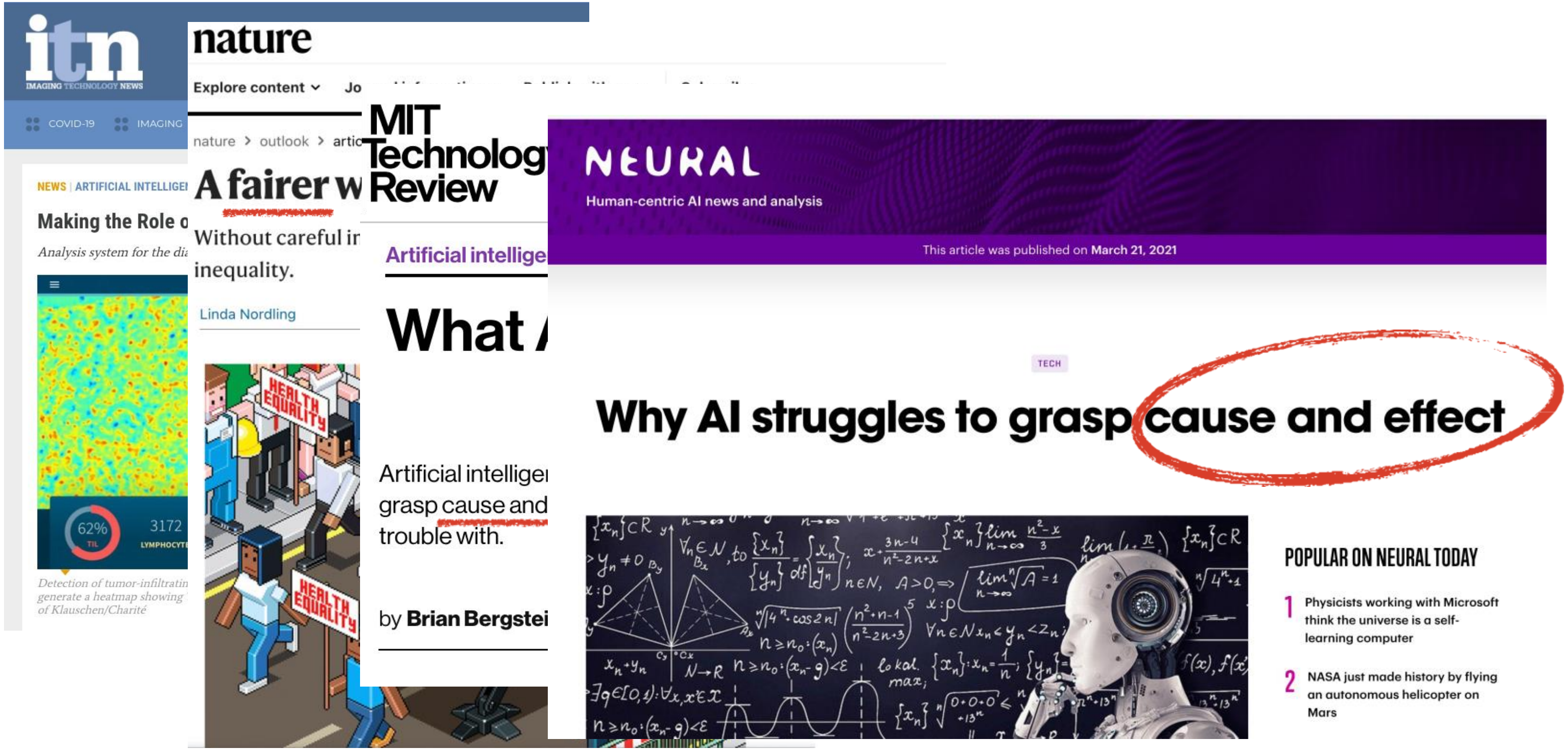
Putting the Causality in Continual Causality

Part I: The Ultimate Hitchhiker's Introduction to Causality for ML

Devendra Singh Dhami

TECHNISCHE UNIVERSITÄT DARMSTADT

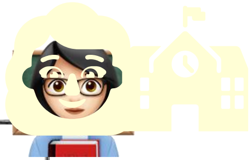
Current Challenges in DS, AI, ML



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Current Challenges in DS, AI, ML

- Data-hungry & sample inefficiency
- Lack of interpretability & explainability
- Lack of robustness & generalizability
- Unfair & unethical decision-making



Lack of Causal Inference Capabilities

What can we achieve with causality?

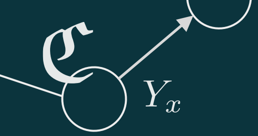
Data Fusion: provides language and theory to cohesively combine prior knowledge and data from multiple and heterogeneous studies.

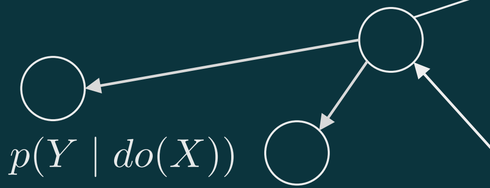
Effect identifiability: can determine the effect of unrealized interventions rather than just predicting an outcome (i.e., can distinguish between association and causation)

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

Explainability: provides a better understanding of the underlying mechanisms.

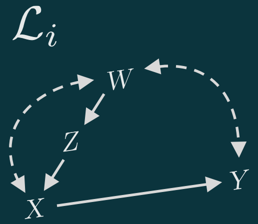
Fairness: captures and disentangles any mechanisms of discrimination that may be present, including direct, indirect-mediated, and indirect-confounded.





I | Why?

Machine Learning needs Causality?

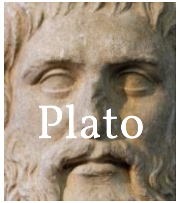


CAUSE

TECHNISCHE UNIVERSITÄT DARMSTADT

What is Causality?

We might want to start here first..

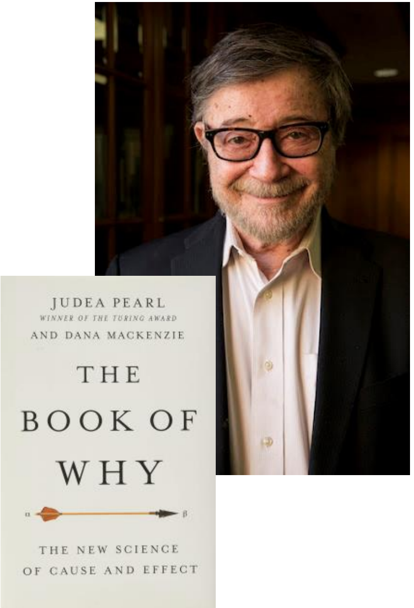


Probably, Plato was the first to state the principle of causality:

"Everything that becomes or changes must do so owing to some cause; for nothing can come to be without a cause."- *Timaeus* 28a

Judea Pearl’s opinion

Pioneer of Causality for AI, Turing awardee



“To Build Truly Intelligent Machines,
Teach Them Cause and Effect”

“All the impressive achievements of deep learning
amount to just curve fitting”

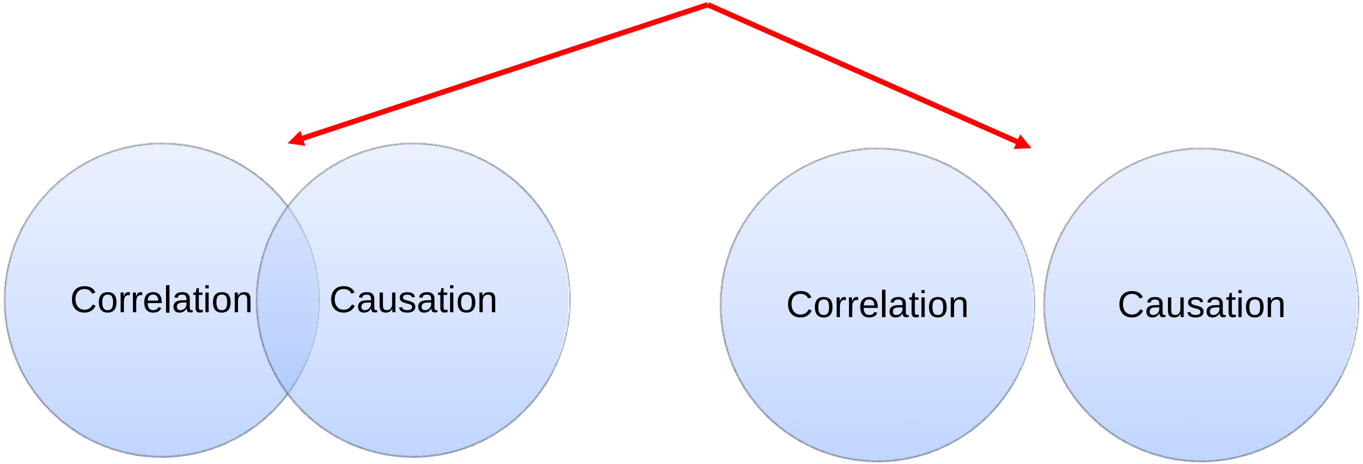
Judea Pearl in “The Book of Why”
and in an interview with quanta magazine in 2018

CAUSE

TECHNISCHE UNIVERSITÄT DARMSTADT

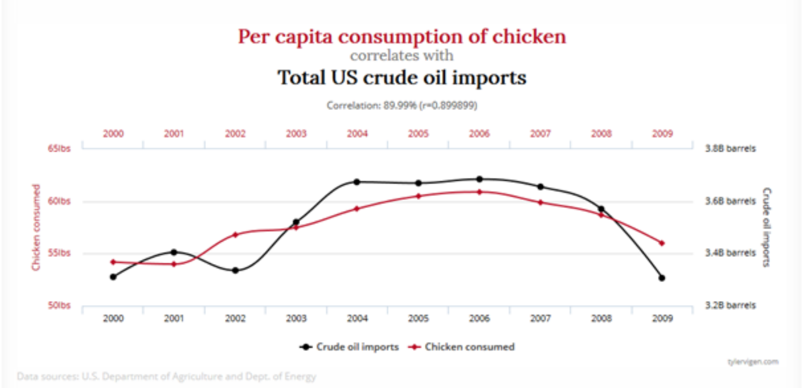
We have All Heard the Phrase

"Correlation does not imply causation"



Pic Credit: Wikipedia

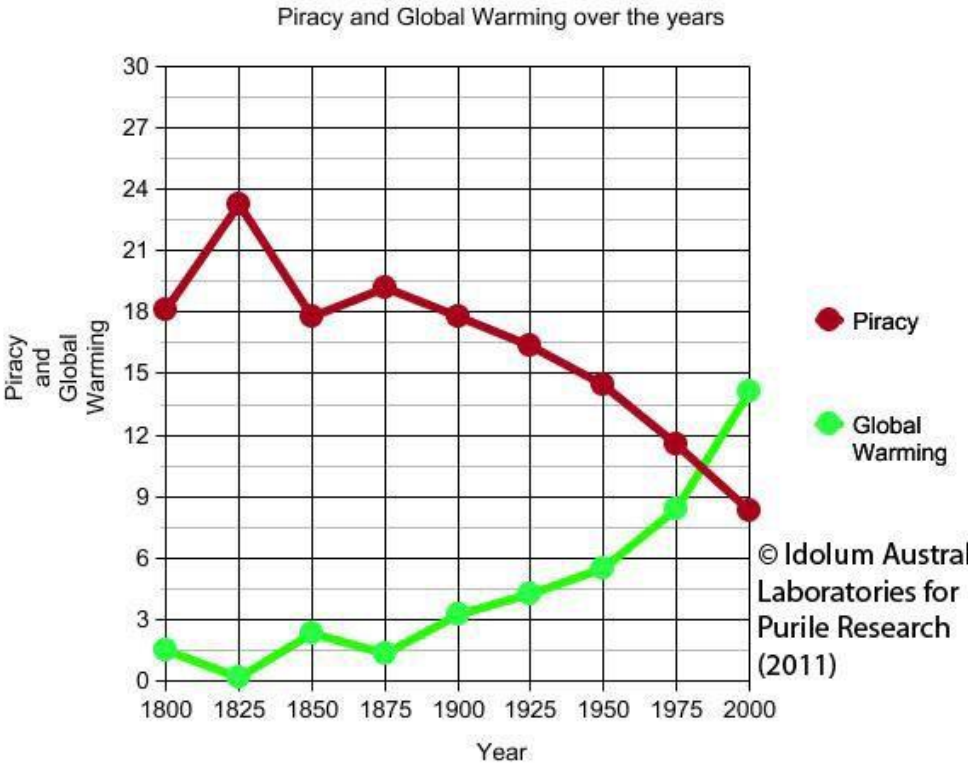
Correlation ∩ Causality = Ø



CAUSE

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Correlation => Causality via a third factor

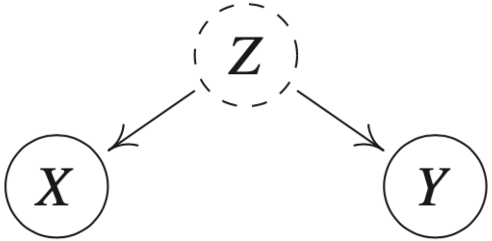


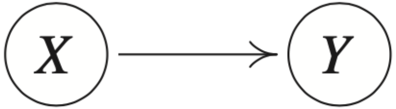
© Idolum Australian Laboratories for Purile Research (2011)

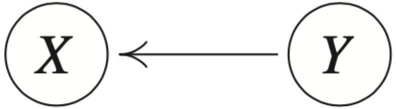
Reichenbach's Common Cause Principle

Defining Confounders

Principle 1. *If two random variables X and Y are statistically dependent ($X \not\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp Y \mid Z$.*



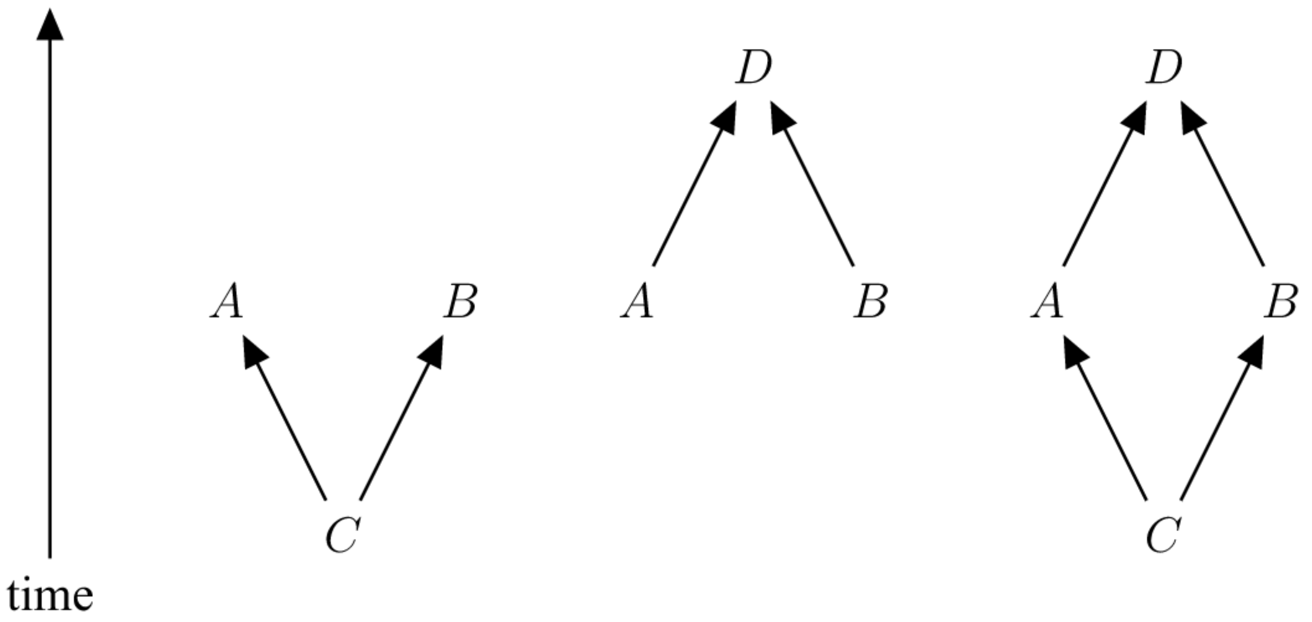




Reichenbach's *Direction of Time* (1956)

TECHNISCHE UNIVERSITÄT DARMSTADT

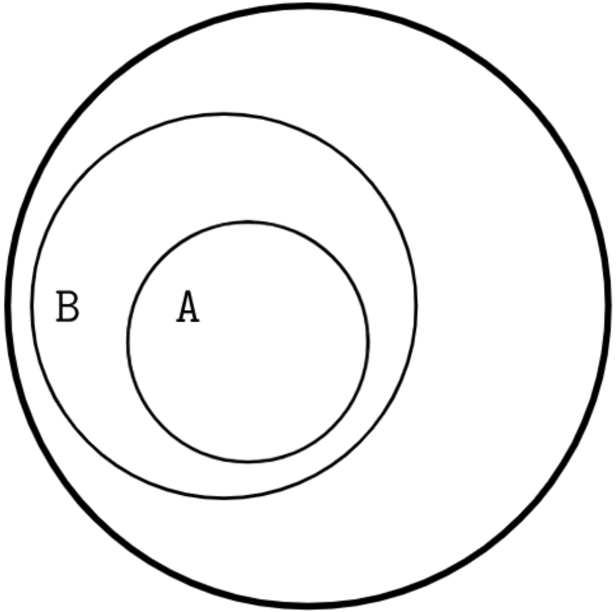
Reichenbach's Common Cause Principle

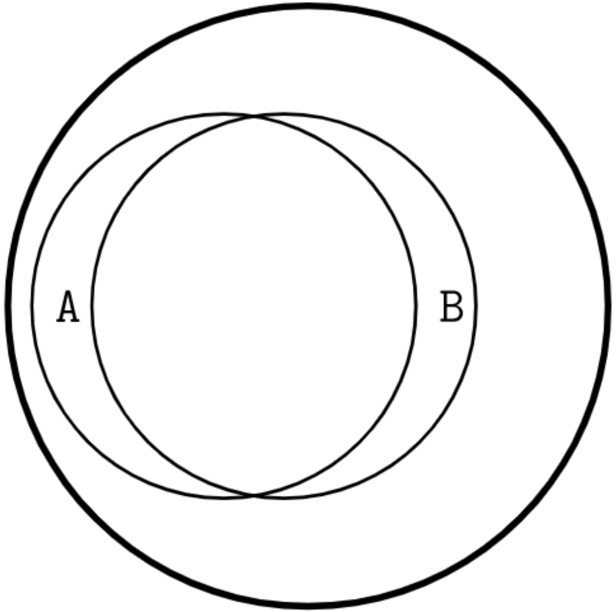


Conjunctive fork: a) open to the future b) open to the past c) closed fork

https://plato.stanford.edu/entries/physics-Rpcc/

Counterexamples to the Reichenbach's Common Cause Principle





https://plato.stanford.edu/entries/physics-Rpcc/

Let's Illustrate Correlation does not imply causation: **Simpson's Paradox**

Condition

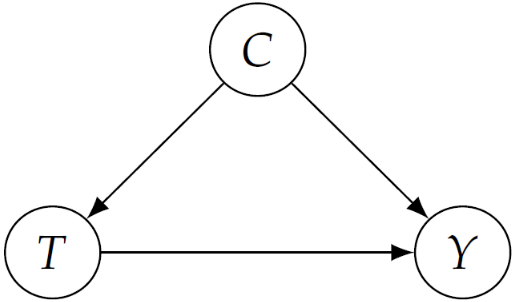
Treatment	Mild	Severe	Total
A	15% (210/1400)	30% (30/100)	**16%** (240/1500)
B	**10%** (5/50)	**20%** (100/500)	19% (105/550)

Credit: Brady Neal Course

more effective treatment is completely dependent on the **causal structure** of the problem

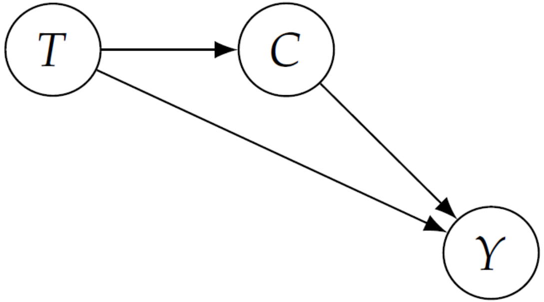
Let's Illustrate Correlation does not imply causation : **Simpson's Paradox**

Scenario 1: Confounders



Treatment A preferable

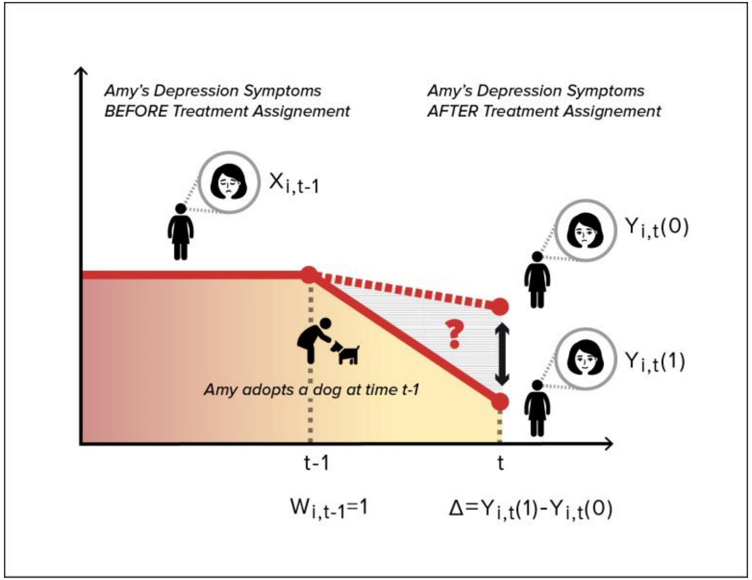
Scenario 2: Treatment causes condition



Treatment B preferable

CAUSE TECHNISCHE UNIVERSITÄT DARMSTADT

2 Fundamental Problems of Causal Inference



Credit: Dominici et al., From Controlled to Undisciplined Data: Estimating Causal Effects in the Era of Data Science Using a Potential Outcome Framework

Definition 1: Individual treatment effect

The individual treatment effect, δ_i, equals $Y_i^1 - Y_i^0$

Definition 2: Average treatment effect (ATE)

The average treatment effect is the population average of all i individual treatment effects

$$
\begin{aligned}
E[\delta_i] &= E[Y_i^1 - Y_i^0] \\
&= E[Y_i^1] - E[Y_i^0]
\end{aligned}
$$

Definition 3: Switching equation

An individual's observed health outcomes, Y, is determined by treatment assignment, D_i, and corresponding potential outcomes:

$$
Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0
$$

$$
Y_i = \begin{cases} Y_i^1 \text{ if } D_i = 1 \\ Y_i^0 \text{ if } D_i = 0 \end{cases}
$$

Definition 4: Fundamental problem of causal inference

It is impossible to observe both Y_i^1 and Y_i^0 for the same individual and so individual causal effects, δ_i, are unknowable.

Credit: Scott Cunningham

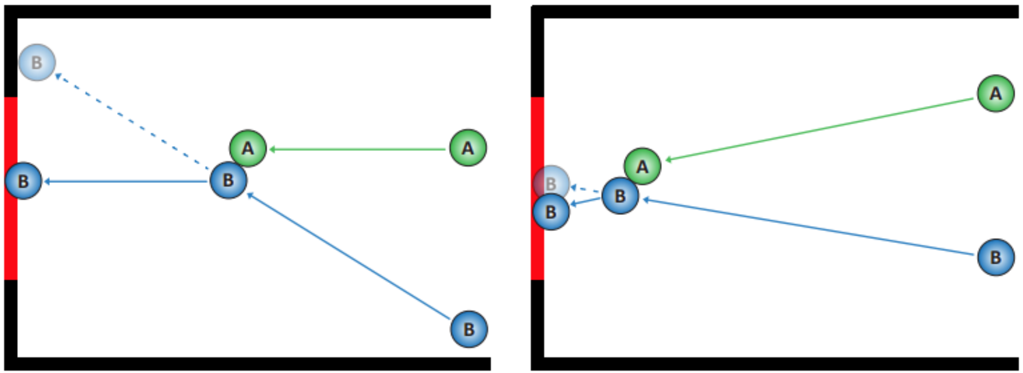
Pearlian Causality

A success story

The formalization with most success in AI/ML so far.

Works in Cognitive Science also in support of the key ideas in the formalism i.e., **humans reason counterfactually**.

(a) A caused B to go into the gate. (b) A didn't cause B to go into the gate.



Gerstenberg. What would have happened?[...] PTRBAE 2022.

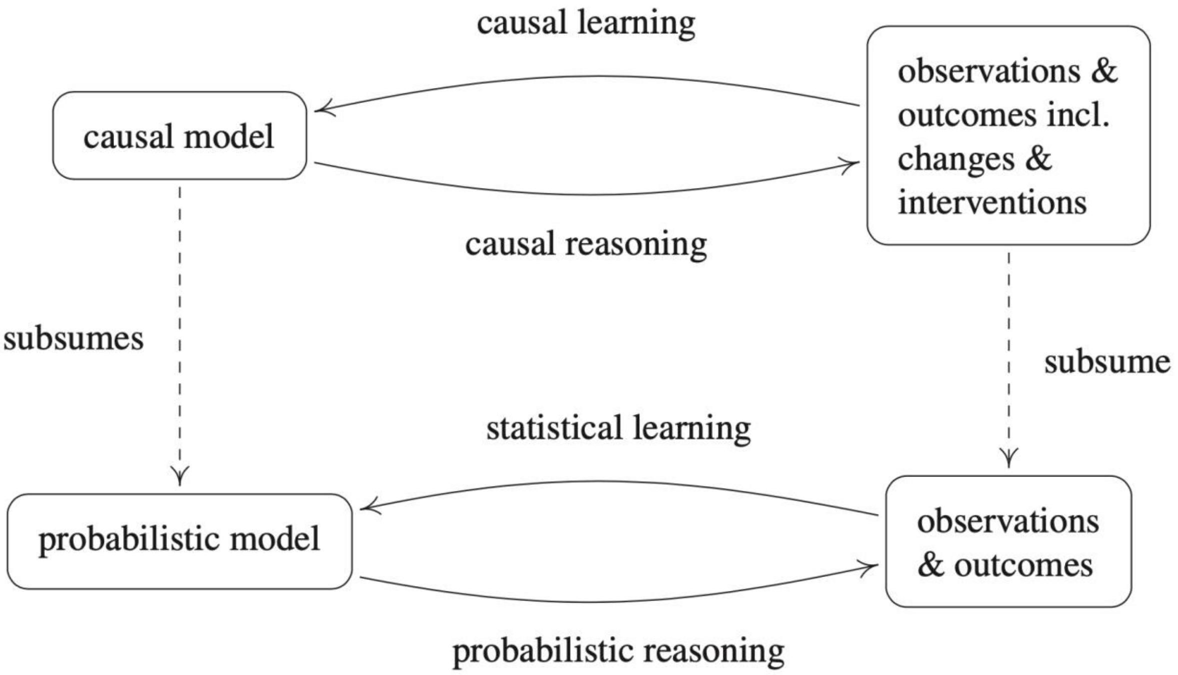
Causality allows us to talk about **modelling assumptions**

Causality allows us to consider not just the joint distribution but the **data generating process** which induces said distribution

2 | What?

Does Pearlian Causality look like?

Causal versus Probabilistic Inference



Peters et al. 2017, Fig.1.1

Pearl Causal Ladder

Level I

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the
election results?

Pearl Causal Ladder

Level II

2. INTERVENTION

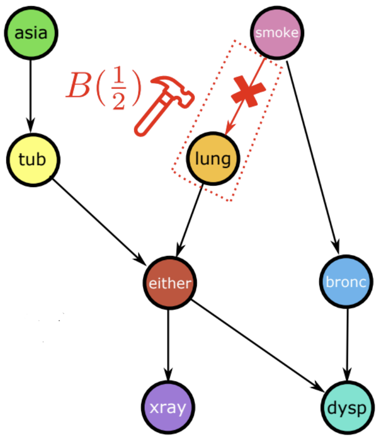
ACTIVITY: Doing, Intervening

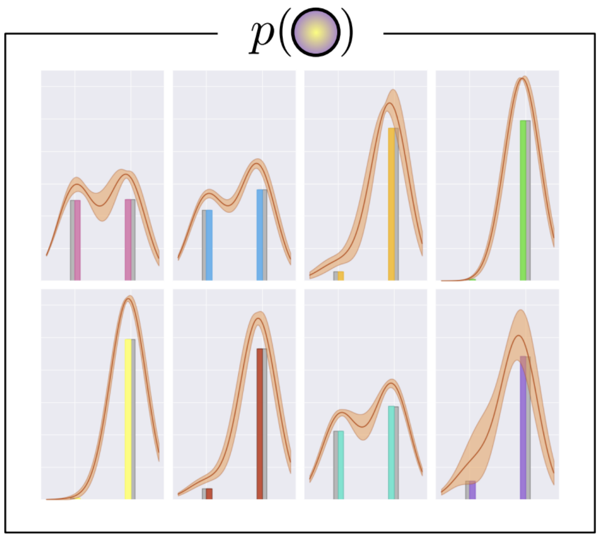
QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

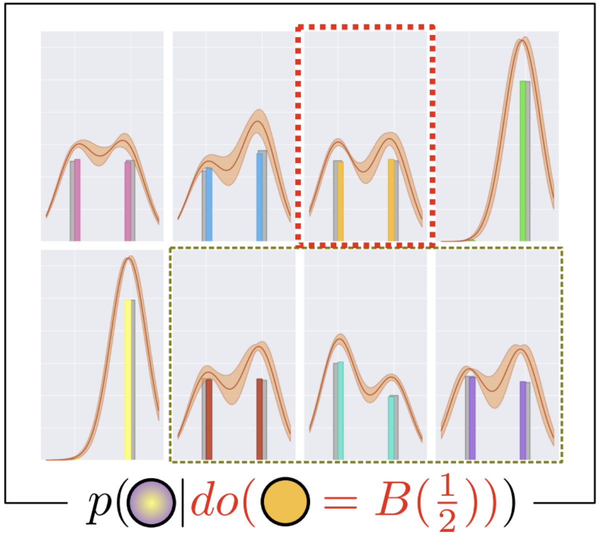
EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

Pearl Causal Ladder

Level II







Pearl Causal Ladder

Level III

3. COUNTERFACTUALS

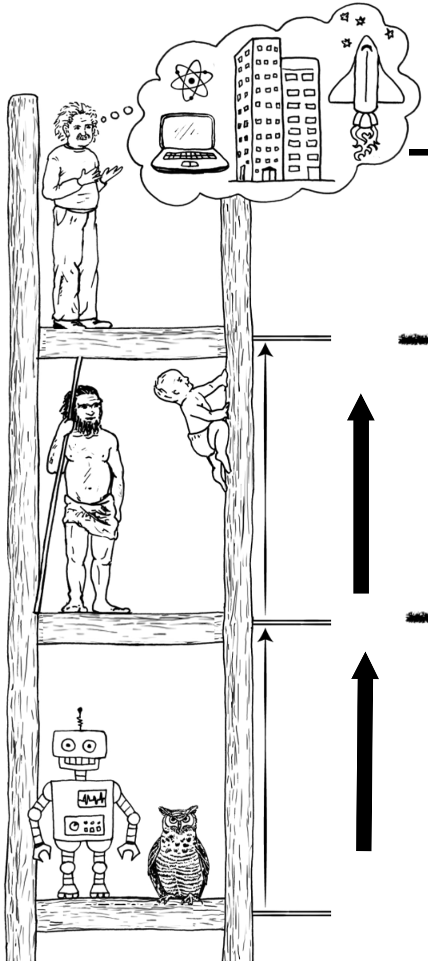
ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done …? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

Pearl Ladder of Causation

* 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



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

TECHNISCHE UNIVERSITÄT DARMSTADT

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

$\mathbf{V}$ are endogenous variables

$\mathbf{U}$ are exogenous variables

$\mathcal{F}$ are functions determining V i.e., $v_i = f_i(\mathbf{pa}_i, \mathbf{u}_i)$

$P_{\mathbf{U}}$ is the probability distribution over U.

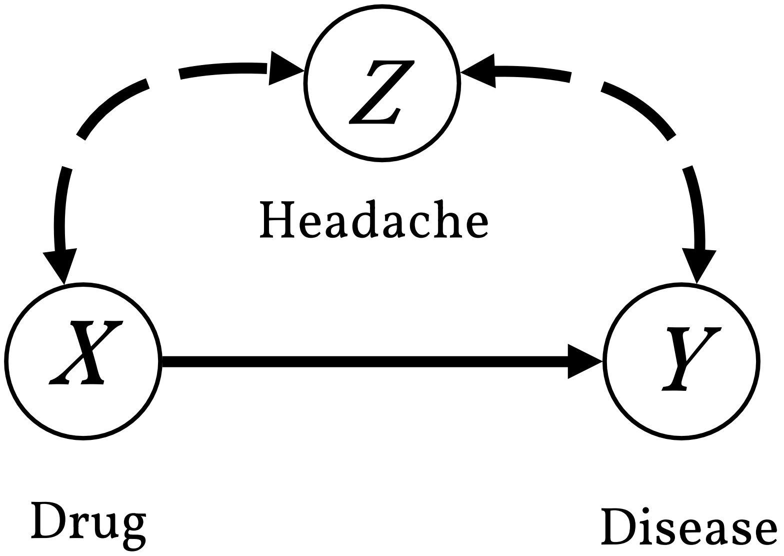
Assumption: $\mathcal{M}$ is recursive i.e., there are no feedback (cyclic) mechanisms

The Causal Graph

An induced property of the SCM

latent

$$\mathcal{F} = \begin{cases} X = f_X(U_X, U_{XZ}) \\ Y = f_Y(X, U_Y, U_{YZ}) \\ Z = f_Z(U_Z) \end{cases}$$

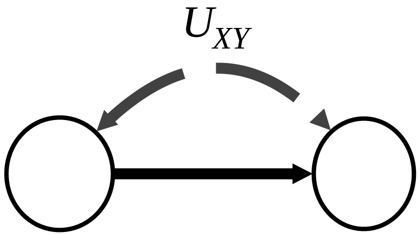


Graphical Representation of an SCM

Structural Causal Model (SCM) | **Graphical Causal Model (Causal Diagram)**

Observational

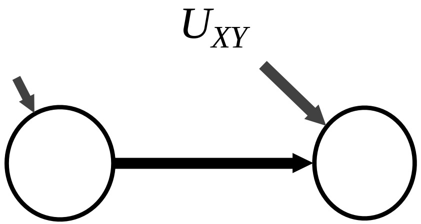
$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} \boxed{X = f_X(U_X, U_{XY})} \\ Y = f_Y(X, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$



$do(X = x)$

Interventional $do(X = x)$

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



TECHNISCHE UNIVERSITÄT DARMSTADT

What is induced by the SCM?

Observational SCM

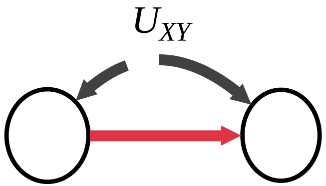
$$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} X = f_X(U_X, U_{XY}) \\ Y = f_Y(X, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$$

$do(X = x)$

Interventional SCM

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

Loss of Information



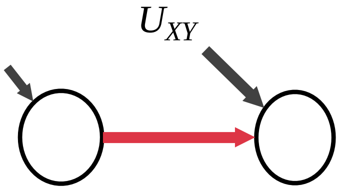
Observational Causal Diagram

$$P(\mathbf{V})$$

Observational Distribution

Observational Data

Loss of Information



Interventional Causal Diagram

$$P(\mathbf{V} \mid do(x))$$

Interventional Distribution

Interventional Data

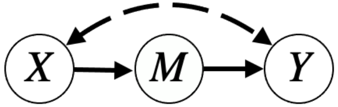
Feasible Cross-layer Inference

Via Constraints on Causal Graph

Using Observational Data From One Population/Domain

1. **Query**
   $P(y \mid do(x))$

2. **Causal Contraints**



3. **Data**
   $P(x, m, y)$

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

Counterfactuals

A 3-step procedure

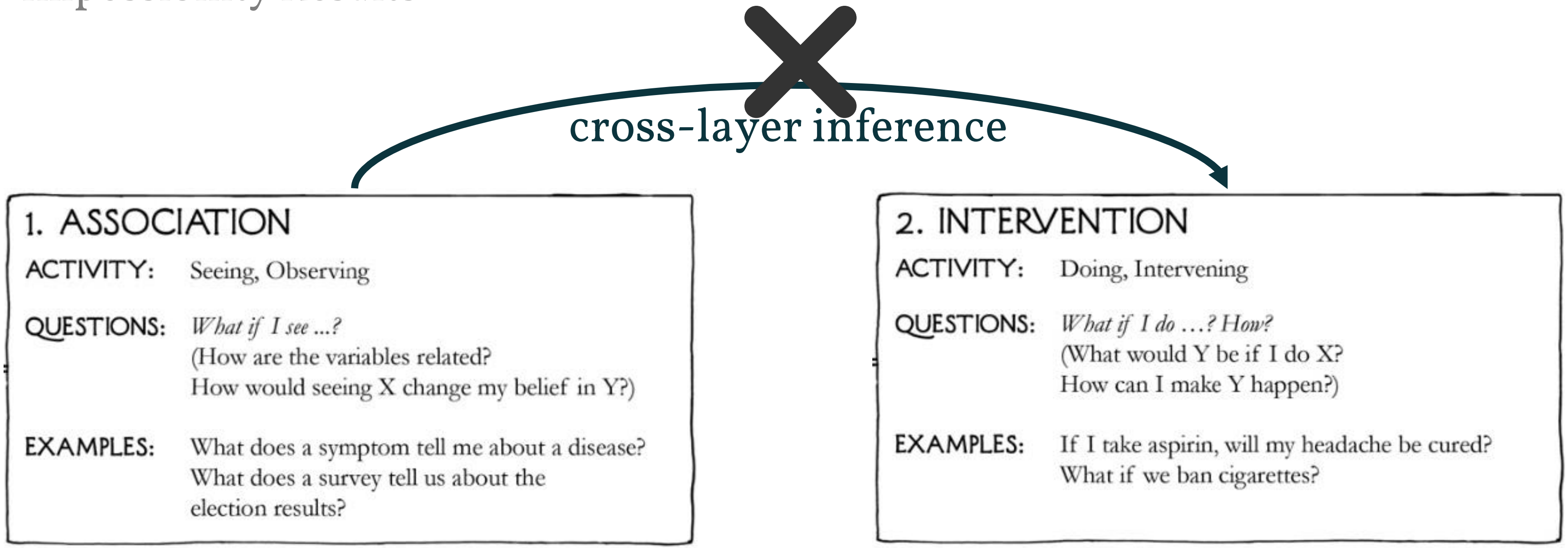
Abduction : Update belief in $P_{\mathbf{U}}$ given evidence $\mathbf{E}$

Action : Change equations accordingly, $do(X = x)$

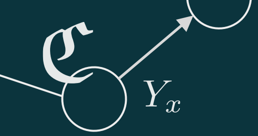
Prediction : Look at variable of interest $P(Y = y)$

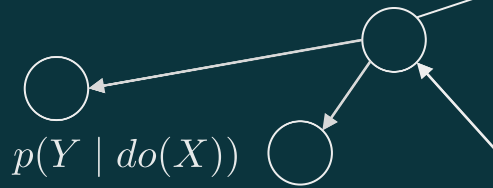
Causal Hierarchy Theorem

Impossibility Results

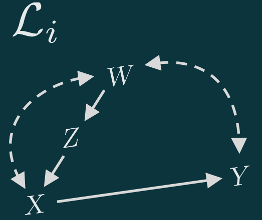


Bareinboim et al., On Pearl's Causal Hierarchy



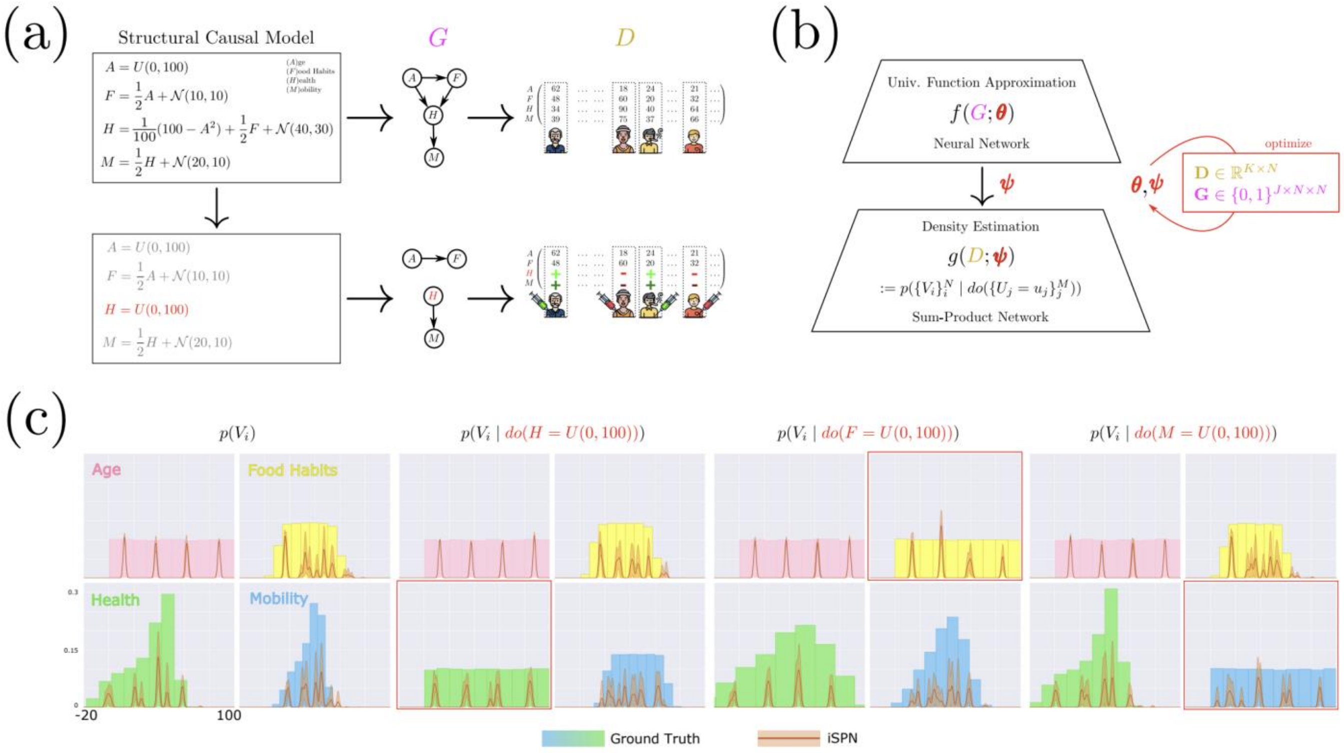


3 | Machine Learning for Causal Effects



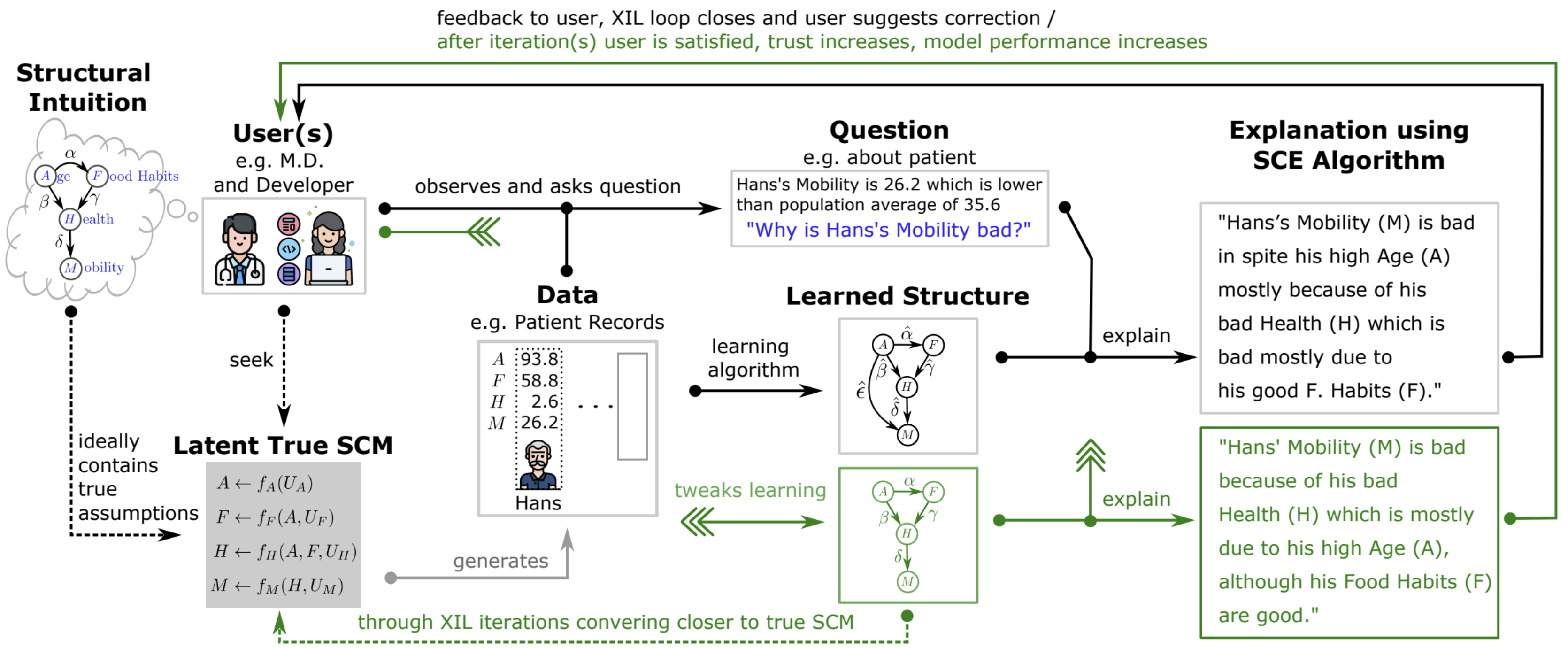
CAUSE
TECHNISCHE UNIVERSITÄT DARMSTADT

Probabilistic Circuits + Causality



Zečević et al. 2021, Interventional SPN

Explanations + Causality



Zečević et al. 2022, Explanations of SCMs

TECHNISCHE UNIVERSITÄT DARMSTADT

Free Code Libraries

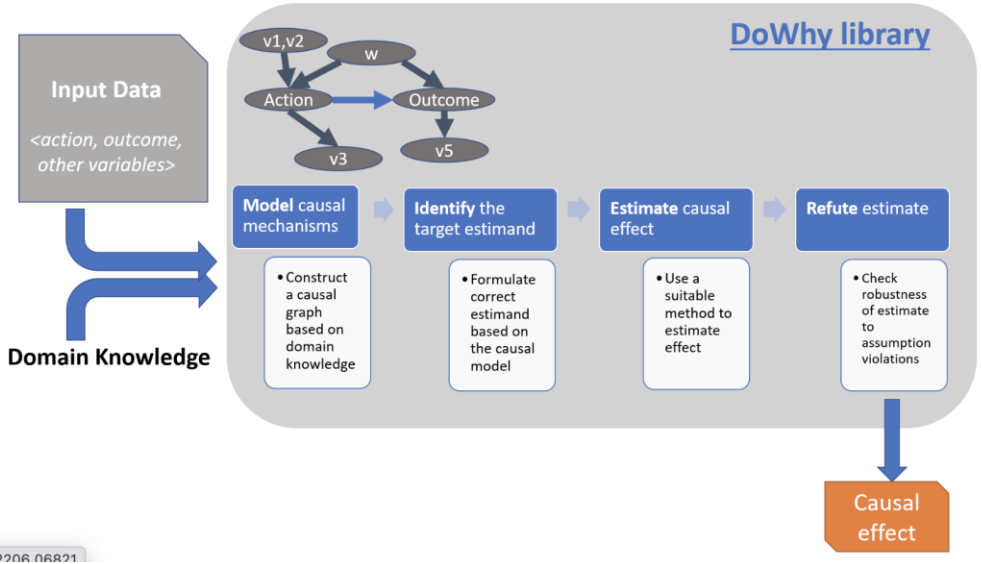
Do it for you

DoWhy | An end-to-end library for causal inference

Introducing DoWhy and the 4 steps of causal inference | Microsoft Research Blog | Video Tutorial | Arxiv Paper | Arxiv Paper (GCM-extension) | Slides

Read the docs | Try it online! launch binder

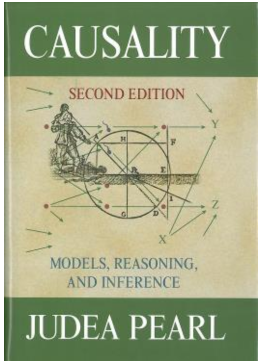
Case Studies using DoWhy: Hotel booking cancellations | Effect of customer loyalty programs | Optimizing article headlines | Effect of home visits on infant health (IHDP) | Causes of customer churn/attrition

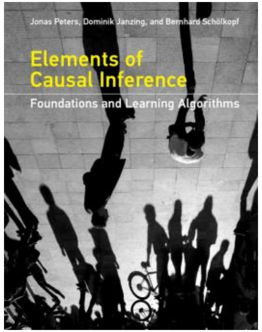


DoWhy, https://github.com/py-why/dowhy

Pointers to Causal Inference References

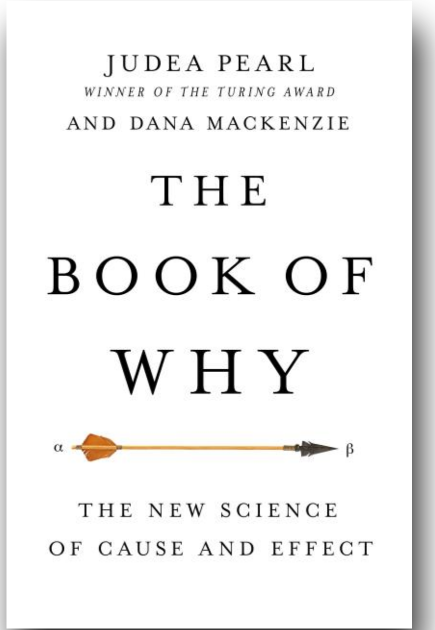
- Judea Pearl, “**Causality**”, Cambridge University Press, 2009.
- Peters et al., “**Elements of Causal Inference**”, MIT Press, 2017.
- Elias Bareinboim Lecture “**Causal Data Science**”, 2019.
  `https://www.youtube.com/watch?v=dUsokjG4DHc`
- Brady Neal’s Free Online Course ”**Introduction to Causal Inference**”, 2020.
  `https://www.bradyneal.com/causal-inference-course`
- Jonas Peters Lecture Series “**Causality**”, 2017.
  `https://www.youtube.com/watch?v=zvrcyqcN9Wo`

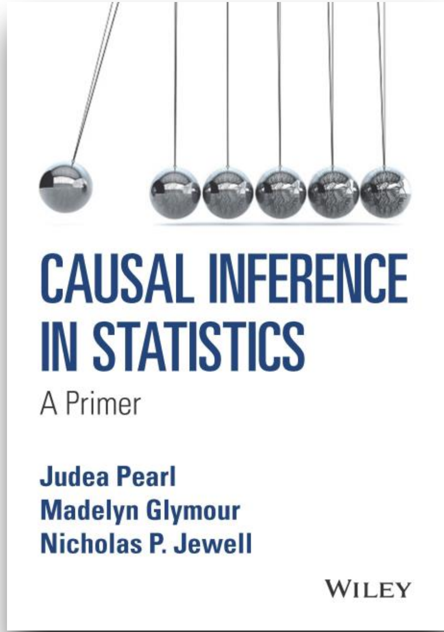


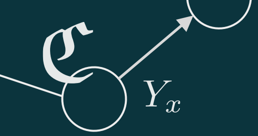


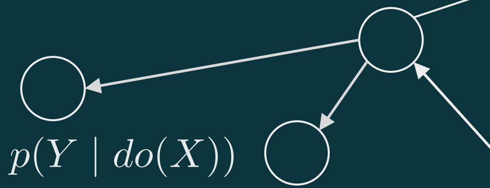
CAUSE
TECHNISCHE UNIVERSITÄT DARMSTADT

Causality Theory by Judea Pearl





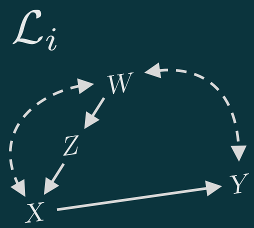




After having seen all this, we realize..

"As X-rays are to the surgeon, <u>graphs</u> are for causation."

-Judea Pearl in Causality (2009)



CAUSE

TECHNISCHE UNIVERSITÄT DARMSTADT