interventional / Interventional SCM

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



Interventional  
Distribution

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

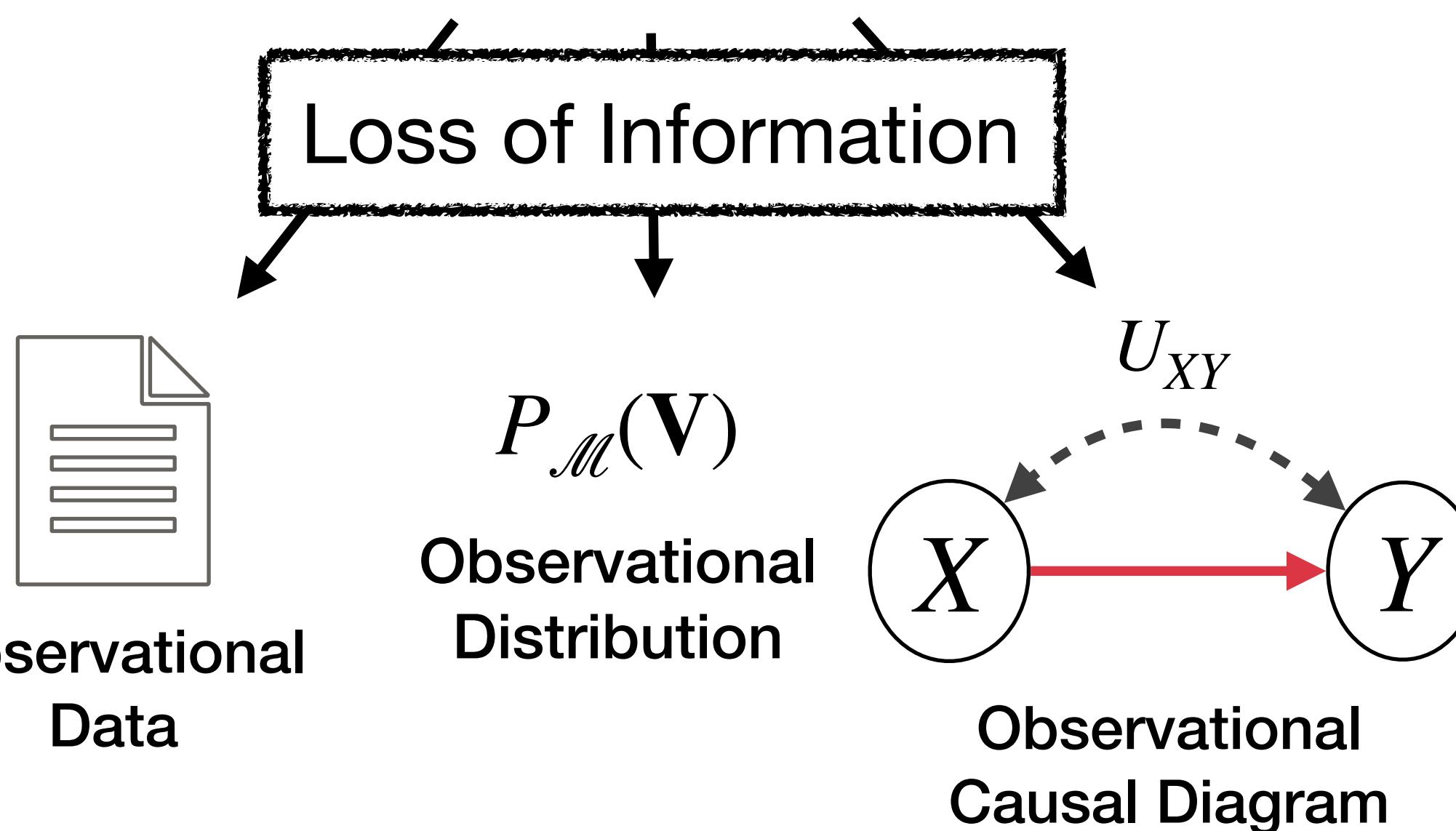
Can we **predict** better the value of  $Y$  after  
**making an intervention**  $do(X = x)$ ?

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

# Statistical Association vs Causation

## Pre-Interventional/ Observational SCM

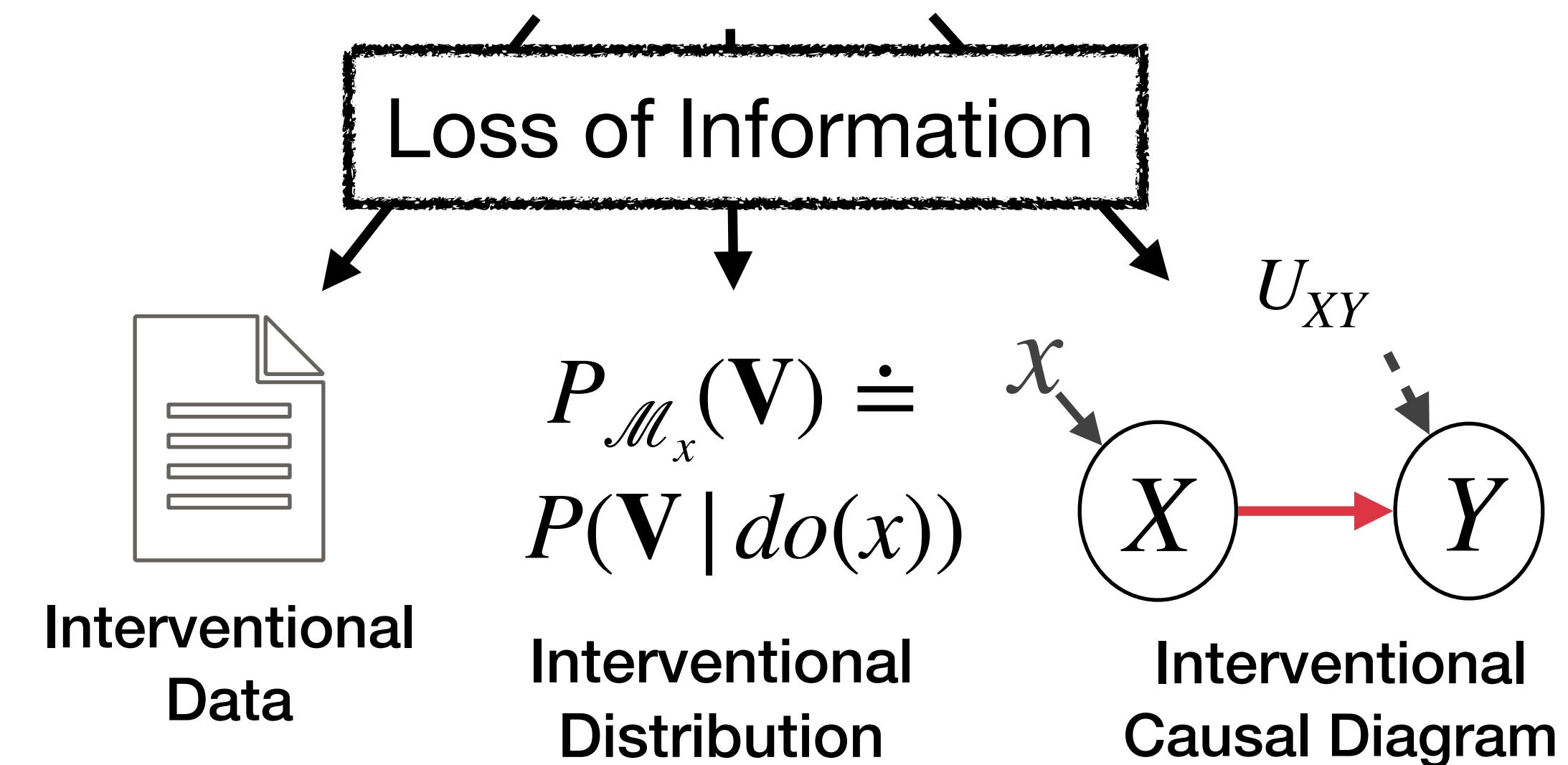
$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \left\{ \begin{array}{l} X = f_X(U_X, U_{XY}) \\ Y = f_Y(X, U_Y, U_{XY}) \end{array} \right. \\ P(\mathbf{U}) \end{cases}$$



$do(X = x)$

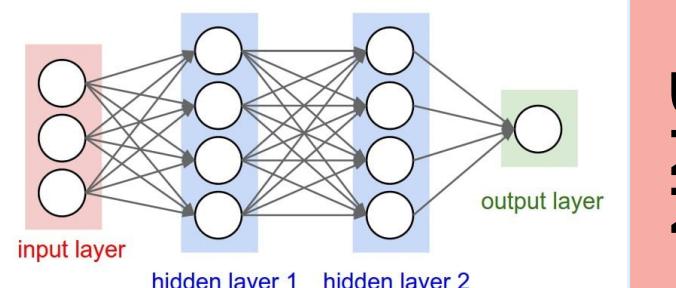
## Post-Interventional / Interventional SCM

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



Observational

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

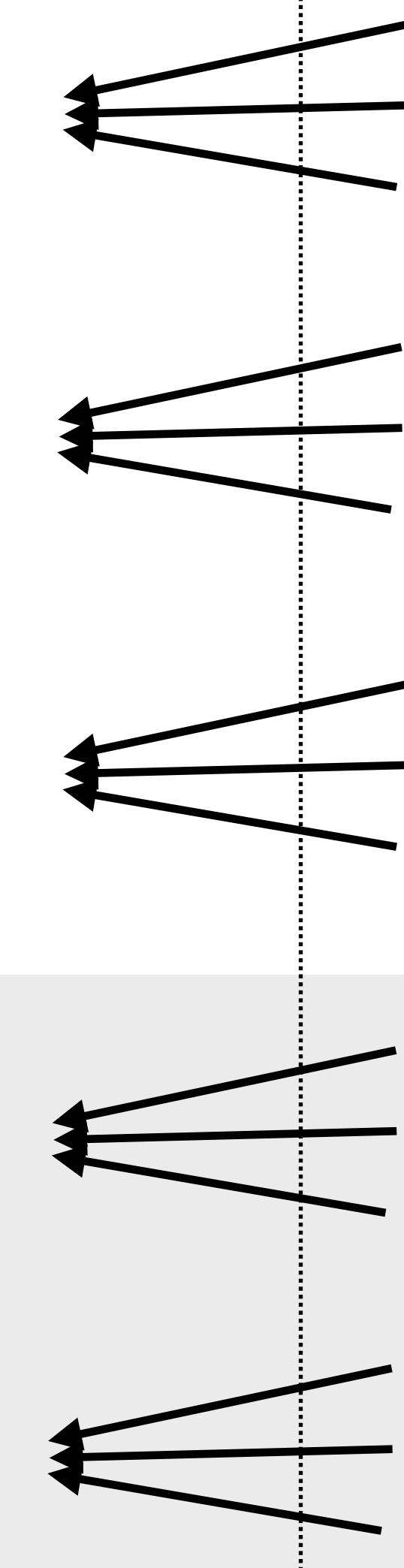
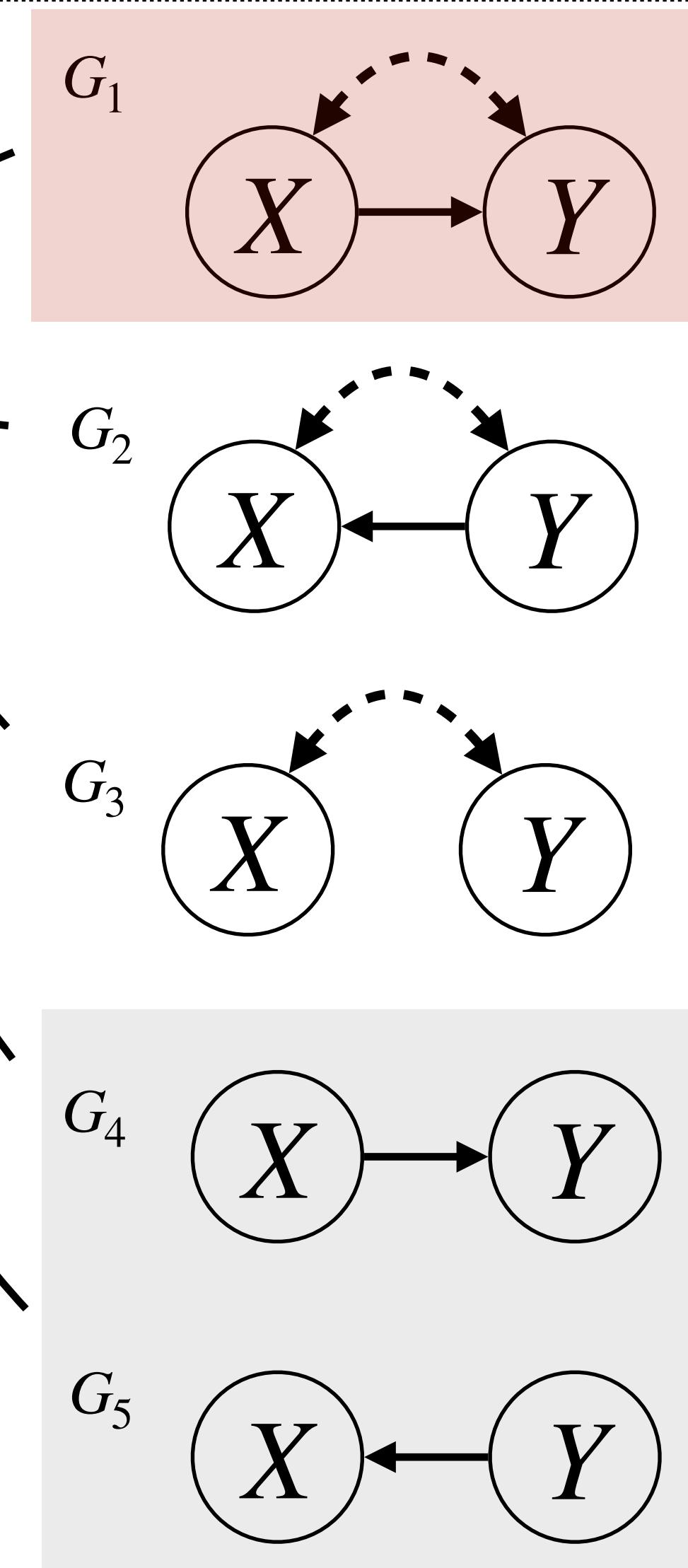


$$P(Y|X=x)$$

Data

Potential Causal Diagrams

Potential SCMs



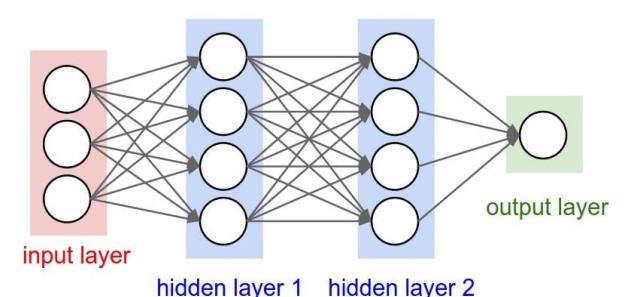
- $\mathcal{M}_{11} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{11}, P_{11}(\mathbf{u}_1) \rangle$
- $\vdots$
- $\mathcal{M}_{1k_1} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{1k_1}, P_{1k_1}(\mathbf{u}_1) \rangle$
- $\mathcal{M}_{21} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{21}, P_{21}(\mathbf{u}_2)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

Parametrization

Encoded Knowledge / Assumptions

Observational

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



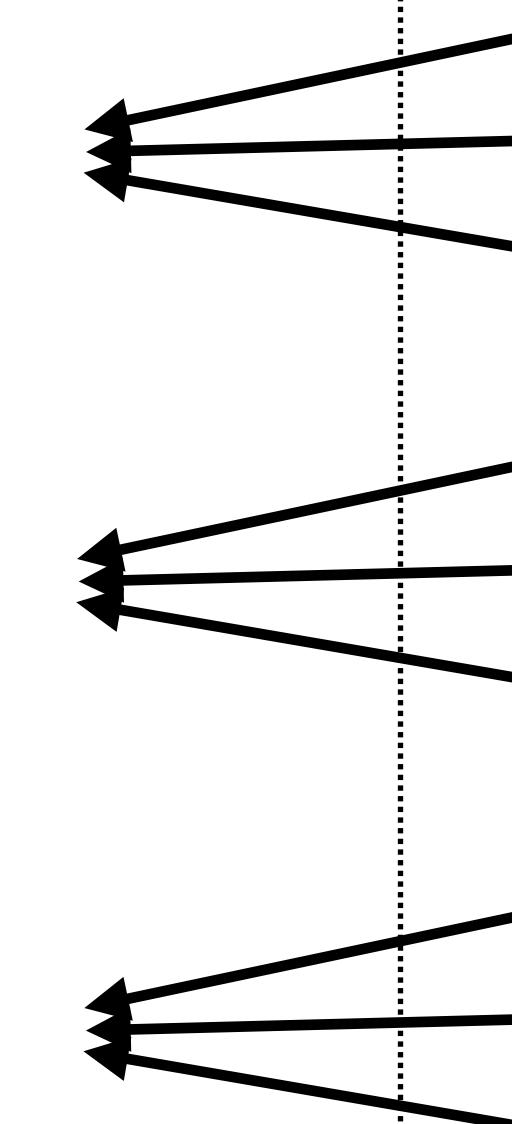
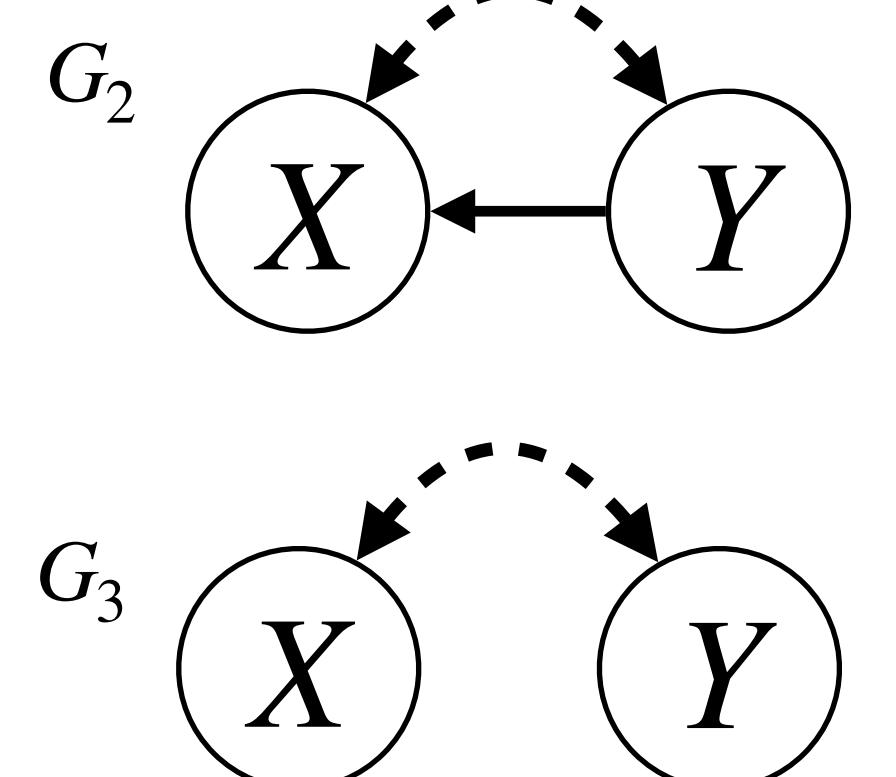
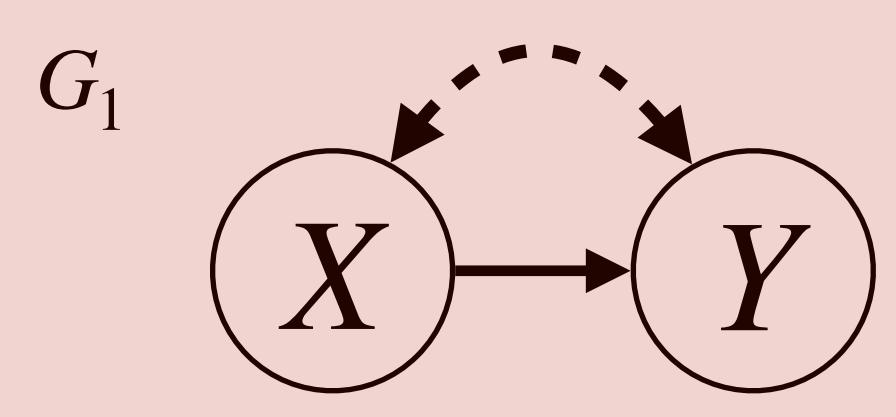
$$P(Y|X=x)$$

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

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

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

⋮

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

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

⋮

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

$$, P_{41}(\mathbf{u}_4) \rangle$$

$$, P_{4k_4}(\mathbf{u}_4) \rangle$$

$$, P_{51}(\mathbf{u}_5) \rangle$$

⋮

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

Parametrization

Encoded Knowledge / Assumptions

# Bayesian Network

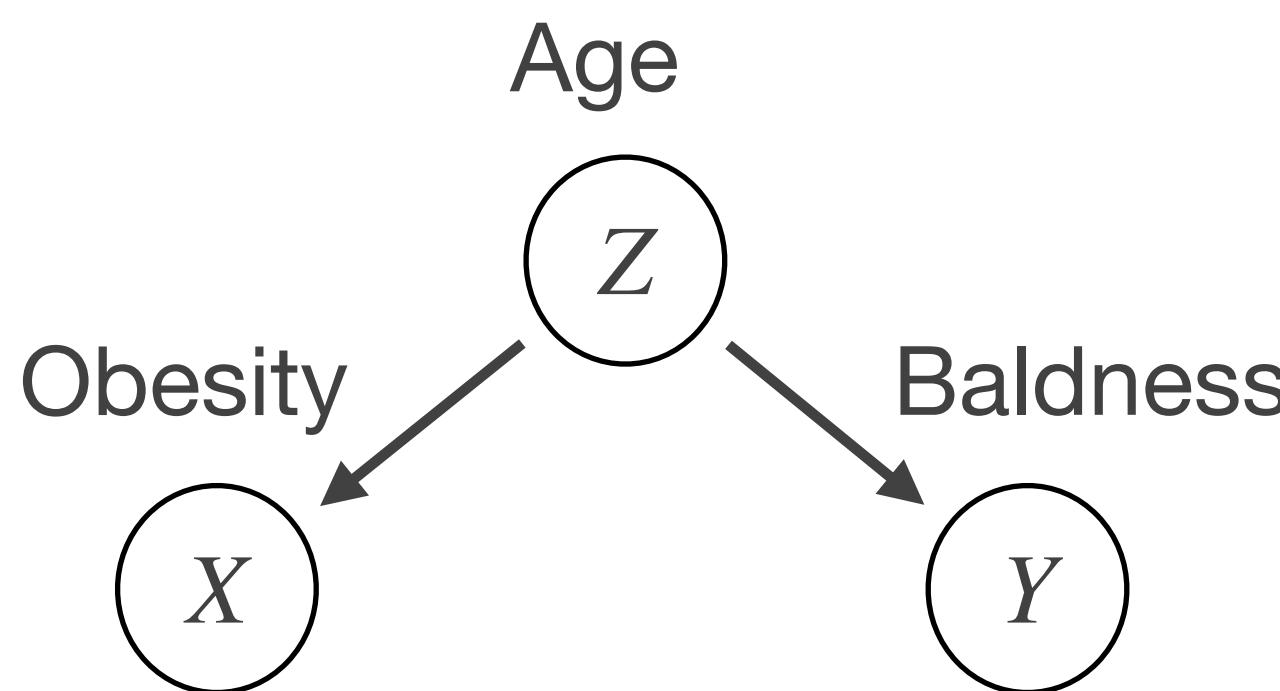
Directed  
Acyclic Graph

A DAG, possibly with latent confounders (ADMG),  
representing the **conditional independences**  
implied by an SCM

Acyclic Directed  
Mixed Graph

# Conditional independencies implied by d-separation

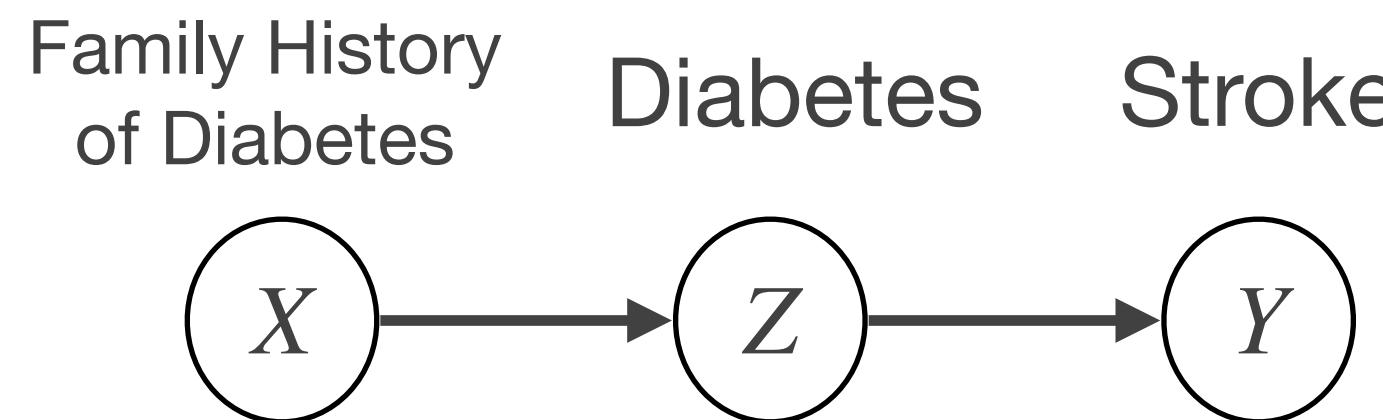
Fork



$$X \perp\!\!\!\perp Y$$

$$X \perp\!\!\!\perp Y | Z$$

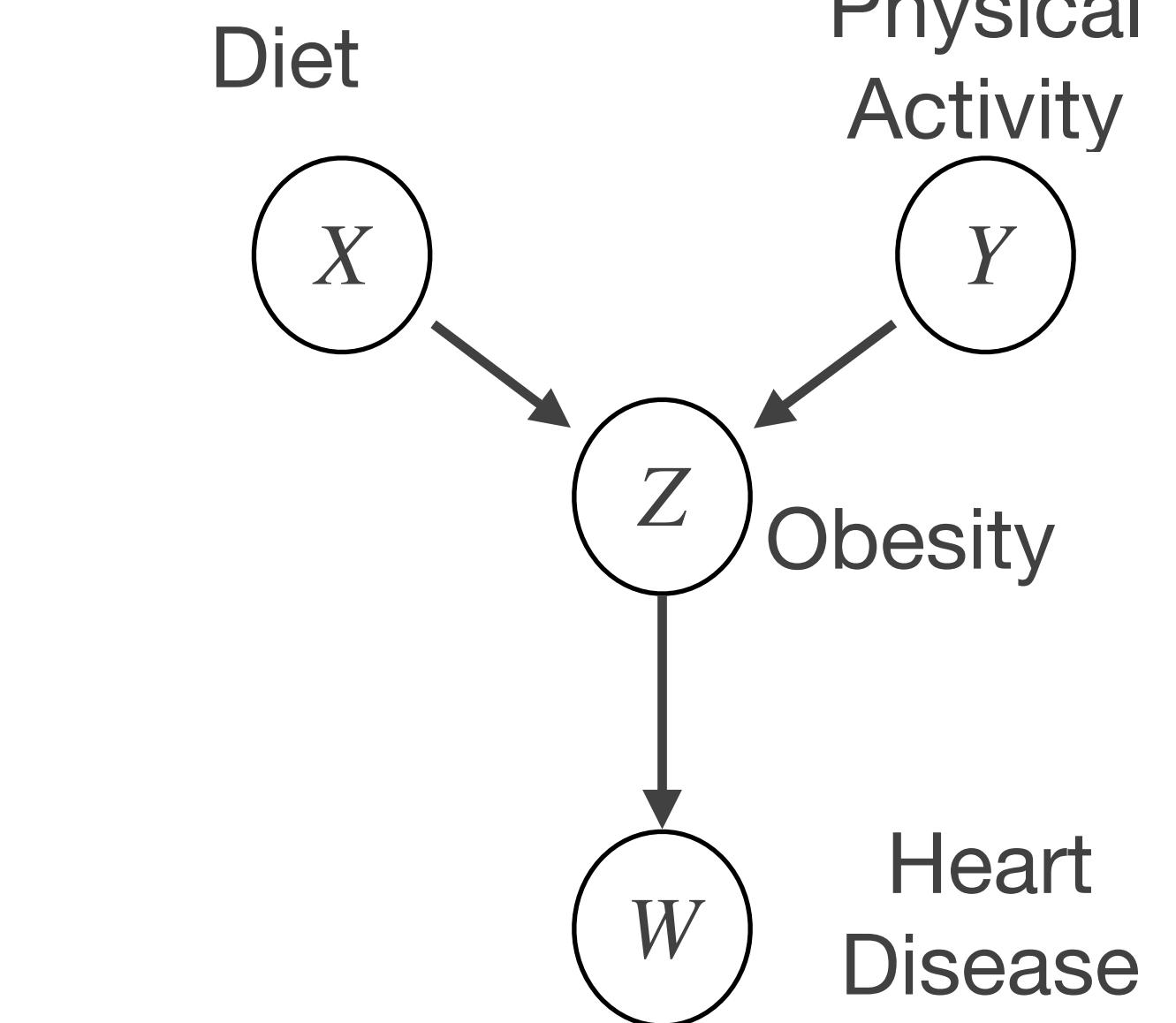
Chain



$$X \perp\!\!\!\perp Y$$

$$X \perp\!\!\!\perp Y | Z$$

V-Structure



$$X \perp\!\!\!\perp Y$$

$$X \perp\!\!\!\perp Y | Z$$

$$X \perp\!\!\!\perp Y | W$$

In both cases, Z is a non-collider!

# BN - Encoder of Conditional Independences

Bayesian Networks (BN) are **Minimal Independence Maps**:

$$(X \perp\!\!\!\perp Y | Z)_G \Rightarrow (X \perp\!\!\!\perp Y | Z)_P$$

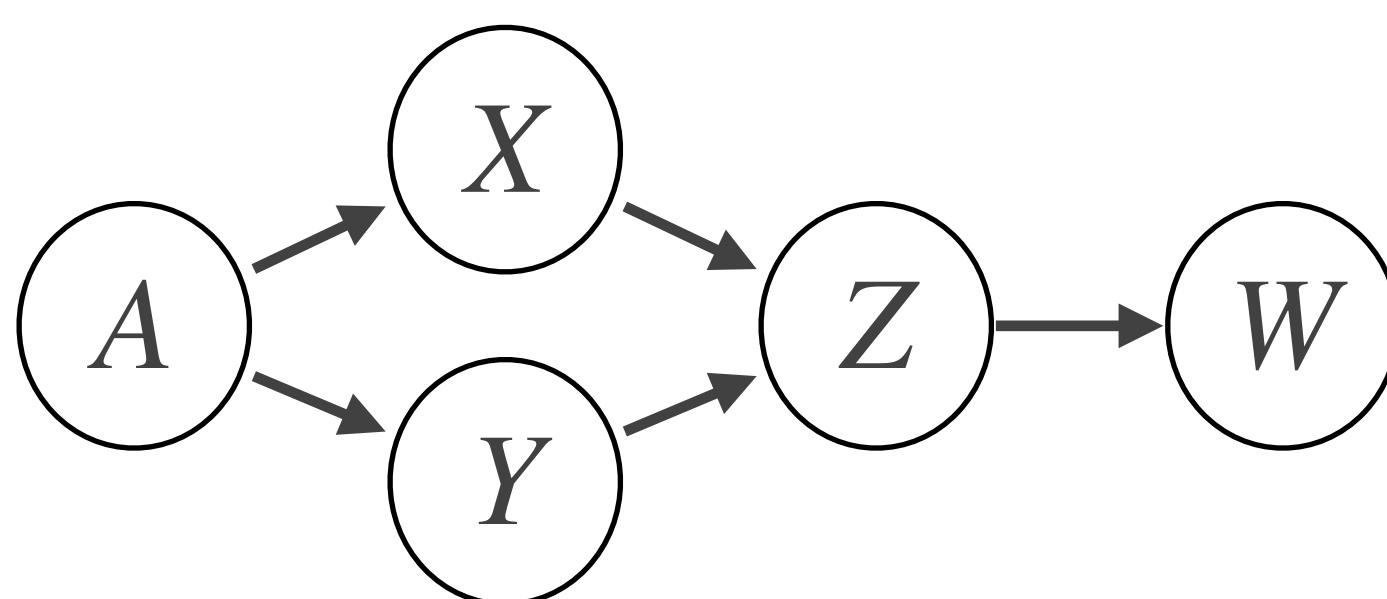
No edges of  $G$  can be removed without ceasing such a property.

**Observational Distribution**

$$P(\mathbf{V}) \doteq P_{\mathcal{M}}(\mathbf{V}) = \sum_{\mathbf{u}} \prod_{V_i \in \mathbf{V}} P(v_i | pa_i, u_i) P(\mathbf{u})$$

**Factorization obtained by Chain Rule and conditional independencies implied by the SCM  $\mathcal{M}$ .**

Edges have no causal semantics!



$$\begin{aligned} P(\mathbf{v}) &= P(w | z, x, y, a) \quad P(z | x, y, a) \quad P(x | y, a) \quad P(y | a) \quad P(a) \\ &= P(w | z) \quad P(z | x, y) \quad P(x | a) \quad P(y | a) \quad P(a) \end{aligned}$$

$$W \perp\!\!\!\perp X, Y, A | Z$$

$$A \perp\!\!\!\perp Z | X, Y$$

$$Y \perp\!\!\!\perp X | A$$

# Markov Equivalence Class

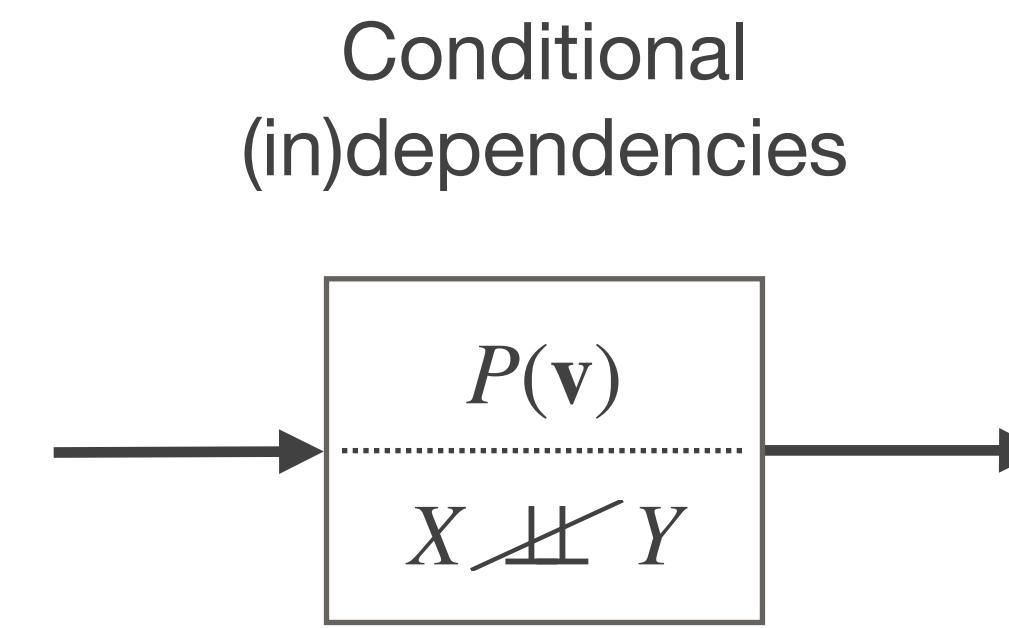
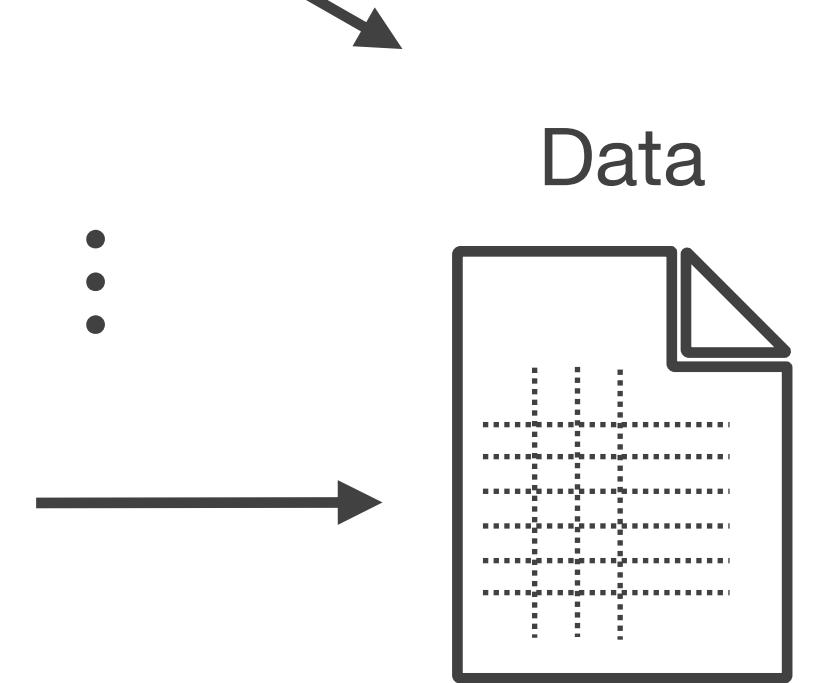
$$\mathcal{M}_1 = \begin{cases} V = \{X, Y\} \\ U = \{U_x, U_Y\} \\ \mathcal{F} = \left\{ f_X(U_X) \atop f_Y(X, U_Y) \right. \\ P(\mathbf{U}) \end{cases}$$

⋮

$$\mathcal{M}_{N-1} = \begin{cases} V = \{X, Y\} \\ U = \{U_x, U_Y, U_{X,Y}\} \\ \mathcal{F} = \left\{ f_X(Y, U_X, U_{X,Y}) \atop f_Y(U_Y, U_{X,Y}) \right. \\ P(\mathbf{U}) \end{cases}$$

⋮

$$\mathcal{M}_N = \begin{cases} V = \{X, Y\} \\ U = \{U_x, U_Y\} \\ \mathcal{F} = \left\{ f_X(U_X) \atop f_Y(U_Y) \right. \\ P(\mathbf{U}) \end{cases}$$

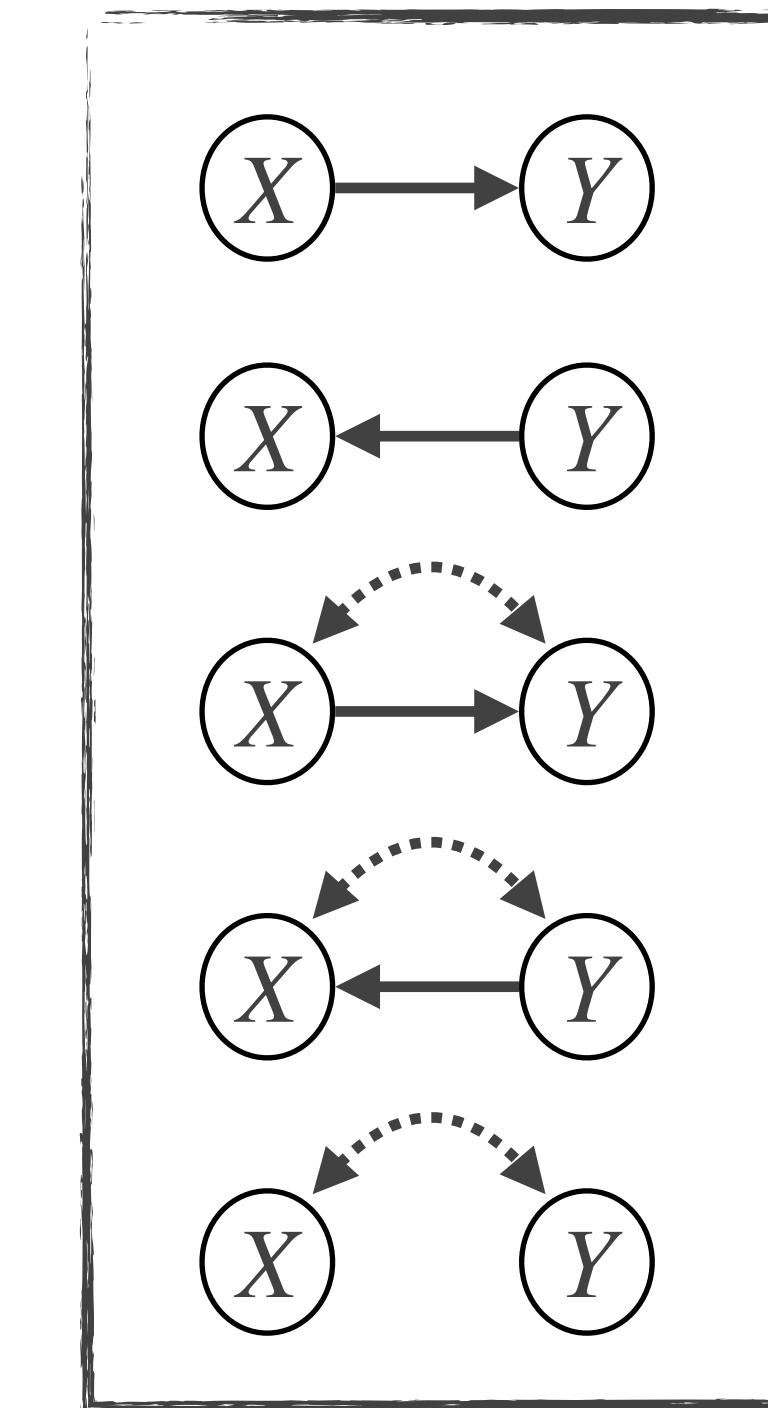


$$P(x, y) = \sum_{u_x, u_y} P(x|y)P(y)P(u_x, u_y)$$

$$P(x, y) = \sum_{u_x, u_y} P(y|x)P(x)P(u_x, u_y)$$

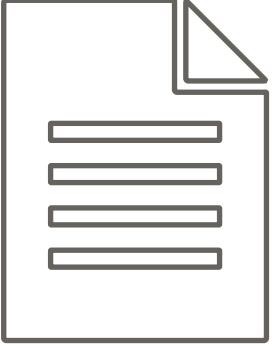
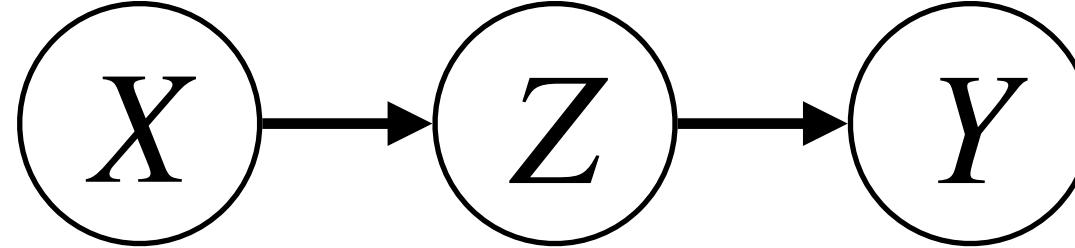
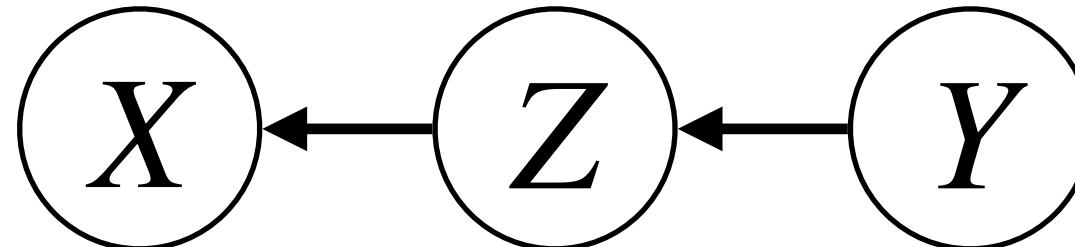
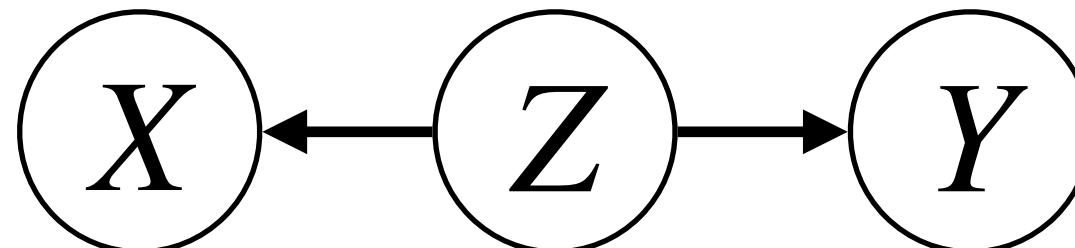
⋮

**Markov Equivalence Class**  
(class of models implying the same set of conditional independencies)



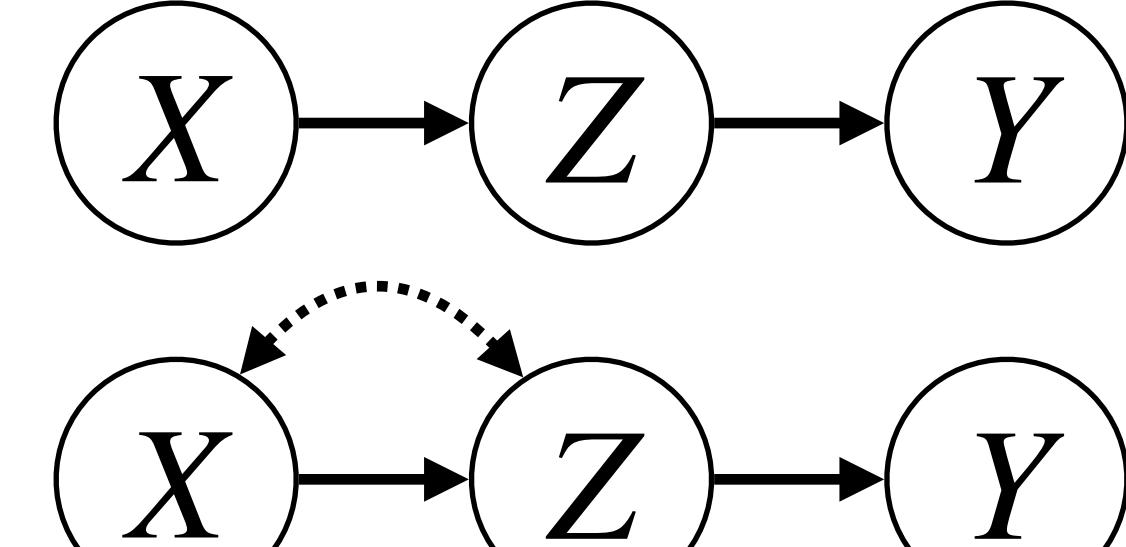
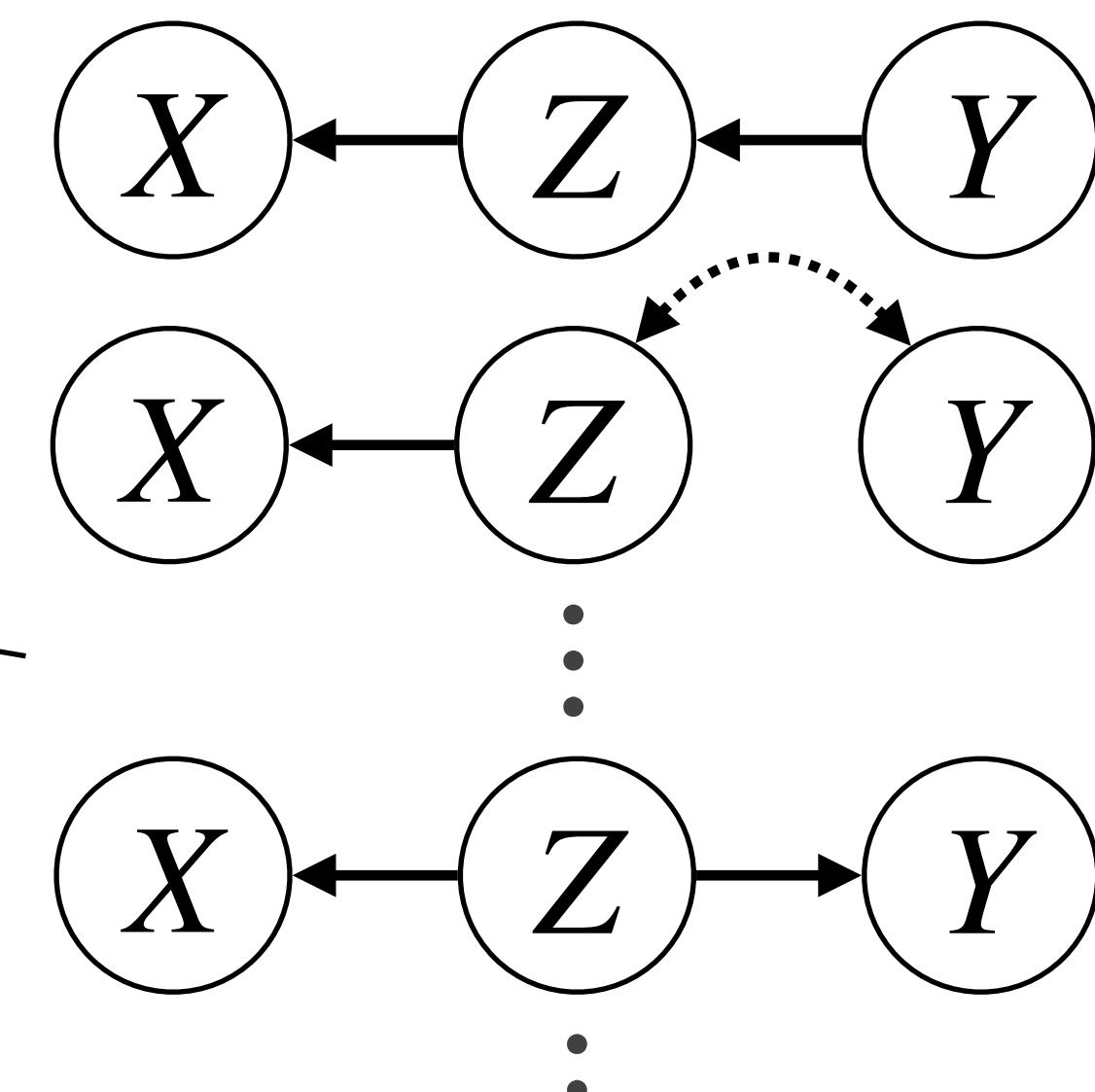
**Correlation does not imply causation!**

# Equivalent Bayesian Networks

Distribution	Factorization	Bayesian Networks
 Observational Data	$P(X, Y, Z)$  with $P(Y X, Z) = P(Y X)$ i.e., $X \perp\!\!\!\perp Y Z$	$P(x, y, z) = P(y x, z)P(z x)P(x)$ $= P(y z)P(z x)P(x)$
	$P(x, y, z) = P(x y, z)P(y z)P(z)$ $= P(x z)P(z y)P(y)$	  
	$P(x, y, z) = P(y x, z)P(x z)P(z)$ $= P(y z)P(x z)P(z)$	<p>Markov Equivalent</p>

Two models are considered **Markov equivalent** if they imply the same conditional independencies.

# Equivalent Bayesian Networks

Distribution	Factorization	Bayesian Networks
$P(X, Y, Z)$ with $P(Y X, Z) = P(Y X)$ i.e., $X \perp\!\!\!\perp Y Z$	$P(x, y, z) = P(y x, z)P(z x)P(x)$ $= P(y z)P(z x)P(x)$	
<b>Invariance:</b> Z is <b>never</b> a collider (either ancestor of X and Y).	$P(x, y, z) = P(y x, z)P(x z)P(z)$ $= P(y z)P(x z)P(z)$	 <p style="text-align: right;">Markov Equivalent</p>

# Equivalent Bayesian Networks

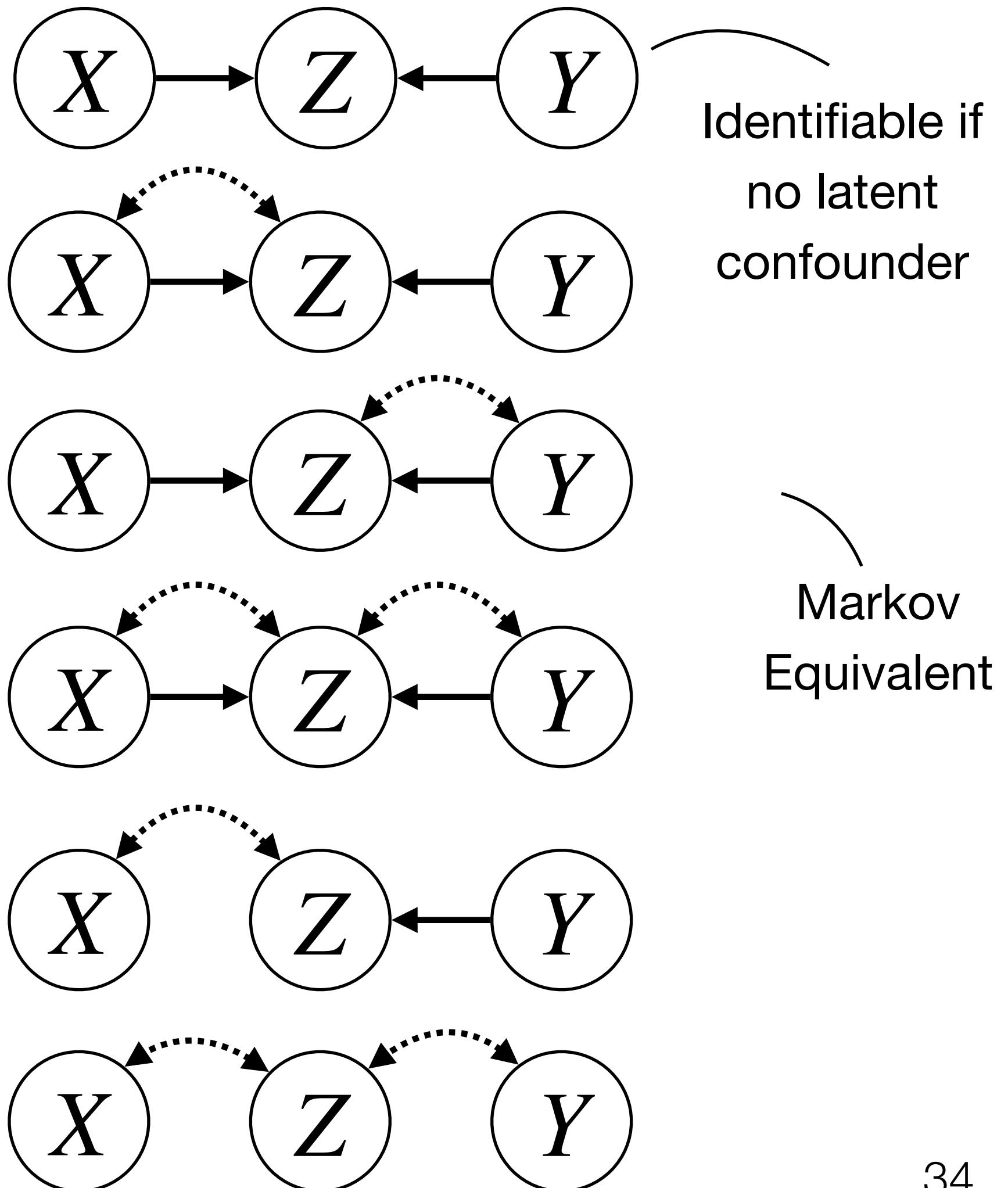
## Distribution

$P(X, Y, Z)$   
with  $P(Y|X) = P(Y)$   
i.e.,  $X \perp\!\!\!\perp Y$

## Factorization

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

## Bayesian Networks



# Equivalent Bayesian Networks

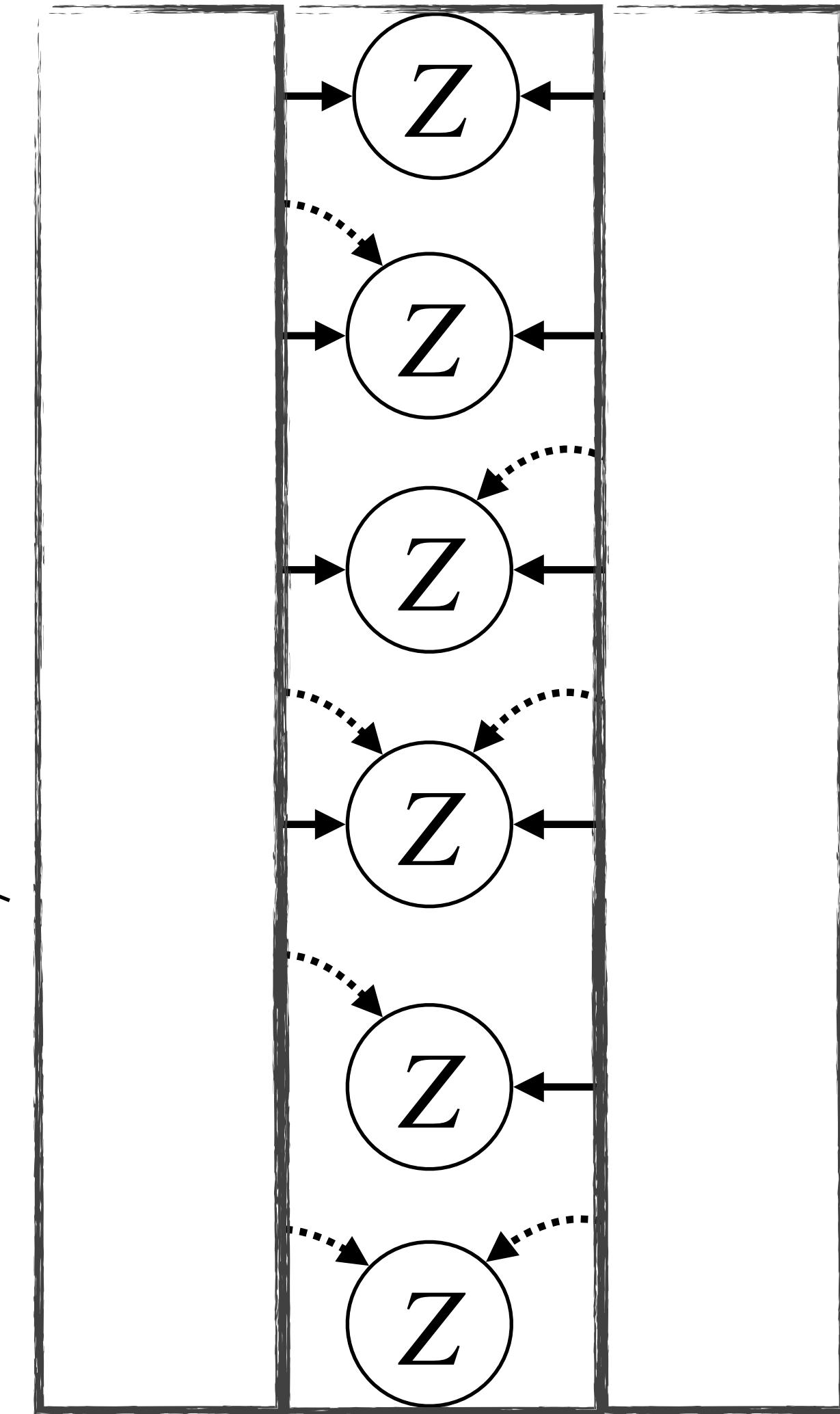
## Distribution

$P(X, Y, Z)$   
with  $P(Y|X) = P(Y)$   
i.e.,  $X \perp\!\!\!\perp Y$

## Factorization

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

## Bayesian Networks



Identifiable if  
no latent  
confounder

Markov  
Equivalent

**Invariance:**  
Z is **always** a collider  
(non-ancestor of  $X$  and  $Y$ ).

# Causal Bayesian Network

A DAG, possibly with latent confounders (ADMG),  
representing the **causal and confounding relationships**  
implied by an SCM

# CBN: Encoder of Structural Causal Knowledge

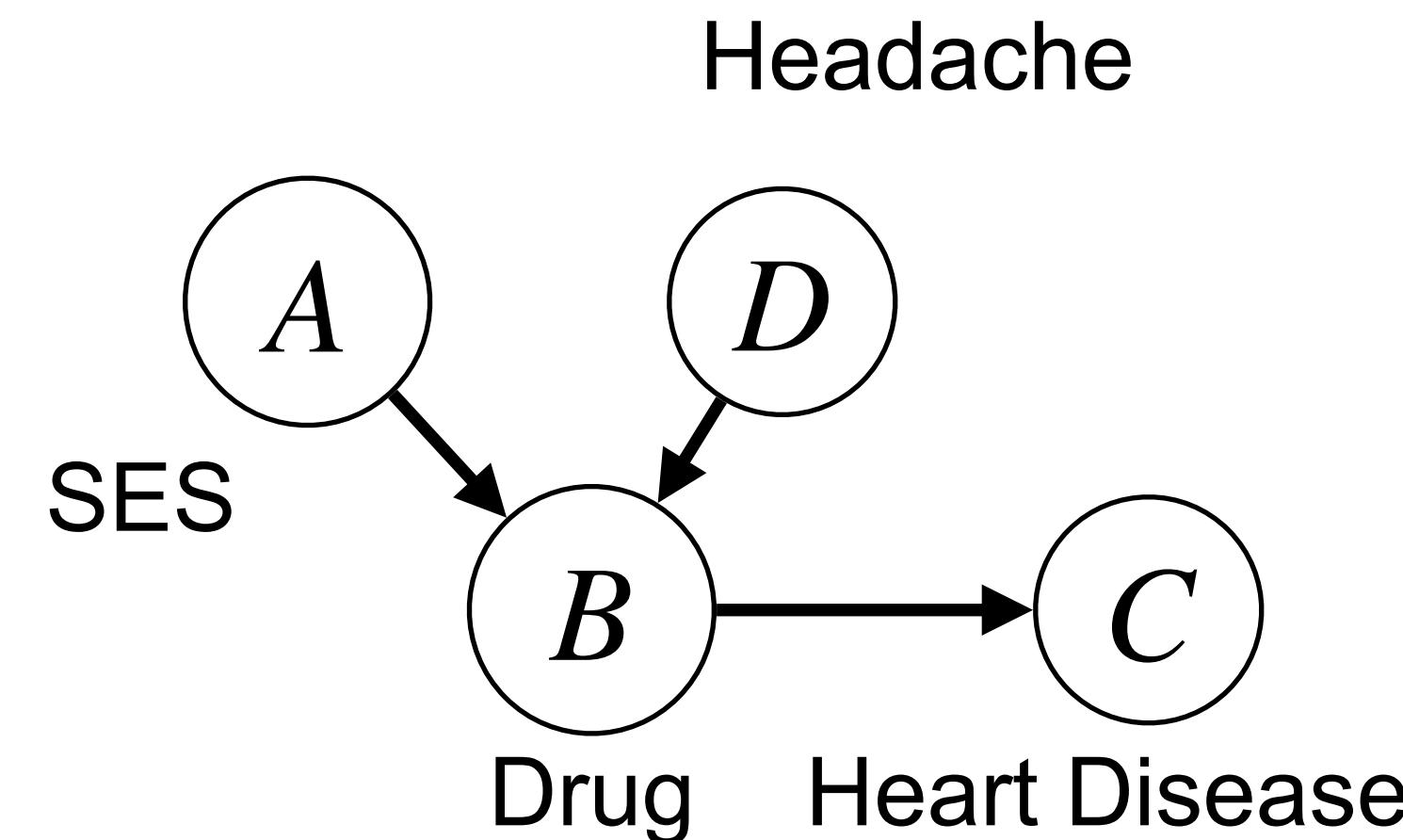
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

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  $f_j \in \mathcal{F}$ .

# CBN: Encoder of Structural Causal Knowledge

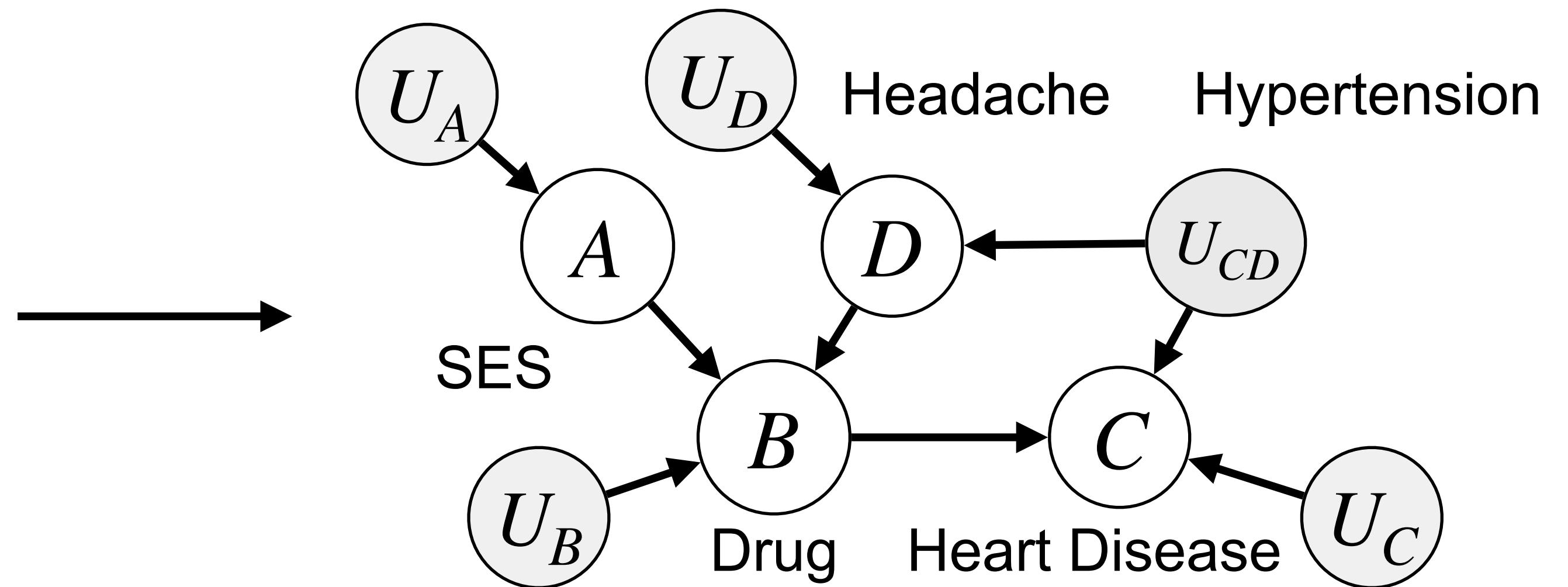
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

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  $f_j \in \mathcal{F}$ .

# CBN: Encoder of Structural Causal Knowledge

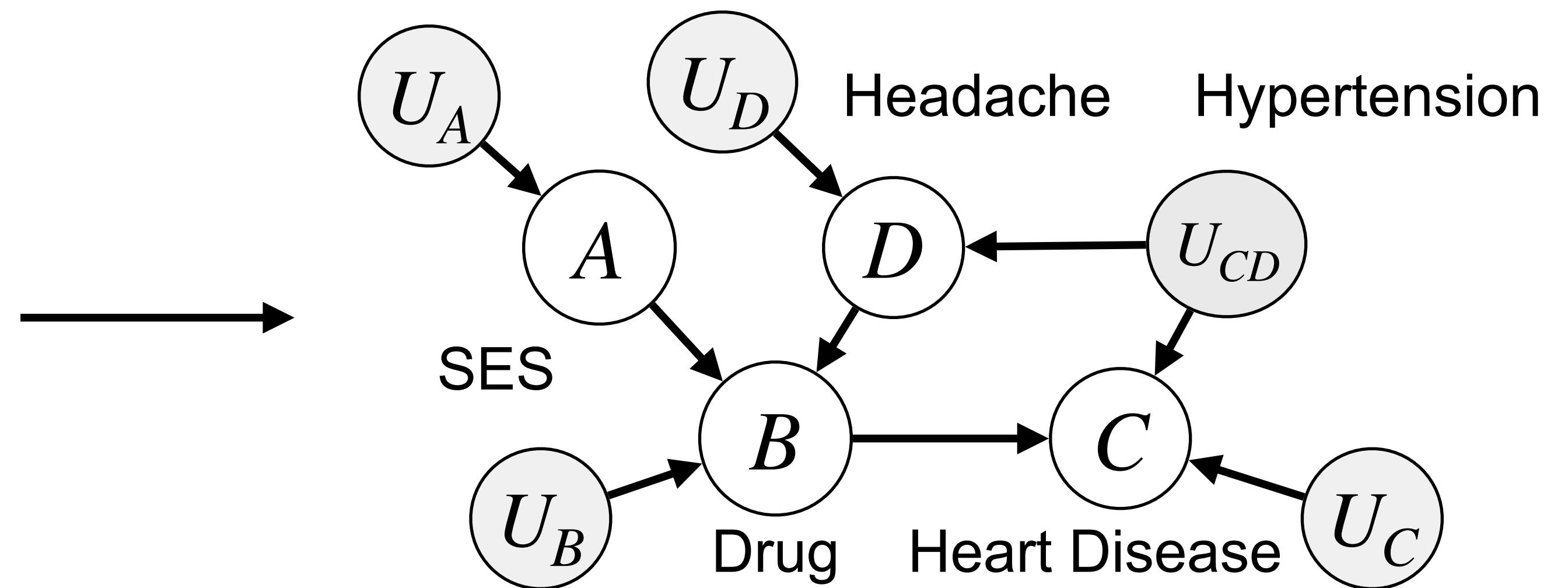
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

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  $U \in \mathbf{U}$ .

# CBN: Encoder of Structural Causal Knowledge

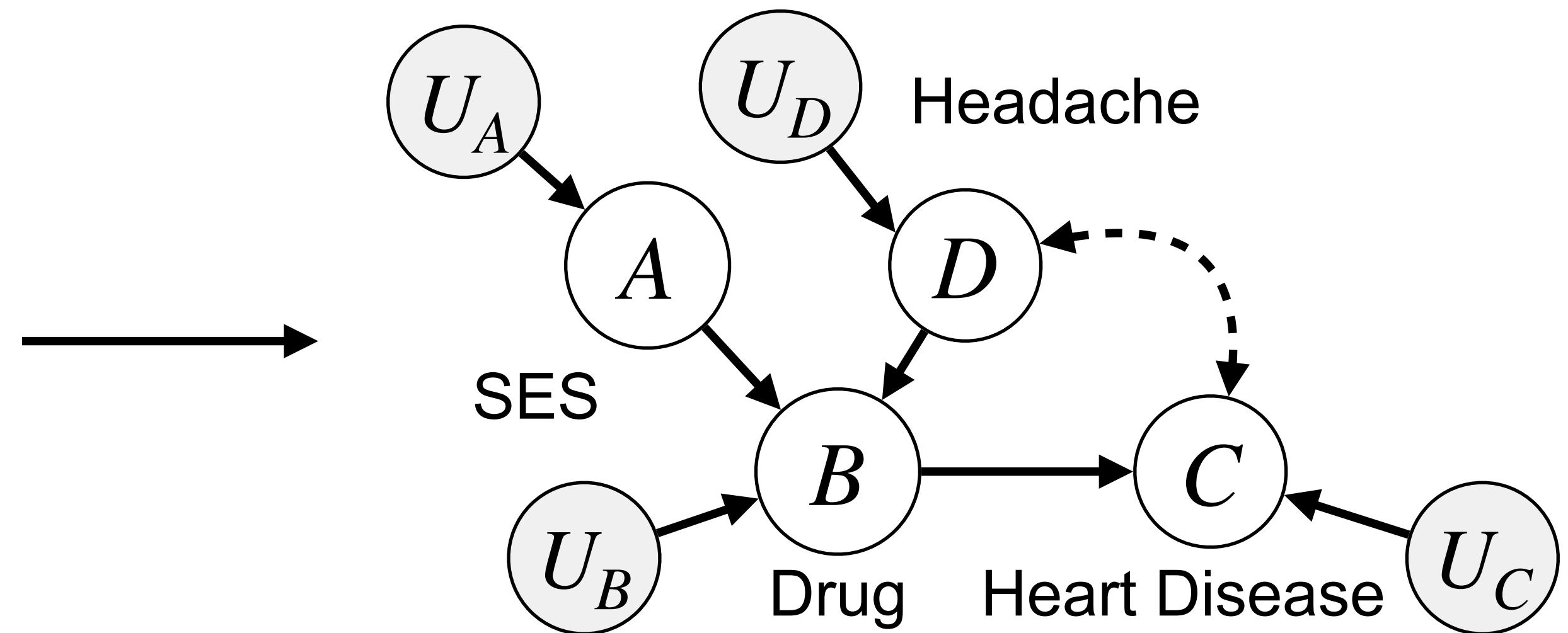
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

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  $U \in \mathbf{U}$ .

# CBN: Encoder of Structural Causal Knowledge

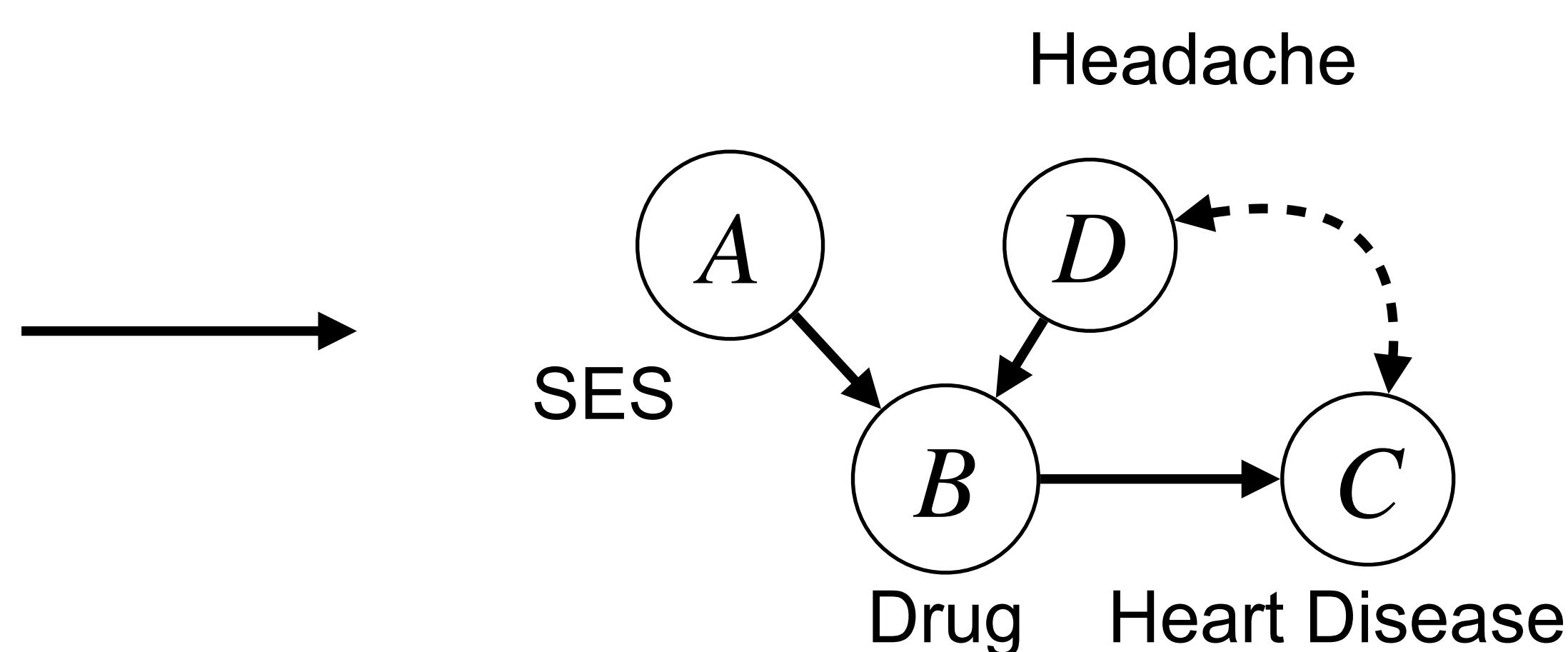
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

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  $U \in \mathbf{U}$ .

# CBN: Encoder of Structural Causal Knowledge

---

Let  $\mathbf{P}_*$  be the collection of all interventional distributions  $P(\mathbf{V} \mid do(\mathbf{x}))$ ,  $\mathbf{X} \subseteq \mathbf{V}$ , including the null (observational) distribution  $P(\mathbf{V})$ .

An Acyclic Directed Mixed Graph (ADMG)  $G$  is a CBN for  $\mathbf{P}_*$  if for every intervention  $do(\mathbf{X} = \mathbf{x})$ ,  $\mathbf{X} \subseteq \mathbf{V}$ , it holds:

**Interventional  
Distribution**

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

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

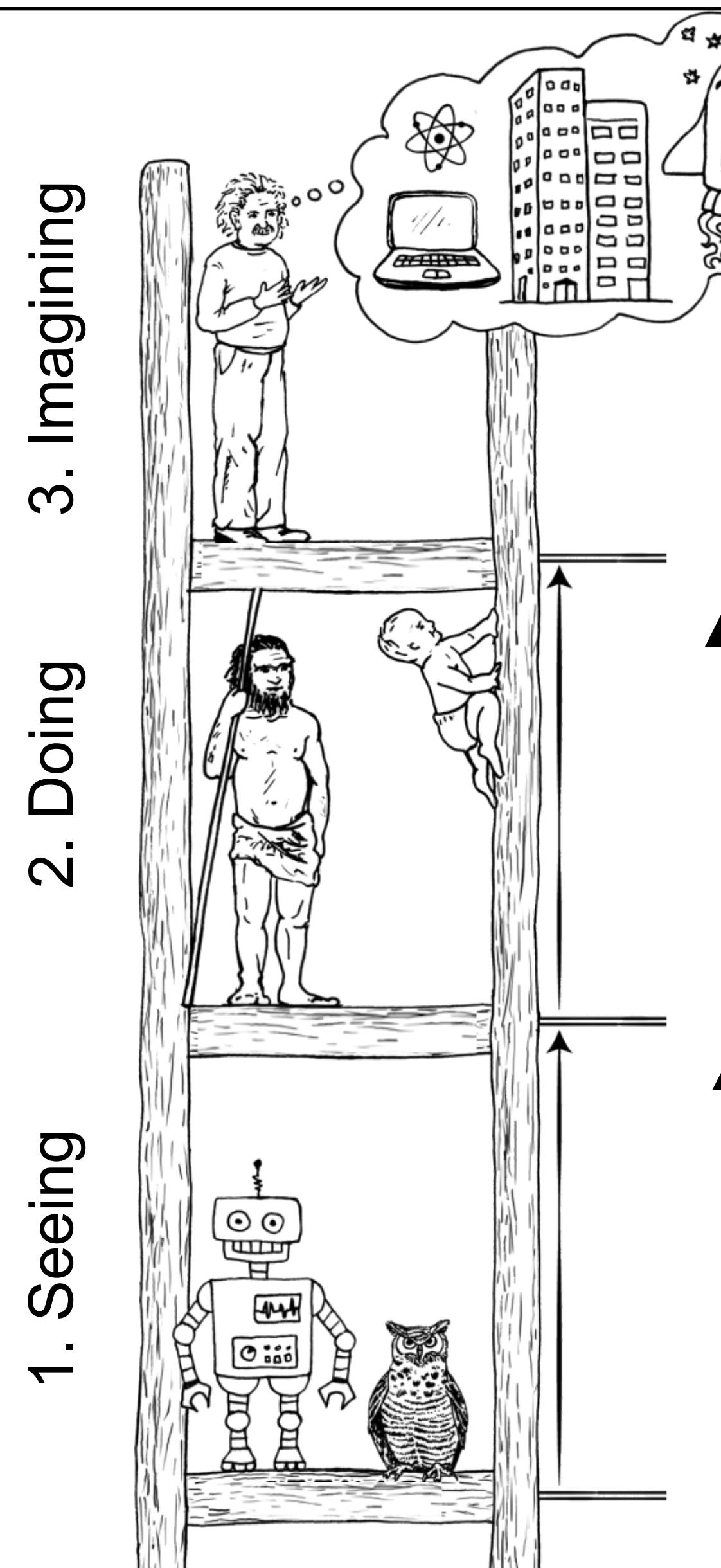
**Truncated factorization  
implied by the SCM  $\mathcal{M}_x$ .**

Semi-Markov relative to  $G_{\overline{\mathbf{X}}}$

# **Pearl's Causal Hierarchy (PCH)**

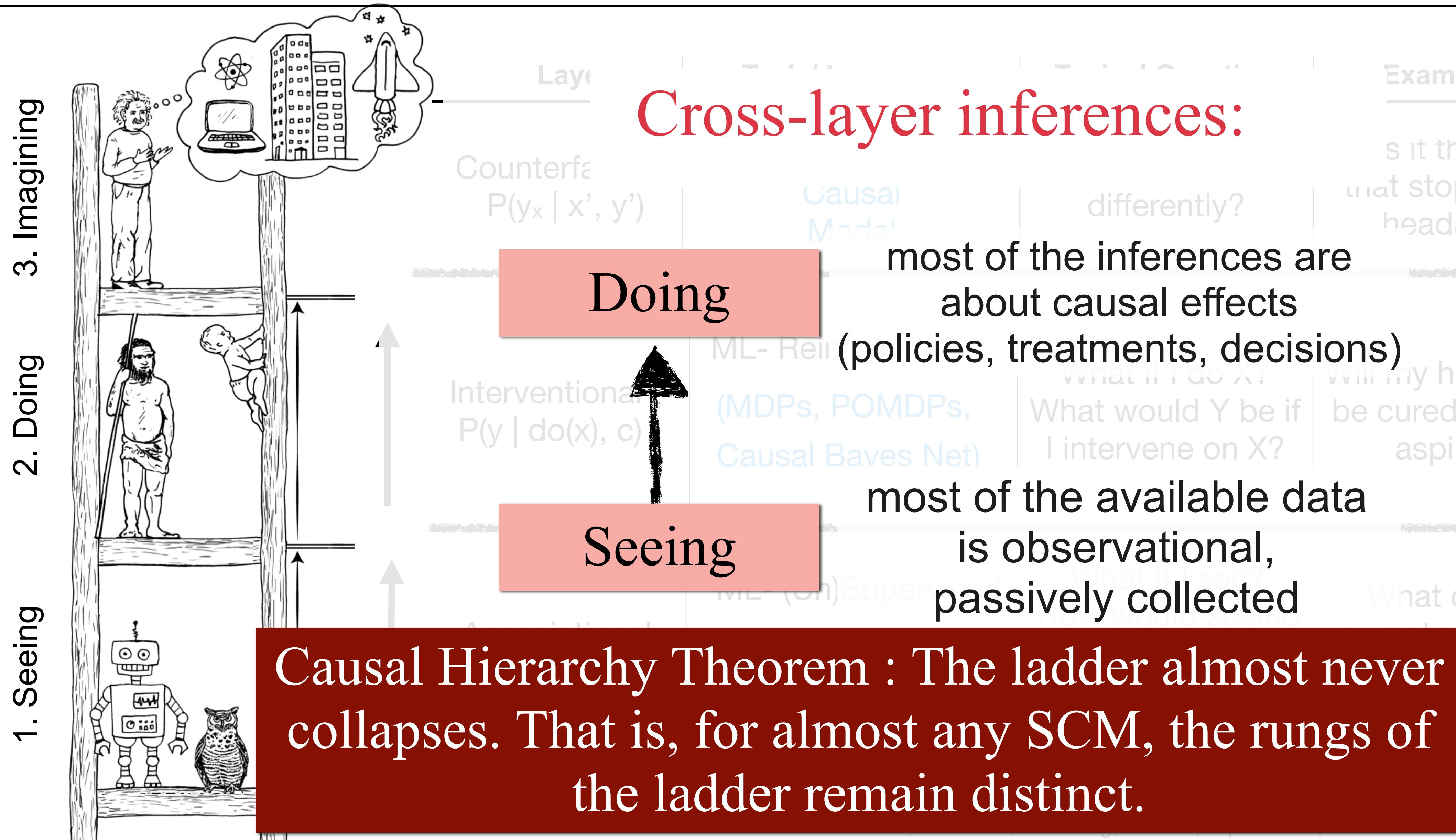
**The Three Inferential Layers**

# Ladder of Causation

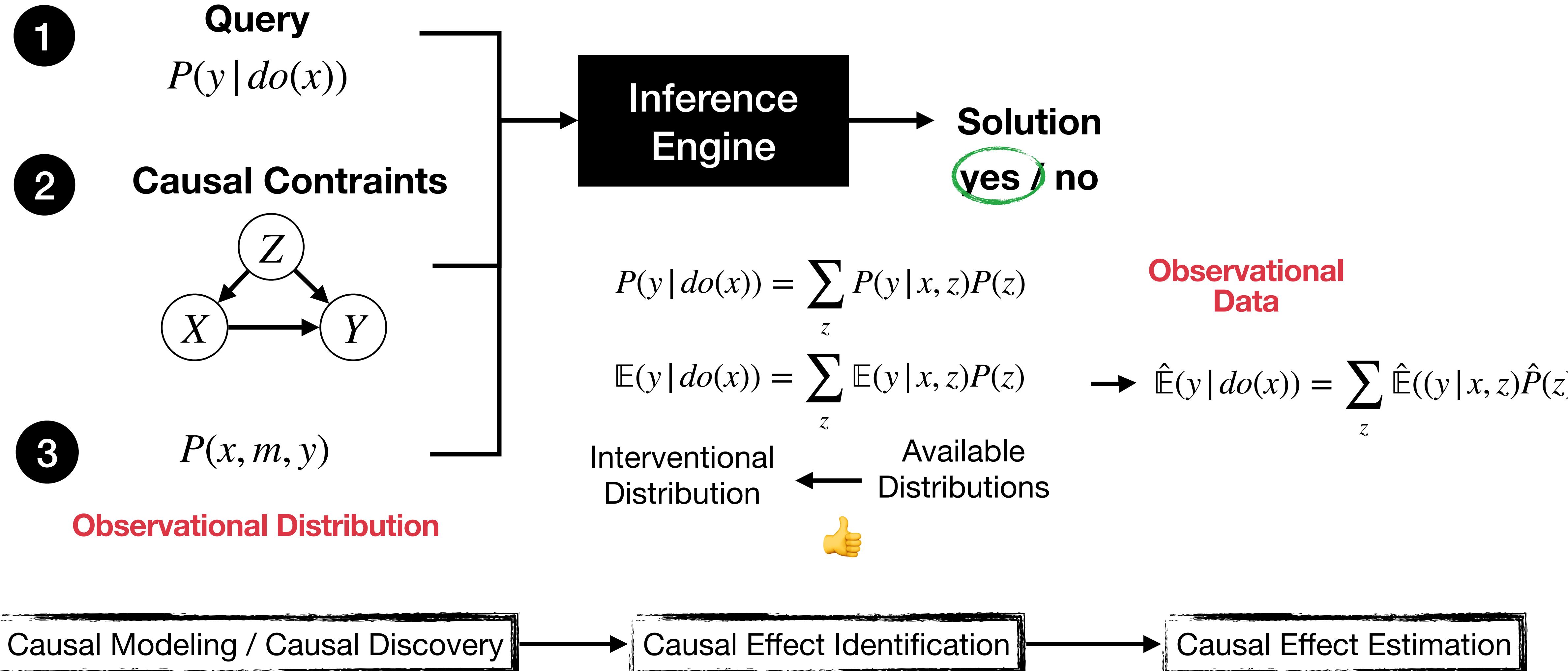


	Layer	Task / Language	Typical Question	Examples
3. Imagining	Counterfactual $P(y_x   x', y')$	Structural Causal Model	What if I had acted differently?	Was it the aspirin that stopped my headache?
2. Doing	Interventional $P(y   \text{do}(x), c)$	ML- Reinforcement (Causal Bayes Net)	What if I do X? What would Y be if I intervene on X?	Will my headache be cured if I take aspirin?
1. Seeing	Associational $P(y   x)$	ML- (Un)Supervised (Bayesian Networks, 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?

# Ladder of Causation



# Causal Pipeline from a Causal Diagram



# **Causal Effect Identification from Causal Diagrams**

# 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,

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

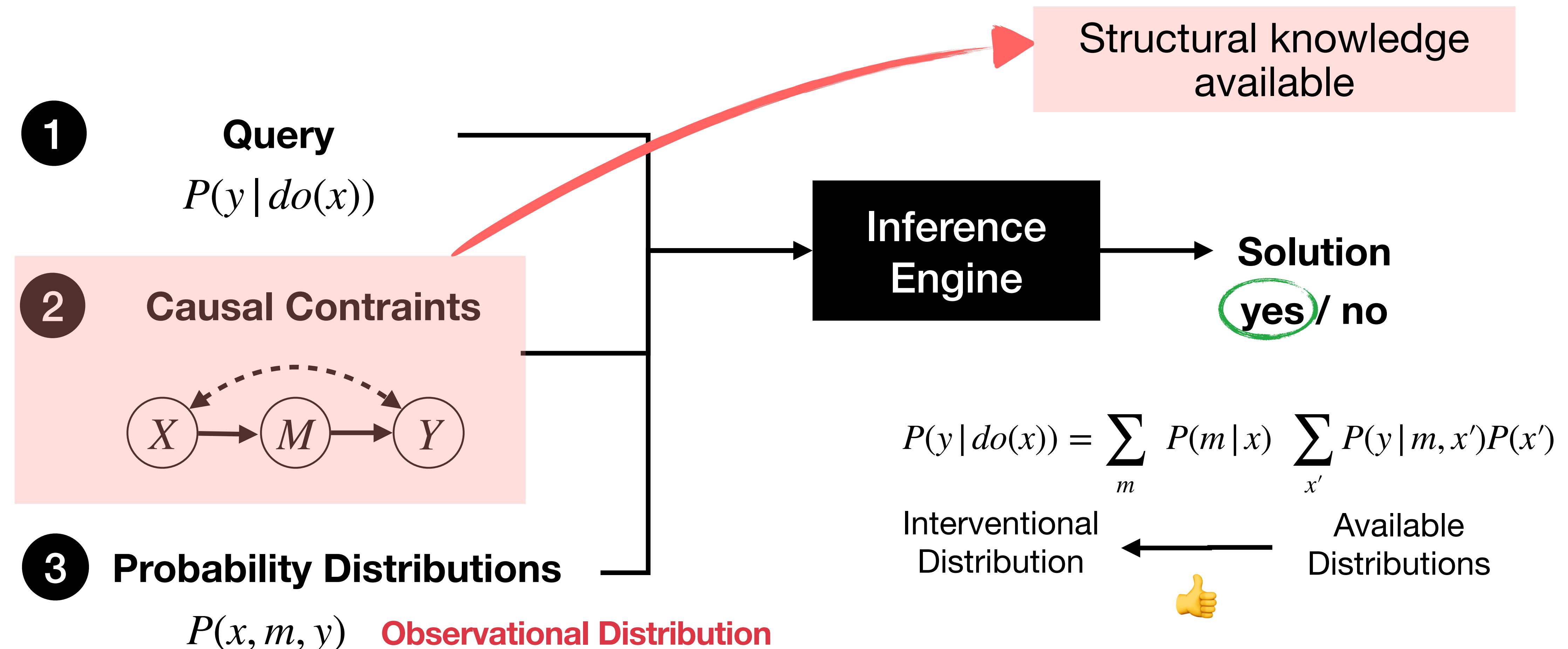
Jacobian of  $\mathbb{E}[\mathbf{Y} | do(\mathbf{X} = \mathbf{x})]$ , where

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

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.  $do(\mathbf{X} = \mathbf{x})$

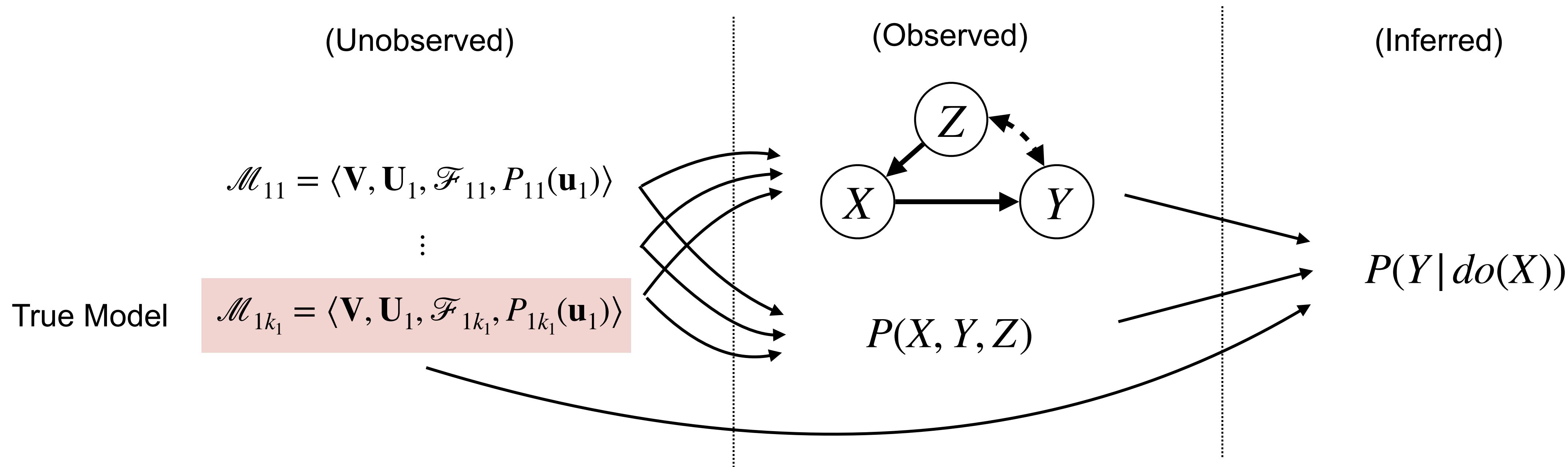
# Classical Causal Effect Identification



- Tian, J. and Pearl, J. (2002) 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

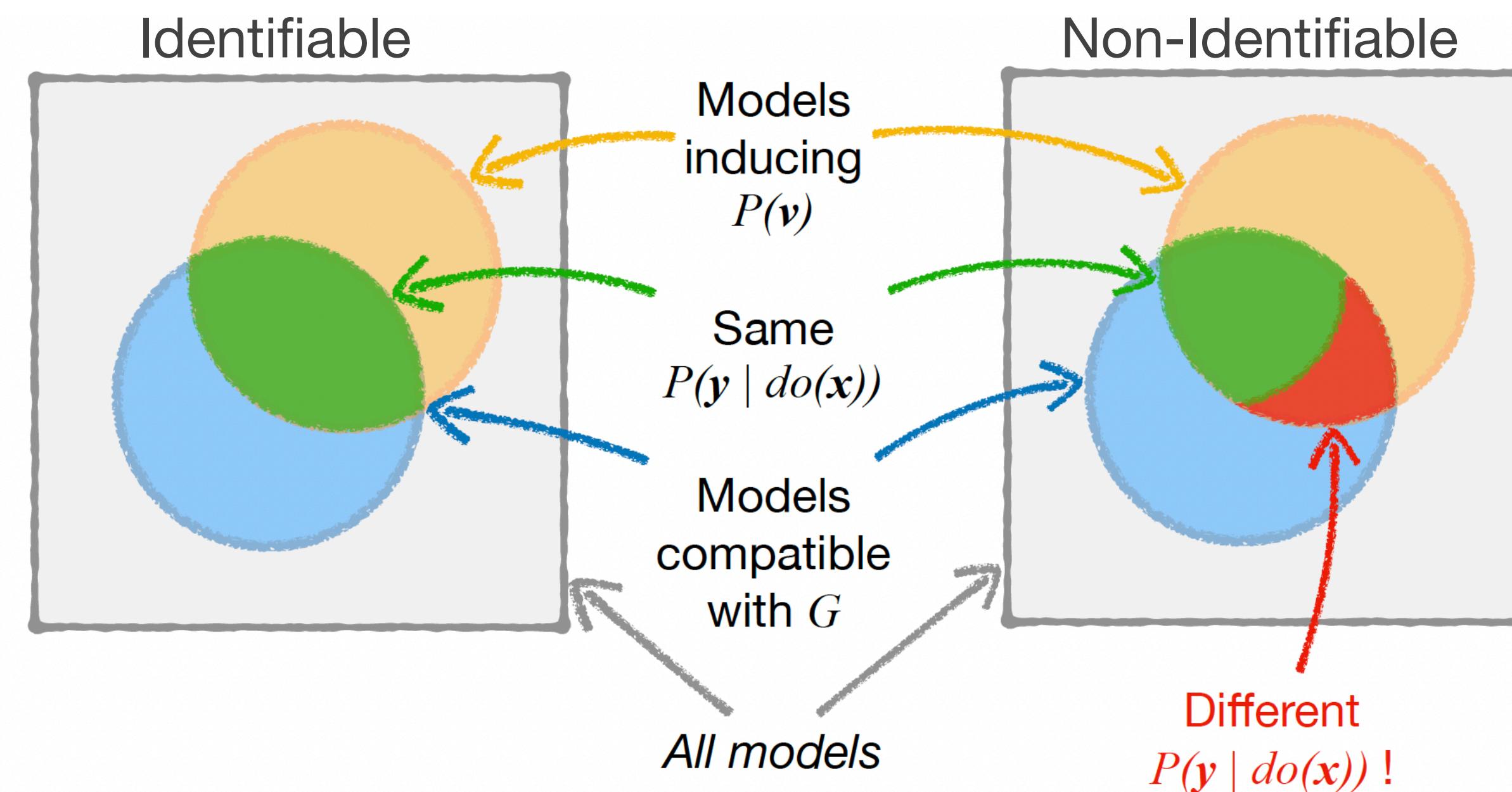
**Causal Effect Identifiability:** The causal effect of a (set of) treatment variable(s)  $\mathbf{X}$  on a (set of) outcome variable(s)  $\mathbf{Y}$  is said to be identifiable from a causal diagram  $G$  and the probability of the observed variables  $P(\mathbf{V})$  if the interventional distribution  $P(\mathbf{Y} | do(\mathbf{X}))$  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  $P(\mathbf{y} | do(\mathbf{x}))$ .

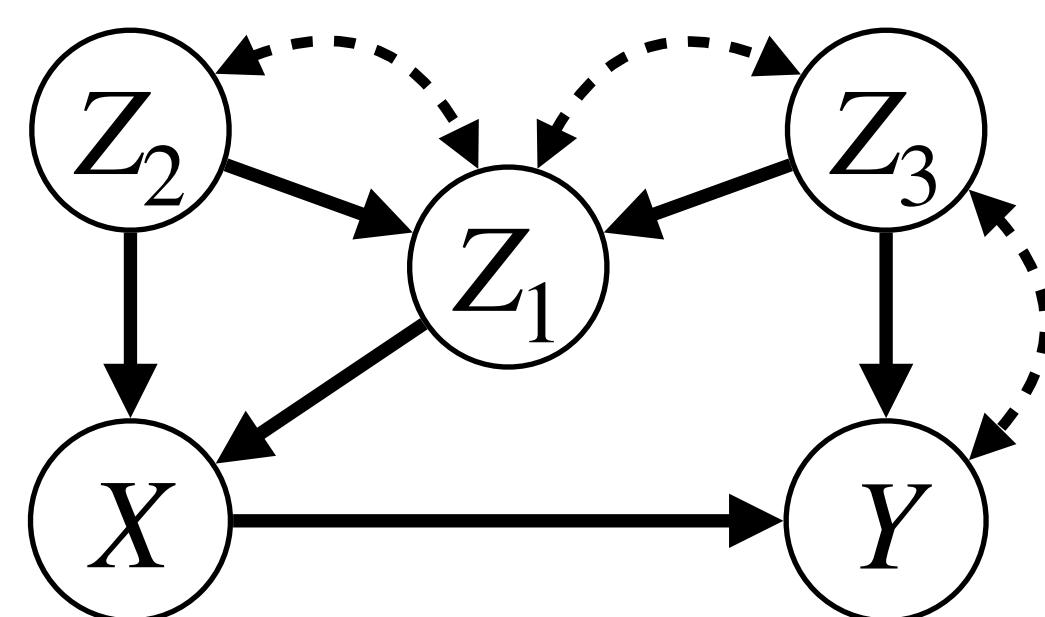
# Identification Via Adjustment over Parents

Let  $G$  be a causal graph with **all parents observed**.

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

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

Proof follows from the truncated factorization for Markovian models.  
Try at home!



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

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

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

# Identification via Backdoor Criterion

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:

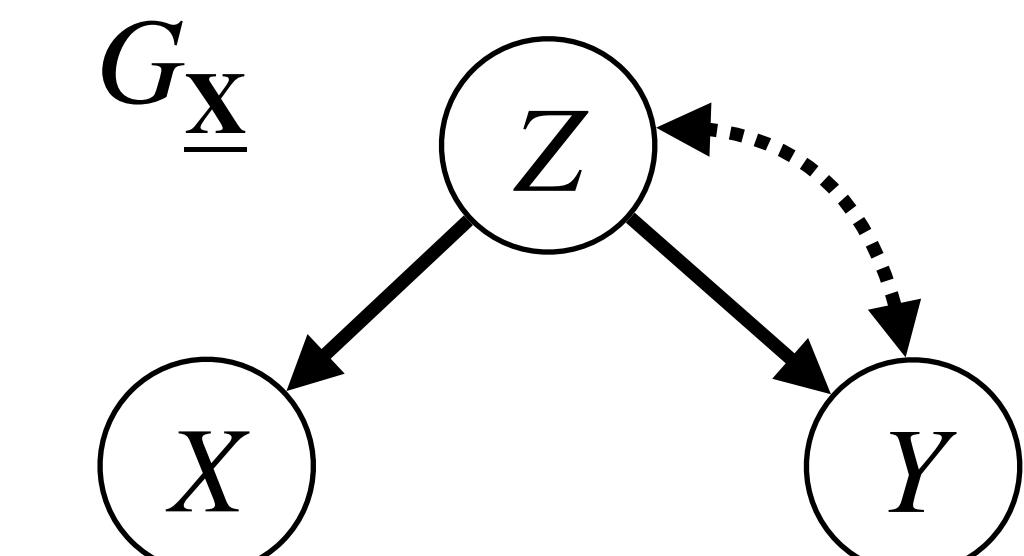
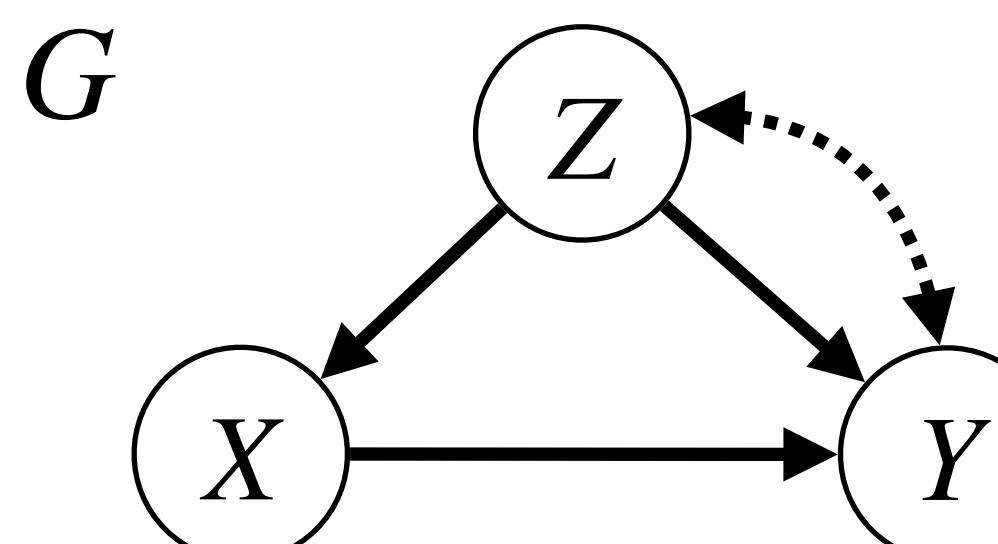
1.  $\mathbf{Z}$  d-separates  $\mathbf{X}$  and  $\mathbf{Y}$  in the graph  $G_{\underline{\mathbf{X}}}$ , i.e., the graph resulting from cutting the arrows out of  $\mathbf{X}$
2. no node in  $\mathbf{Z}$  is a descendant of a variable  $X \in \mathbf{X}$  in  $G$  (all variables in  $\mathbf{Z}$  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:

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

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

$\mathbf{Z}$ , a set of covariates, admissible for backdoor adjustment



In  $G_{\underline{\mathbf{X}}}$ , all non-backdoor paths are severed

# Identification via Backdoor Criterion

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:

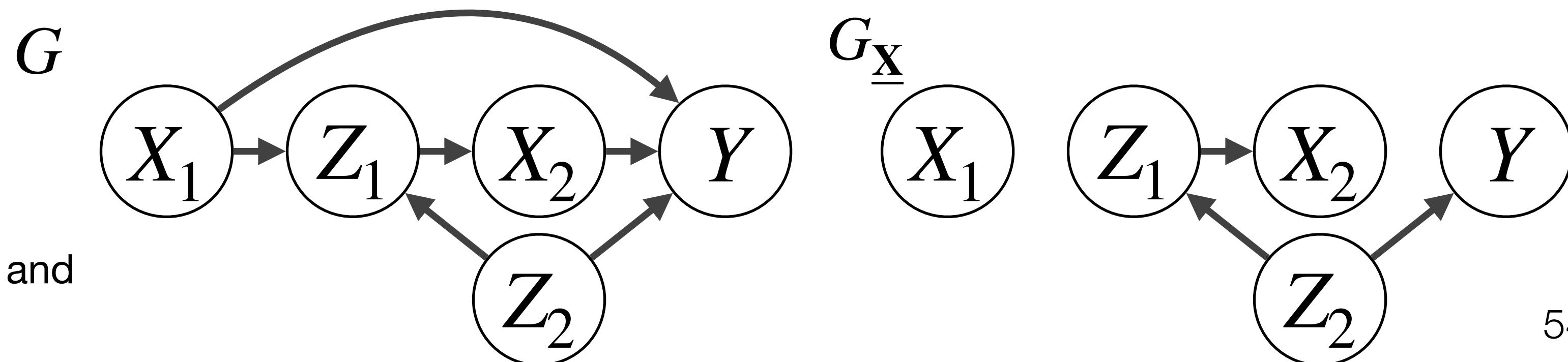
1.  $\mathbf{Z}$  d-separates  $\mathbf{X}$  and  $\mathbf{Y}$  in the graph  $G_{\underline{\mathbf{X}}}$ , i.e., the graph resulting from cutting the arrows out of  $\mathbf{X}$
2. no node in  $\mathbf{Z}$  is a descendant of a variable  $X \in \mathbf{X}$  in  $G$  (all variables in  $\mathbf{Z}$  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:

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

$$\begin{aligned}\mathbf{X} &= \{X_1, X_2\} \\ \mathbf{Y} &= \{Y\} \\ \mathbf{Z} &= \{Z_1, Z_2\}\end{aligned}$$

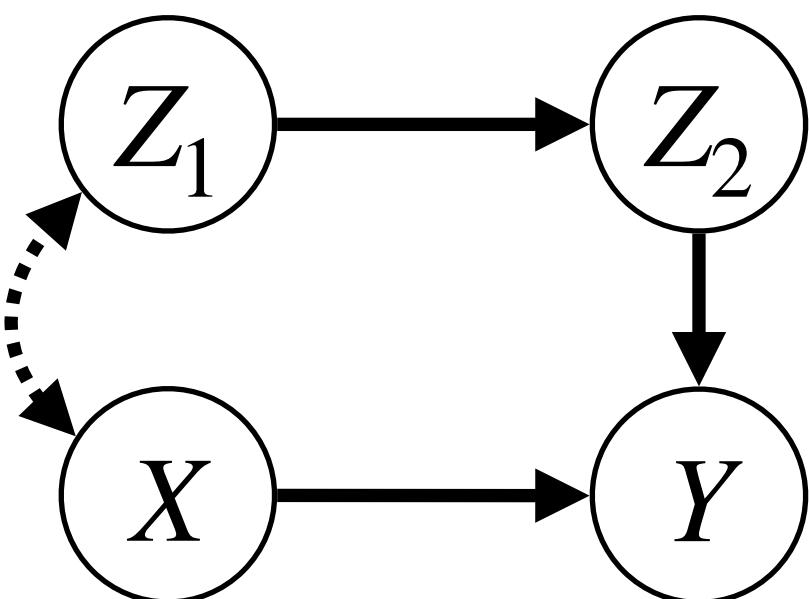
$\mathbf{Z}$ , a set of covariates, admissible for backdoor adjustment



# Admissible Sets for BD Adjustment

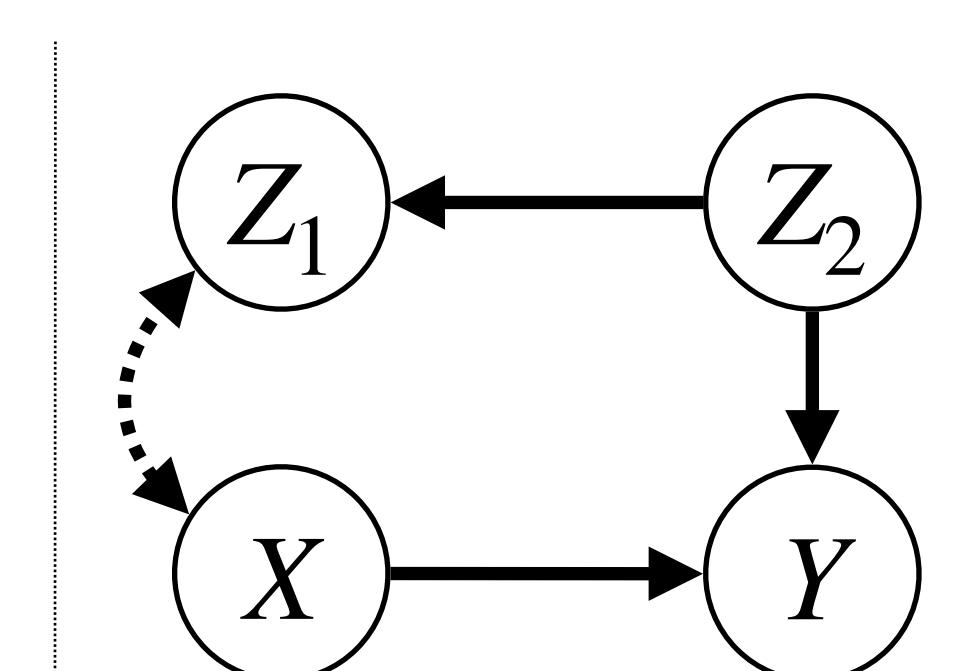
$Z$  satisfies the **backdoor criterion** for or  $(X, Y)$  in the causal graph  $G$  if:

1.  $Z$  d-separates  $X$  and  $Y$  in the graph  $\underline{G}_X$ , i.e., the graph resulting from cutting the arrows out of  $X$
2. no node in  $Z$  is a descendant of a variable  $X \in X$  in  $G$  (all variables in  $Z$  are pre-treatment)



Minimal BD  
Adjustment Sets  $\left\{ \begin{array}{l} \{Z_1\}, \\ \{Z_2\}, \\ \{Z_1, Z_2\} \end{array} \right.$

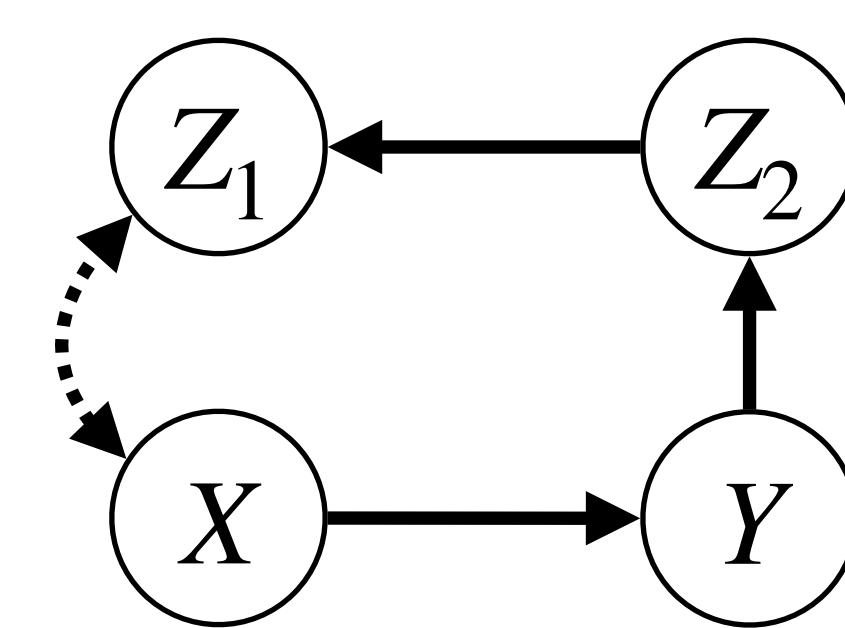
$$P(y|do(x)) = \sum_{z_1} P(y|x, z_1) P(z_1)$$



$\{\}$ ,

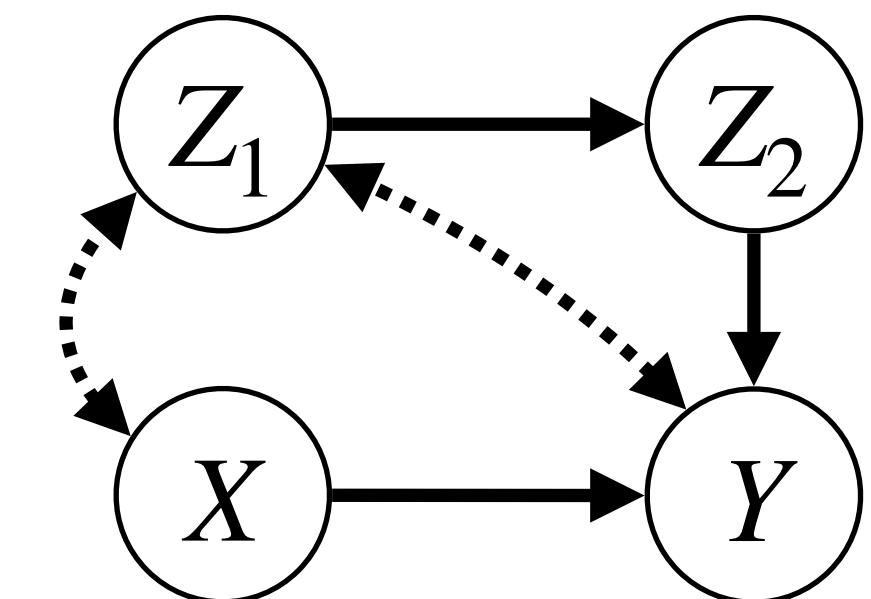
$\{Z_2\}$ ,  
 $\{Z_1, Z_2\}$

$$P(y|do(x)) = P(y|x)$$



$\{\}$

$$P(y|do(x)) = P(y|x)$$



There is no BD  
Adjustment Set!

$P(y|do(x))$  is  
non-identifiable

The screenshot shows three parallel causal fusion interfaces, each with a summary panel, editor panel, and a central graph area.

**Summary Panel:**

- Treatment:  $X$
- Outcome:  $Y$
- Adjusted:
- Query:  $P(Y|do(X))$

**Editor Panel:**

- Graphical tab (selected)
- Structural tab
- Nodes and Edges definitions (repeated across three panels):
  - <NODES>
  - 1 X -45,-15
  - 2 Y 45,-15
  - 3 Z 0,-60
  - <EDGES>
  - 7 X -> Y
  - 8 Z -> X
  - 9 Z -> Y
- Populations tab
- Compute button

**Central Graph Area:**

```

    graph LR
        X((X)) --> Y((Y))
        Z((Z)) -.-> Y
        Z((Z)) -.-> Y
    
```

**Right-hand Side Panels:**

- Confounding Analysis
  - Admissible Sets
  - Admissibility Test
  - Instrumental Variables
  - IV Admissibility Test
- Path Analysis
  - D-Separation
  - Causal Paths
  - Confounding Paths
  - Biassing Paths
- Do-Calculus Analysis
  - Do-Inspector
  - Do-Separation
- $\sigma$ -Calculus Analysis
  - $\sigma$ -Inspector
  - $\sigma$ -Separation
- Load, Estimation, Derivation, Remove buttons

**Bottom Input/Output:**

The causal effect of  $X$  on  $Y$  conditional on  $Z$  with do :  $\equiv$  (Query:  $P(Y|do(X))$  from  $P(v)$ )

Non-Parametric  Clear

$P(Y|do(X)) = \sum_Z P(Y|X, Z) P(Z)$

causalfusion.net/app

### Fusion<sup>(β)</sup>

Summary

Treatment :  $X$   
Outcome :  $Y$   
Adjusted :  
Query :  $P(Y|do(X))$

Show More Details

Editor

Graphical Structural Refresh

```
1 <NODES>
2 X -45,-15
3 Y 45,-15
4 Z 0,-60
5
6 <EDGES>
7 X -> Y
8 Z -> X
9 Z -> Y
```

Populations

Confounding Analysis

Admissible Sets  
Admissibility Test  
Instrumental Variables  
IV Admissibility Test

Path Analysis

D-Separation  
Causal Paths  
Confounding Paths  
Biasing Paths

Do-Calculus Analysis

Do-Inspector  
Do-Separation

$\sigma$ -Calculus Analysis

$\sigma$ -Inspector  
 $\sigma$ -Separation

Compute The causal effect of  $X$  on  $Y$  conditional on  with do :  $\equiv$  (Query:  $P(Y|do(X))$  from  $P(v)$ ) Non-Parametric  Clear

1  $P(Y|do(X))$  is not identifiable from  $P(X, Y, Z)$ .

Load Remove

```
graph LR; X((X)) -- "solid" --> Y((Y)); X((X)) -.-> Z((Z)); Z((Z)) -.-> X((X)); Z((Z)) -.-> Y((Y)); X((X)) -.-> Y((Y))
```

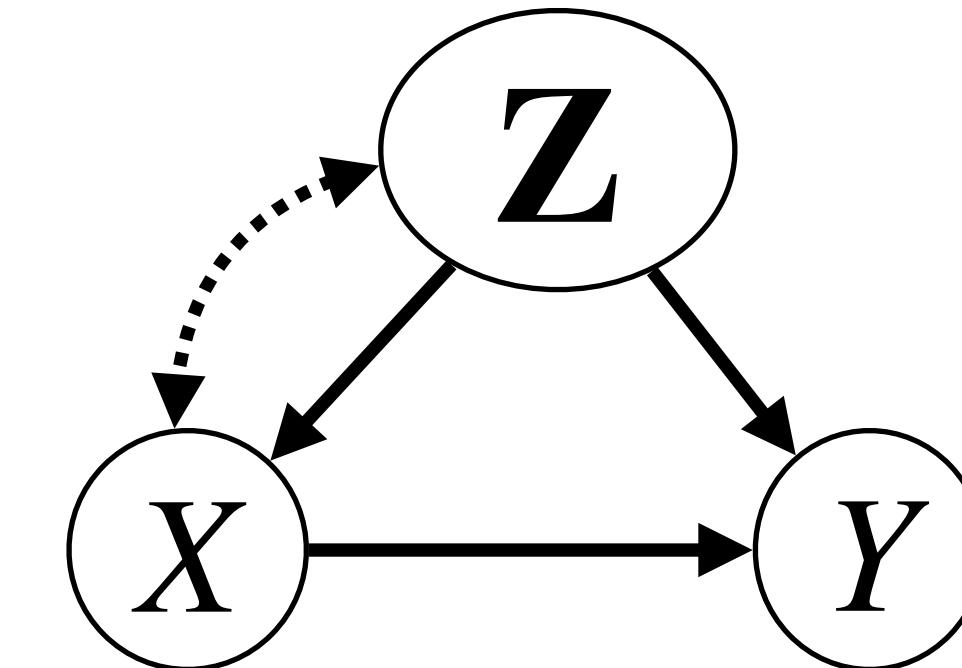
# Counterfactual Interpretation of Backdoor

Theorem 4.3.1, Pearl's Primer Book

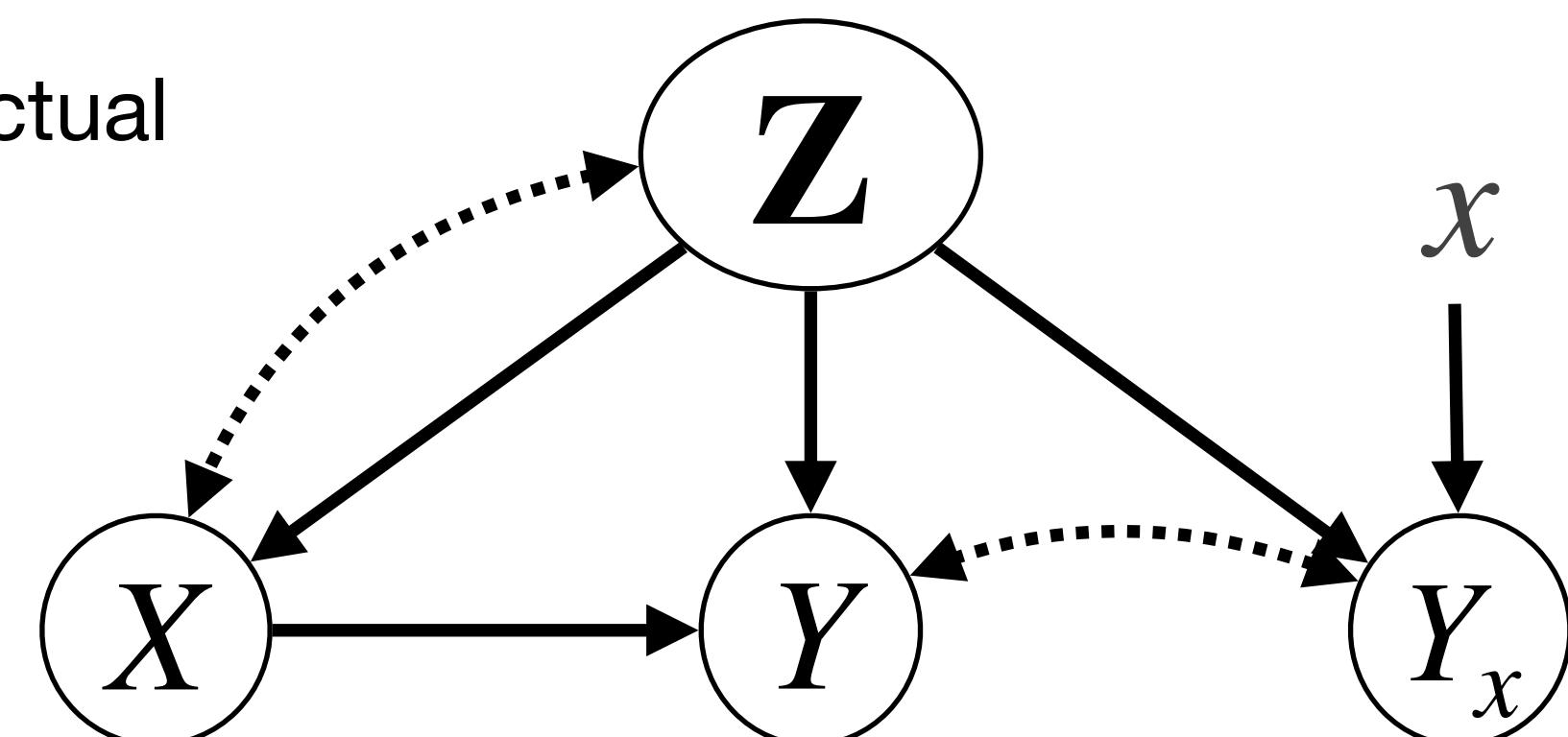
**Theorem:** If a set  $Z$  satisfies the *backdoor criterion* w.r.t. the ordered pair  $(X, Y)$ , then, for all  $x$ , it holds that  $Y_x \perp\!\!\!\perp X | Z$ .

Although the satisfiability of  $Z$  to the *backdoor criterion* can be tested given a causal diagram or a PAG, the condition  $Y_x \perp\!\!\!\perp X | Z$  is sometimes framed as an assumption, referred to as **(conditional) ignorability, exchangeability or unconfoundedness**.

Observational  
Graph

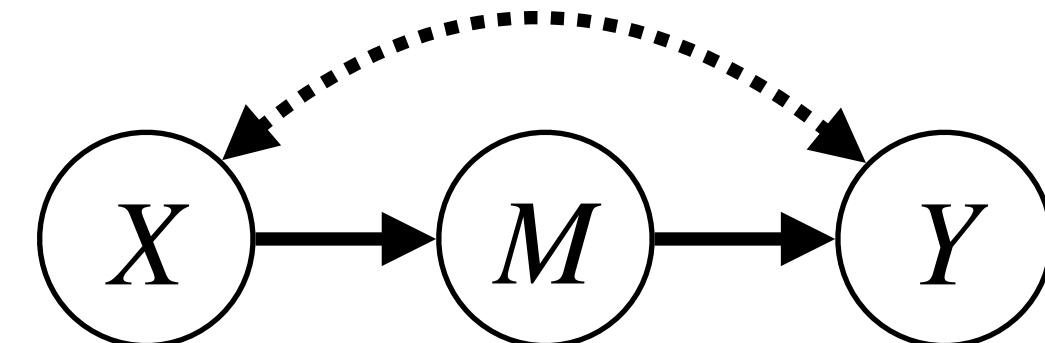


Counterfactual  
Graph



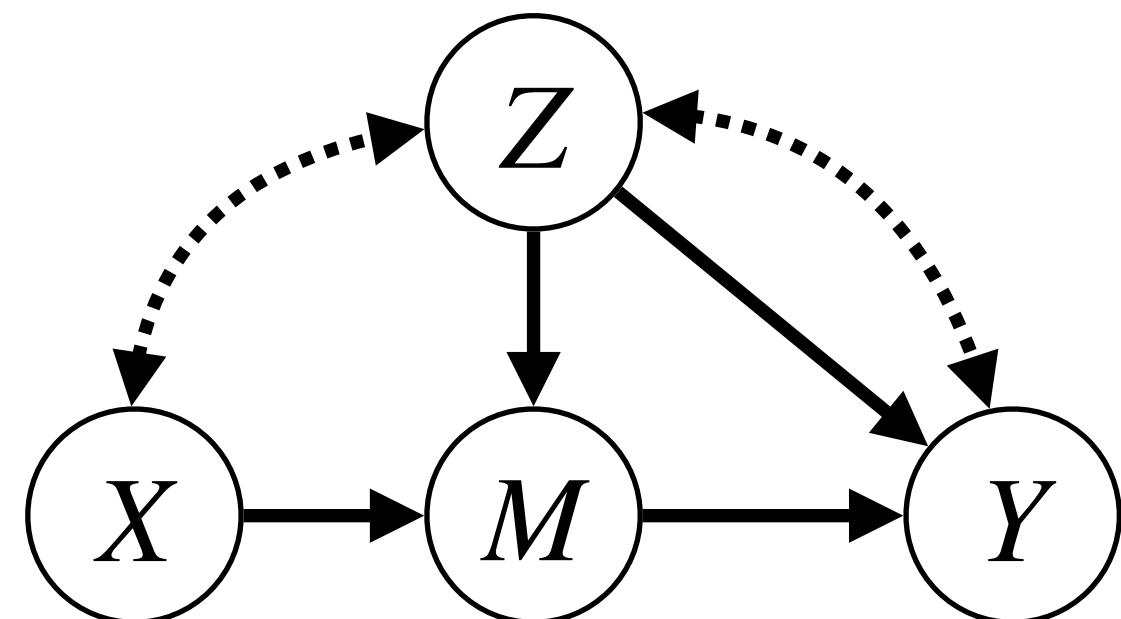
$$Y_x \perp\!\!\!\perp X | Z$$

# Many Scenarios Beyond Adjustment!

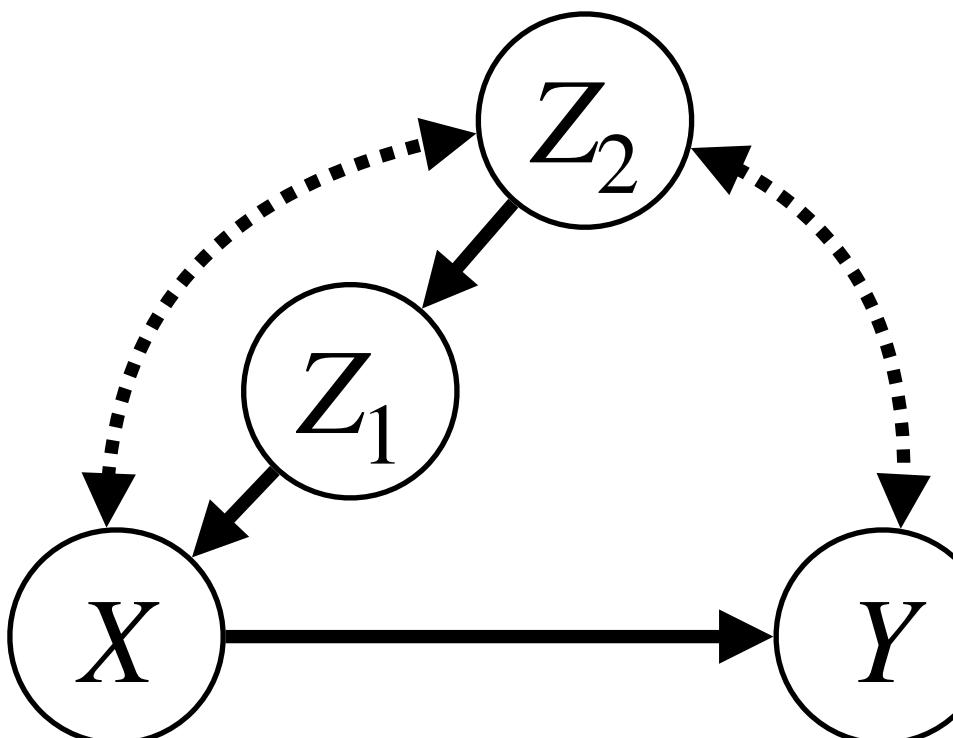


$$P(y | do(x)) = \sum_m P(m | x) \sum_{x'} P(y | m, x') P(x')$$

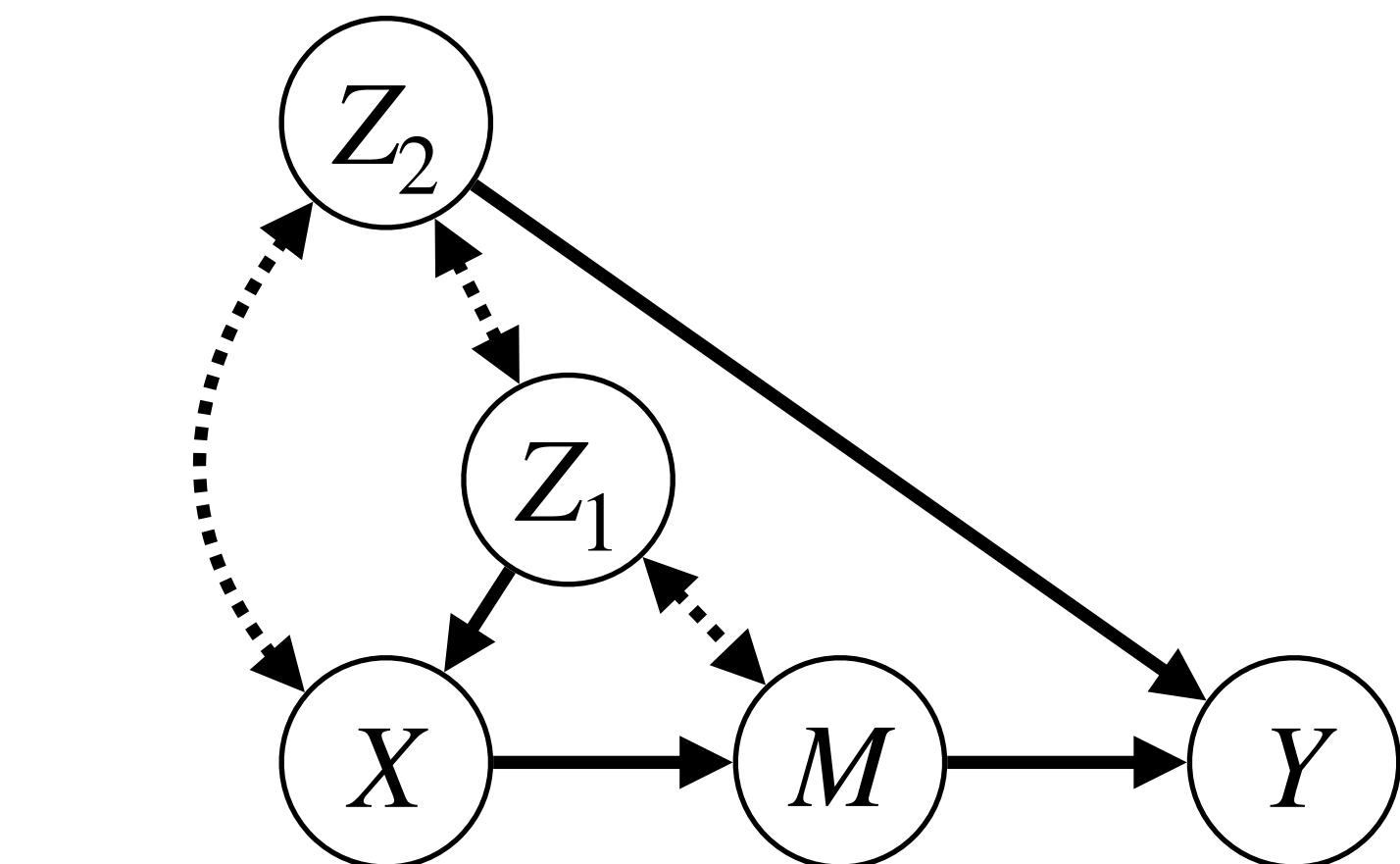
Front-Door



Conditional Front-Door



Napkin



Unnamed

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

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

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

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

$$\sum_{z_1} P(z_3 | x, z_1) P(z_1)$$

And many others....

# Tools for Causal Identification

---

1. Markovian Models (No Unobserved Confounders)
  - i. Truncated Factorization / G-computation or G-formula
2. Adjustment over Parents (No Unobserved Parents)
3. Non-Markovian Models (Under the Presence of Unobserved Confounders)
  - i. Graphical criteria (Backdoor Adjustment, Generalized Adjustment, Front-door Adjustment)
  - ii. Do-Calculus (a.k.a Causal Calculus)
  - iii. 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.

# Advances on Effect Identification given a Causal Diagram

---

## **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.

J. Correa, S. Lee, E. Bareinboim. (2021) Nested Counterfactual Identification from Arbitrary Surrogate Experiments. In Proceedings of the 35th Annual Conference on Neural Information Processing Systems

## **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.

## **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.

Xia, K., Pan, Y., and Bareinboim, E. (2023) Neural Causal Models for Counterfactual Identification and Estimation. In Proceedings of the 11th International Conference on Learning Representations.

# Coding Exercises

---

## Causality Tutorial:

- Available at <https://github.com/adele/Causality-Tutorial/> → Nordic ProbAI 2024
- Google Colab Notebook: ([Link](#))

## Check Part I:

1. Causal Modeling
2. Causal Effect Identification from Causal Diagrams
  1. Adjustment Criterion -- pcalg R package
  2. ID Algorithm -- causaleffect R package



---

**What if domain knowledge does not allow  
you construct a causal diagram?**



# Super-Exponential Growth

---

The space of DAGs grows super-exponentially with the number  $n$  of variables, as shown by the following recurrence relation (Robinson, 1973):

$$|DAG(n)| = \sum_{i=1}^n \binom{n}{i} 2^{i(n-i)} |DAG(n-1)|$$

Inference through enumeration  
is not a good idea!

$n$	$ DAG(n) $
2	3
3	27
4	729
5	59,049
6	$1.4349 \times 10^7$
7	$1.0460 \times 10^{10}$
8	$2.2877 \times 10^{13}$

# Super-Exponential Growth

---

The space of ADMGs also grows super-exponentially with the number  $n$  of variables, and it is much bigger than the space of DAGs:

$$|ADMG(n)| = |DAG(n)| \times 2^{n(n-1)/2}$$

$$|ADMG(n)| \gg |DAG(n)|$$

$n$	$ DAG(n) $	$ ADMG(n) $
2	3	6
3	27	216
4	729	46,656
5	59,049	$6.0457 \times 10^7$
6	$1.4349 \times 10^7$	$4.7019 \times 10^{11}$
7	$1.0460 \times 10^{10}$	$2.1936 \times 10^{16}$
8	$2.2877 \times 10^{13}$	$6.1410 \times 10^{21}$

# Learning the Markov Equivalence Class

---

## Causal Discovery:

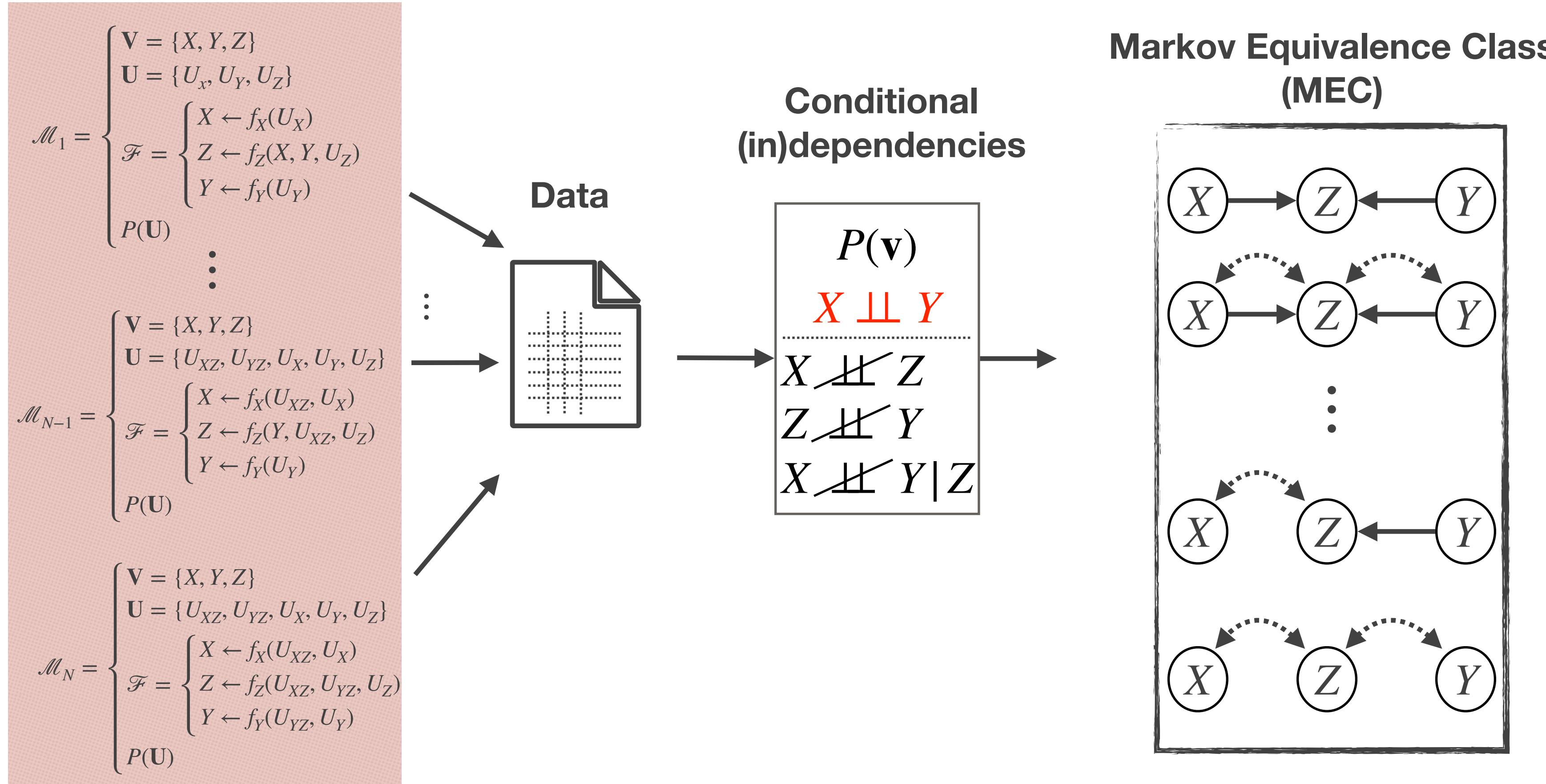
Many models are statistically indistinguishable without additional parametric / distributional assumptions.

In non-parametric settings, causal discovery algorithms can only learn a graphical representation of its *Markov equivalence class* (MEC)!

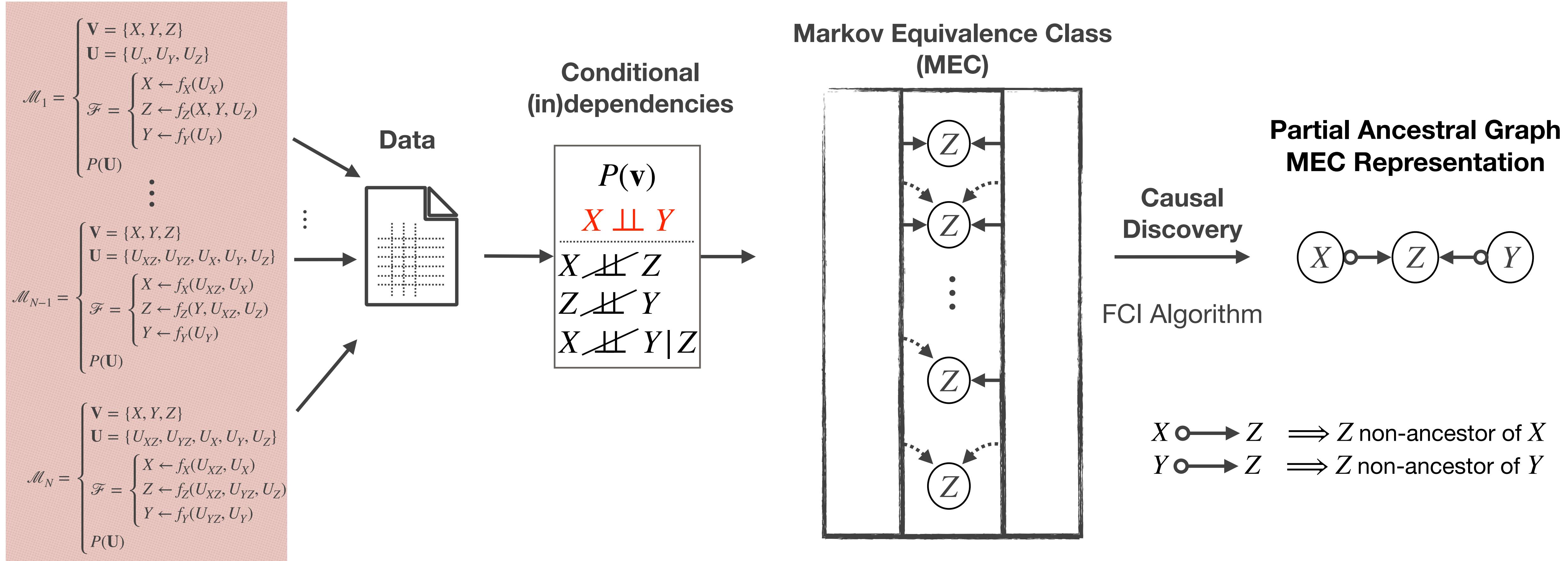
**Fast Causal Inference (FCI):** Sound and complete causal discovery algorithm, even in the presence of unobserved confounders and selection bias.

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: Learning Structural Invariances

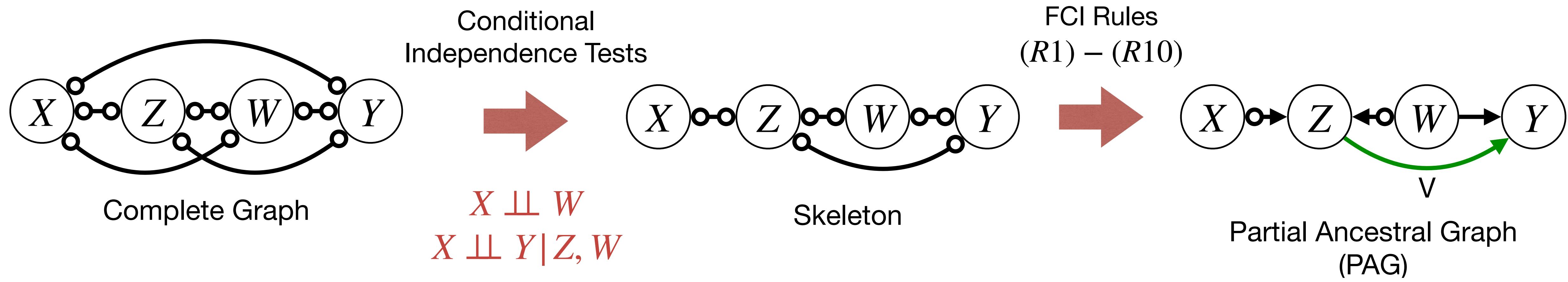
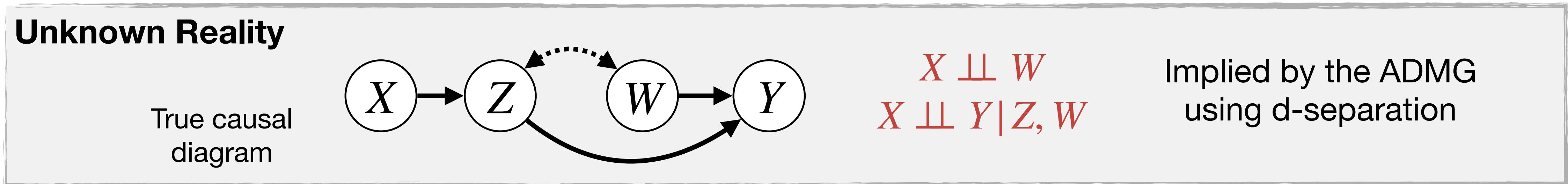


# Causal Discovery: Learning Structural Invariances



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](#)

# FCI Algorithm - Pipeline



By **faithfulness**, are correctly observed in the data

$A \circlearrowleft B \implies$  B non-ancestor of A

$A \longrightarrow B \implies$  A ancestor of B

$A \longleftrightarrow B \implies$  spurious association

$A — B \implies$  selection bias

Implied by the PAG using m-separation

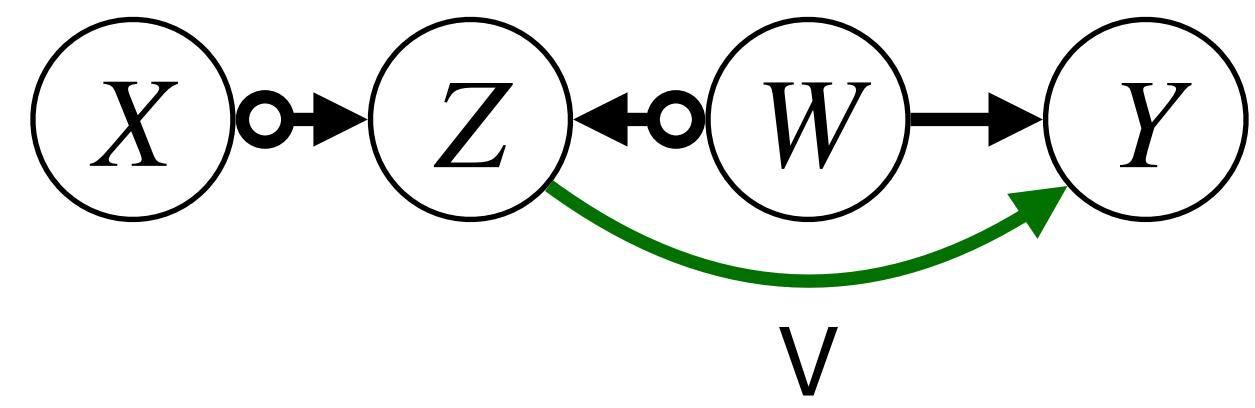
$X \perp\!\!\!\perp W$   
 $X \perp\!\!\!\perp Y | Z, W$

**Z is not an ancestor of X or W.**

**Z and W are ancestors of Y.**

**Z is not confounded with Y.**

# PAG: Representation of the Markov Equivalence Class

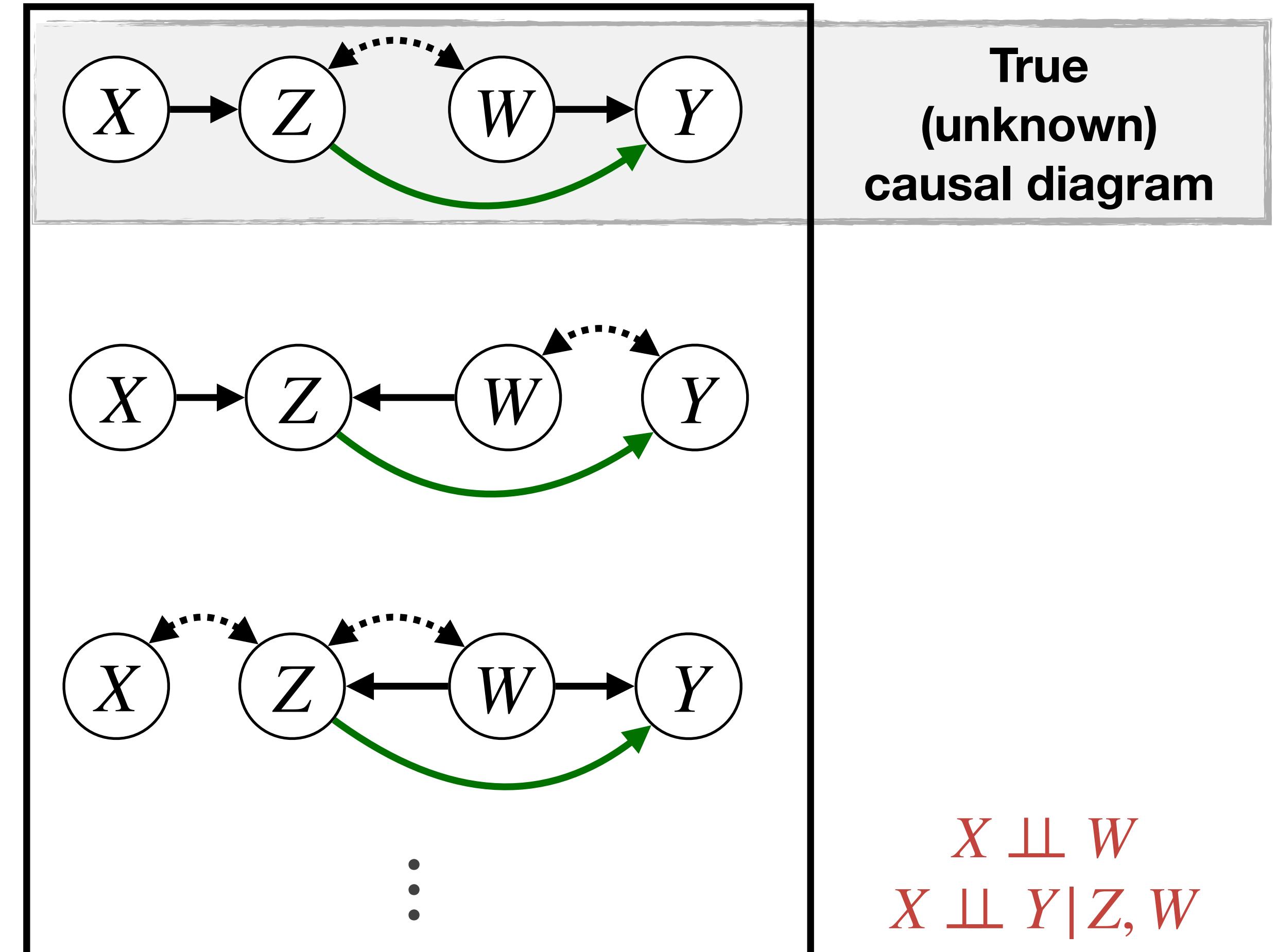


Partial Ancestral Graph  
(PAG)

**Z is not an ancestor of X or W.**

**Z and W are ancestors of Y.**

**Z is not confounded with Y.**

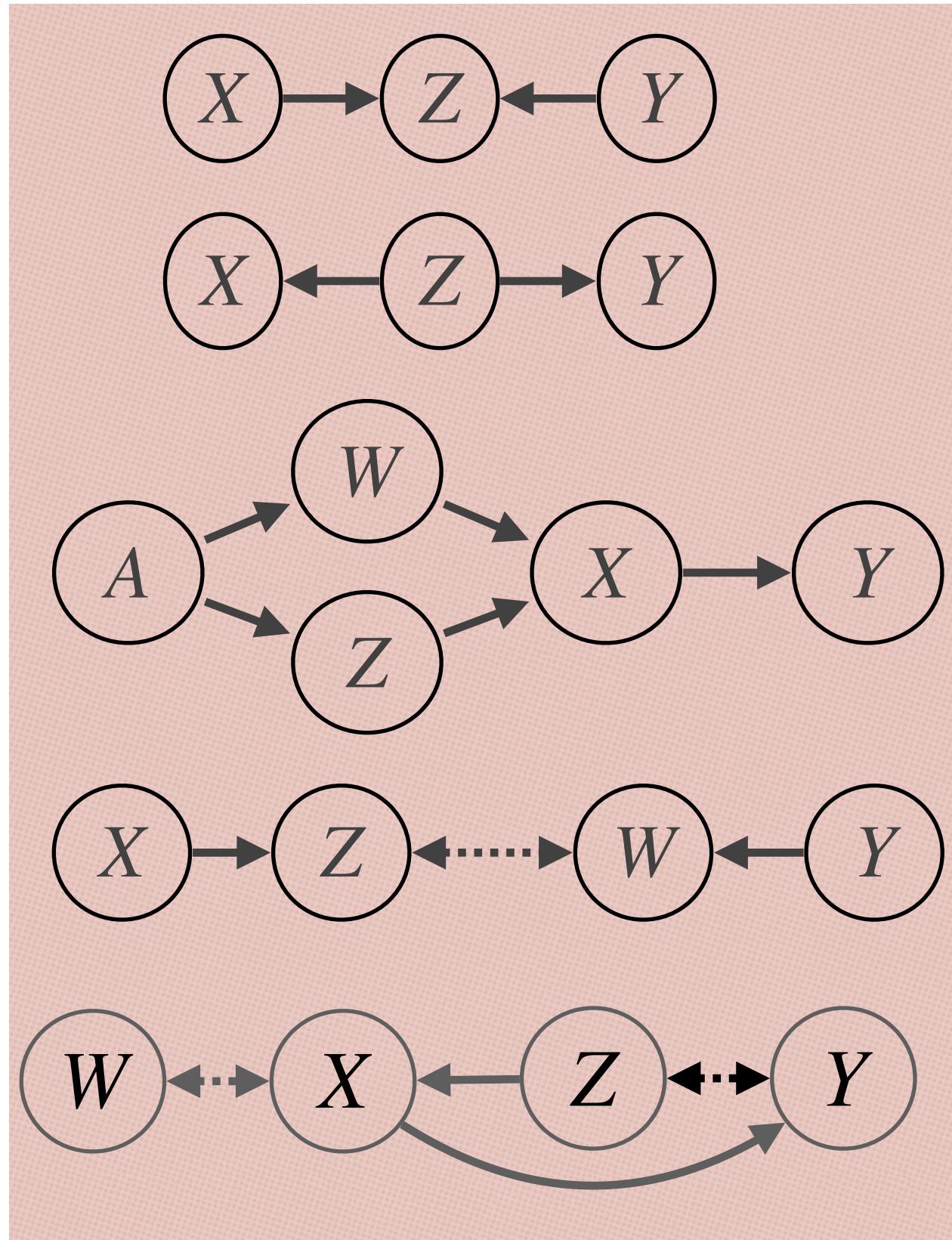


$X \perp\!\!\!\perp W$

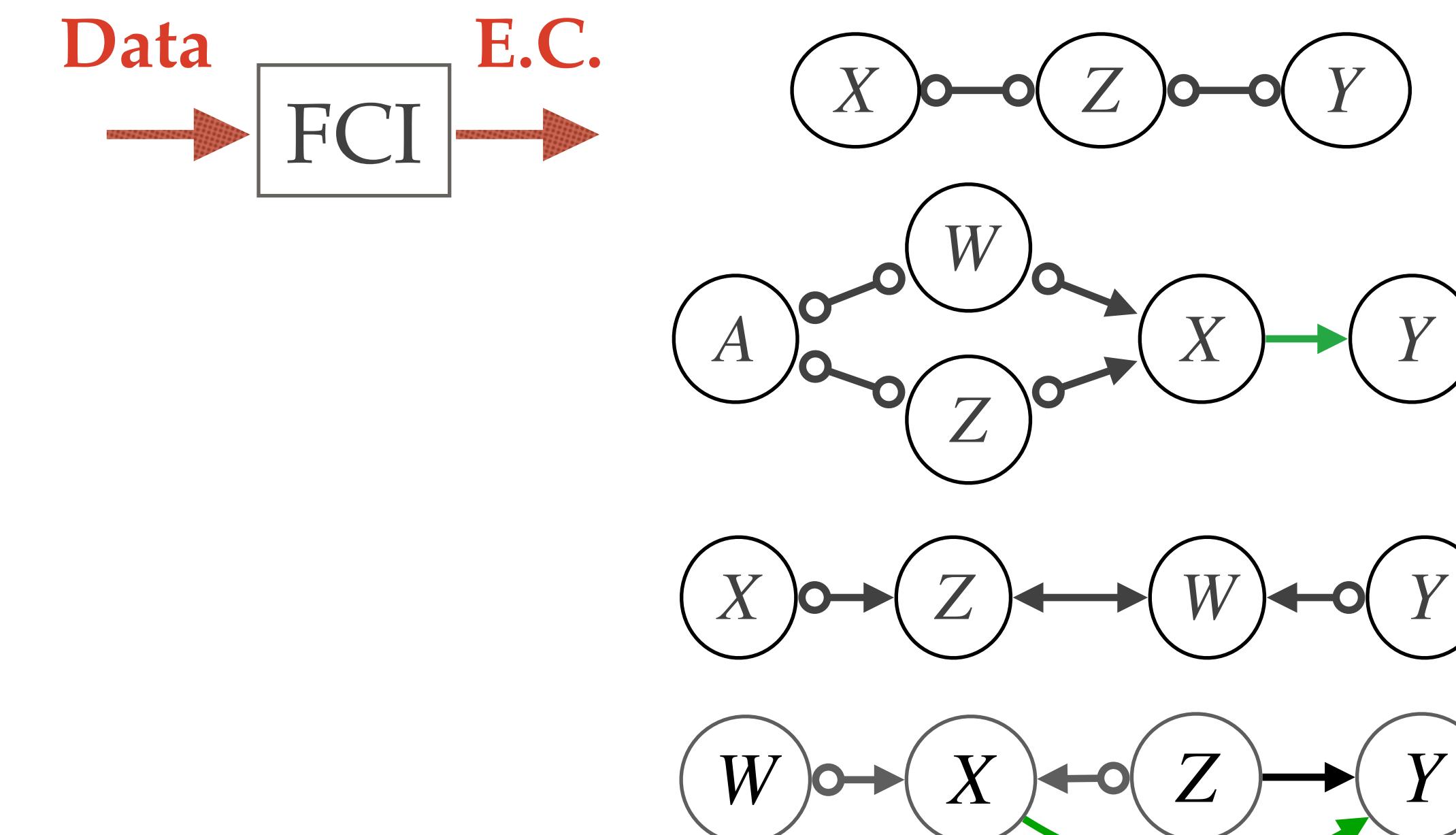
$X \perp\!\!\!\perp Y | Z, W$

# Fast Causal Inference (FCI) Algorithm

Underlying Causal Diagram



Partial Ancestral Graph



# 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>

# Causal Identification from PAGs

---



Can we identify causal effects from the equivalence class?

## Effect Identification:

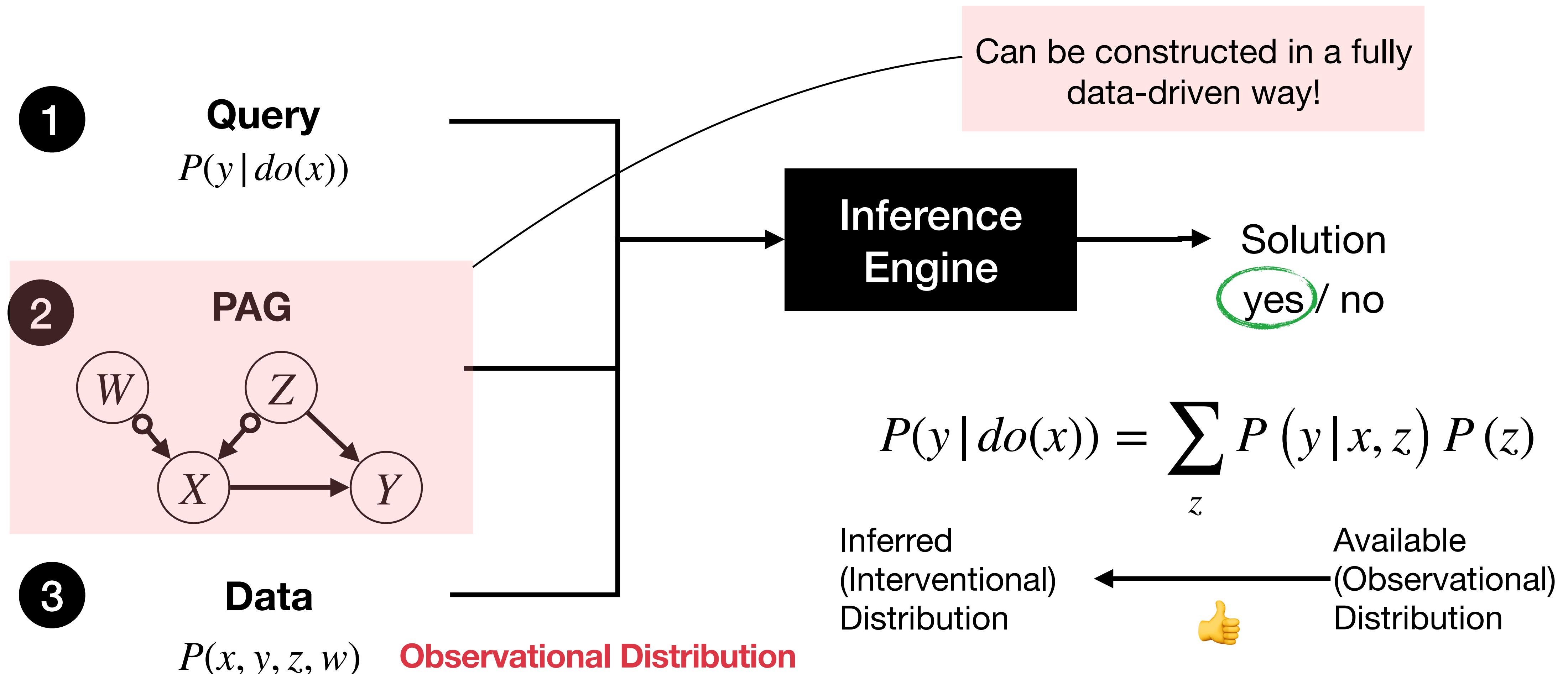
For Covariate Adjustment, we can use the Generalized Adjustment Criterion.

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

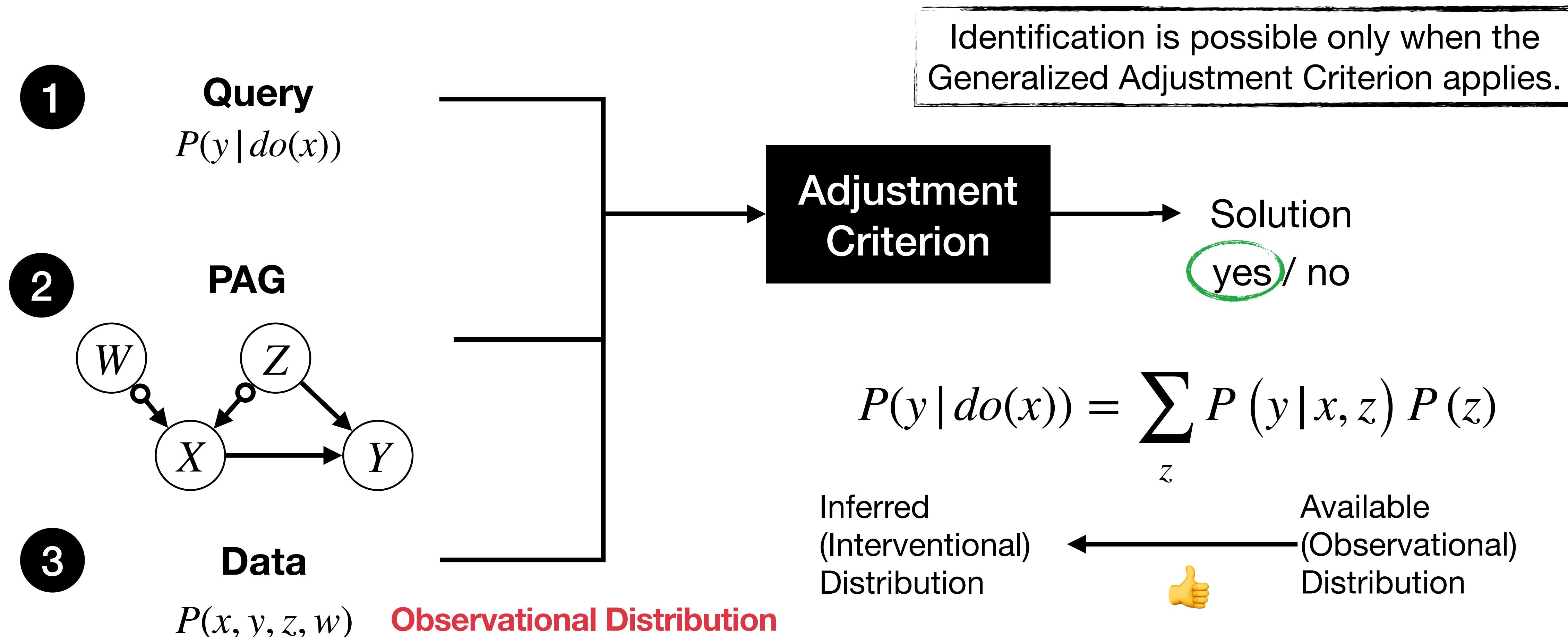
Perkovic, E., Textor, J. C., Kalisch, M., & Maathuis, M. H. (2018). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. Journal of Machine Learning Research 18 (2018) 1-62

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](#))

# Effect Identification in Markov Equivalence Classes

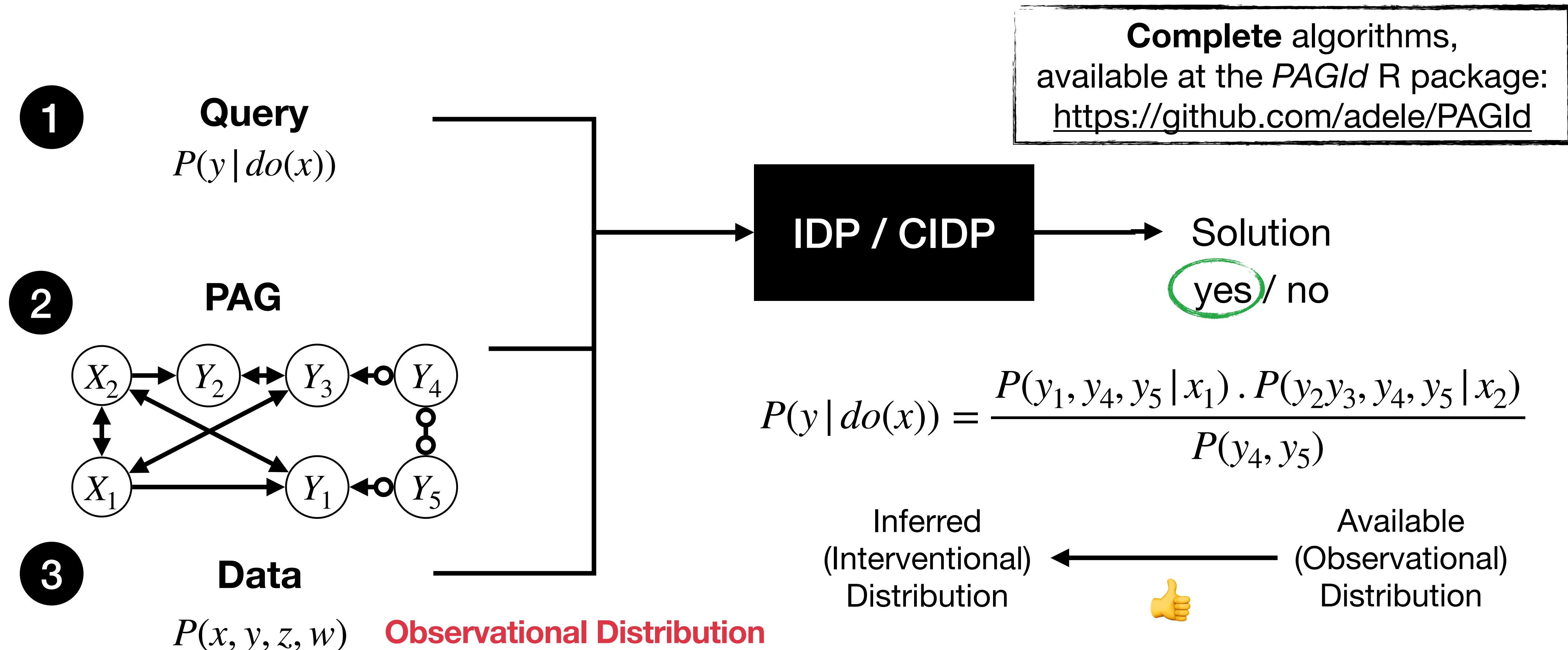


# Identification via Adjustment in Markov Equivalence Classes



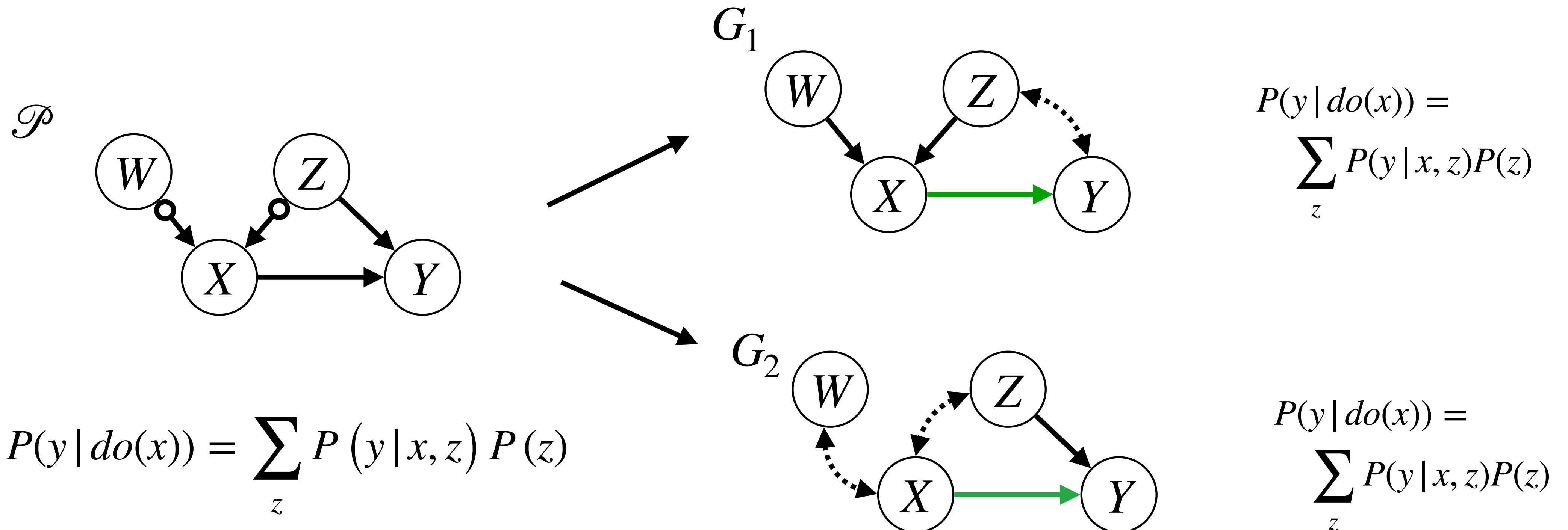
Perkovic, E., Textor, J. C., Kalisch, M., & Maathuis, M. H. (2018). [Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs](#). Journal of Machine Learning Research 18 (2018) 1-62

# General Identification in Markov Equivalence Classes



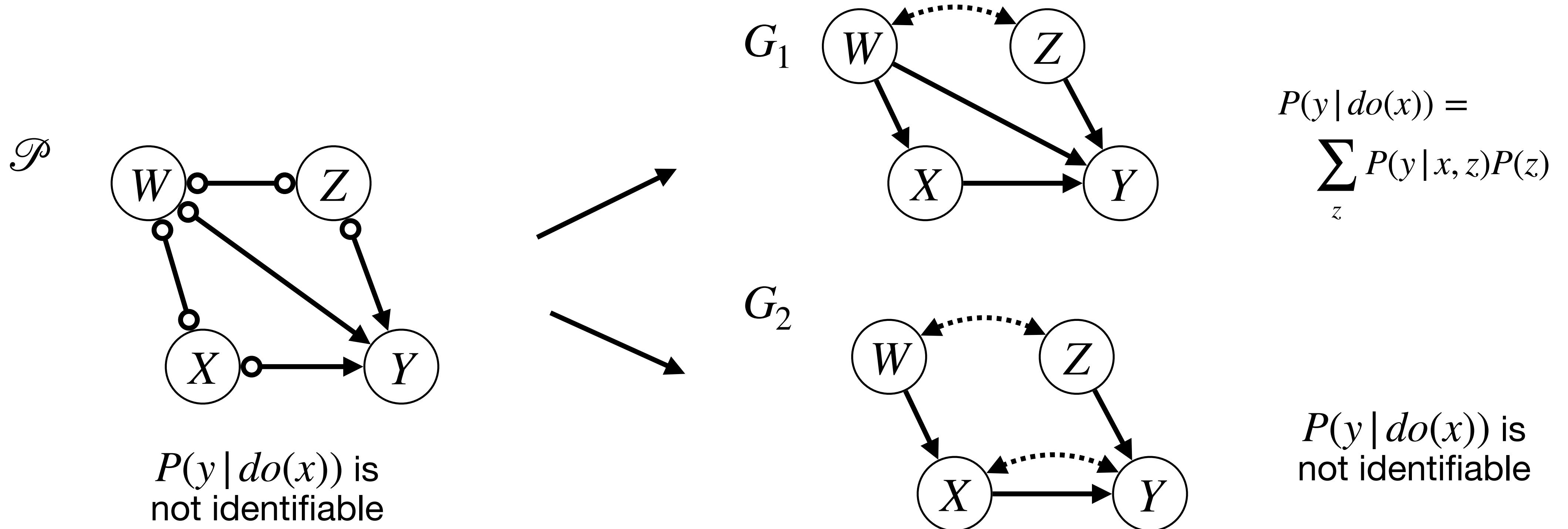
Jaber A., Ribeiro A. H., Zhang, J., Bareinboim, E. (2022) Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS 2022).

# 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

# Many other Topics in Causal Inference

---

1. Causal Discovery with Cycles
2. Causal Discovery from Time-Series Data
3. Causal Discovery in Linear Models
4. Causal Discovery with Prior Knowledge
5. Causal Discovery for Additive Noise Models
6. Causal Discovery with Interventional Data
7. Data-Driven Covariate Selection for Adjustment
8. Partial Effect Identification
9. Fairness and Mediation Analysis
10. Individual Treatment Effect (ITE) Estimation
11. Many more...

# Causal Discovery with Cycles

---

October 2021

## Foundations of structural causal models with cycles and latent variables

Stephan Bongers, Patrick Forré, Jonas Peters, Joris M. Mooij

Author Affiliations +

Ann. Statist. 49(5): 2885-2915 (October 2021). DOI: 10.1214/21-AOS2064

DE GRUYTER

Research Article

S. Bongers\*, T. Blom, and J.M. Mooij

## Causal Modeling of Dynamical Systems

Structural Dynamical Causal Models  
(SDCMs)

---

## Establishing Markov Equivalence in Cyclic Directed Graphs

---

Tom Claassen<sup>1</sup>

Joris M. Mooij<sup>2</sup>

<sup>1</sup>Institute for Computing and Information Sciences, Radboud University, Nijmegen, Netherlands

<sup>2</sup>Korteweg-deVries Institute, University of Amsterdam, Amsterdam, Netherlands

**Claassen, T. & Mooij, J.M.. (2023).**  
*Establishing Markov equivalence in cyclic directed graphs. Proceedings of the Thirty-Ninth Conference on Uncertainty in Artificial Intelligence, PMLR 216:433-442, 2023.*

# Causal Discovery from Time-Series Data

---

---

## High-recall causal discovery for autocorrelated time series with latent confounders

---

**Andreas Gerhardus**  
German Aerospace Center  
Institute of Data Science  
07745 Jena, Germany  
[andreas.gerhardus@dlr.de](mailto:andreas.gerhardus@dlr.de)

**Jakob Runge**  
German Aerospace Center  
Institute of Data Science  
07745 Jena, Germany  
[jakob.runge@dlr.de](mailto:jakob.runge@dlr.de)

**Gerhardus, A., & Runge, J.** (2020). High-recall causal discovery for autocorrelated time series with latent confounders. *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 33, 12615-12625.

## CHARACTERIZATION OF CAUSAL ANCESTRAL GRAPHS FOR TIME SERIES WITH LATENT CONFOUNDERS

BY ANDREAS GERHARDUS<sup>1,a</sup>

<sup>1</sup>*German Aerospace Center, Institute of Data Science, [a.andreas.gerhardus@dlr.de](mailto:a.andreas.gerhardus@dlr.de)*

# Causal Discovery and Identification in Linear Models

---

ParceLiNGAM: A causal ordering method robust  
against latent confounders

Tatsuya Tashiro\*, Shohei Shimizu†, Aapo Hyvärinen‡ and Takashi Washio§

**Tashiro, T., Shimizu, S., Hyvärinen, A., & Washio, T.** (2014).  
ParceLiNGAM: A causal ordering method robust against latent  
confounders. *Neural computation*, 26(1), 57-83.

Causal Discovery with Unobserved Confounding and  
Non-Gaussian Data

**Y. Samuel Wang**

*Department of Statistics and Data Science  
Cornell University  
Ithaca, NY 14853, USA*

YSW7@CORNELL.EDU

**Mathias Drton**

*Department of Mathematics & Munich Data Science Institute  
Technical University of Munich  
85748 Garching bei München, Germany*

MATHIAS.DRTON@TUM.DE

**Editor:** Joris Mooij

**Wang, Y. S., & Drton, M.** (2023). Causal discovery with  
unobserved confounding and non-Gaussian data. *Journal  
of Machine Learning Research*, 24(271), 1-61.

---

## Causal Effect Identification in LiNGAM Models with Latent Confounders

---

Daniele Tramontano<sup>1</sup> Yaroslav Kivva<sup>2</sup> Saber Salehkaleybar<sup>3</sup> Mathias Drton<sup>1,4</sup> Negar Kiyavash<sup>2</sup>

# Causal Discovery with Prior Knowledge

---

---

## Sound and Complete Causal Identification with Latent Variables Given Local Background Knowledge

---

Tian-Zuo Wang, Tian Qin, Zhi-Hua Zhou

National Key Laboratory for Novel Software Technology,  
Nanjing University, Nanjing, 210023, China.  
{wangtz, qint, zhouzh}@lamda.nju.edu.cn

Wang, T.Z., Qin, T. and Zhou, Z.H., 2022. Sound and complete causal identification with latent variables given local background knowledge. *Advances in Neural Information Processing Systems*, 35, pp.10325-10338.



Contents lists available at [ScienceDirect](#)

Artificial Intelligence

journal homepage: [www.elsevier.com/locate/artint](http://www.elsevier.com/locate/artint)

---

Sound and complete causal identification with latent variables given local background knowledge

Tian-Zuo Wang, Tian Qin, Zhi-Hua Zhou \*

*National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023, China*

# Causal Discovery for Additive Noise Models

---

## Beyond the Markov Equivalence Class: Extending Causal Discovery under Latent Confounding

Mirthe M. van Diepen  
Ioan Gabriel Bucur  
Tom Heskes  
Tom Claassen

*Institute for Computing and Information Sciences, Radboud Universiteit Nijmegen*

MIRTHE.VANDIEPEN@RU.NL  
G.BUCUR@CS.RU.NL  
TOM.HESKES@RU.NL  
TOMC@CS.RU.NL

**Van Diepen, M. M., Bucur, I. G., Heskes, T., & Claassen, T.** (2023, August). Beyond the Markov Equivalence Class: Extending Causal Discovery under Latent Confounding. In *Conference on Causal Learning and Reasoning* (pp. 707-725). PMLR.

**Editors:** Mihaela van der Schaar, Dominik Janzing and Cheng Zhang

And interesting papers cited within, such as:

**Isabelle Guyon, Olivier Goudet, and Diviyan Kalainathan.** Evaluation methods of cause-effect pairs. In *Cause Effect Pairs in Machine Learning*, pages 27–99. Springer, 2019.

# Causal Discovery with Interventional Data

---

---

## Causal Discovery from Soft Interventions with Unknown Targets: Characterization and Learning

---

**Amin Jaber**  
Department of Computer Science  
Purdue University, USA  
[jaber@purdue.edu](mailto:jaber@purdue.edu)

**Karthikeyan Shanmugam**  
MIT-IBM Watson AI Lab  
IBM Research NY, USA  
[kartikeyan.shanmugam2@ibm.com](mailto:kartikeyan.shanmugam2@ibm.com)

**Murat Kocaoglu**  
MIT-IBM Watson AI Lab  
IBM Research MA, USA  
[murat@ibm.com](mailto:murat@ibm.com)

**Elias Bareinboim**  
Department of Computer Science  
Columbia University, USA  
[eb@cs.columbia.edu](mailto:eb@cs.columbia.edu)

---

## Causal discovery from observational and interventional data across multiple environments

---

**Adam Li**  
Department of Computer Science, Columbia University  
[adam.li@columbia.edu](mailto:adam.li@columbia.edu)

**Amin Jaber**  
Synlico Inc.  
[amin.jaber@synlico.com](mailto:amin.jaber@synlico.com)

**Elias Bareinboim**  
Department of Computer Science, Columbia University  
[eb@cs.columbia.edu](mailto:eb@cs.columbia.edu)

**Jaber, A., Kocaoglu, M., Shanmugam, K. and Bareinboim, E.,**  
(2020). Causal discovery from soft interventions with unknown  
targets: Characterization and learning. *Advances in neural  
information processing systems*, 33, pp.9551-9561.

**A. Li, A. Jaber, E. Bareinboim.** Causal discovery from observational  
and interventional data across multiple environments. (2023)  
In *Proceedings of the 37th Annual Conference on Neural Information  
Processing Systems — NeurIPS-23*.

# Data-Driven Covariate Selection for Adjustment

---

---

## Finding Valid Adjustments under Non-ignorability with Minimal DAG Knowledge

---

**Abhin Shah**  
MIT  
[abhin@mit.edu](mailto:abhin@mit.edu)

**Karthikeyan Shanmugam**  
IBM Research  
[karthikeyan.shanmugam2@ibm.com](mailto:karthikeyan.shanmugam2@ibm.com)

**Kartik Ahuja**  
Mila  
[kartik.ahuja@mila.quebec](mailto:kartik.ahuja@mila.quebec)

**Abhin Shah, Karthikeyan Shanmugam, and Kartik Ahuja.** Finding valid adjustments under non-ignorability with minimal DAG knowledge. In *International Conference on Artificial Intelligence and Statistics (AISTATS - 2022)*, pages 5538–5562. PMLR, 2022.

---

## Differentiable Causal Backdoor Discovery

---

**Limor Gultchin**  
University of Oxford  
The Alan Turing Institute

**Matt J. Kusner**  
University College London  
The Alan Turing Institute

**Varun Kanade**  
University of Oxford  
The Alan Turing Institute

**Ricardo Silva**  
University College London  
The Alan Turing Institute

# Partial Effect Identification

---

## Stochastic Causal Programming for Bounding Treatment Effects

**Kirtan Padh**

*Helmholtz AI, Helmholtz Munich & Technical University Munich*

KIRTAN.PADH@TUM.DE

**Jakob Zeitler**

*University College London*

**David Watson**

*King's College London*

**Matt Kusner**

*University College London*

**Ricardo Silva**

*University College London*

**Niki Kilbertus**

*Helmholtz AI, Helmholtz Munich & Technical University Munich*

**Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner, Ricardo Silva, Niki Kilbertus;** *Proceedings of the Second Conference on Causal Learning and Reasoning*, PMLR 213:142-176

Related: **Jakob Zeitler, and Ricardo Silva.** (2022) The Causal Marginal Polytope for Bounding Treatment Effects arXiv preprint arXiv:2202.13851 - <https://arxiv.org/pdf/2202.13851.pdf>

# Fairness and Mediation Analysis

---

---

## A Causal Framework for Decomposing Spurious Variations

**Drago Plecko** and **Elias Bareinboim**  
Department of Computer Science  
Columbia University  
dp3144@columbia.edu, eb@cs.columbia.edu

**D. Plecko, E. Bareinboim.** A Causal Framework for Decomposing Spurious Variations. In *Proceedings of the 37th Annual Conference on Neural Information Processing Systems — NeurIPS-23*.

## Foundations and Trends® in Machine Learning Causal Fairness Analysis

### A Causal Toolkit for Fair Machine Learning

---

**Suggested Citation:** Drago Plečko and Elias Bareinboim (2024), "Causal Fairness Analysis", Foundations and Trends® in Machine Learning: Vol. 17, No. 3, pp 1–238. DOI: 10.1561/2200000106.

**Drago Plečko**  
Seminar für Statistik, ETH Zürich  
drago.plecko@stat.math.ethz.ch

**Elias Bareinboim**  
Department of Computer Science, Columbia University  
eb@cs.columbia.edu

# Individual Treatment Effect (ITE) Estimation

---

## Generalization Bounds and Representation Learning for Estimation of Potential Outcomes and Causal Effects

**Fredrik D. Johansson**

*Chalmers University of Technology  
Göteborg, 412 96, Sweden*

FREDRIK.JOHANSSON@CHALMERS.SE

**Uri Shalit**

*Technion - Israel Institute of Technology  
Haifa, 3200003, Israel*

URISHALIT@TECHNION.AC.IL

**Nathan Kallus**

*Cornell University  
New York, NY 10044, USA*

KALLUS@CORNELL.EDU

**David Sontag**

*Massachusetts Institute of Technology  
Cambridge, MA 02139, USA*

DSONTAG@CSAIL.MIT.EDU

Other related works cited within, such as:

### Estimating individual treatment effect: generalization bounds and algorithms

*Uri Shalit, Fredrik D. Johansson, David Sontag* Proceedings of the 34th International Conference on Machine Learning, PMLR 70:3076–3085, 2017.

---

### Learning Representations for Counterfactual Inference

---

**Fredrik D. Johansson\***

CSE, Chalmers University of Technology, Göteborg, SE-412 96, Sweden

FREJOHK@CHALMERS.SE

**Uri Shalit\***

**David Sontag**

CIMS, New York University, 251 Mercer Street, New York, NY 10012 USA

SHALIT@CS.NYU.EDU

DSONTAG@CS.NYU.EDU

\* Equal contribution

# Causal Representation Learning

---

## Toward Causal Representation Learning

*This article reviews fundamental concepts of causal inference and relates them to crucial open problems of machine learning, including transfer learning and generalization, thereby assaying how causality can contribute to modern machine learning research.*

By BERNHARD SCHÖLKOPF<sup>ID</sup>, FRANCESCO LOCATELLO<sup>ID</sup>, STEFAN BAUER<sup>ID</sup>, NAN ROSEMARY KE,  
NAL KALCHBRENNER, ANIRUDH GOYAL, AND YOSHUA BENGIO<sup>ID</sup>

Schölkopf, B., Locatello, F., Bauer, S., Ke, N. R., Kalchbrenner, N., Goyal, A., & Bengio, Y. (2021). Toward causal representation learning. *Proceedings of the IEEE*, 109(5), 612-634.

## Coarse-grained causal models:

---

### Causal Consistency of Structural Equation Models

Paul K. Rubenstein<sup>\*12</sup>, Sebastian Weichwald<sup>\*13</sup>, Stephan Bongers<sup>4</sup>, Joris M. Mooij<sup>4</sup>  
Dominik Janzing<sup>1</sup>, Moritz Grosse-Wentrup<sup>1</sup>, Bernhard Schölkopf<sup>1</sup>  
\*Equal contribution  
<sup>1</sup>Empirical Inference, MPI for Intelligent Systems, <sup>2</sup>Machine Learning Group, University of Cambridge,  
<sup>3</sup>Max Planck ETH Center for Learning Systems, <sup>4</sup>Informatics Institute, University of Amsterdam

The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

### Causal Effect Identification in Cluster DAGs

Tara V. Anand<sup>\*1</sup>, Adele H. Ribeiro<sup>\*2</sup>, Jin Tian<sup>3</sup>, Elias Bareinboim<sup>2</sup>

<sup>1</sup>Department of Biomedical Informatics, Columbia University

<sup>2</sup>Department of Computer Science, Columbia University

<sup>3</sup>Department of Computer Science, Iowa State University

tara.v.anand@columbia.edu, adele@cs.columbia.edu, jtian@iastate.edu, eb@cs.columbia.edu

### Neural Causal Abstractions

**Kevin Xia and Elias Bareinboim**

Causal Artificial Intelligence Lab  
Columbia University  
{kevinmxia, eb}@cs.columbia.edu

# And many more....

---

1. Effect Identificationf from surrogate experimental distributions such as  $P(\mathbf{y} | do(\mathbf{w}))$
2. Identification of effects from soft-interventions:
  - i. Hard /  $do$ -interventions vs soft /  $\sigma$ -interventions.
3. Counterfactual Identification:
  1. Axioms and Identification Algorithms of  $\mathcal{L}_3$  quantities, such as  $P(\mathbf{y}_{\mathbf{x}} | \mathbf{x}', \mathbf{z})$ .
4. More on estimation of the adjustment formula:
  1. Variety of frameworks leveraging machine learning
  2. Doubly robustness
  3. Balancing
5. Optimal Causal Experimental Design
6. Causal Reinforcement Learning

# Coding Exercises

---

## Causality Tutorial:

- Available at <https://github.com/adele/Causality-Tutorial/> → Nordic ProbAI 2024
- Google Colab Notebook: ([Link](#))

## Check Part II:

1. Causal Discovery using FCI
2. Causal Effect Identification from PAGs



# **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)