Advancing Causal Inference for Scientific Discovery

From Modeling and Discovery to Effect Identification and Experimental Design

Adèle H. Ribeiro

https://adele.github.io/ | adele.ribeiro@uni-muenster.de

Institute for Medical Informatics
University of Münster

3rd TACsy PhD School & Consortium Meeting Leipzig
February 19th, 2025

Recent Breakthroughs in Artificial Intelligence

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

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

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

NEWS | ARTIFICIAL INTELLIGENCE

Making the Role of AI

Analysis system for the diagnos

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

nature

Explore content Journal information Publish with us Subscribe

nature > outlook > article

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

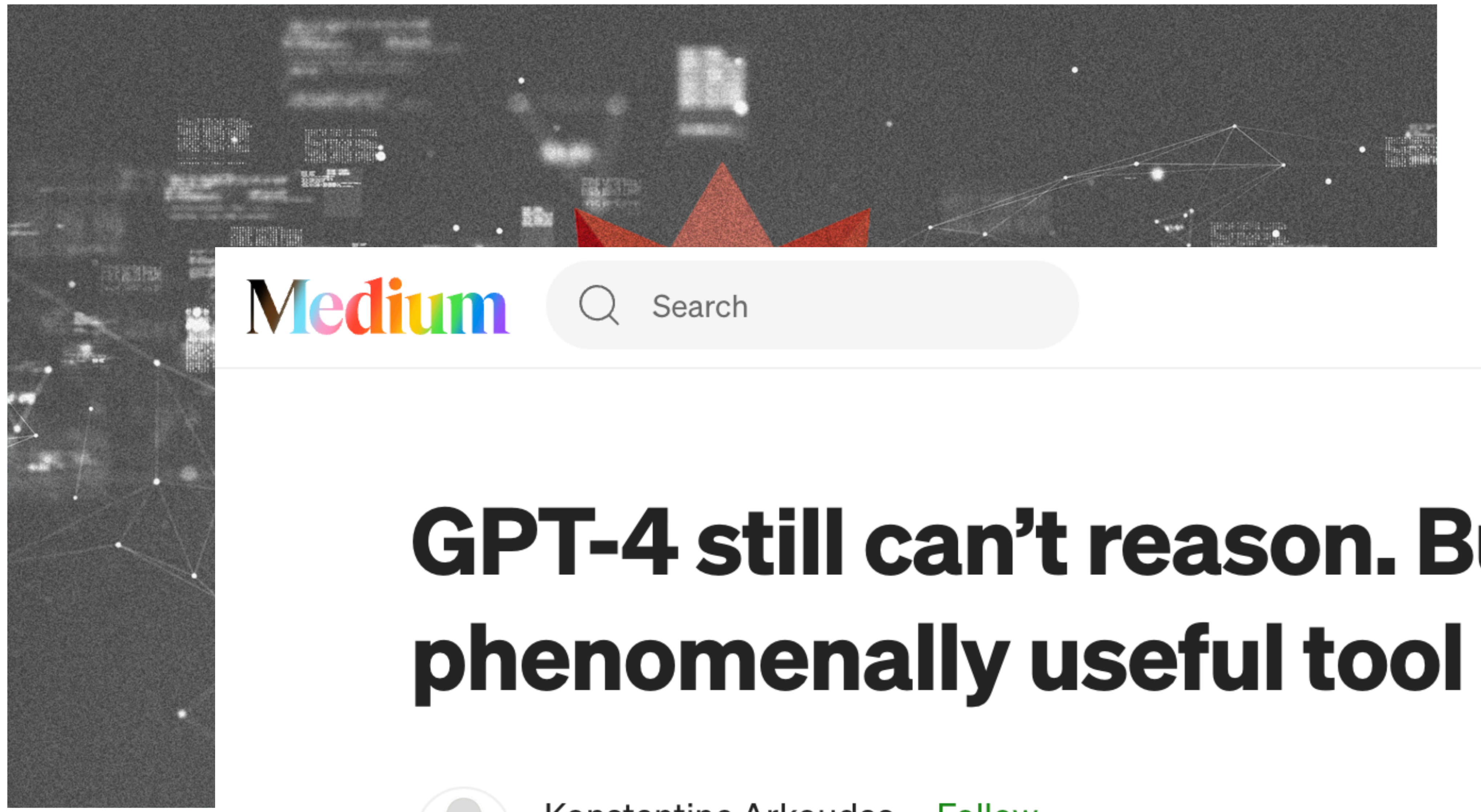
Chat GPT - Impressive Abilities:

- Debug code
- Play Games
- Translate text
- Explain complex topics
- Create AI art
- Grade essays
- Write a resume
- Brainstorm ideas
- Play games with ChatGPT
- Create chatbots
- Language tutor
- Create jokes
- Solve coding errors
- Write poetry

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.



[Image: RICHARD JON

BY CHR

Medium

Search

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

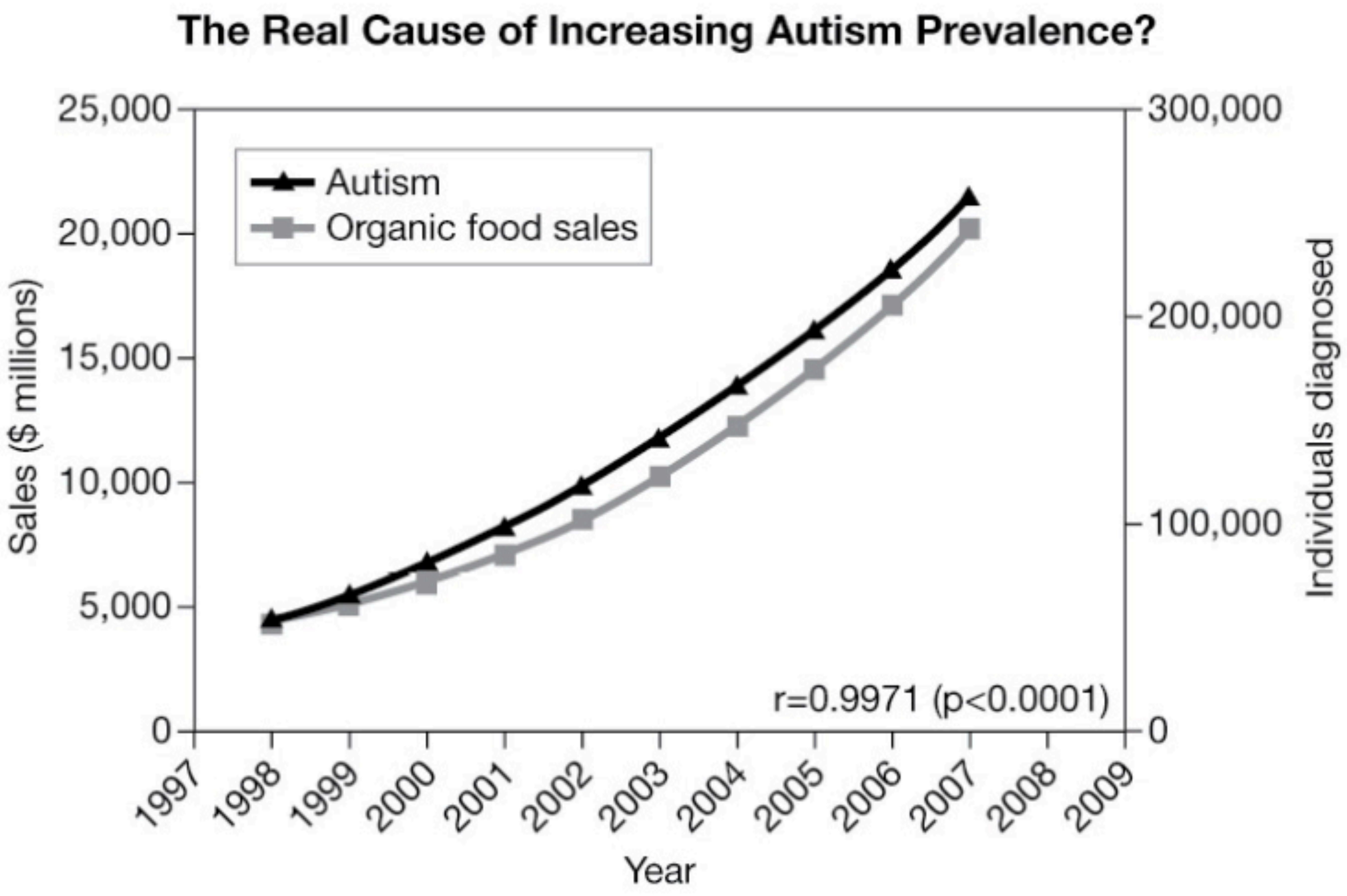
Why artificial intelligence needs to understand consequences

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

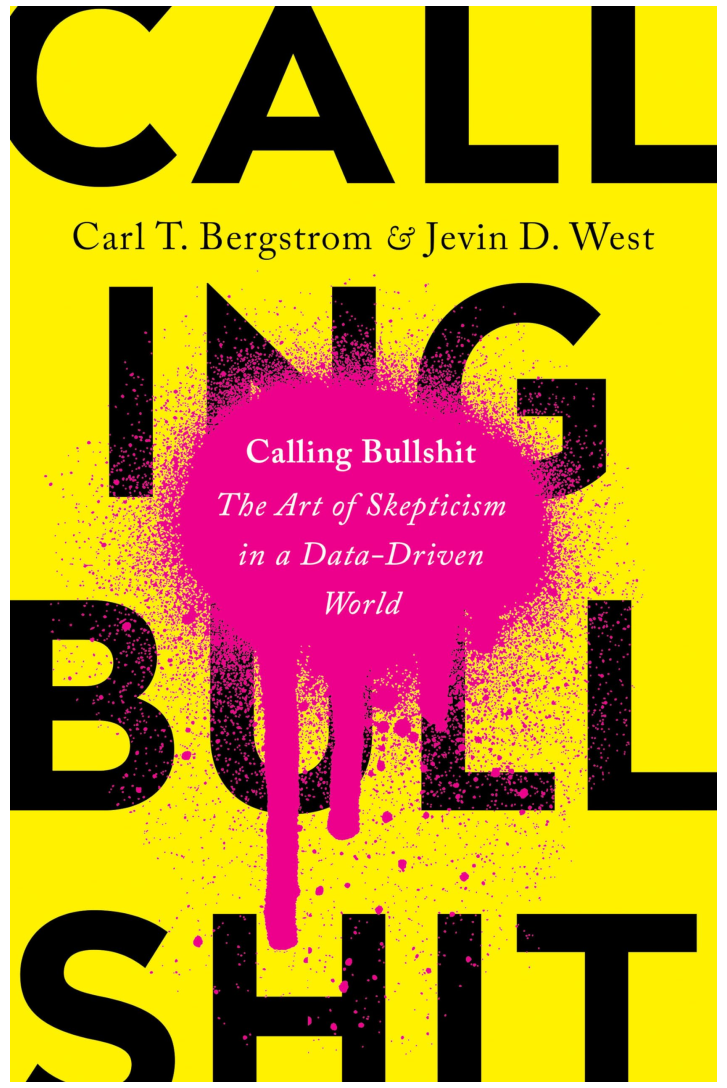
Prediction vs Effect of Interventions

Statistical Association vs Causation

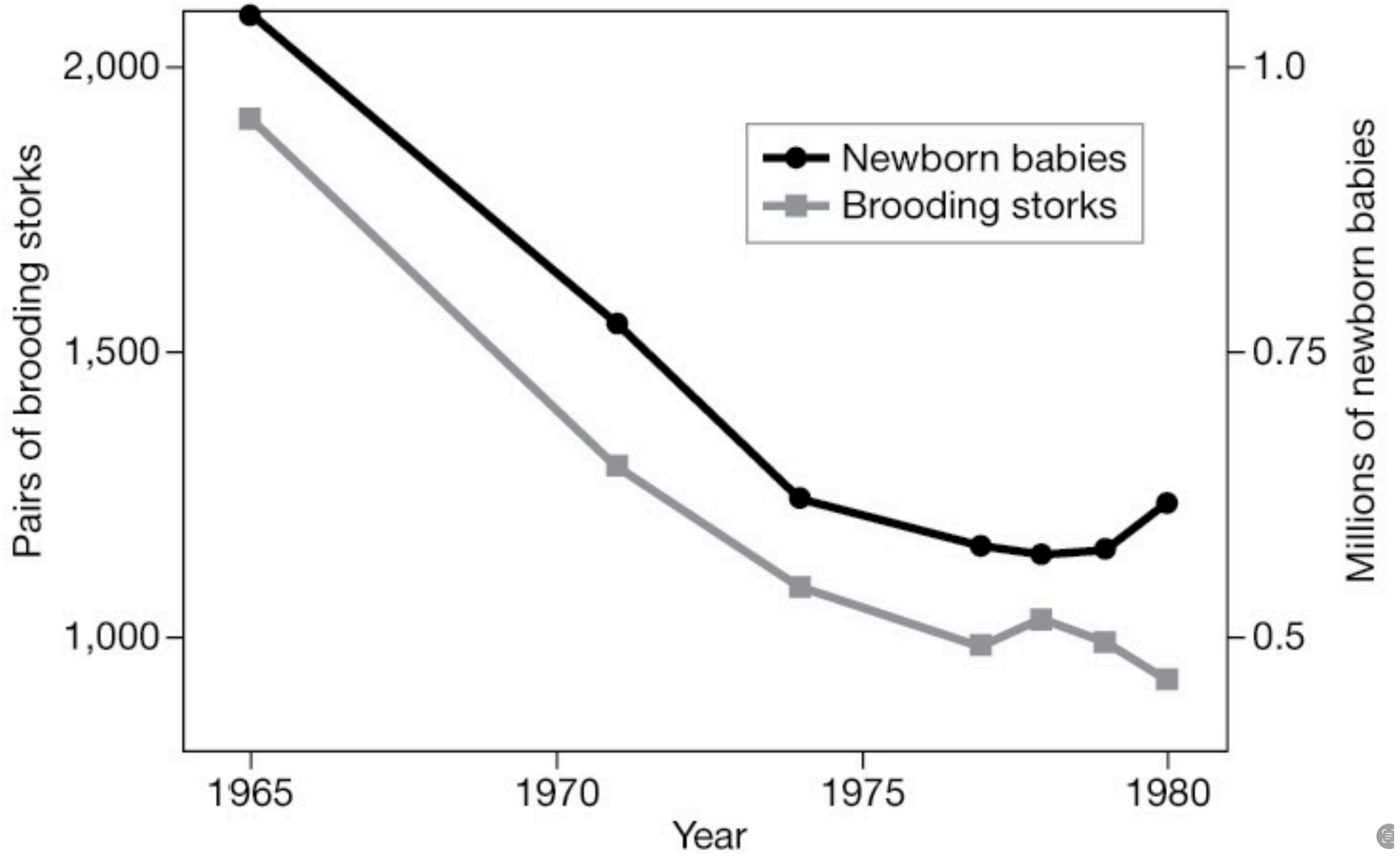
What Do Statistical Associations Reveal?



“one user of the Reddit website posted the following graph”



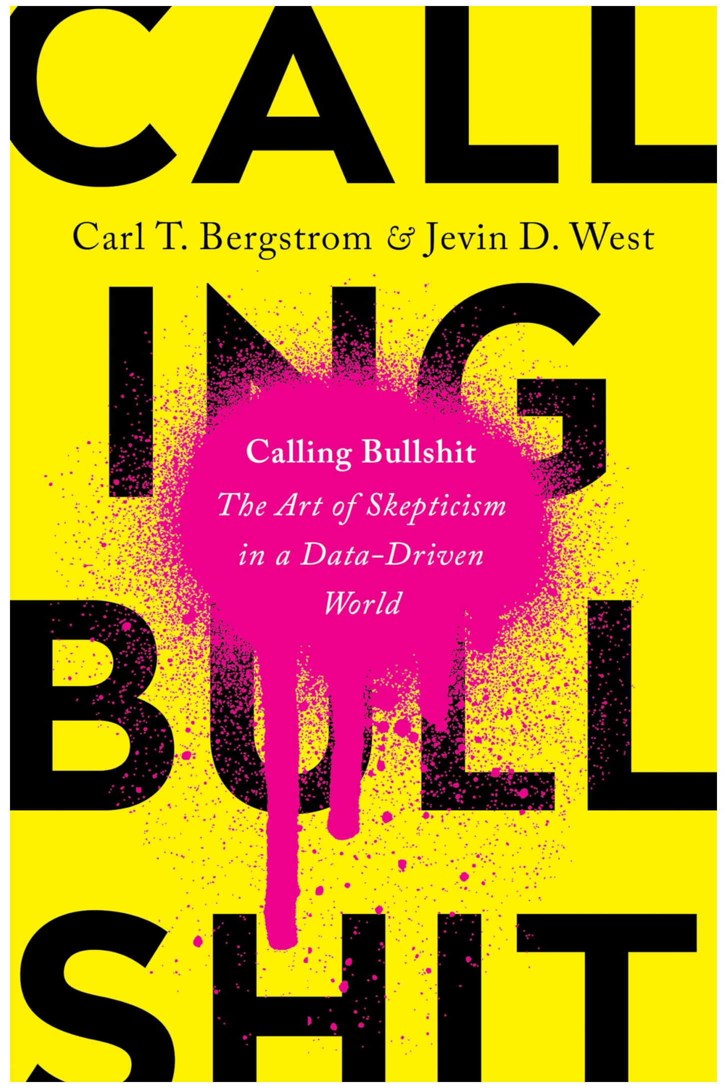
What Do Statistical Associations Reveal?



“Pairs of brooding storks in West Germany and the number of newborn human babies.”

The graph, titled “A New Parameter for Sex Education,” appeared in a humorous publication in Nature.

“Perhaps the old tall tale is right: Perhaps storks do bring babies after all.”



Predictive Tasks

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

X: **Number of firefighters in action**

Y: **Severity of the (initial) fire**

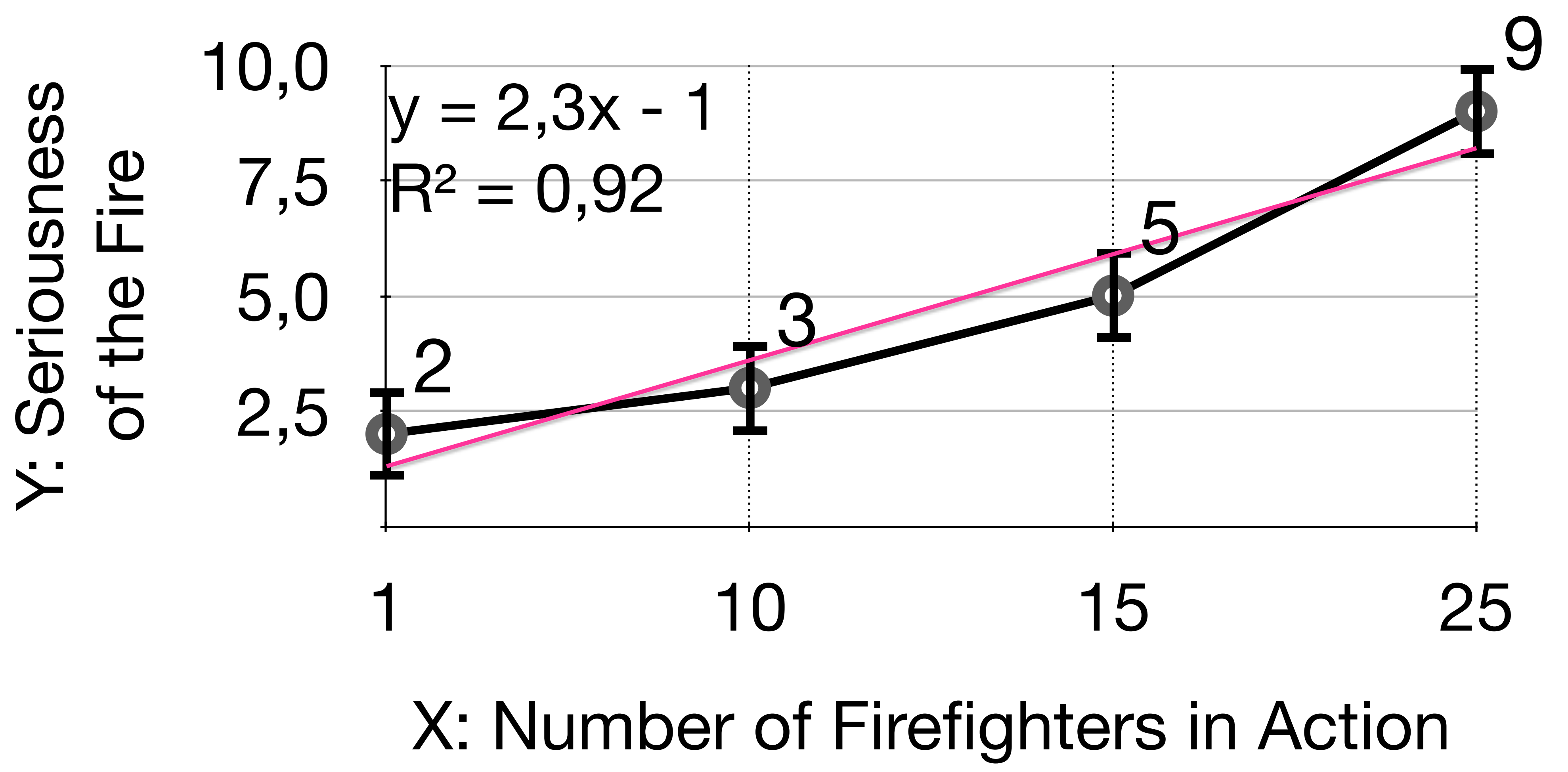
Yes!

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

$$P(Y = y \mid X = x) \neq P(Y = y)$$

Observational Probability Distribution

Correlation between severity of fire and number of firefighters in action



Positive Correlation:

The more firefighters, the stronger the fire!

The less firefighters, the weaker the fire!

Prediction ⇒ Decision-Making / Reasoning?

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

Misleading correlation: It is the size of the fire that determines the number of firefighters needed, not the other way around.

Causal Effect ≡ Effect of an Intervention

The causal direction is determined by understanding the underlying reality.

X: **Number of firefighters in action**

Y: **Seriousness of fire**

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

Underlying
Structural Causal Model (SCM)

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

$$P(Y = y \mid do(X = x)) = P(Y = y)$$

Interventional
Probability Distribution

In this case, $P(Y = y \mid see(X = x)) \neq P(Y = y)$

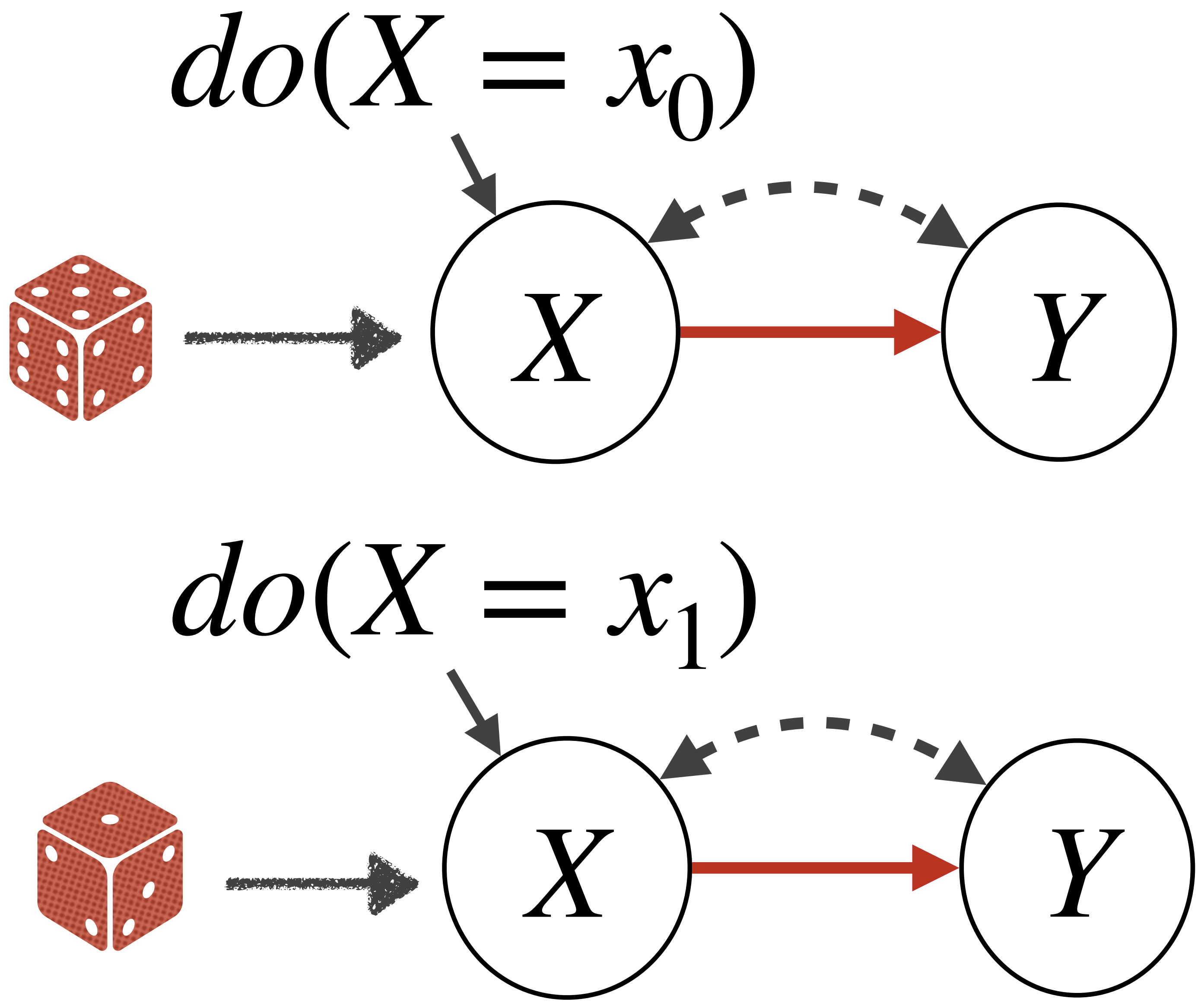
but $\forall x, P(Y = y \mid do(X = x)) = P(Y = y)$

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

Randomized Experiments

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

Randomization of the X's assignment

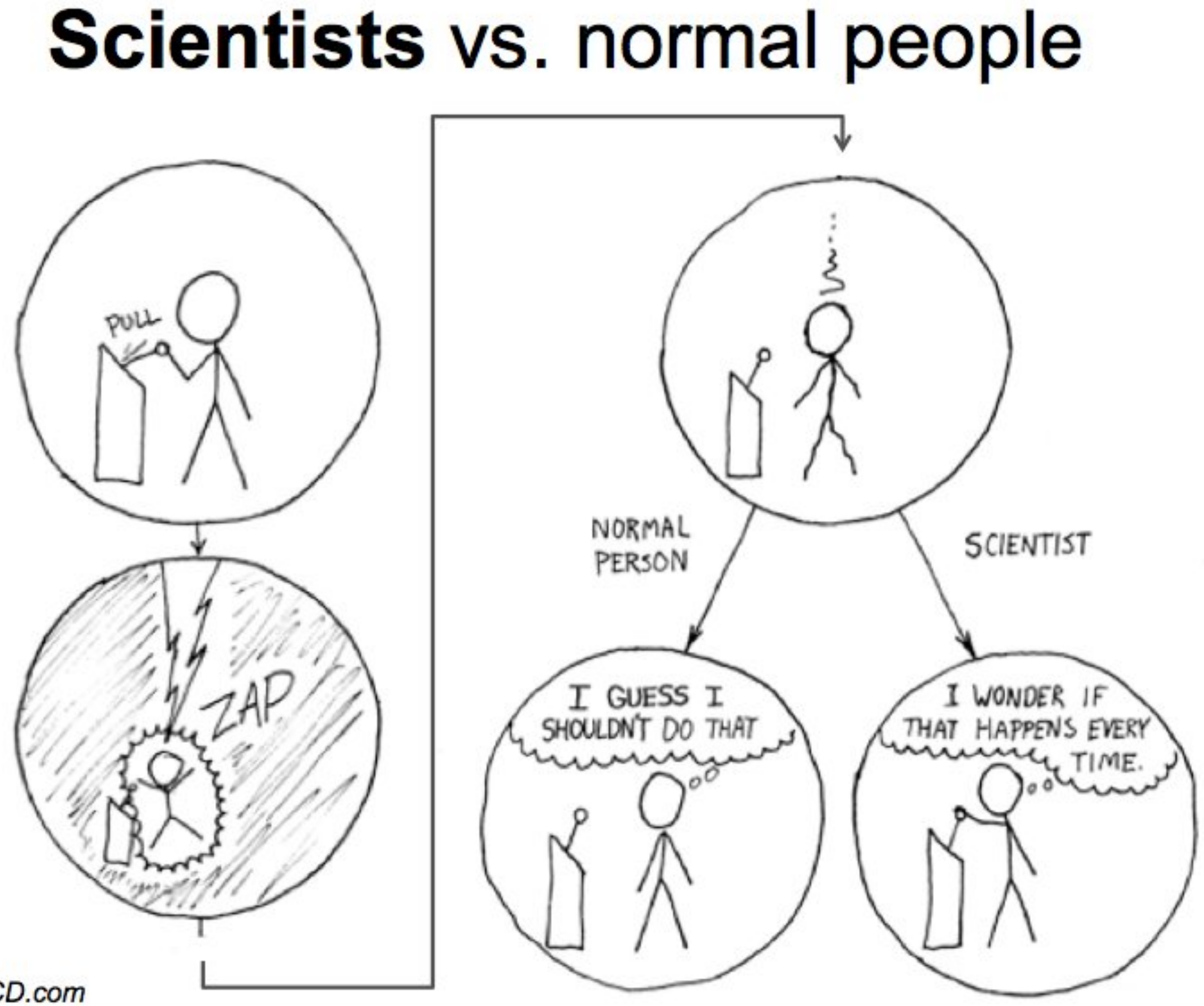


$\mathbb{E}[Y|do(X=x_0)]$

$\mathbb{E}[Y|do(X=x_1)]$

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

Can we always conduct randomized experiments?



From XKCD.com

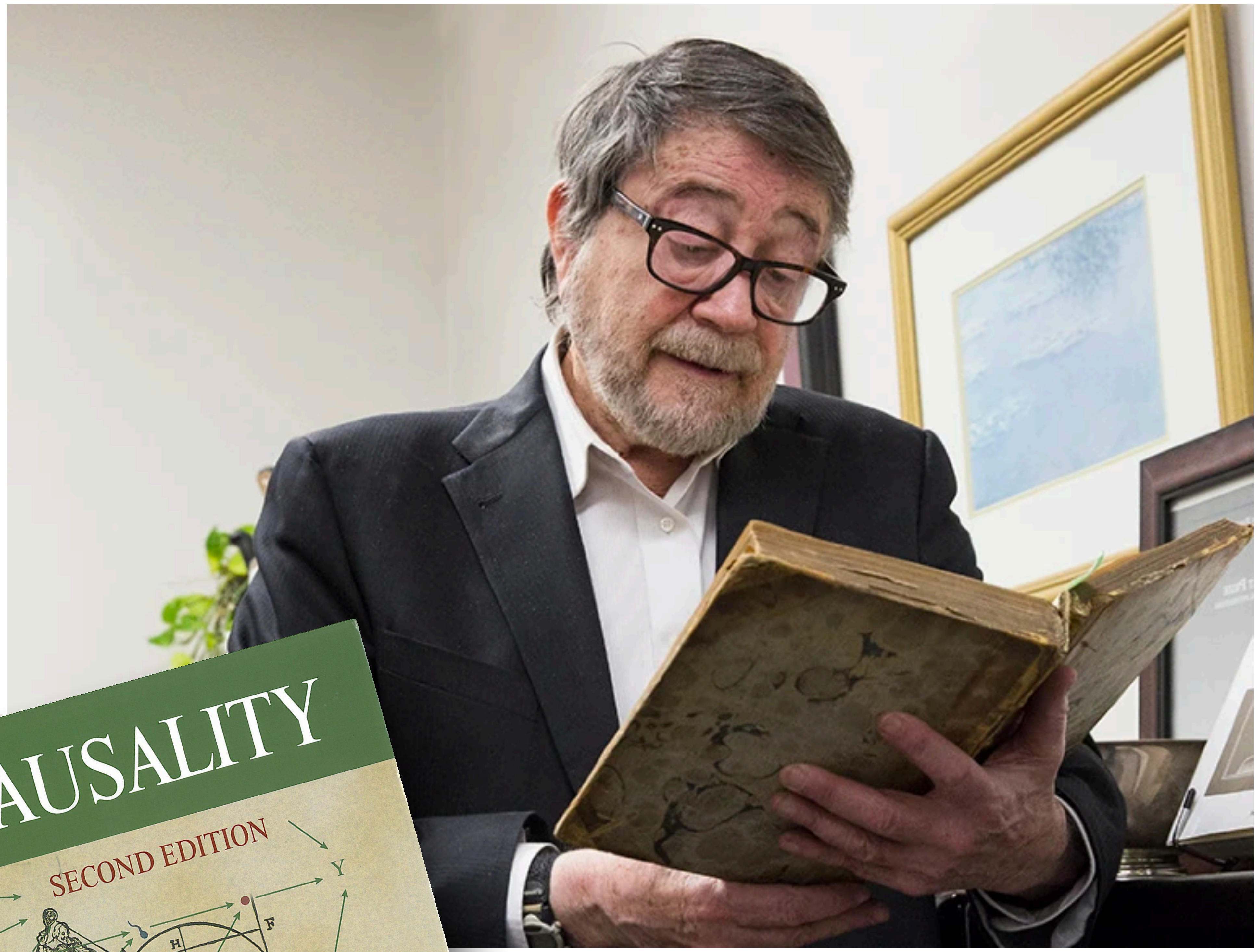
- Safety, ethical concerns
- Practical limitations
- Logistical challenges

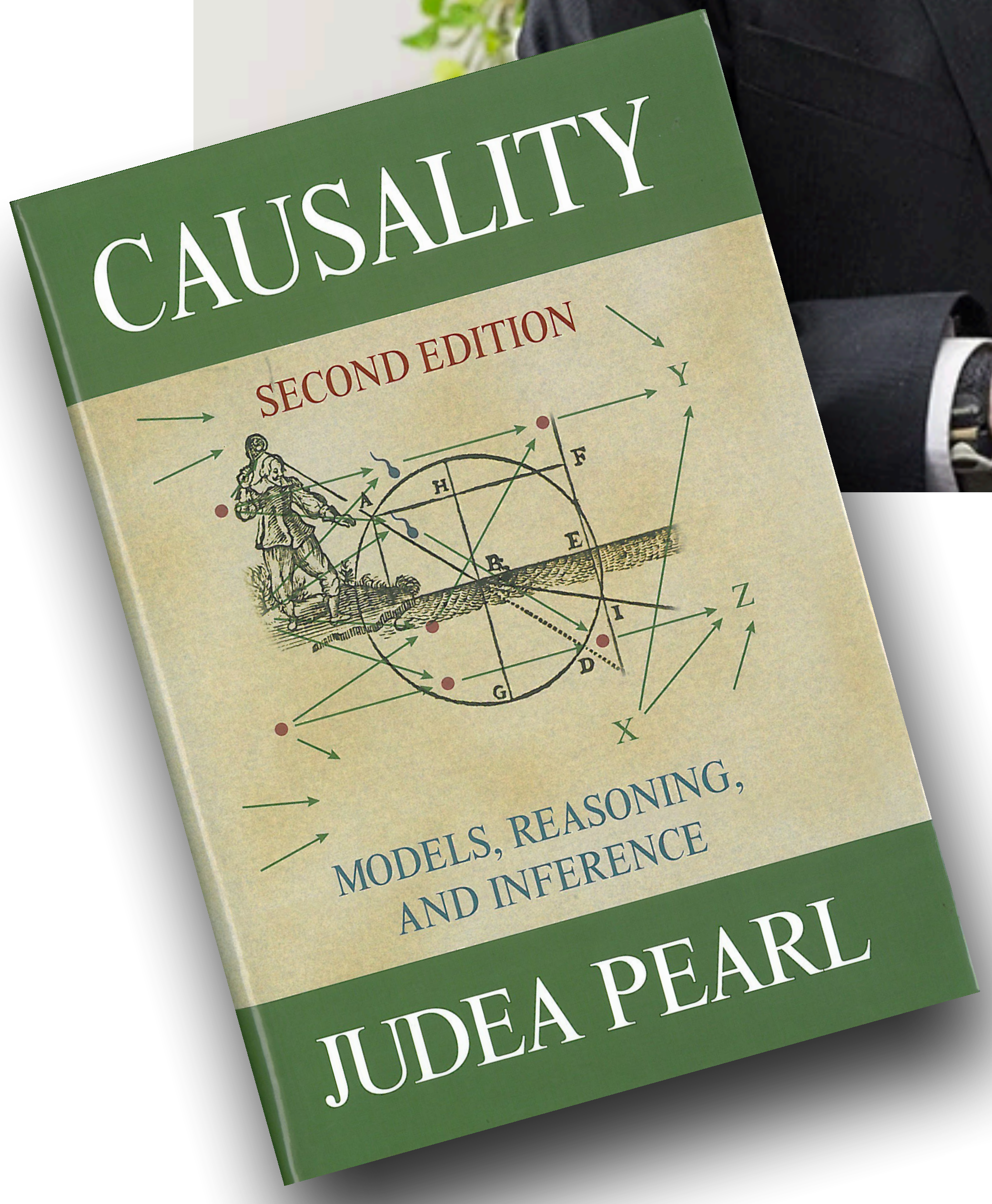
Was pulling the lever really the cause of the lightning?

The Mathematical Framework of Causal Data Science

Can we estimate causal effects without / minimizing randomized experiments?

Judea Pearl — Causality





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

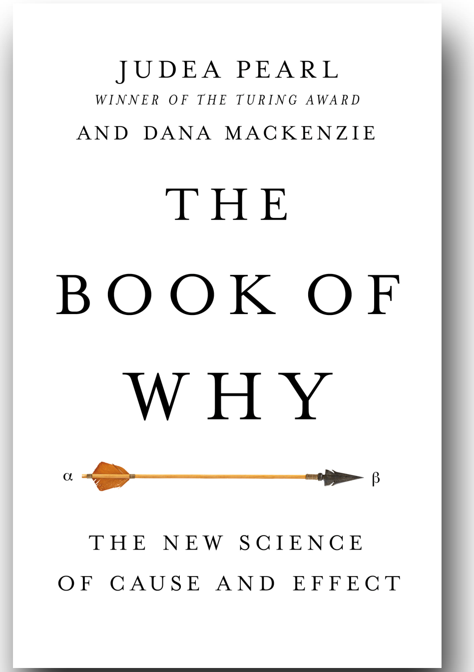
In 2011, he won the A. M. Turing Award:

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

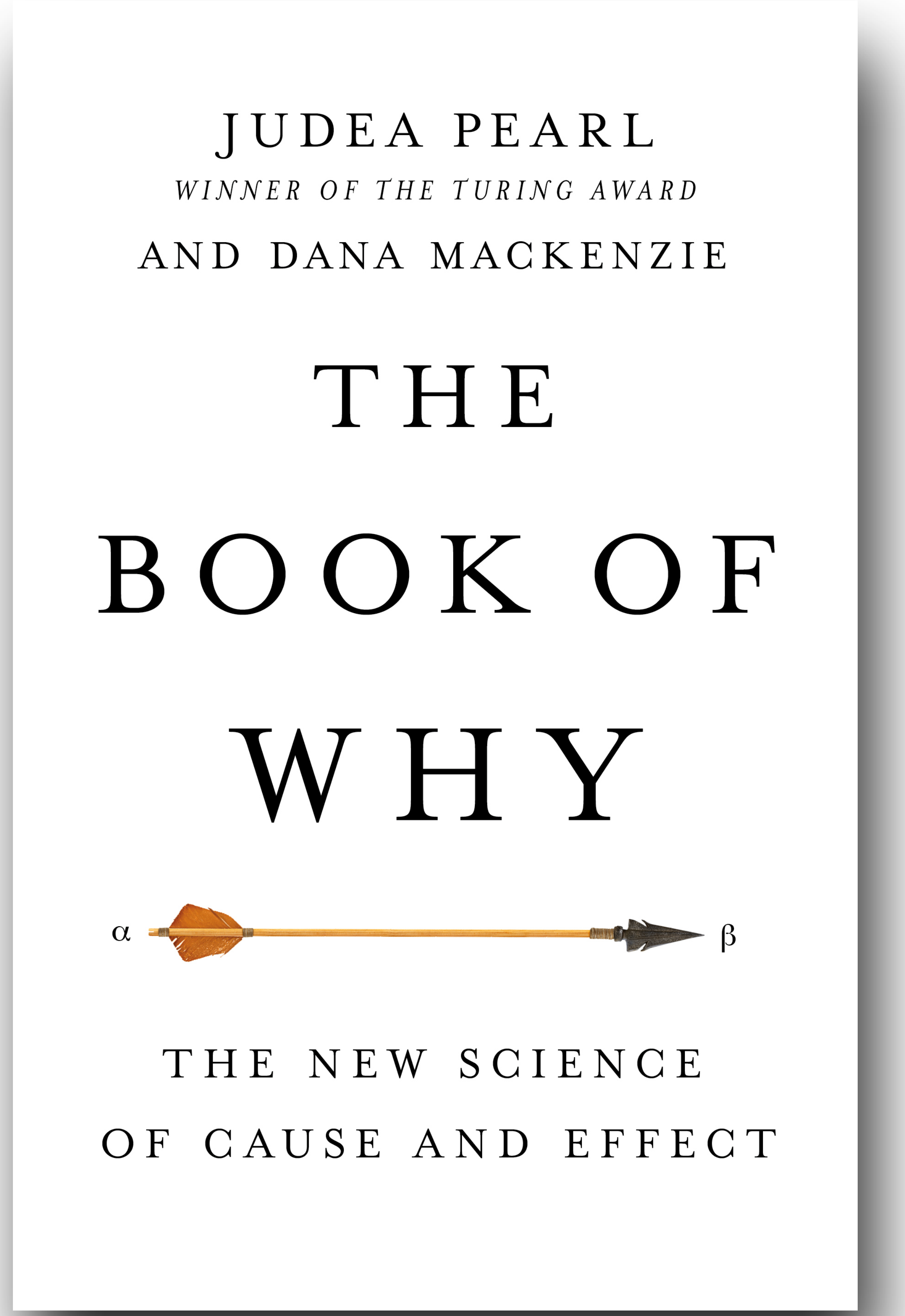
— Association for Computing Machinery (ACM)

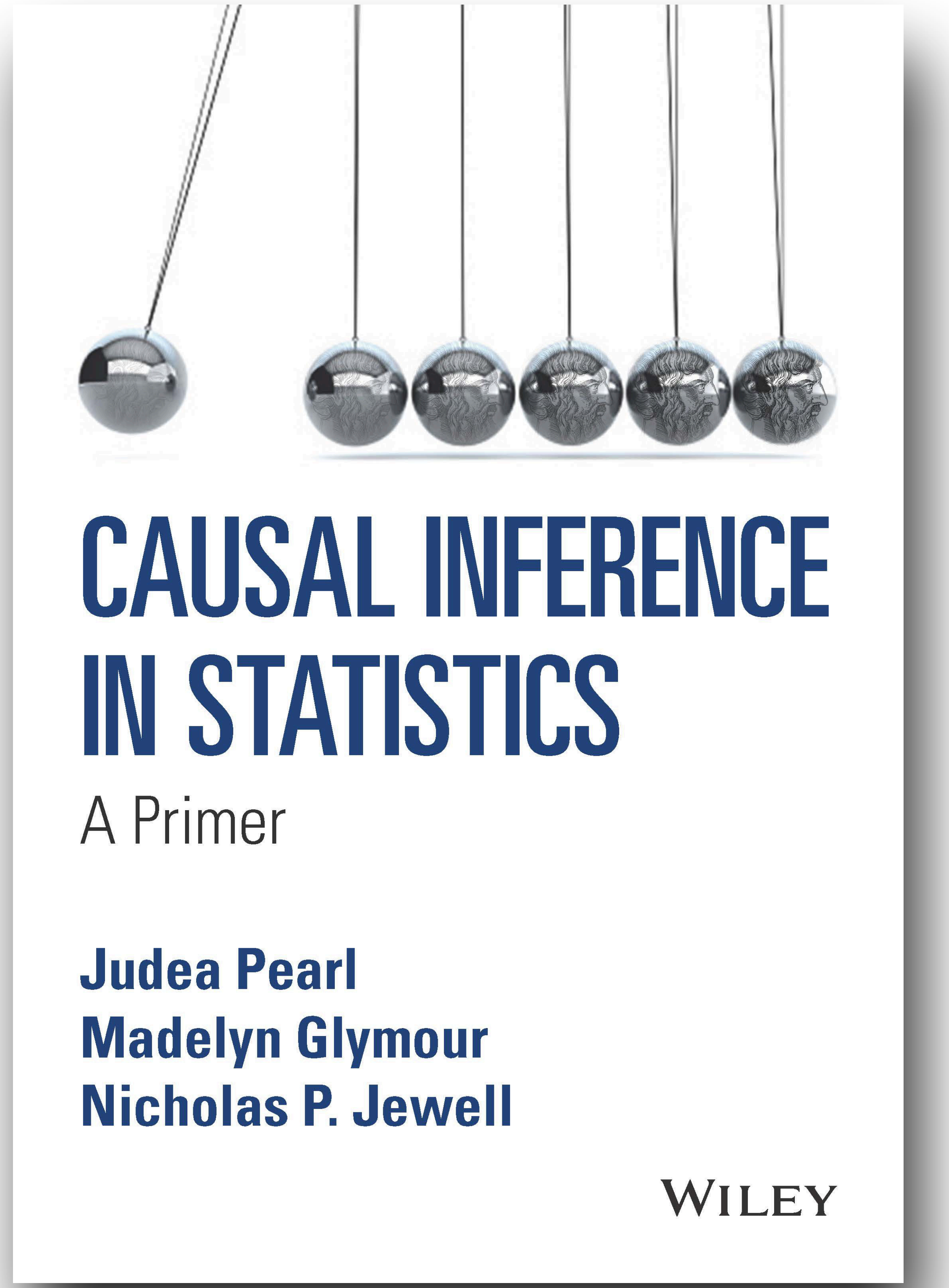
"Deep learning has instead given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality."

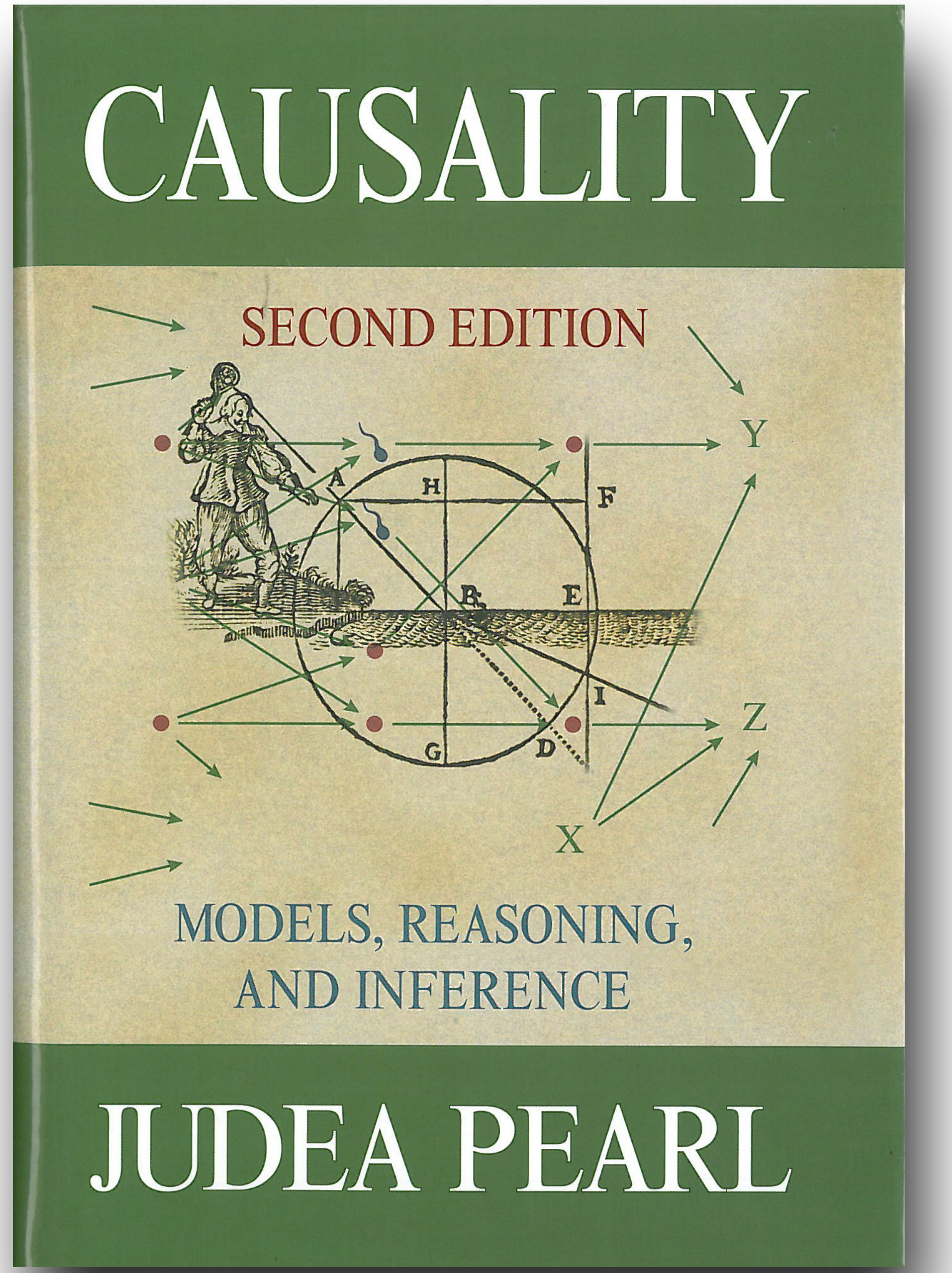
— The Book of Why: The New Science of Cause and Effect



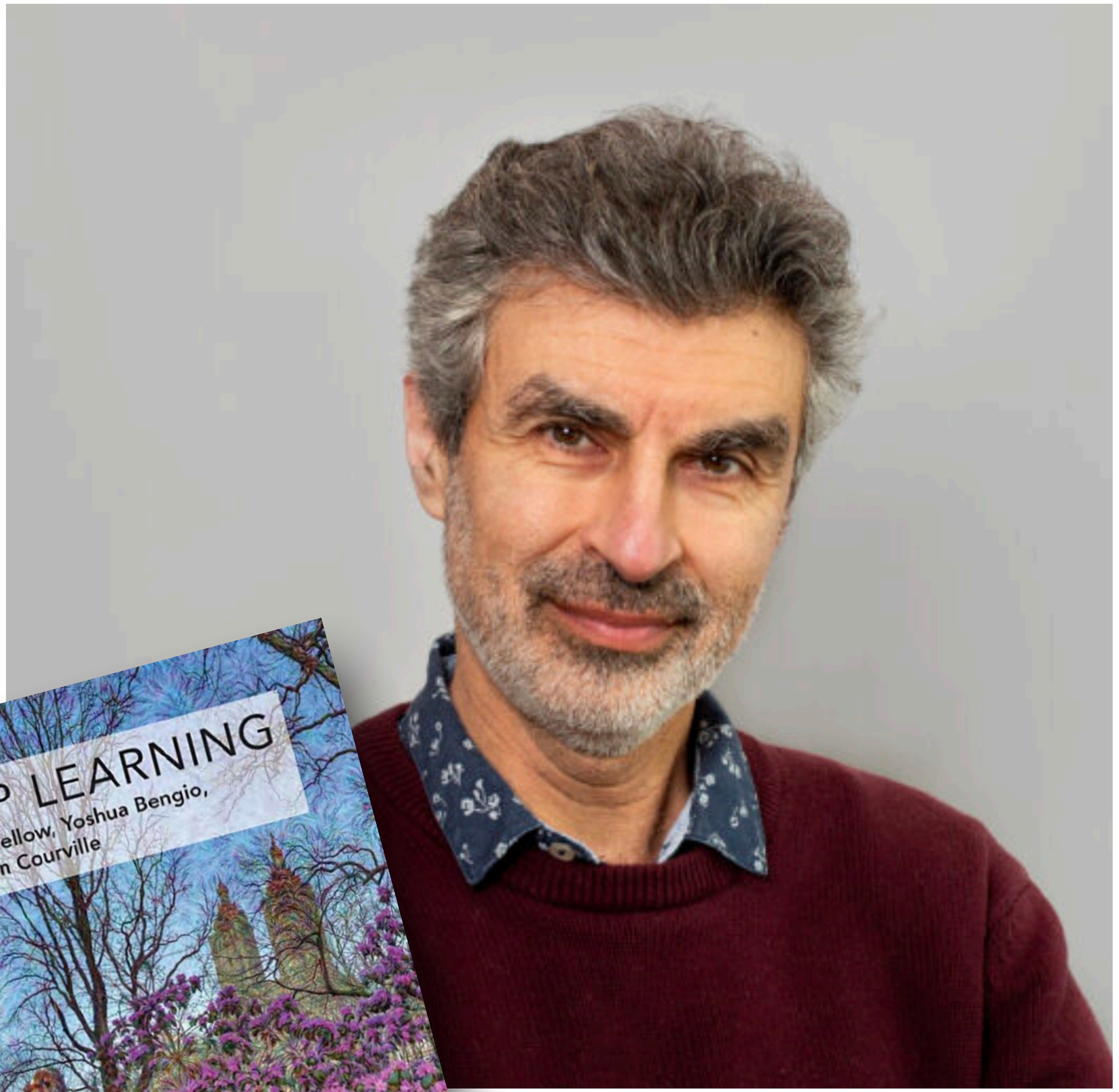
Causality Theory by Judea Pearl

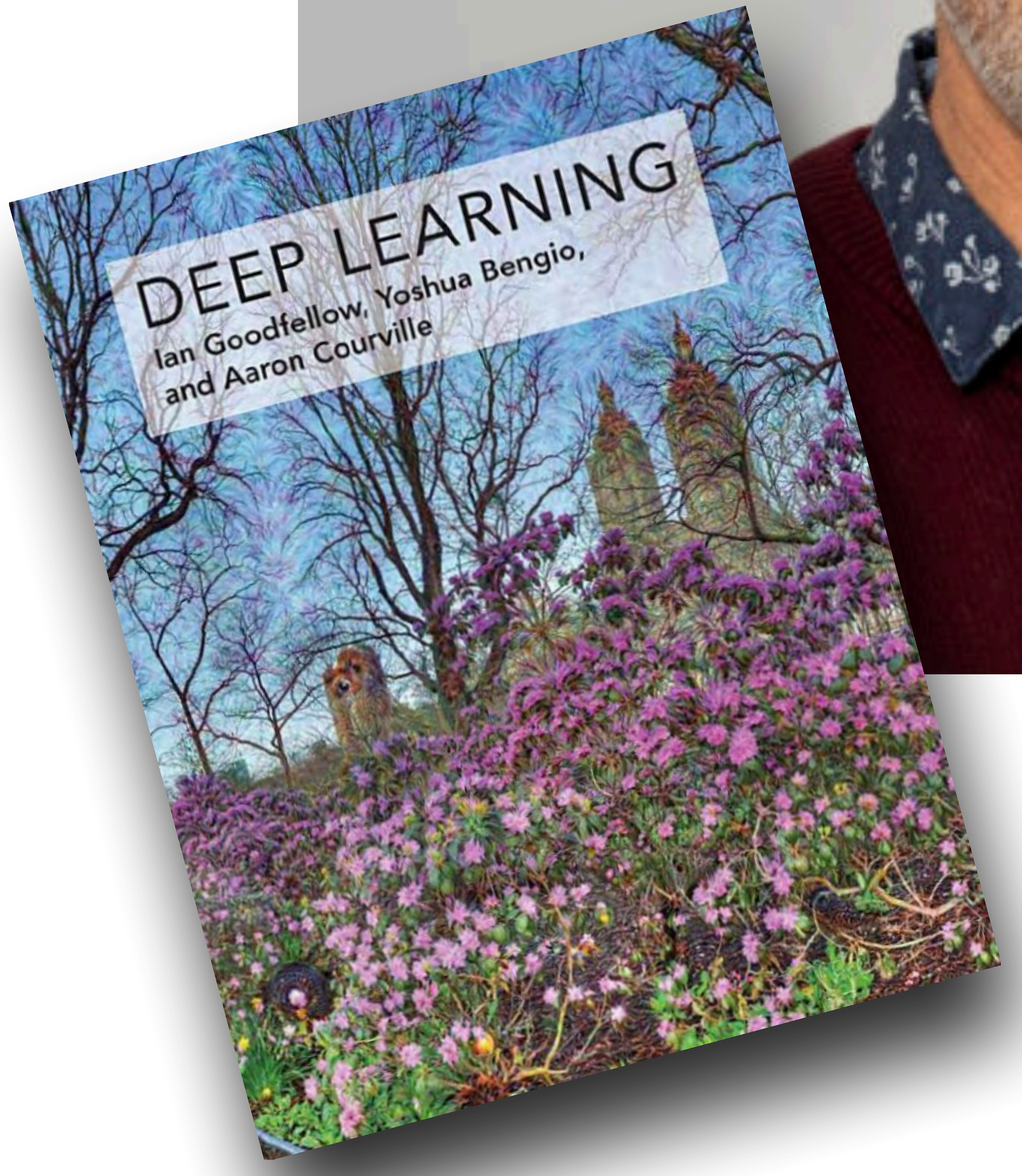






Yoshua Bengio — Deep Learning





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*, 2020.

Guido W. Imbens & Joshua D. Angrist

Guido W. Imbens

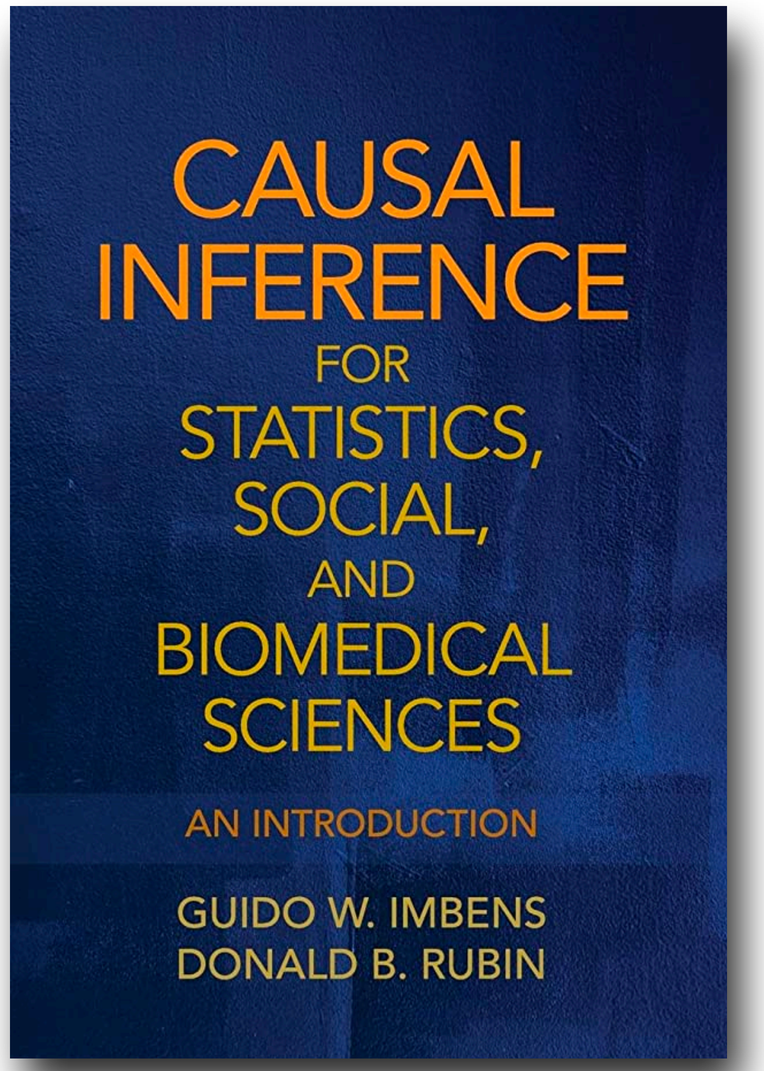
Professor of Applied Econometrics in Stanford University

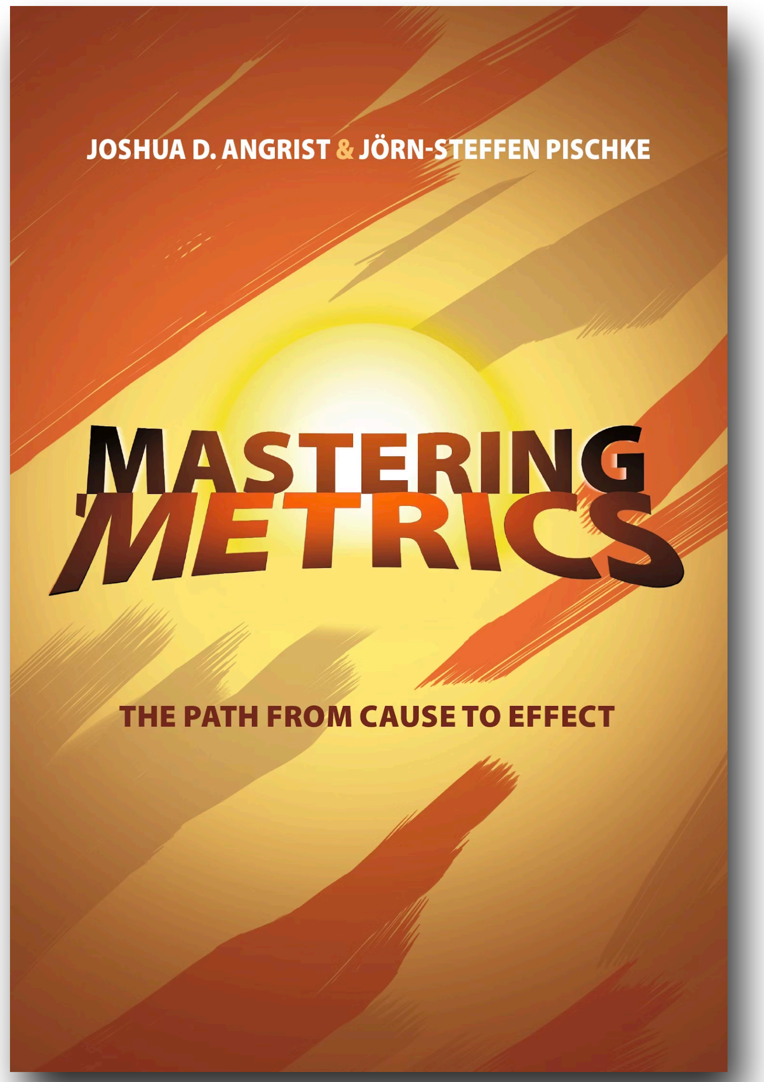
Joshua D. Angrist

Professor of Economics at the Massachusetts Institute of Technology

In 2021, they won the Nobel Prize in Economics (about $1 million)

“for their methodological contributions to the analysis of causal relationships”

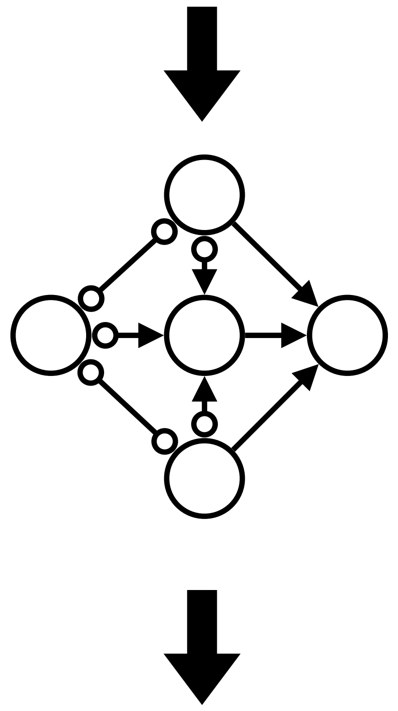




Causality: A Key to Overcoming AI's Greatest Challenges

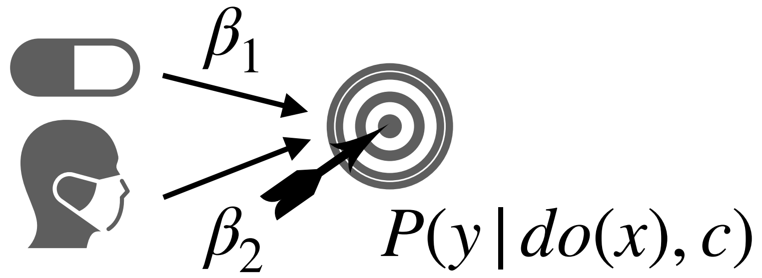
Data Fusion: Provides language and inferential machinery to cohesively combine prior knowledge and data from multiple and heterogeneous studies.

- **Causal Modeling, Causal Representation Learning and Causal Abstraction**



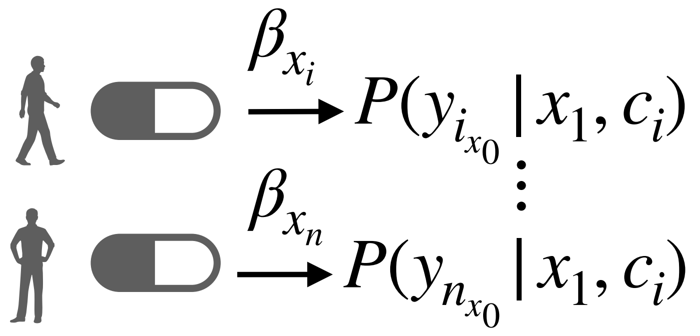
Explainability: Provides a better understanding of the true underlying mechanisms

- **Causal Discovery**

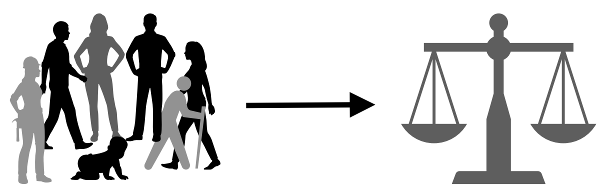


Optimal Decision Making: Can determine the *unbiased* effect of *unrealized* interventions, distinguishing between association and causation, rather than just predicting outcomes.

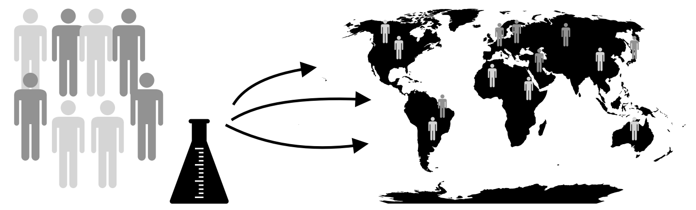
- **Causal Effect Identification and Estimation**



Personalized Inferences: Enables ***counterfactual reasoning*** by considering alternate scenarios and individual variability.



Fairness: Identifies and disentangles any mechanisms of discrimination, whether direct or indirect (potentially mediated or confounded).



Generalizability: Enables effect *transportability* across different populations.

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

- $\mathbf{V} = \{V_1, \ldots, V_n\}$: are endogenous variables
- $\mathbf{U} = \{U_1, \ldots, U_m\}$: are exogenous variables
- $\mathcal{F} = \{f_1, \ldots, f_n\}$: 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
    - $U_i \subseteq \mathbf{U}$ 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

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

$do(X = x)$

Post-Interventional / Interventional SCM

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

Observational Distribution

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

$\neq$

Interventional Distribution

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

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

$P(Y = y \mid X = x) \neq P(Y = y) \implies X$ is ***correlated*** to Y

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

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

Structural Equation Model (SEM)

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

- **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 <u>requires</u> 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!

Modeling Structural Causal Knowledge

Acyclic Directed Acyclic Graph (ADMG)
Causal Diagrams

CBN: Encoder of Structural Causal Knowledge

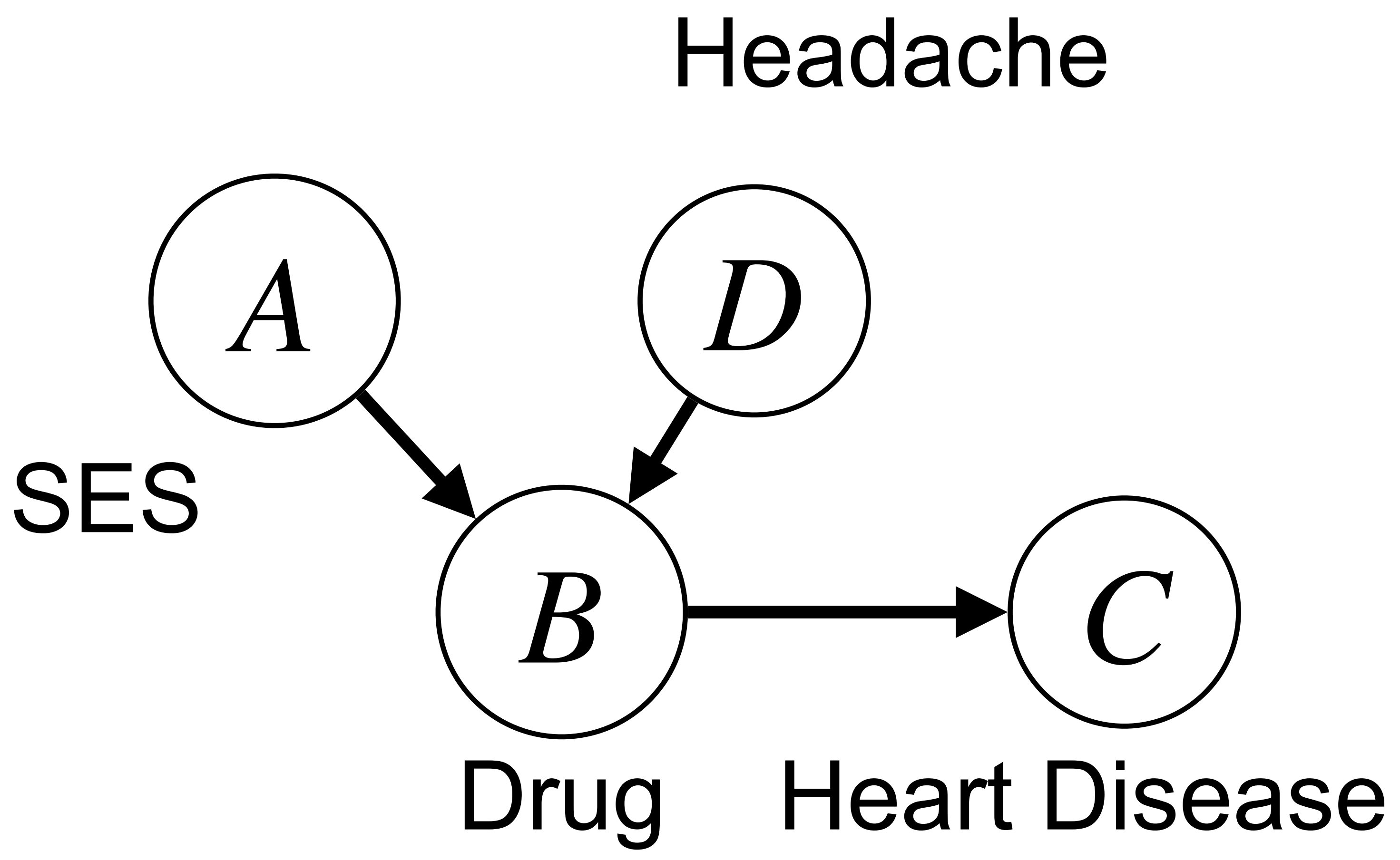
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

Causal Diagram



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ 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 $f_j \in \mathscr{F}$.

CBN: Encoder of Structural Causal Knowledge

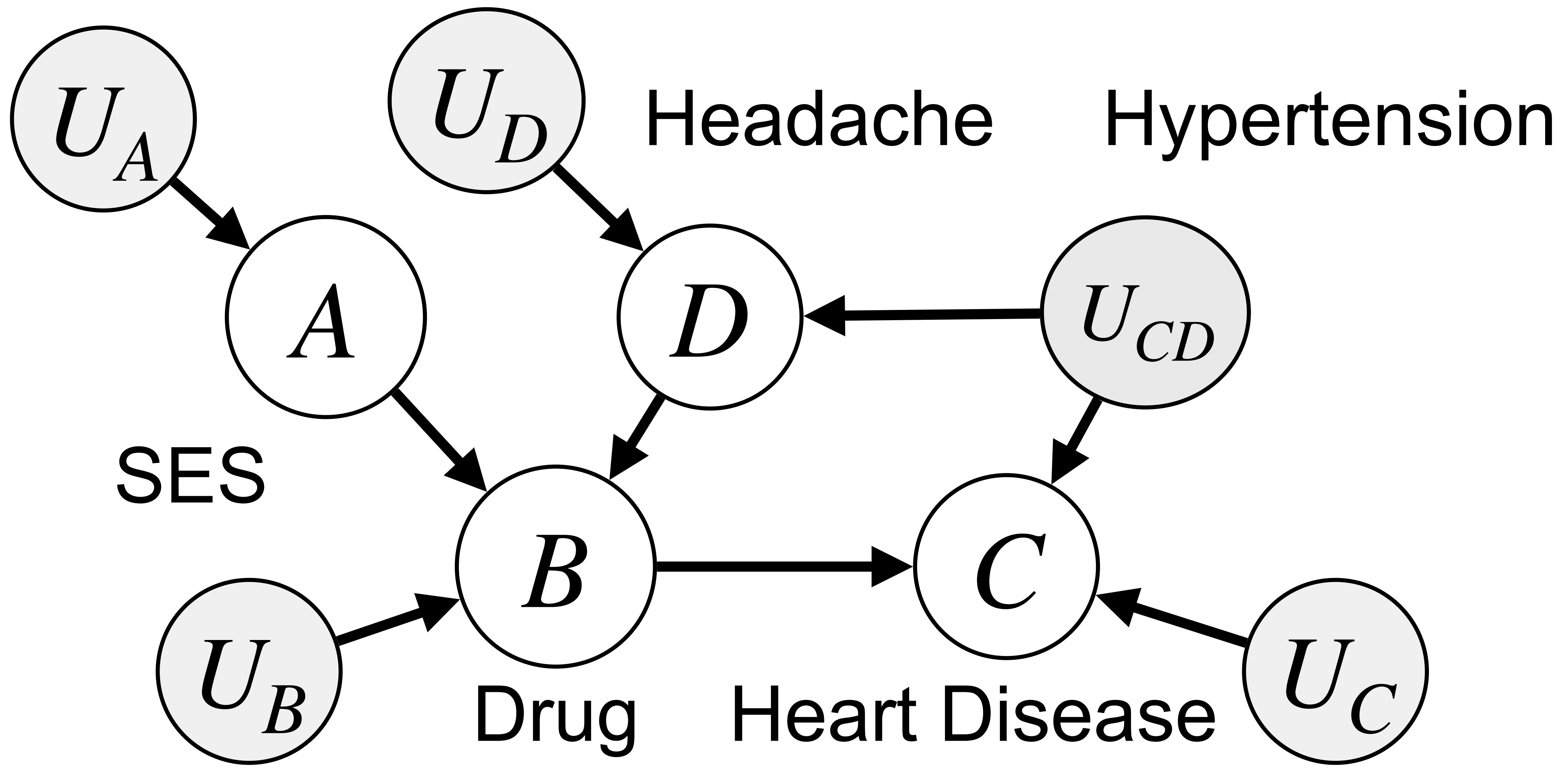
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

Causal Diagram



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

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

CBN: Encoder of Structural Causal Knowledge

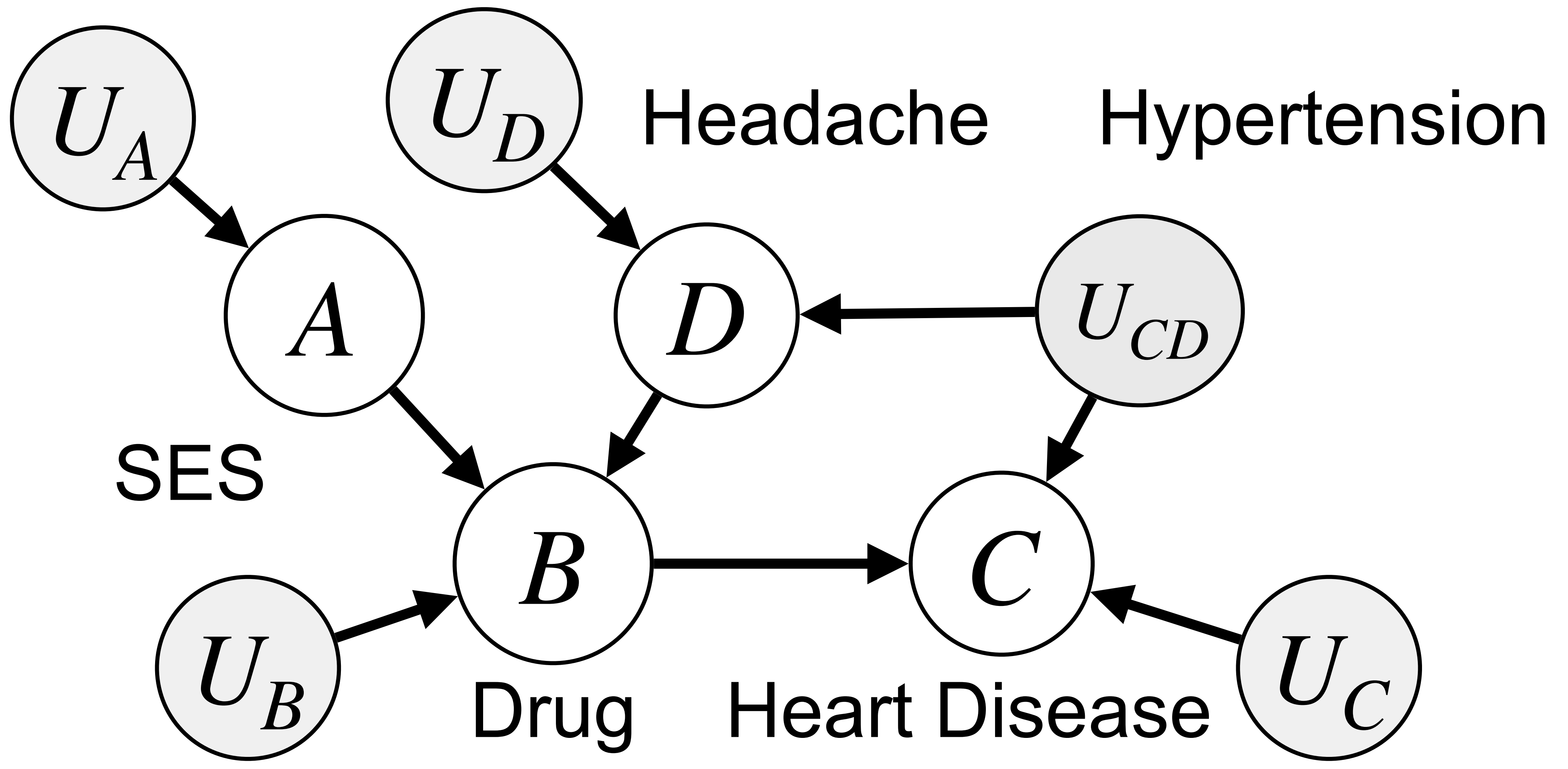
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

Causal Diagram



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ 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 $f_j \in \mathscr{F}$.

$V_i \dashleftarrow\dashrightarrow V_j$ 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

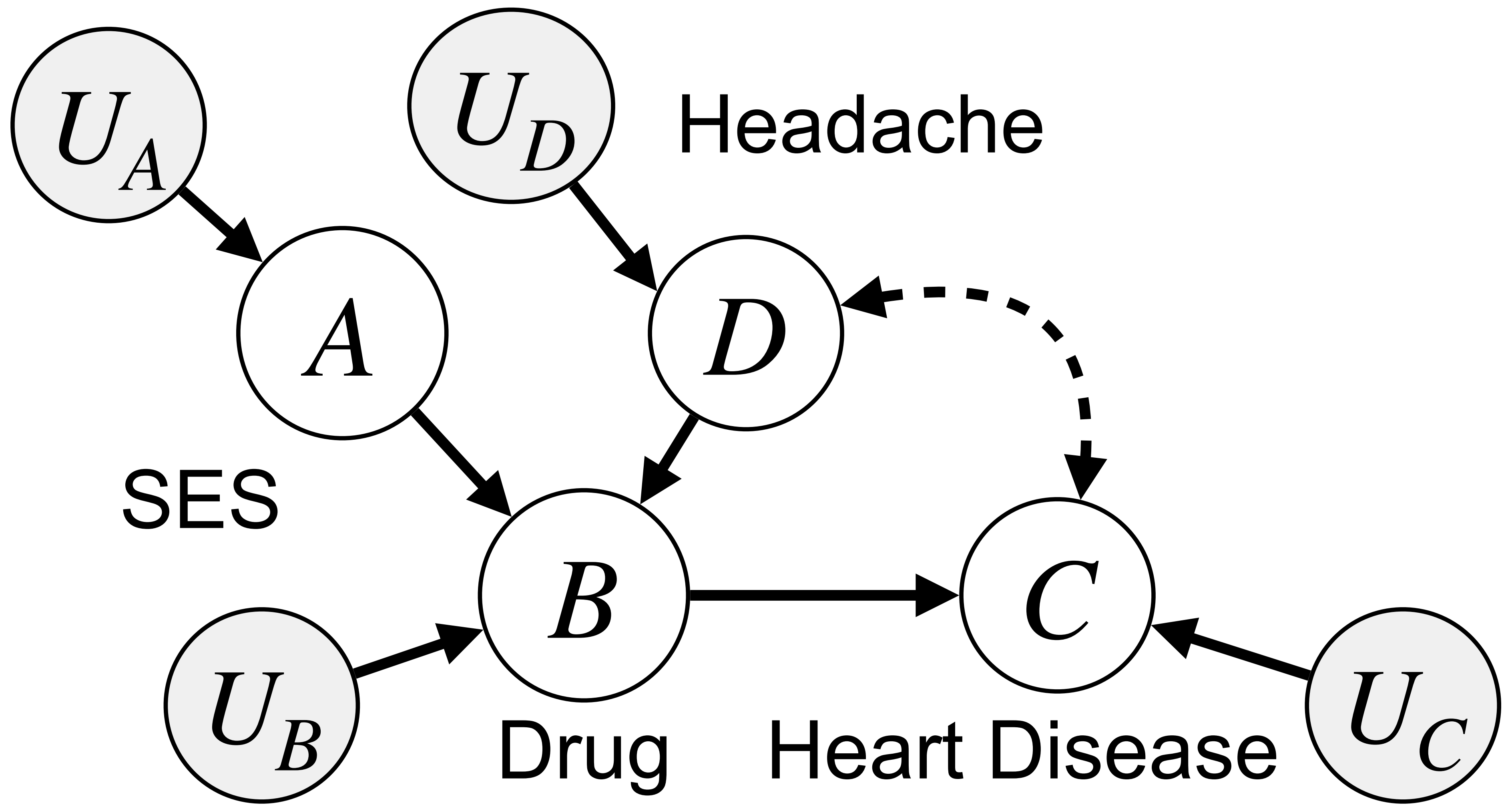
Structural Causal Model (SCM)

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

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

Induced Causal Bayesian Network (CBN)

Causal Diagram



An SCM $\mathscr{M} = \langle \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) \rangle$ 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 $f_j \in \mathscr{F}$.

$V_i \leftarrow\!\!\dashrightarrow V_j$ 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

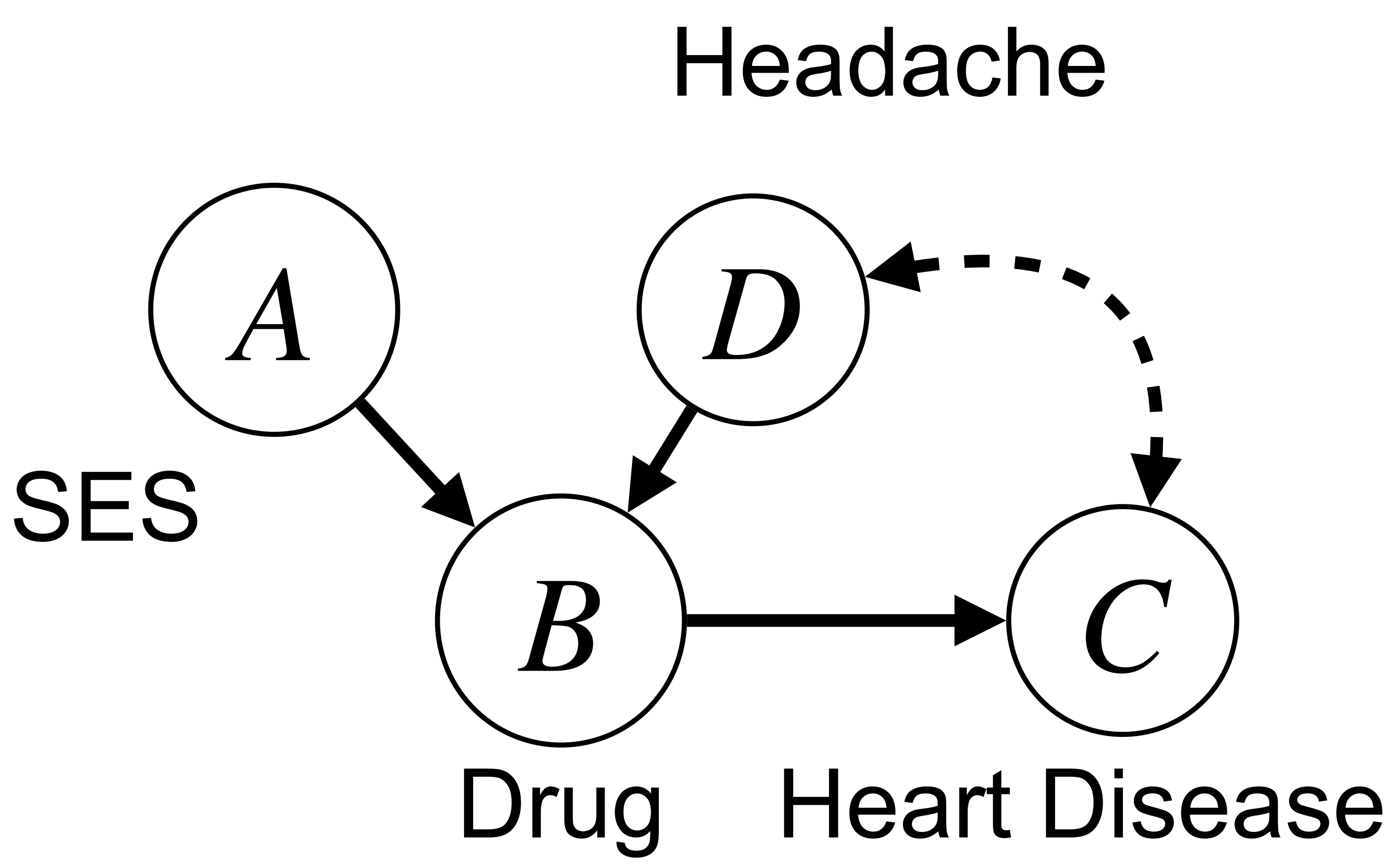
Structural Causal Model (SCM)

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

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

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 $f_j \in \mathcal{F}$.

$V_i \dashleftarrow\dashrightarrow V_j$ 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 in a subpath $\langle V_i, V_m, V_j \rangle$ is said to be ***inactive*** relative to a set $\mathbf{Z}$ if V_m:

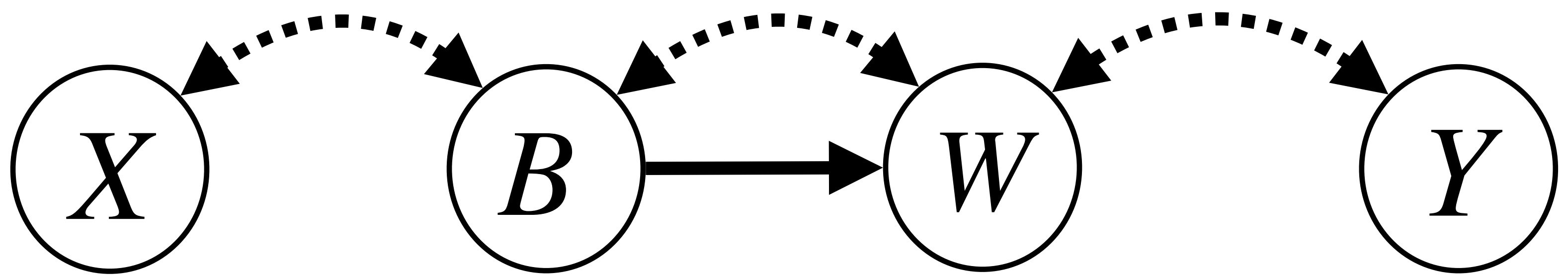
1. Is a non-collider and is in $\mathbf{Z}$; or
2. Is a collider and neither it nor any of its descendants in $\mathbf{Z}$.

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

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

Under the **global Markov condition**: $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_P$

D-separations in G imply conditional independencies in P



Does $\mathbf{Z}$ d-separates X and Y ?

$\mathbf{Z}$: ✓ $\{\}$ ✓ $\{B\}$ ✗ $\{W\}$ ✗ $\{B, W\}$

We have that $(X \perp\!\!\!\perp Y)_G$, $(X \perp\!\!\!\perp Y \mid B)_G$, and $(X \not\perp\!\!\!\perp Y \mid W)_G$, but $(X \not\perp\!\!\!\perp Y \mid B, W)_G$

Causal Effect Identification Given a Causal Diagram

Graphical Criteria, Do-Calculus, and ID-Algorithm

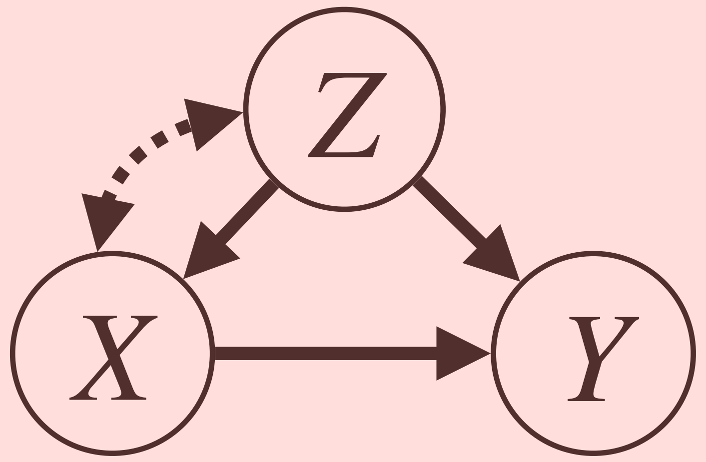
Classical Causality Pipeline from a Causal Diagram

Structural knowledge available

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 $P(x, m, y)$

Observational Distribution

Inference Engine

Solution

yes / no

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

$$\mathbb{E}(y \mid do(x)) = \sum_z \mathbb{E}(y \mid x, z) P(z)$$

Interventional Distribution ← Available Distributions

Observational Data

$$\rightarrow \hat{\mathbb{E}}(y \mid do(x)) = \sum_z \hat{\mathbb{E}}((y \mid x, z) \hat{P}(z)$$

Causal Modeling → Causal Effect Identification → Causal Effect Estimation

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 $P(\mathbf{Y} \mid do(\mathbf{X}))$ 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:

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

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

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

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

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

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

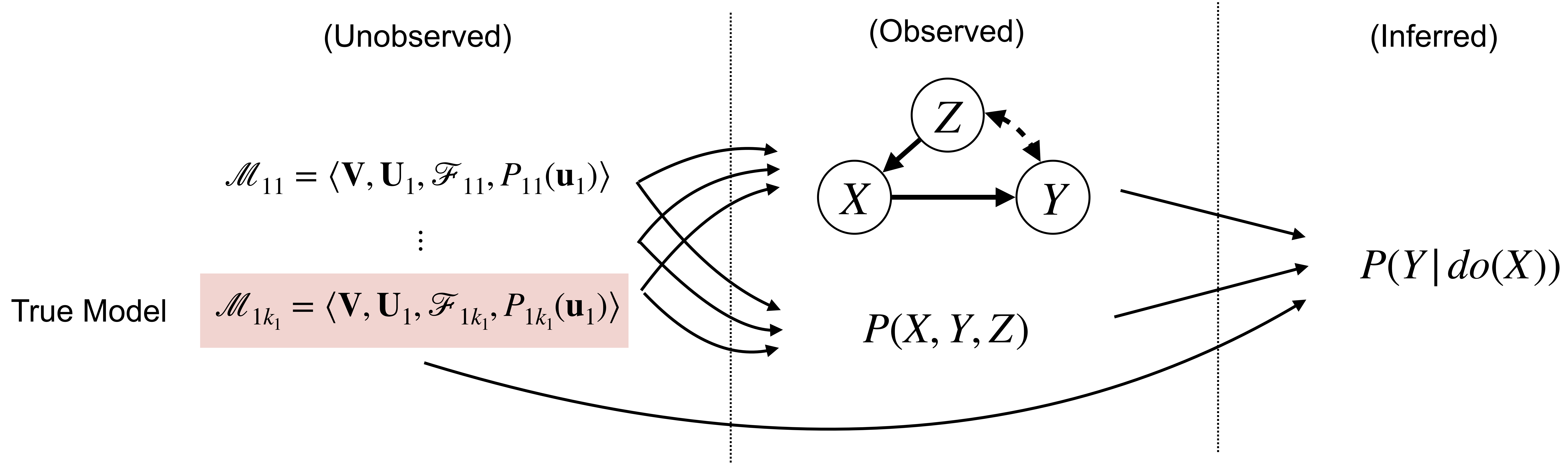
$$\mathbb{E}[\mathbf{Y} \mid do(\mathbf{X} = \mathbf{x})] = \int_{\Omega_{\mathrm{y}}} \mathbf{y} P(\mathbf{y} \mid do(\mathbf{x})) d\mathbf{y},$$

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

The Effect Identification Problem

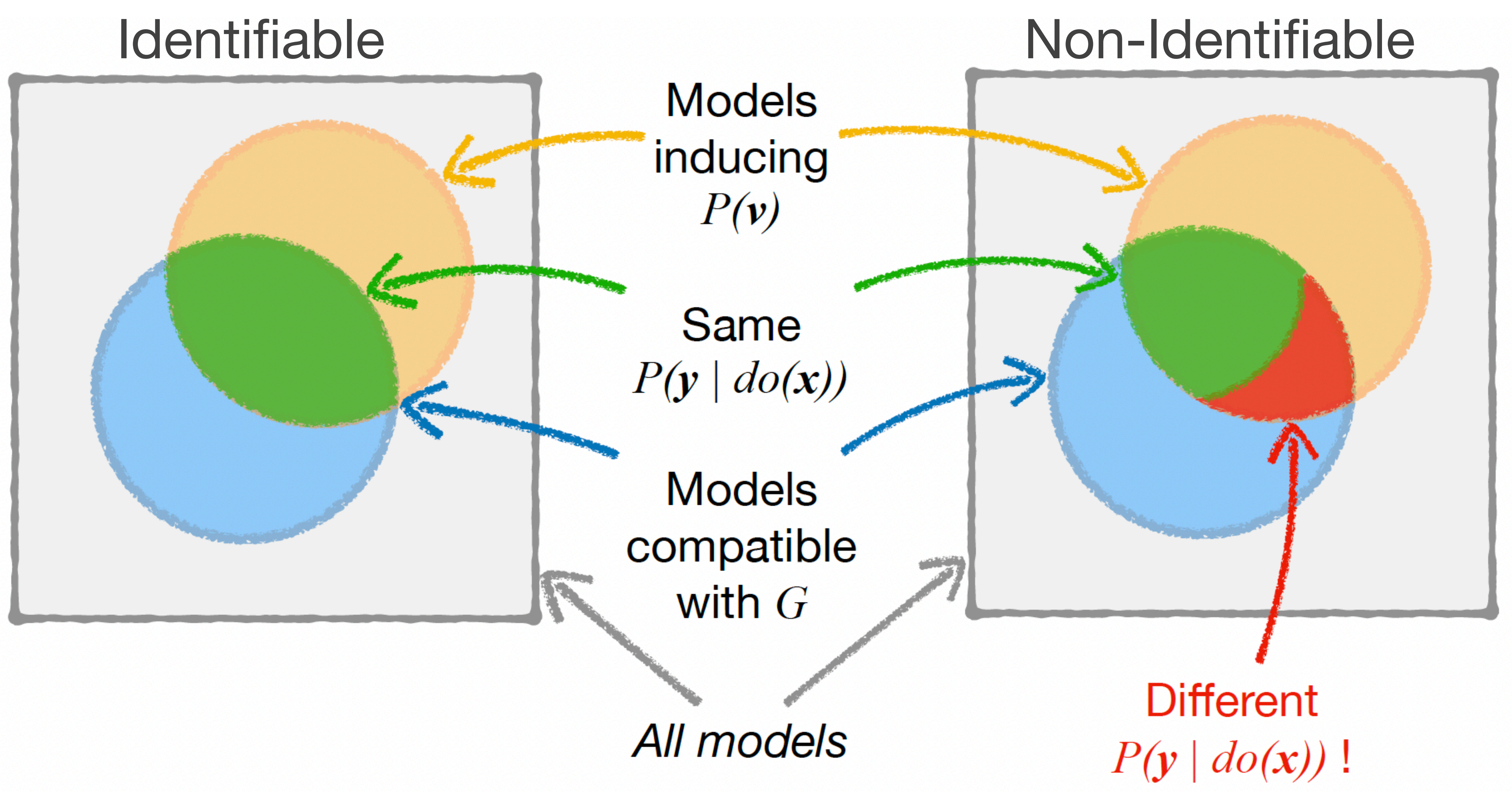
Causal Effect Identifiability: The effect of $\mathbf{X}$ on $\mathbf{Y}$ is said to be *identifiable* from a causal diagram G and the probability distribution $P(\mathbf{V})$ if $P(\mathbf{Y} \mid do(\mathbf{X}))$ is *uniquely computable,* i.e., if for every pair of SCMs $\mathscr{M}_1$ and $\mathscr{M}_2$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



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

The Effect Identification Problem

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



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 $P(\mathbf{y} \mid do(\mathbf{x}))$.

Tools for Causal Identification

1. Truncated Factorization / G-computation formula

Markovian Models

2. Graphical criteria
    1. Parent adjustment
    2. Backdoor Adjustment
    3. Front-door Adjustment

A few interesting (albeit still constrained) scenarios

3. Do-Calculus (a.k.a Causal Calculus)
4. Identify Algorithm (a.k.a. ID algorithm)

General Semi-Markovian Scenarios

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 via Backdoor Criterion

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

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:

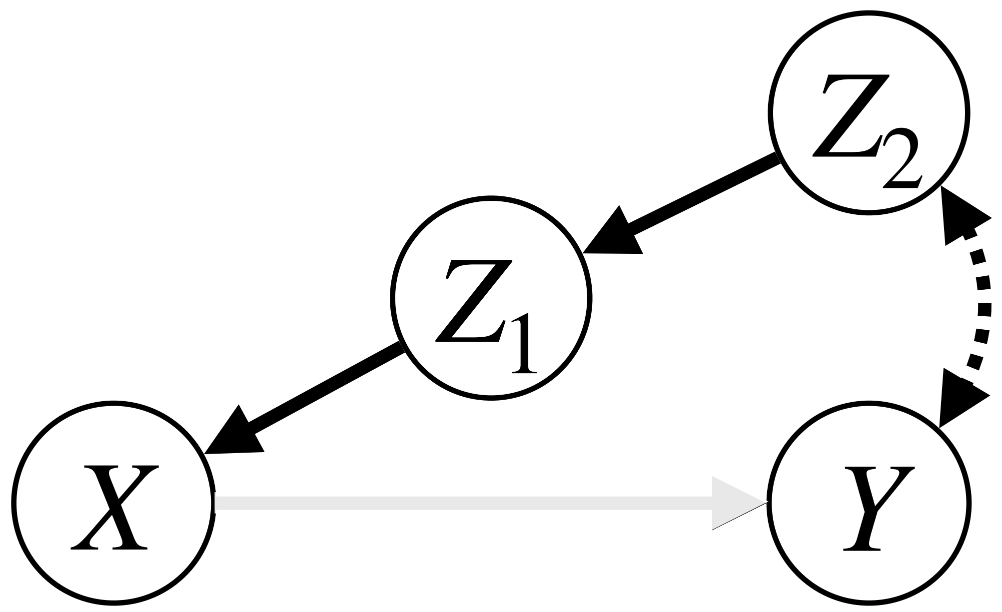
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P(\mathbf{z})$$

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

$\mathbf{Z} = \{Z_1\}$

$\mathbf{Z} = \{Z_1, Z_2\}$

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



Judea Pearl. Comment: Graphical models, causality and intervention. Stat. Sci., 8:266–269, 1993.

Identification via Backdoor Criterion

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

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:

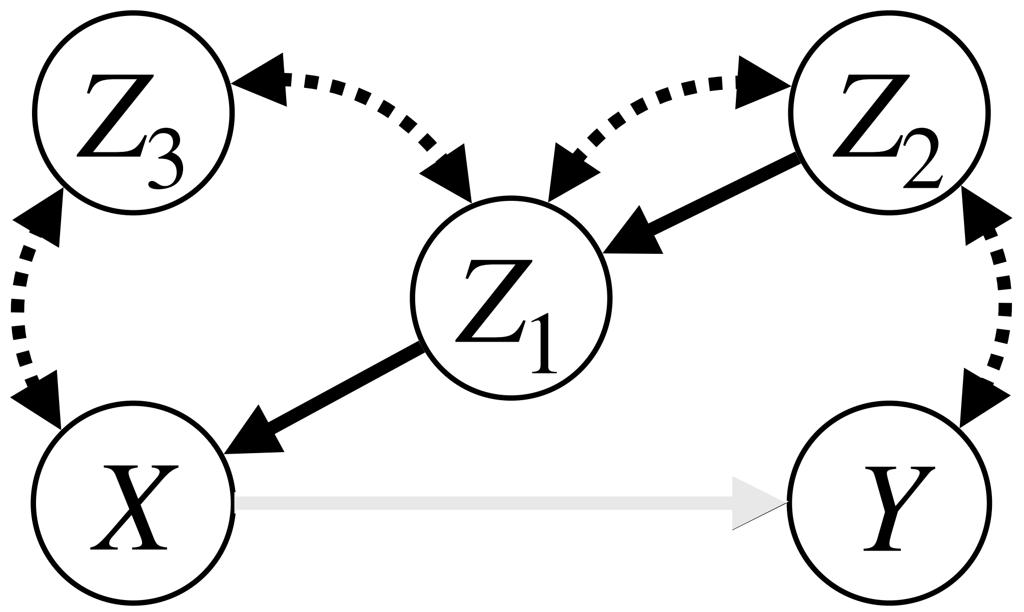
1. $\mathbf{Z}$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the graph $G_{\underline{\mathbf{X}}}$, i.e., the graph resulting from cutting the arrows out of $\mathbf{X}$
2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ in G (all variables in $\mathbf{Z}$ are pre-treatment)

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:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

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

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



$\mathbf{Z} = \{Z_1\}$
$\mathbf{Z} = \{Z_1, Z_2\}$

Judea Pearl. Comment: Graphical models, causality and intervention. Stat. Sci., 8:266–269, 1993.

Identification via Backdoor Criterion

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

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 $\mathbf{X}$

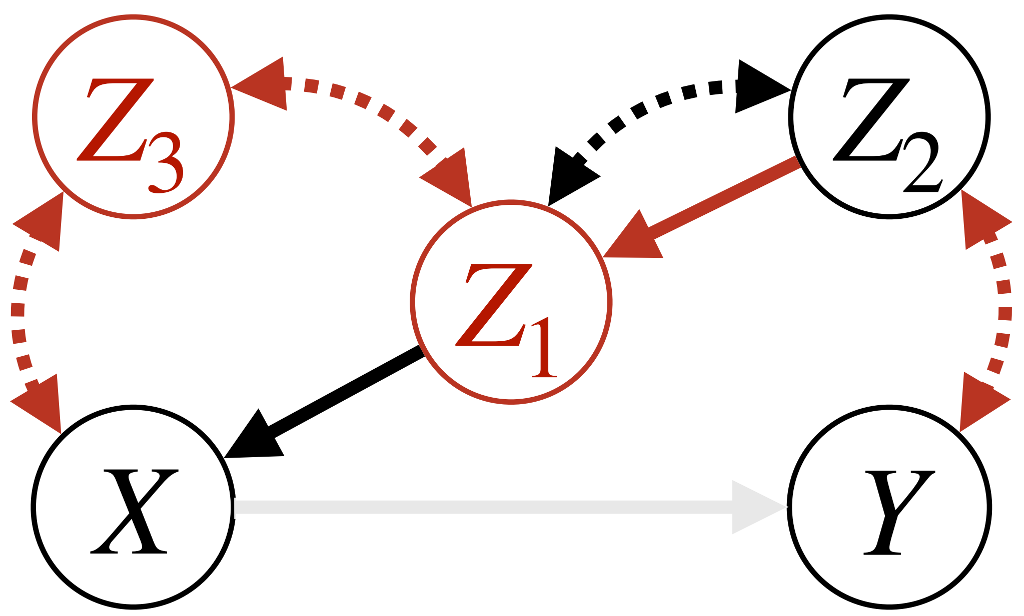
2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ in G (all variables in $\mathbf{Z}$ are pre-treatment)

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:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

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

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



$\mathbf{Z} = \{Z_1\}$

$\mathbf{Z} = \{Z_1, Z_3\}$

$\mathbf{Z} = \{Z_1, Z_3\}$

Judea Pearl. Comment: Graphical models, causality and intervention. Stat. Sci., 8:266–269, 1993.

Identification via Backdoor Criterion

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

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)

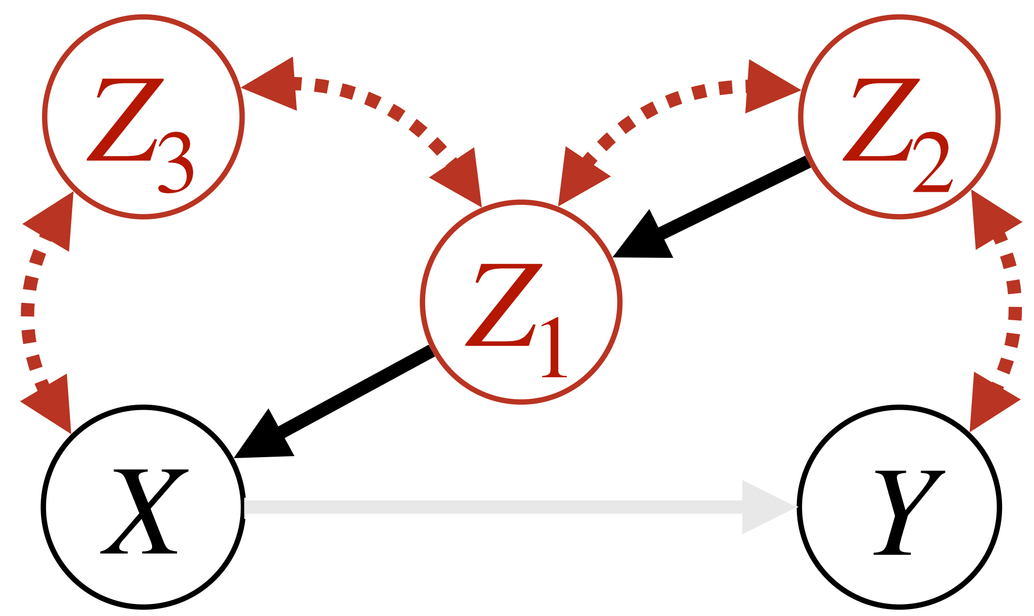
Then, $\mathbf{Z}$ satisfies the ***backdoor criterion*** for $(\mathbf{X}, \mathbf{Y})$ and, then the effect of $\mathbf{X}$ on $\mathbf{Y}$ is given by:

$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

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

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

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



$\mathbf{Z} = \{Z_1\}$

$\mathbf{Z} = \{Z_1, Z_3\}$

$\mathbf{Z} = \{Z_1, Z_2, Z_3\}$ ✗

Judea Pearl. Comment: Graphical models, causality and intervention. Stat. Sci., 8:266–269, 1993.

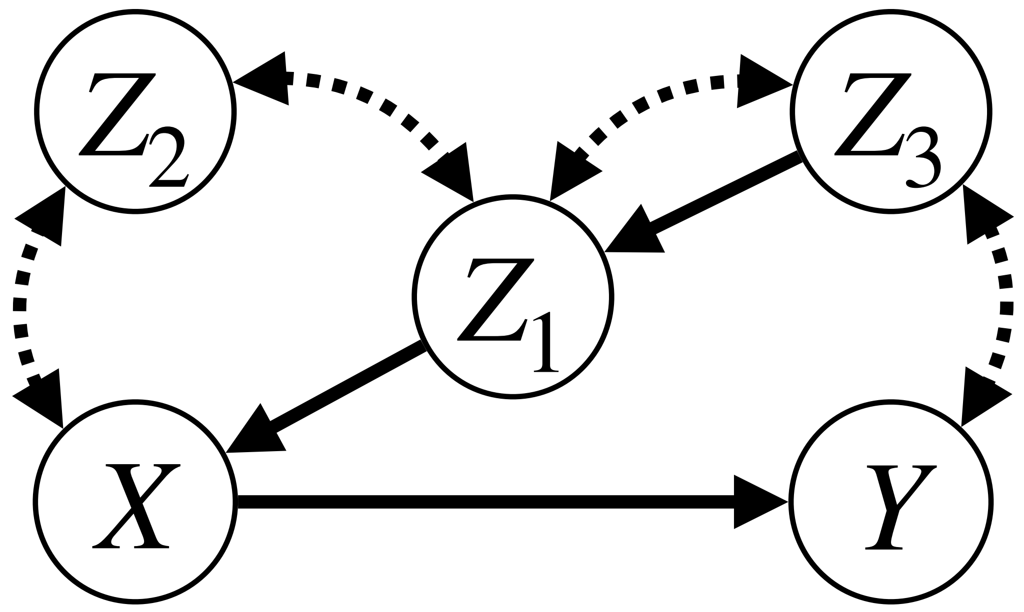
Estimation via Propensity Scores

Consider the case in which the causal effect of X on Y is identifiable through adjustment over a set of variables $\mathbf{Z}$, i.e.,

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

$$= \sum_{\mathbf{z}} \frac{P(y \mid x, \mathbf{z}) P(x \mid \mathbf{z}) P(\mathbf{z})}{P(x \mid \mathbf{z})}$$

$$= \sum_{\mathbf{z}} \frac{P(y, x, \mathbf{z})}{\boxed{P(x \mid \mathbf{z})}}$$



$\mathbf{Z} = \{Z_1\}$

$\mathbf{Z} = \{Z_1, Z_3\}$

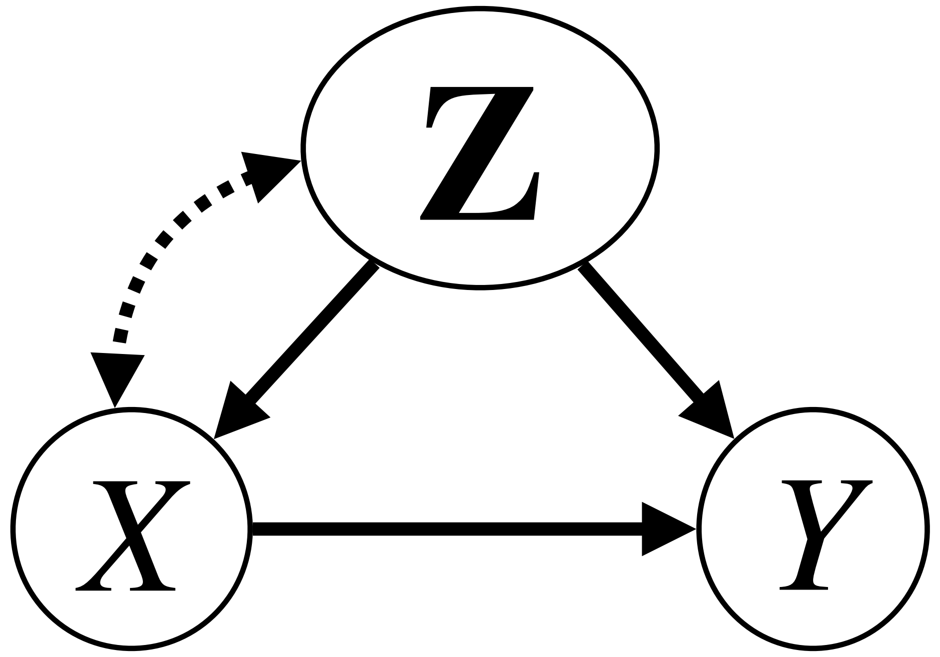
Only if $\mathbf{Z}$ is admissible for adjustment, Propensity Score can be used to estimate $P(y \mid do(x))$.

For X is binary/categorial: logistic/multinomial regression or ML-based classification
For X continuous: ML-based regression techniques.

The interventional joint distribution can be easily derived by reweighing the observational joint distribution with the inverse of the propensity score!

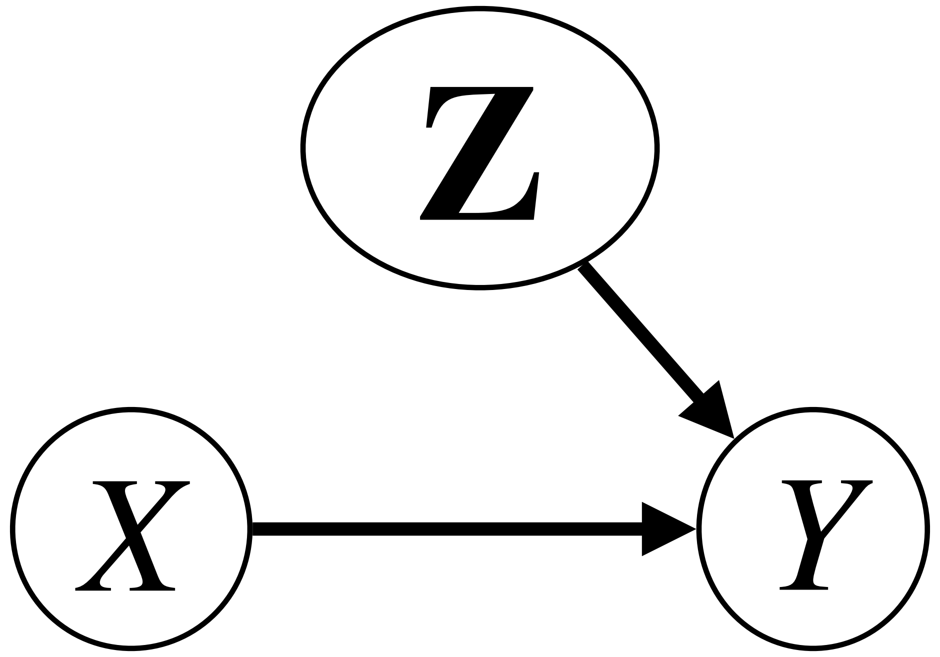
Inverse Probability Weighting (IPW)

After reweighing the observational samples, we obtain *pseudo* interventional samples:



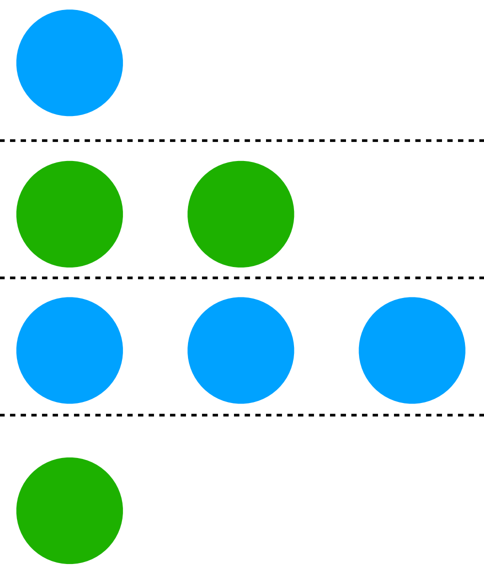
$$P(X|\mathbf{Z}) \neq P(X)$$

Reweighing samples with $\frac{1}{P(X|\mathbf{Z})}$



$$P(X|\mathbf{Z}) = P(X)$$

Original Sample



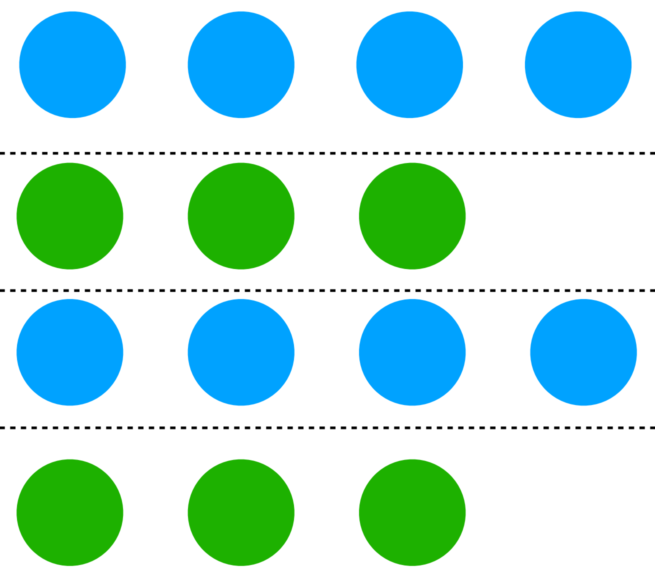
	$P(X|Z)$	$\frac{1}{P(X|Z)}$
X=0 (Control Group)	1/4	4
	2/3	1.5
X = 1 (Treated Group)	3/4	1.33
	1/3	3

Imbalanced

Pseudo interventional Sample

X=0 (Control Group)

X = 1 (Treated Group)



Balanced

Inverse Probability Weighting (IPW)

This gives us the following estimator of $E(Y\,|\,do(x))$, from a sample $\{x_i, y_i, \mathbf{z_i}\}_{i=1}^{N}$:

$$\hat{E}(Y\,|\,do(x)) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i\mathbf{1}_{\{x_i=x\}}}{\hat{P}(x_i\,|\,\mathbf{z}_i)}$$

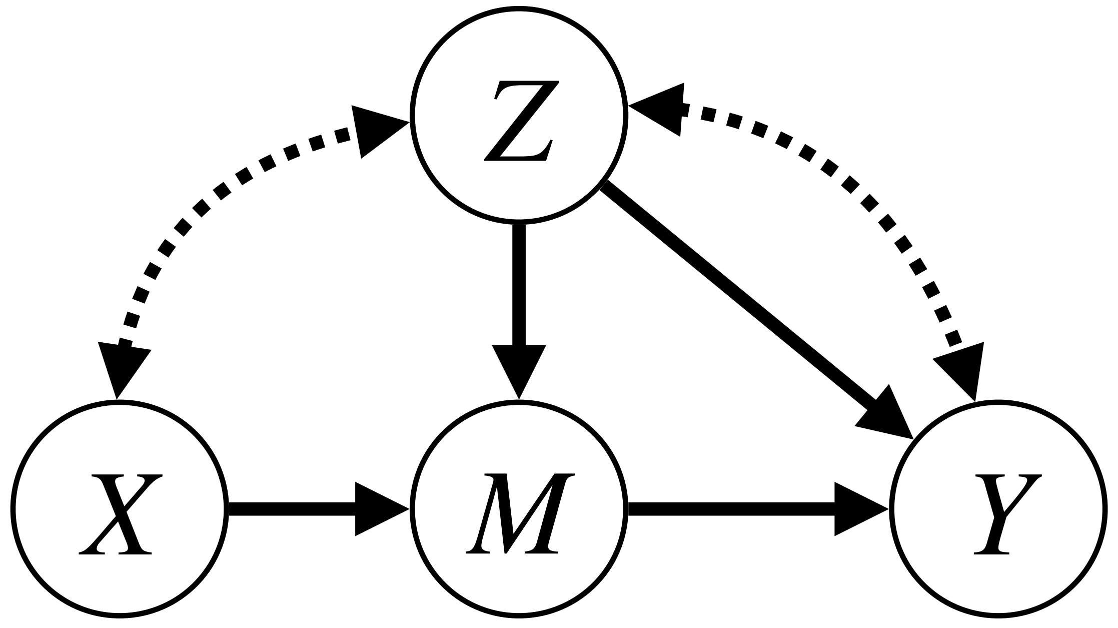
The mean of all values y_i, inversely weighted according to the propensity score.

The Average Treatment Effect (ATE) of a binary treatment can be estimated as:

$$\hat{E}(Y\,|\,do(X=1)) - \hat{E}(Y\,|\,do(X=0))$$

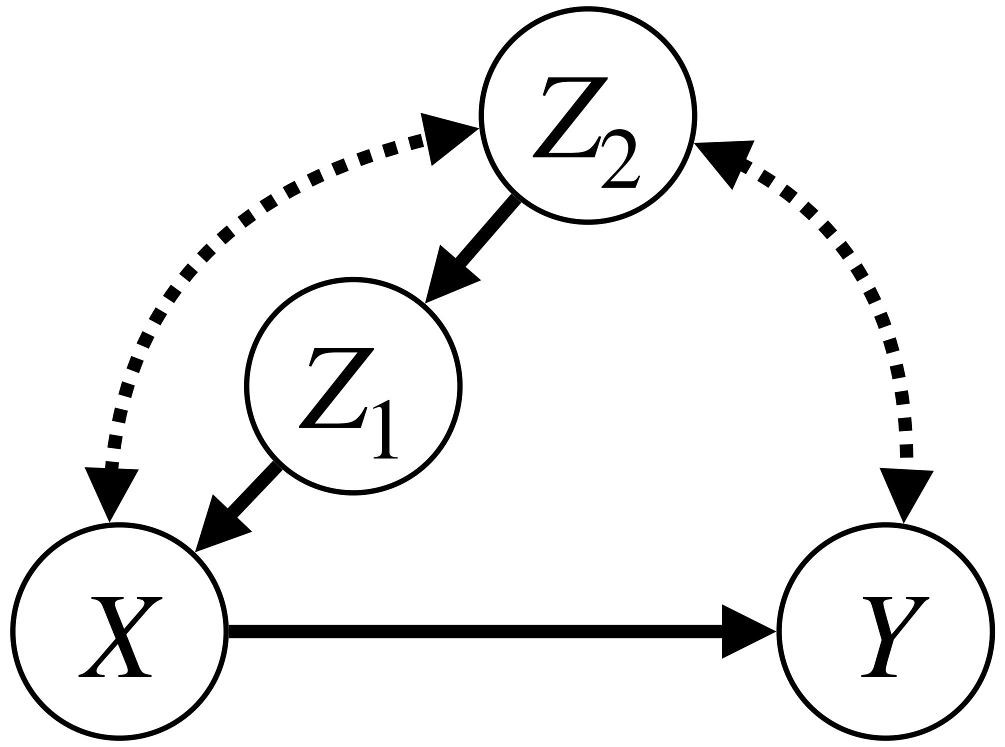
$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i\mathbf{1}_{\{x_i=1\}}}{\hat{P}(X=1\,|\,\mathbf{z}_i)} - \frac{y_i\mathbf{1}_{\{x_i=0\}}}{\hat{P}(X=0\,|\,\mathbf{z}_i)}\right)$$

Many Scenarios Beyond Backdoor …



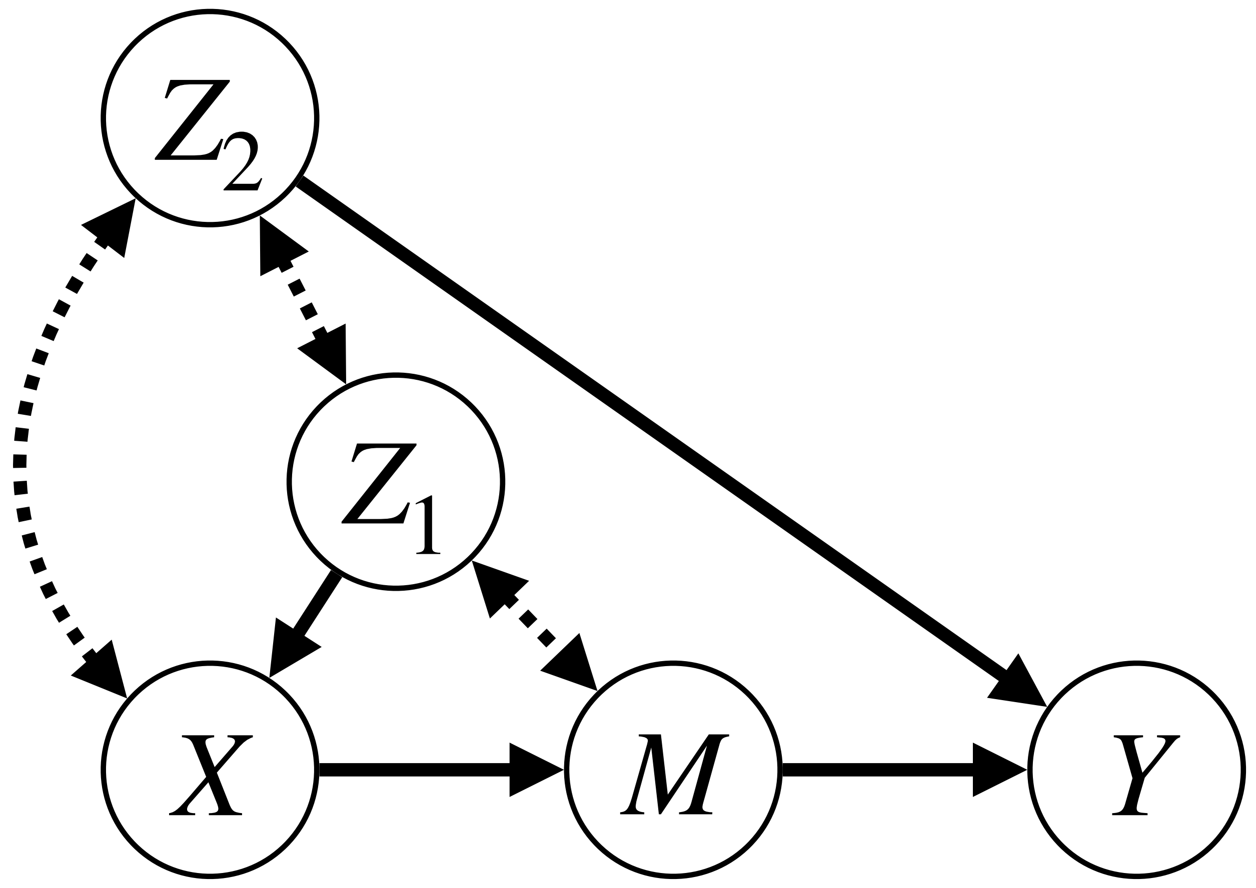
Conditional Front-Door

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



Napkin

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



Unnamed

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

And many others….

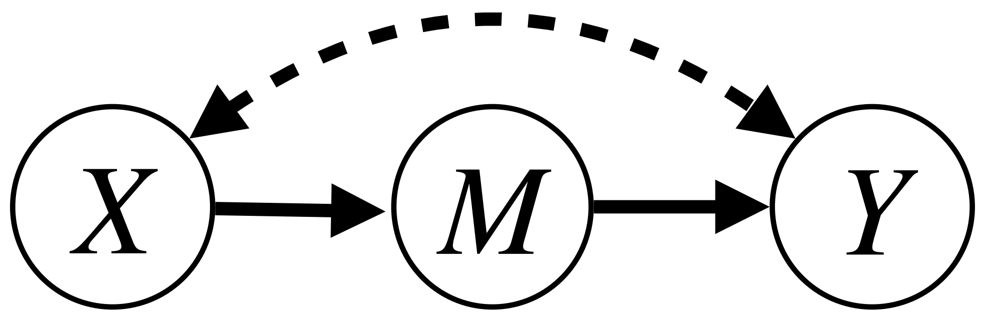
Causal Effect Identification (ID) Algorithm

ID-Algorithm and many recent generalizations.

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Probability Distributions**

$P(x, m, y)$ **Observational Distribution**

Inference Engine

Solution

yes / no

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

Interventional Distribution ← Available Distributions

- 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.

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.

Advances on Effect Identification given a Causal Diagram

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.

Partial Effect Identification:

Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner, Ricardo Silva, Niki Kilbertus; Stochastic Causal Programming for Bounding Treatment Effect. Proceedings of the Second Conference on Causal Learning and Reasoning, PMLR 213:142-176

Coding Exercises

Causality Tutorial:

- **Google Colab Notebook:** (Link)

  Please, open and make a local copy: File → Save a copy in Drive

- Slides and Google Colab link also available on GitHub → TACsy School 2025

Setup: Download the required packages — it takes a few minutes.

Check Part I: Causal Modeling

Check Part II: Causal Effect Identification from Causal Diagrams

1. Backdoor / Adjustment Criterion -- pcalg R package
2. ID Algorithm -- causaleffect R package

Can we relax the causal constraints encoded in the causal diagram?

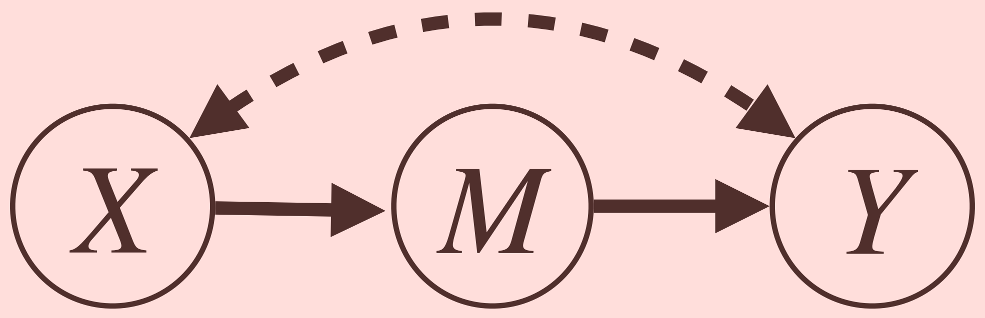
Causal Effect Identification

Requires structural causal knowledge for **every pair** of variables!

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Probability Distributions**

$P(x, m, y)$ **Observational Distribution**

Inference Engine

Solution

yes / **no**

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

Interventional Distribution ← Available Distributions 👍

- 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.

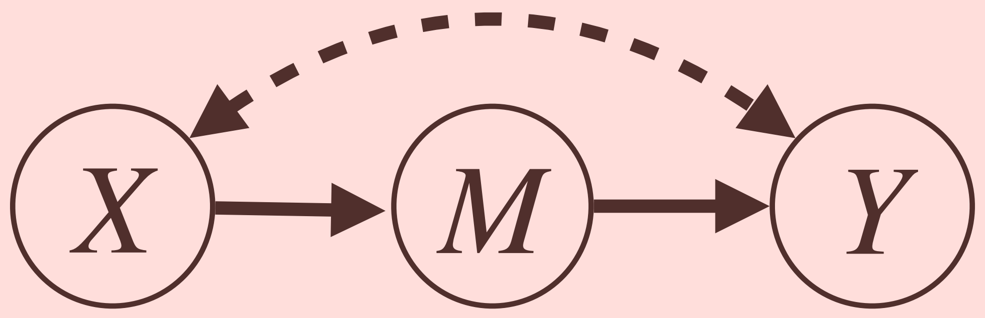
Causal Effect Identification

Can we relax such a requirement?

1 **Query**

$P(y \mid do(x))$

2 **Causal Contraints**



3 **Probability Distributions**

$P(x, m, y)$ **Observational Distribution**

Inference Engine

Solution

yes / no

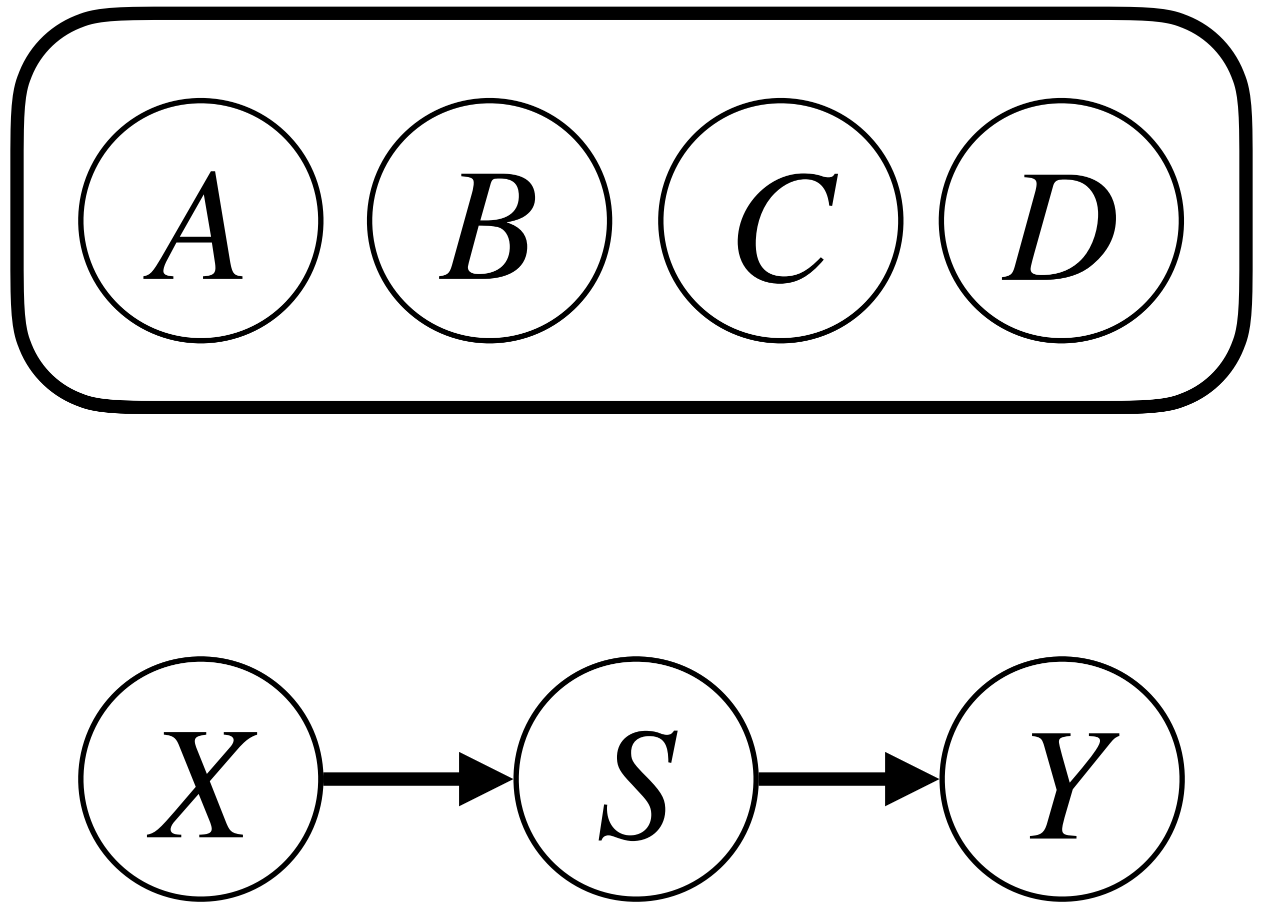
$$P(y \mid do(x)) = \sum_m P(m \mid x) \sum_{x'} P(y \mid m, x')P(x')$$

Interventional Distribution ← Available Distributions

- 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.

Partially Understood Systems

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

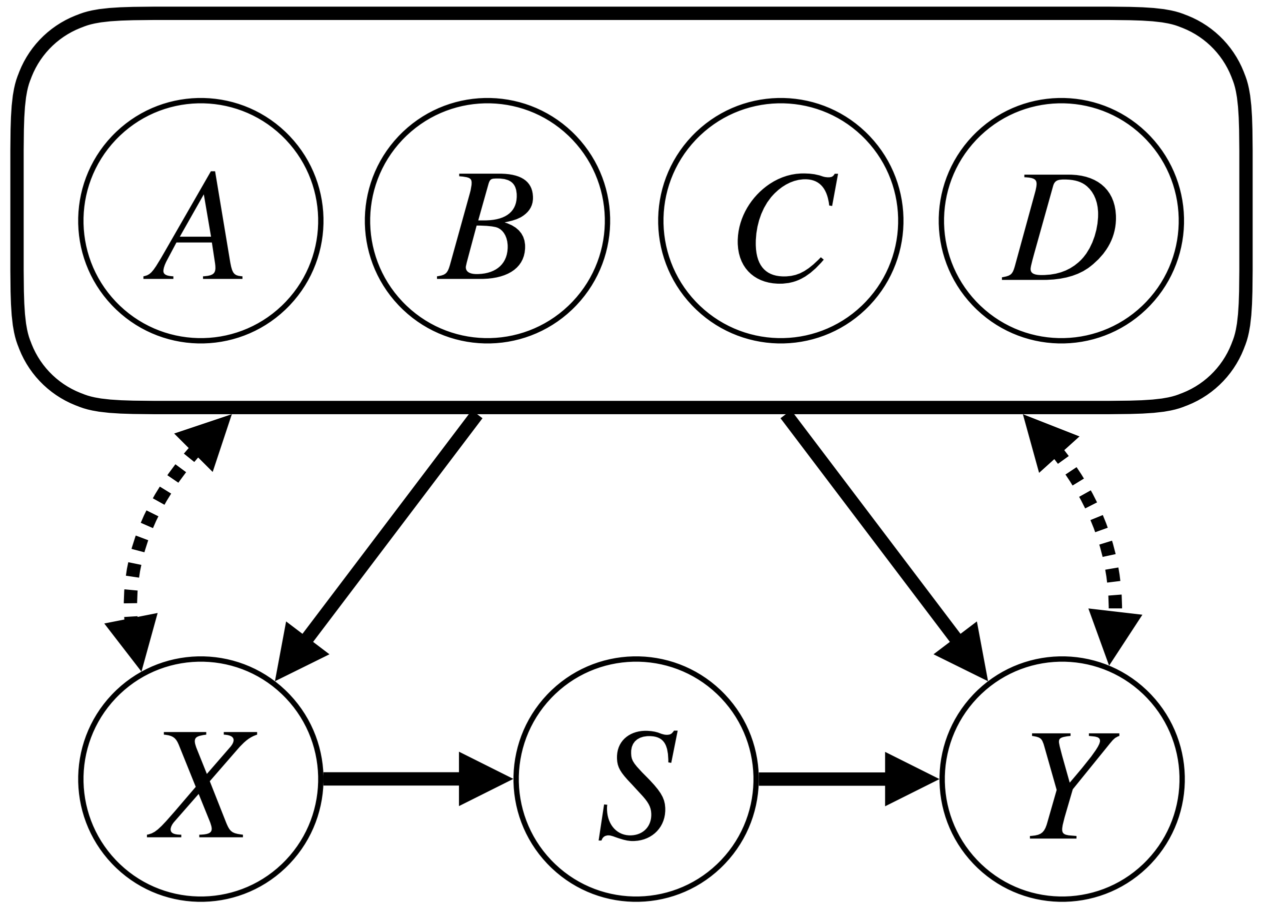


A causal diagram cannot be specified given the existing knowledge!

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

Cluster DAGs (C-DAGs)

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



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

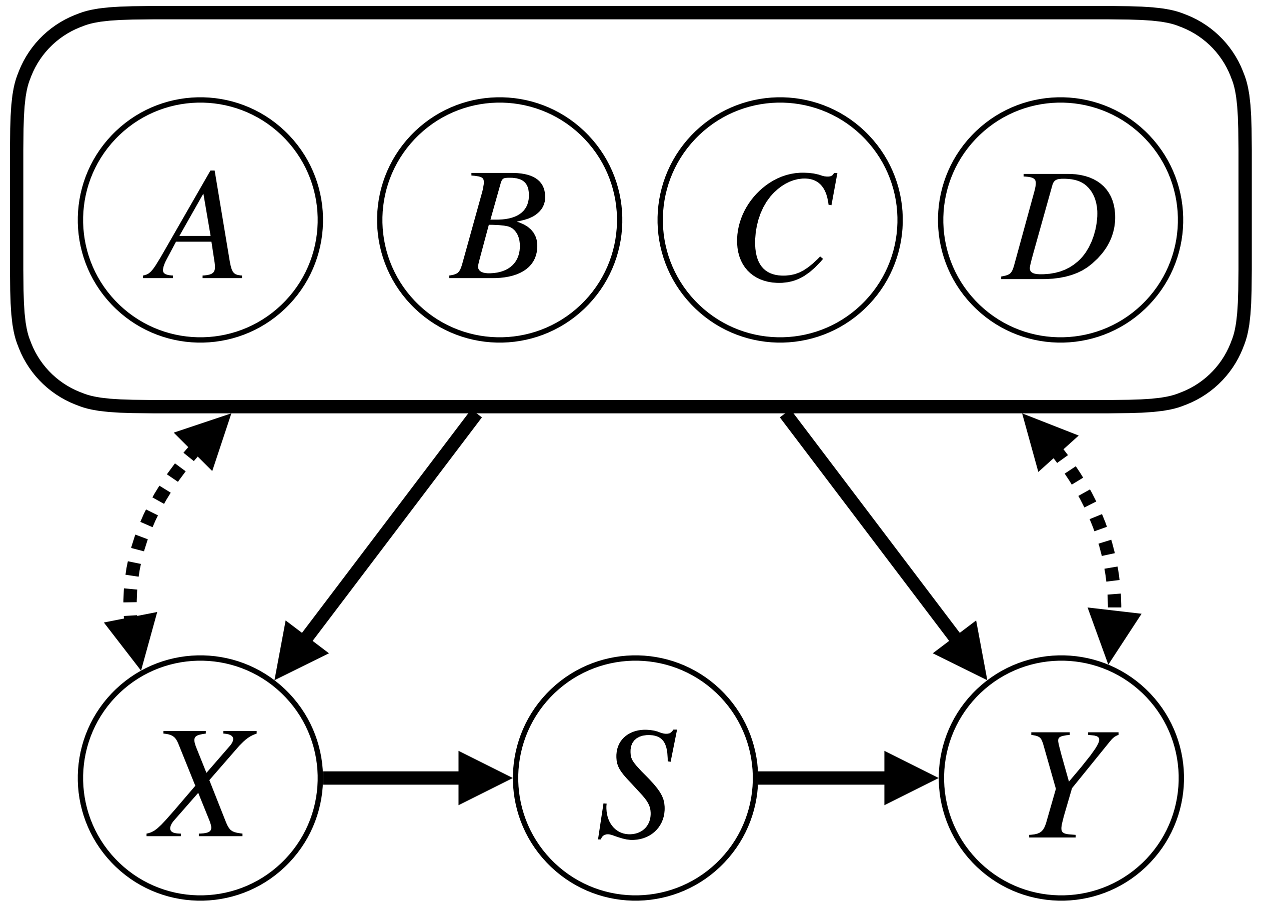
A *cluster DAG* $G_{\mathbf{C}}$ over a given partition $\mathbf{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_k\}$ of $\mathbf{V}$ is compatible with a causal diagram G over $\mathbf{V}$ if **for every $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$**:

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

and $G_{\mathbf{C}}$ contains no cycles.

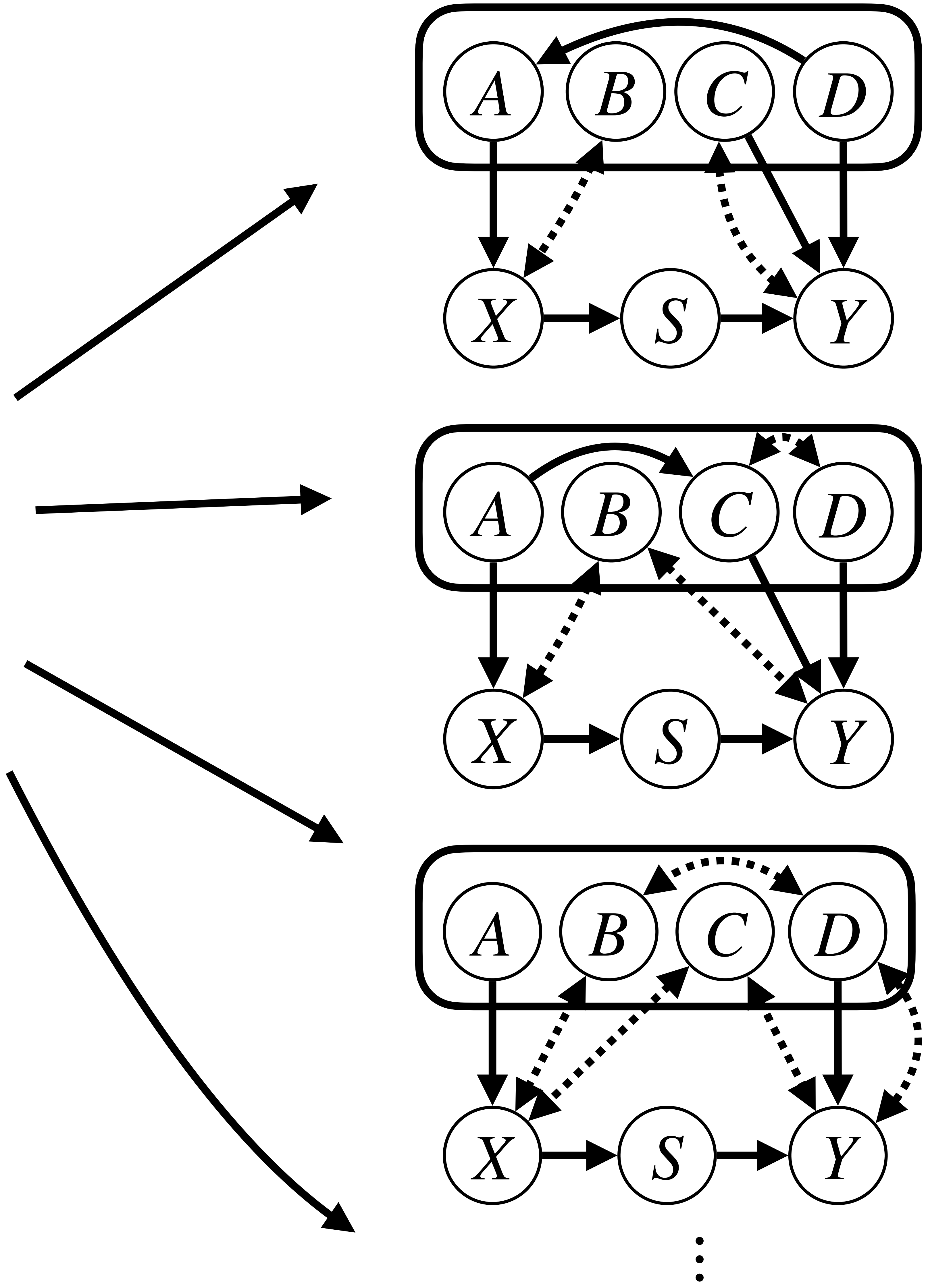
Partially Understood Systems

Many causal diagrams are compatible with the current knowledge!

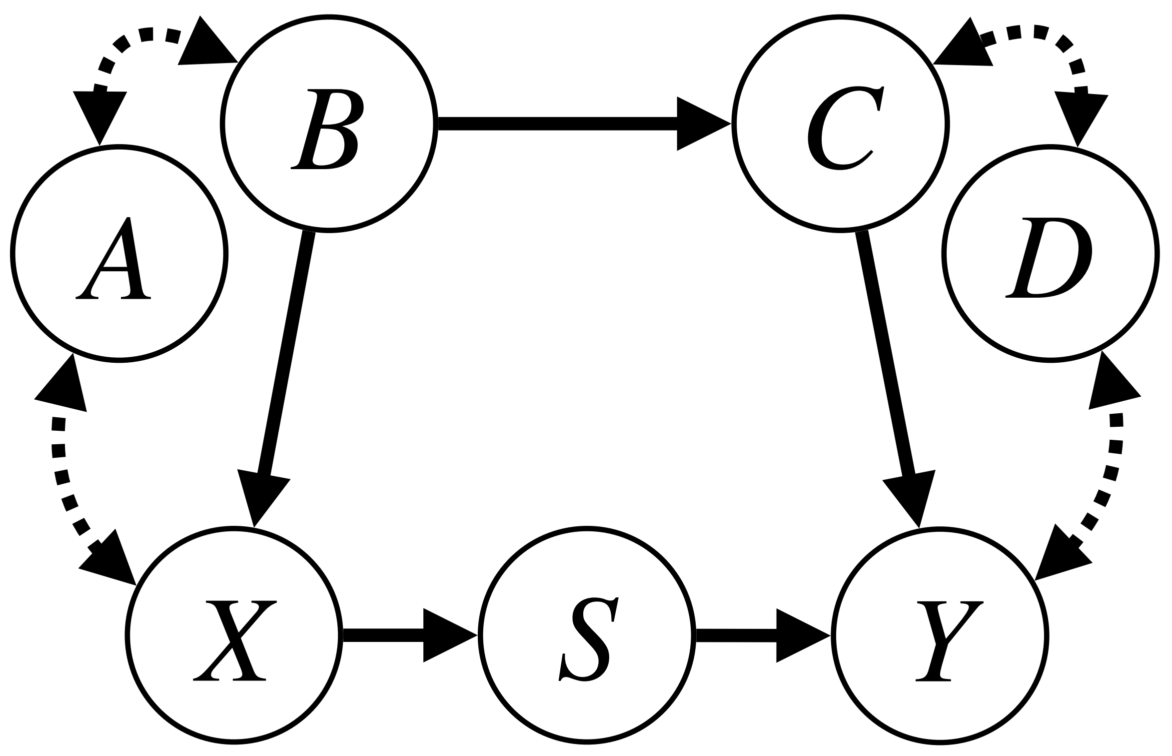


Can be seen as an *equivalence class* of causal diagrams, where any relationships are allowed among the variables within each cluster.

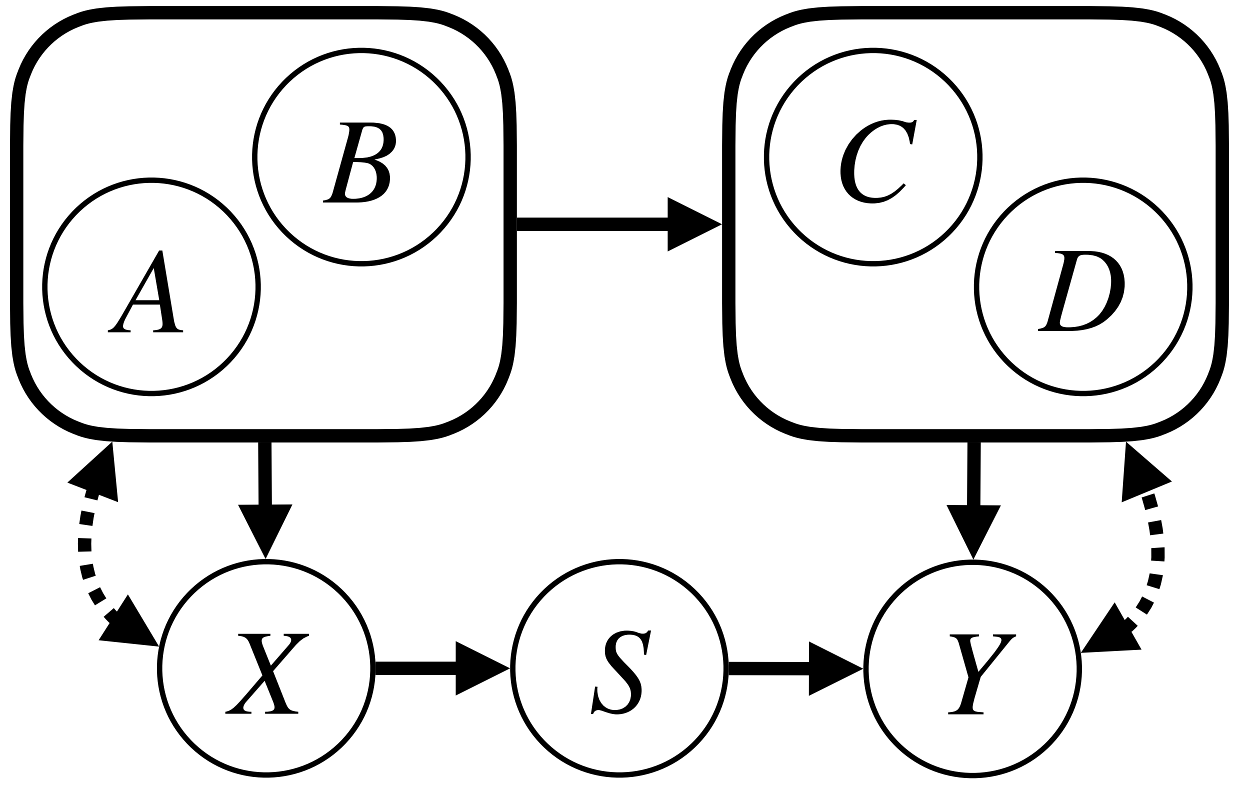
Can we infer causal effects without deciding on any one particular causal diagram?



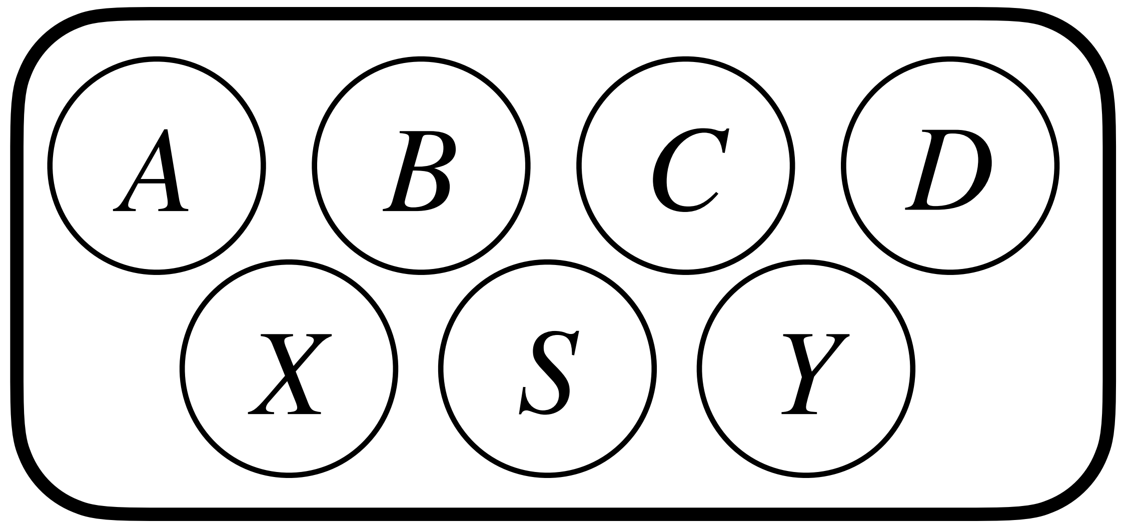
C-DAG: Flexible Encoder of Model Assumptions



N clusters of size one
(full knowledge - DAG)

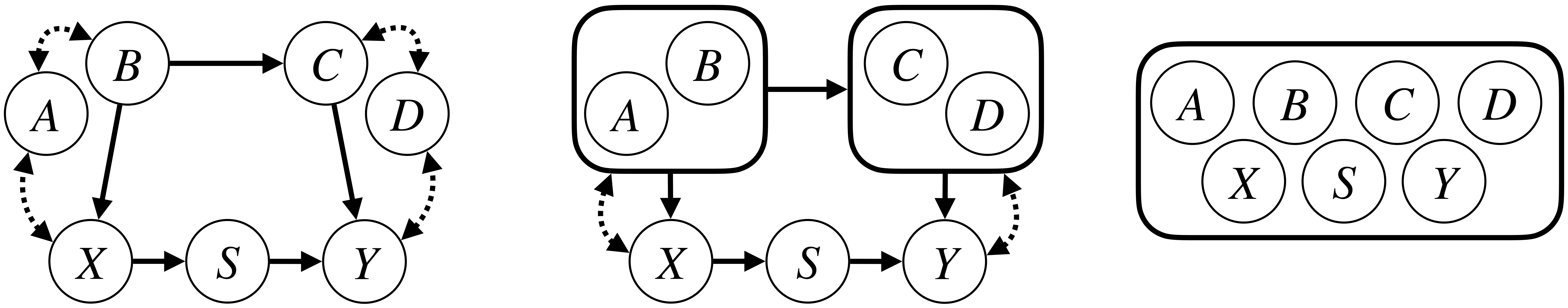


…
(partial knowledge - C-DAG)



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.

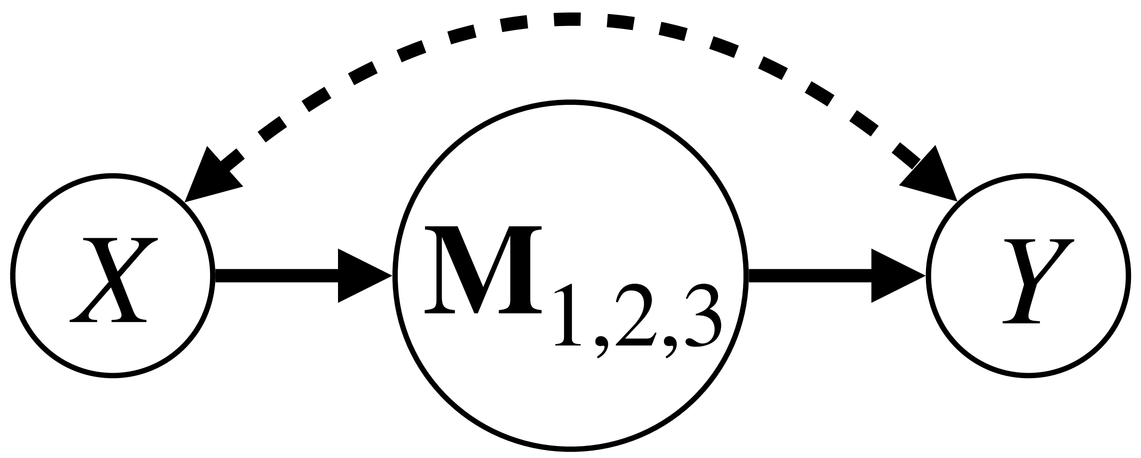
Identification of Causal Effects from C-DAGs

Extension of the ID and Ctf-ID algorithms to C-DAGs.

1 **Query**

$P(y \mid do(x))$

2 **C-DAG**



3 **Data**

$P(x, m_1, m_2, m_3, y)$

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_{\mathbf{m}_{123}} P(\mathbf{m}_{123} \mid x) \sum_{x'} P(y \mid \mathbf{m}_{123}, x')P(x')$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

Anand, T. V.*, **Ribeiro A. H.*,** Tian, J., & Bareinboim, E. (2023). Causal Effect Identification in Cluster DAGs. In Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence.

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

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

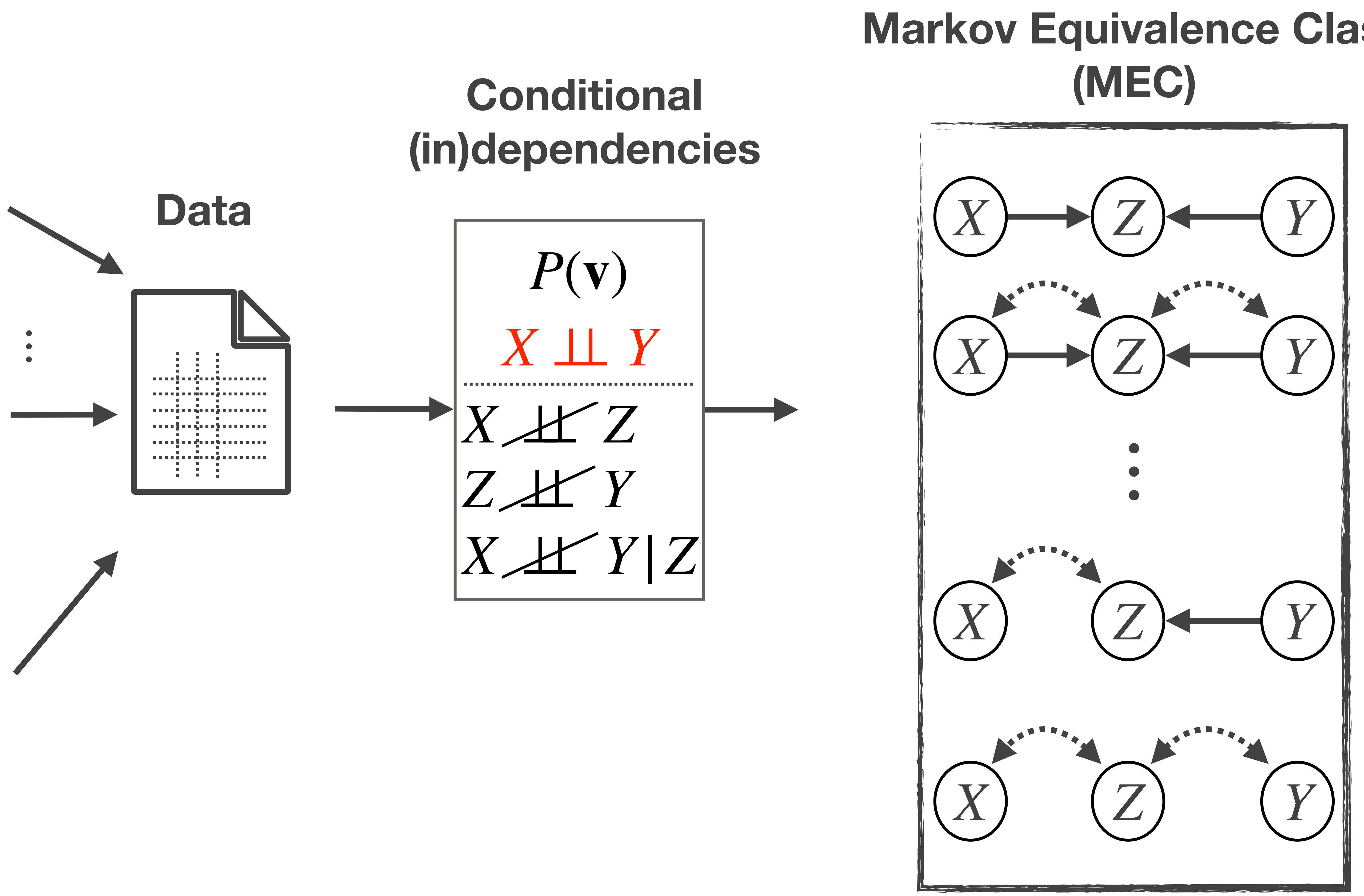
Causal Discovery: Learning Structural Invariances

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

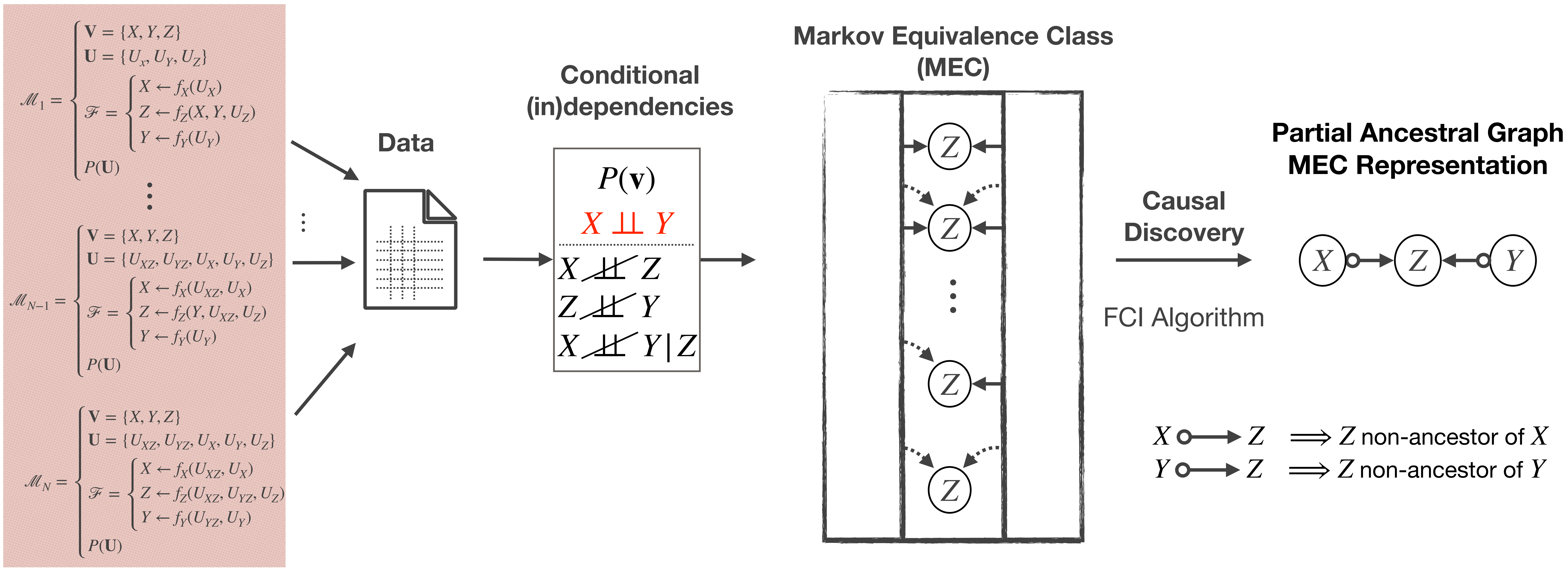
$$\vdots$$

$$\mathcal{M}_{N-1} = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(Y, U_{XZ}, U_Z) \\ Y \leftarrow f_Y(U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$$\mathcal{M}_N = \begin{cases} \mathbf{V} = \{X, Y, Z\} \\ \mathbf{U} = \{U_{XZ}, U_{YZ}, U_X, U_Y, U_Z\} \\ \mathcal{F} = \begin{cases} X \leftarrow f_X(U_{XZ}, U_X) \\ Z \leftarrow f_Z(U_{XZ}, U_{YZ}, U_Z) \\ Y \leftarrow f_Y(U_{YZ}, U_Y) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



Causal Discovery: Learning Structural Invariances

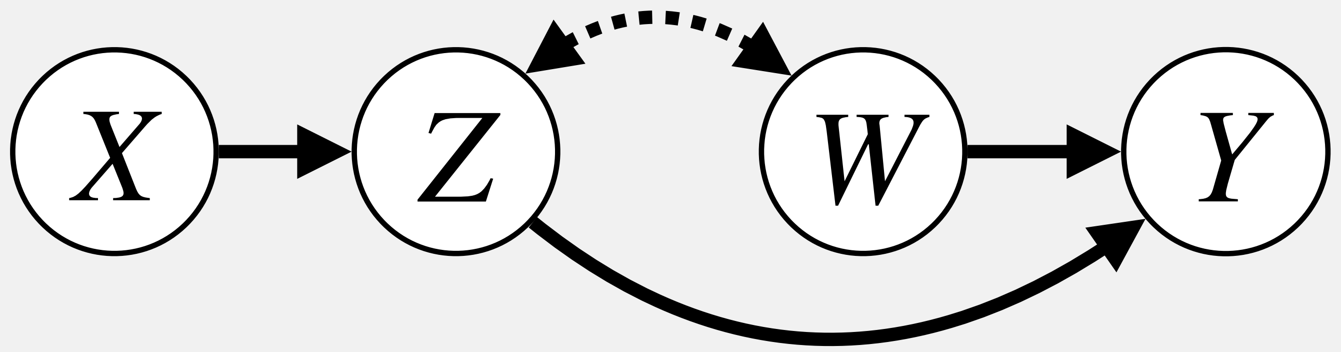


Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. Link

FCI Algorithm - Pipeline

Unknown Reality

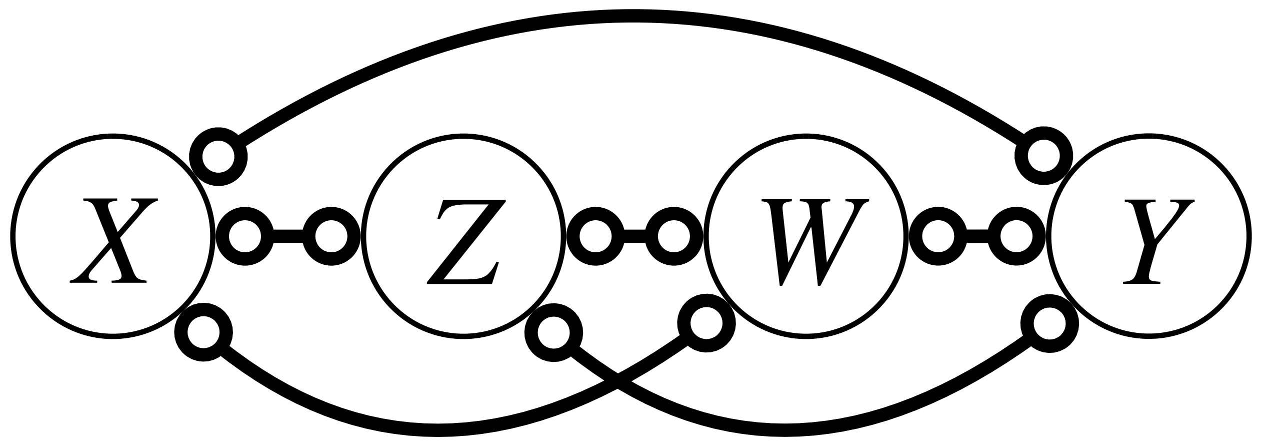
True causal diagram



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

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

Implied by the ADMG using d-separation



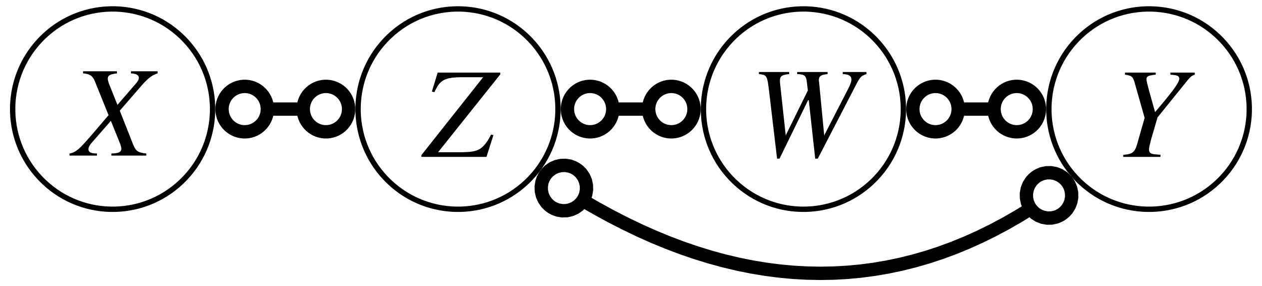
Complete Graph

Conditional Independence Tests

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

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

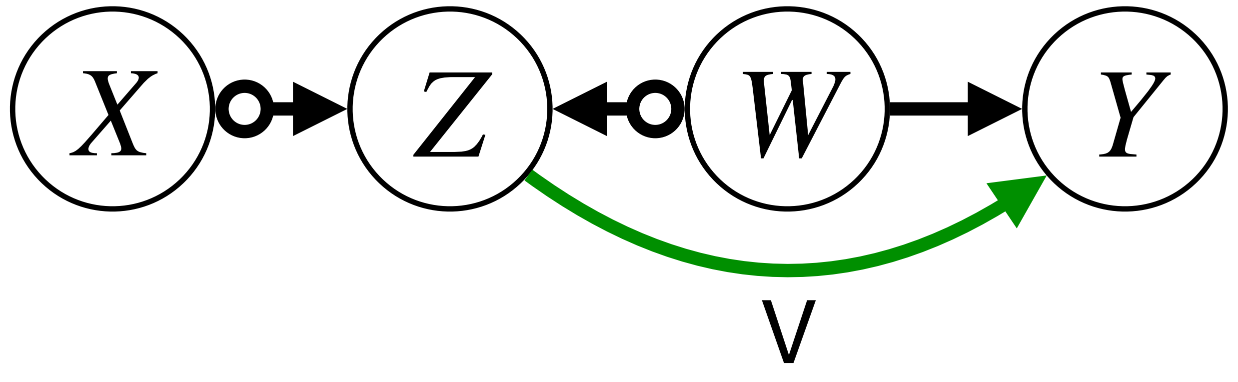
By **faithfulness**, are correctly observed in the data



Skeleton

FCI Rules

$(R1) - (R10)$



Partial Ancestral Graph (PAG)

Implied by the PAG using m-separation

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

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

$A \circ\!\!\rightarrow B \implies$ B non-ancestor of A

$A \rightarrow B \implies$ A ancestror of B

$A \leftrightarrow B \implies$ spurious association

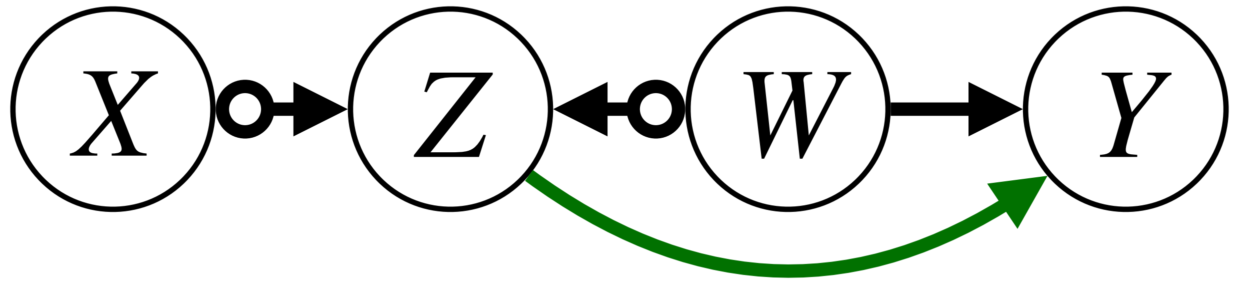
$A - B \implies$ selection bias

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



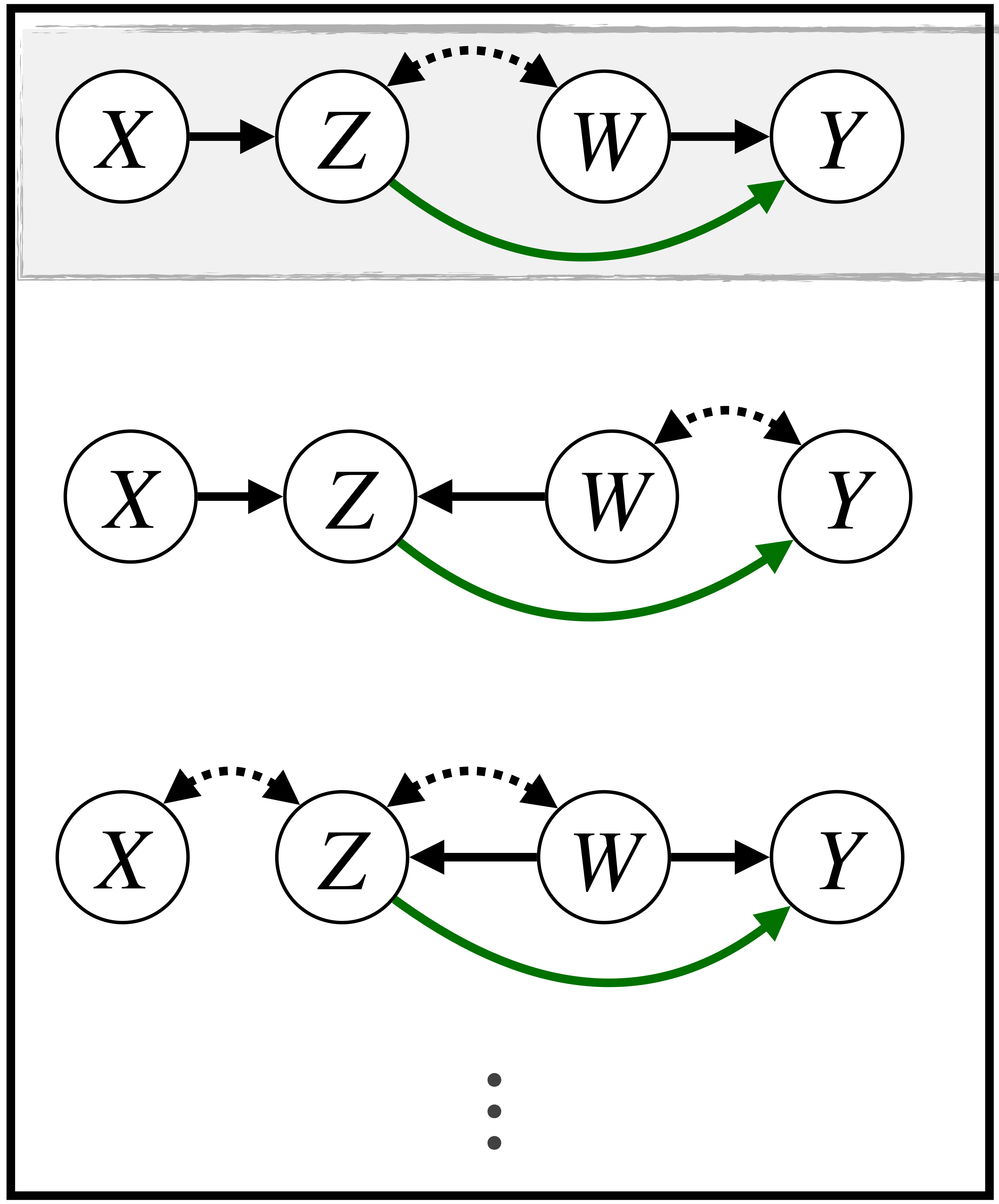
V

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.



True (unknown) causal diagram

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

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

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

Available Implementations of the FCI

R Packages:

- pcalg R package:
  - https://cran.r-project.org/web/packages/pcalg/
  - https://github.com/cran/pcalg/
- RPy-Tetrad (Wrapper in R): https://github.com/cmu-phil/py-tetrad/tree/main/pytetrad/R

Python Packages:

- Do-discover in PyWhy: https://github.com/py-why/dodiscover
- Causal-Learn: https://causal-learn.readthedocs.io/en/latest/index.html
- Py-Tetrad (Wrapper in Python): https://github.com/bd2kccd/py-causal

Other Causal Discovery Algorithms

Hybrid approaches accounting for latent confounding:

- **M3HC:** Max-Min Hill Climbing, by Tsirlis et al., 2018 — extends GSMAG by introducing a constraint-based first phase that greatly reduces the space of structures to investigate.
- **GFCI:** Greedy FCI, by Ogarrio et al., 2016 — combines FGES and FCI. The skeleton and orientation phases are firstly performed using FGES and then refined by using the FCI.

Parametric approaches under causal sufficiency — Identifiable Structure

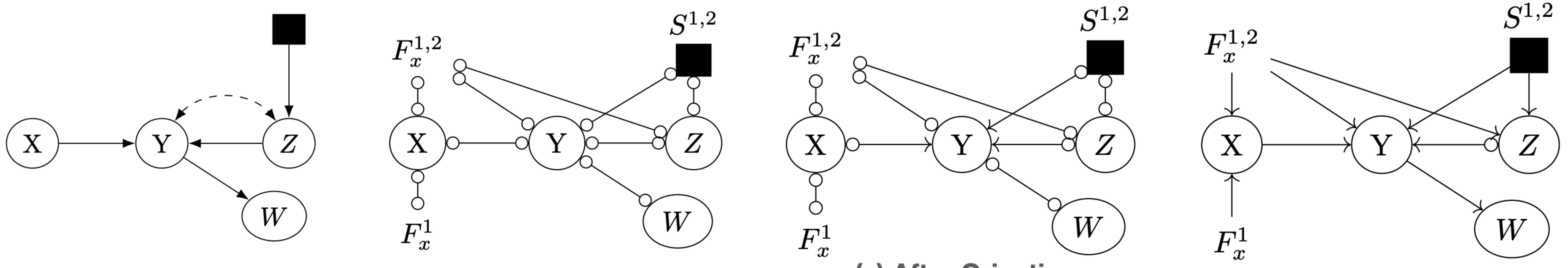
- **LiNGAM:** Linear, non-Gaussian, and Acyclic Model, by Shimizu et al., 2006 — leverage distributional asymmetries with linear causal mechanisms are non-Gaussian error terms.
- **ANM:** Non-linear additive noise model (Hoyer et al., 2009; Zhang and Hyvärinen, 2009a) — leverage distributional asymmetries with non-linear mechanisms and additive noise.

Developments in Causal Discovery with Unobserved Confounding

Going *Beyond* the Markov Equivalence Class:

1. Causal Discovery with Interventional Data

   - **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.



(a) G_S (b) Skeleton (c) After Orienting Unshielded Colliders (d) Final S-PAG

Developments in Causal Discovery with Unobserved Confounding

Going *Beyond* the Markov Equivalence Class:

2. Causal Discovery with Background Knowledge

    - **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.

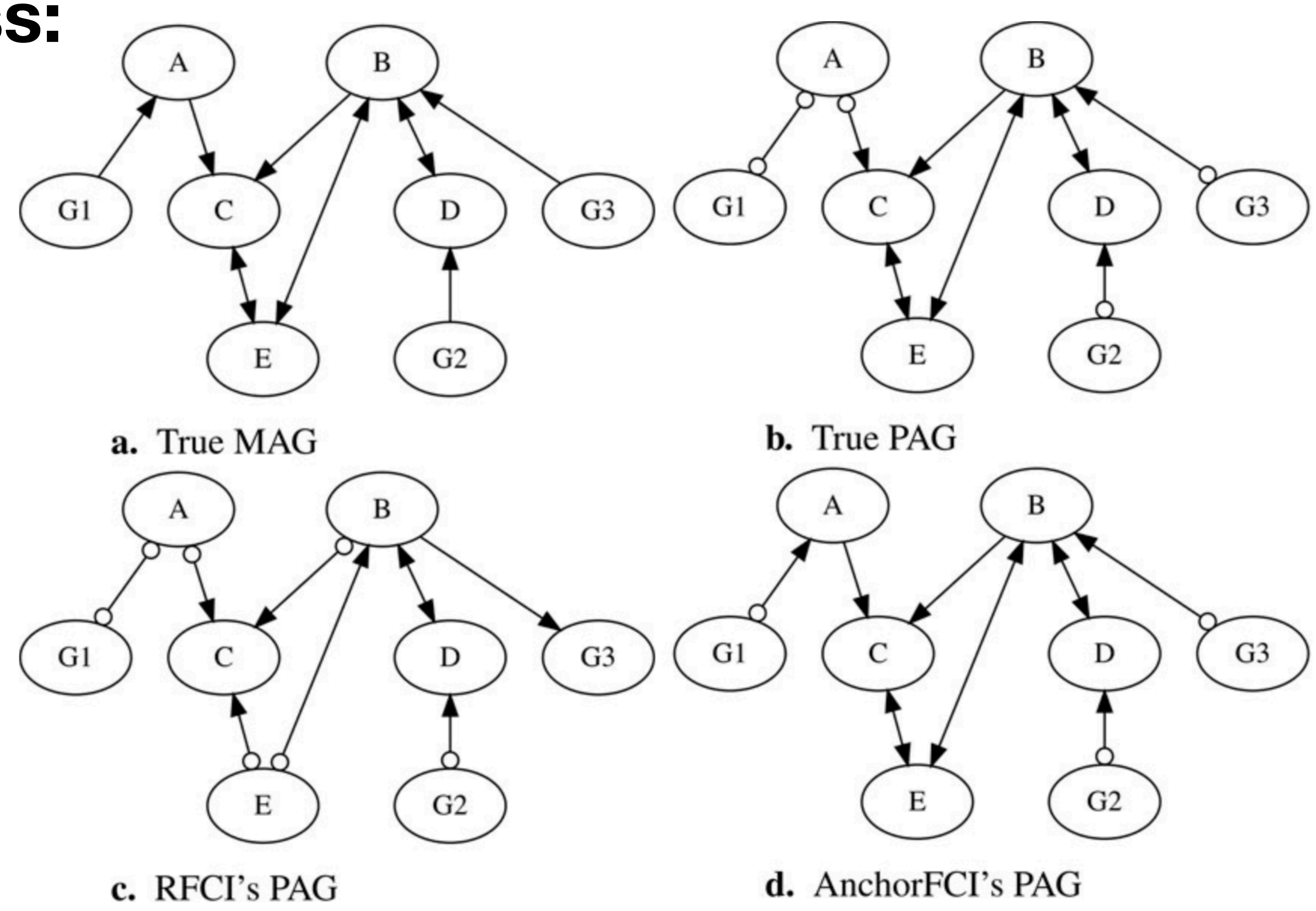
    - **Bryan Andrews, Peter Spirtes, Gregory F. Cooper** (2020). On the Completeness of Causal Discovery in the Presence of Latent Confounding with Tiered Background Knowledge. Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, PMLR 108:4002-4011, 2020.

Developments in Causal Discovery with Unobserved Confounding

Going *Beyond* the Markov Equivalence Class:

Leveraging Non-Ancestral Knowledge

- AnchorFCI: strategically selects and integrates *reliable anchors,* i.e. variables not caused by the variables of interest, forming unambiguous triplets.
- Integrates known non-ancestral relations (e.g., In Genetics, Genotypes $\prec$ Phenotypes).
- AnchorFCI R package — *GitHub repository:* @adele/anchorFCI



Ribeiro, A. H. Crnkovic, M., Pereira, J. L., Fisberg, R. M., Sarti, F. M., Rogero, M. M., Heider, D., and Cerqueira, A. (2024). AnchorFCI: Harnessing Genetic Anchors for Enhanced Causal Discovery of Cardiometabolic Disease Pathways. Frontiers in Genetics 15:1436947. DOI: 10.3389/fgene.2024.1436947 (Link)

Developments in Causal Discovery with Unobserved Confounding

4. Causal Discovery in Linear Models

   - **Tashiro, T., Shimizu, S., Hyvärinen, A., & Washio, T.** (2014). ParceLiNGAM: A causal ordering method robust against latent confounders. *Neural computation*, *26*(1), 57-83.
   - **Wang, Y. S., & Drton, M.** (2023). Causal discovery with unobserved confounding and non-Gaussian data. *Journal of Machine Learning Research*, *24*(271), 1-61.

Relax the causal sufficiency assumption of LinGAN by Shimizu et al., 2006:
order / ancestral identifiability under linear systems with non-gaussian error terms

5. Causal Discovery for Additive Noise Models

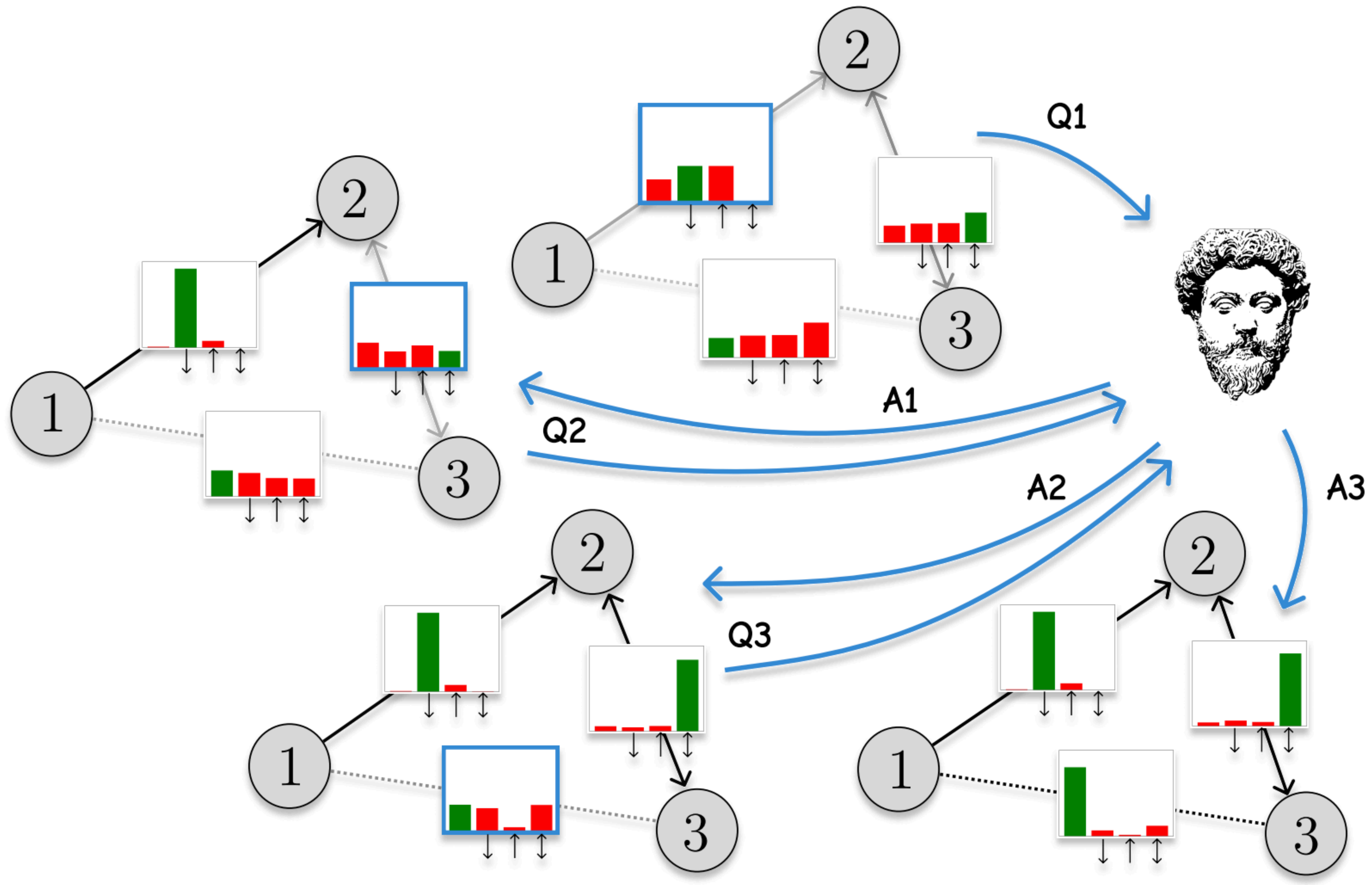
   - **Van Diepen, M. M., Bucur, I. G., Heskes, T., & Claassen, T.** (2023). Beyond the Markov Equivalence Class: Extending Causal Discovery under Latent Confounding. In *Conference on Causal Learning and Reasoning* (pp. 707-725). PMLR.

FCI-CDC: causal direction criterion (CDC) allows pairwise orientation in (weakly) additive noise models with independent causal mechanisms.

Developments in Causal Discovery with Unobserved Confounding

Enhanced Robustness, Uncertainty Modeling, and Knowledge Integration.

- Ancestral Generative Flow Networks (AFGN): **probabilistic causal discovery with an expert-in-the-loop (EITL).**
- Provides **optimal strategy for elicitation** of an expert's feedback regarding the nature of a specific causal relation.
- Significantly outperforms traditional causal discovery algorithms after just a few expert interactions.



Learned AGFN is progressively refined with feedbacks from a (simulated) human expert.

da Silva, T., Silva, E., Góis, A., Heider, D., Kaski, S., Mesquita, D., **Ribeiro, A.H.** (2024). Human-Aided Discovery of Ancestral Graphs. LXAI Workshop at Neural Information Processing Systems (NeurIPS 2024) — (Link)

da Silva, T., Silva, E., Góis, A., Heider, D., Kaski, S., Mesquita, D., **Ribeiro, A.H.** (2024). Human-in-the-Loop Causal Discovery under Latent Confounding using Ancestral GFlowNets. *arXiv:2309.12032* (Link) — Under Review.

Developments in Causal Discovery with Unobserved Confounding

Learning Dynamic Systems:

1. Causal Discovery with Cycles

    - **Bongers, S., Forré, P., Peters, J., & Mooij, J. M.** (2021). Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, *49*(5), 2885-2915.

    - **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.

2. Causal Discovery from Time-Series Data

    - **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.

Causal Identification from Equivalence Classes

Can we identify causal effects from an equivalence class?

Effect Identification from PAGs:

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)

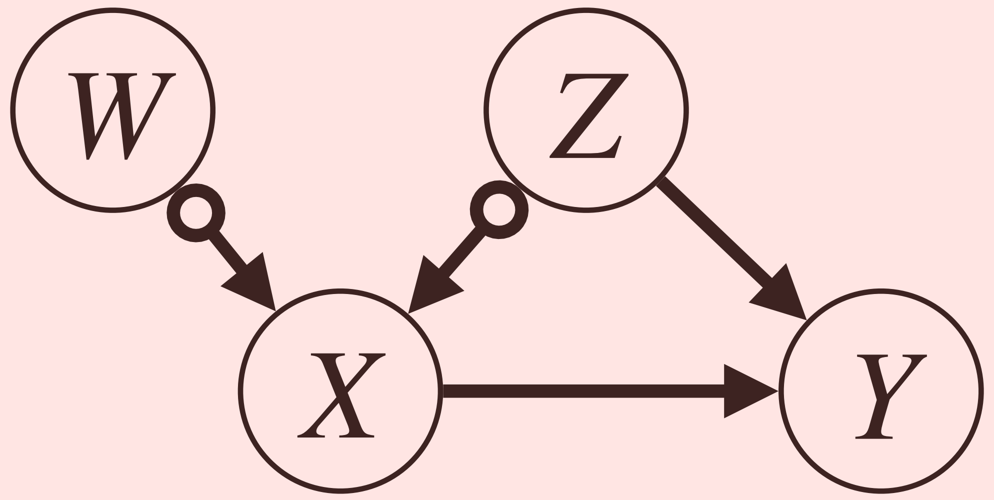
Effect Identification in Markov Equivalence Classes

Can be constructed in a fully data-driven way!

1 **Query**

$P(y \mid do(x))$

2 **PAG**



3 **Data**

$P(x, y, z, w)$ **Observational Distribution**

Inference Engine

Solution

yes / no

$$P(y \mid do(x)) = \sum_{z} P\left(y \mid x, z\right) P\left(z\right)$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution

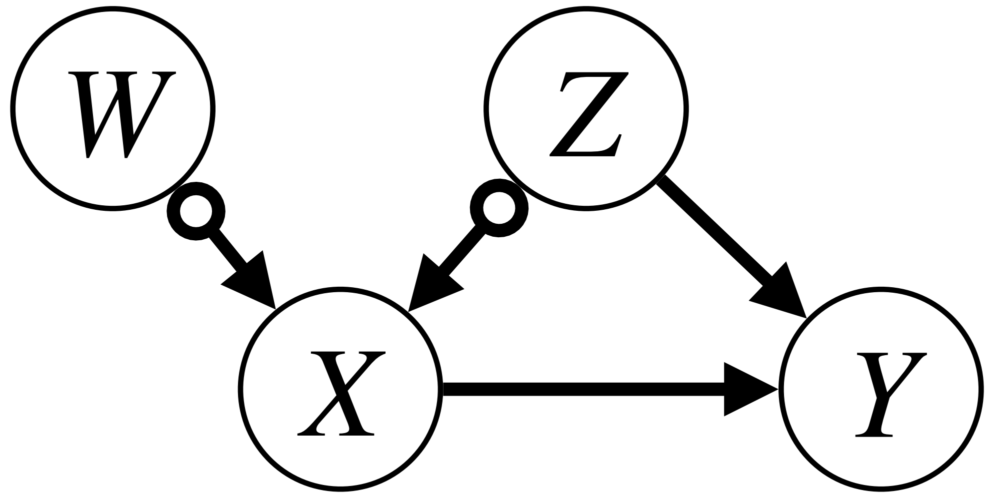
Identification via Adjustment in Markov Equivalence Classes

Identification is possible only when the Generalized Adjustment Criterion applies.

1 **Query**

$P(y \mid do(x))$

2 **PAG**



3 **Data**

$P(x, y, z, w)$ **Observational Distribution**

Adjustment Criterion

Solution

yes / no

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

Inferred (Interventional) Distribution ← Available (Observational) Distribution

👍

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

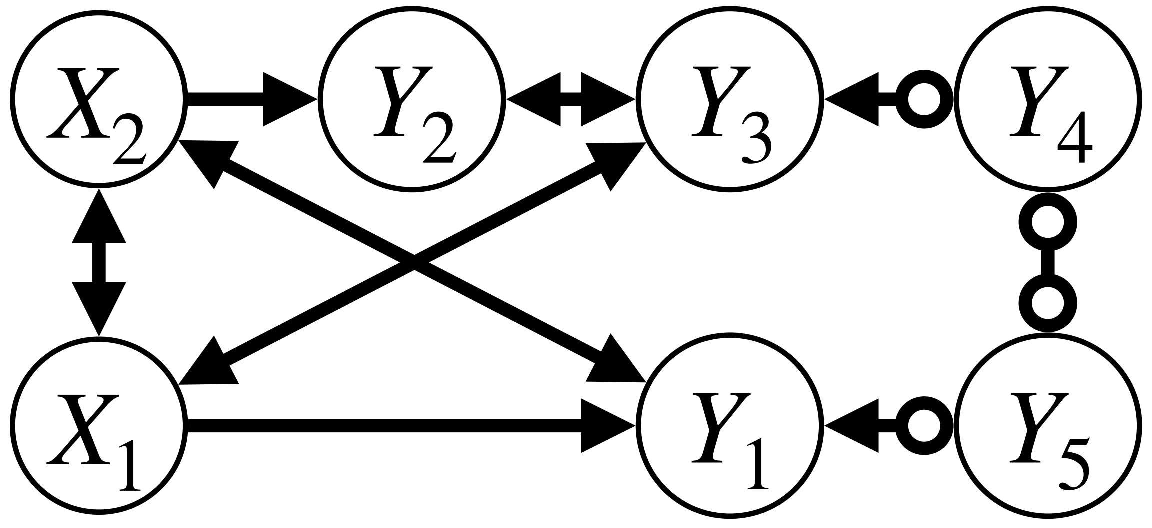
General Identification in Markov Equivalence Classes

Complete algorithms,
available at the *PAGId* R package:
https://github.com/adele/PAGId

1 **Query**

$P(y \mid do(x))$

2 **PAG**



3 **Data**

$P(x, y, z, w)$ **Observational Distribution**

IDP / CIDP

Solution

yes / no

$$P(y \mid do(x)) = \frac{P(y_1, y_4, y_5 \mid x_1) \,.\, P(y_2 y_3, y_4, y_5 \mid x_2)}{P(y_4, y_5)}$$

Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

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).

Uncovering Causal Relationships in a Real-World Application

Application to the DFG FOR2107 dataset

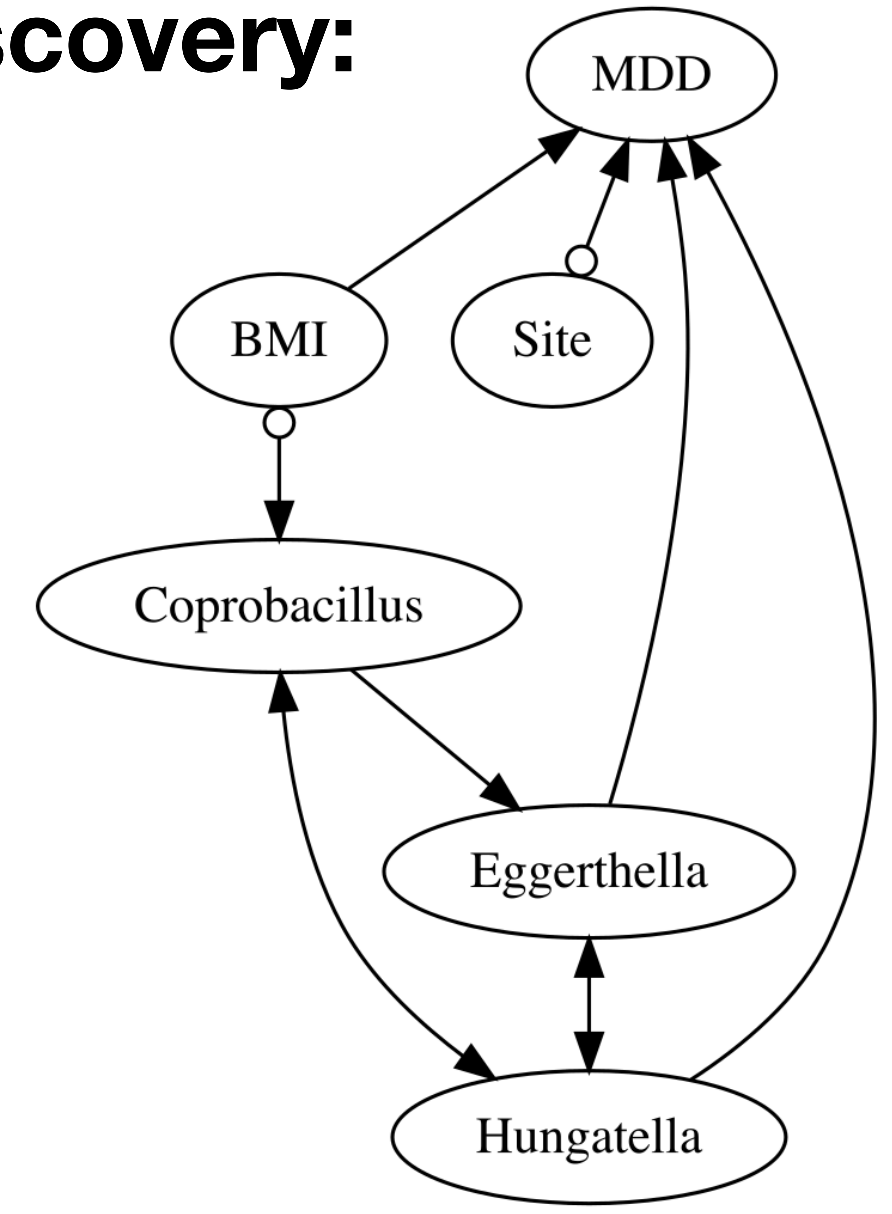
- Collaboration with the Institute for Translational Psychiatry, University of Münster.
- Gut microbiome and clinical data from 1,269 patients.
- Conducted Differential Abundance Analysis to identify gut bacteria genera linked to Major Depressive Disorder (MDD).
- Applied the FCI algorithm for causal discovery and conducted multiple robustness assessments for reliability.
- Used the PAGId algorithm to identify causal effects.
- Identified *Eggerthella and Hungatella* as causal contributors to Major Depressive Disorder (MDD).

Differential Abundance Analysis:

Genus	FDR-corr. p-values	
	LinDA	ZicoSeq
Hungatella	0.0002	0.0071
Eggerthella	0.0063	0.0071
Coprobacillus	0.0070	0.0071
Lachnospiraceae FCS020 group	0.0063	0.011

Causal Discovery:

FCI with a robustness-enhancing strategy.



Fehse L.*, **Ribeiro, A.H.***, Winter, N. R., …, Dominik Heider, Tim Hahn. (2024) From Gut to Mind: Dissecting the Causal Contribution of Gut-Microbiota to Major Depressive Disorder in Humans. MedRxiv Preprint, DOI: 10.1101/2024.12.05.24318549

Gut Microbiota's Causal Role in Major Depressive Disorder

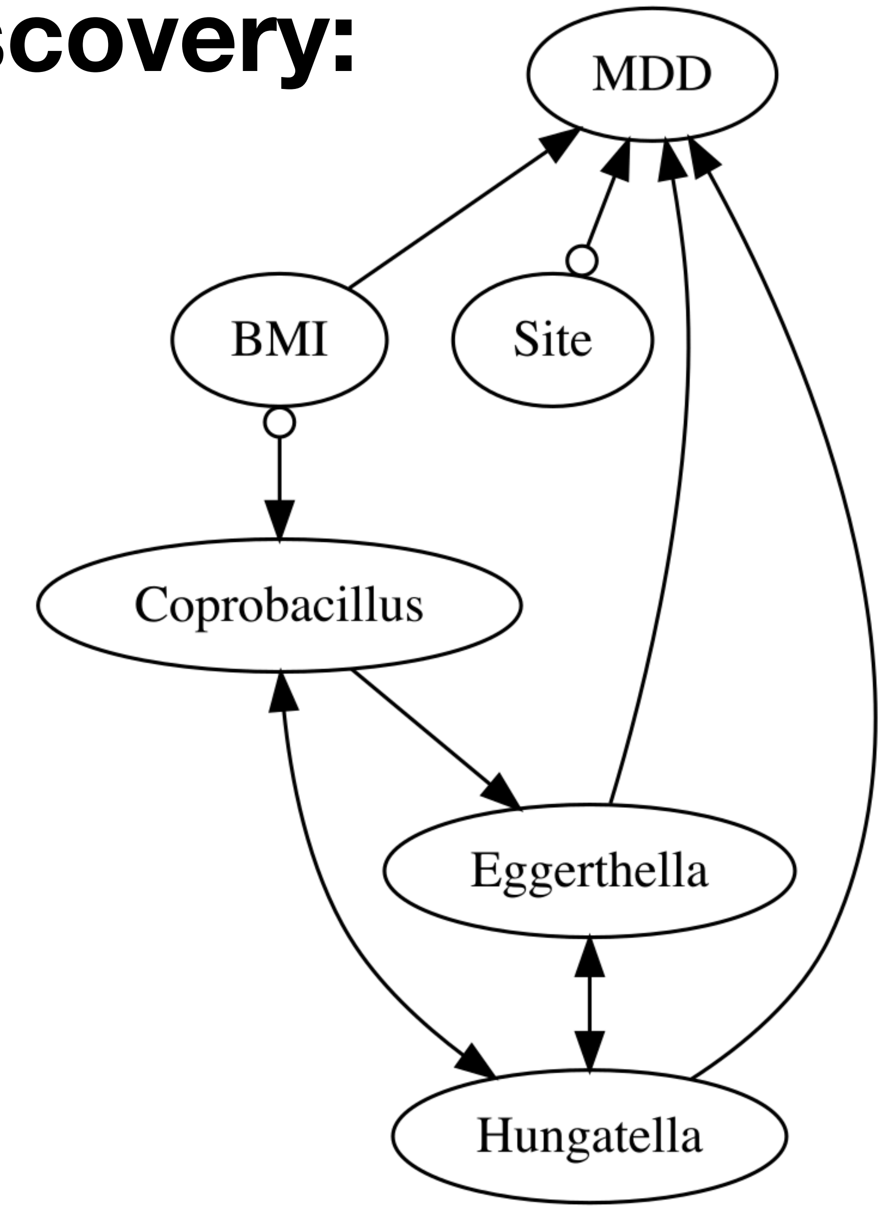
DFG FOR2107 dataset, including microbiome and clinical data from 1,269 patients.

Differential Abundance Analysis:

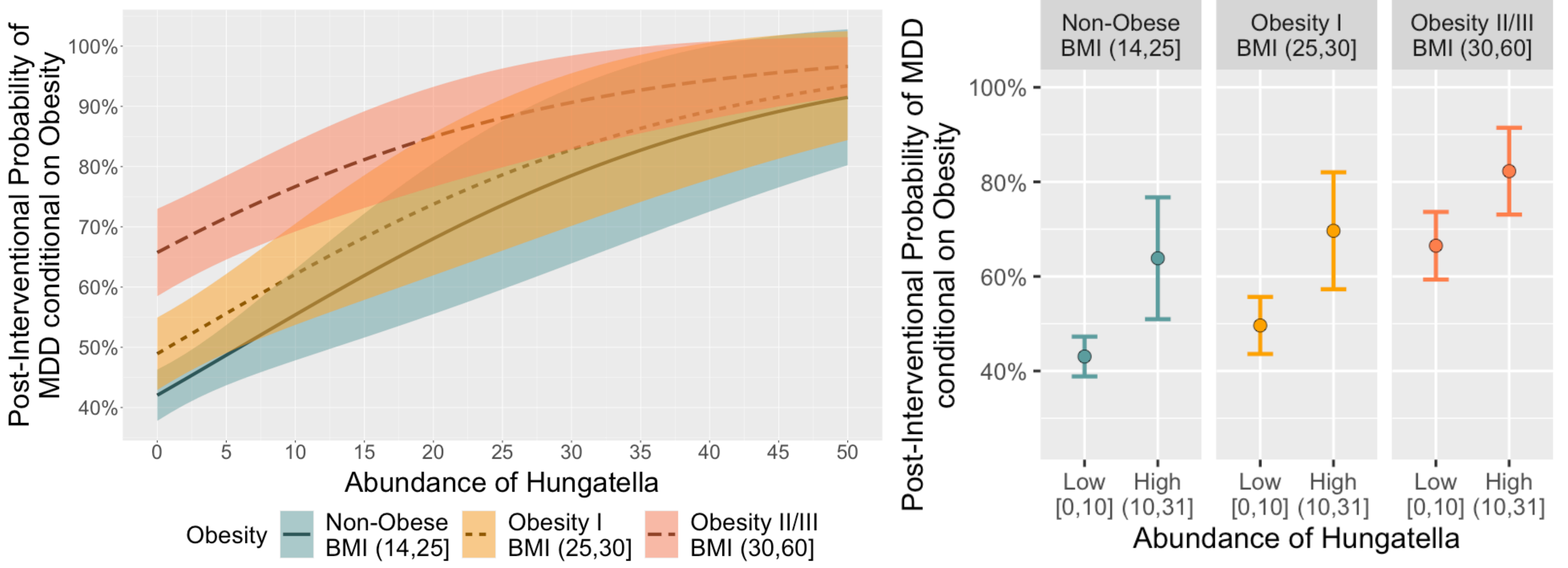
Genus	FDR-corr. p-values	
	LinDA	ZicoSeq
Hungatella	0.0002	0.0071
Eggerthella	0.0063	0.0071
Coprobacillus	0.0070	0.0071
Lachnospiraceae FCS020 group	0.0063	0.011

Causal Discovery:

FCI with a robustness-enhancing strategy.



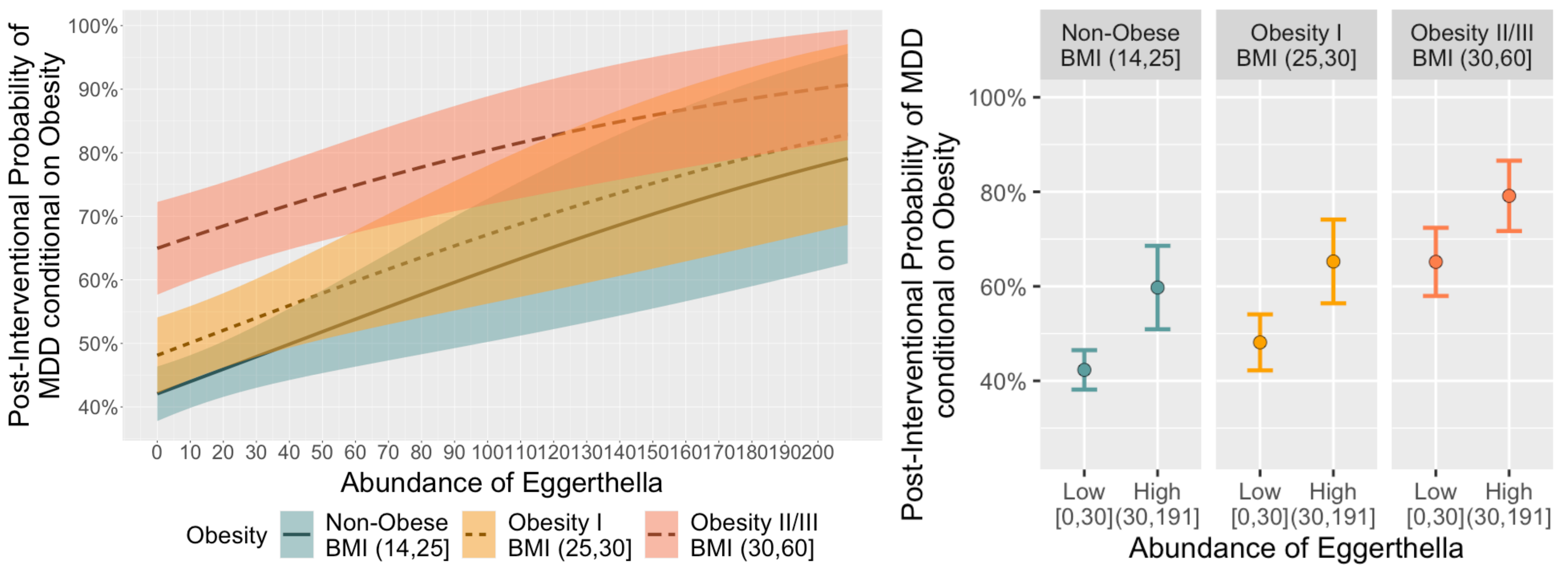
Obesity-specific causal effect of *Hungatella* on MDD



Obesity	Effect	95% C.I.	
No	0.2078	0.07230	0.3433
I	0.2002	0.06261	0.3379
II/II	0.1577	0.04138	0.2741

ACE: 0.1984
95% CI: [0.06607, 0.3307]

Obesity-specific causal effect of *Eggerthella* on MDD



Obesity	Effect	95% C.I.	
No	0.1742	0.07659	0.2718
I	0.1712	0.06460	0.2778
II/II	0.1397	0.03607	0.2433

ACE: 0.1683
95% CI: [0.06812, 0.2684]

Fehse L.*, **Ribeiro, A.H.***, Winter, N. R., …, Dominik Heider, Tim Hahn. (2024) From Gut to Mind: Dissecting the Causal Contribution of Gut-Microbiota to Major Depressive Disorder in Humans. MedRxiv Preprint, DOI: 10.1101/2024.12.05.24318549 * Equal contribution

Coding Exercises

Causality Tutorial:

- **Google Colab Notebook:** (Link)

  Please, open and make a local copy: File → Save a copy in Drive

- Slides and Google Colab link also available on GitHub → TACsy School 2025

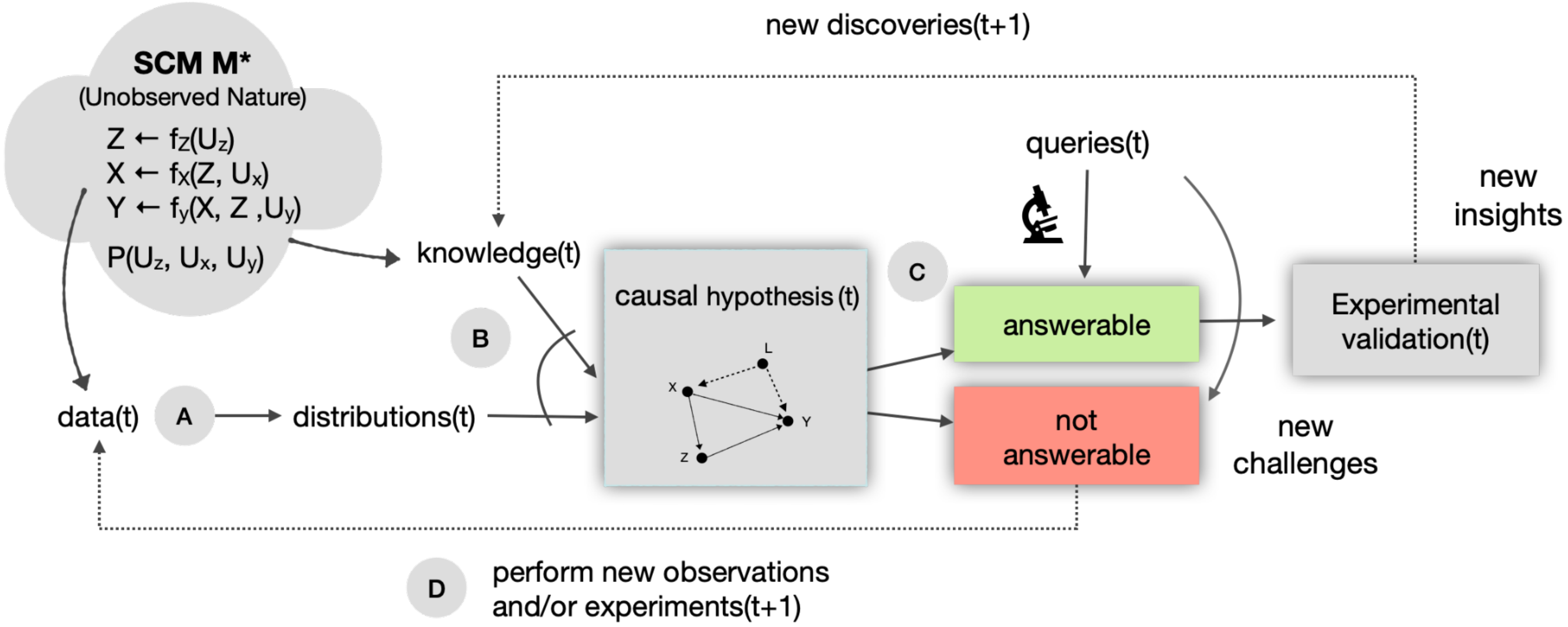
Check Part III: Causal Discovery using FCI

Check Part IV: Causal Effect Identification from the Markov Equivalence Class

1. Generalized Backdoor / Adjustment Criterion -- pcalg R package
2. CIDP Algorithm -- pagID R package

Causal Inference Workflow

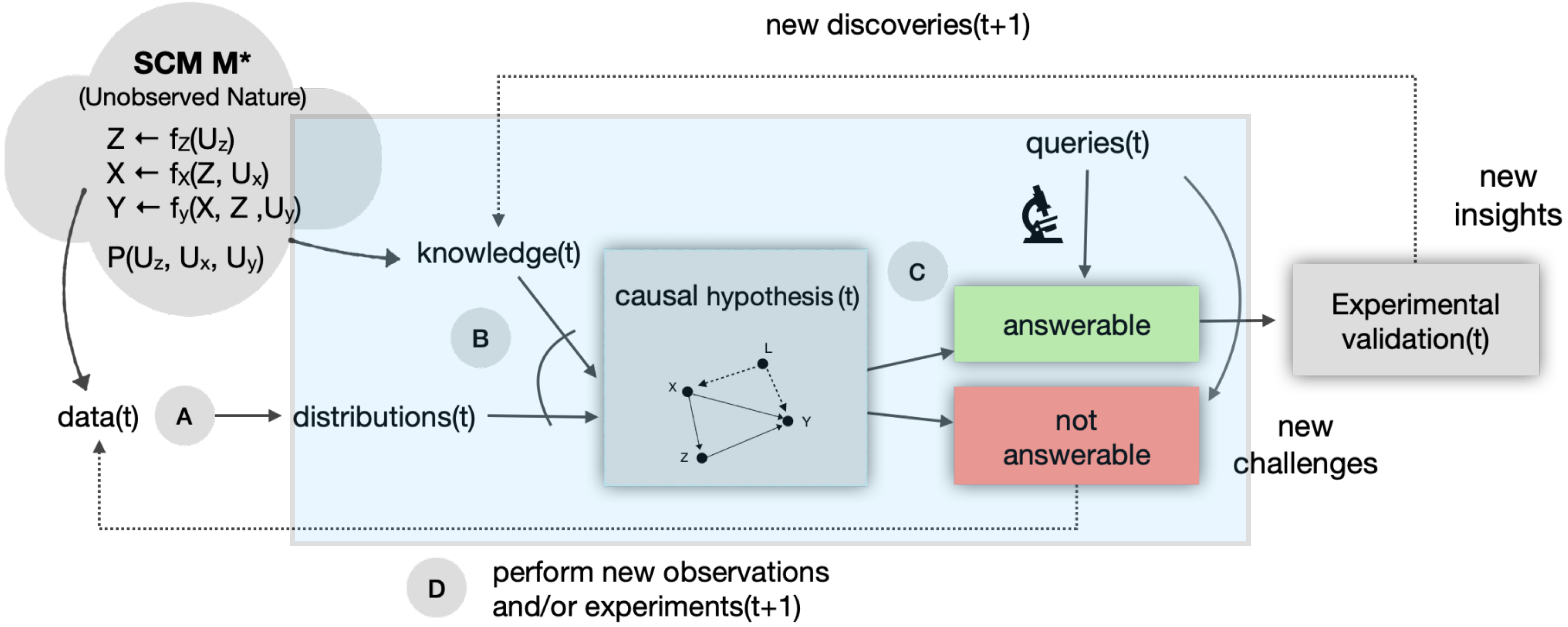
Continuous Process of Scientific Discovery and Causal Hypothesis Refinement



A Statistical Learning | B Causal Learning | C Causal Inference | D Causal Exp. Design

Causal Inference Workflow

Continuous Process of Scientific Discovery and Causal Hypothesis Refinement



A Statistical Learning | B Causal Learning | C Causal Inference | D Causal Exp. Design

Many other Topics in Causal Inference

1. Optimal Causal Experimental Design
2. Causal Representation Learning & Causal Abstraction
3. Causal Reinforcement Learning
4. Fairness & Mediation Analysis
5. Individual Treatment Effect (ITE) Estimation
6. Data-Driven Covariate Selection for Adjustment
7. Partial Effect Identification
8. Many more…

Optimal Experimental Design

Focusing on Causal Discovery:

Experimental Design for Learning Causal Graphs with Latent Variables

Murat Kocaoglu*
Department of Electrical and Computer Engineering
The University of Texas at Austin, USA
mkocaoglu@utexas.edu

Karthikeyan Shanmugam*
IBM Research NY, USA
karthikeyan.shanmugam2@ibm.com

Elias Bareinboim
Department of Computer Science and Statistics
Purdue University, USA
eb@purdue.edu

Kocaoglu, M., Shanmugam, K., & Bareinboim, E. (2017). Experimental design for learning causal graphs with latent variables. *Advances in Neural Information Processing Systems*, *30*.

Focusing on Decision-Making:

Structural Causal Bandits: Where to Intervene?

Sanghack Lee
Department of Computer Science
Purdue University
lee2995@purdue.edu

Elias Bareinboim
Department of Computer Science
Purdue University
eb@purdue.edu

Lee, S., & Bareinboim, E. (2018). Structural causal bandits: Where to intervene?. *Advances in neural information processing systems*, *31*.

Causal Representation Learning & Causal Abstraction

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, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio

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

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

Causal Effect Identification in Cluster DAGs

Tara V. Anand*[1], **Adele H. Ribeiro***[2], **Jin Tian**[3], **Elias Bareinboim**[2]
[1]Department of Biomedical Informatics, Columbia University
[2]Department of Computer Science, Columbia University
[3]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

Causal Reinforcement Learning

http://crl.causalai.net

TASK 1

Generalized Policy Learning

combining online + offline learning

Learn policy Π by systematically combining offline (L_1) and online (L_2) modes of interaction.

TASK 2

When and Where to Intervene?

refining the policy space

Identify subset of L_2 to refine the policy space do(Π(X)) based on topological constraints implied by *M* on *G*.

TASK 3

Counterfactual Decision-Making

changing optimization function based on intentionality, free will, and autonomy

Optimization criterion based on counterfactuals and L_3-based randomization (instead of L_2/do()-counterpart).

TASK 4

Generalizability & Robustness of Causal Claims

transportability & structural invariances

Generalize policy based on structural invariances shared across training (SCM *M*) and deployment environments (*M**).

TASK 5

Learning Causal Models

discovering the causal structure with observation and experiments

Learn the causal graph *G* (of *M*) by systematically combining observations (L_1) and experimentation (L_2).

TASK 6

Causal Imitation Learning

policy learning with unobserved rewards

Construct L_2-policy based on partially observable L_1-data coming from an expert with unknown reward function.

By Elias Bareinboim's Research Group

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 FREDRIK.JOHANSSON@CHALMERS.SE
Chalmers University of Technology
Göteborg, 412 96, Sweden

Uri Shalit URISHALIT@TECHNION.AC.IL
Technion - Israel Institute of Technology
Haifa, 3200003, Israel

Nathan Kallus KALLUS@CORNELL.EDU
Cornell University
New York, NY 10044, USA

David Sontag DSONTAG@CSAIL.MIT.EDU
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Johansson, F.D., Shalit, U., Kallus, N. and Sontag, D., 2022. Generalization bounds and representation learning for estimation of potential outcomes and causal effects. *Journal of Machine Learning Research*, *23*(166), pp.1-50.

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* FREJOHK@CHALMERS.SE
CSE, Chalmers University of Technology, Göteborg, SE-412 96, Sweden

Uri Shalit* SHALIT@CS.NYU.EDU
David Sontag DSONTAG@CS.NYU.EDU
CIMS, New York University, 251 Mercer Street, New York, NY 10012 USA

* **Equal contribution**

Data-Driven Covariate Selection for Adjustment

Finding Valid Adjustments under Non-ignorability with Minimal DAG Knowledge

Abhin Shah
MIT
abhin@mit.edu

Karthikeyan Shanmugam
IBM Research
karthikeyan.shanmugam2@ibm.com

Kartik Ahuja
Mila
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

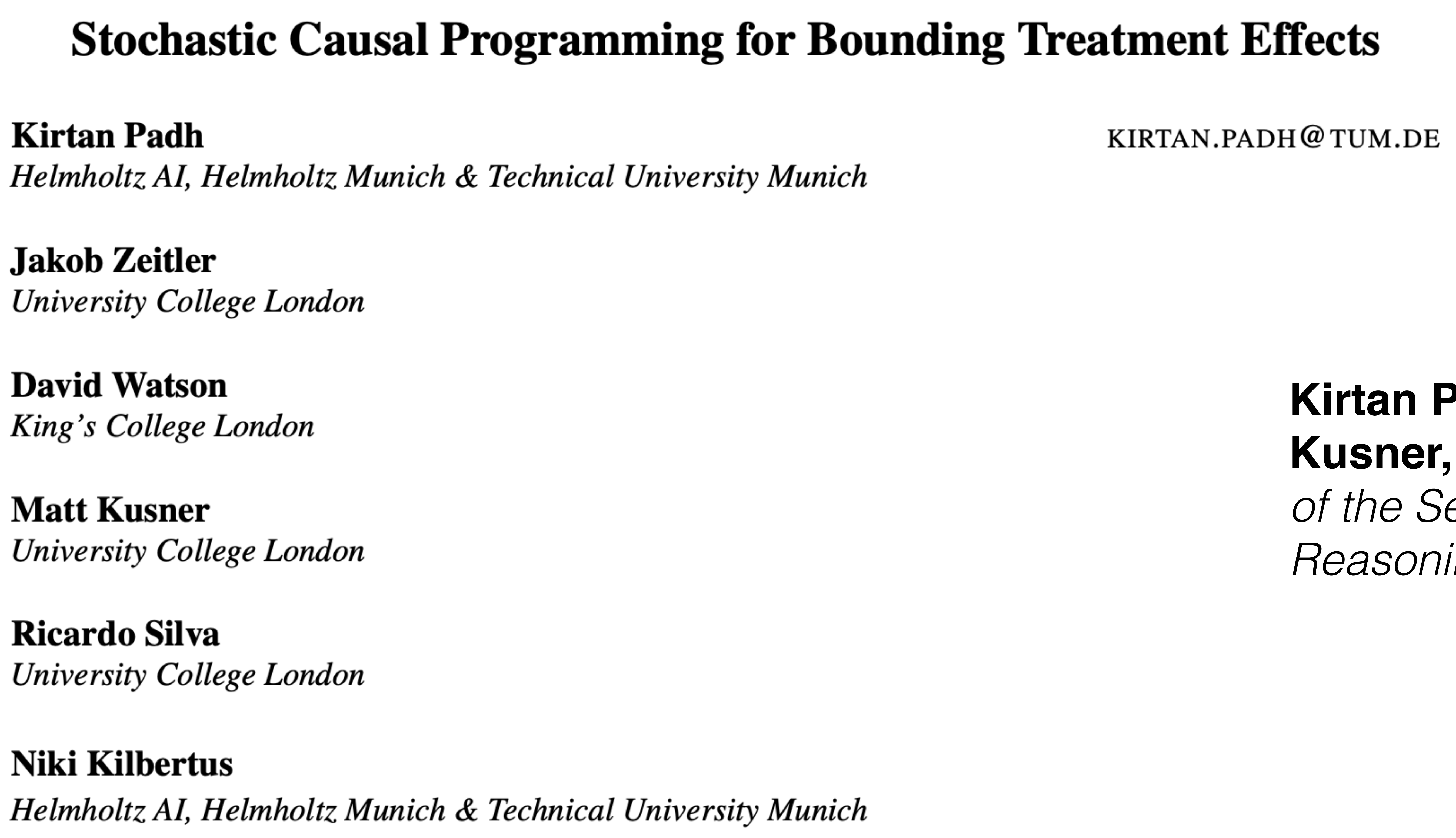
Limor Gultchin
University of Oxford
The Alan Turing Institute

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 KIRTAN.PADH@TUM.DE
Helmholtz AI, Helmholtz Munich & Technical University Munich

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

Current Challenges & Open Problems

- Robustness in real-world scenarios, with small (unfaithful) datasets.
- Uncertainty modeling, lack of ground-truth in real-world applications.
- Integration of expert / human knowledge
- Scalability in insufficient systems — development of adaptive approaches.
- Effect identification in more general equivalence classes.
- Causal experimental design — what if a causal relation or effect is not identified?
- Causal effects among abstractions: connection with causal abstraction and causal representation learning.
- Learning from multi-modal datasets — connection with causal abstraction and causal representation learning.
- Continual causal discovery

Thank you! :)

If any of these topics interest you, feel free to reach out:

adele.ribeiro@uni-muenster.de