Introduction to Causal Inference

Adèle H. Ribeiro

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

Institute of Medical Informatics
University of Münster

Lisbon Machine Learning School (LxMLS)
July 24, 2025

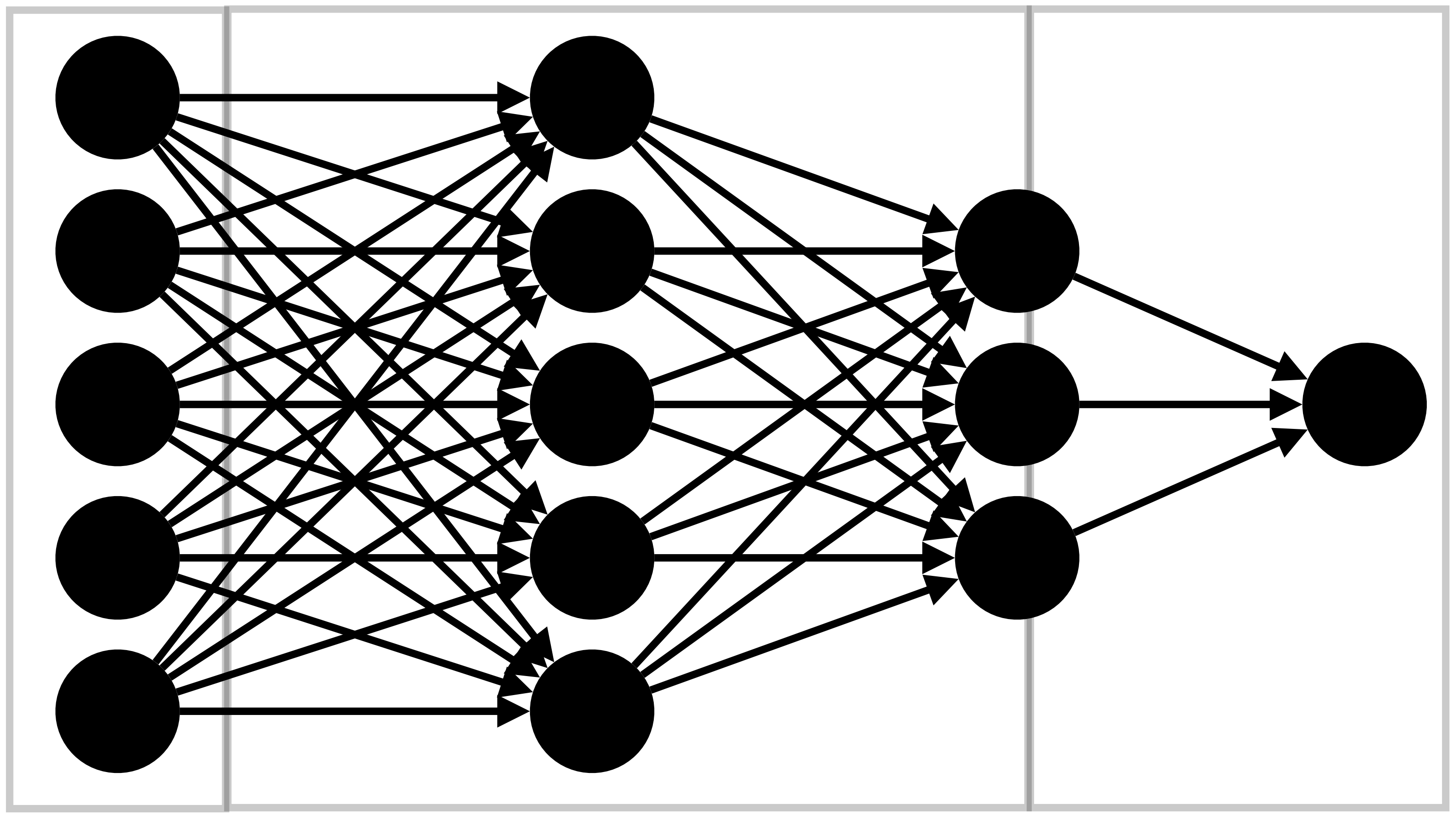
Today's AI: Powerful Predictors Built on Correlation

Representation Learning

Regression

$\mathbf{X}$: Input

Observed features



$\mathbf{Z}$: Learned features

Prediction

$P(Y = y \,|\, \mathbf{X})$

$E(Y \,|\, \mathbf{X})$

Medical Diagnosis

$P(Y = \text{disease} \,|\, X = \text{patient features})$

Image Classification

$P(Y = \text{dog} \,|\, X = \text{pixel array})$

Language Modeling

$P(Y = w_t \,|\, X = w_1, \ldots, w_{t-1})$

Speech Recognition

$P(Y = \text{text} \,|\, X = \text{audio})$

Weather Forecasting

$P\left(Y = T_{t+h} \,|\, T_{t-1}, T_{t-2}, \ldots T_{t-p}\right)$

We learn a model $f : \mathcal{X} \mapsto \mathcal{Y}$, such that $\hat{Y} = f(\mathbf{x})$ is the optimal prediction of Y by capturing the conditional distribution $P(Y \,|\, \mathbf{X})$

Remarkable advances in estimating $P(Y \,|\, \mathbf{X})$ include DNNs, transformers, GNNs, …

Does a predictive model explain the world?

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

X: **Number of firefighters in action**

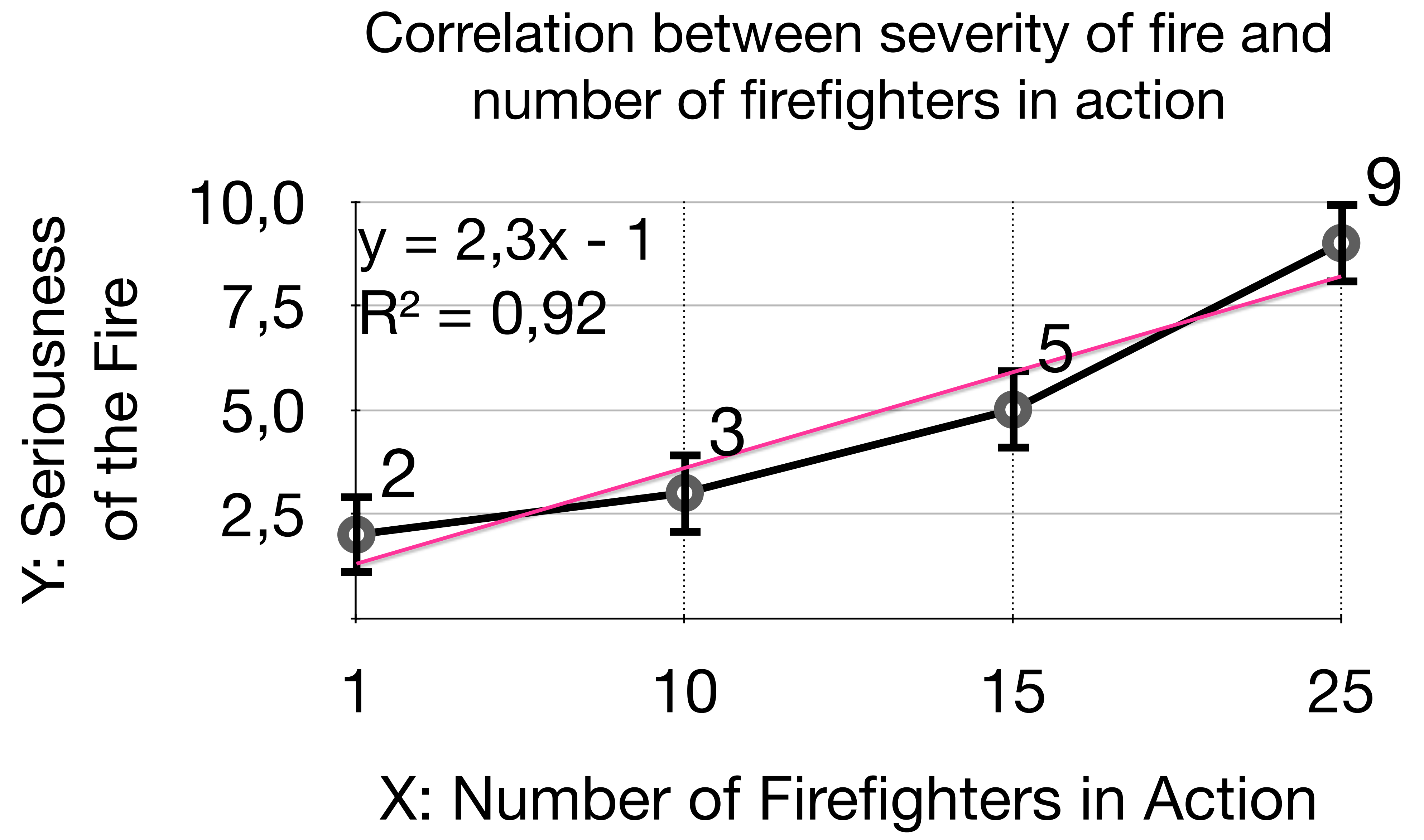
Y: **Severity of the (initial) fire**

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



Positive Correlation:

Changing X will change our prediction for Y:

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

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

Structural Causal Model (SCM)

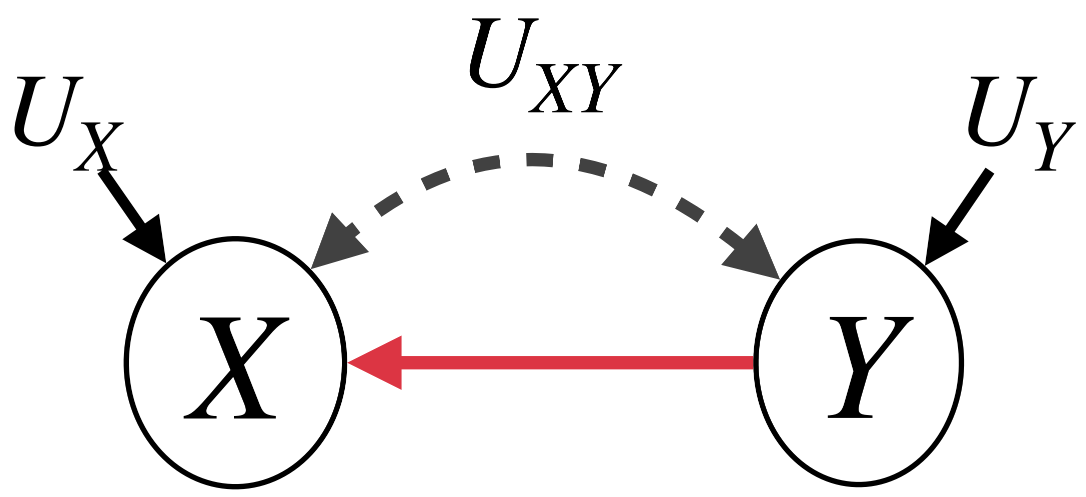
X does not cause Y,
it is the other way around!

Changing X won't change the value of Y in reality!

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

Interventional Probability Distribution

Reverse Causation!



Causal Diagram

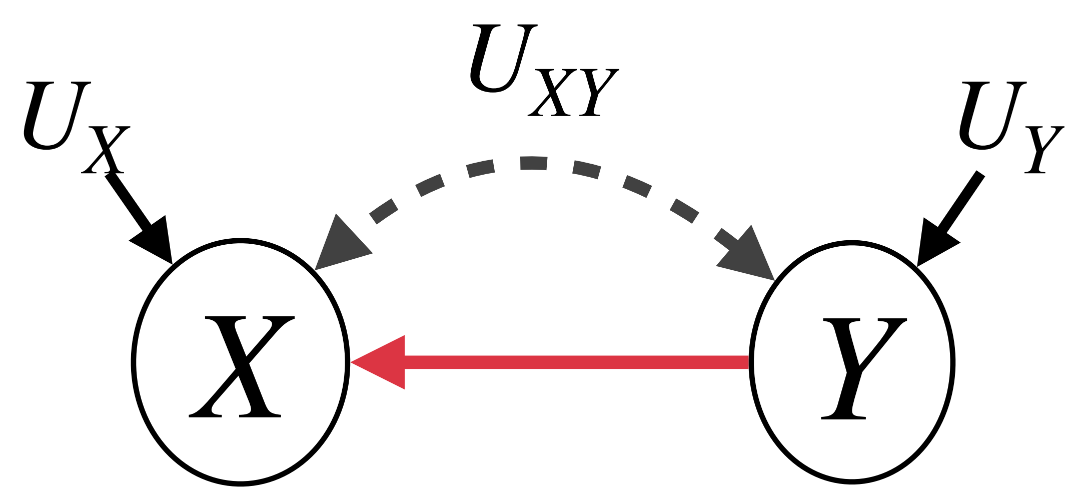
In this case, $P(Y = y \mid see(X = x)) \neq P(Y = y)$

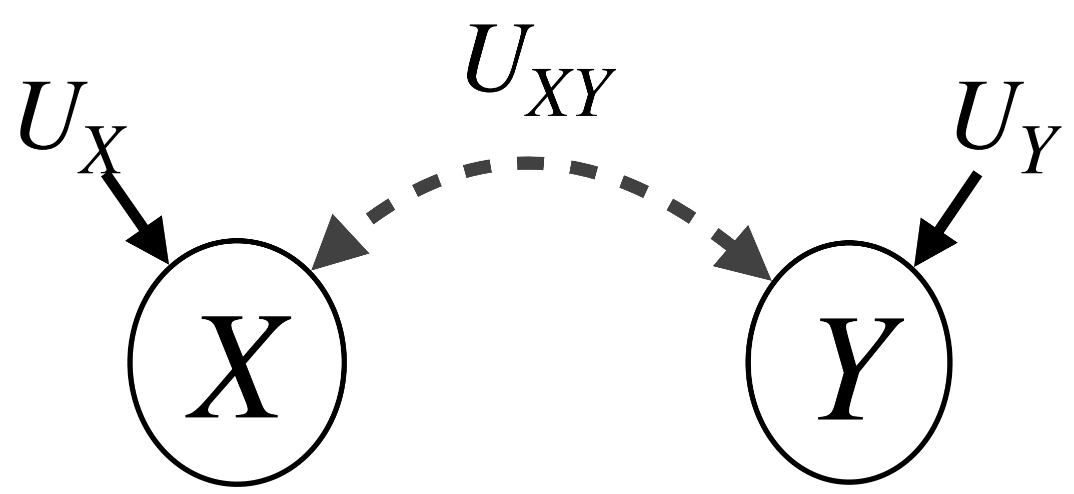
but $\forall x, P(Y = y \mid do(X = x)) = P(Y = y)$

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

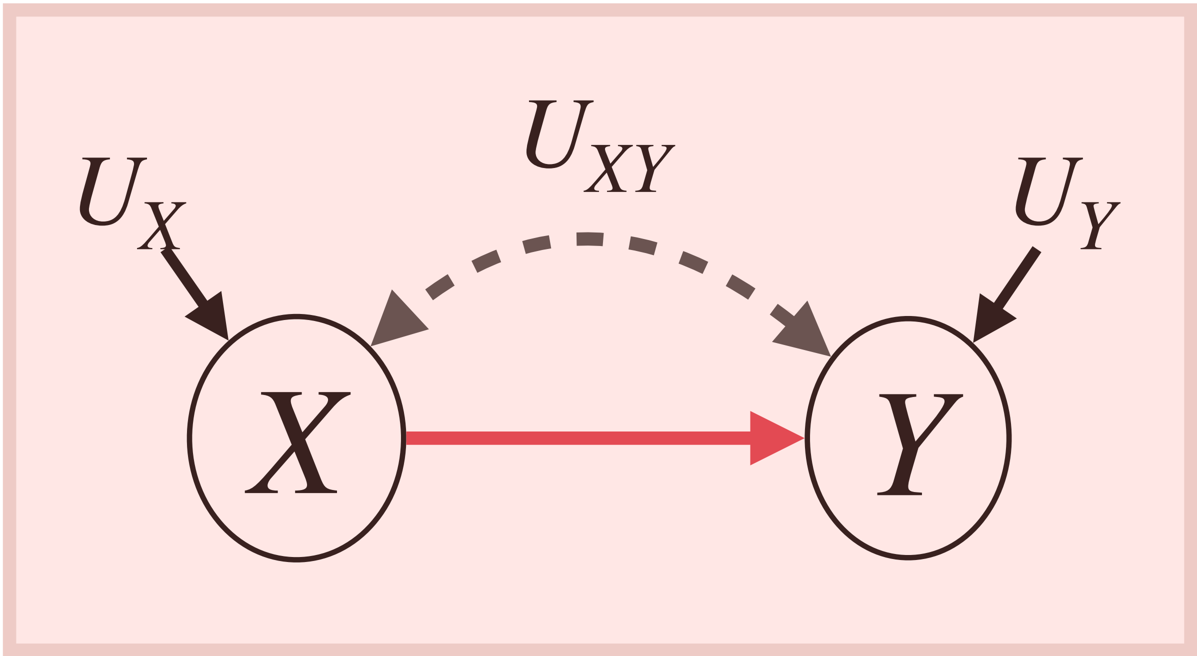
Interpreting the Link: Nurses and Patient Mortality

X: **Number of nurses in a hospital**
Y: **Patient mortality**





Hospitals with more nurses on staff sometimes report higher patient mortality rates.



Typically, a negative causal effect from improved patient care.

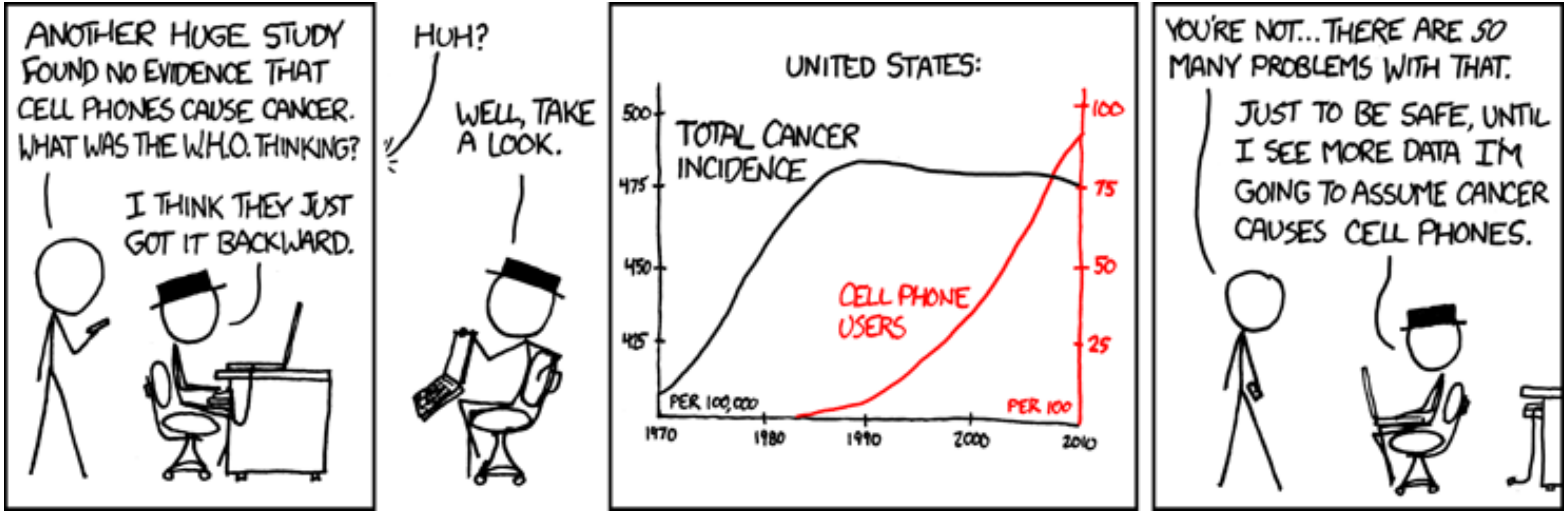
X is positively correlated to Y

↓ Regression / Machine Learning

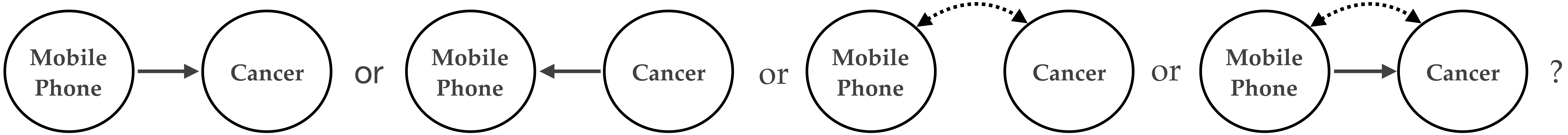
$$\hat{Y} = f(X)$$

U_{XY}: Hospital size, patient severity, or emergencies (latent confounders) can increase both nurse demand and mortality, causing a (positive) spurious correlation.

Even with Time, Can Associations be Misleading?



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



AI predicts everything, but does it explain the world?

Challenges emerge in more complex tasks that demand careful consideration of **biases** and **underlying mechanisms**, including:

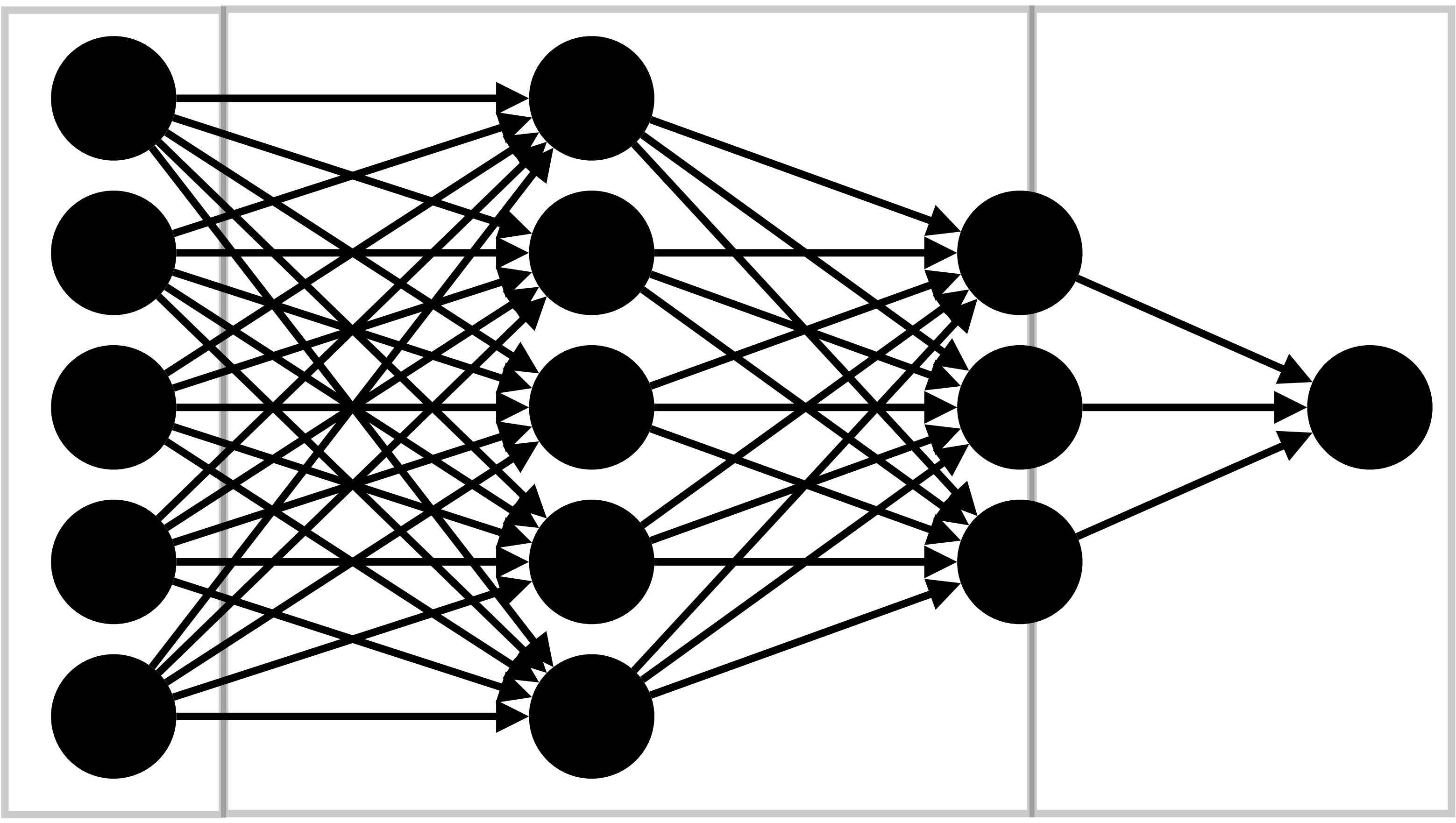
- **Explaining** the underlying data-generating processes,
- Providing **unbiased** estimates of effects of interventions,
- Identifying **optimal** and **personalized** approaches,
- Ensuring **fairness** in clinical decision support systems,
- Achieving **generalizability** across diverse domains / populations.

Those require causal / real-world insights!!

What can be done to overcome these challenges?

Can Model Explainability Provide Causal Insights?

X: Observed features



$\hat{Y} = f(\mathbf{x})$

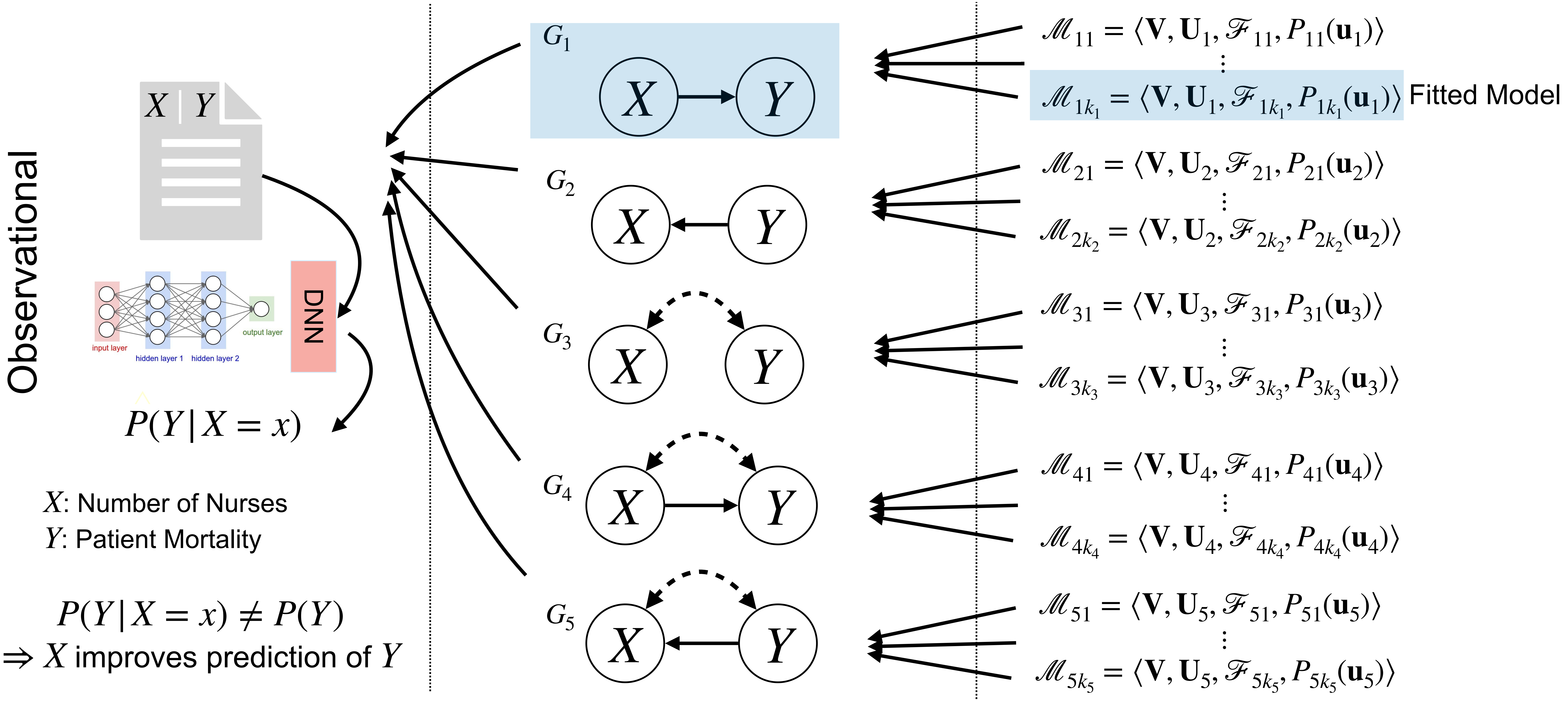
Optimal prediction of Y, capturing $P(Y \mid \mathbf{X})$

$f : \mathscr{X} \mapsto \mathscr{Y}$

The promise of model explainability:

- **Feature Importance:** SHAP, LIME, Permutation Importance
- **Visualization Techniques:** Partial Dependence Plots, Saliency Maps
- **Counterfactuals Explanations:** What-If Analysis
- **Model Probing:** Ablation Studies, Sensitivity Analysis

Why Doesn't Model Explainability Imply Causality?



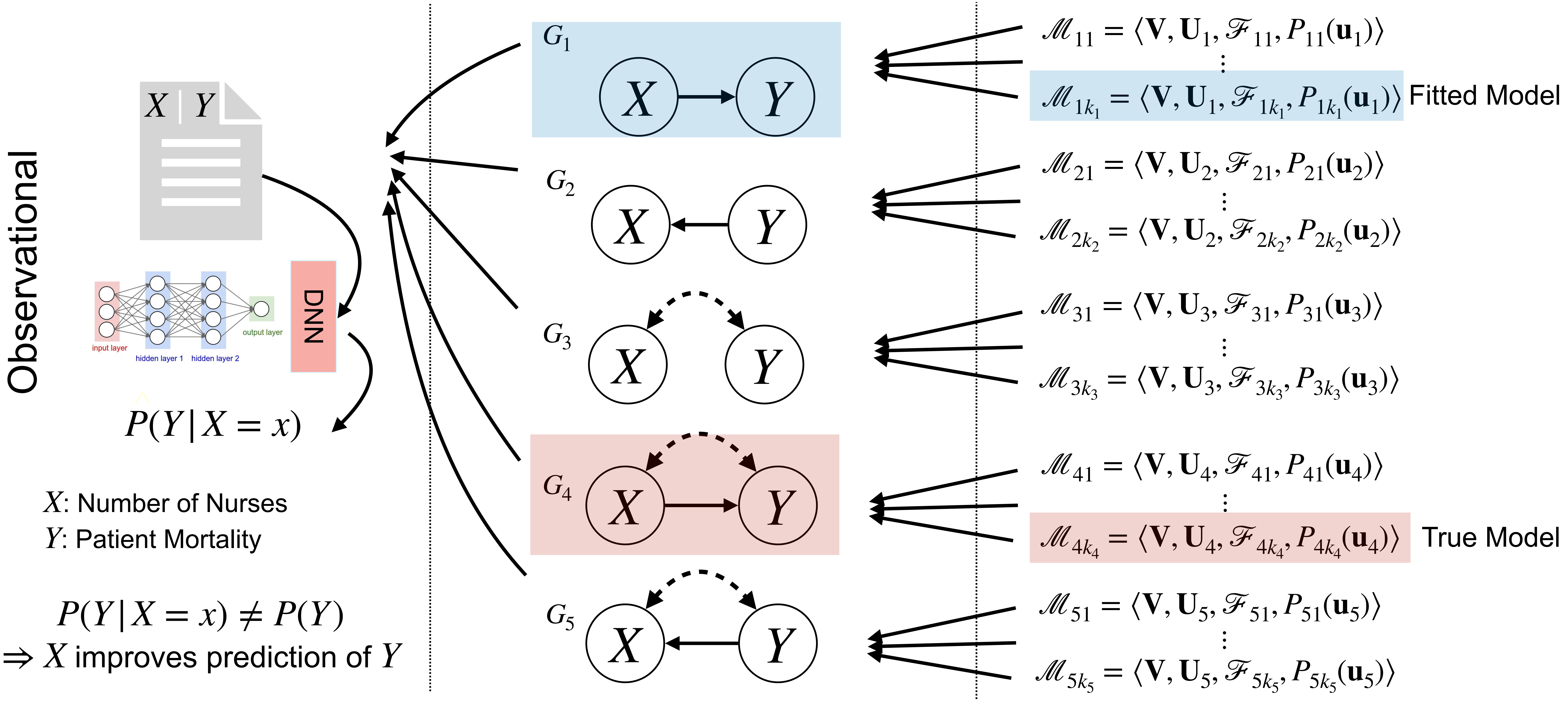
X: Number of Nurses

Y: Patient Mortality

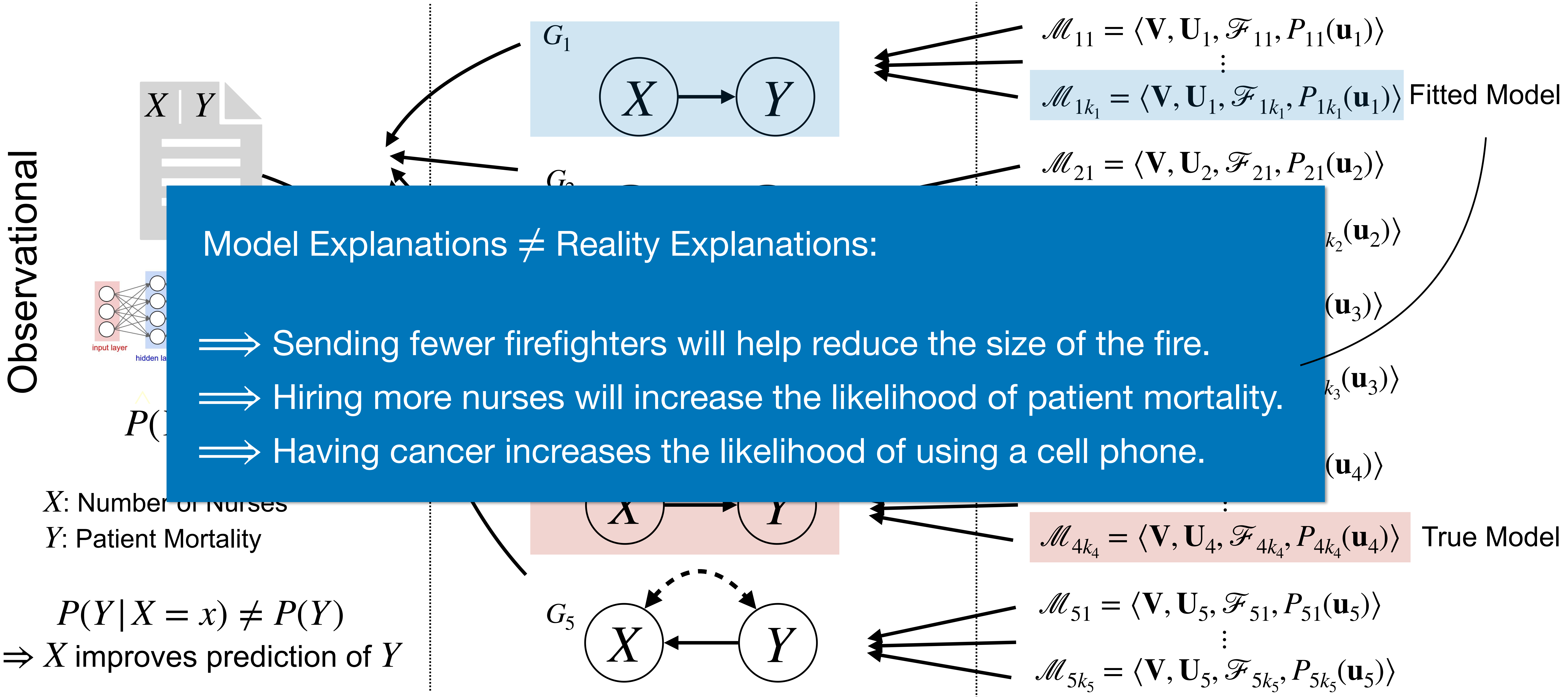
$P(Y|X=x) \neq P(Y)$

$\Rightarrow X$ improves prediction of Y

Why Doesn't Model Explainability Imply Causality?



Why Doesn't Model Explainability Imply Causality?



Model Explanations $\neq$ Reality Explanations:

$\Longrightarrow$ Sending fewer firefighters will help reduce the size of the fire.

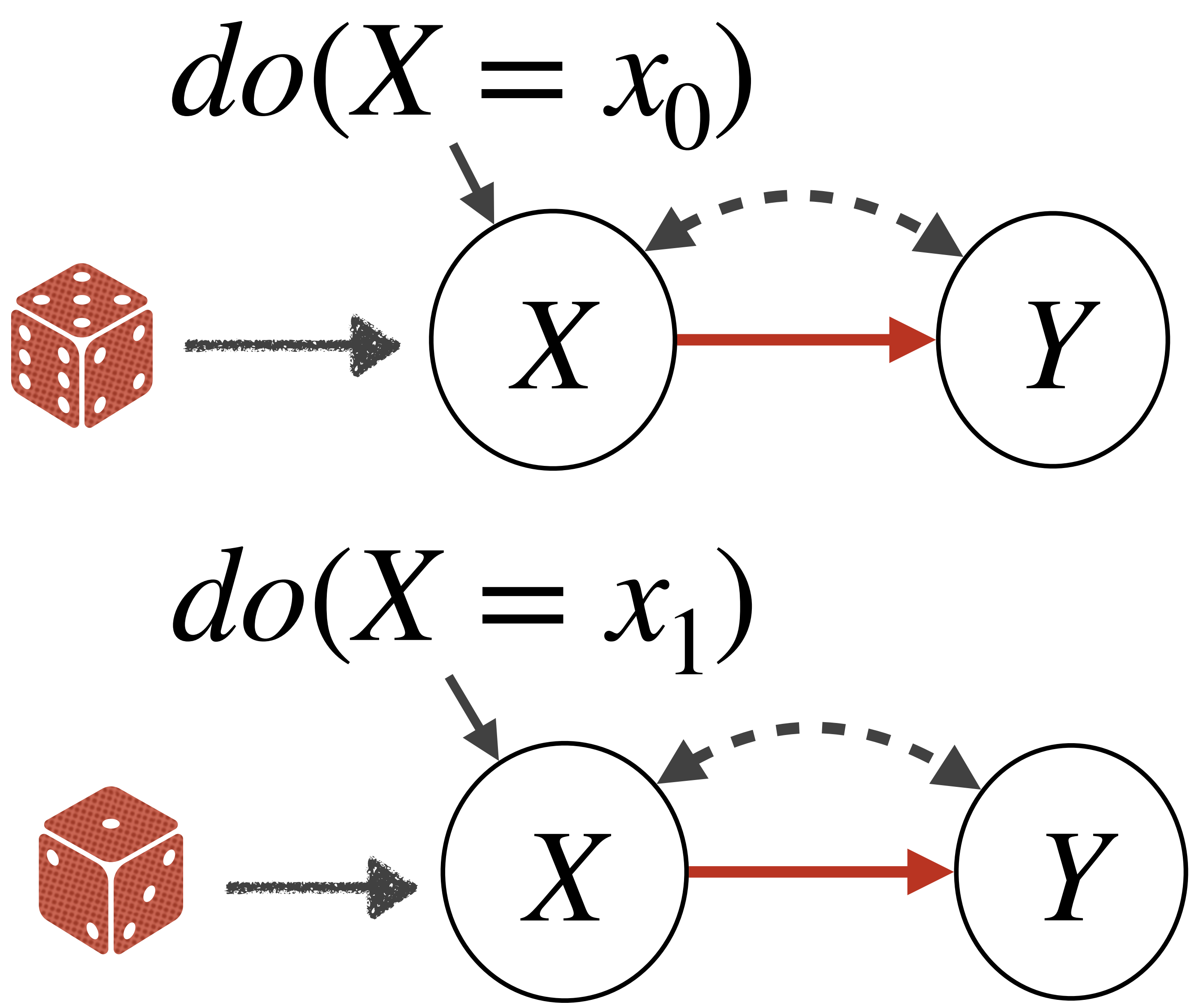
$\Longrightarrow$ Hiring more nurses will increase the likelihood of patient mortality.

$\Longrightarrow$ Having cancer increases the likelihood of using a cell phone.

Randomized Experiments

A well accepted way to access $P(Y|do(X=x))$ 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: $\mathbb{E}[Y|do(X=x_0)] - \mathbb{E}[Y|do(X=x_1)]$

How do we move from prediction to
true understanding of reality
without randomized experiments?

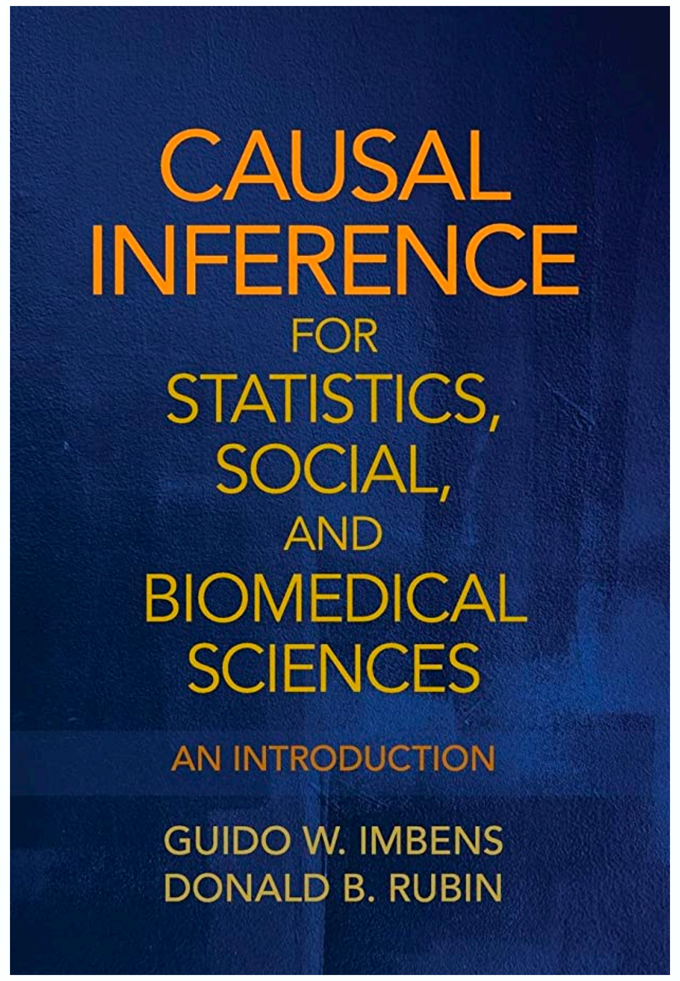
Donald B. Rubin, Guido W. Imbens & Joshua D. Angrist

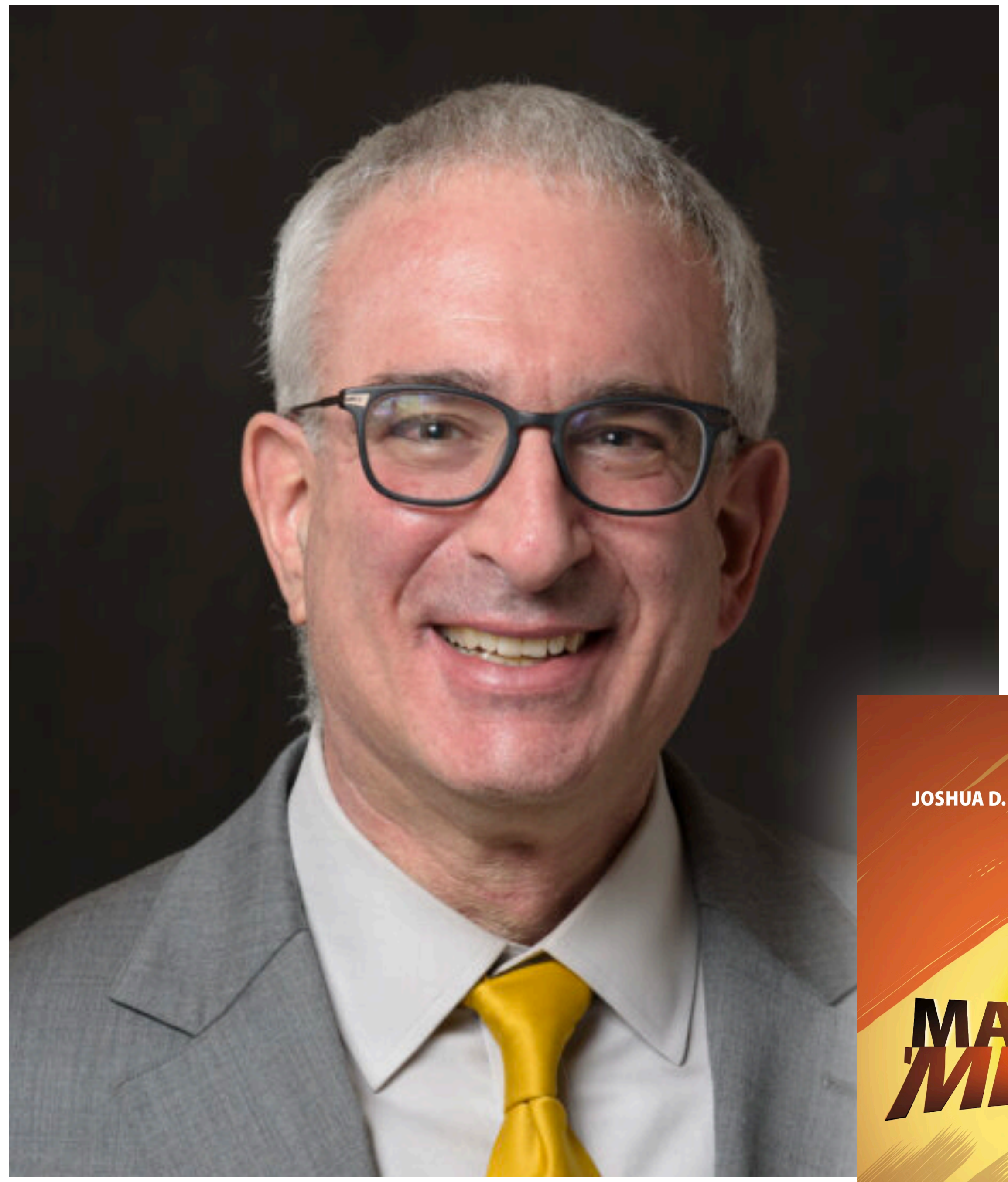
Donald B. Rubin

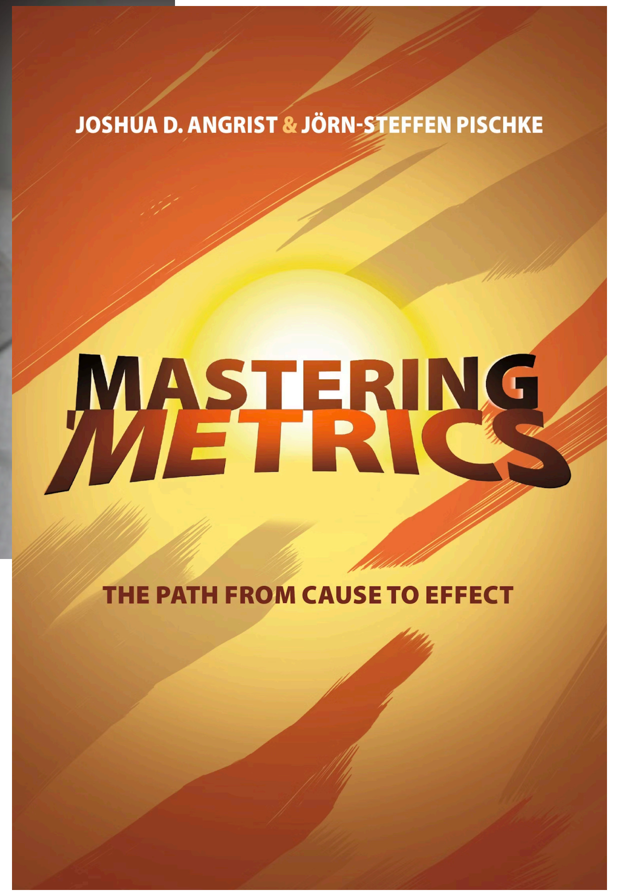
Professor of Statistics at Harvard University

Guido W. Imbens

Professor of Applied Econometrics at Stanford University





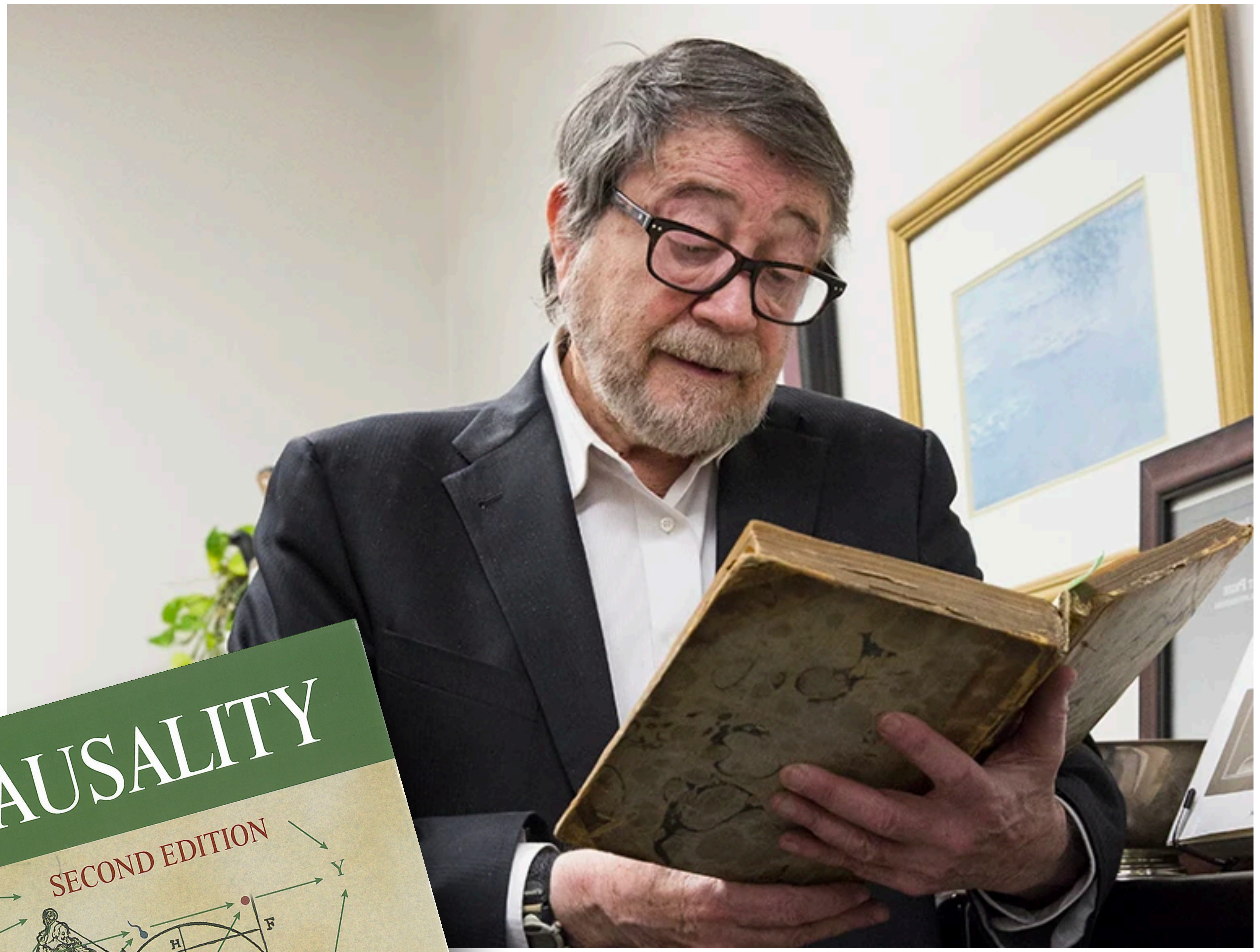


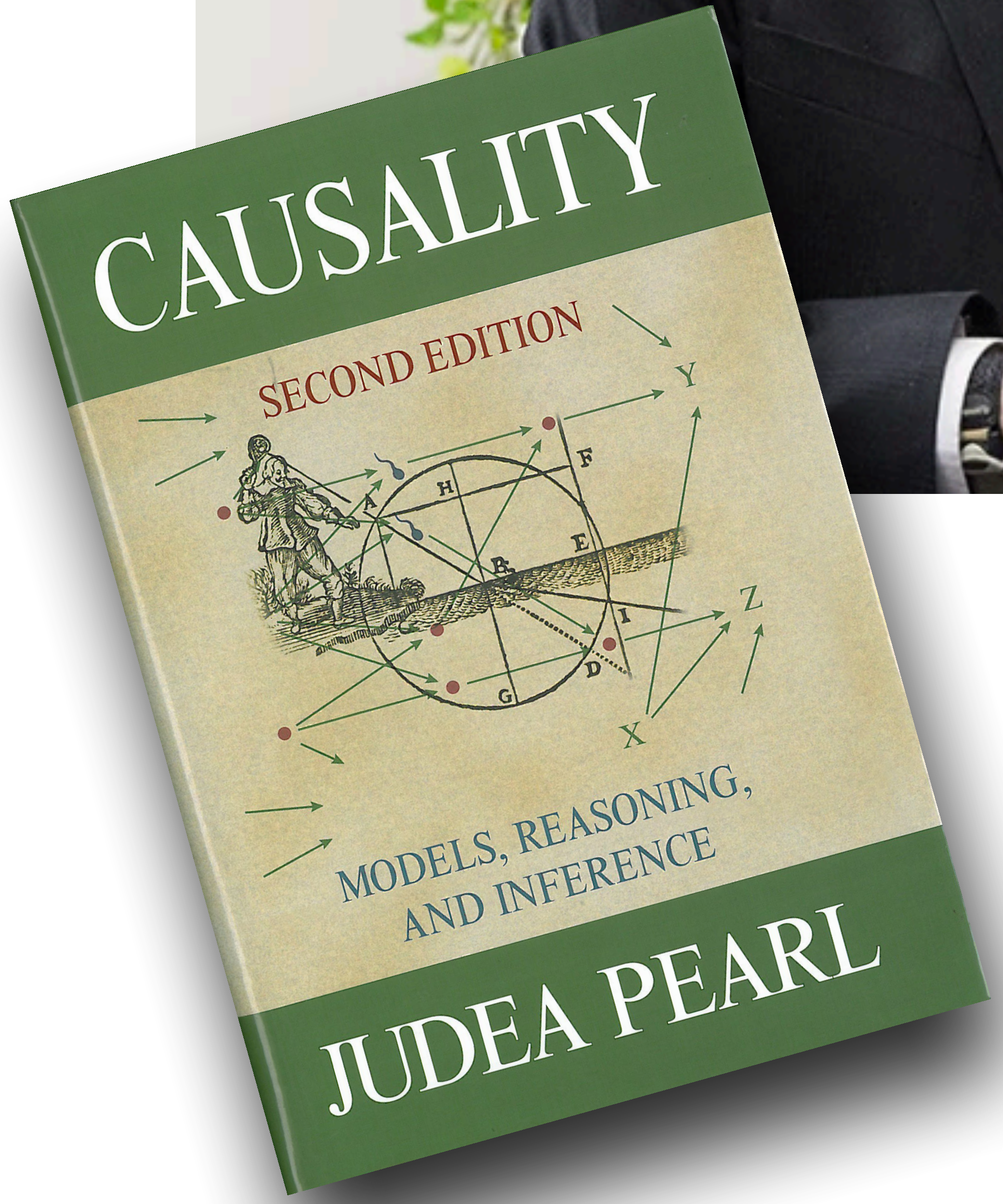
Joshua D. Angrist

Professor of Economics at MIT

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

Judea Pearl — Causal Artificial Intelligence





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

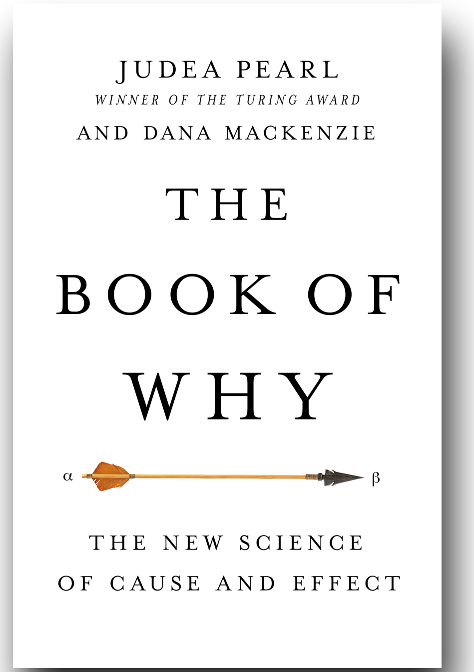
In 2011, he won the A. M. Turing Award (the highest distinction in computer science and a $250,000 prize)

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

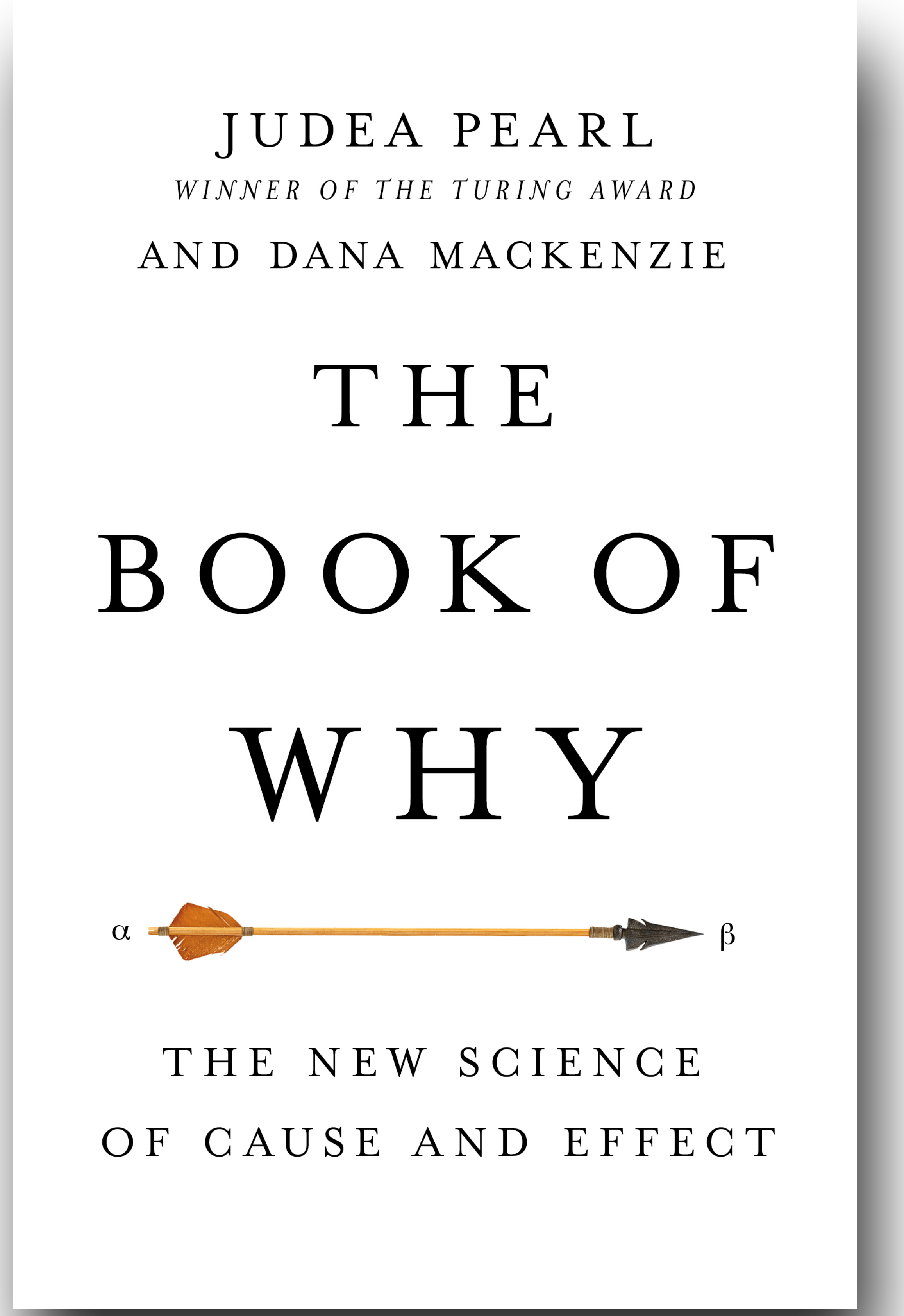
— Association for Computing Machinery (ACM)

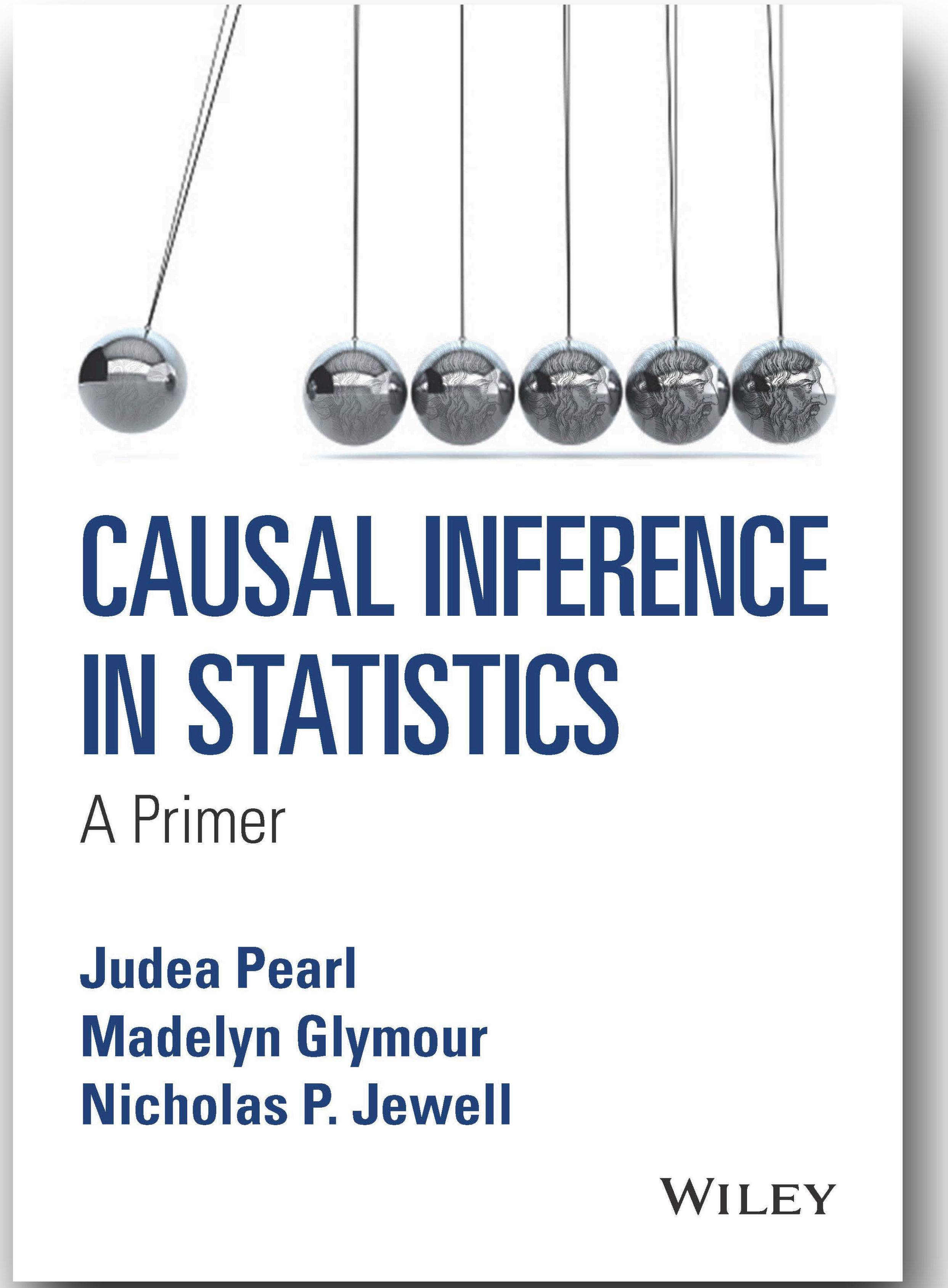
“Deep learning has instead given us machines with truly impressive abilities but no intelligence. The difference is profound and lies in the absence of a model of reality.”

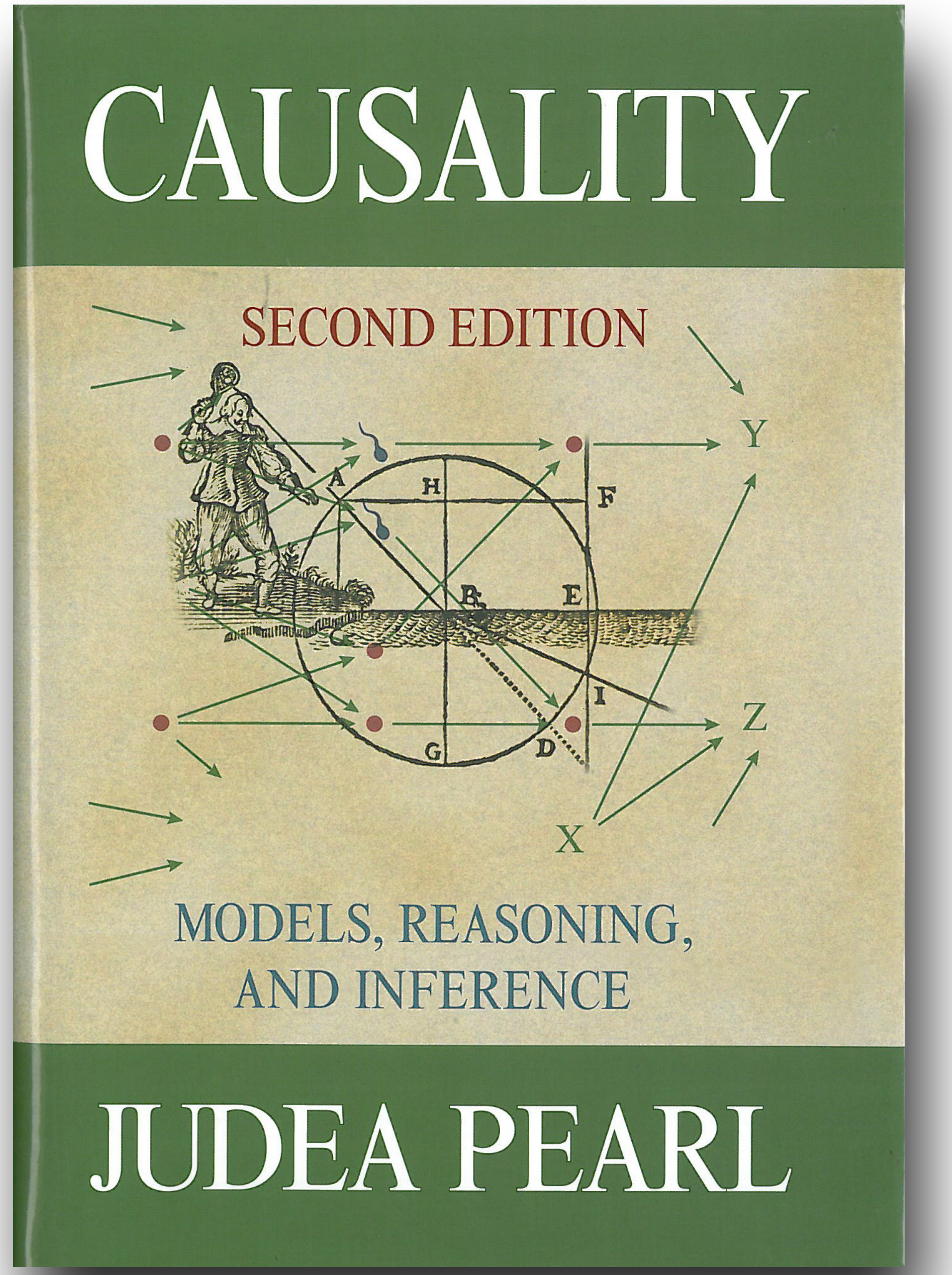
— The Book of Why: The New Science of Cause and Effect



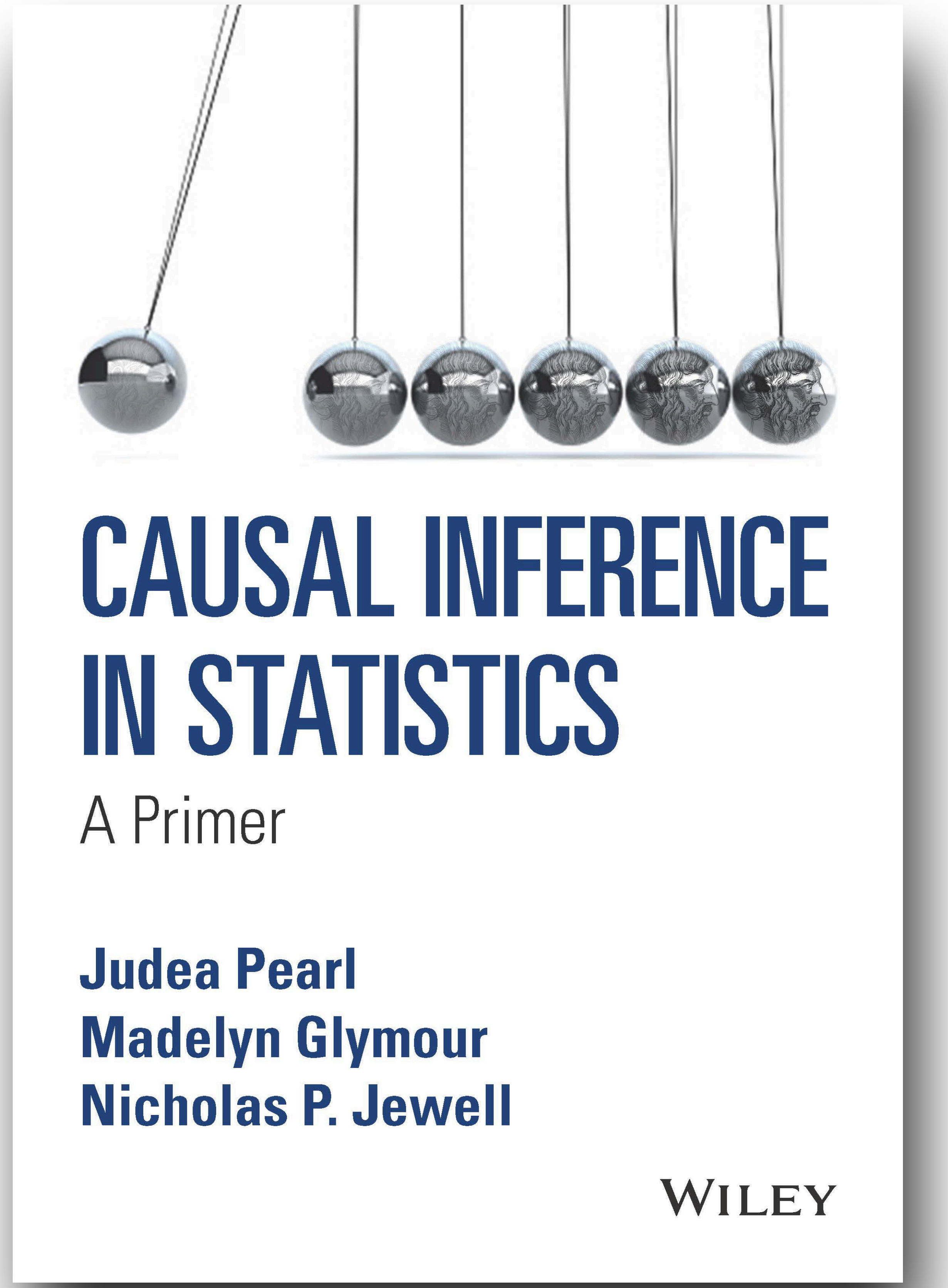
Causality Theory by Judea Pearl



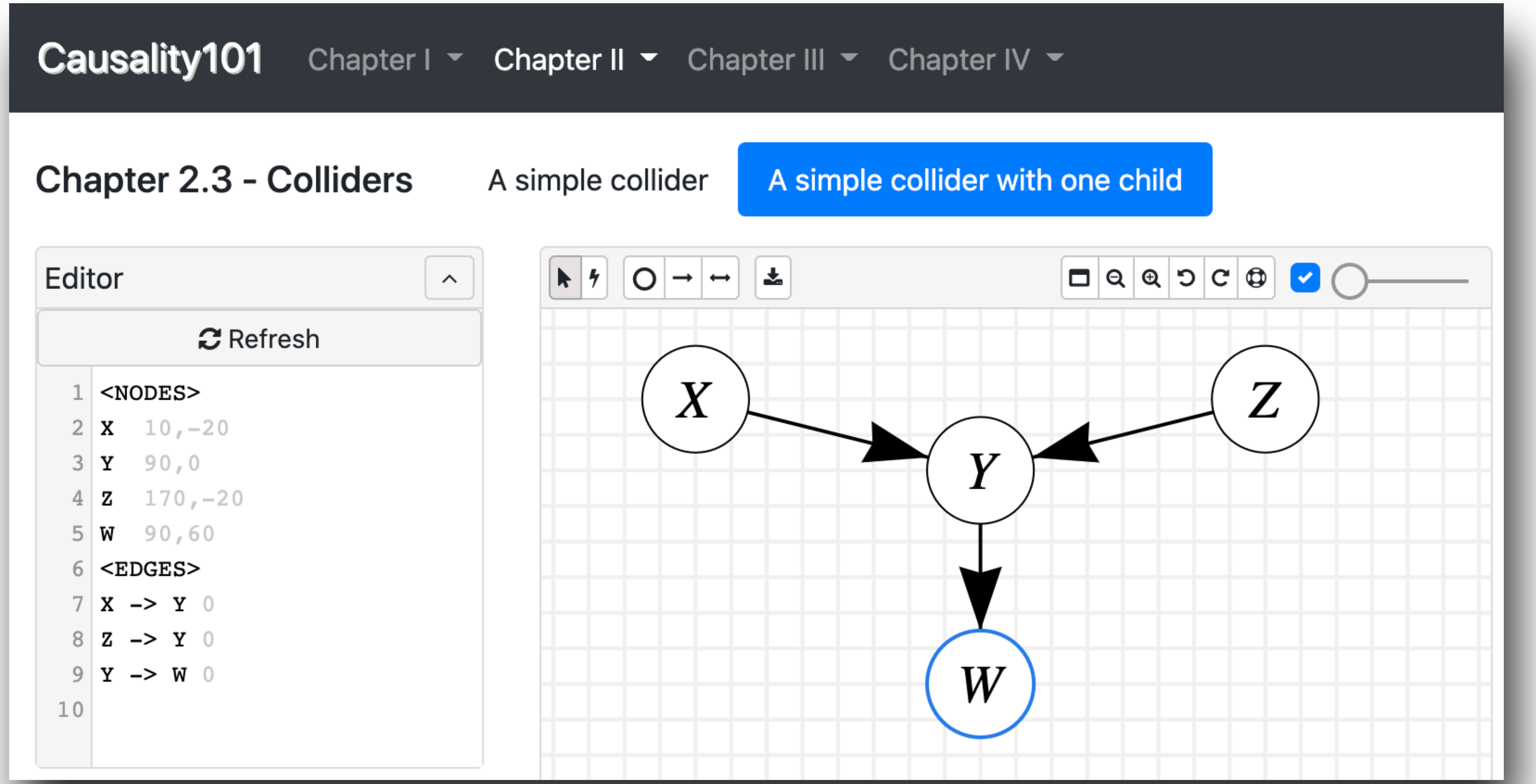




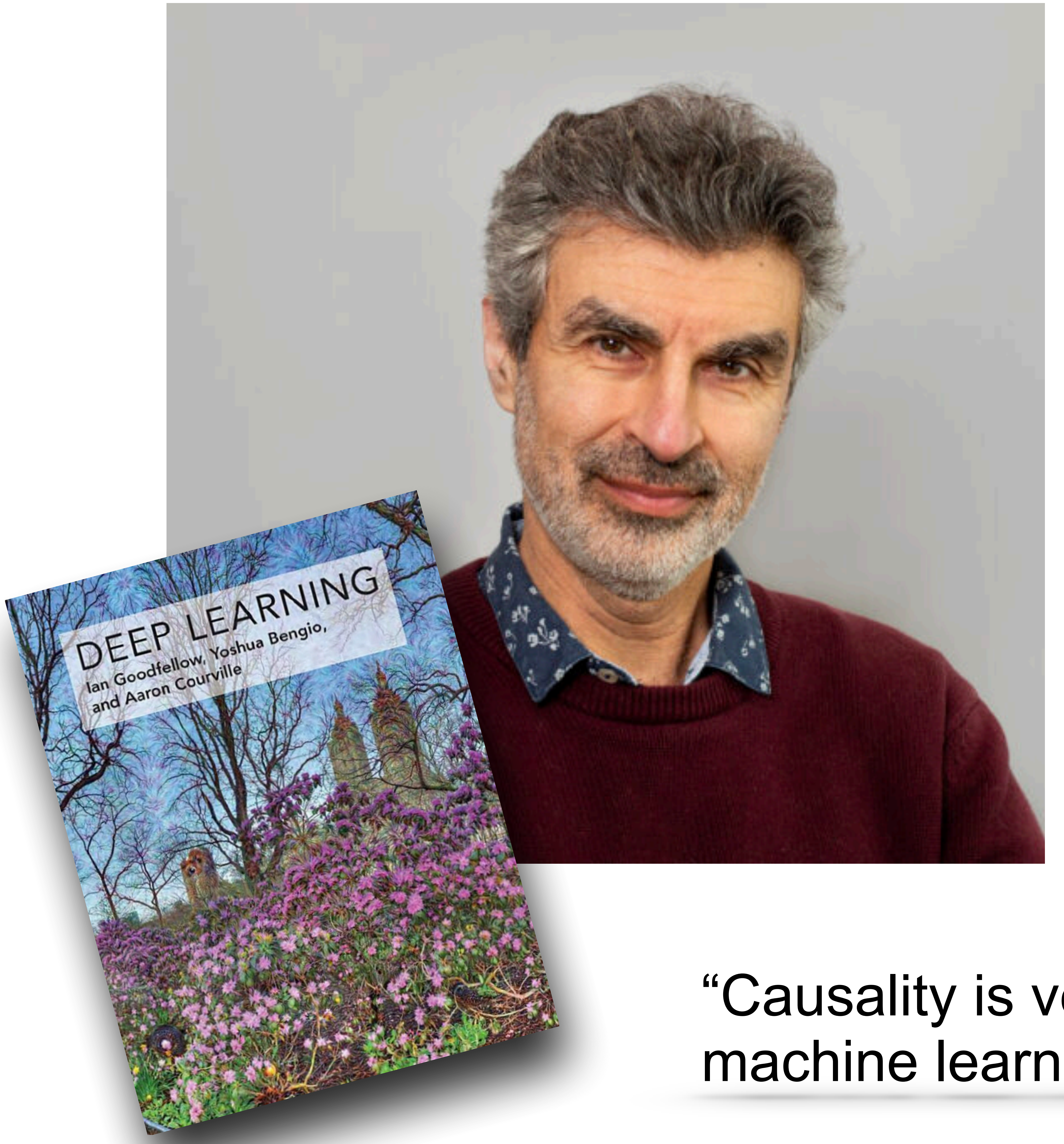
Causality Theory by Judea Pearl



https://causality101.net/



Yoshua Bengio — Deep Learning



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

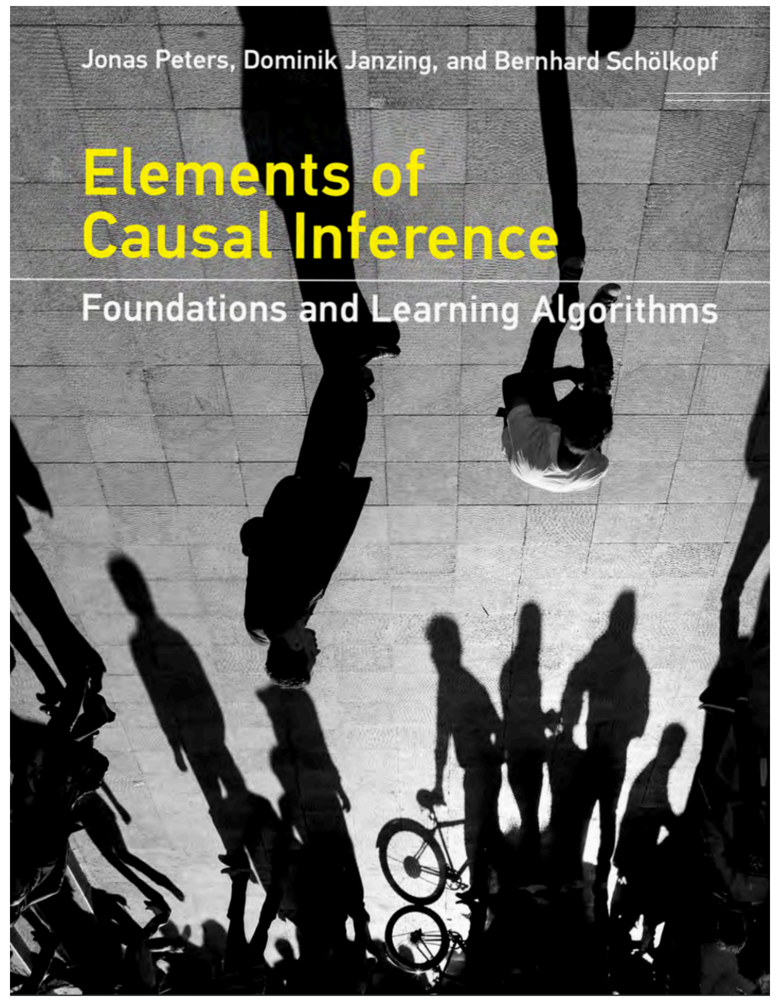
In 2018, he won the A. M. Turing Award, with Geoffrey Hinton, and Yann LeCun

“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*.

Notable Books and Lecture Notes



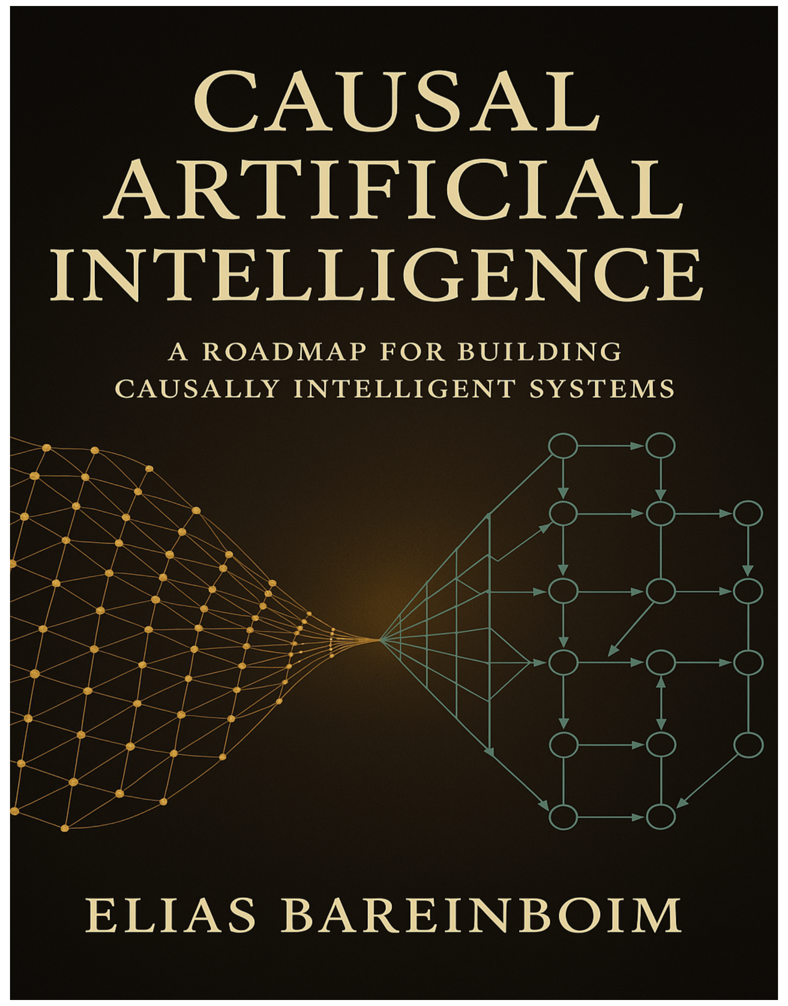
A Mathematical Introduction to Causality

Lecture Notes

Patrick Forré & Joris M. Mooij

July 18, 2025

University of Amsterdam



https://causalai-book.net/

Advanced Data Analysis from an Elementary Point of View

Cosma Rohilla Shalizi

Carnegie Mellon University

Introduction to Causal Inference

from a Machine Learning Perspective

Brady Neal

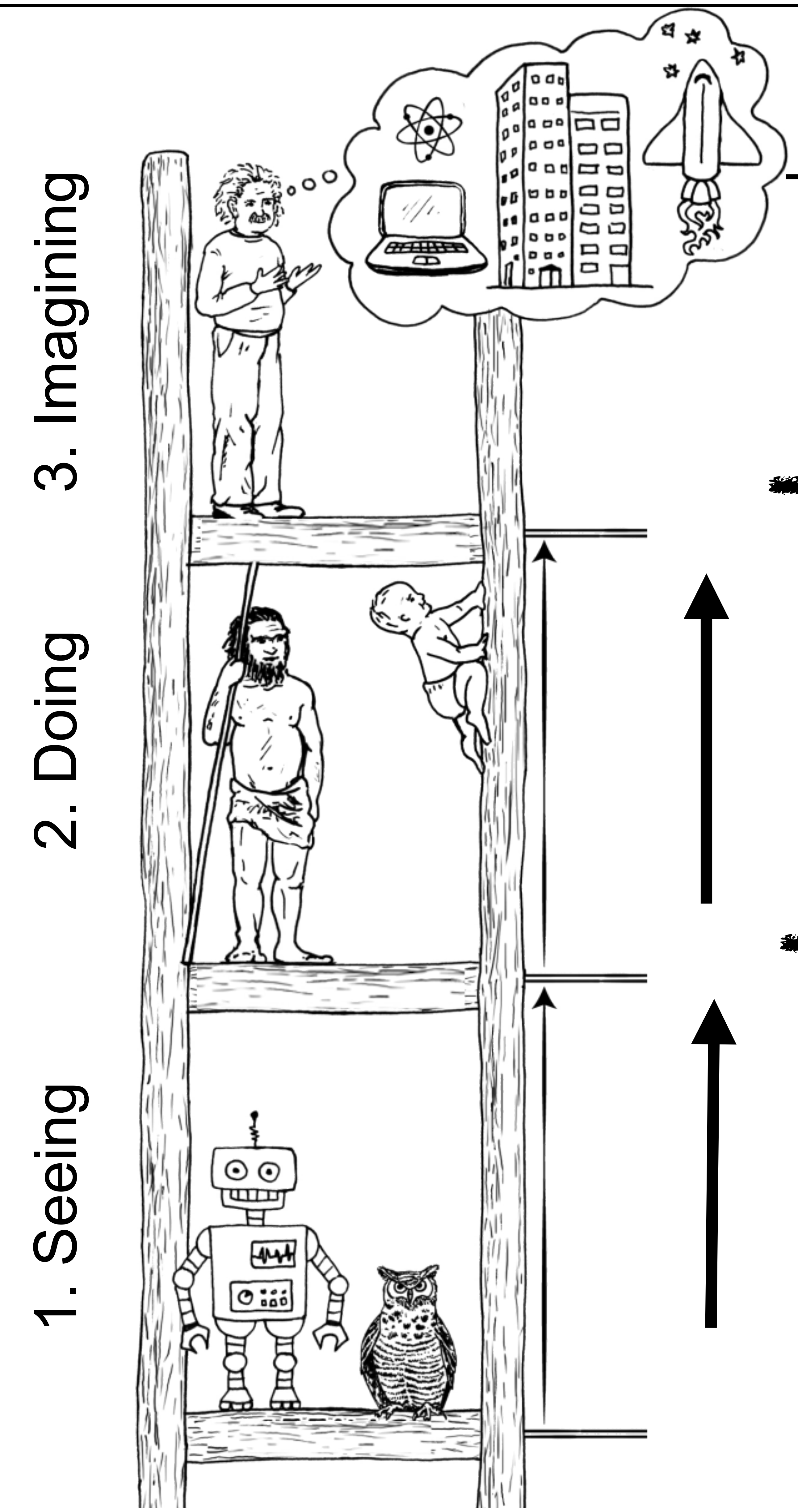
December 17, 2020

Mila - Quebec AI Institute

Pearl’s Causal Hierarchy (PCH)

The Three Inferential Layers

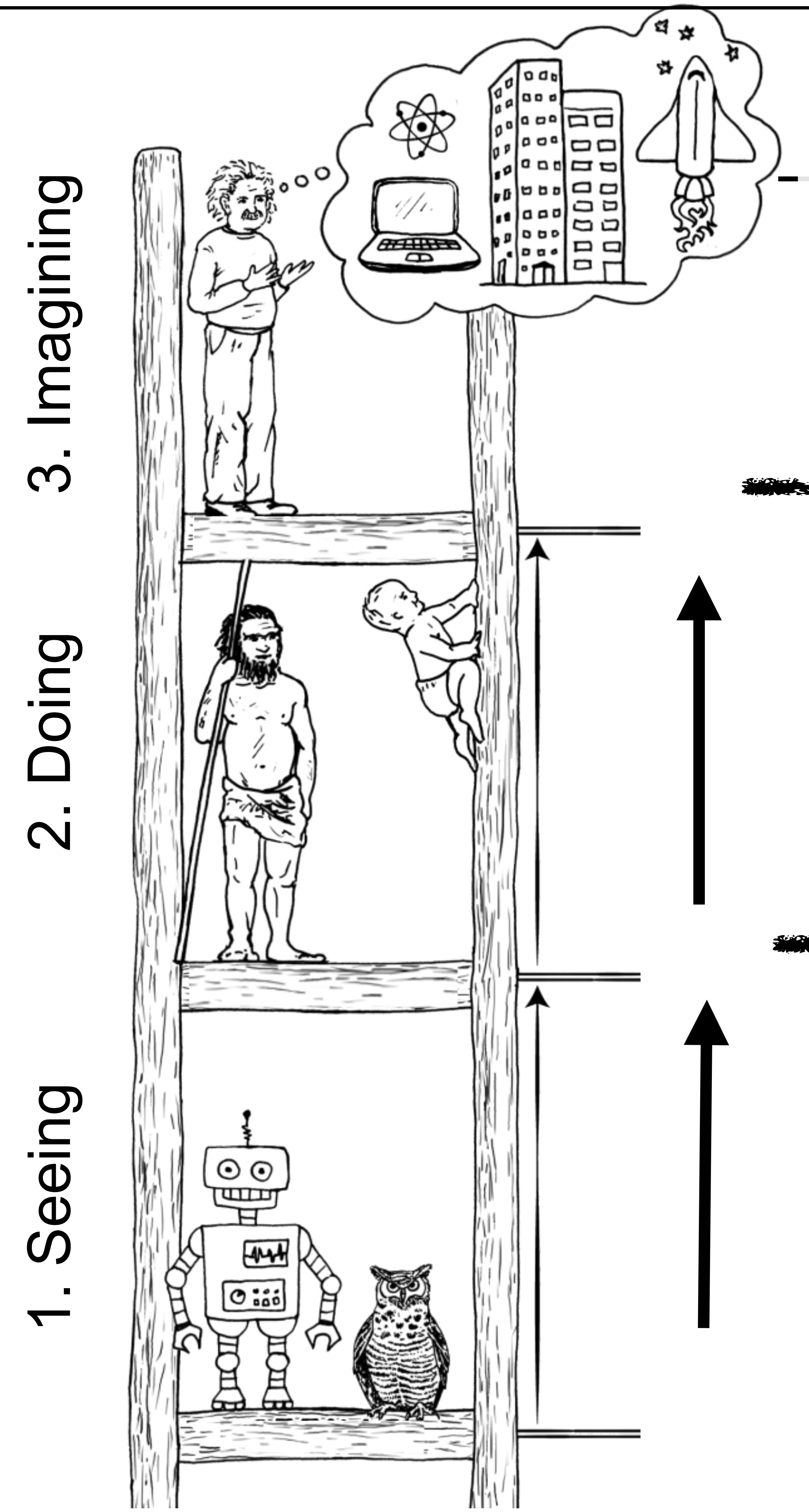
Ladder of Causation



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

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

Ladder of Causation



Cross-layer inferences:

Doing

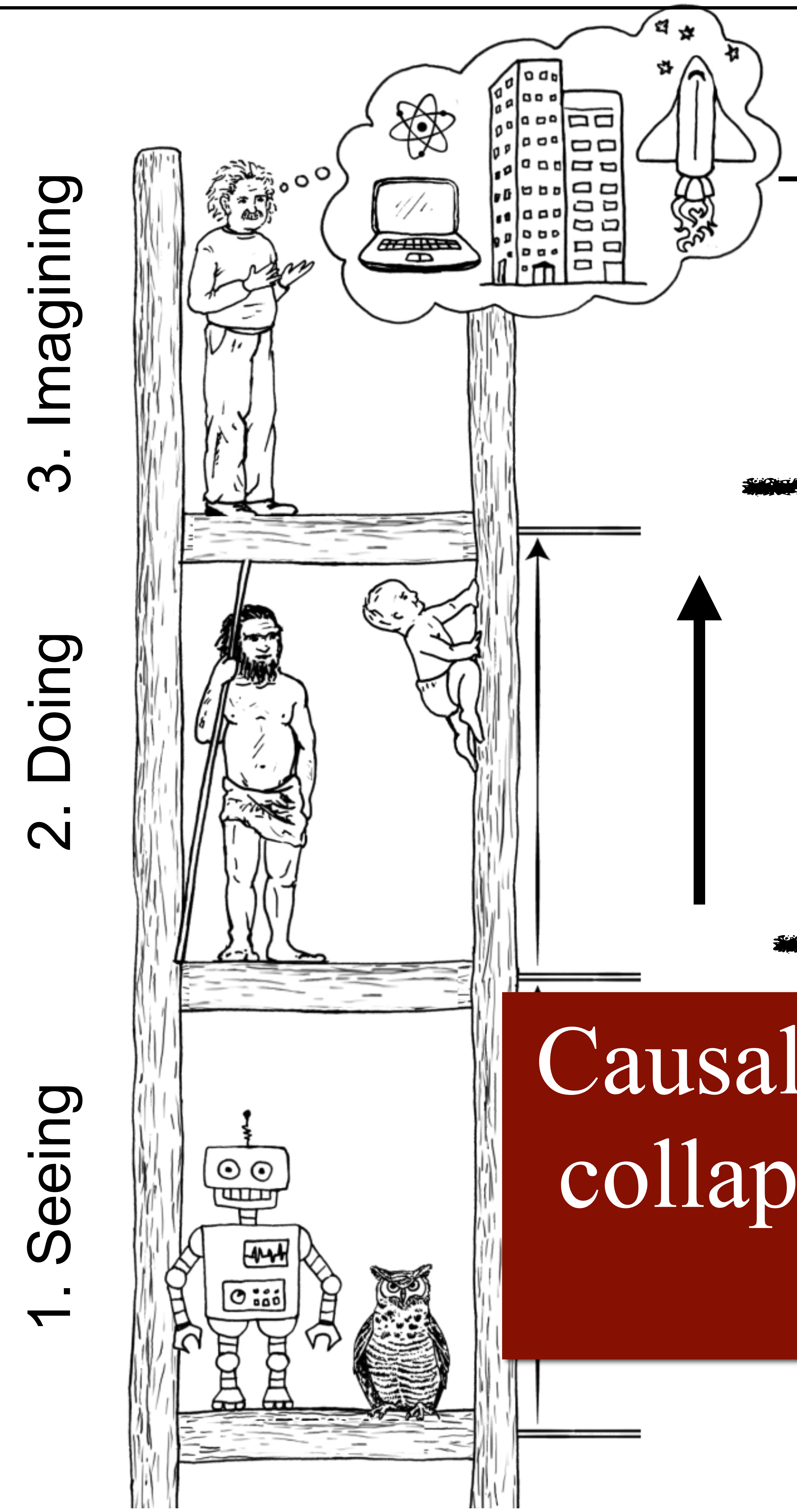
most of the inferences are about causal effects (policies, treatments, decisions)

Seeing

most of the available data is observational, passively collected

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

Ladder of Causation



Cross-layer inferences:

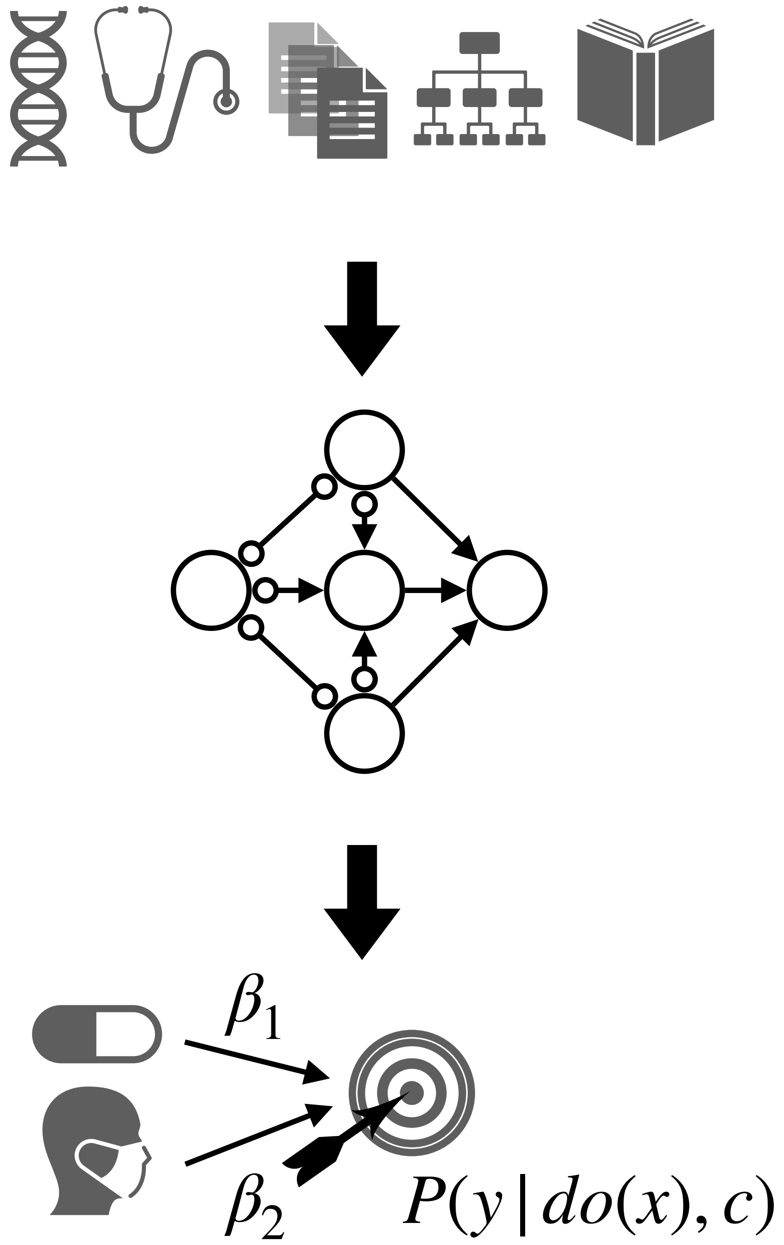
Doing

most of the inferences are about causal effects (policies, treatments, decisions)

Causal Hierarchy Theorem : The ladder almost never collapses. That is, for almost any SCM, the rungs of the ladder remain distinct.

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

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.

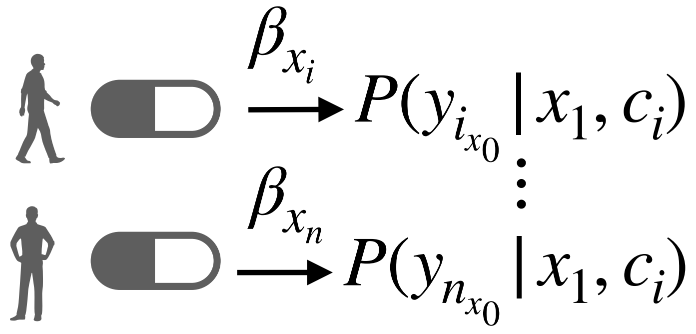
- **Causal Modeling, Causal Representation Learning and Causal Abstraction**

Explainability: Provides a better understanding of the true underlying mechanisms

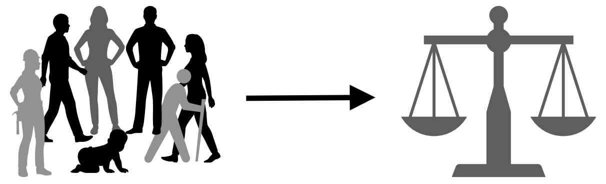
- **Causal Discovery**

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

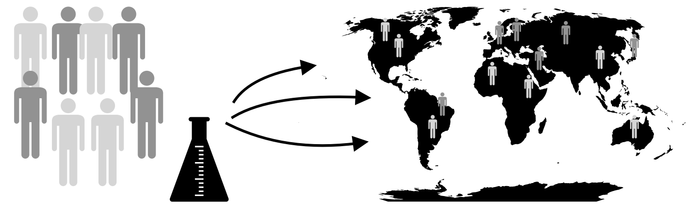
- **Causal Effect Identification and Estimation**



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)

The true model behind the data

Full explainability for all layers of the causal hierarchy

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.

Statistical Association vs Causation

Pre-Interventional/ Observational SCM

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

$do(X = x)$

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*** $do(X = x)$?

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

Structural Equation Model (SEM)

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

- **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 requires parametric and distributional assumptions.

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

SCM: Encoder of Functional Knowledge

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

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

Yes, with structural knowledge encoded as a causal diagram!

Causal Bayesian Network

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

Directed Acyclic Graph

Acyclic Directed Mixed Graph

CBN: Encoder of <u>Structural</u> <u>Causal</u> Knowledge

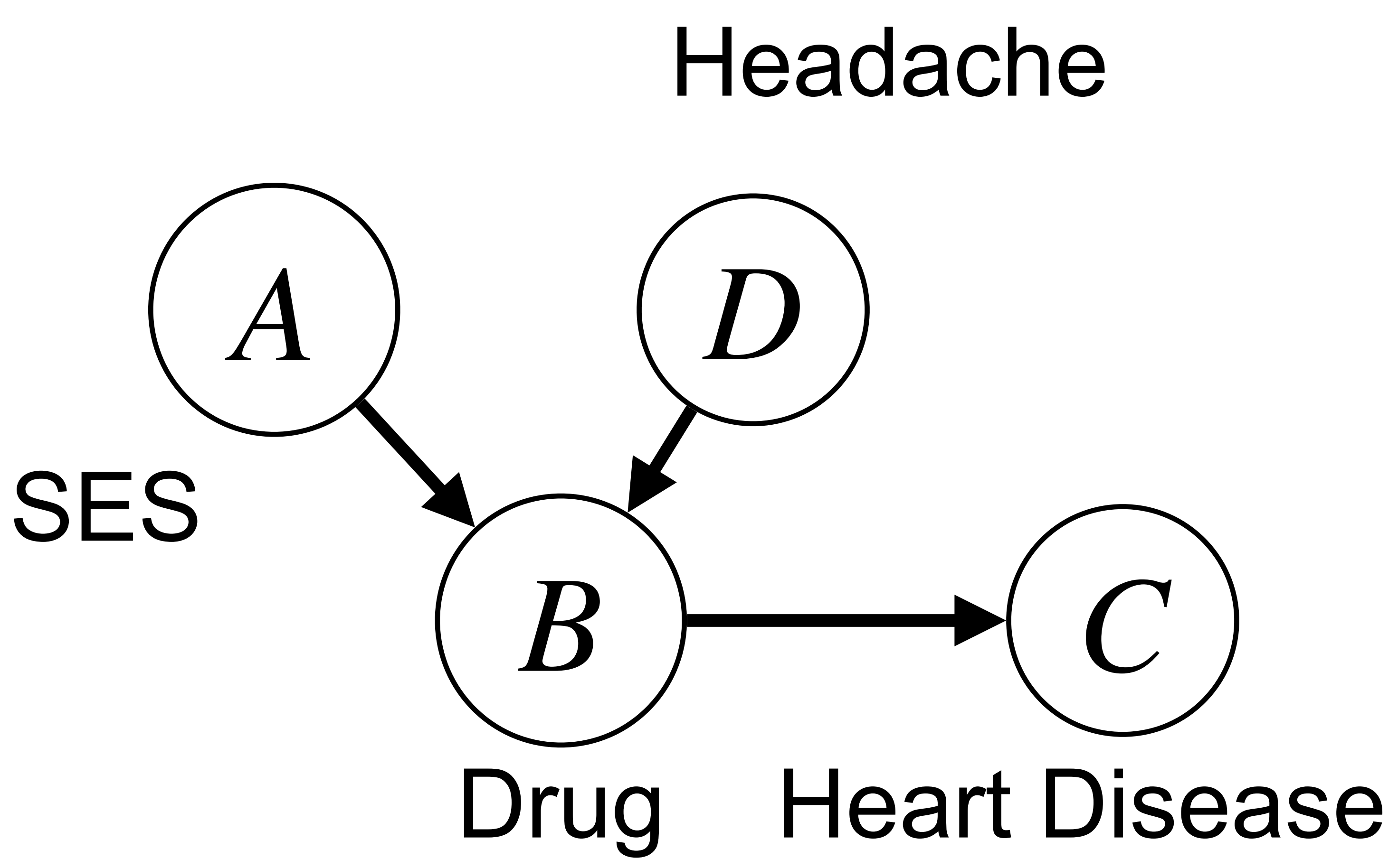
Structural Causal Model (SCM)

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

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

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

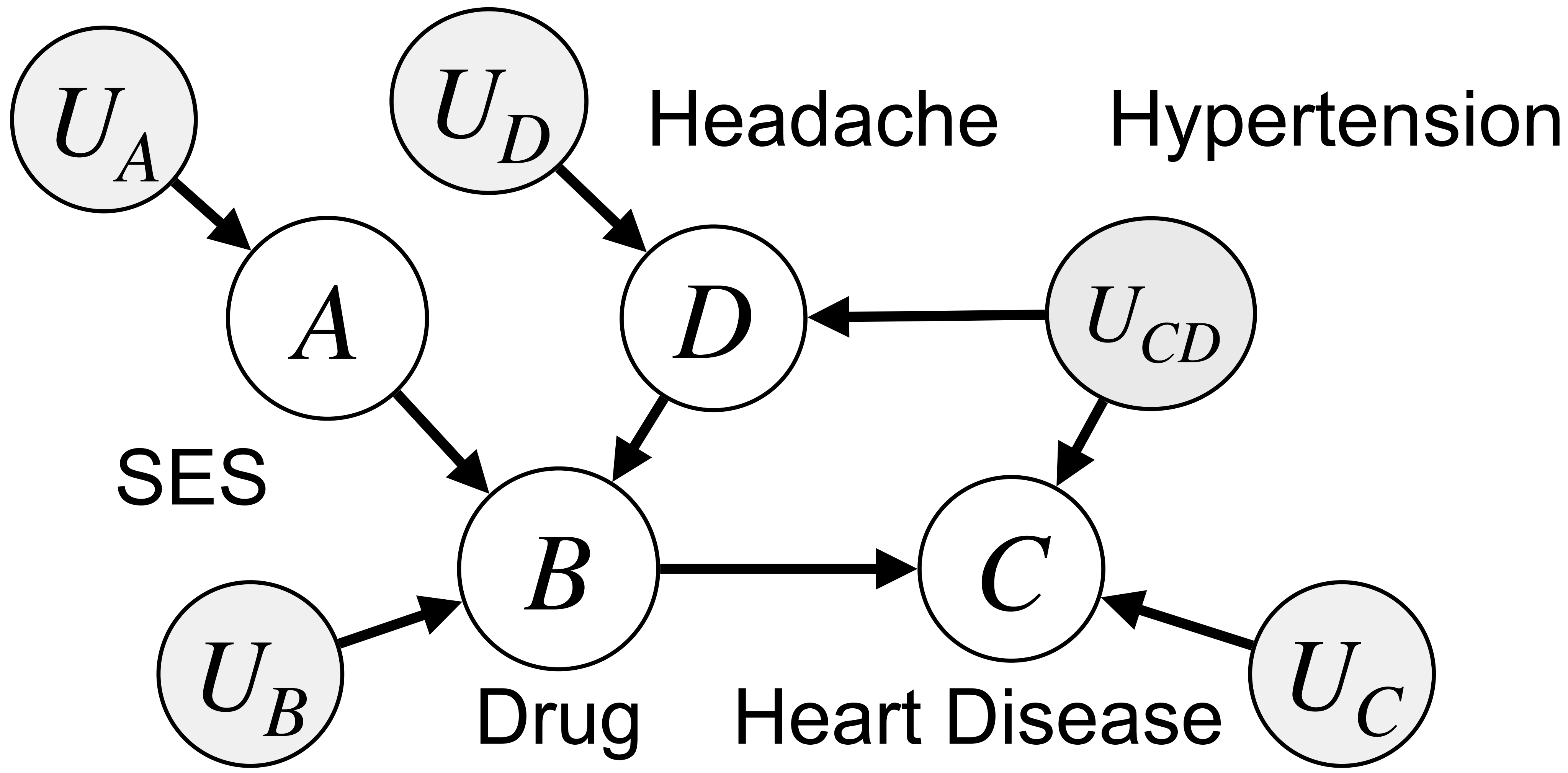
Structural Causal Model (SCM)

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

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

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 $f_j \in \mathcal{F}$.

CBN: Encoder of Structural Causal Knowledge

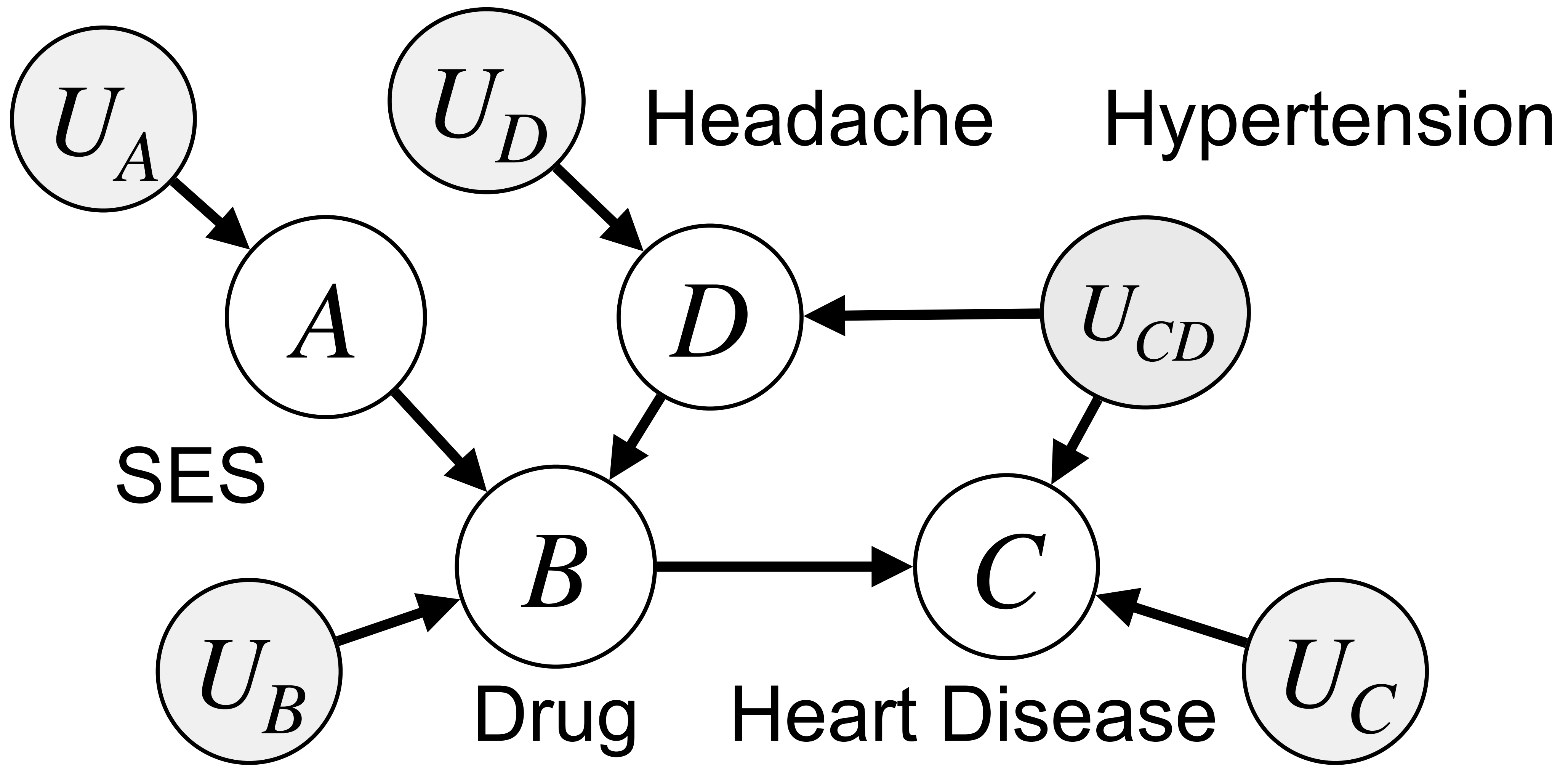
Structural Causal Model (SCM)

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

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

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 $U \in \mathbf{U}$.

CBN: Encoder of Structural Causal Knowledge

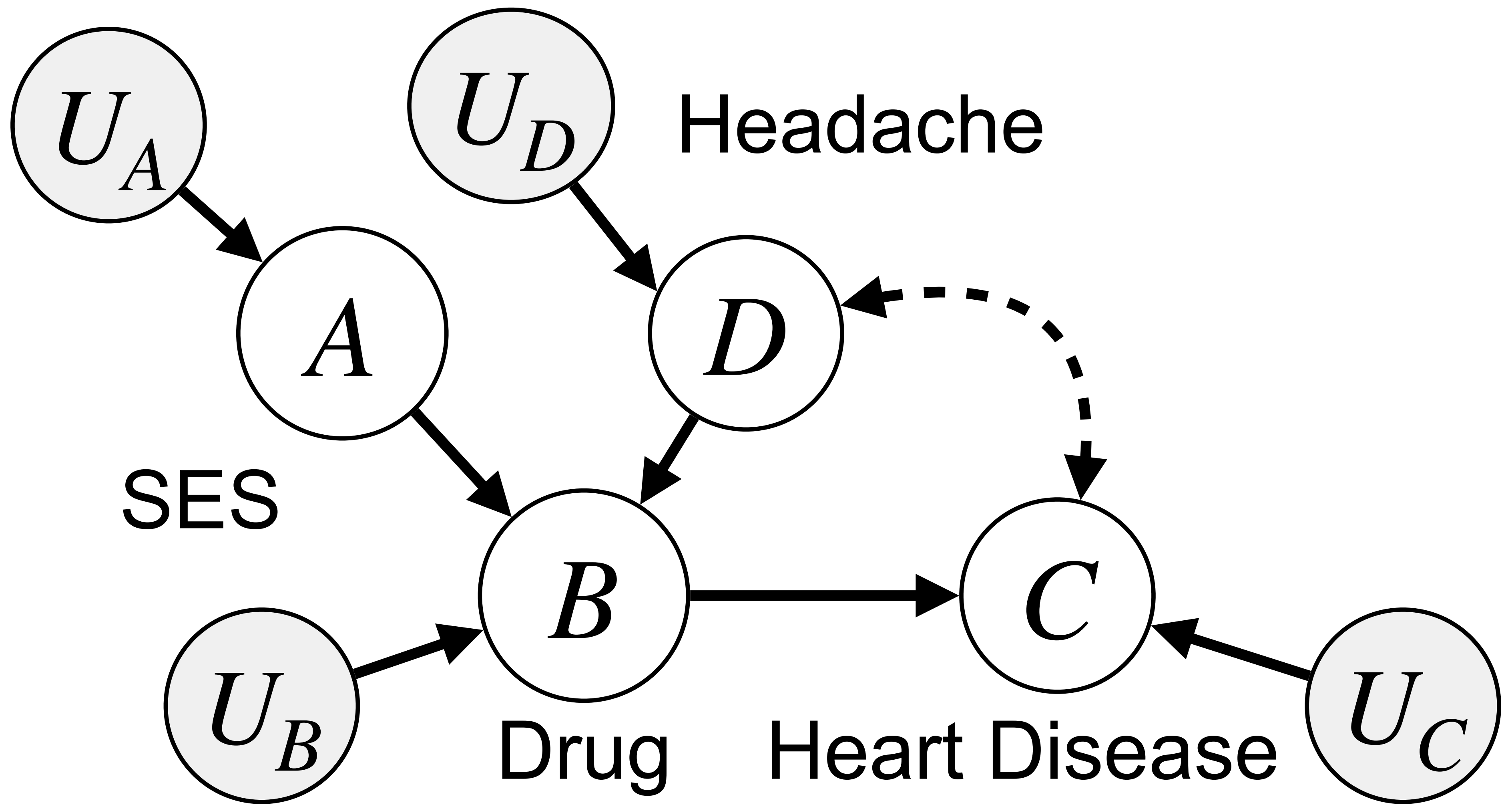
Structural Causal Model (SCM)

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

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

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

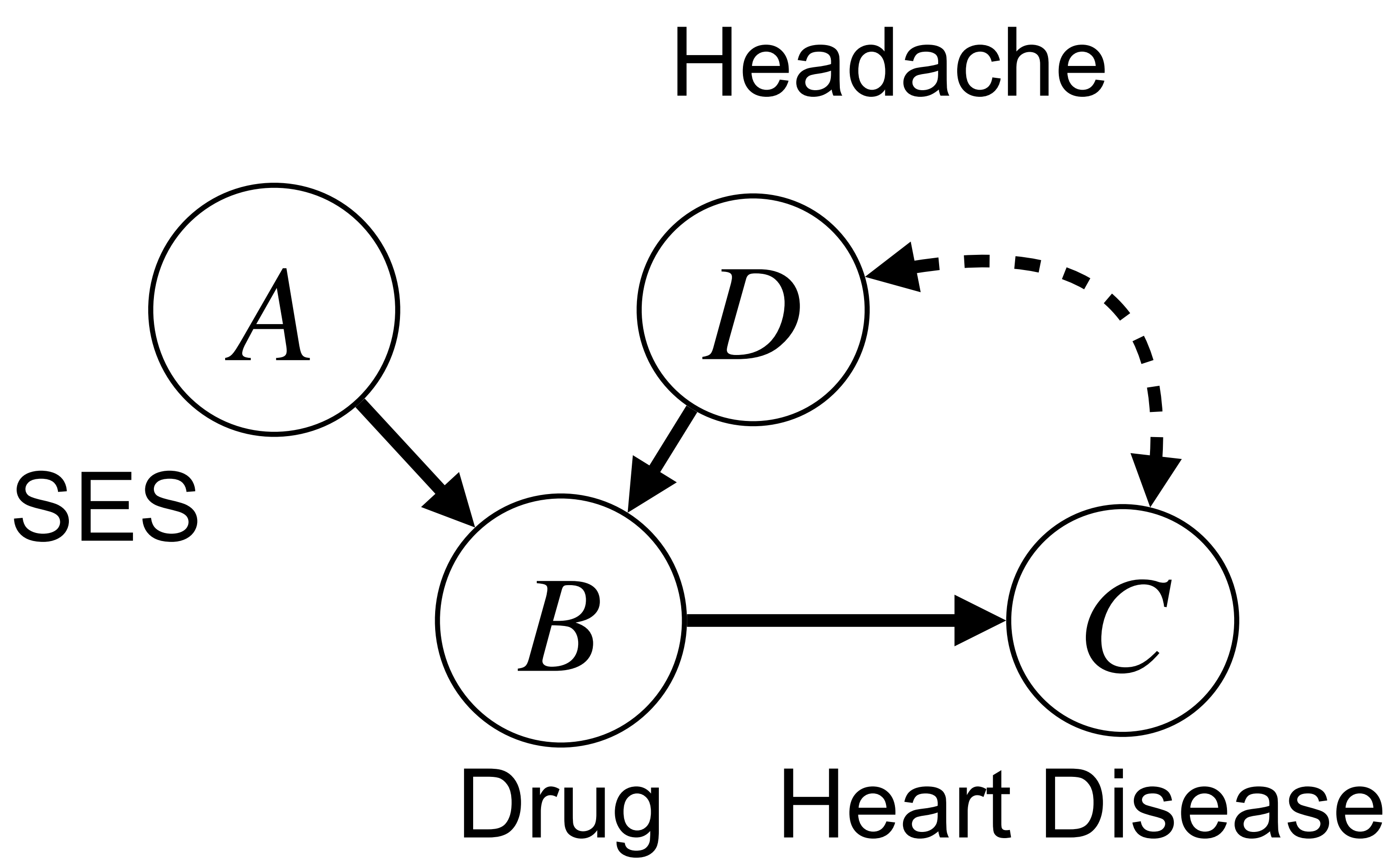
Structural Causal Model (SCM)

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

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

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 $U \in \mathbf{U}$.

Challenges in Causal Modeling

- **Dimensionality and complexity:**
  - Specifying a correct causal diagram in high-dimensional settings is often infeasible due to combinatorial explosion.
- **Limited and fragmented prior knowledge:**
  - Expert knowledge is often partial, inconsistent, and insufficient for full model specification.
- **Mismatch between observed and the true causal variables:**
  - Observed variables may be proxies, composites, or irrelevant to the underlying causal structure.

Need for **causal abstraction** and **causal representation learning** to represent meaningful, tractable causal variables and structures.

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

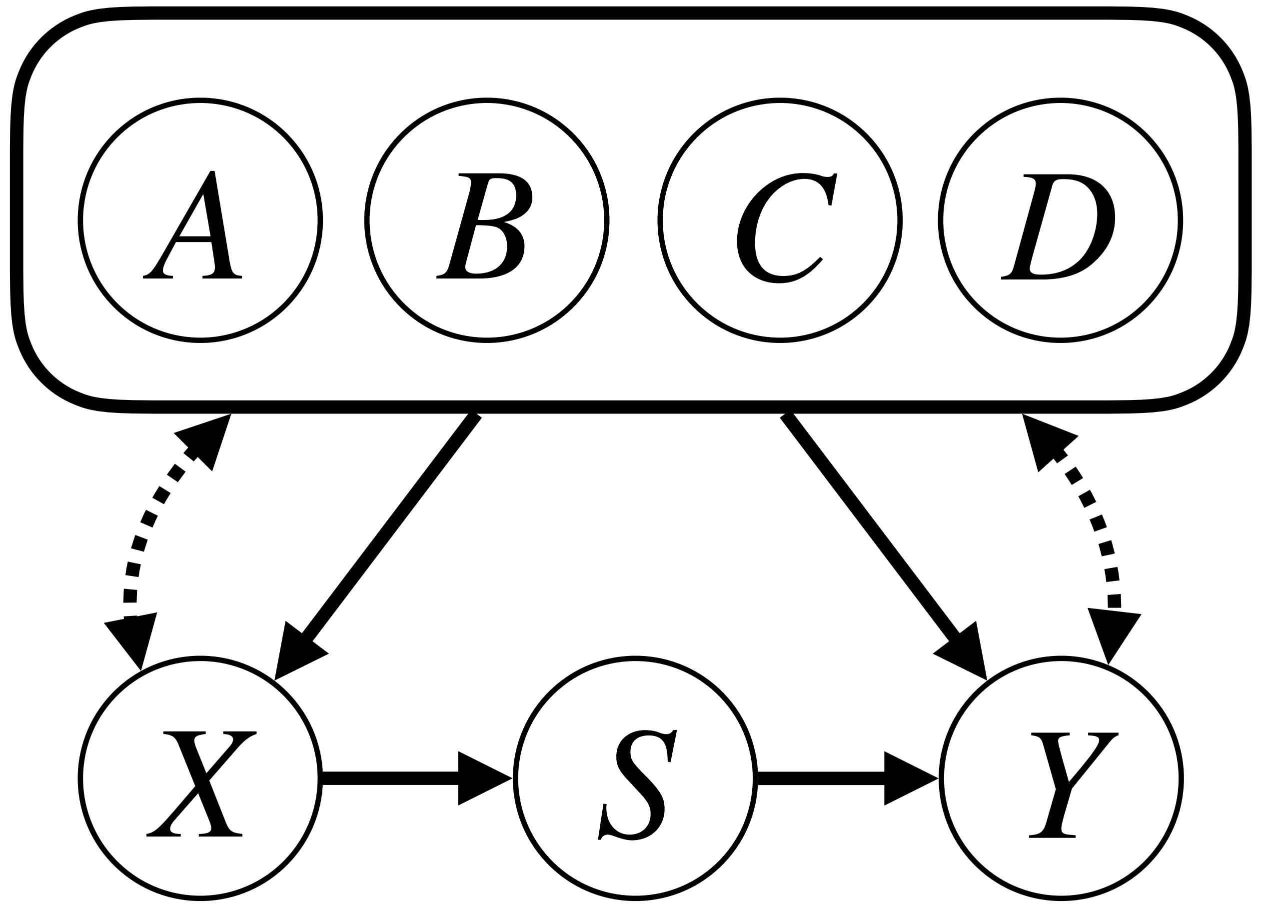
The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)

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

C-DAGs for Partially Understood Causal Systems

(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* (C-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 $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$**:

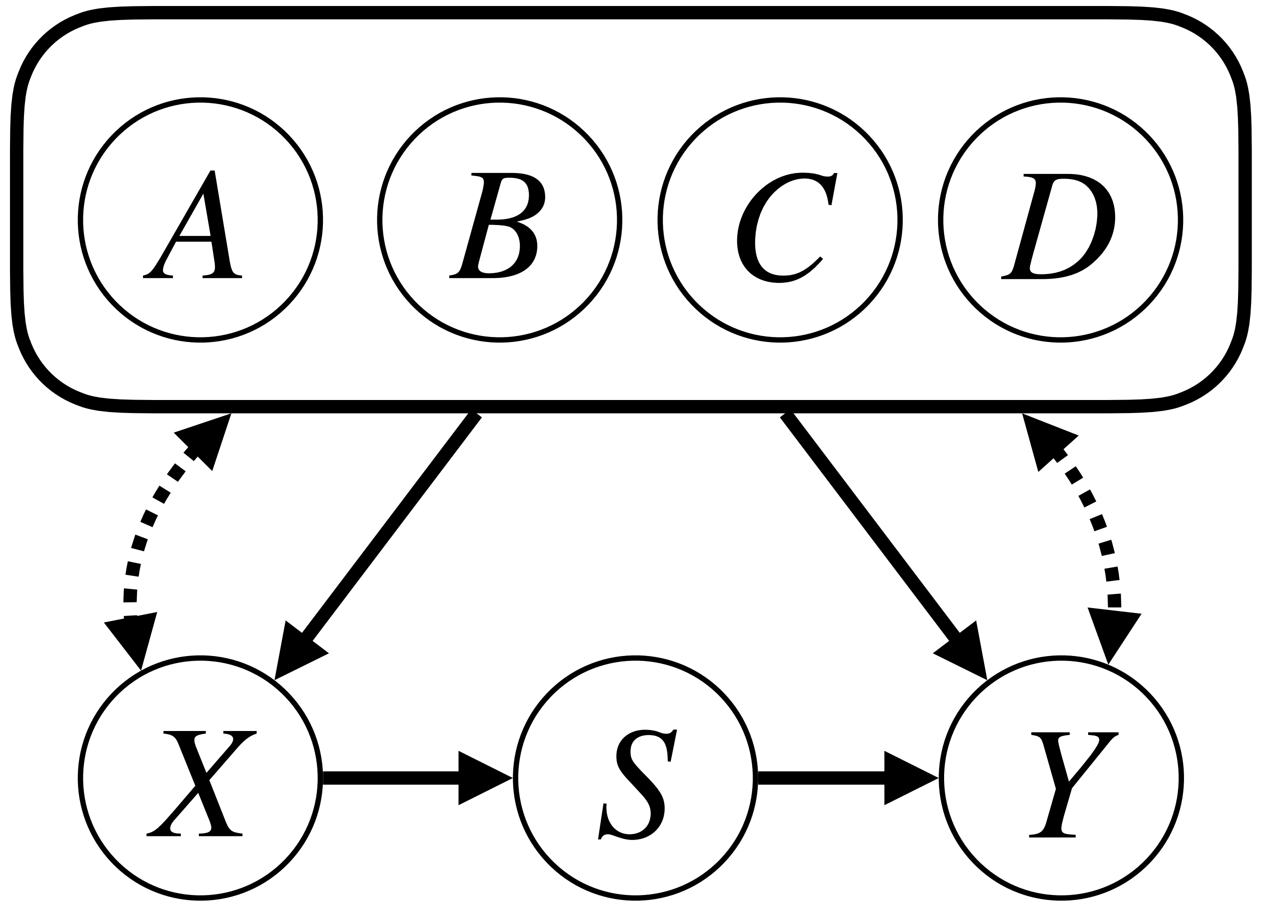
- $\mathbf{C}_i \rightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \rightarrow V_j$
- $\mathbf{C}_i \leftarrow\!\!\dashrightarrow \mathbf{C}_j$ if $\exists V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \leftarrow\!\!\dashrightarrow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

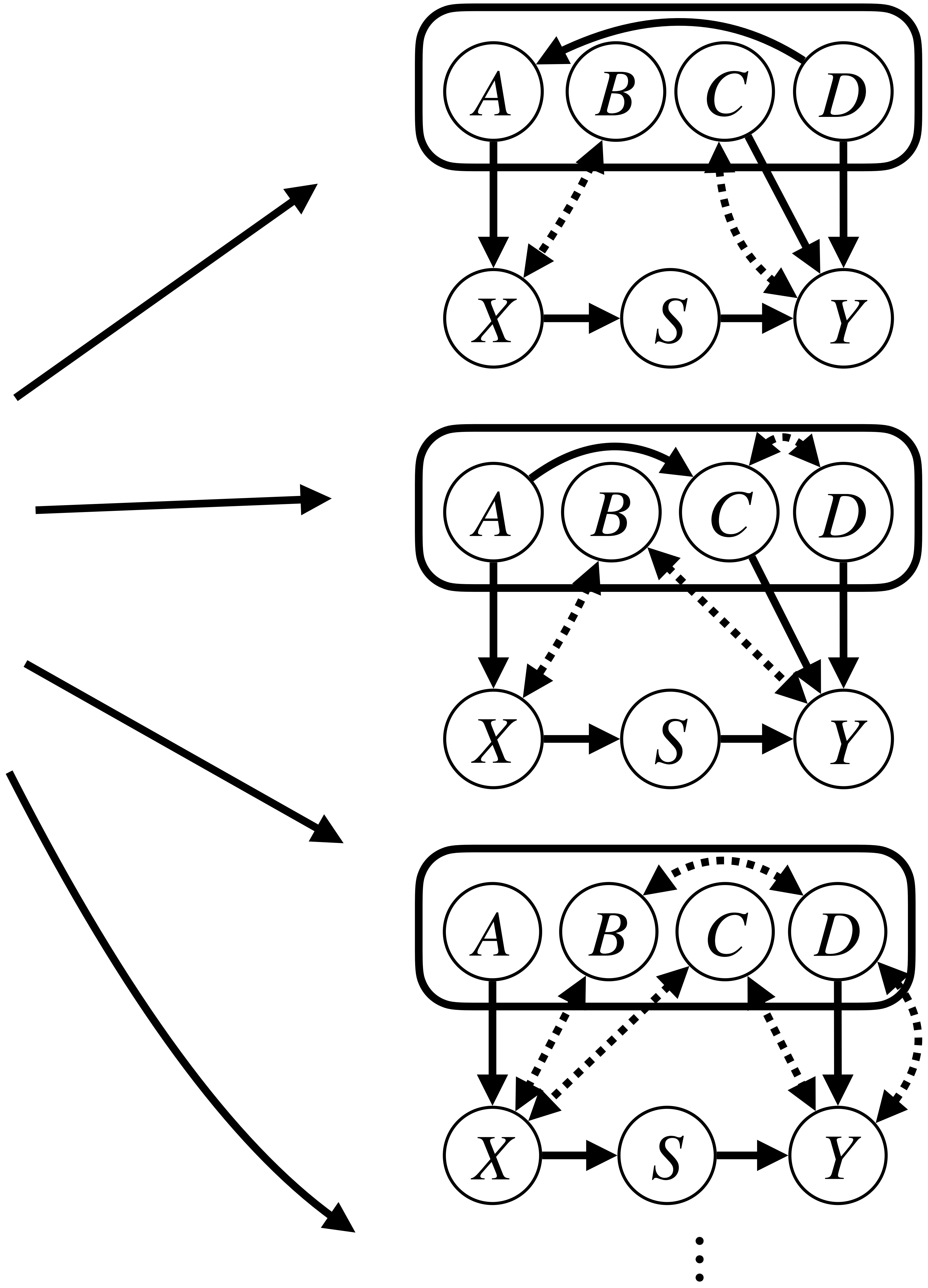
and $G_{\mathbf{C}}$ contains no cycles.

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.

C-DAGs for Partially Understood Causal Systems

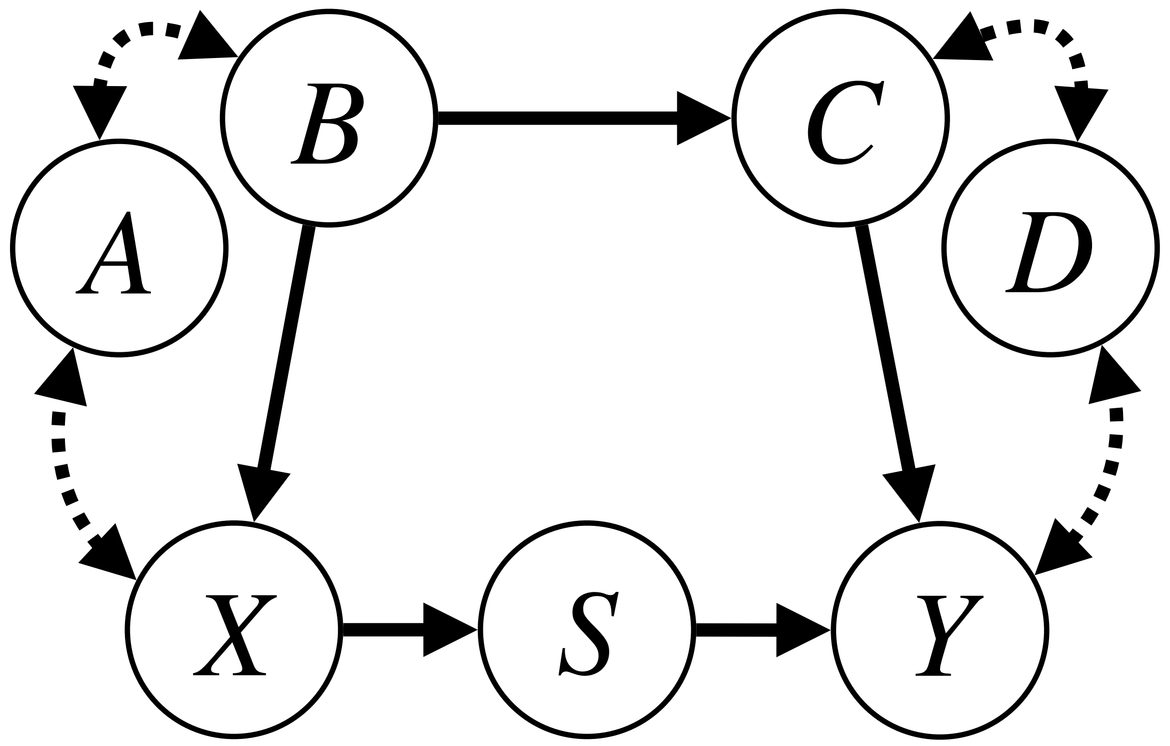
Many causal diagrams are compatible with the current knowledge!



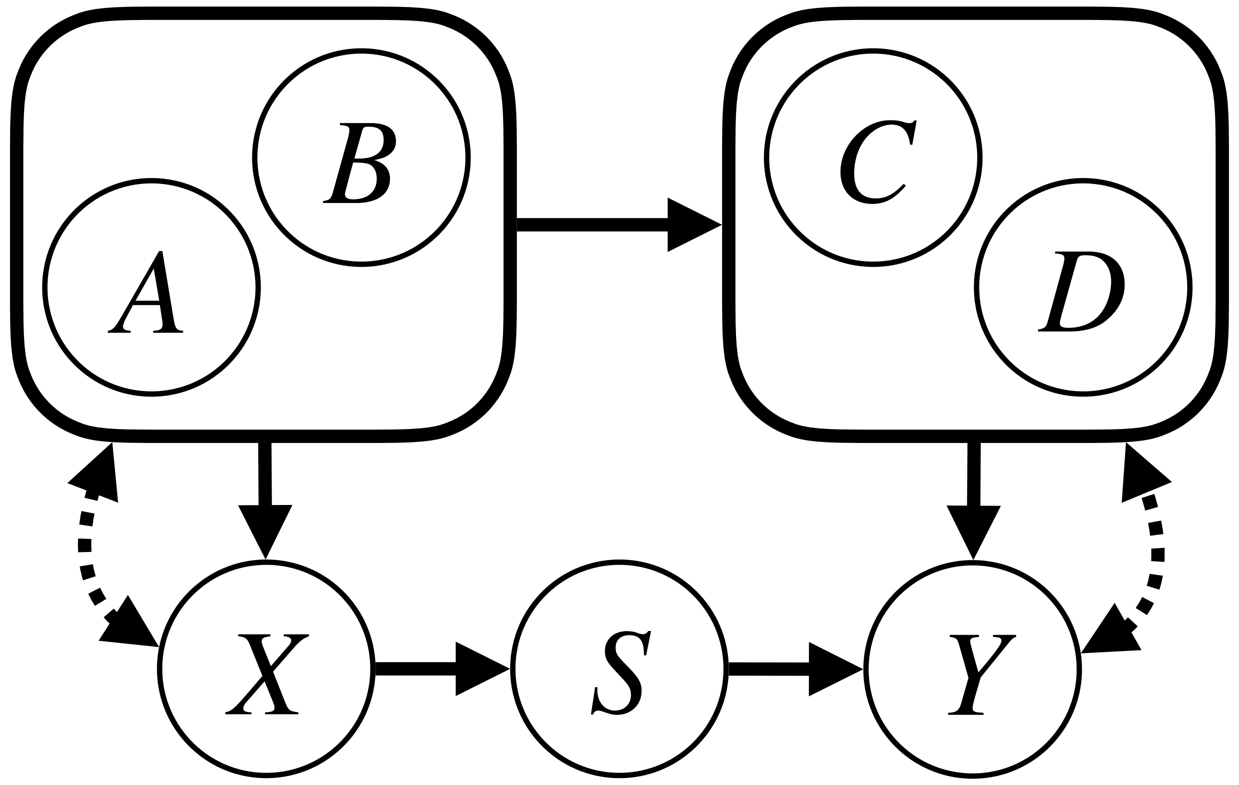


Can be seen as an *equivalence class* of causal diagrams, where any relationships are allowed among the variables within each cluster.

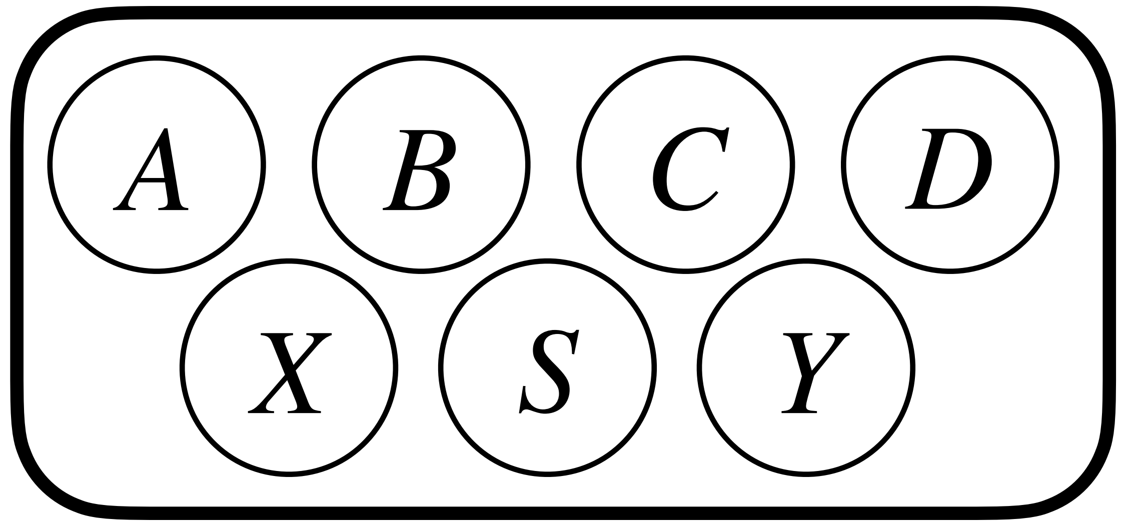
C-DAG: Flexible Encoder of Model Assumptions



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.

In a C-DAG, clusters represent macro-variables of an abstracted, coarser SCM.

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

Bayesian Network

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

Bayesian Networks & Markov Condition

A DAG G over $\mathbf{V}$ is a *Bayesian Network* for a joint probability distribution $P(\mathbf{V})$ if, for every $V_i \in \mathbf{V}$, it holds that $V_i \perp\!\!\!\perp NDesc_i \mid Pa_i$ and, therefore, $P(\mathbf{v})$ factorizes as follows:

P is satisfies the
Markov Condition
w.r.t. G

$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid v_{i-1}, \ldots, v_1)$$

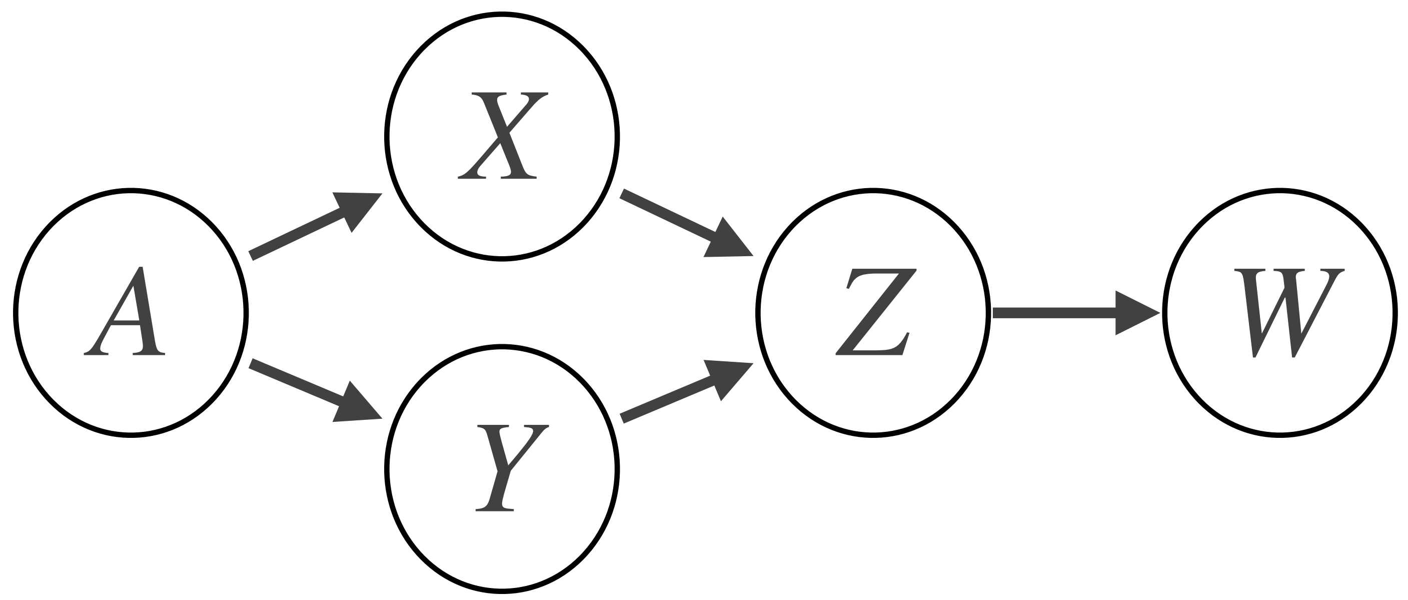
$$= \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i)$$

Chain Rule:

It holds for any topological order of G

$V_i \perp\!\!\!\perp NDesc_i \mid Pa_i, U_i$

Edges have no causal semantics!



$$P(\mathbf{v}) = P(w \mid z, x, y, a)\ P(z \mid x, y, a)\ P(x \mid y, a)\ P(y \mid a)\ P(a)$$
$$= P(w \mid z)\ P(z \mid x, y)\ P(x \mid a)\ P(y \mid a)\ P(a)$$

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

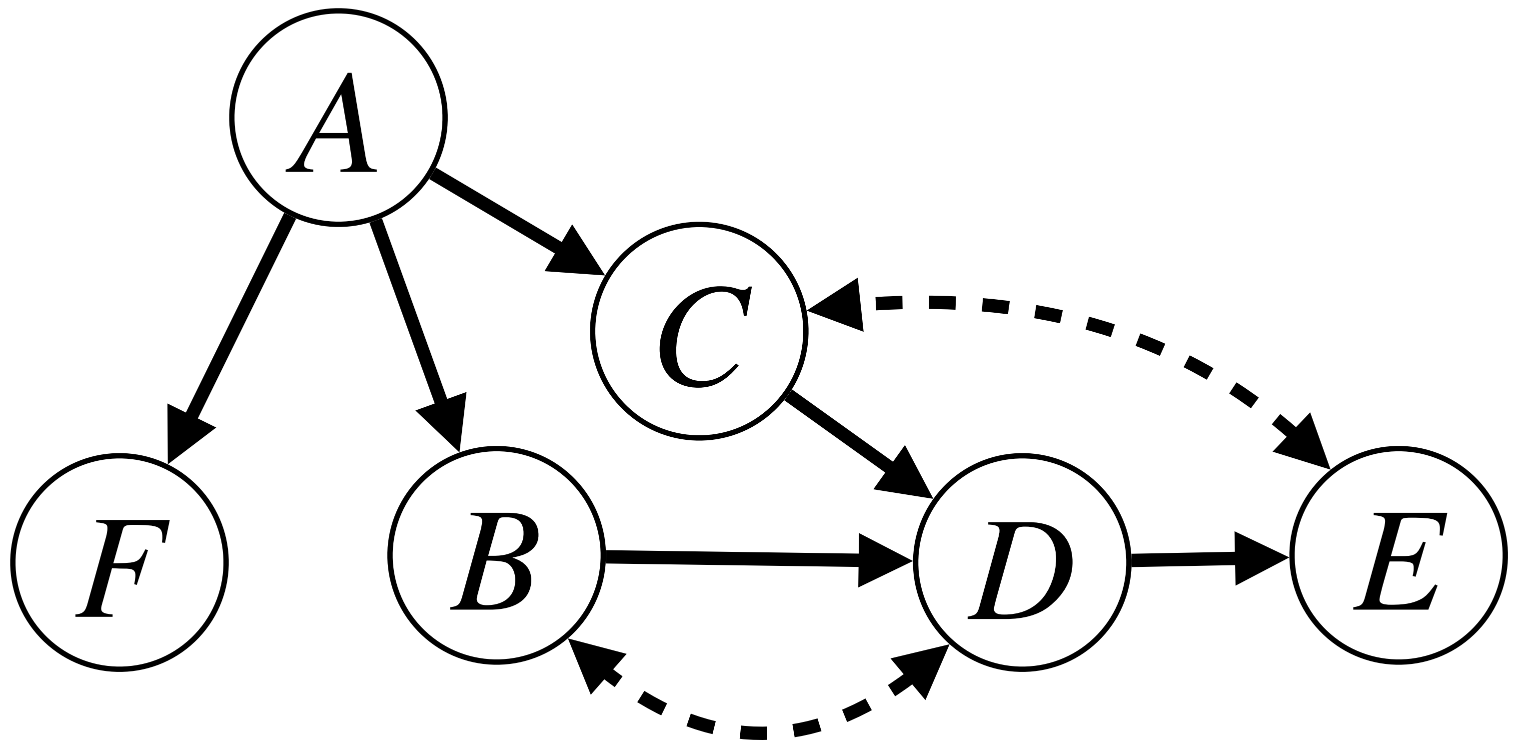
Bayesian Networks & Semi-Markov Condition

An ADMG G over $\mathbf{V}$ is a *Bayesian Network* for a joint probability distribution $P(\mathbf{V})$ if, for every $V_i \in \mathbf{V}$, it holds that $V_i \perp\!\!\!\perp NDesc_i \mid Pa_i^+$ and, therefore, $P(\mathbf{v})$ factorizes as follows:

P is satisfies the **Semi-*Markov* Condition** *w.r.t.* G

$$P(\mathbf{v}) = \prod_{V_i \in \mathbf{V}} P(v_i \mid pa_i^+).$$

The extended parents of V_i is defined as $Pa_i^+ = Pa^1(\{V \in \mathbf{C}(V_i) : V \leq V_i\}) \setminus \{V_i\}$, where $Pa^1(V) = Pa(V) \cup V$ and $\mathbf{C}(V_i)$ is a maximal path entirely made of bidirected edges.



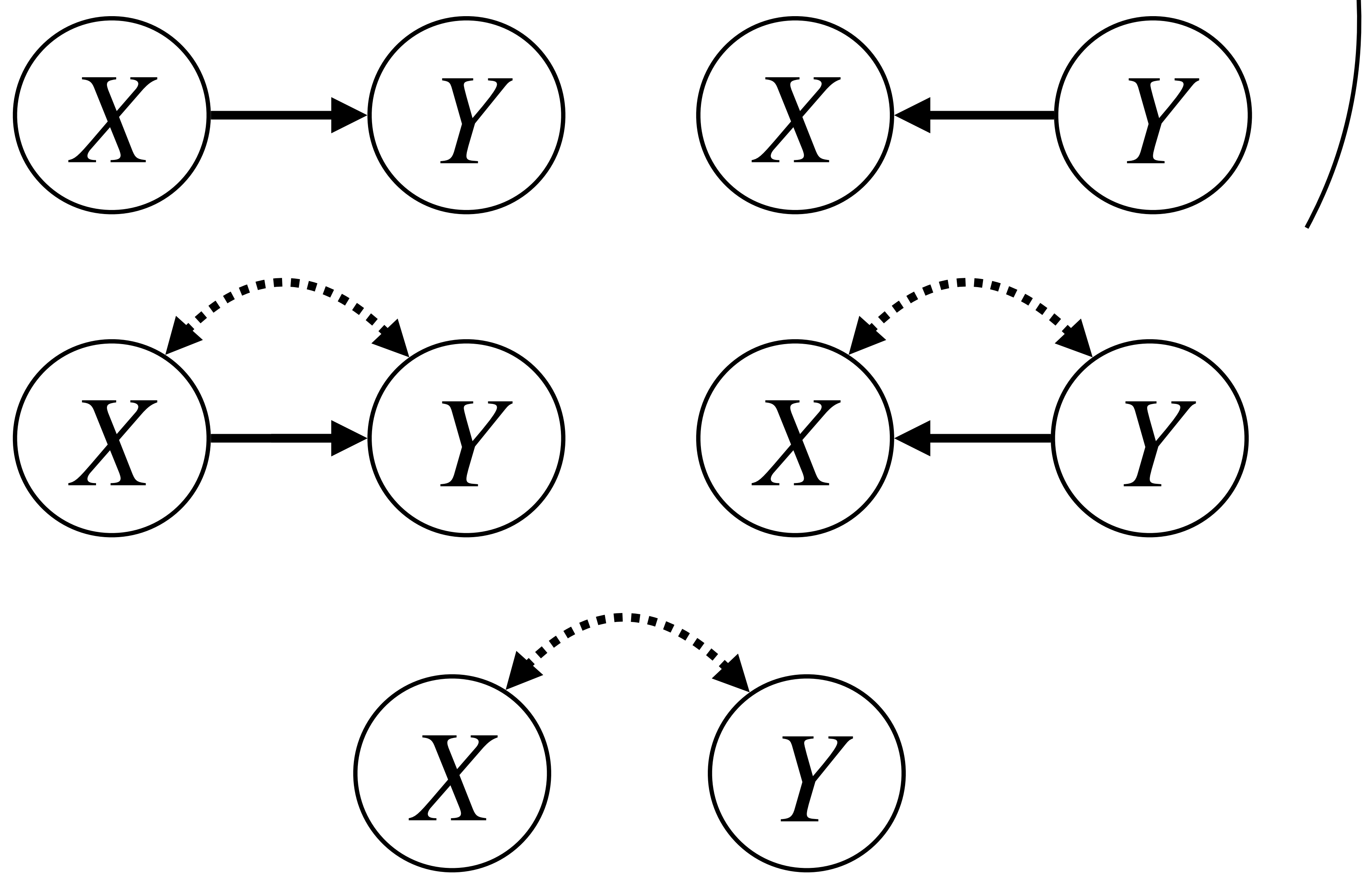
$$P(\mathbf{v}) = P(e \mid d, c, b, a, f)\ P(d \mid c, b, a, f)\ P(c \mid b, a, f)\ P(b \mid a, f)\ P(f \mid a)\ P(a)$$
$$= P(e \mid d, c, a)\ P(d \mid c, b, a)\ P(c \mid a)\ P(b \mid a)\ P(f \mid a)\ P(a)$$

$E \perp\!\!\!\perp F, B \mid D, C, A$ $\quad$ $D \perp\!\!\!\perp F \mid B, C, A$ $\quad$ $C \perp\!\!\!\perp F, B \mid A$ $\quad$ $B \perp\!\!\!\perp F \mid A$

Markov Equivalence Class

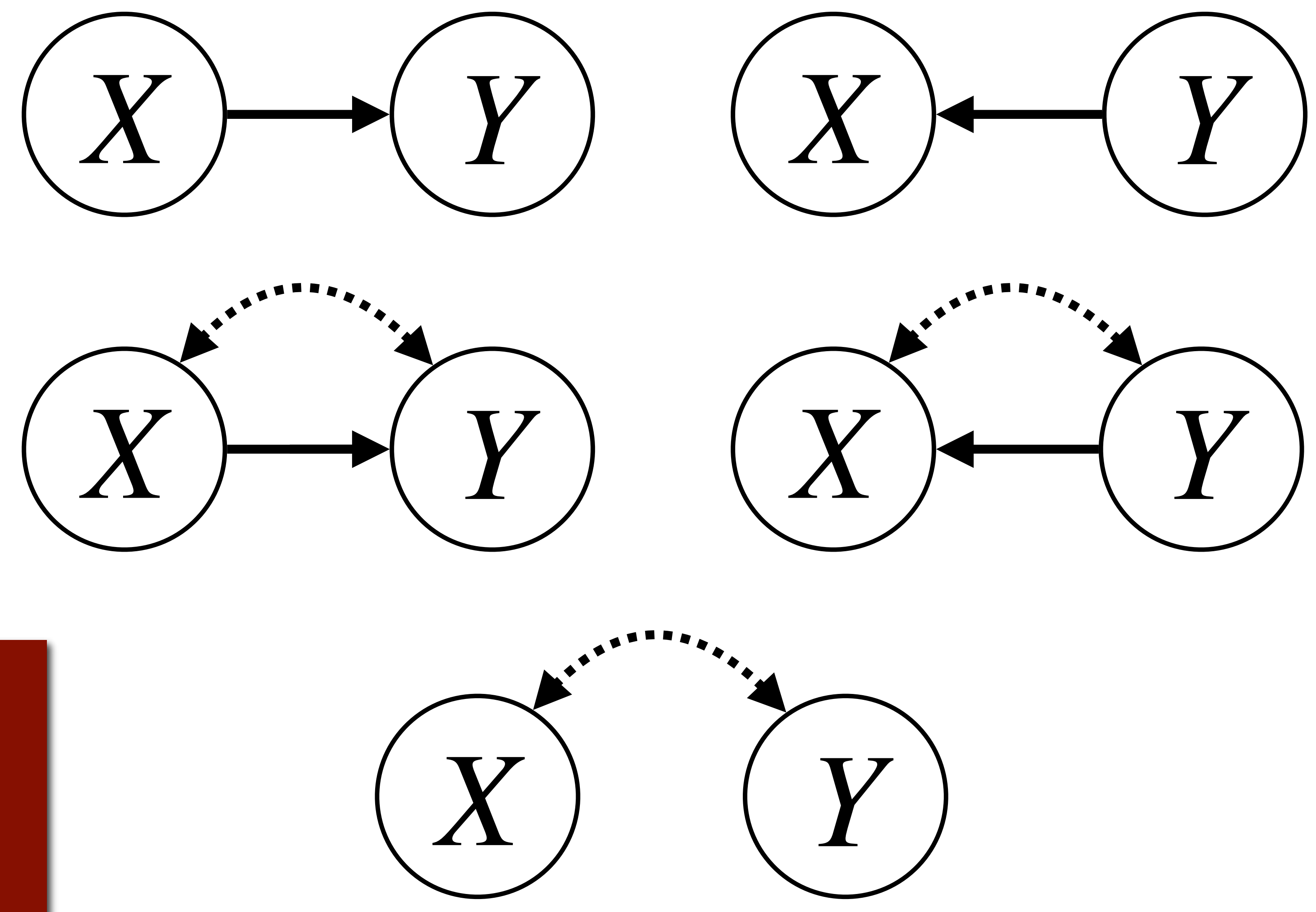
Definition (Markov Equivalence Class, MEC for short): A Markov Equivalence Class is a set of models that encode the same set of conditional independencies.

Markov Equivalent BNs

Distribution	Factorization	Bayesian Networks
$P(X, Y)$ with $P(Y \mid X) \neq P(Y)$ i.e., $X \not\perp\!\!\!\perp Y$	$P(x, y) = P(y \mid x)P(x)$ $P(x, y) = P(x \mid y)P(y)$	

Markov Equivalence Class

Definition (Markov Equivalence Class, MEC for short): A Markov Equivalence Class is a set of models that encode the same set of conditional independencies.

Distribution	Factorization	Bayesian Networks
$P(X, Y)$ with $P(Y \mid X) \neq P(Y)$ i.e., $X \not\perp\!\!\!\perp Y$	$P(x, y) = P(y \mid x)P(x)$ $P(x, y) = P(x \mid y)P(y)$	

Markov Equivalent BNs

All models imply no independence and no other invariance

Markov Equivalence Class

Distribution	Factorization	Bayesian Networks

Distribution

$$P(X, Y, Z)$$

$$\text{with } P(Y \mid X, Z) = P(Y \mid X)$$

$$\text{i.e., } X \perp\!\!\!\perp Y \mid Z$$

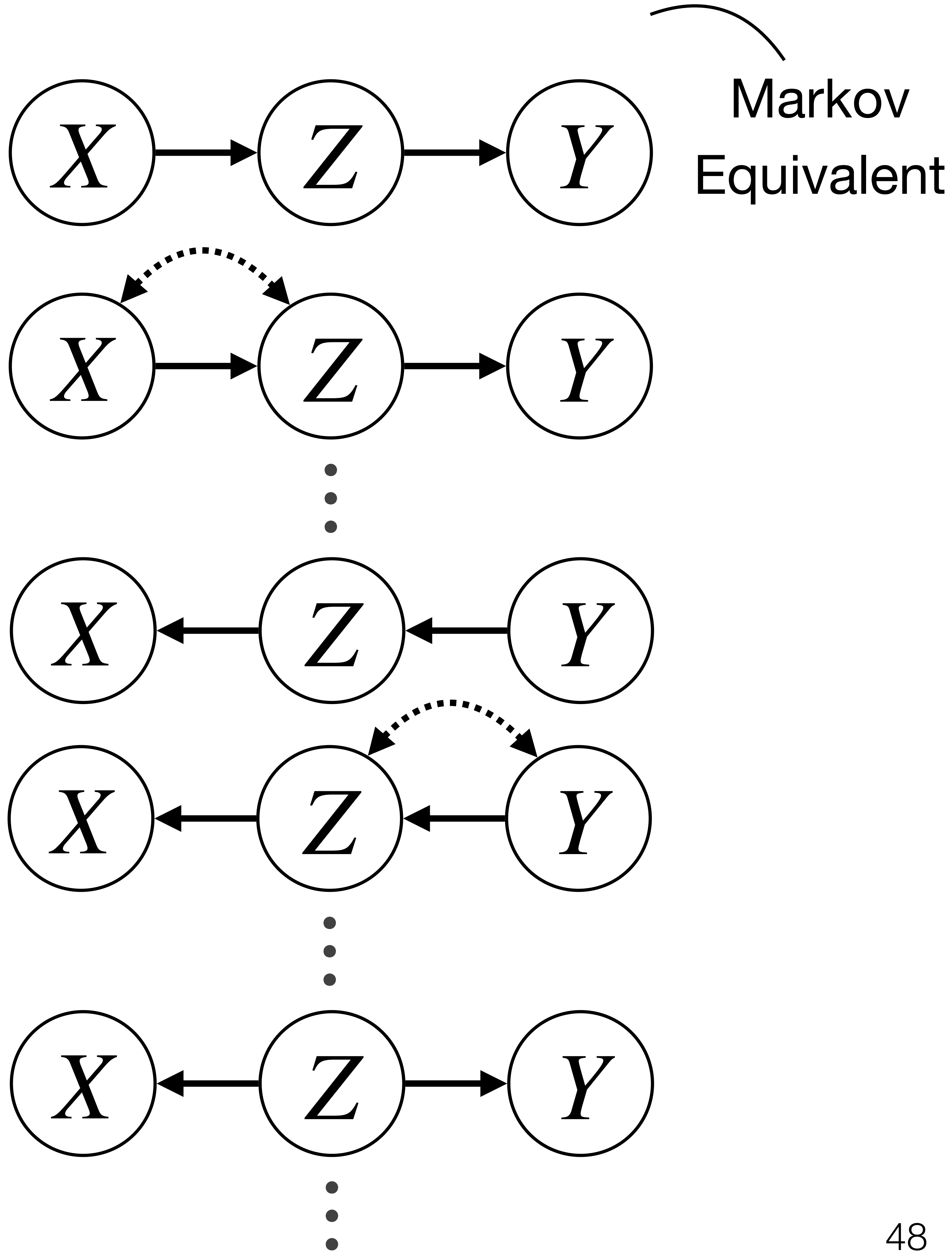
Factorization

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

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

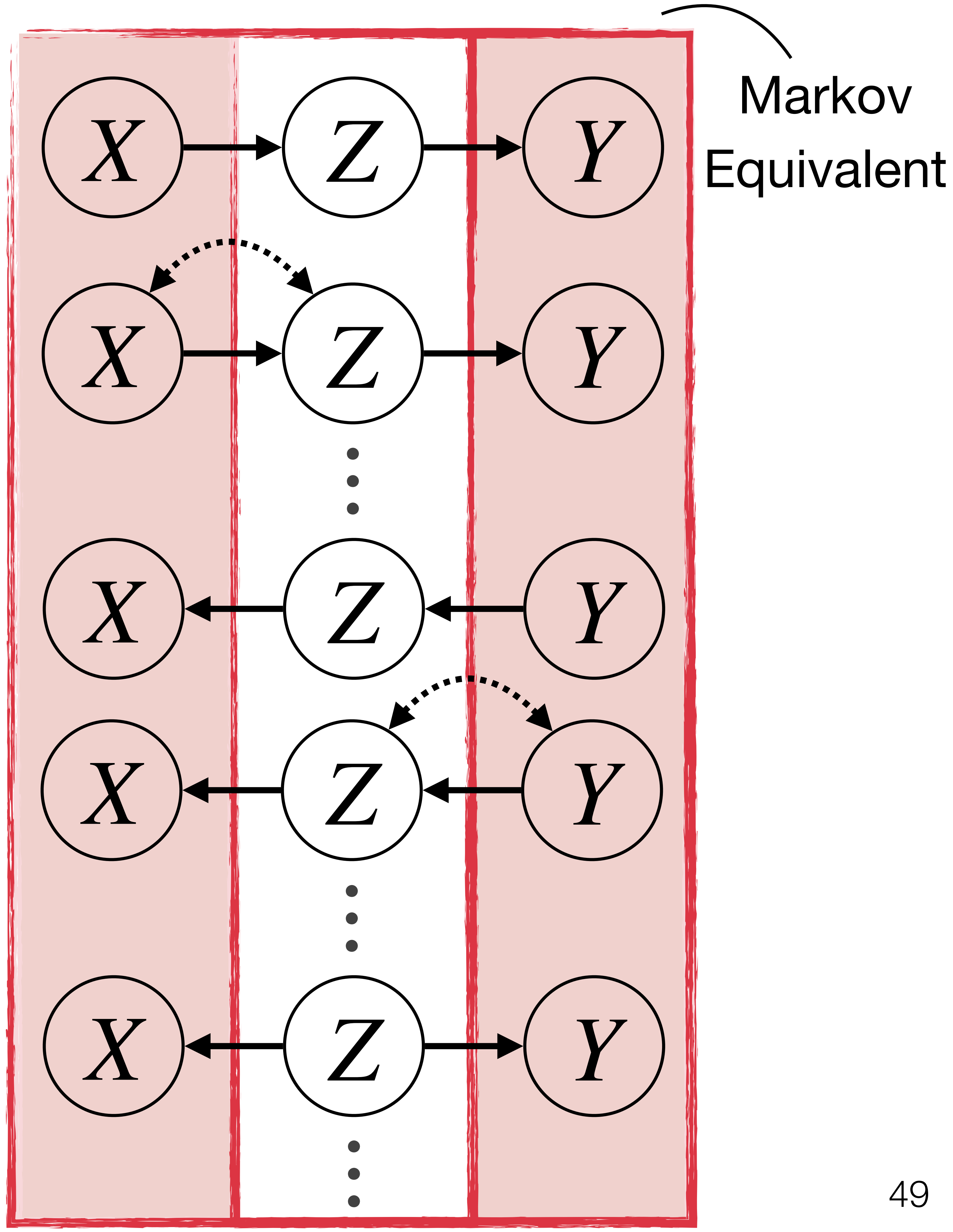
$$P(x, y, z) = {\color{red}P(y \mid x, z)}P(x \mid z)P(z) = {\color{red}P(y \mid z)}P(x \mid z)P(z)$$

Bayesian Networks



Markov Equivalence Class

Distribution	Factorization	Bayesian Networks
$P(X, Y, Z)$ with $P(Y \mid X, Z) = P(Y \mid X)$ i.e., $X \perp\!\!\!\perp Y \mid Z$	$P(x, y, z) = P(y \mid x, z) P(z \mid x) P(x)$ $= P(y \mid z) P(z \mid x) P(x)$	
	$P(x, y, z) = P(x \mid y, z) P(y \mid z) P(z)$ $= P(x \mid z) P(z \mid y) P(y)$	
	$= P(y \mid z) P(x \mid z) P(z)$	



All models imply *only* $X \perp\!\!\!\perp Y \mid Z$ and
Z is always a *non-collider* in such models.

Markov Equivalence Class

Distribution

$P(X, Y, Z)$

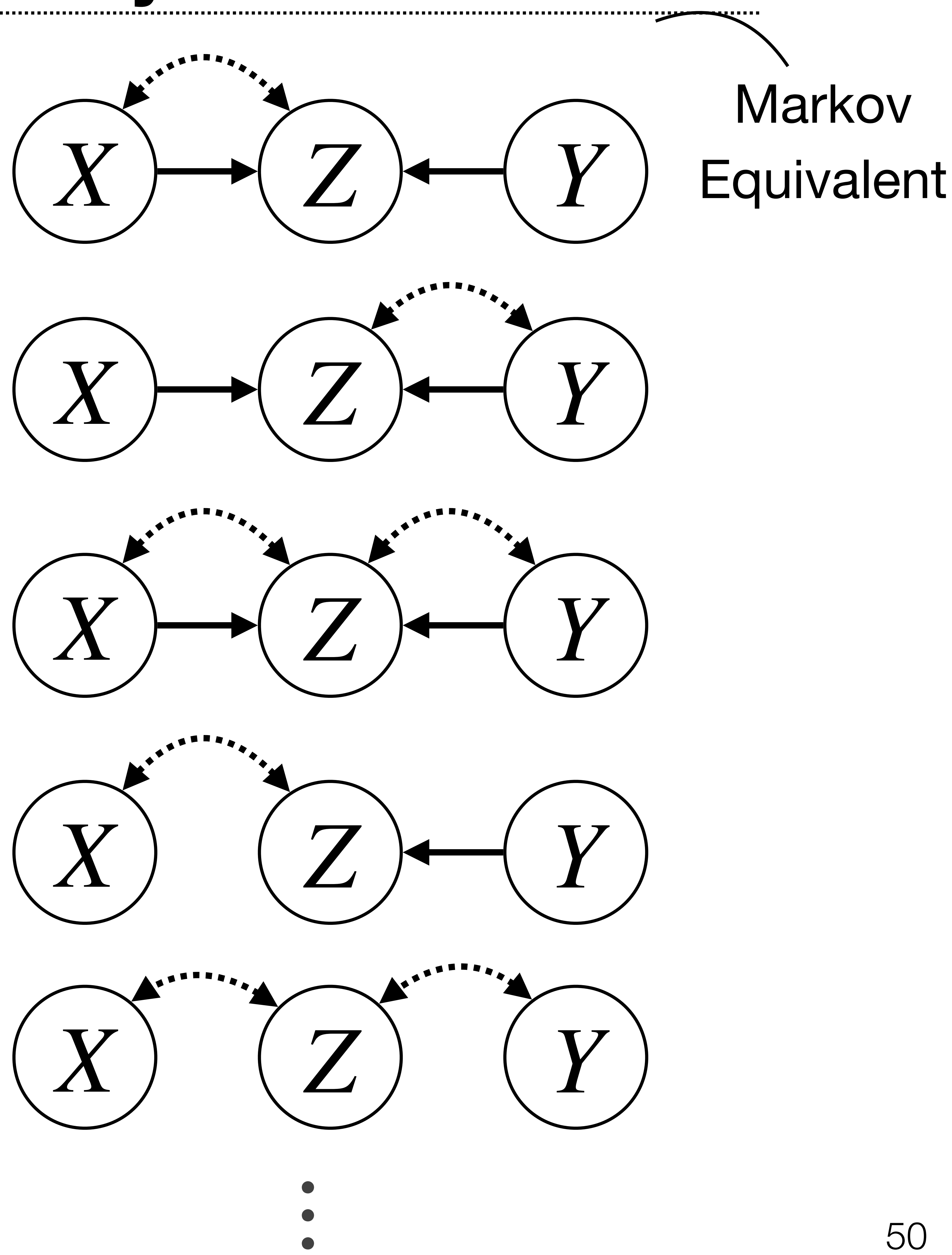
with $P(Y|X) = P(Y)$

i.e., $X \perp\!\!\!\perp Y$

Factorization

$$P(x, y, z) = P(z|x, y)P(x|y)P(y)$$
$$= P(z|x, y)P(x)P(y)$$

Bayesian Networks



Markov Equivalence Class

Distribution

$P(X, Y, Z)$

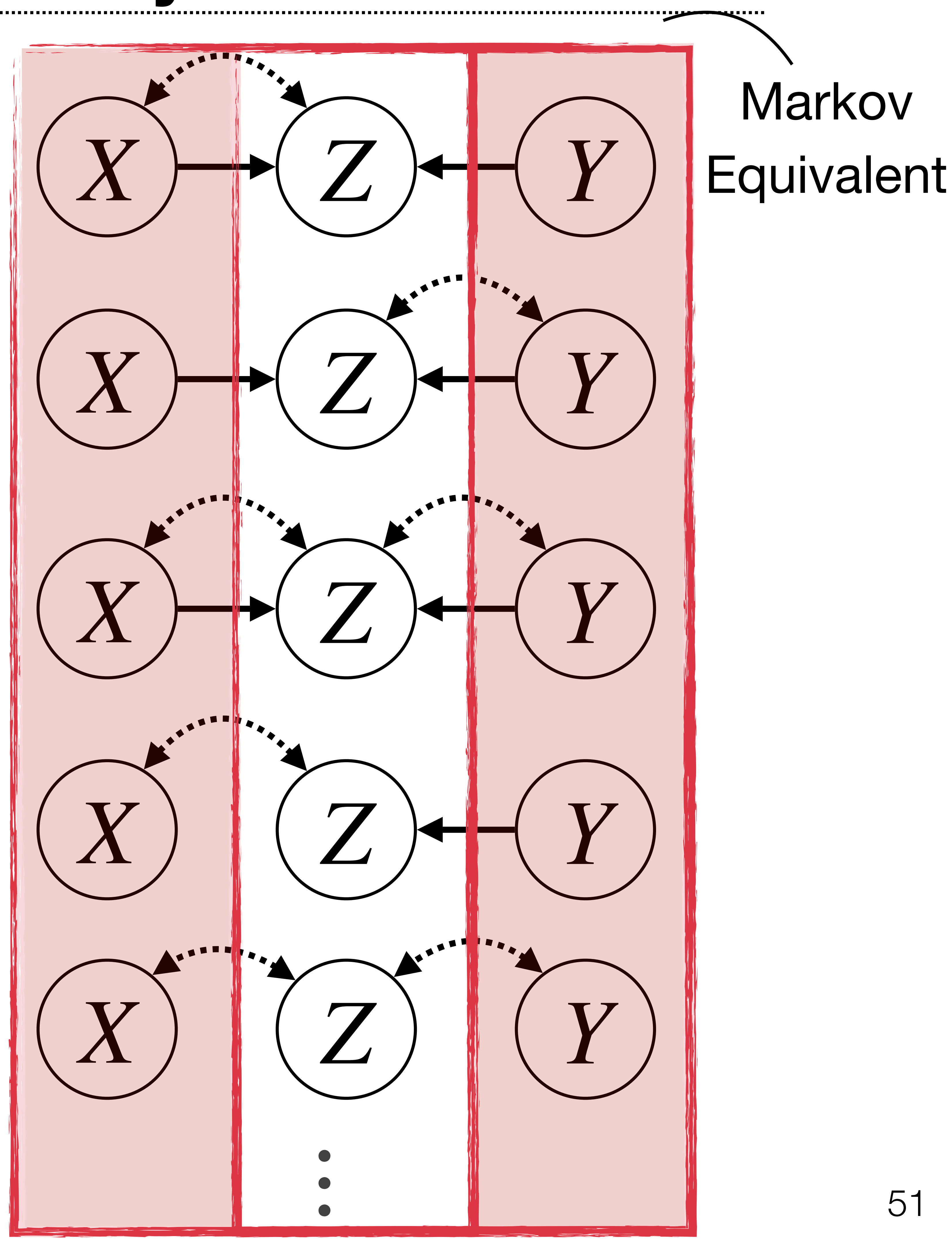
with $P(Y|X) = P(Y)$

i.e., $X \perp\!\!\!\perp Y$

Factorization

$$P(x, y, z) = P(z|x, y)P(x|y)P(y) = P(z|x, y)P(x)P(y)$$

Bayesian Networks



All models imply *only* $X \perp\!\!\!\perp Y$ and
Z is always a *collider* in such models,

Note: Z is never an ancestor of X or Y

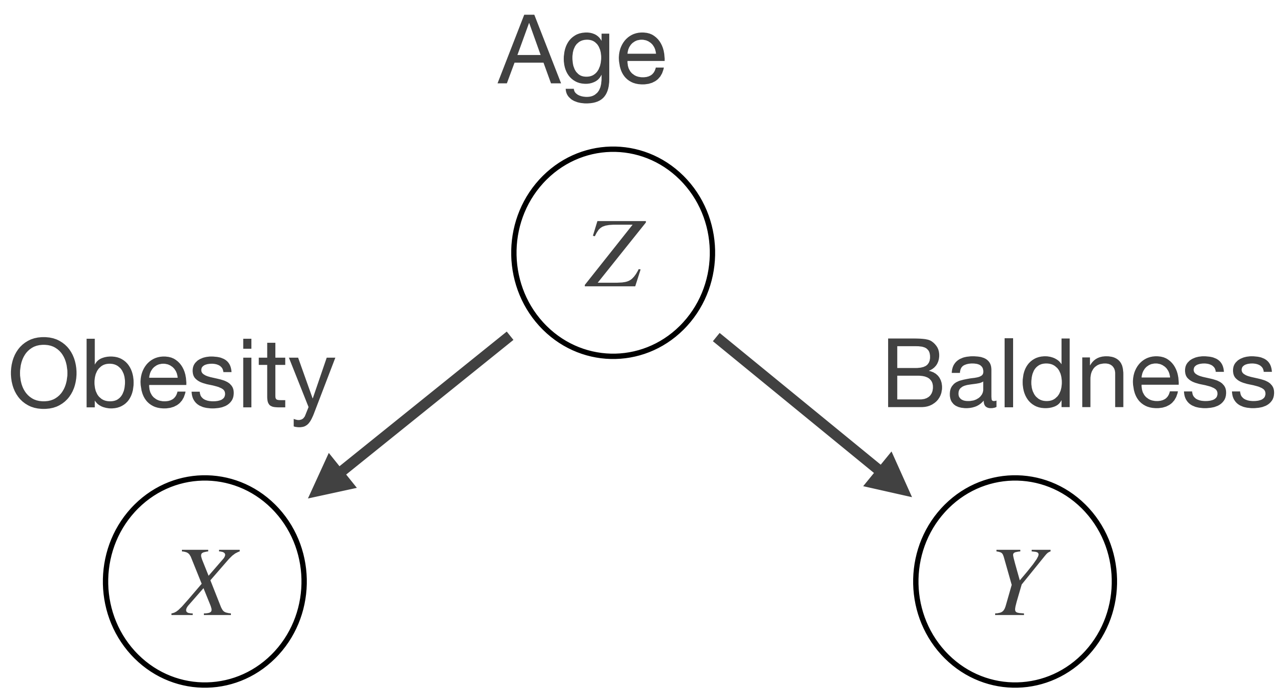
D-Separation

Graphical Tool for Identifying Conditional Independencies implied by Bayesian Networks

Implied Conditional independencies

Fork

Z **as a common cause**

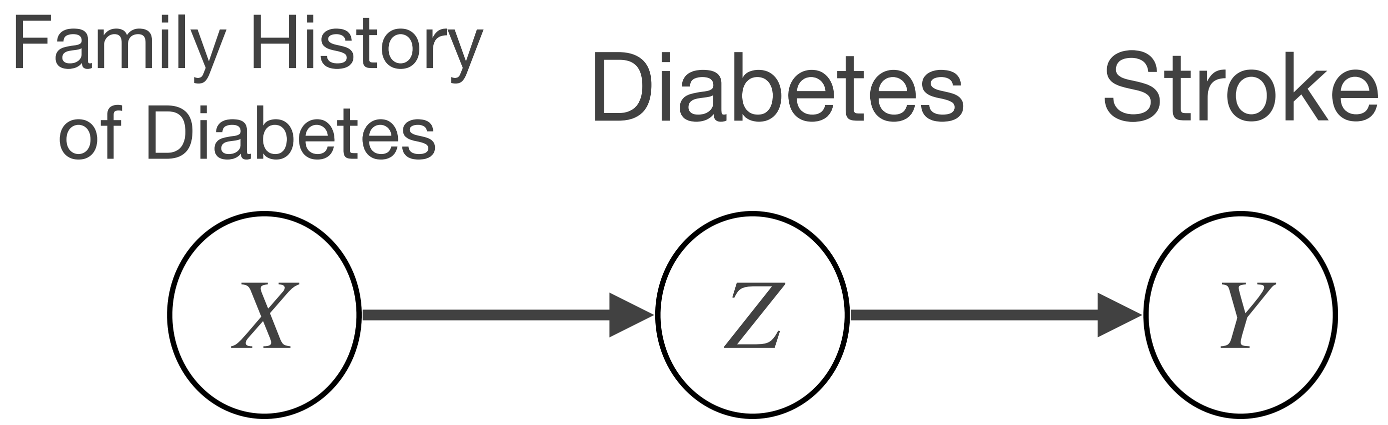


$$X \not\perp\!\!\!\perp Y$$

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

Chain

Z **as a mediator**

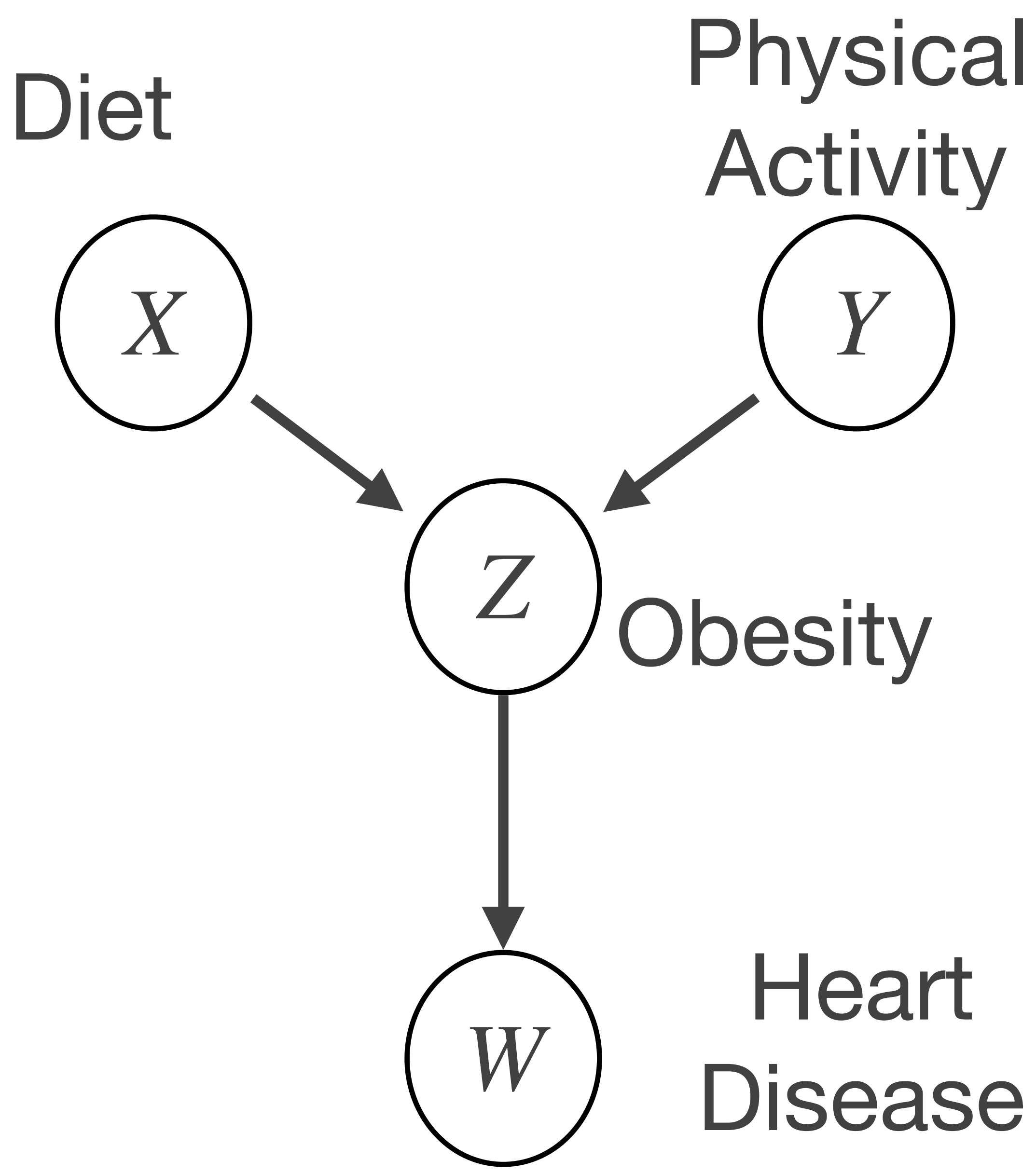


$$X \not\perp\!\!\!\perp Y$$

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

V-Structure

Z **as a collider or common effect**



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

$$X \not\perp\!\!\!\perp Y | Z$$

$$X \not\perp\!\!\!\perp Y | W$$

Two Markov-equivalent models.

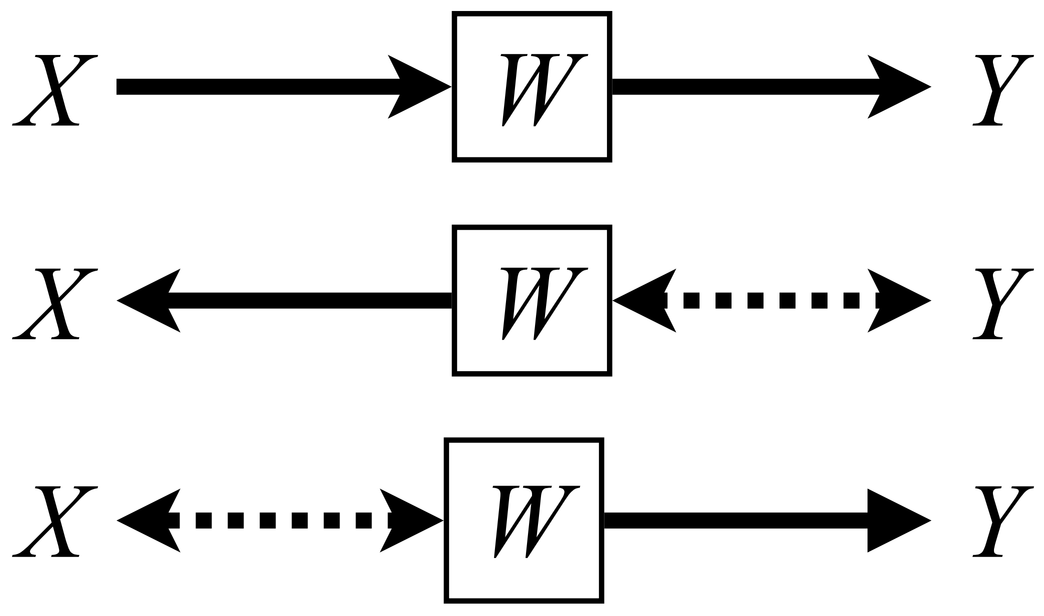
Note that in both cases Z is a non-collider!

Active and Inactive Triplets

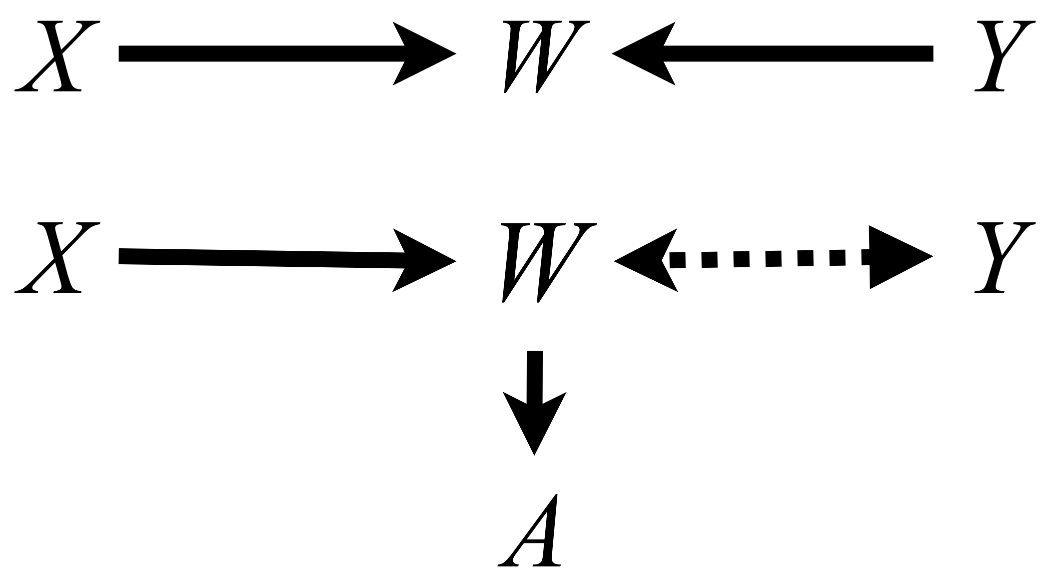
Definition (inactive): A triplet $\langle V_i, V_m, V_j \rangle$ is said to be ***inactive*** relative to a set $\mathbf{Z}$ if the middle node V_m:

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

W is non-collider and $W \in \mathbf{Z}$



W is (descendant of) a collider and $W, A \notin \mathbf{Z}$

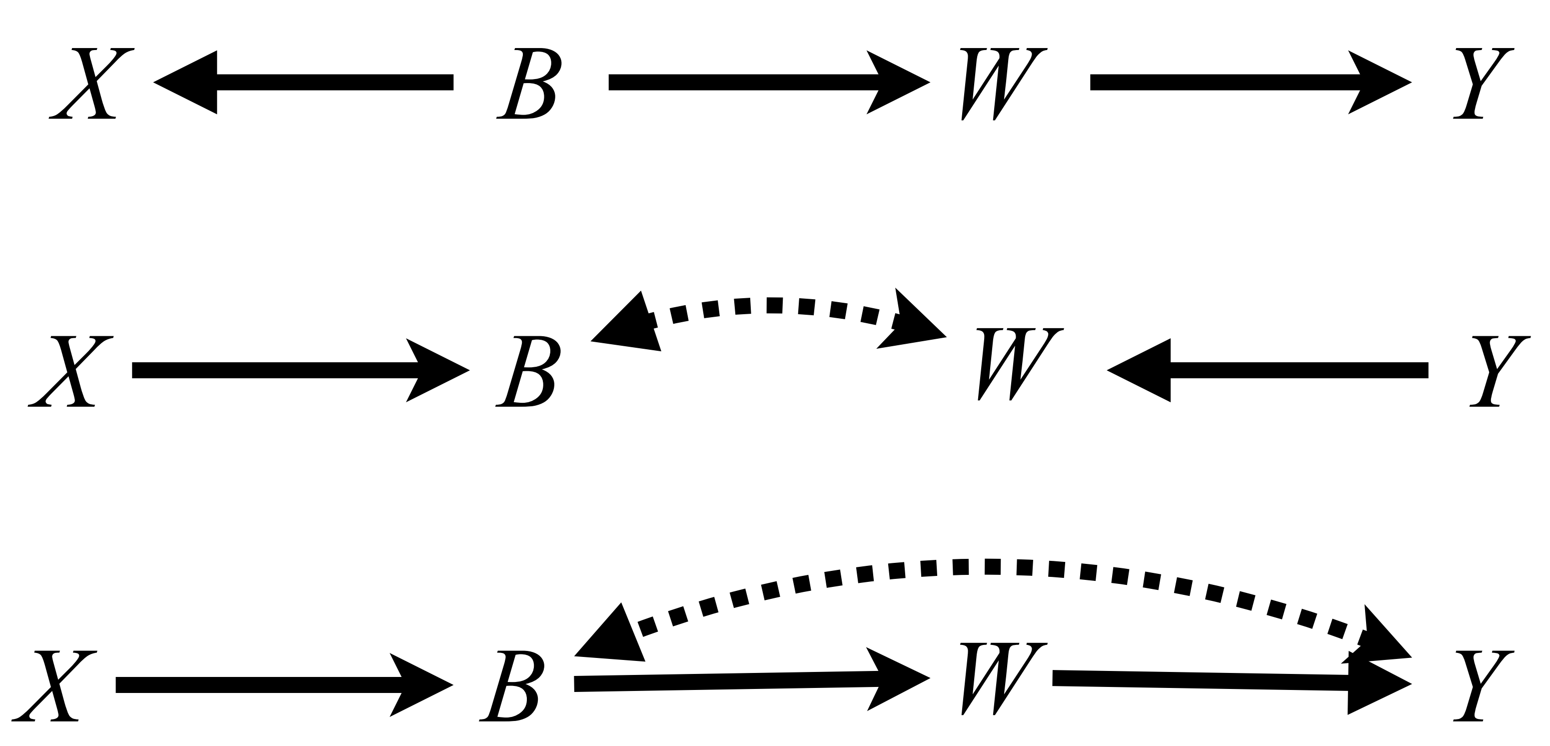


D-Separation

Definition (d-separation): A path p in an ADMG 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 $\mathbf{Y}$. We denote that by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G$.

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



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

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

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

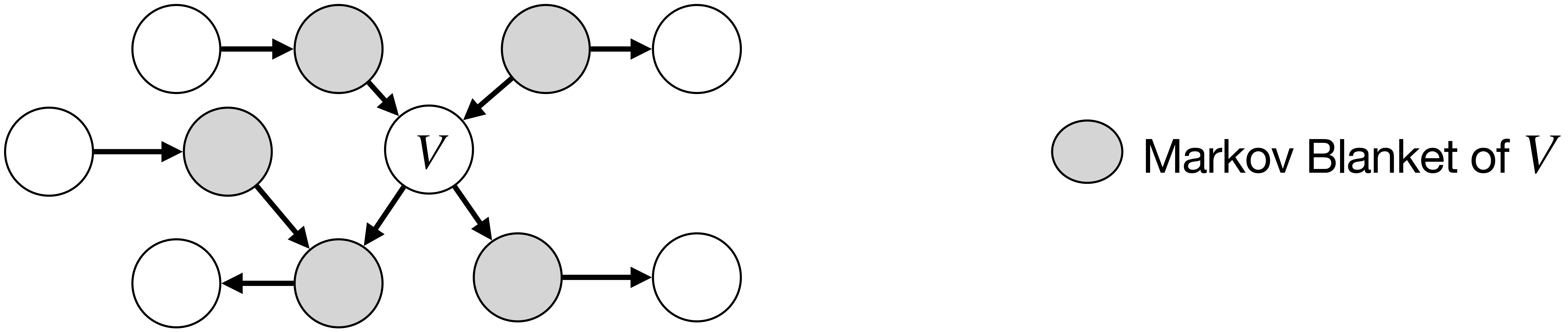
$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G \Rightarrow (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_P$$

D-separations in G correspond to conditional independencies in P

Markov Blanket (Markovian)

Markov Blanket (MB) of a *Markovian* BN over V: the union of parents, children, and parents of the children V.

$$\text{mb}_G(V) = Pa(V) \cup Ch(V) \cup Pa(Ch(V))$$



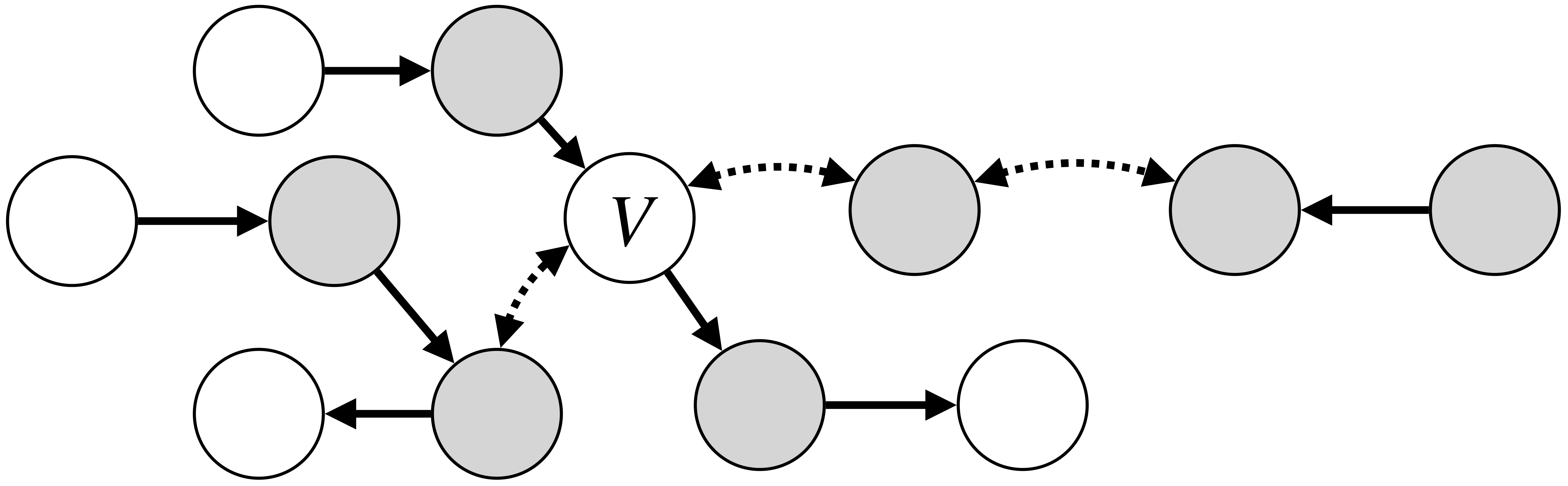
$$V \perp\!\!\!\perp \mathbf{V} \setminus \text{mb}_G(V) \mid \text{mb}_G(V)$$

Markov Blanket (Semi-Markovian)

Markov Blanket (MB) of a *Semi-Markovian* BN over V: is the district of V and the parents of the district of V (excluding V itself) i.e.:

$$\text{mb}_G(V) = \text{dis}_G(V) \cup \text{Pa}_G(\text{dis}_G(V)) \backslash \{V\}$$

District of V, $\text{dis}_G(V)$, is the set of variables connected with V through an edge or a bidirected path.



Markov Blanket of V

$$V \perp\!\!\!\perp \mathbf{V} \backslash \text{mb}_G(V) \,|\, \text{mb}_G(V)$$

Richardson, T. (2003). Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics*, *30*(1), 145–157

Can we learn the
Markov Equivalence Class from
observational data?

Learning the Markov Equivalence Class

Identifiability: In non-parametric settings (i.e., without making parametric or distributional assumptions) and solely from observational data, causal discovery algorithms can only learn a graphical representation of a *Markov equivalence class*!

Algorithms: Constraint-Based vs Score-Based

Systems: Causal Sufficient vs Causal Insufficient

Causal Sufficiency: assumption that all confounding variables have been observed — although strong, it has been widely employed to simplify causal discovery and inference.

Score-Based Causal Discovery Algorithms

Strategy: search for the most probable causal structure by assessing goodness-of-fit scores of different possible structures.

Common Scores: Bayesian Information Criterion (BIC) for Gaussian variables and the BDeu score for multinomial variables.

Under causal sufficiency:

- **GES:** Greedy Equivalence Search, by Chickering, 2003.
- **FGES:** Fast GES, by Ramsey et al., 2017 — extension of the GES that improves the runtime of the algorithm by using parallelization.

Score-Based Causal Discovery Algorithms

Accounting latent confounding:

Use BIC, assuming linear Gaussian models

- **GSMAG:** a greedy search algorithm for learning MAGs, by Triantafillou, S. and Tsamardinos, I., 2016.
- **MAGSL:** search based on **dynamic programming and branch and bound**, by Rantanen et al., 2021 — it is guaranteed to find a globally optimal MAG.
- **Diff-discovery:** solves a **continuous optimization problem** with differentiable procedures to find the best fitting ADMG, by Bhattacharya et al., 2021.
- **N-ADMG:** Neural ADMG Learning, by Ashman et al., 2013 — extends Diff-discovery to the setting where the true causal diagram is bow-free and corresponds to a non-linear SCM with additive noise.

Constraint-Based Causal Discovery Algorithms

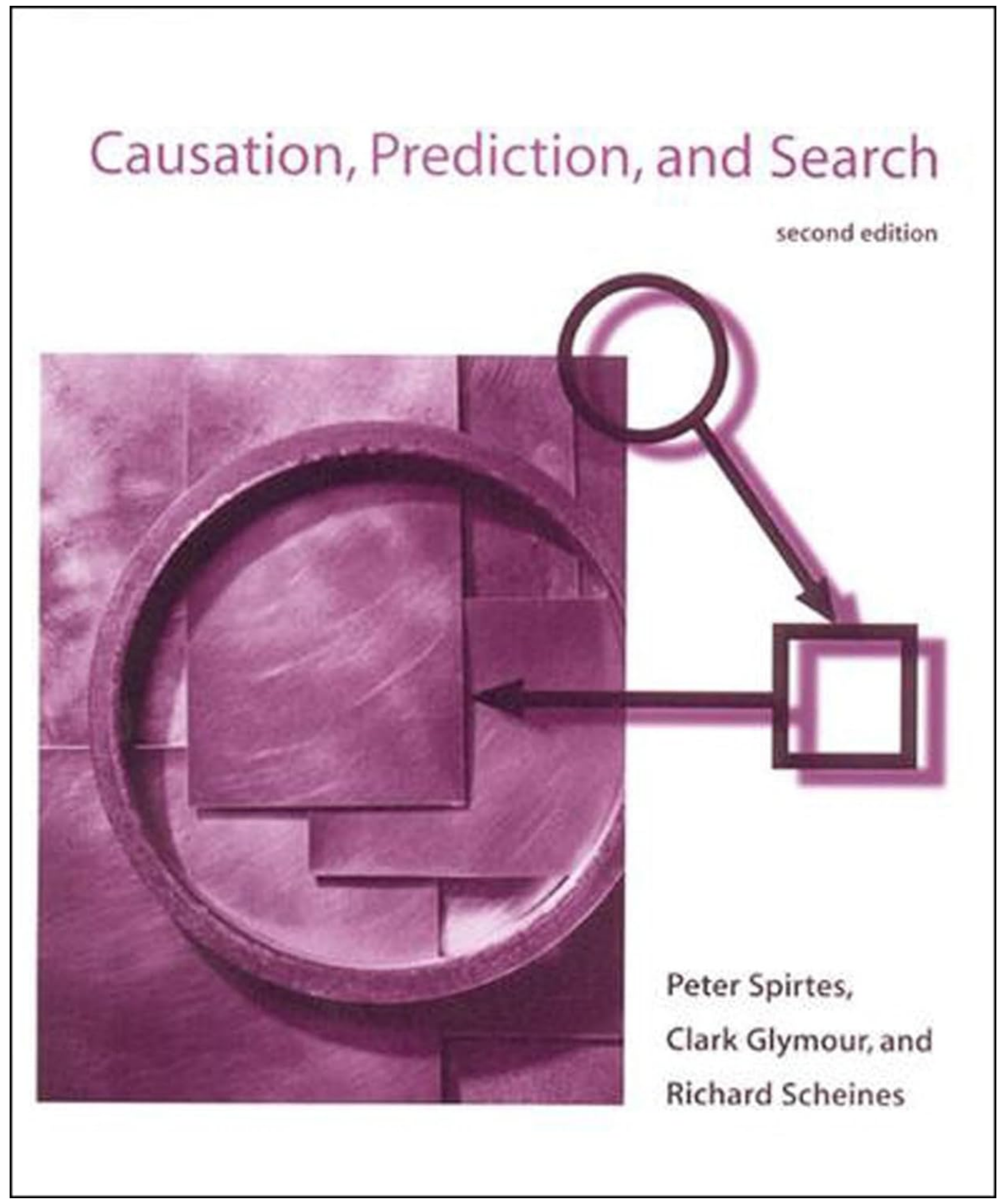
Strategy: construct a causal structure that aligns with all observed conditional independencies, identified using conditional independence tests.

Under causal sufficiency:

IC: Inductive Causation, by Verma and Pearl, 1990.

PC: Peter-Clark, by Spirtes and Glymour, 1991.

They start with an adjacency (skeleton) phase, based on conditional independence tests, followed by an orientation phase.



Spirtes, P., Glymour, C., and Scheines, R. (2001).
Causation, Prediction, and Search, 2nd edn. Cambridge, MA: MIT Press.

Constraint-Based Causal Discovery Algorithms

Accounting for latent confounding:

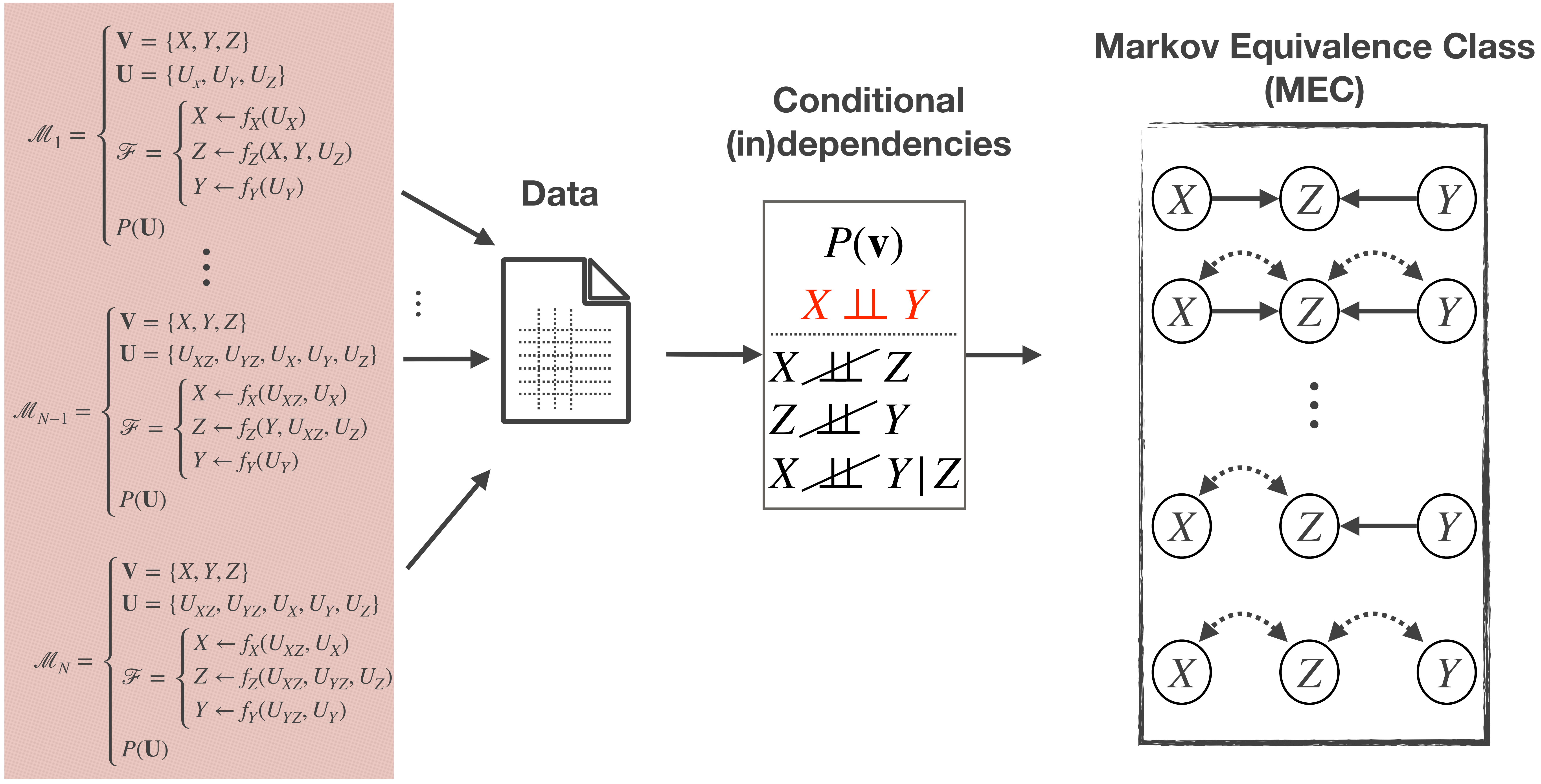
- **FCI:** Fast Causal Inference, by Spirtes et al., 1995 — most prominent extension of the PC and IC/IC* algorithms. Together with the additional rules by Zhang, J. (2008), is a complete algorithm accounting for both latent confounding and selection bias.

- **FCI variants:** Anytime FCI (**AFCI**), by Spirtes P., 2001, Conservative FCI (**CFCI)** and Really FCI (**RFCI**), by Colombo et al. 2012; and **FCI+**, by Claassen et al. 2013.

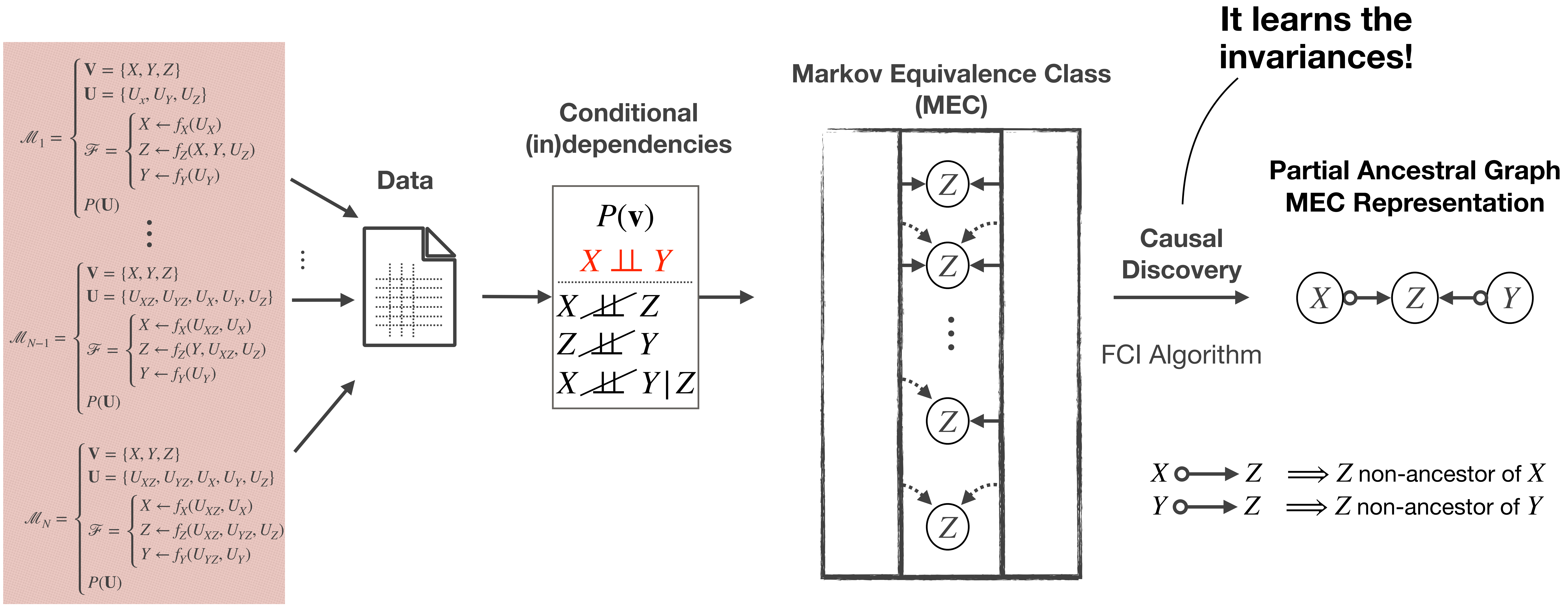
- **SAT-Based:** uses a Answer Set Programming (ASP) solver to find a causal structure that most satisfies the minimal observed conditional independencies, by Hyttinen et al., 2014.

- **ACI:** Ancestral Causal Inference — a logic-based algorithm by Magliacane et al., 2016.

Constraint-Based Causal Discovery Algorithms



Constraint-Based Causal Discovery Algorithms

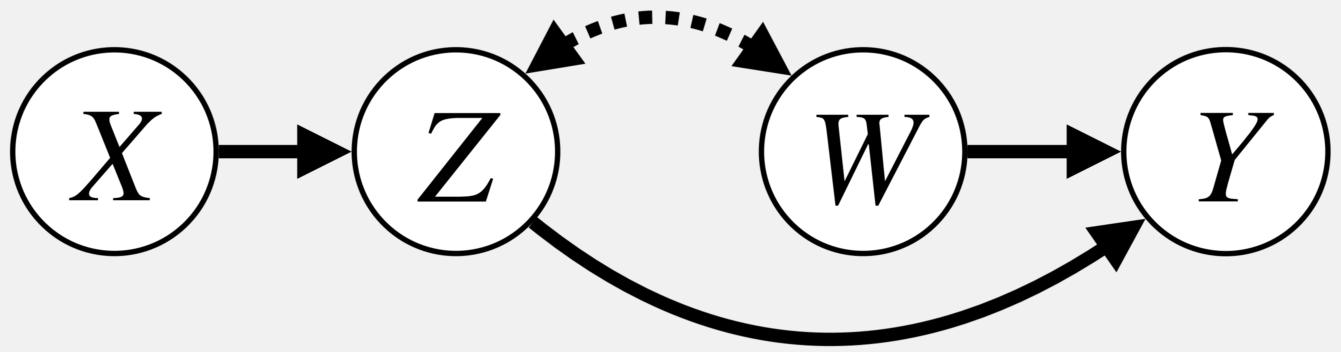


Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. Link

FCI Algorithm - Pipeline

Unknown Reality

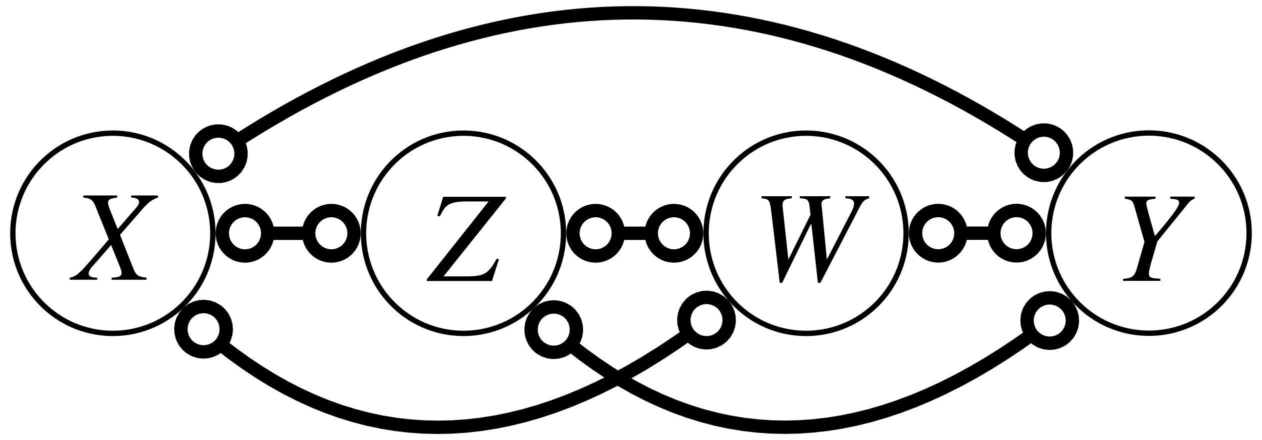
True causal diagram



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

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

Implied by the ADMG using d-separation



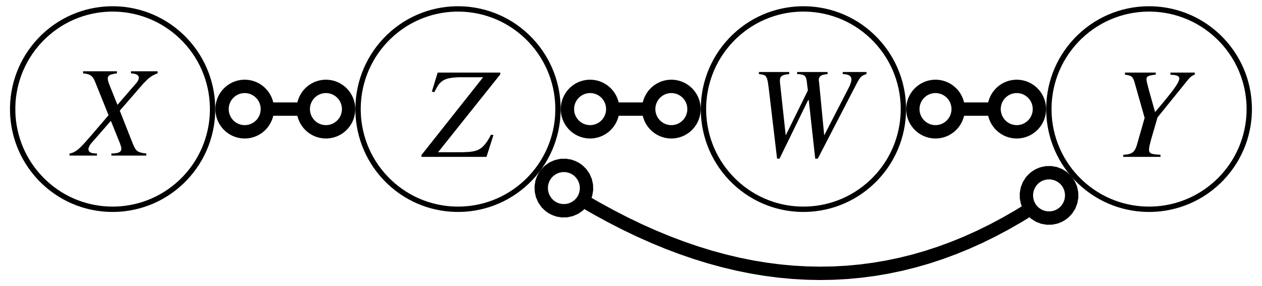
Complete Graph

Conditional Independence Tests

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

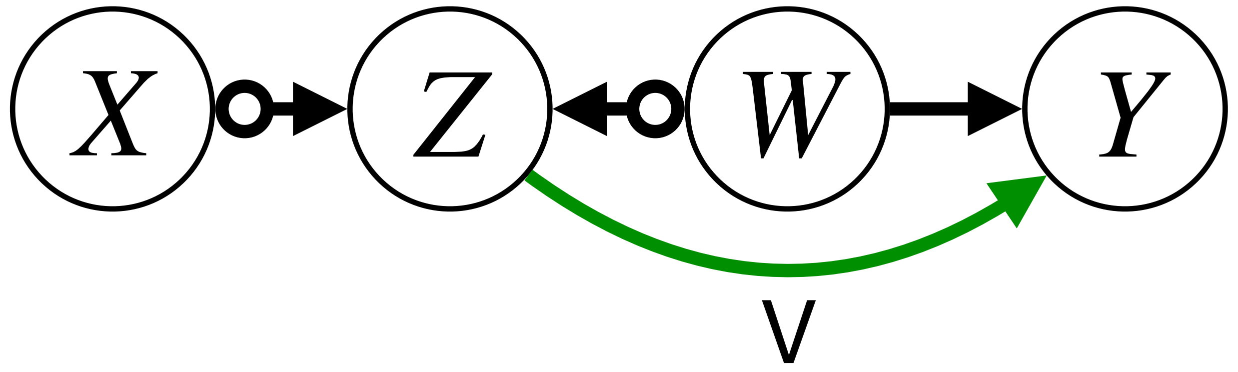
$X \perp\!\!\!\perp Y \mid Z, W$

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 o→ B $\Longrightarrow$ B non-ancestor of A

$A \longrightarrow B$ $\Longrightarrow$ A ancestror of B

$A \longleftrightarrow B$ $\Longrightarrow$ spurious association

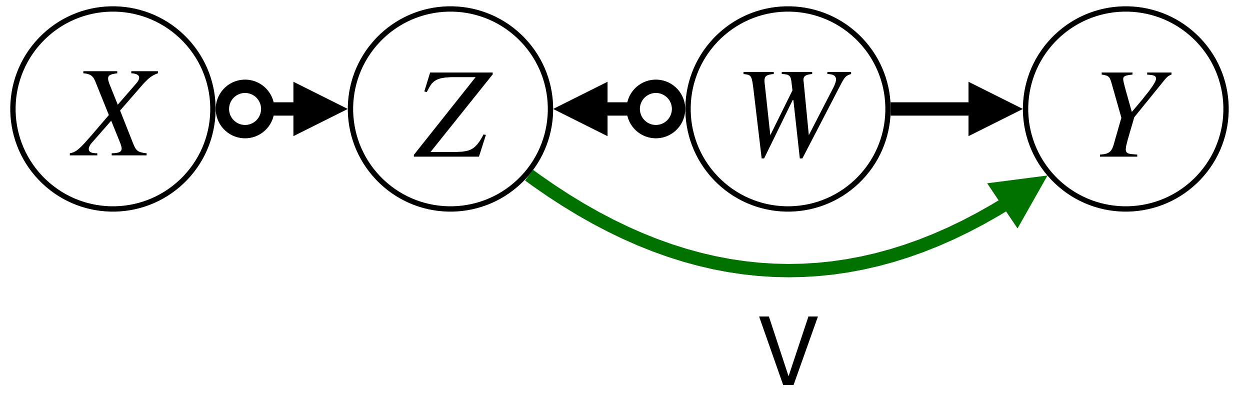
A — B $\Longrightarrow$ selection bias

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

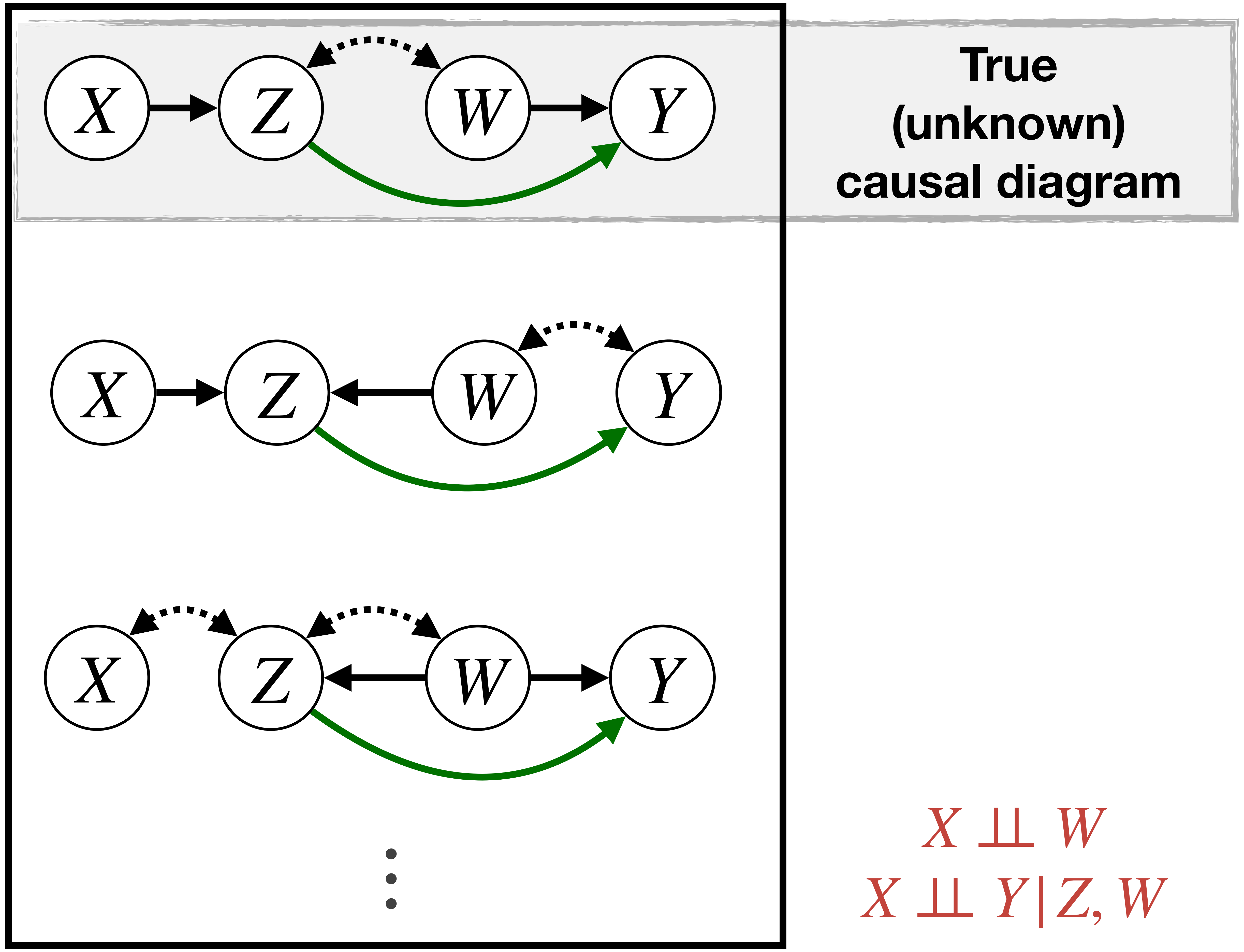


Partial Ancestral Graph
(PAG)

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

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

Z **is not confounded with** Y**.**

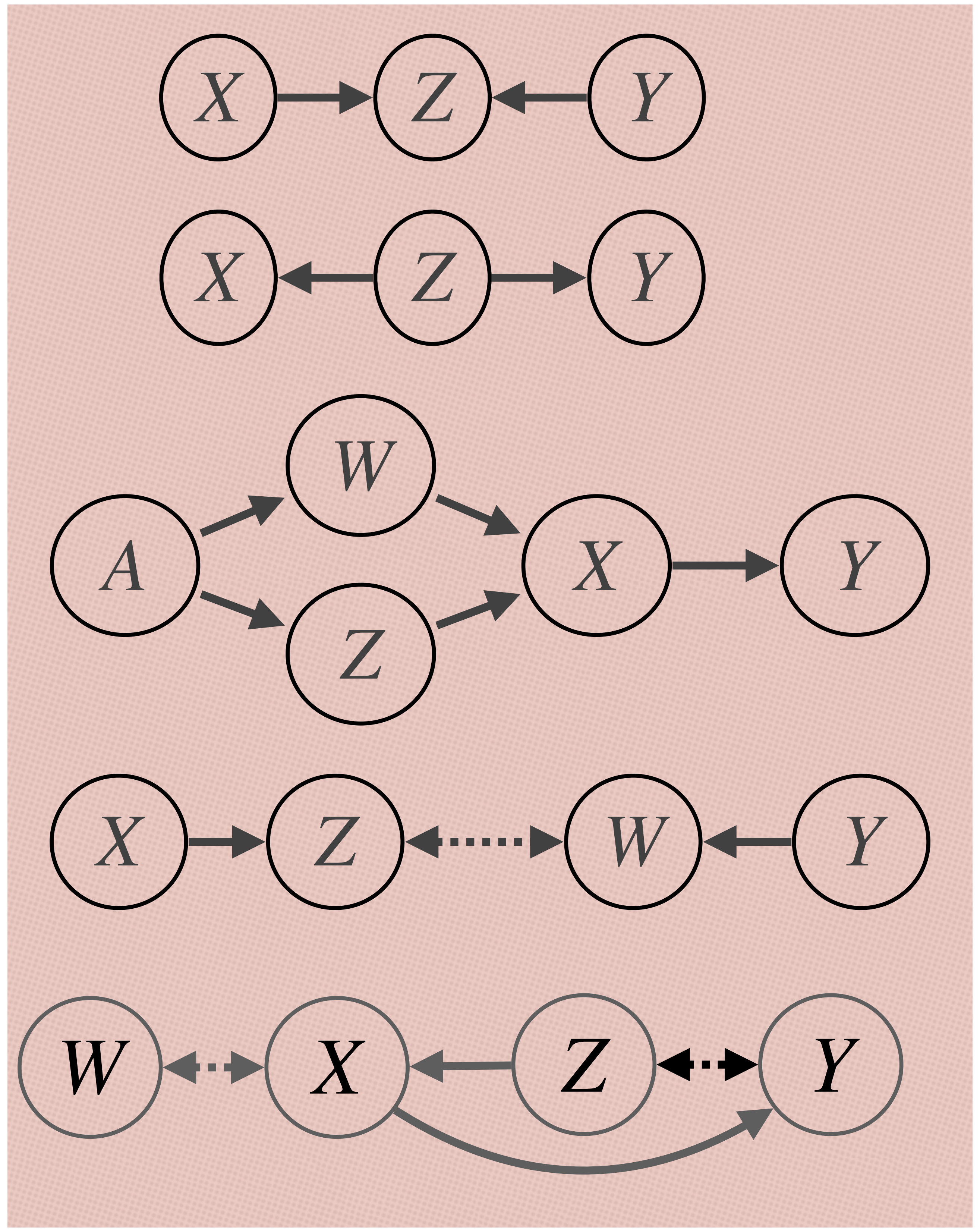


$X \perp\!\!\!\perp W$

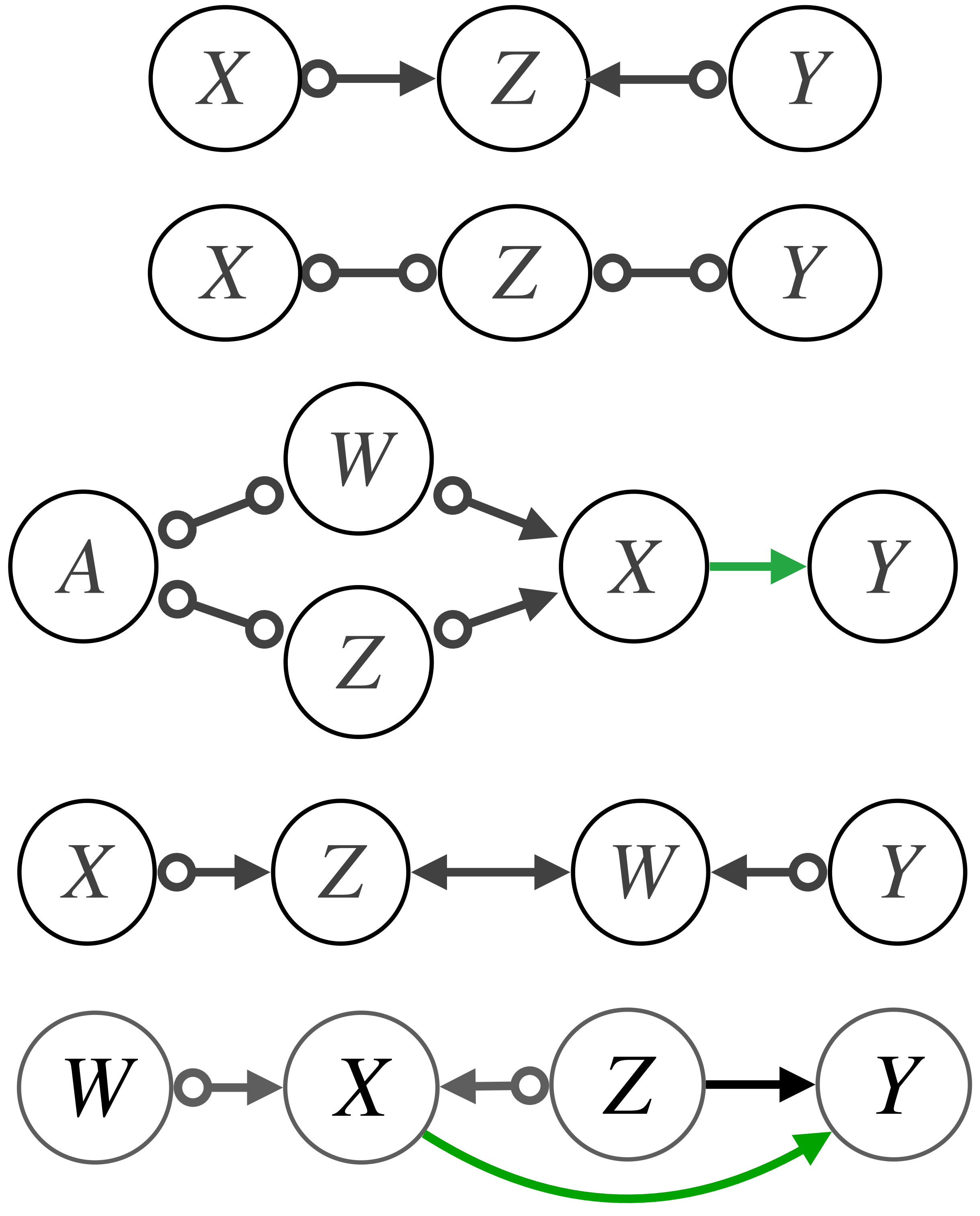
$X \perp\!\!\!\perp Y \mid Z, W$

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) - Likelihood ratio tests based on GLM

- **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

Developments in Causal Discovery with Unobserved Confounding

Causal Discovery with Robustness to Empirical Unfaithfulness

- **Ribeiro, A. H., & Heider, D.** (2025). dcFCI: Robust Causal Discovery Under Latent Confounding, Unfaithfulness, and Mixed Data. *arXiv preprint arXiv:2505.06542 (link)*.
  dcFCI R package: *GitHub repository:* @adele/dcFCI

Leverages a PAG-data compatibility score that supports heterogeneous variable types.

Causal Discovery of Cluster DAGs

- **Anand, T. , Ribeiro, A. H., Tian, J., Hripcsak, G. & E. Bareinboim.** (2025). Causal Discovery over Clusters of Variables in Markovian Systems. Columbia CausalAI Laboratory, Technical Report (R-128), June, 2025. (link).

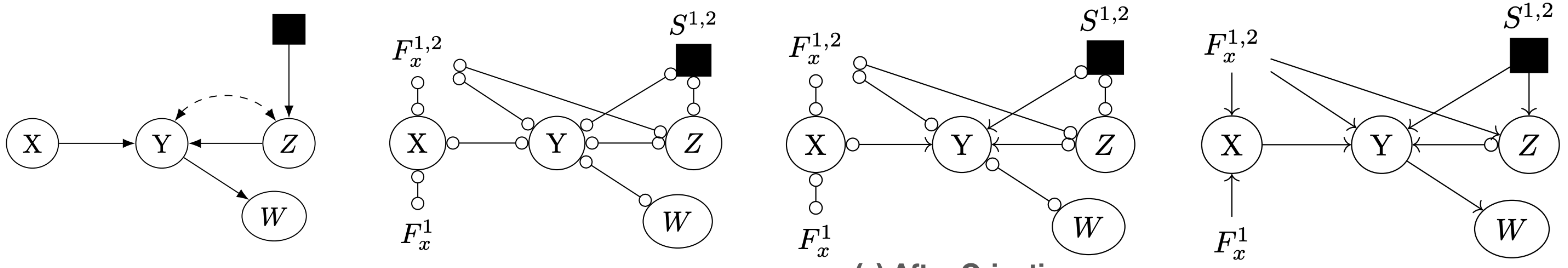
Learns Equivalence Classes of Cluster DAGs (C-DAGs) under causal sufficiency

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.

- **Ribeiro, A. H.** Crnkovic, M., …, Heider, D., and Cerqueira, A. (2024).
  AnchorFCI: Harnessing Genetic Anchors for Enhanced Causal Discovery of Cardiometabolic Disease Pathways.
  Frontiers in Genetics 15:1436947. (Link)
  AnchorFCI R package — GitHub*:* @adele/anchorFCI

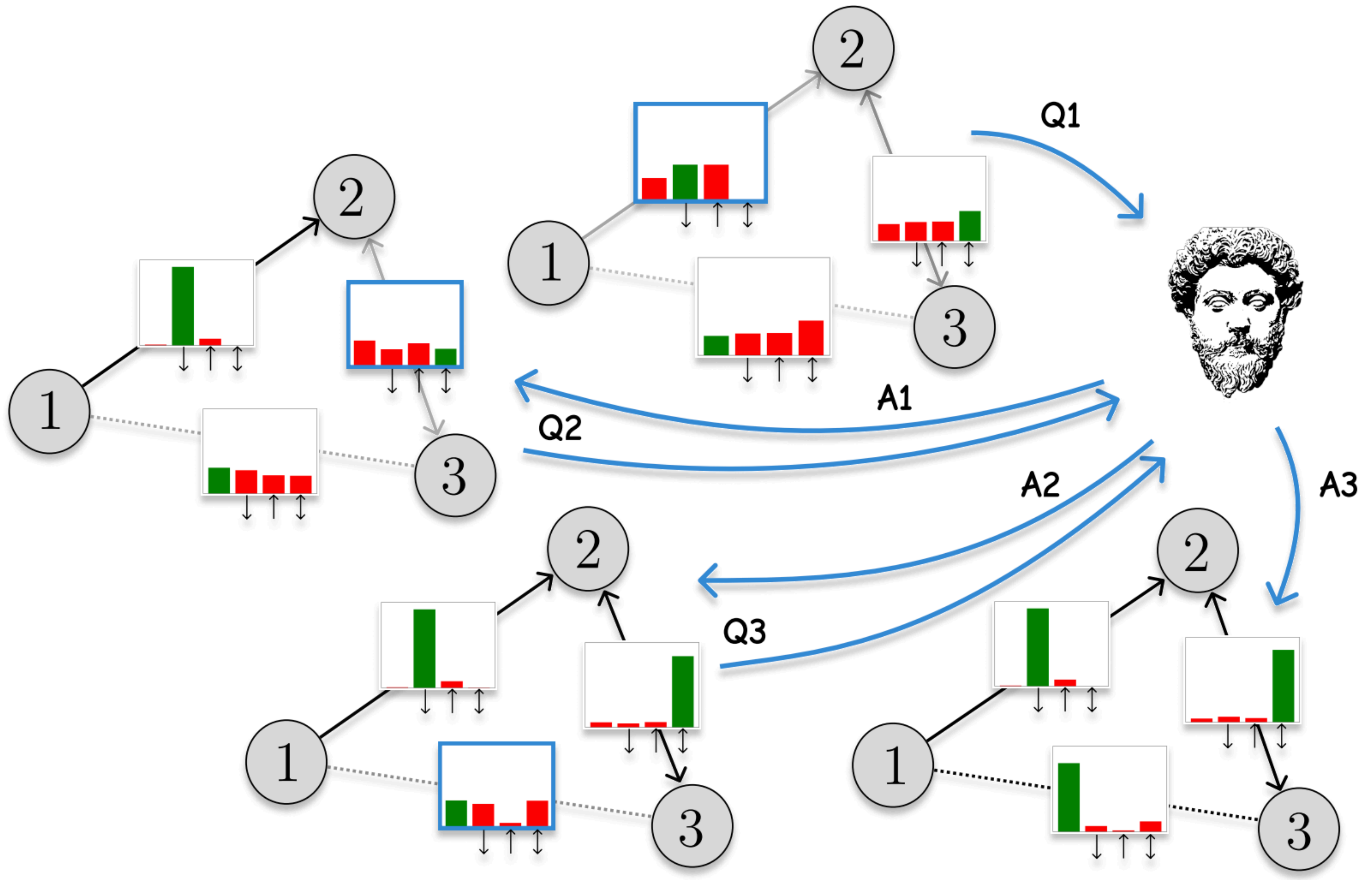
  - Integrates known non-ancestralities
    e.g., Genotypes $\prec$ Phenotypes

Developments in Causal Discovery with Unobserved Confounding

3. Human-in-the-Loop Probabilistic Causal Discovery

   - **da Silva, T., Silva, E., Góis, A., Heider, D., Kaski, S., Mesquita, D., Ribeiro, A.H.** (2023). Human-in-the-Loop Causal Discovery under Latent Confounding using Ancestral GFlowNets. *arXiv:2309.12032* (Link)

   - **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).



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

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 Effect Identification Given a Causal Diagram

Graphical Criteria, Do-Calculus, and ID-Algorithm

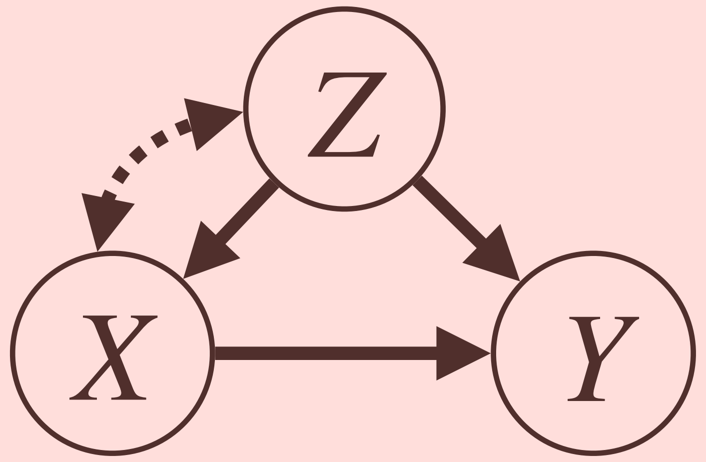
Classical Causality Pipeline from a Causal Diagram

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

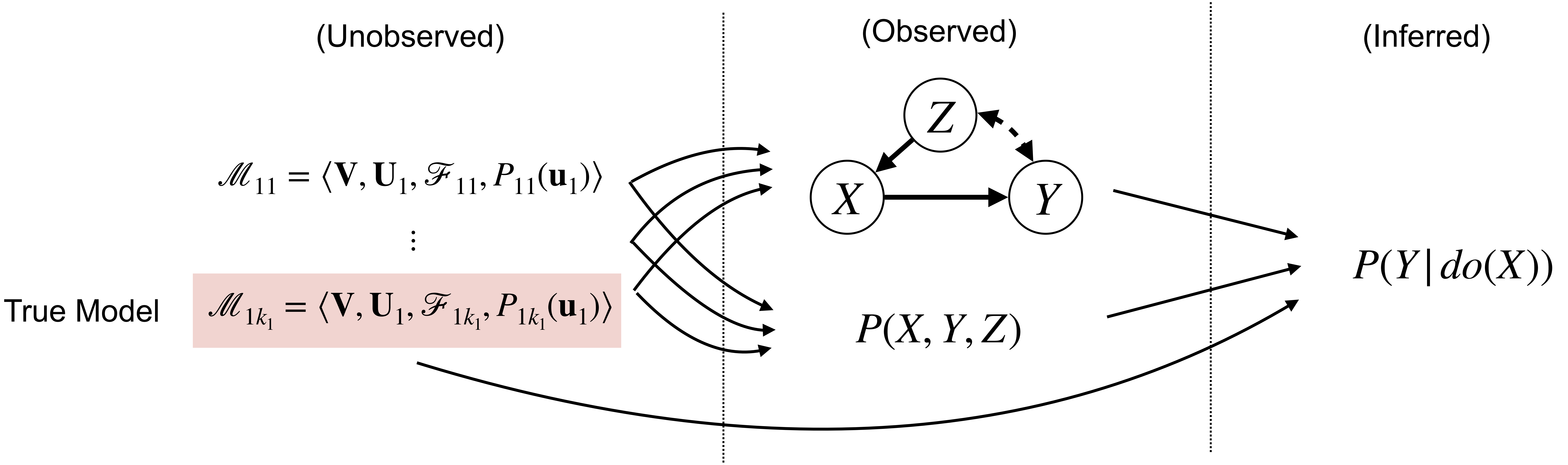
Observational Data

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

Causal Modeling → Causal Effect Identification → Causal Effect Estimation

The Effect Identification Problem

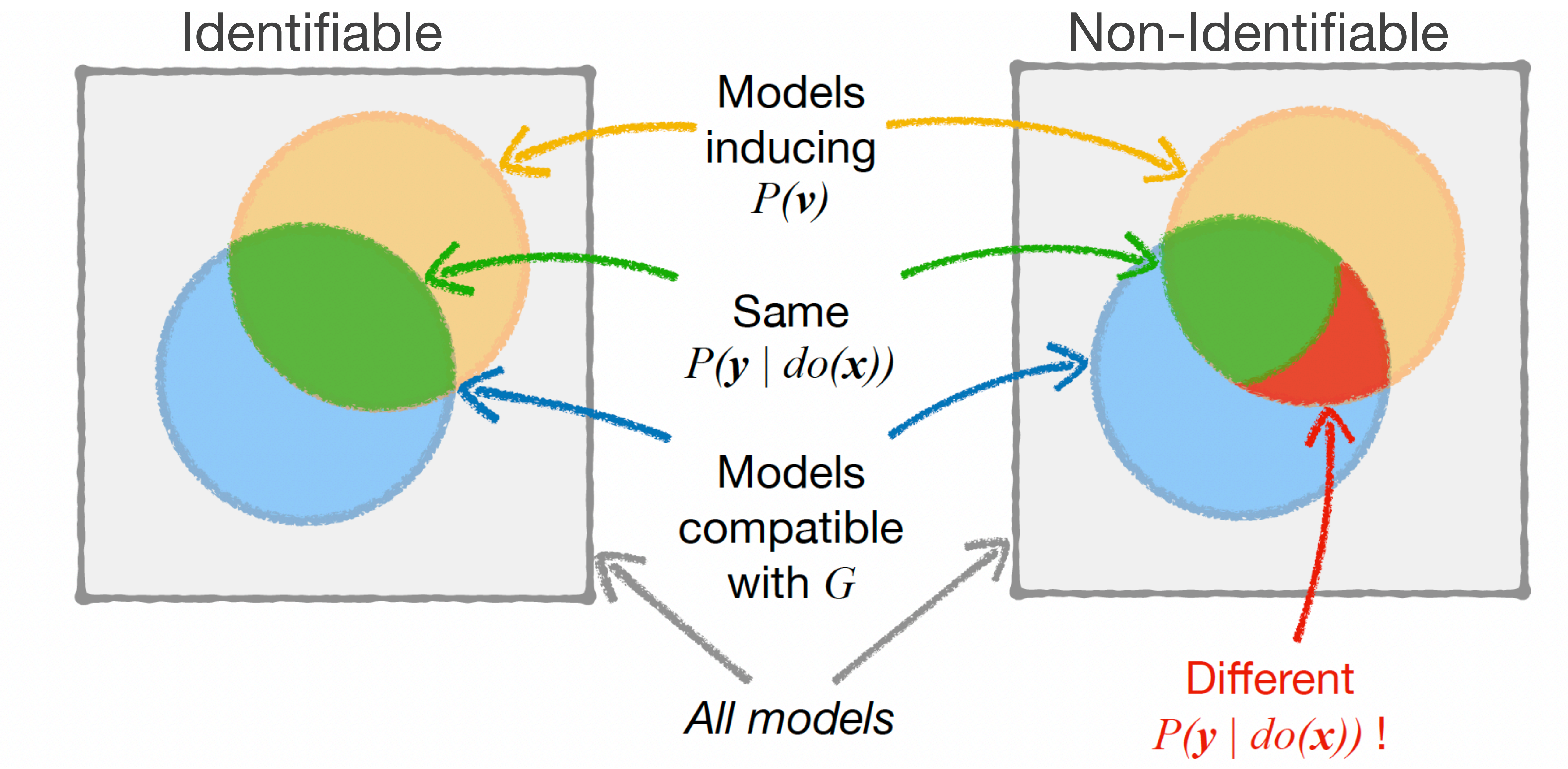
Causal Effect Identifiability: The effect of $\mathbf{X}$ on $\mathbf{Y}$ is said to be *identifiable* from a causal diagram G and the probability distribution $P(\mathbf{V})$ if $P(\mathbf{Y} \mid do(\mathbf{X}))$ is *uniquely computable,* i.e., if for every pair of SCMs $\mathcal{M}_1$ and $\mathcal{M}_2$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

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:

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

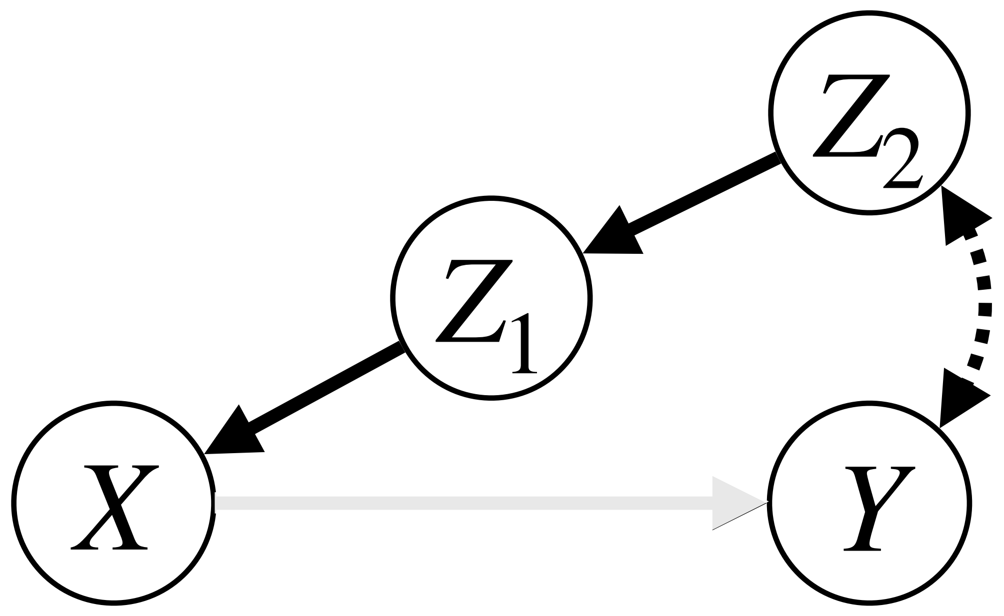
$\mathbf{Z}$, a set of covariates, admissible for backdoor adjustment

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

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

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

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



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:

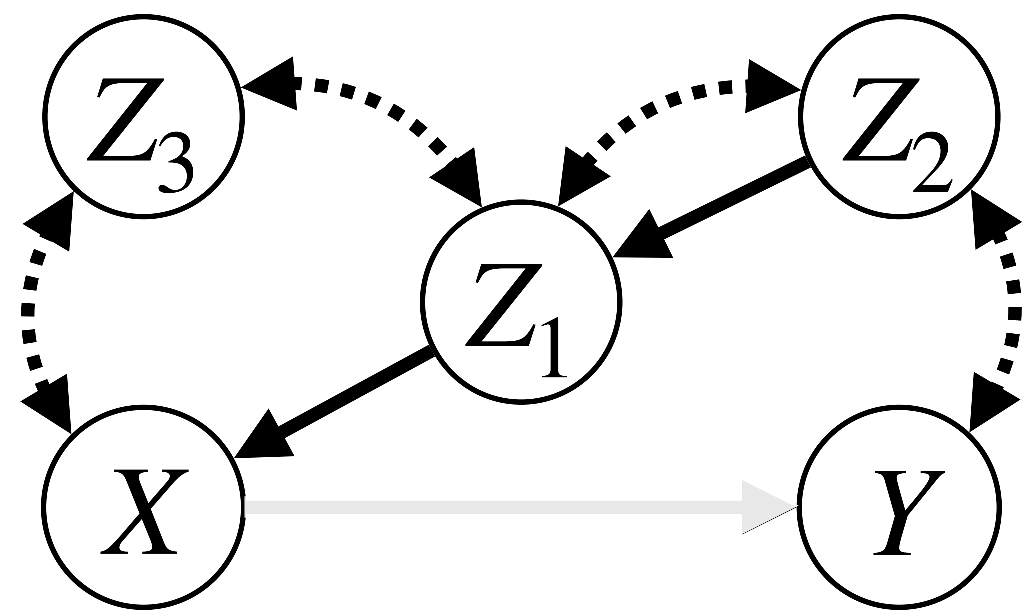
$$P(\mathbf{y} \mid do(\mathbf{x})) = \sum_{\mathbf{z}} P\left(\mathbf{y} \mid \mathbf{x}, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

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

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

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

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



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 $\mathbf{X}$

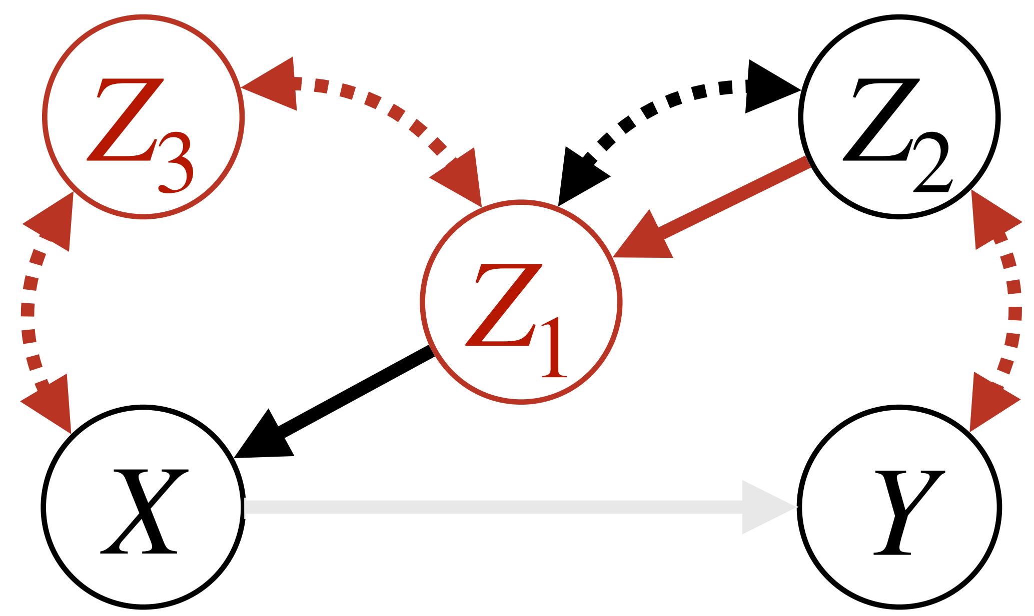
2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ in G (all variables in $\mathbf{Z}$ are pre-treatment)

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:

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

$\mathbf{Z}$, 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:

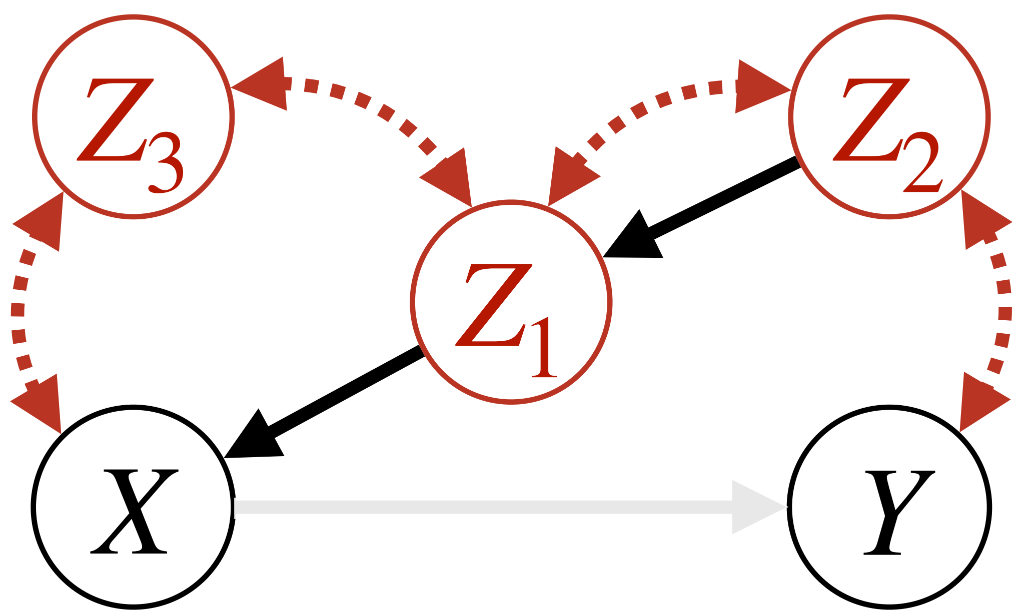
1. $\mathbf{Z}$ d-separates $\mathbf{X}$ and $\mathbf{Y}$ in the graph $G_{\underline{\mathbf{X}}}$, i.e., the graph resulting from cutting the arrows out of $\mathbf{X}$
2. no node in $\mathbf{Z}$ is a descendant of a variable $X \in \mathbf{X}$ in G (all variables in $\mathbf{Z}$ are pre-treatment)

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

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



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

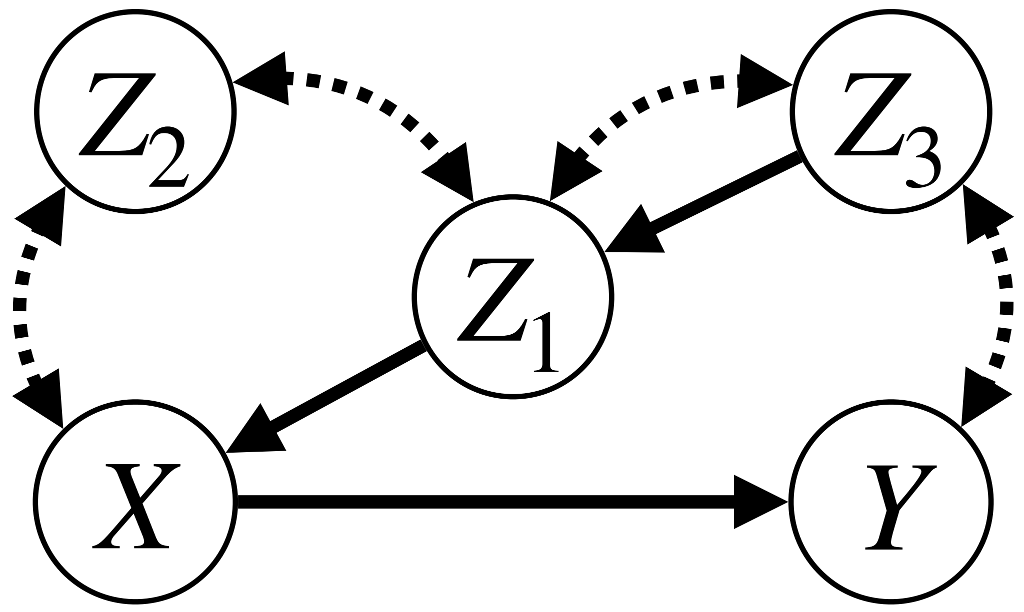
$\mathbf{Z} = \{Z_1, Z_3\}$

$\mathbf{Z} = \{Z_1, Z_2, Z_3\}$ ✗

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.,



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

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

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

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

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

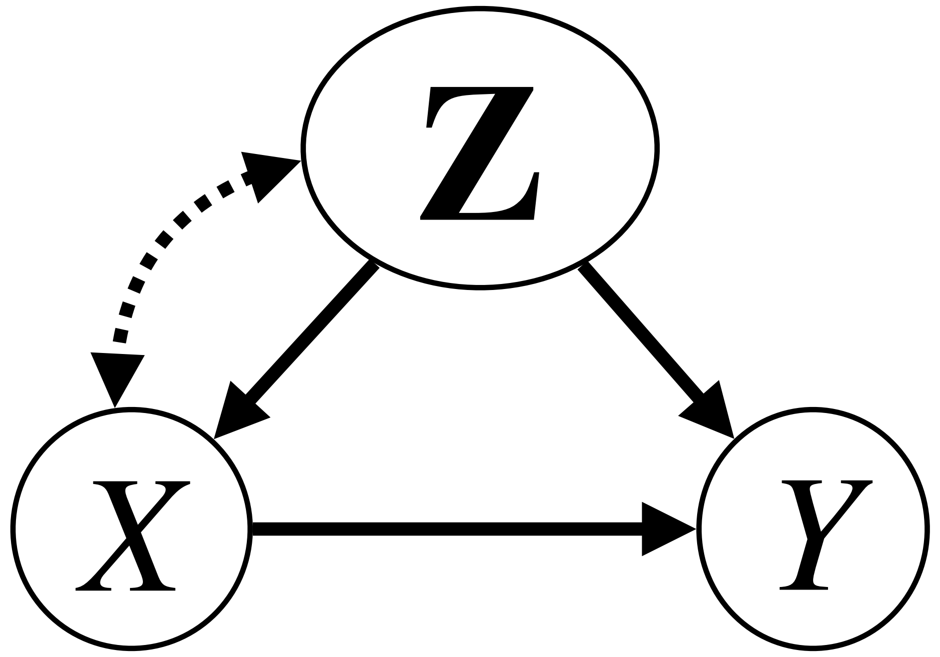
Only if $\mathbf{Z}$ is admissible for adjustment, Propensity Score can be used to estimate $P(y \mid do(x))$.

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!

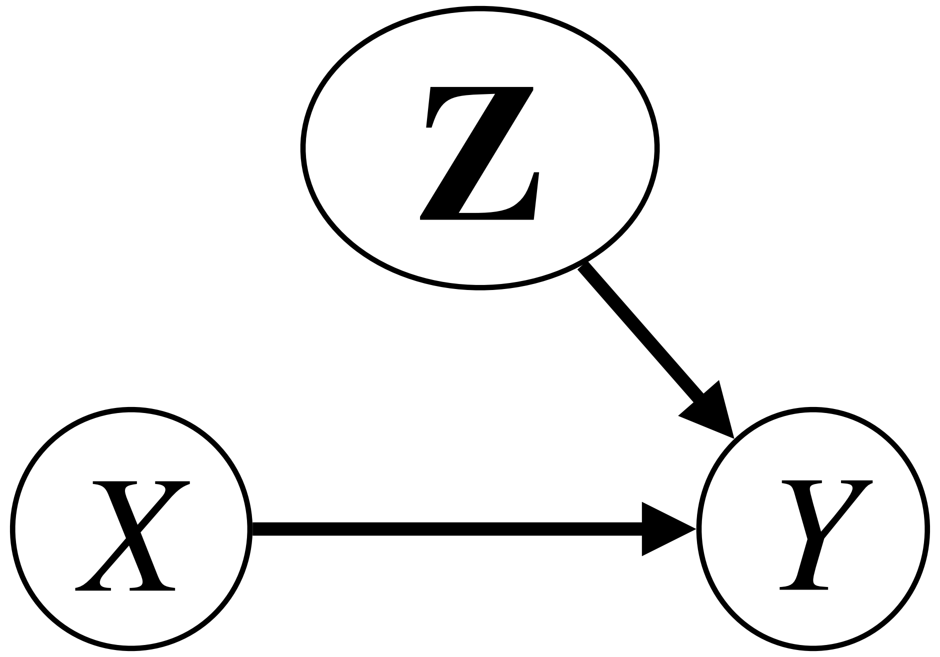
Inverse Probability Weighting (IPW)

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



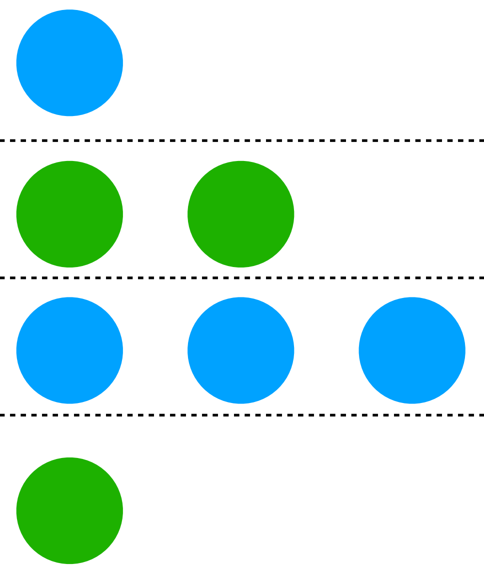
$P(X|\mathbf{Z}) \neq P(X)$

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



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

Original Sample



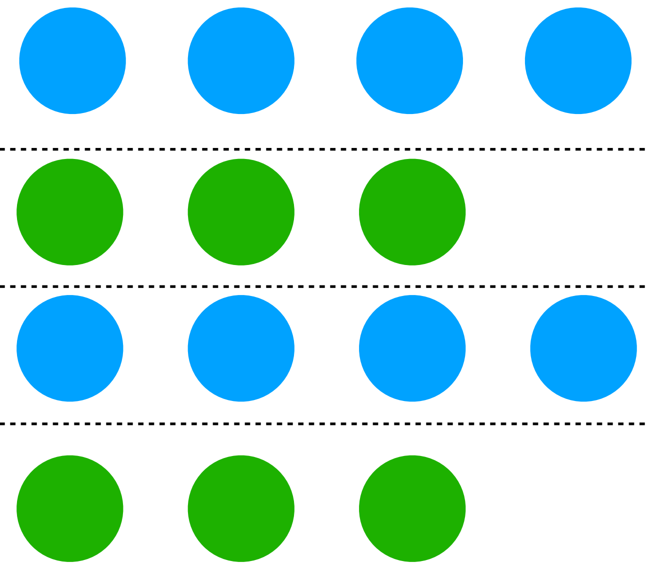
	$P(X\|Z)$	$\frac{1}{P(X\|Z)}$
X=0 (Control Group)	1/4	4
	2/3	1.5
X = 1 (Treated Group)	3/4	1.33
	1/3	3

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 $\{x_i, y_i, \mathbf{z_i}\}_{i=1}^{N}$:

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

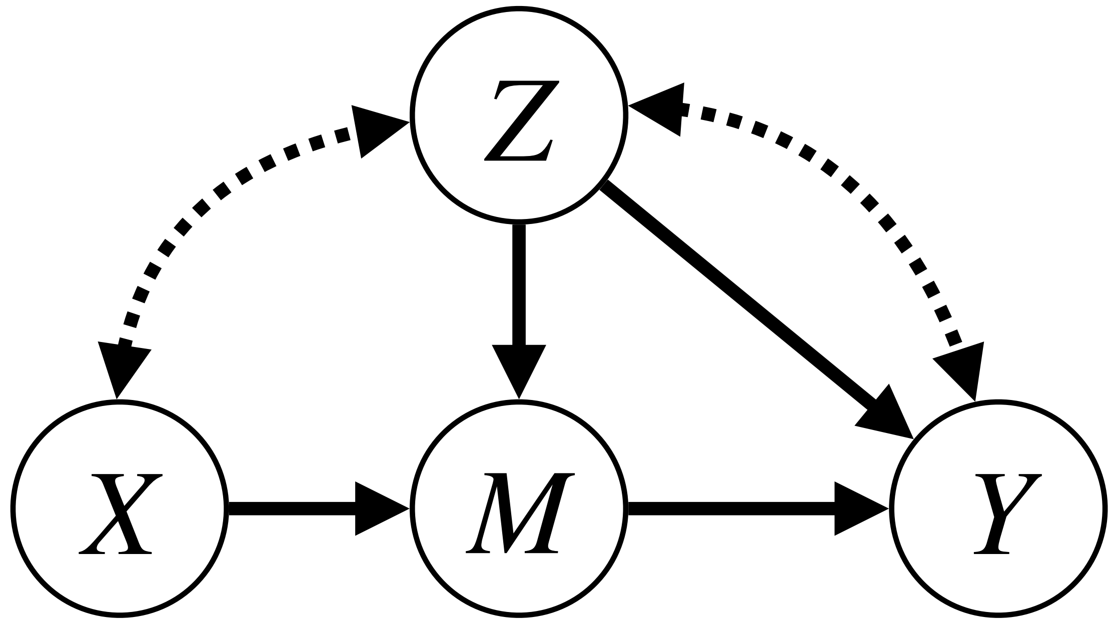
The mean of all values y_i, inversely weighted according to the propensity score.

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

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

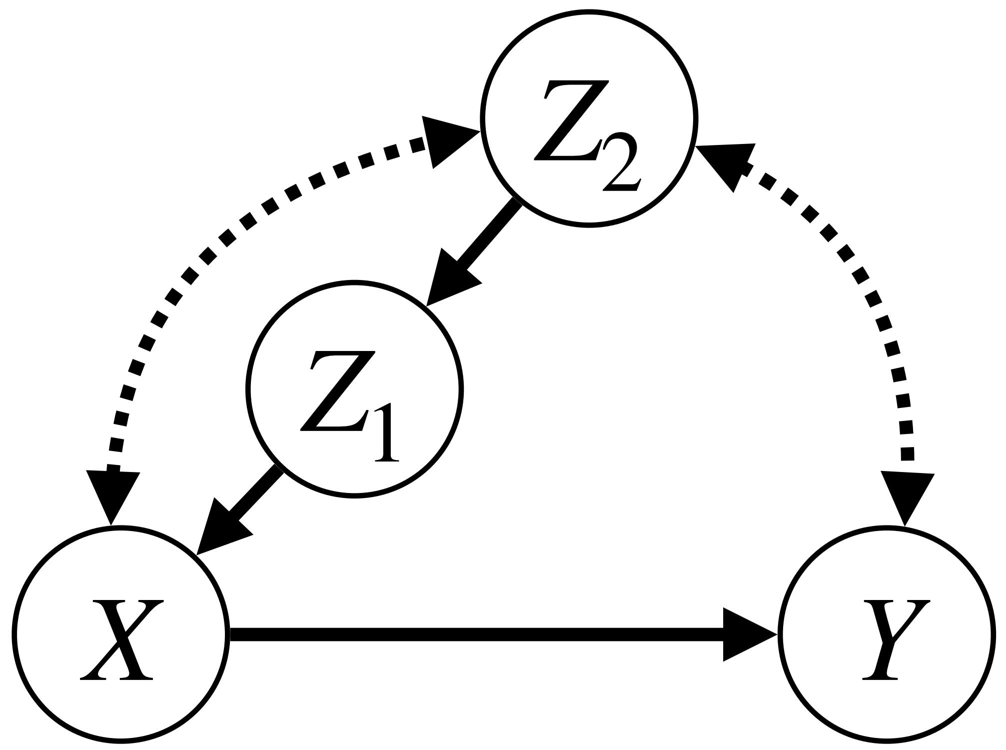
$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i\mathbf{1}_{\{x_i=1\}}}{\hat{P}(X=1|\mathbf{z}_i)} - \frac{y_i\mathbf{1}_{\{x_i=0\}}}{\hat{P}(X=0|\mathbf{z}_i)}\right)$$

Many Scenarios Beyond Backdoor …



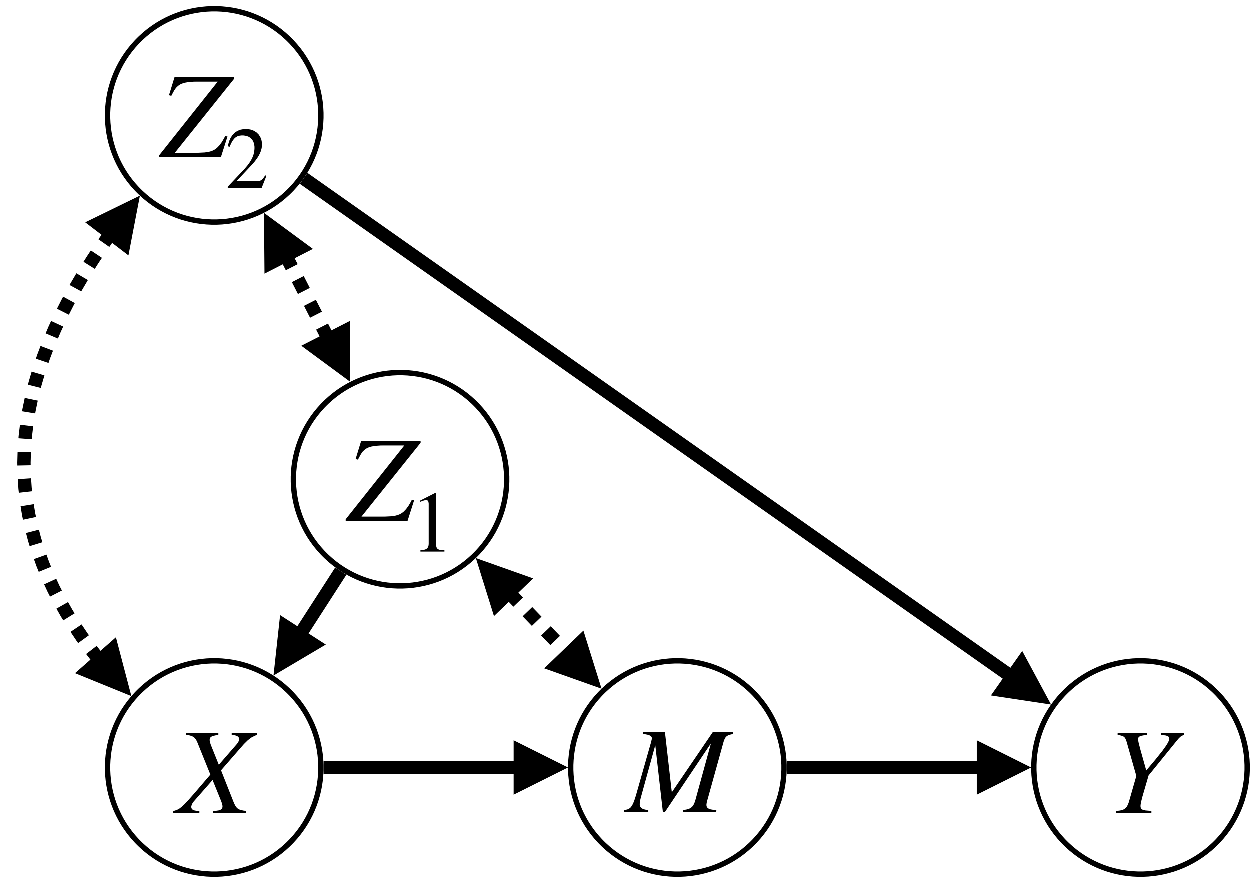
Conditional Front-Door

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



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….

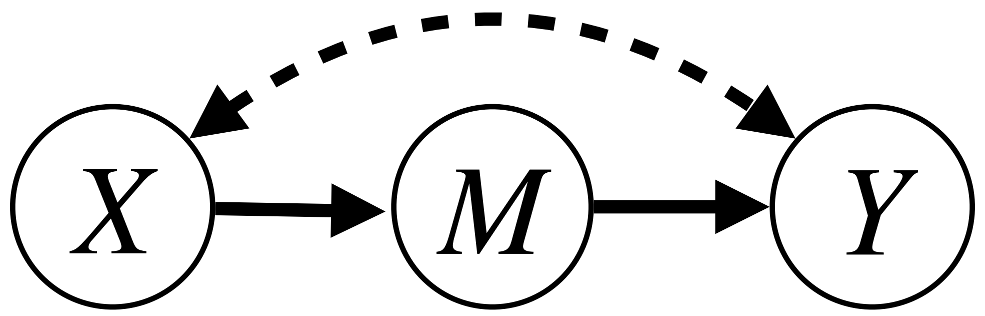
Causal Effect Identification (ID) Algorithm

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

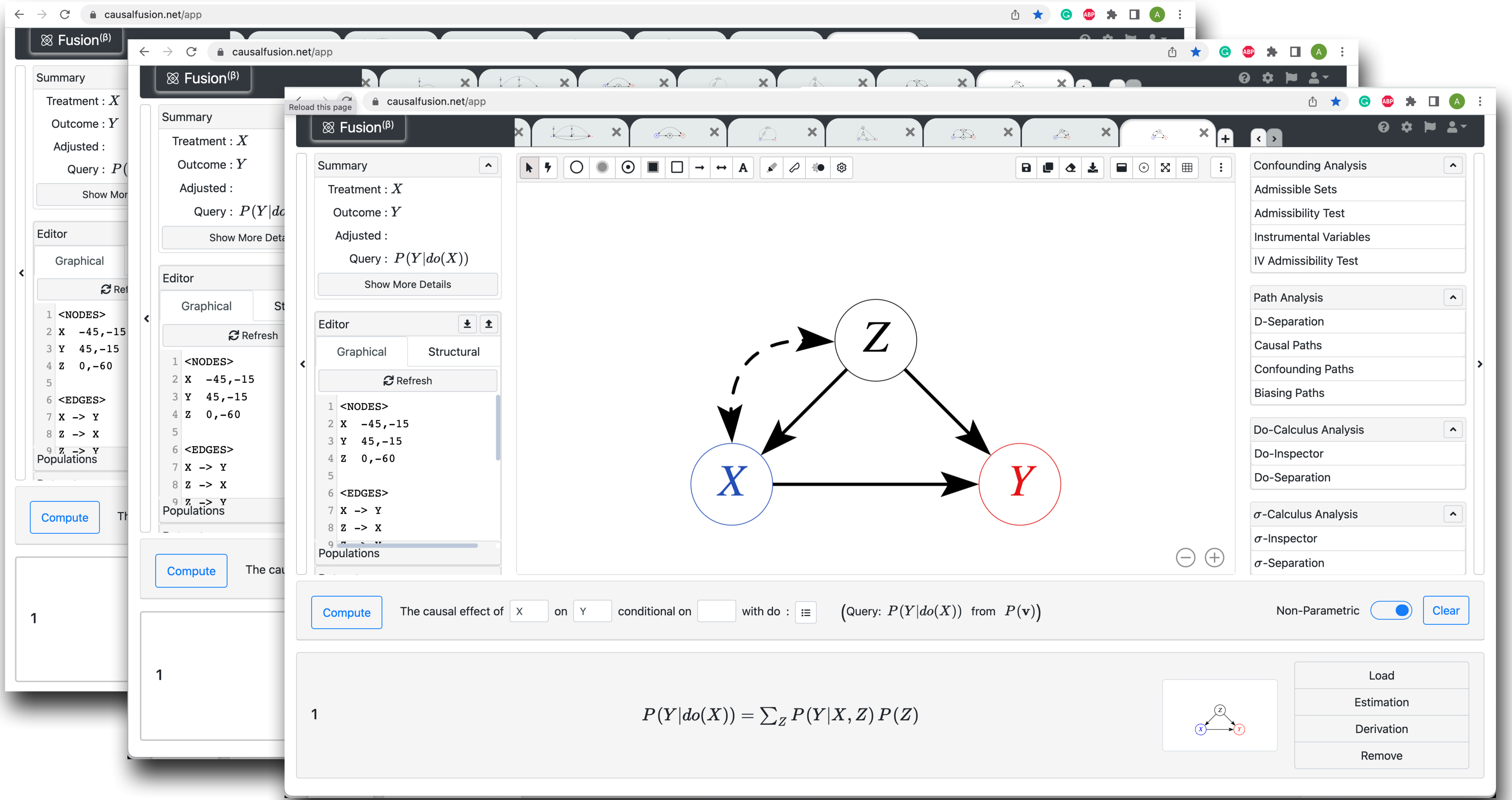
yes / no

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

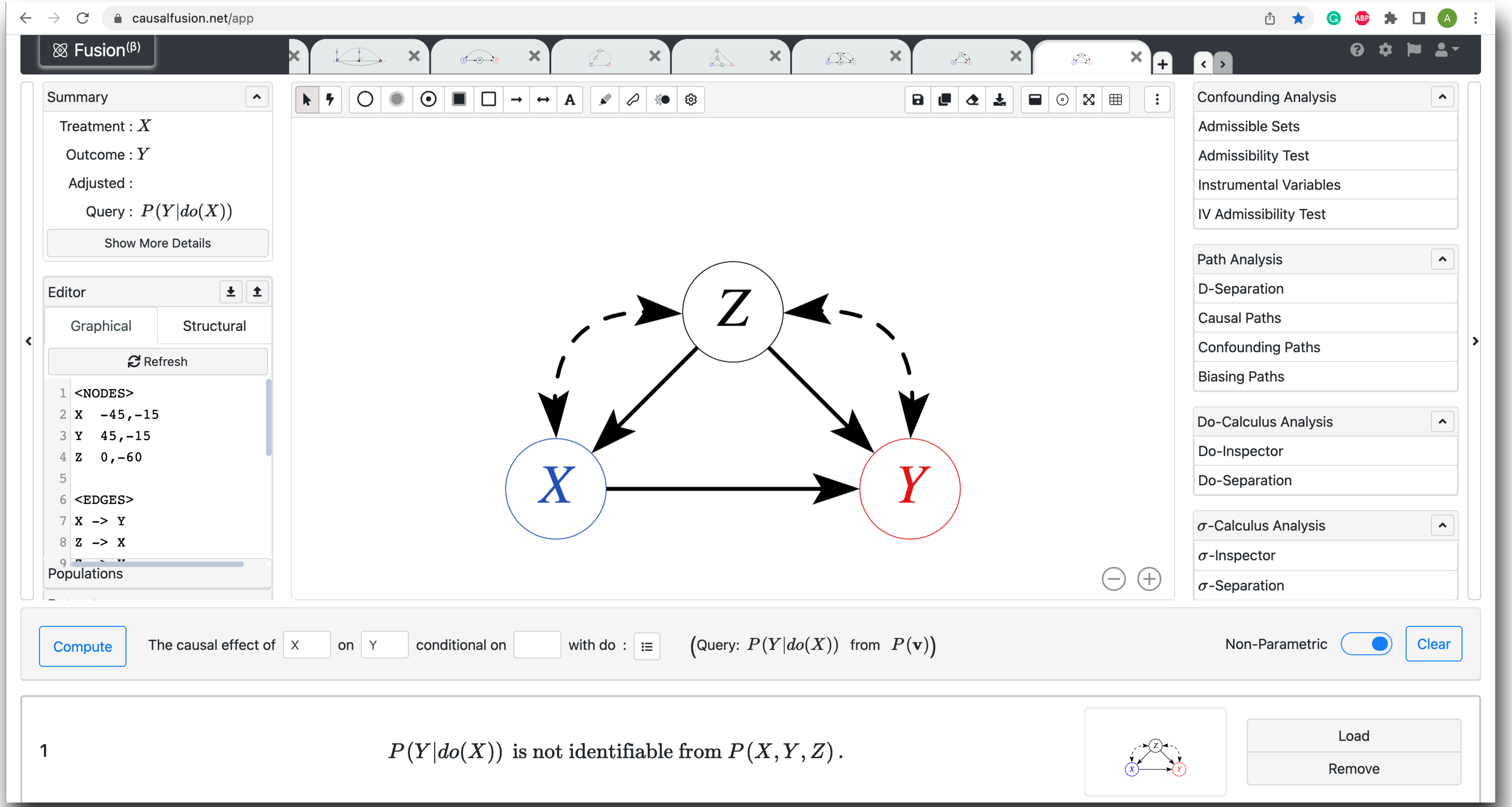
Interventional Distribution ← Available Distributions

- 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.

http://causalfusion.net



http://causalfusion.net



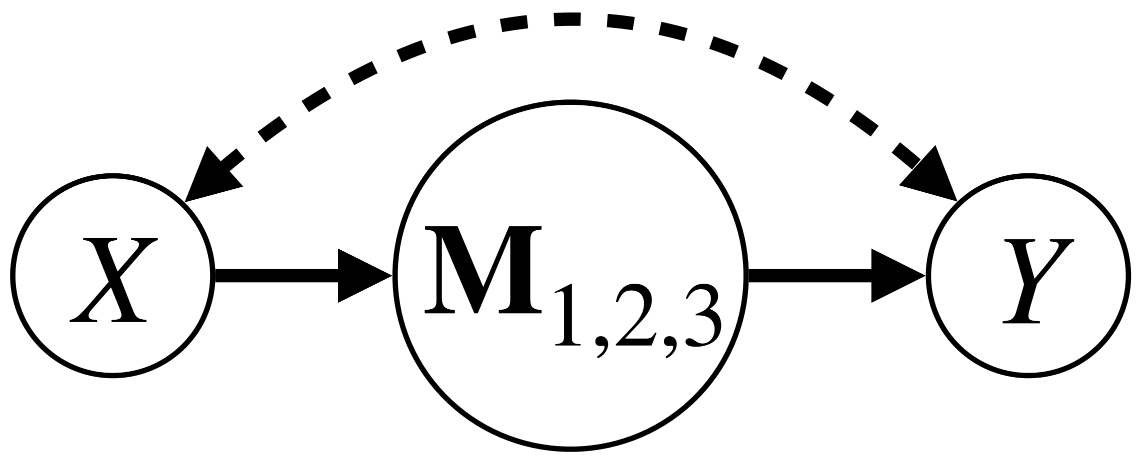
Identification of Causal Effects from C-DAGs

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

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

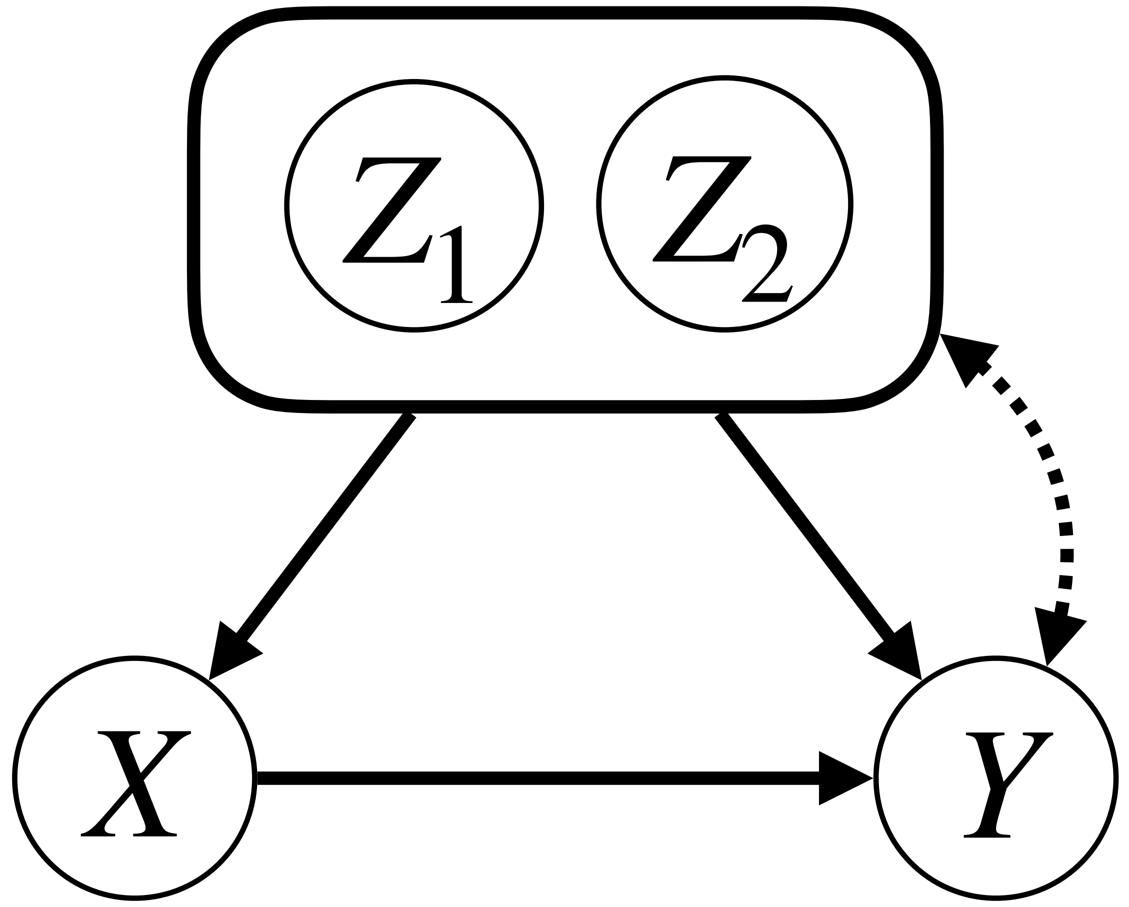
Inferred (Interventional) Distribution ← Available (Observational) Distribution 👍

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.

Effect Identifiabiliy given a C-DAG

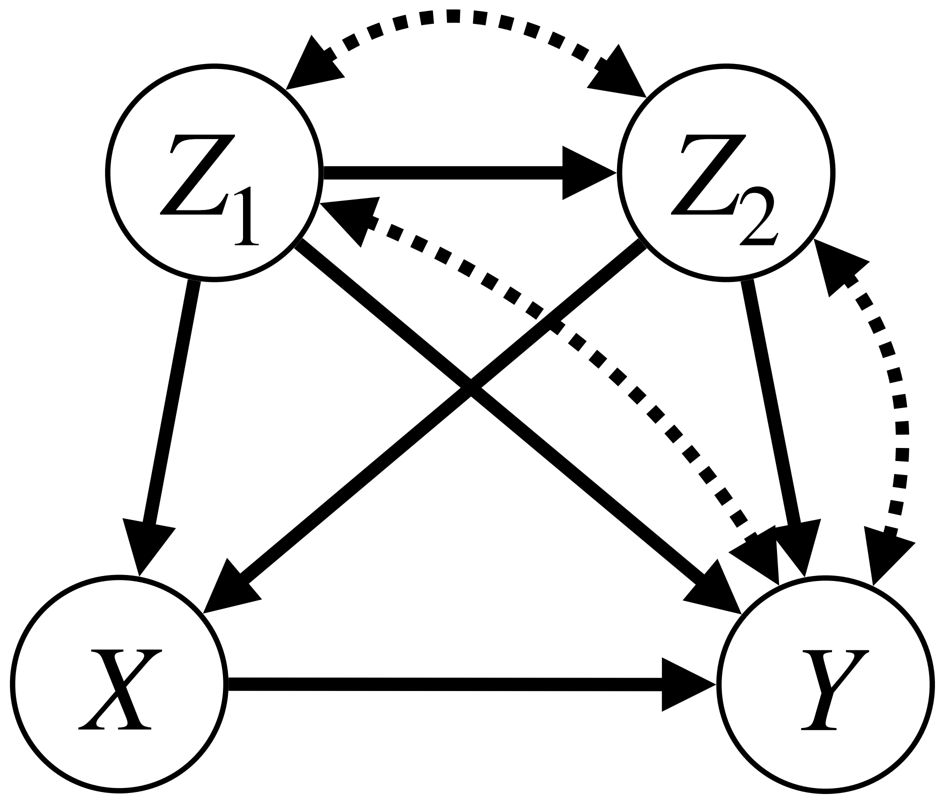
Simple evaluation of the **validity** of the Backdoor Criterion / Conditional Exchangeability

$G_{\mathbf{C}}$



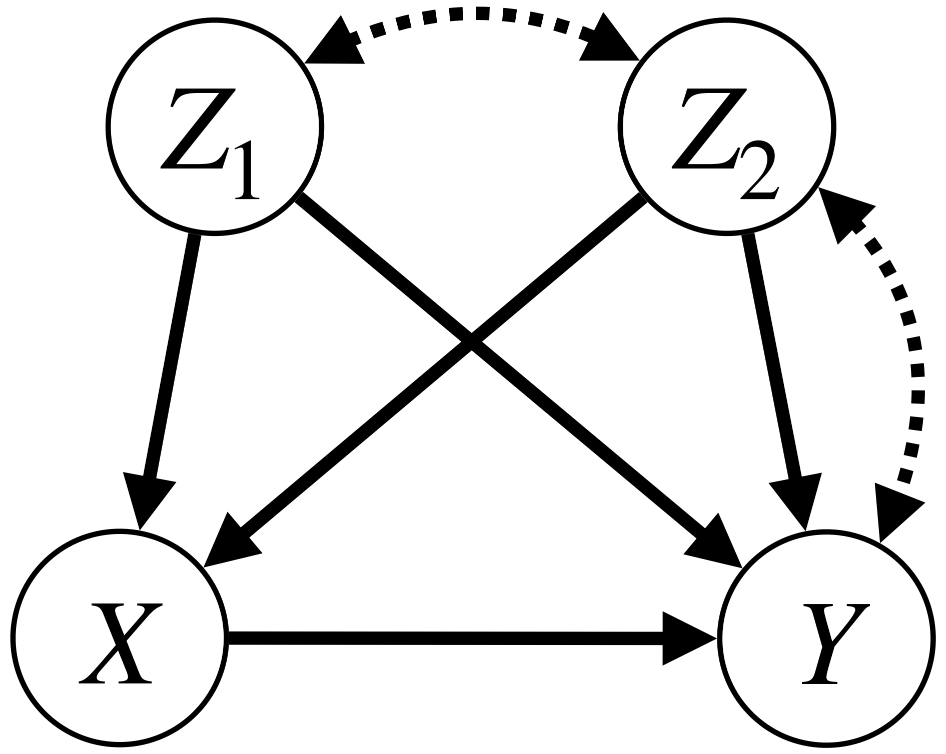
$$P(y \mid do(x)) = \sum_{\mathbf{z}} P\left(y \mid x, \mathbf{z}\right) P\left(\mathbf{z}\right)$$

G_1



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

G_2



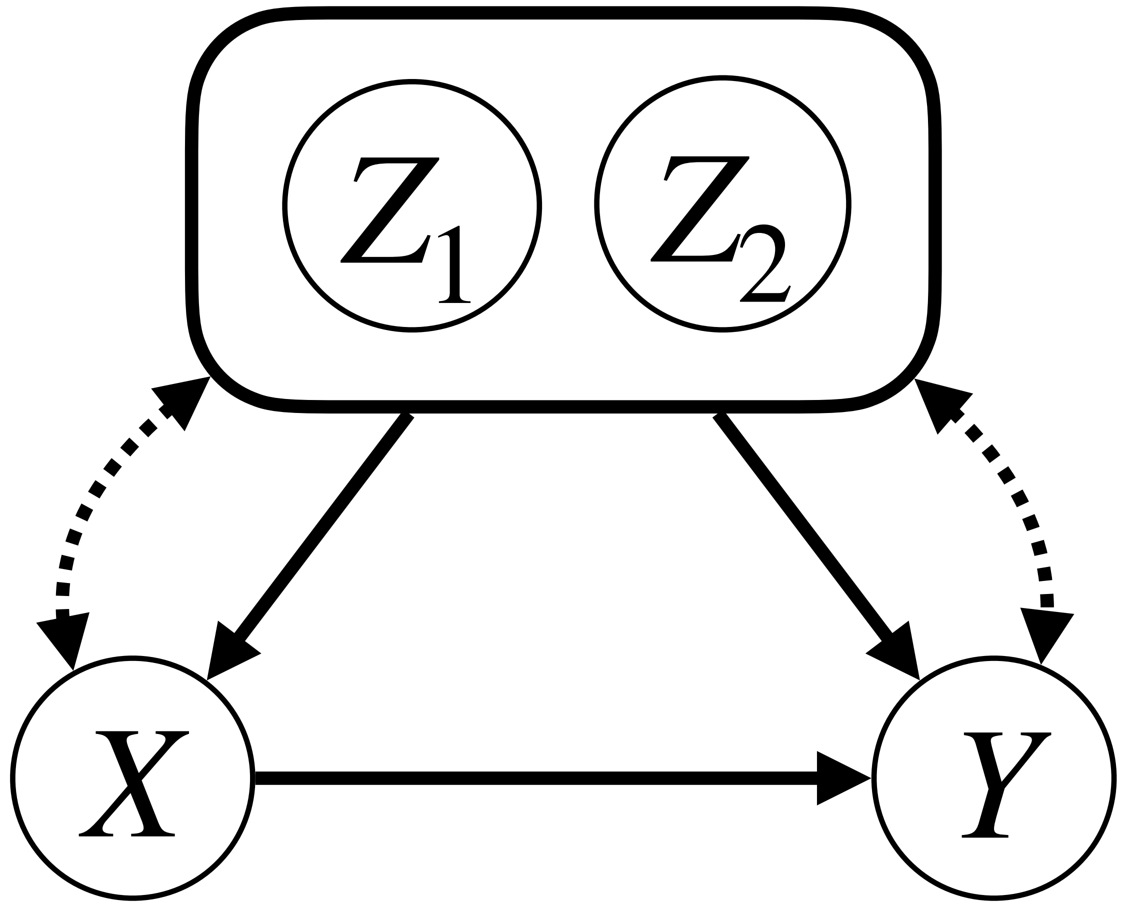
$$P(y \mid do(x)) = \sum_{z_1, z_2} P(y \mid x, z_1, z_2) P(z_1, z_2)$$

An identifiable effect in a C-DAG $G_{\mathbf{C}}$ is identifiable in all compatible causal diagrams G using the same identification formula!

Effect Non-Identifiabiliy given a C-DAG

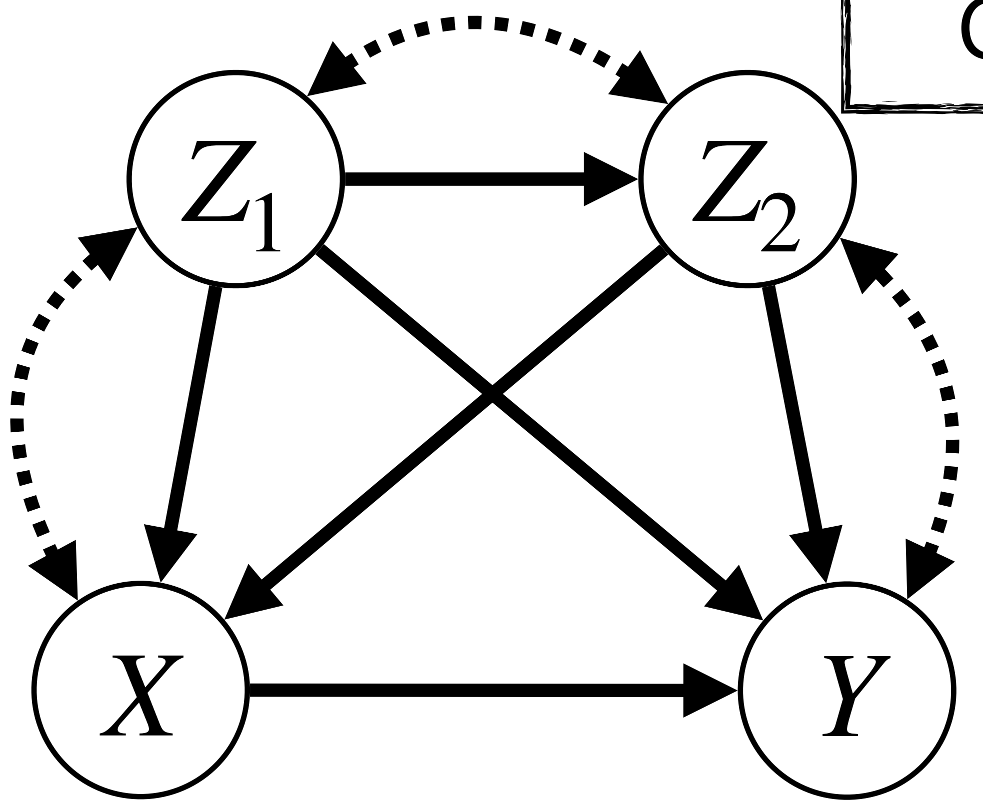
Simple evaluation of a **violation** of the Backdoor Criterion / Conditional Exchangeability

$G_{\mathbf{C}}$



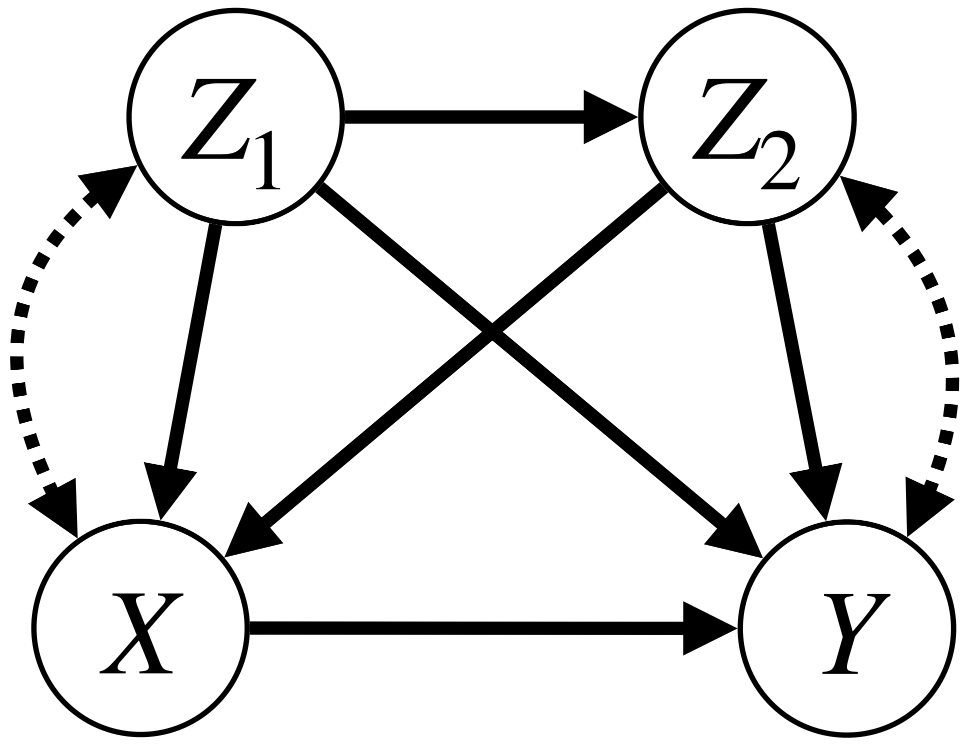
$P(y \mid do(x))$ is not identifiable

G_1



$P(y \mid do(x))$ is not identifiable

G_2



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

A non-identifiable effect in a C-DAG $G_{\mathbf{C}}$ implies that there exists at least one compatible causal diagrams G in which the effect is non-identifiable.

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.

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.

General graphical counterfactual identification:

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

Correa, J. D., & Bareinboim, E. (2025). Counterfactual graphical models: Constraints and inference.
In *Forty-second International Conference on Machine Learning*.

Advances on Effect Identification given a Causal Diagram

Identification and Estimation via Deep Neural Networks:

Xia, K., & Bareinboim, E. (2024, March). Neural causal abstractions. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 38, No. 18, pp. 20585-20595).

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.

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.

Partial Effect Identification:

Kirtan Padh, Jakob Zeitler, David Watson, Matt Kusner, Ricardo Silva, Niki Kilbertus. (2022). Stochastic Causal Programming for Bounding Treatment Effect. Proceedings of the Second Conference on Causal Learning and Reasoning, PMLR 213:142-176

Zhang, J., Tian, J. & Bareinboim, E.. (2022). Partial Counterfactual Identification from Observational and Experimental Data. Proceedings of the 39th International Conference on Machine Learning.

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)

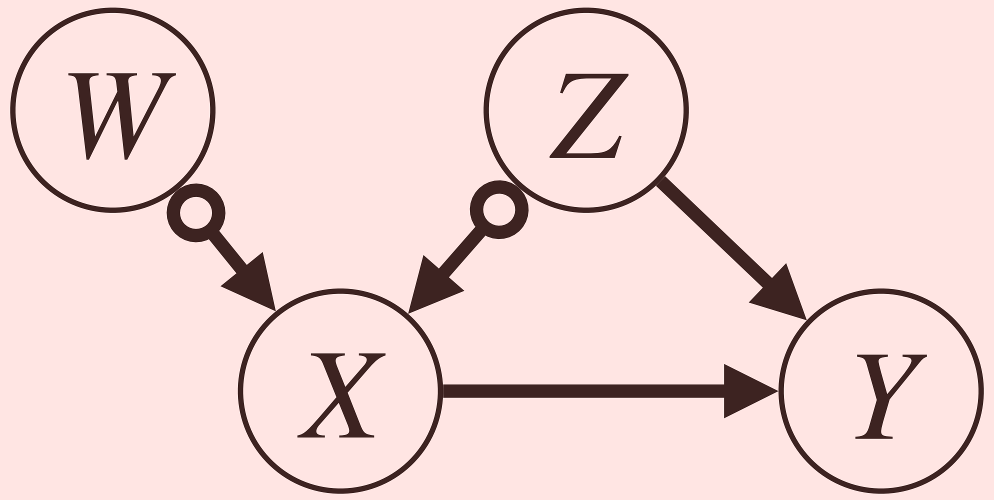
Effect Identification in Markov Equivalence Classes

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 👍

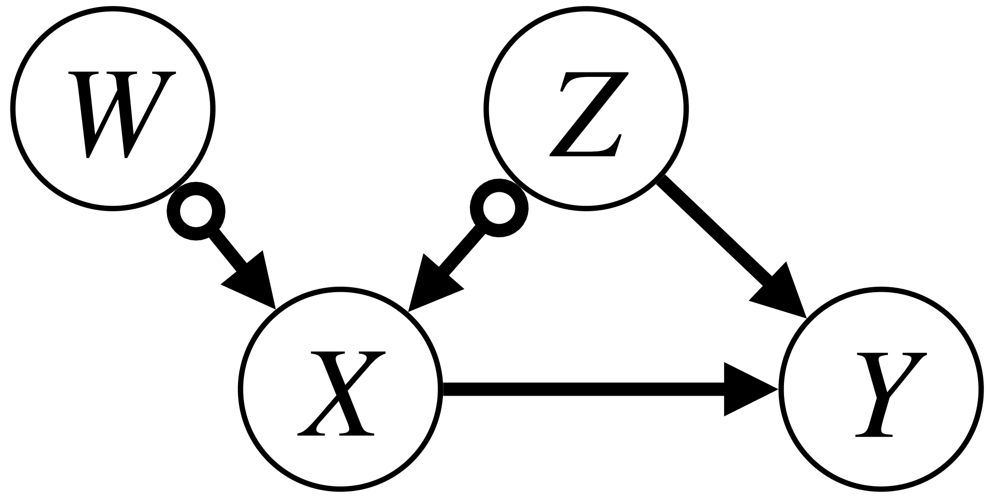
Identification via Adjustment in Markov Equivalence Classes

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

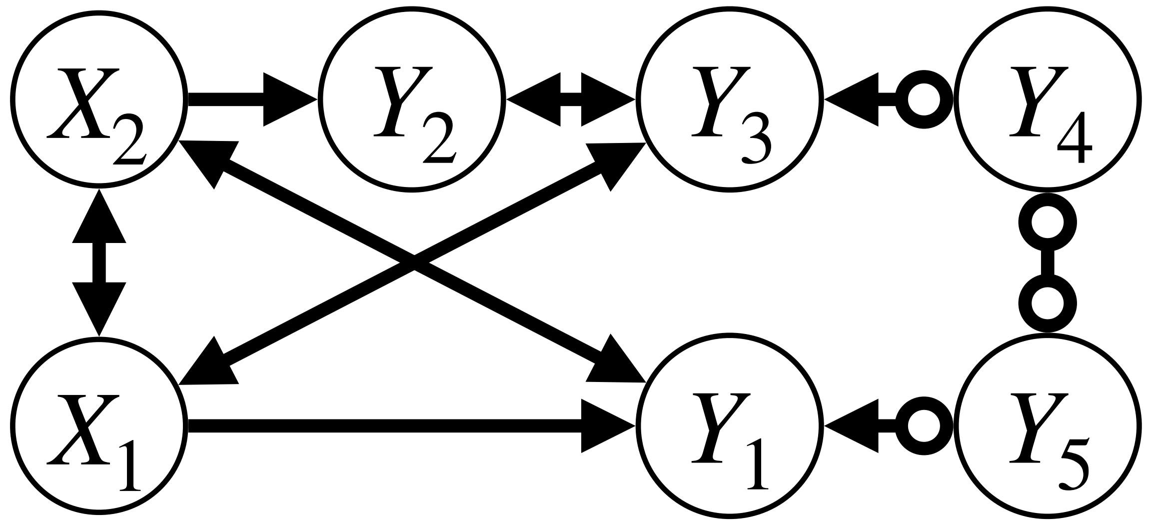
General Identification in Markov Equivalence Classes

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).

Gut Microbiota's Causal Role in Major Depressive Disorder

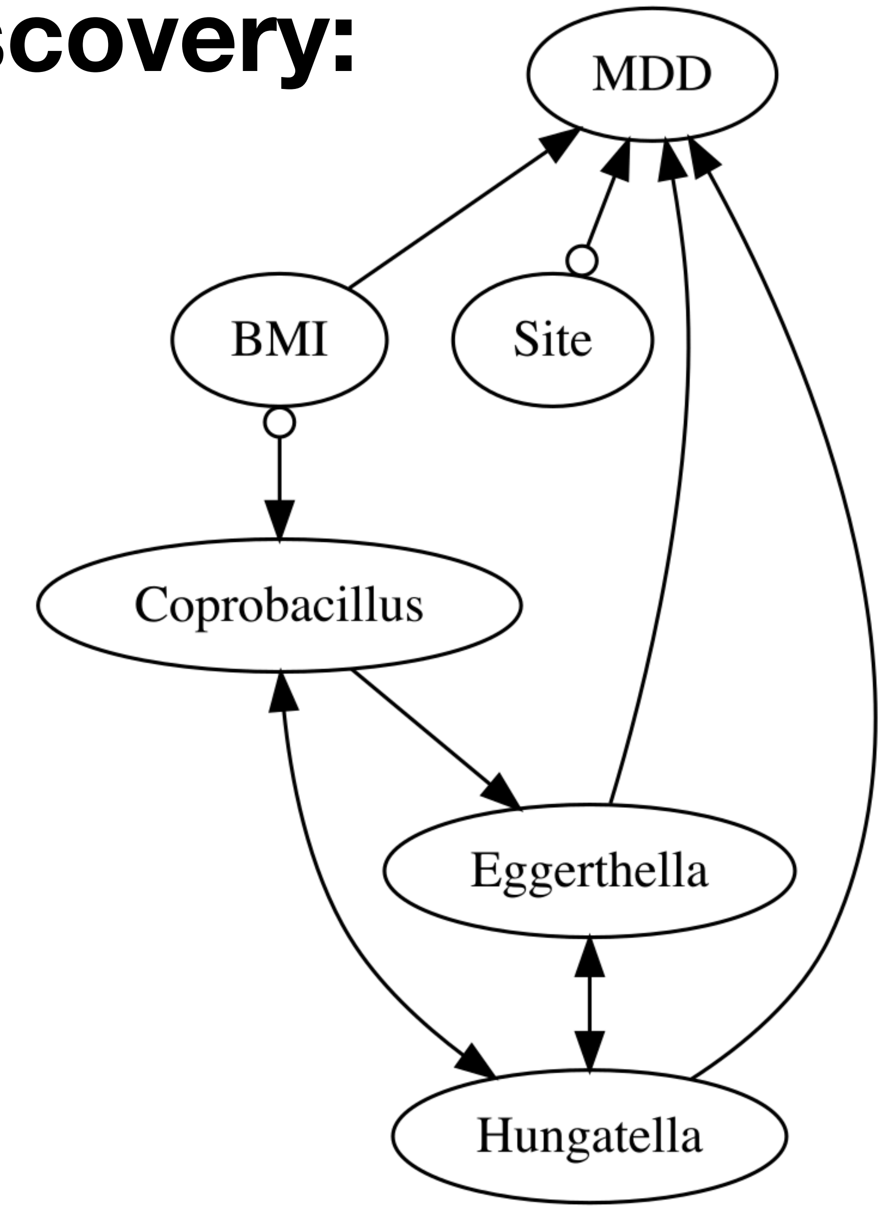
DFG FOR2107 dataset, including microbiome and clinical data from 1,269 patients.

Differential Abundance Analysis:

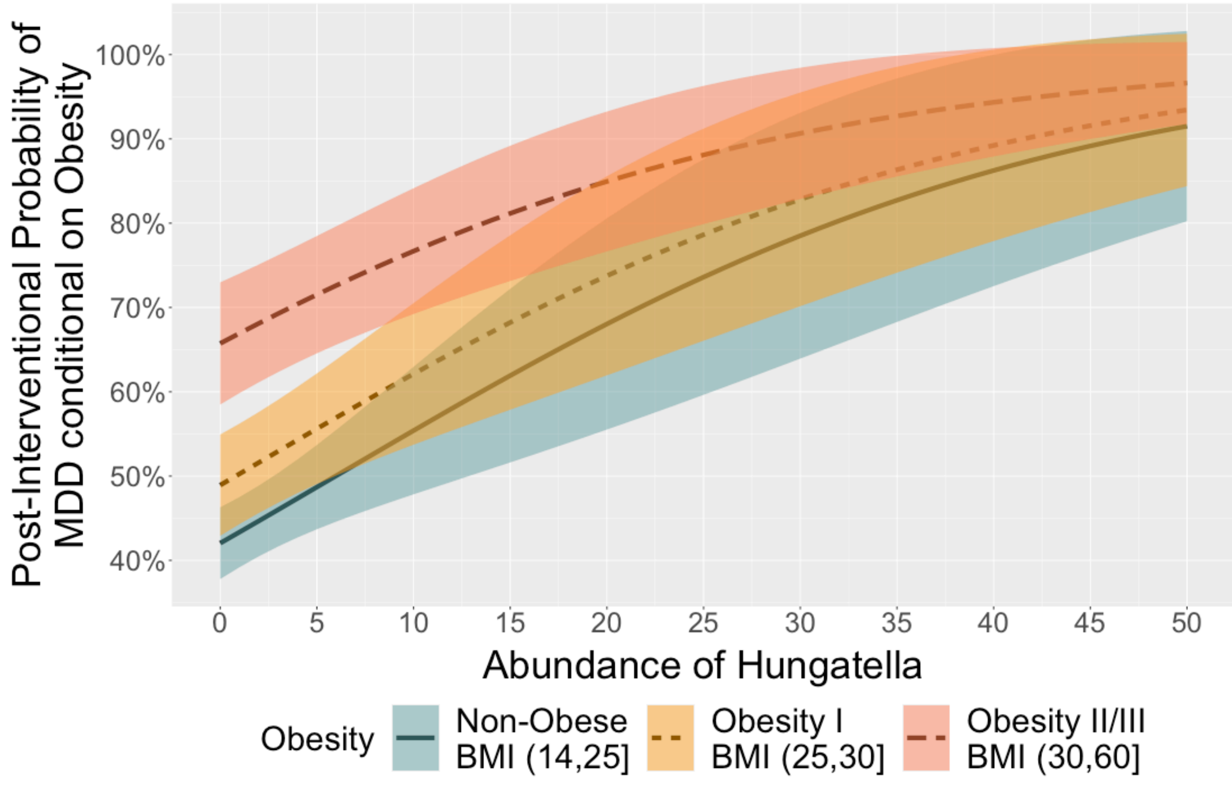
Genus	FDR-corr. p-values	
	LinDA	ZicoSeq
Hungatella	0.0002	0.0071
Eggerthella	0.0063	0.0071
Coprobacillus	0.0070	0.0071
Lachnospiraceae FCS020 group	0.0063	0.011

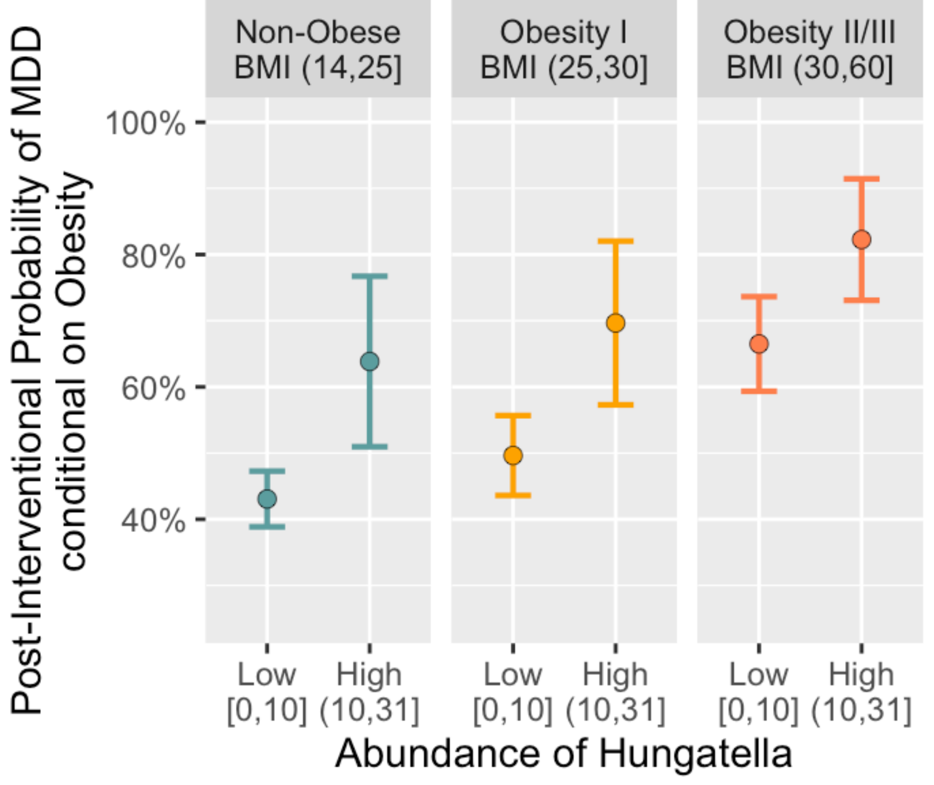
Causal Discovery:

FCI with a robustness-enhancing strategy.



Obesity-specific causal effect of *Hungatella* on MDD

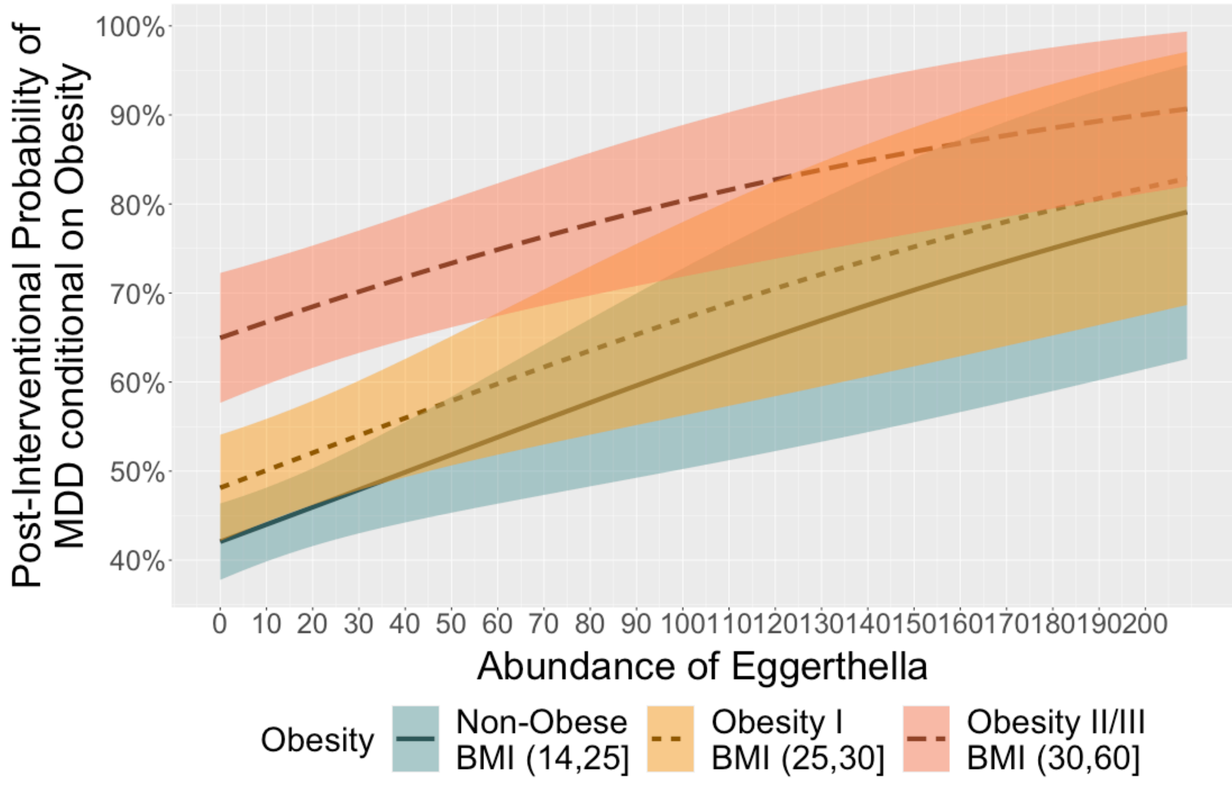


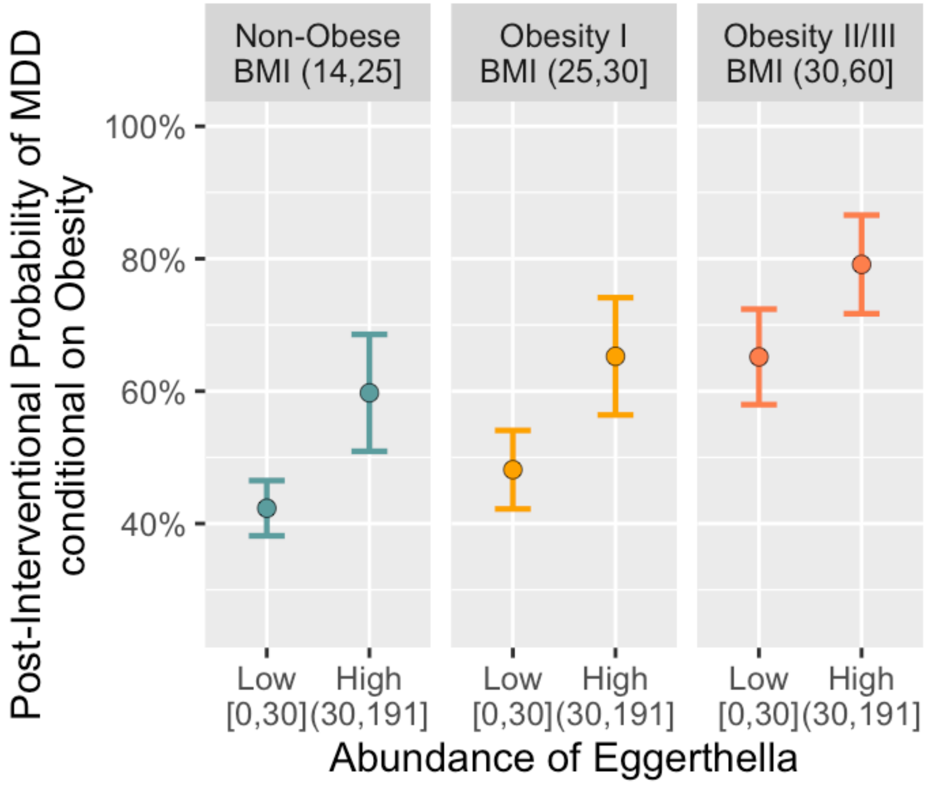


Obesity	Effect	95% C.I.	
No	0.2078	0.07230	0.3433
I	0.2002	0.06261	0.3379
II/II	0.1577	0.04138	0.2741

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

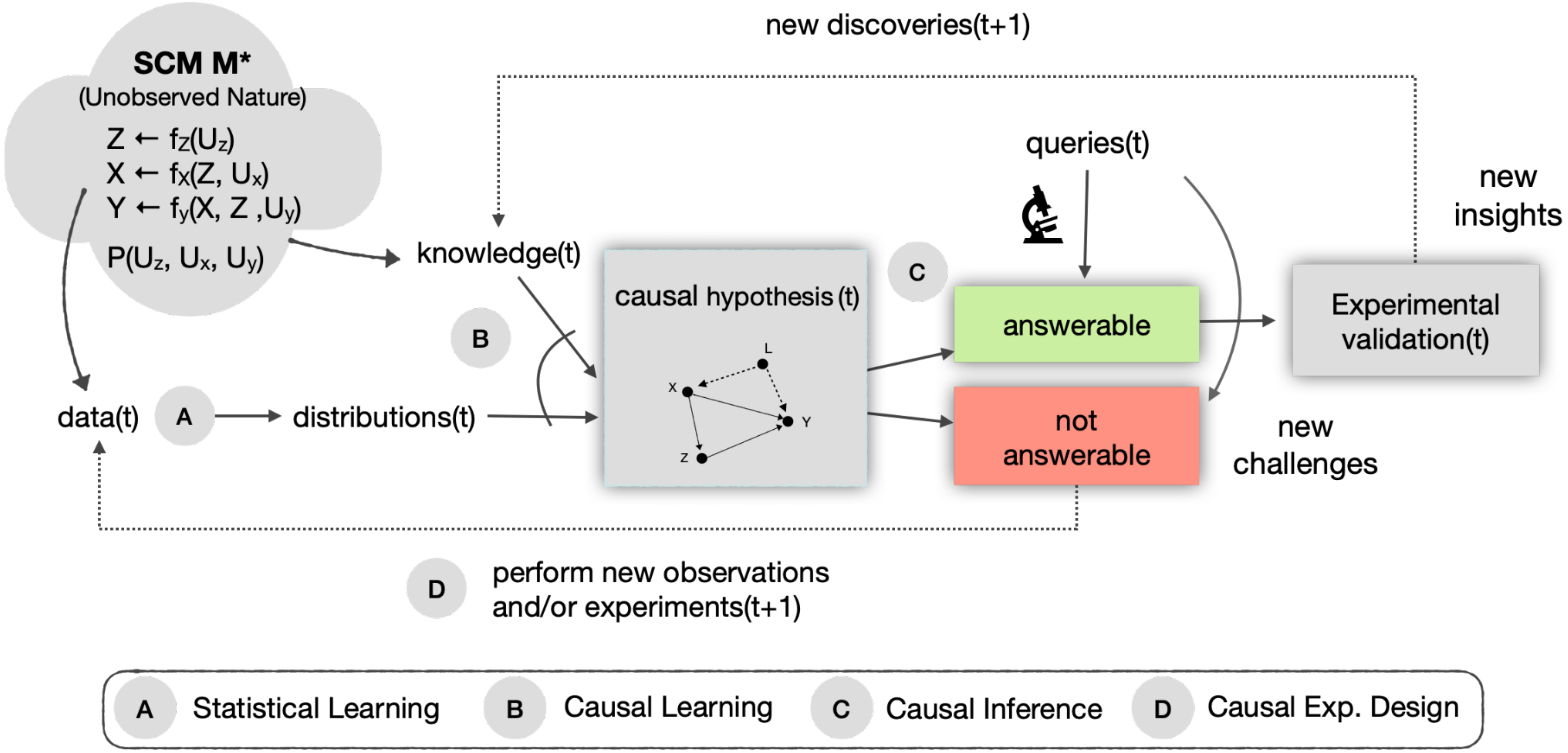
ACE: 0.1683
95% CI: [0.06812, 0.2684]

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

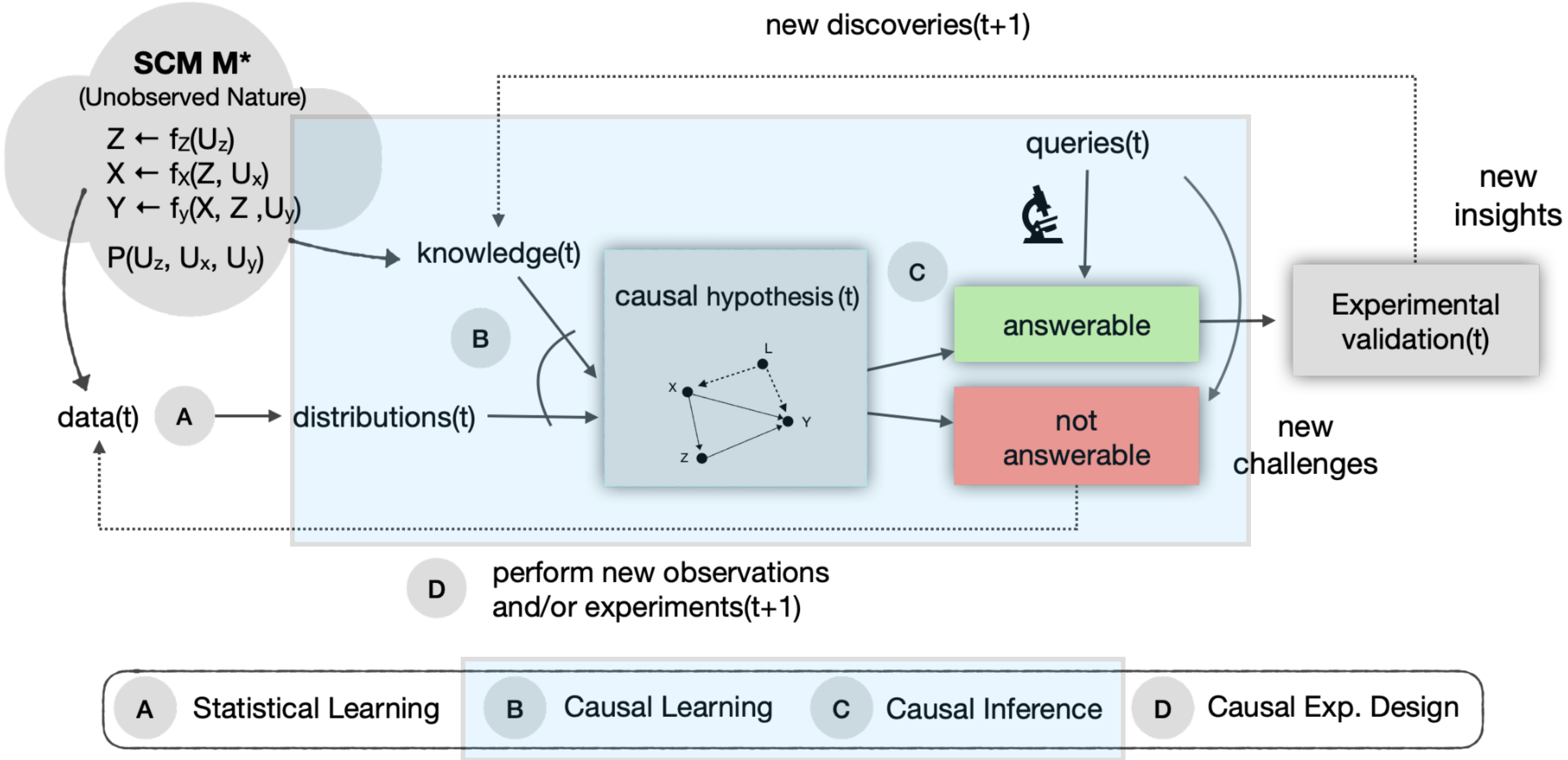
Causal Inference Workflow

Continuous Process of Scientific Discovery and Causal Hypothesis Refinement



Causal Inference Workflow

Continuous Process of Scientific Discovery and Causal Hypothesis Refinement



Many other Topics in Causal Inference

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

I am happy to discuss more if you are interested! :)

Educational Resources

- Causality Lectures on YouTube : @adelehelena
  - Complete course (13 lectures at HHU): Playlist
  - 3-hour tutorials at summer schools:
    - Lisbon Machine Learning School (LxMLS): Playlist (2021-2025)
    - Nordic Probabilistic AI School (ProbAI): Playlist (2023-2024)
    - European Summer School on Artificial Intelligence (ESSAI 2024): Playlist

- Tutorial on GitHub : @adele → Causality-Tutorial
  - Causal Discovery — Google Colab Notebook: Link
  - Causal Effect Identification — Google Colab Notebook: Link

Thank you! :)

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

adele.ribeiro@uni-muenster.de