https://www.aboutwayfair.com/tech-innovation/the-importance-of-covariates-in-causal-inference

CAUSE

TU/e
EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

PHILIPPS-UNIVERSITÄT MARBURG 1527

Machines Climbing Pearl’s Ladder of Causation

Devendra Singh Dhami
Adele Ribeiro
Matej Zečević

ESSAI & ACAI 2024
ATHENS - GREECE

- Devendra Singh Dhami
  - Assistant Professor at Eindhoven University of Technology

- Adèle Helena Ribeiro
  - Postdoctoral Researcher at Philipps-Universität Marburg

- Matej Zečević
  - PhD Candidate at Technical University of Darmstadt

5 Thrilling Days

- Day 1: Introduction to Causality (Devendra)
- Day 2: Causal Discovery (Adèle)
- Day 3: Causal Identification and Estimation (Adèle)
- Day 4: Causal Representation Learning (Devendra)
- Day 5: Causality + Large Language Models (Devendra)

Disclaimer

"Don't be fooled. Chaos reigns."

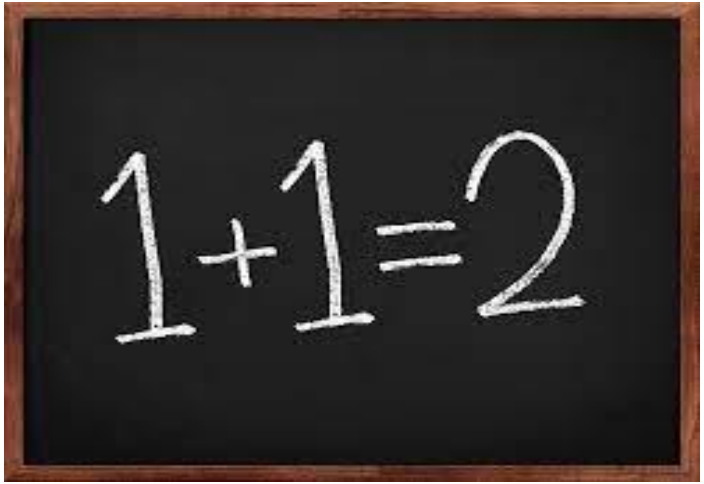
The Ultimate Hitchhiker’s Introduction to Causality for ML

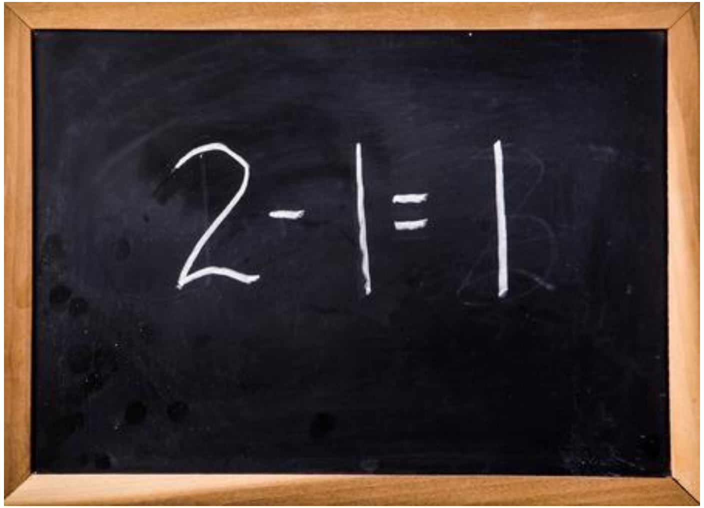
Devendra Singh Dhami


KMBC
NEWS
9


KMBC
NEWS
9

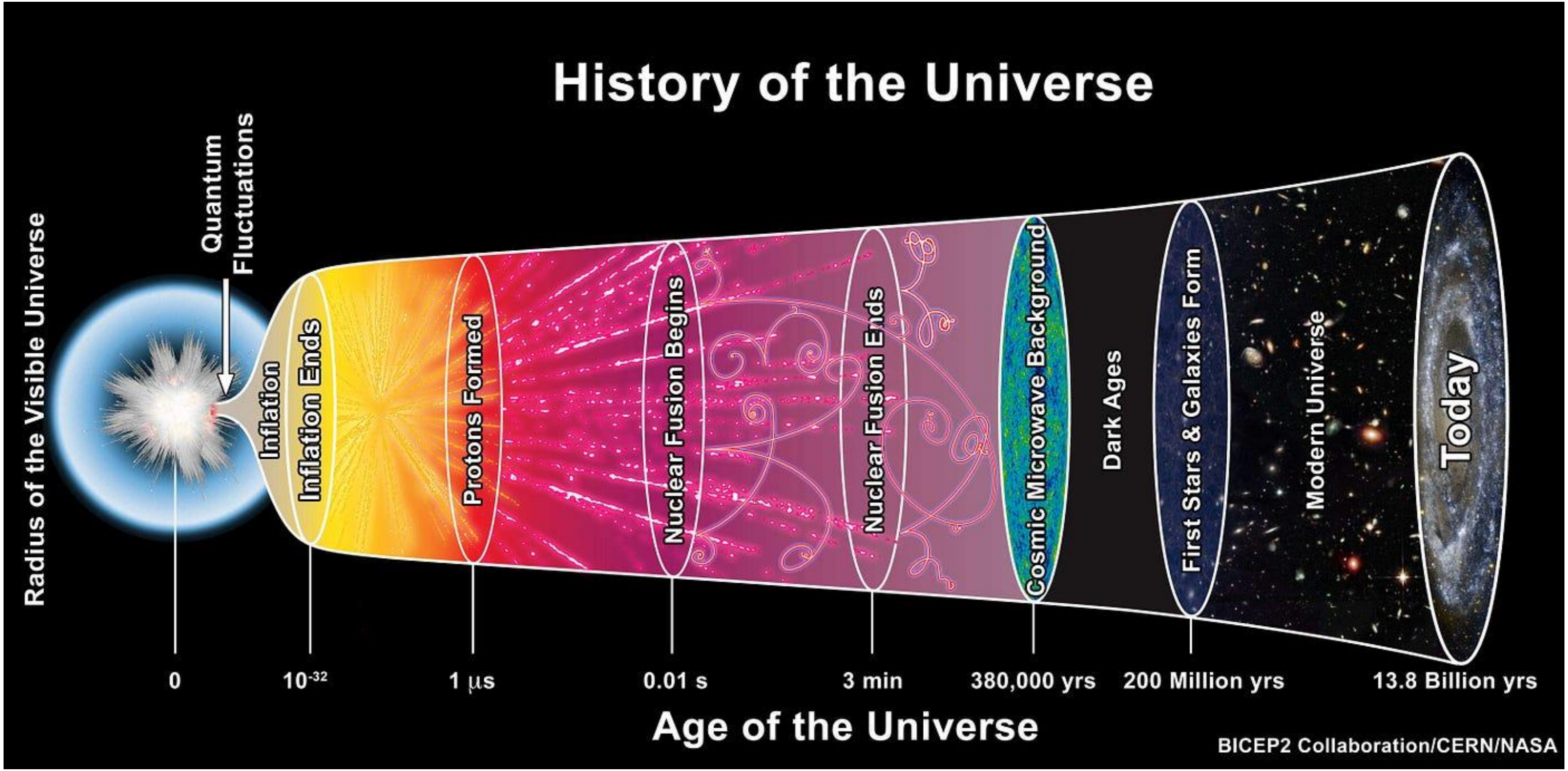
Causality is hard





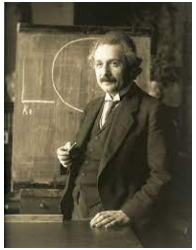
Causality is hard(er)

Causality is hard(er)

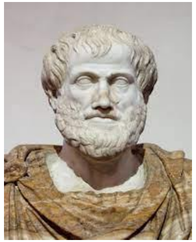


Carl Sagan: If you wish to make an apple pie from scratch, you must first invent the universe

Causality is Omnipresent!



Classical Physics: An effect cannot occur from a cause that is not in the back (past) light cone of that event. Similarly, a cause cannot have an effect outside its front (future) light cone

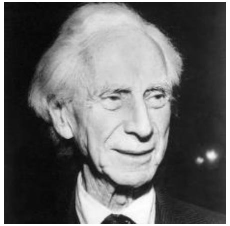


Philosophy: The material cause of a being is its physical properties or makeup. The formal cause is the structure or direction of a being. The efficient cause is the thing or agent, which actually brings it about. And the final cause is the ultimate purpose for its being

Economics: Develops explicit models of outcomes where the causes of effects are investigated and the mechanisms governing the choice of treatment are analyzed

Causality is Omnipresent?

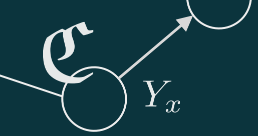
David Hume: Causation is a relationship between two impressions or ideas in the mind. Because causation is defined by experience, any cause-and-effect relationship could be incorrect because thoughts are subjective and therefore causality cannot be proven

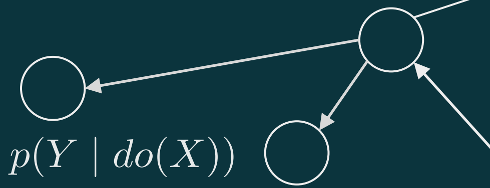


Barber's Paradox: There is a village where the barber shaves all those and only those who do not shave themselves. Who shaves the barber?

A causal problem → Are the non-shavers actually non-shavers because they're shaved by the barber or whether they're shaved by the barber because they're non-shavers. If the first is true, the barber is a member of both groups: he shaves himself and is shaved by the barber

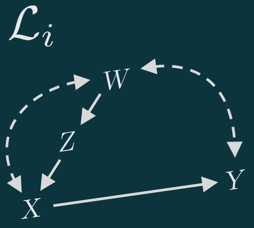
Can you define
“What is Causality?”

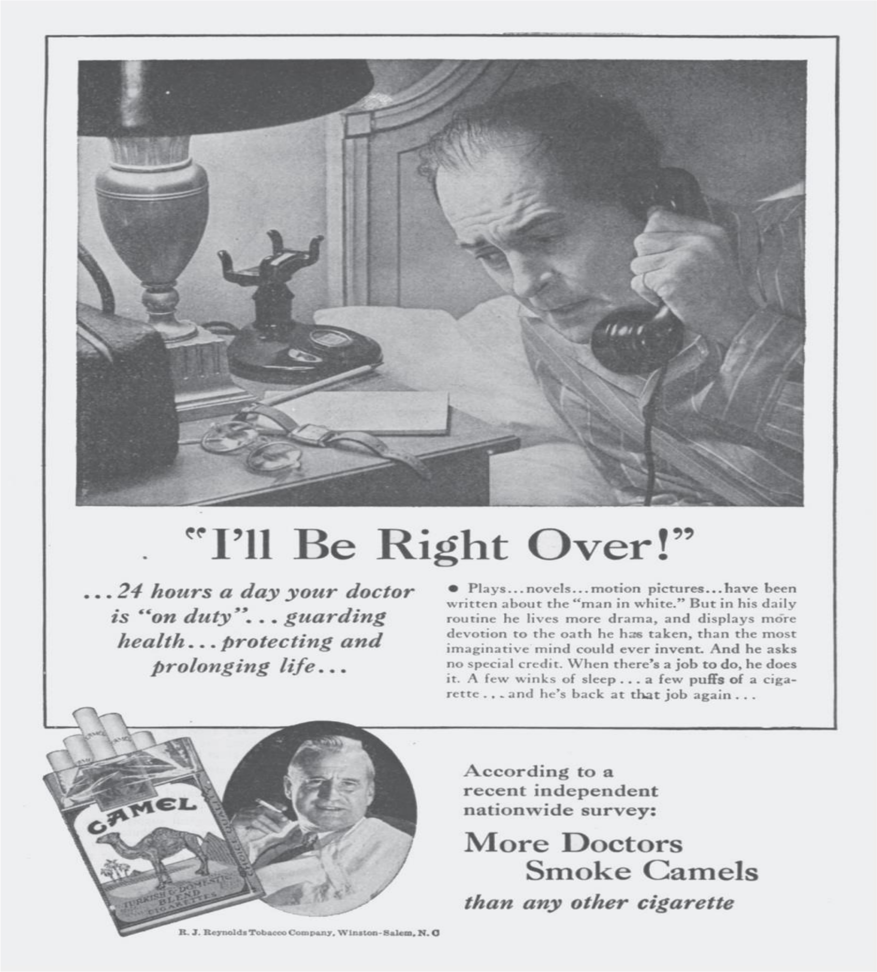




I | Why?

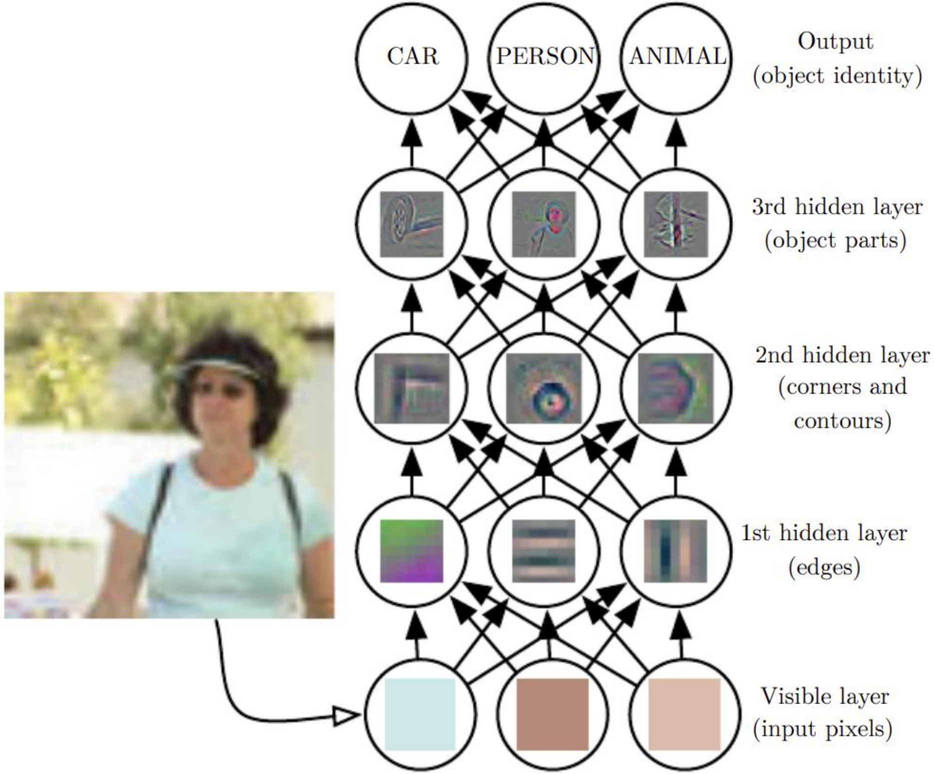
Machine Learning needs Causality?





Pearl and Mackenzie, The Book of Why, 2017

Current Trend: Deep Learning



Goodfellow et al., Deep Learning, MIT Press, 2017

Current Trend: Deep Learning

DE

an example causal graph for unexperienced audience

Sure! Let's create a simple causal graph with an example that is easy to understand for an inexperienced audience. Suppose we want to demonstrate the relationship between sleep and academic performance in students.

Example Causal Graph: Sleep and Academic Performance

In this example, we have two variables: "Sleep" and "Academic Performance." We will use arrows to represent causal relationships between these variables.

scss   Copy code

```
     +-----------------+
     |                 |
     |    Sleep        |
     |                 |
     +---(causes)-----> +---(causes)--->
     |                 |                      Academic
     |                 |                      Performance
     +-----------------+
```

Current Trend: Deep Learning

Explanation:

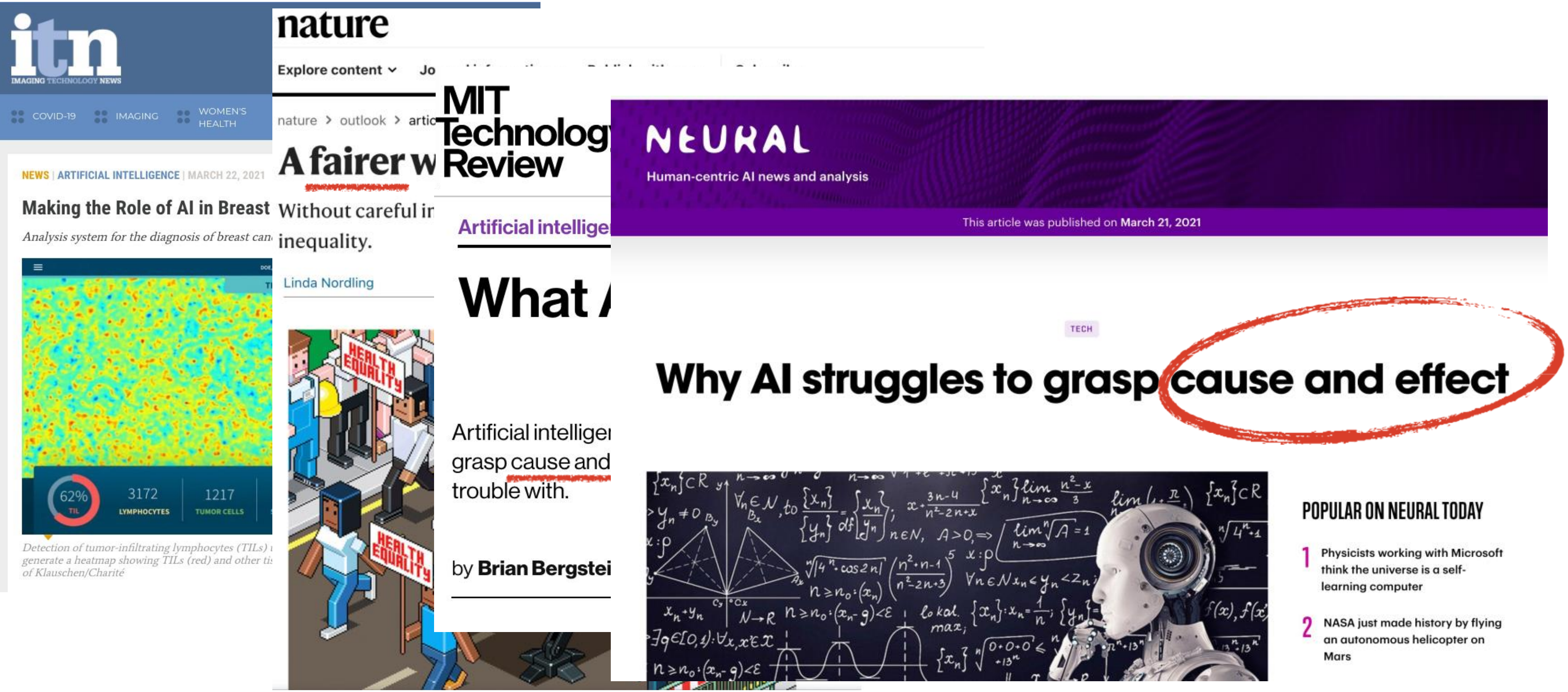
1. Sleep: This variable represents the number of hours a student sleeps each night.
2. Academic Performance: This variable represents a student's academic performance, which can be measured by their grades or exam scores.

Causal Relationship:

The arrow pointing from "Sleep" to "Academic Performance" indicates a causal relationship. In this example, we are hypothesizing that sleep directly affects academic performance. More sleep may lead to better academic performance, while less sleep may result in poorer academic performance.

Current Trend: Deep Learning

Current Challenges in DS, AI, ML



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.

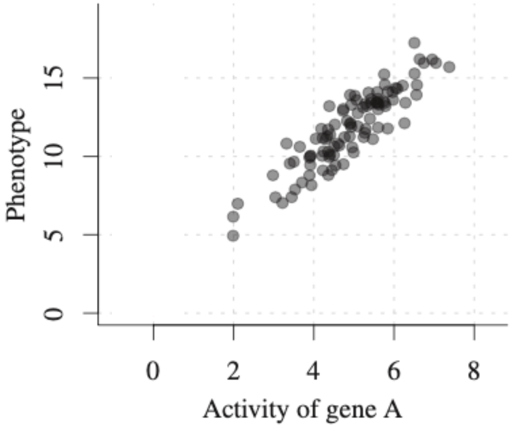
Fairness: captures and disentangles any mechanisms of discrimination that may be present, including direct, indirect-mediated, and indirect-confounded.

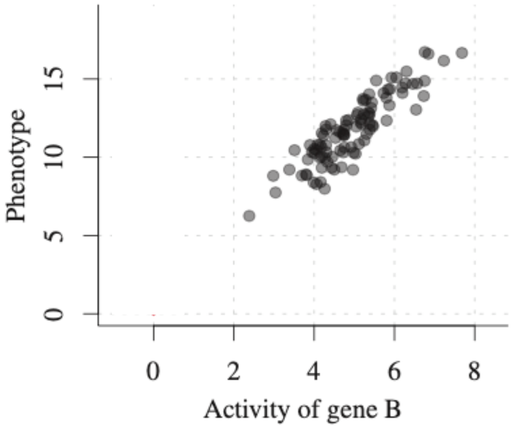
Prompt: Make it a dark **start**-lit night.

Prompt: Make it a dark **star-lit** night.

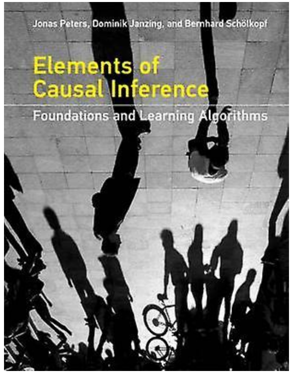
Prompt: Make it a dark **star-lit** night.

Why do we care about causality in AI & ML?



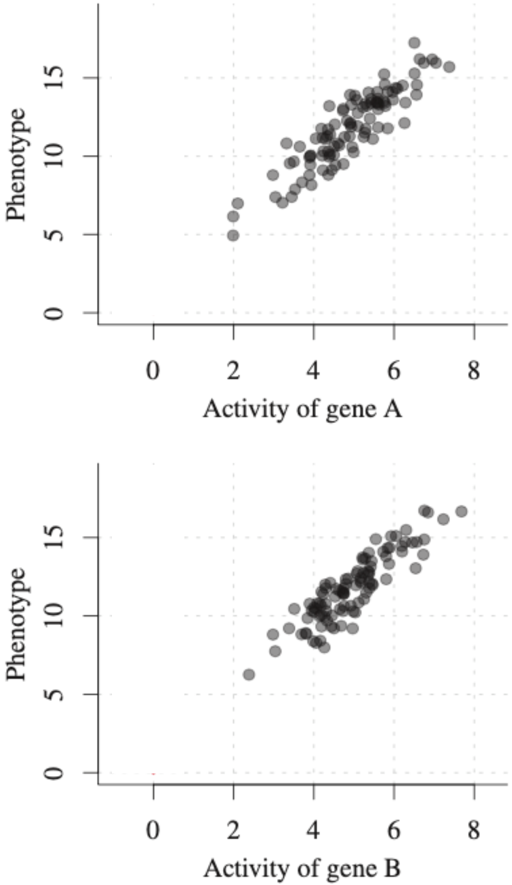


In ML for instance, we'll encounter data sets like on the left…



Example adapted from "Elements of Causal Inference" (2017) by Jonas Peters, Dominik Janzing and Bernhard Schölkopf

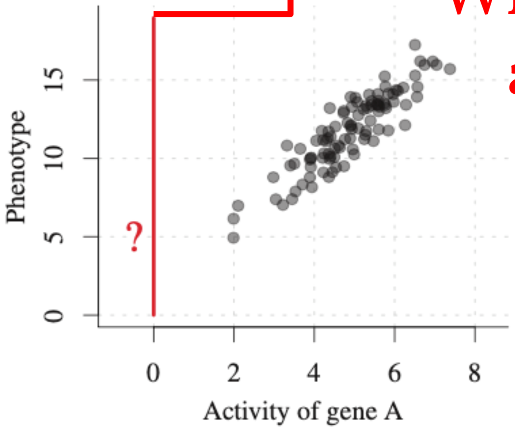
Dataset I



Dataset II

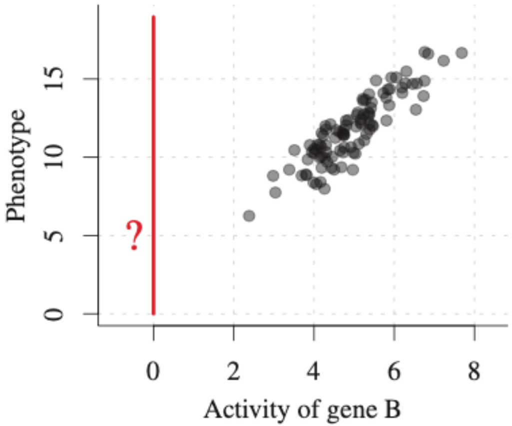
But, what is the difference between datasets I & II w.r.t. learning?

Dataset I



What if we kill the activity of the gene?

Dataset II

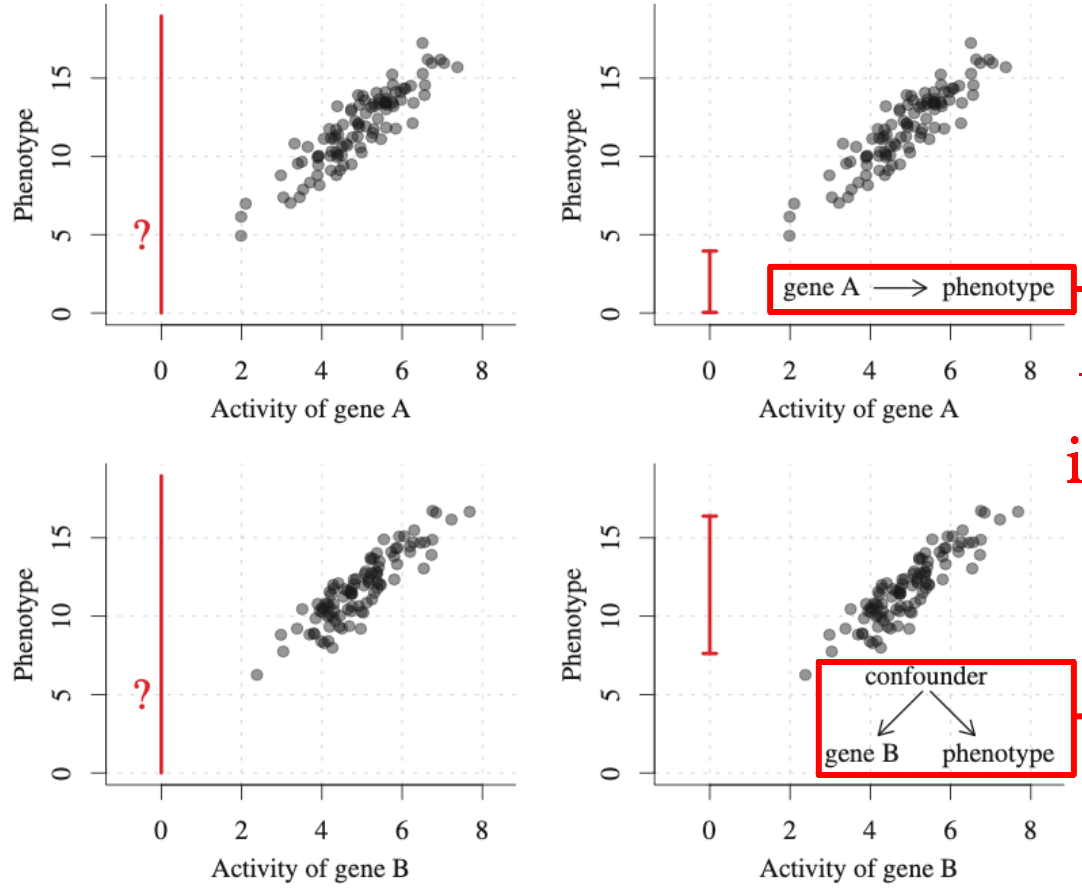


We might ask about the difference w.r.t. generalization!

What if we kill the activity of the gene? Then we expect the following range

Dataset I

Dataset II



the difference is in the underlying **Causality!**

In Conclusion:

Without considering causality, the best that our algorithms could do is to say "I don't know"*

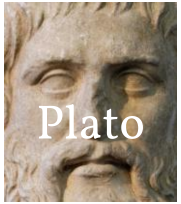
* and we really really really do want to know

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

What is Causality?

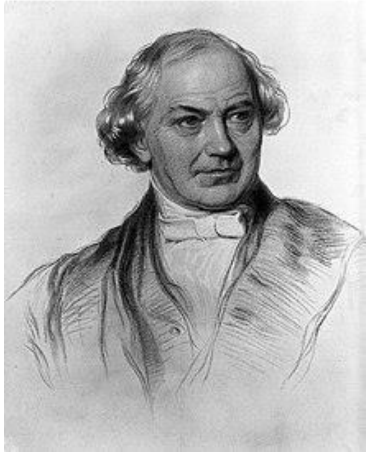
We might want to start here first..

Probably, 

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

Axioms of Causality



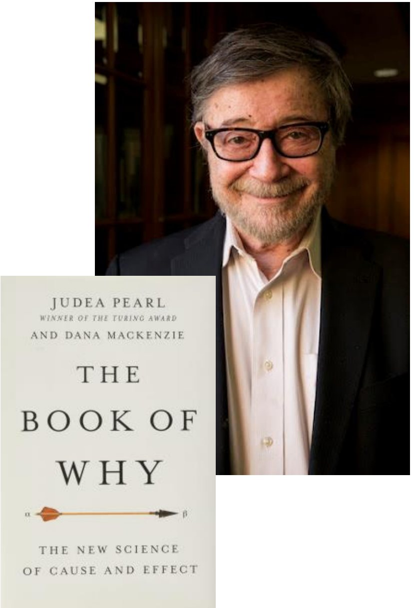
- Nothing takes place without a cause.
- The magnitude of an effect is proportional to the magnitude of its cause.
- To every action there is an equal and opposed reaction.

Example: if a baseball is moving through the air, it must be moving this way because of a previous interaction with another object, such as being hit by a baseball bat

Principle of Universal Causation: Every phenomenon has a cause, which it invariably follows; and from this are derived other invariable sequences among the successive stages of the same effect, as well as between the effects resulting from causes which invariably succeed one another

Judea Pearl’s opinion

Pioneer of Causality for AI, Turing awardee



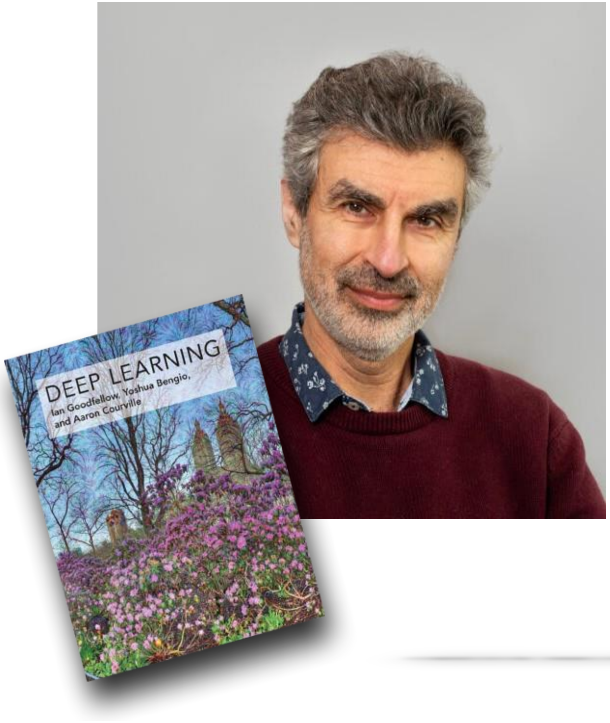
“To Build Truly Intelligent Machines,
Teach Them Cause and Effect”

“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

Yoshua Bengio's opinion

Pioneer of Deep Learning, Turing awardee



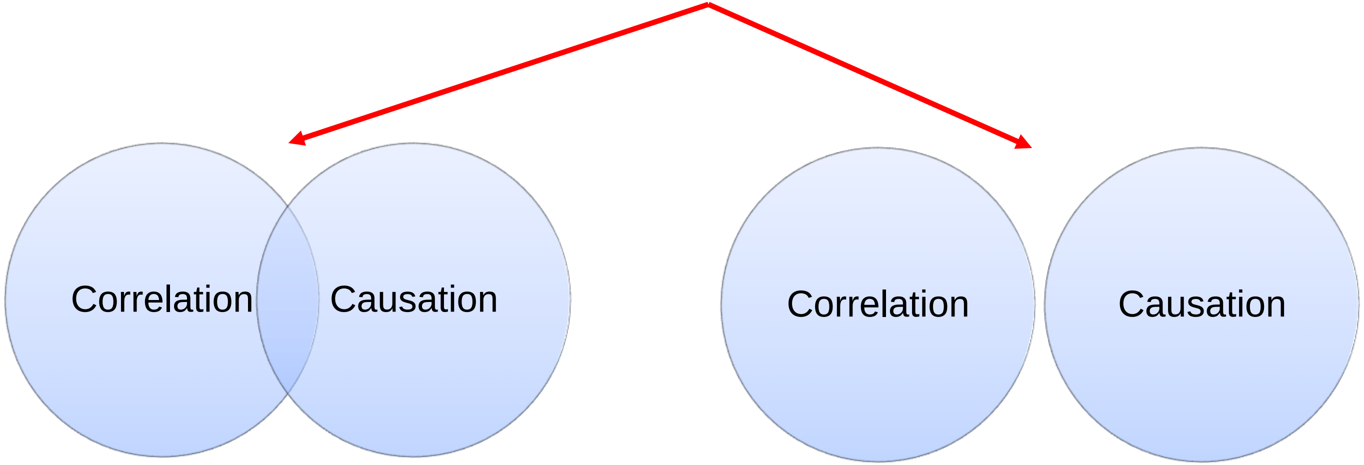
"..for conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing."

"Causality is very important for the next steps of progress of machine learning,"

Yoshua Bengio in an interview with IEEE Spectrum, 2020

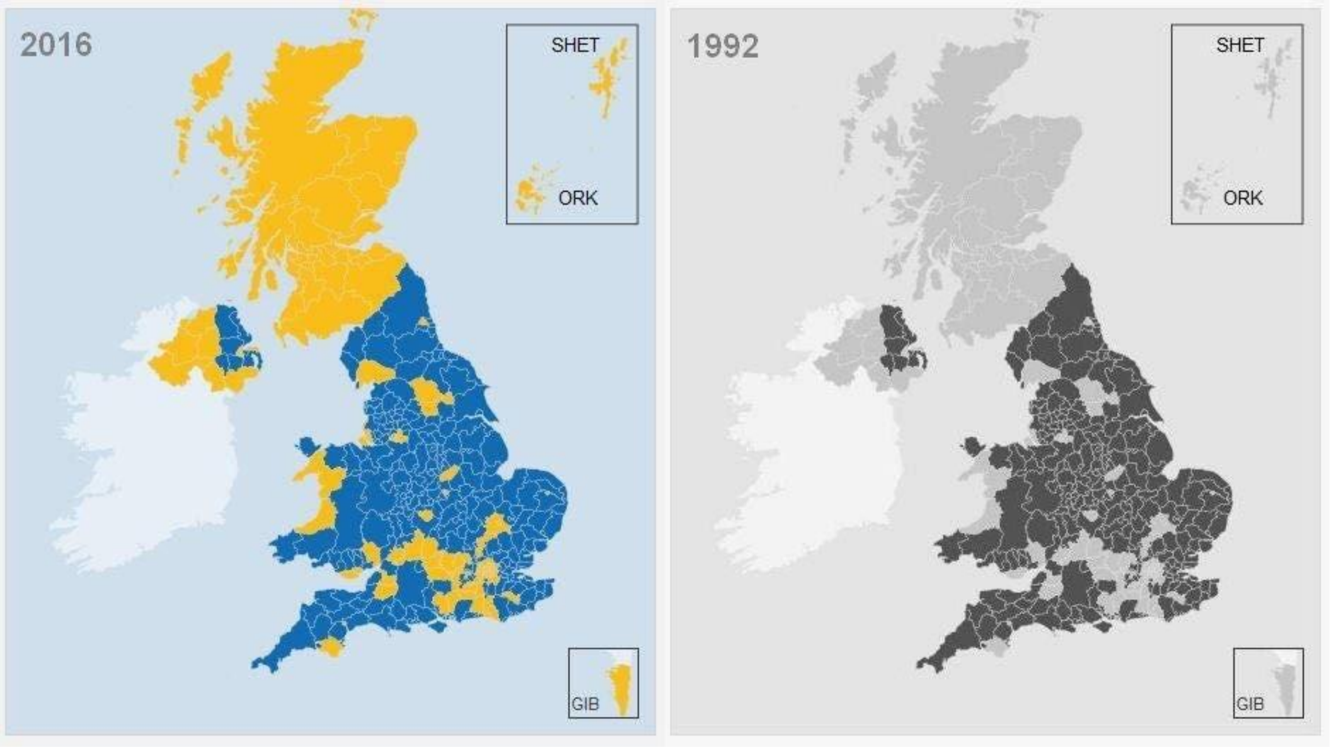
We have All Heard the Phrase

“*Correlation does not imply causation*”



Pic Credit: Wikipedia

Correlation ∩ Causality = Ø



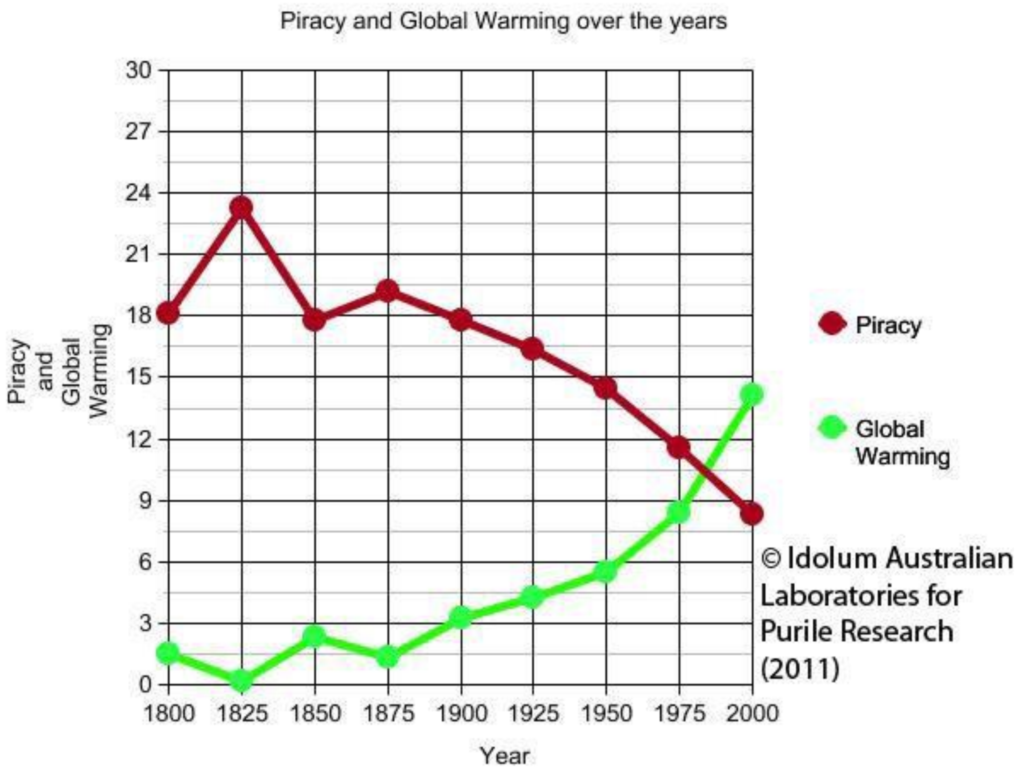
Key:

Majority leave Majority remain

Key:

BSE-Areas BSE-Free-Areas

Correlation => Causality via a third factor



DE

what is the effect of global warming on pirate ships

As of my last update in September 2021, there is no direct scientific evidence or research specifically focused on the effect of global warming on pirate ships. Global warming primarily refers to the long-term increase in Earth's average surface temperature due to human activities, such as the emission of greenhouse gases like carbon dioxide.

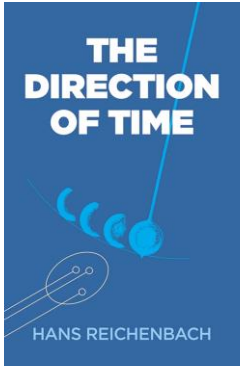
However, we can speculate on some potential indirect effects that global warming might have on pirate ships:

1. Rising sea levels: Global warming contributes to the melting of polar ice caps and glaciers, leading to rising sea levels. Higher sea levels can alter coastlines and navigational routes, making it challenging for pirate ships to access certain areas or ports.
2. Changes in weather patterns: Global warming can lead to shifts in weather patterns, including more intense storms, hurricanes, and cyclones. These extreme weather events can be dangerous for pirate ships at sea and may increase the risk of accidents or shipwrecks.
3. Ocean acidification: The increased concentration of carbon dioxide in the atmosphere not only causes global warming but also results in more CO2 being absorbed by the oceans. This process leads to ocean acidification, which can have implications for marine life, including the plankton and other organisms that form the base of the marine food chain. Changes in marine ecosystems could indirectly impact pirate activity if their traditional hunting grounds are affected.
4. Temperature changes: While global warming generally leads to higher temperatures globally, localized effects can vary. Some regions may experience altered wind patterns and temperature shifts, which can influence the timing and strength of monsoons, trade winds, and other maritime conditions that pirates historically relied on for navigation.
5. Economic and geopolitical impacts: Global warming can disrupt economies and lead to resource scarcity in certain regions. These changes may influence trade routes and shift the focus of piracy to areas that were previously less affected.

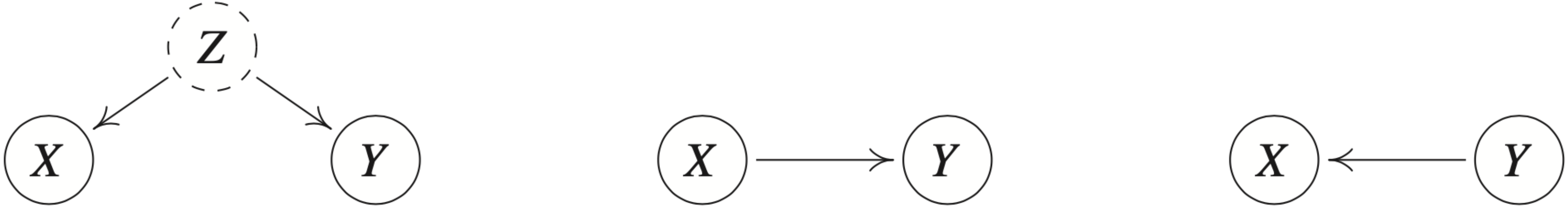
It's essential to note that these are speculative effects, and any actual impact would likely depend on various factors such as the severity and pace of global warming, regional climate changes, and human responses to these challenges.

Reichenbach's Common Cause Principle

Defining Confounders

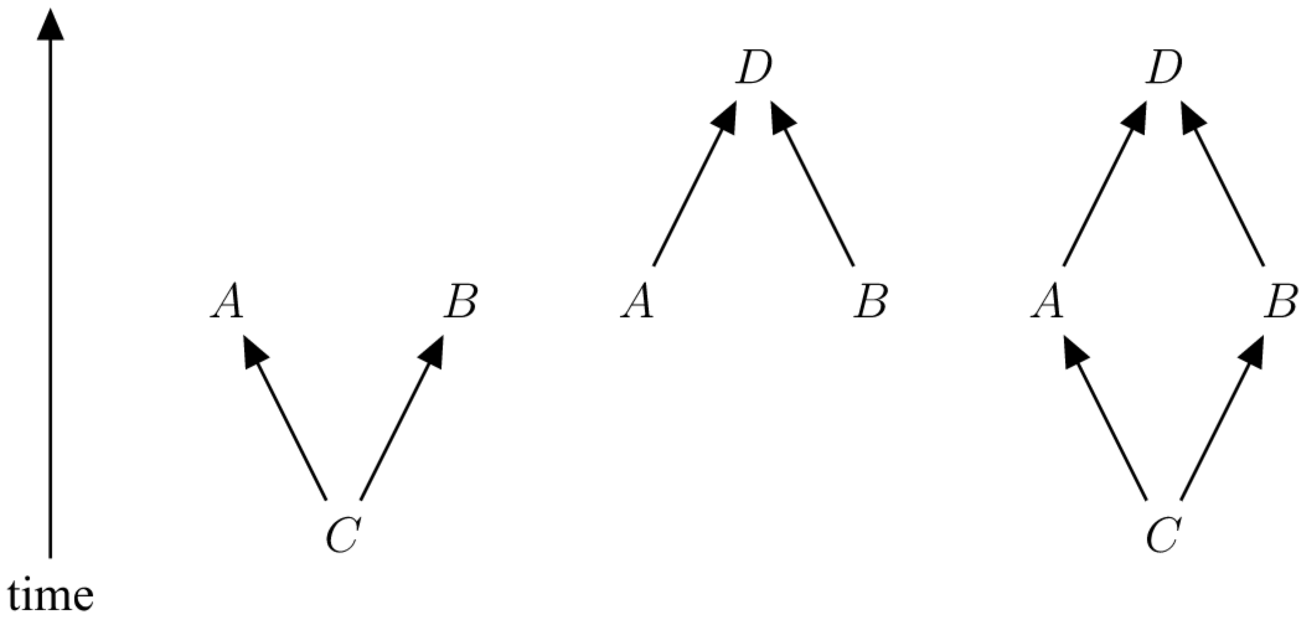


Principle 1. *If two random variables X and Y are statistically dependent ($X \not\perp Y$), then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given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



Reichenbach's *Direction of Time* (1956)

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/

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

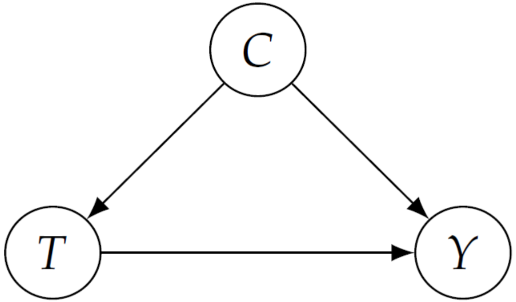
Condition

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

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

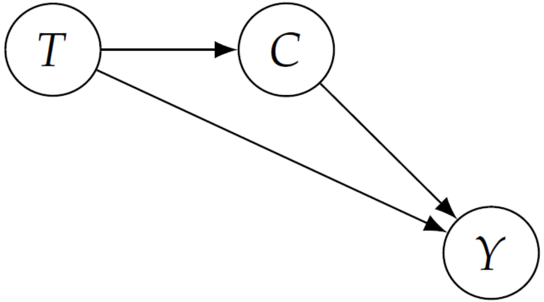
Let's Illustrate Correlation does not imply causation : **Simpson's Paradox**

Scenario 1: Confounders



Treatment A preferable

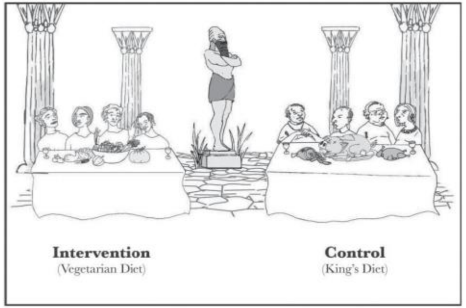
Scenario 2: Treatment causes condition



Treatment B preferable

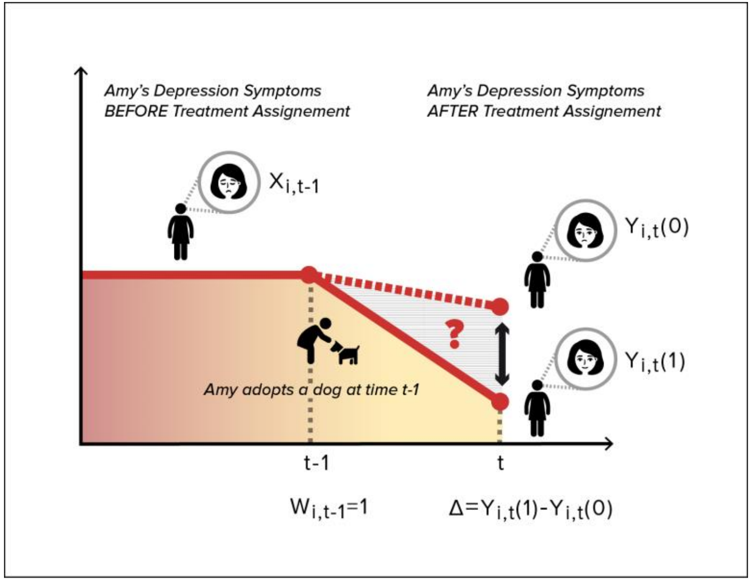
From Neyman-Rubin…..

- Potential Outcome theory



The biblical story of Daniel, often cited as the first controlled experiment. Daniel (third from left?) realized that a proper comparison of two diets could only be made when they were given to two groups of similar individuals, chosen in advance. King Nebuchadnezzar (rear) was impressed with the results. (*Source:* Drawing by Dakota Harr.)

Fundamental Problem 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 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 2: Average treatment effect (ATE)

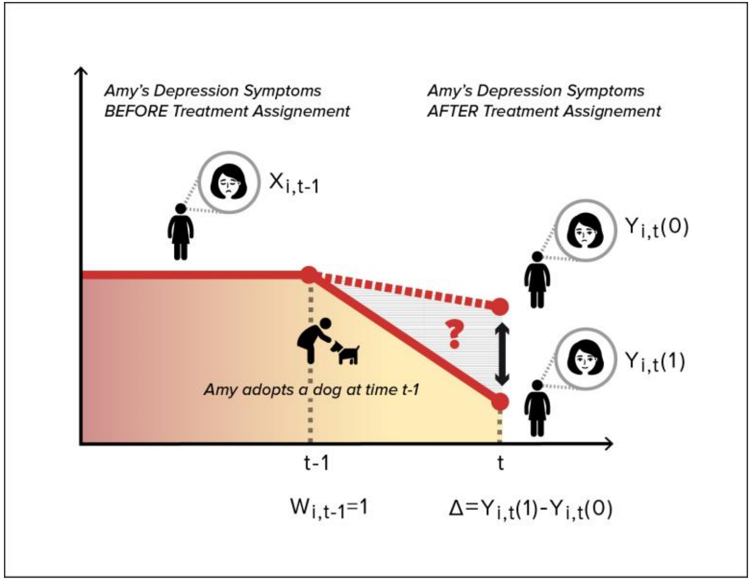
The average treatment effect is the population average of all i individual treatment effects

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

What is the Apparent Problem?

Credit: Scott Cunningham

Fundamental Problem 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 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 2: Average treatment effect (ATE)

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

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

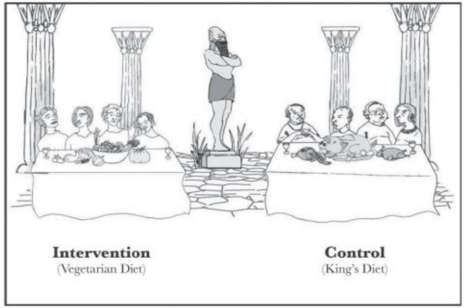
Definition 4: Fundamental problem of causal inference

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.

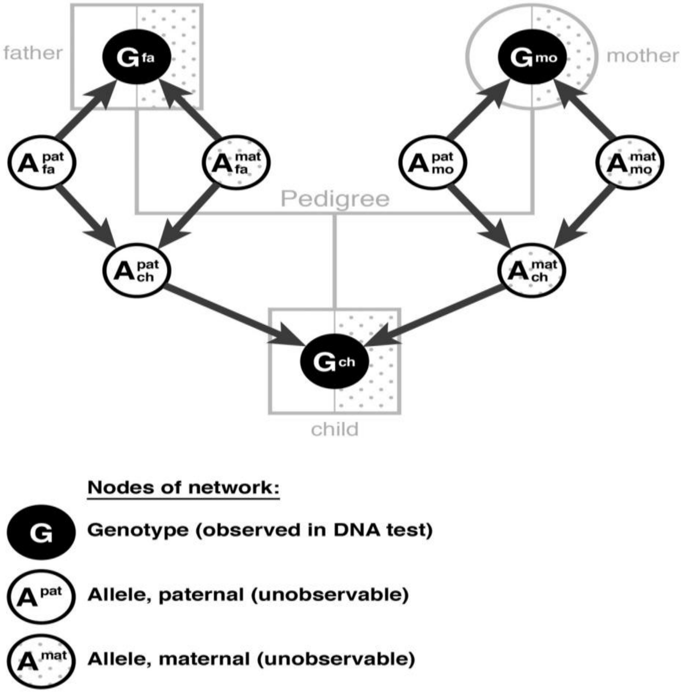
Credit: Scott Cunningham

From Neyman-Rubin to Pearl

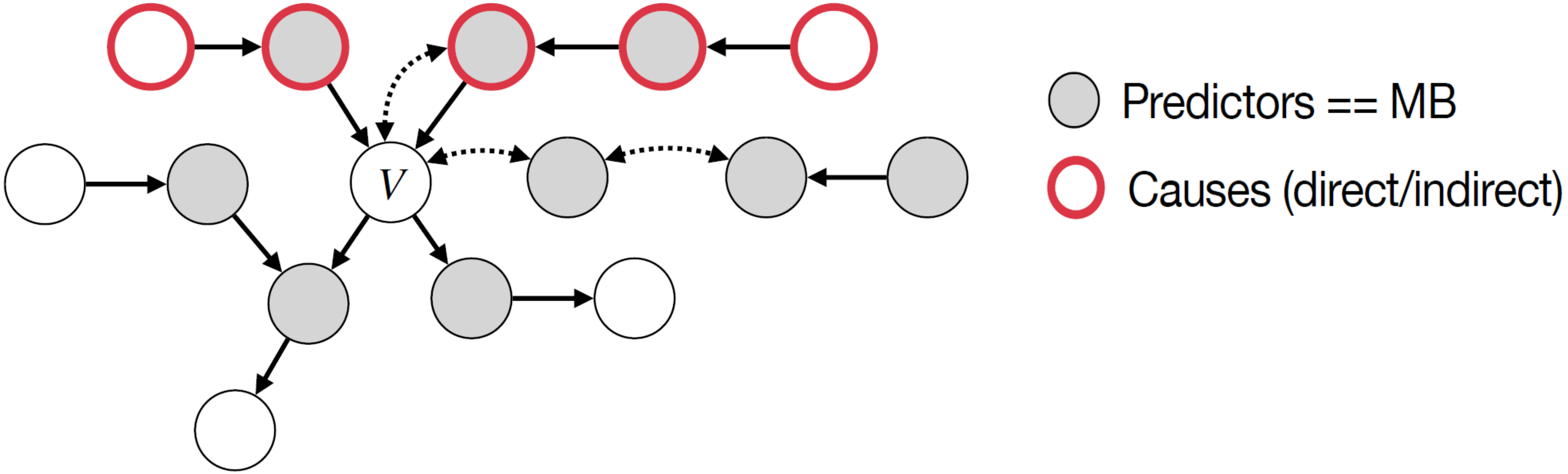
- Potential Outcome theory to Bayesian networks and DAG's



The biblical story of Daniel, often cited as the first controlled experiment. Daniel (third from left?) realized that a proper comparison of two diets could only be made when they were given to two groups of similar individuals, chosen in advance. King Nebuchadnezzar (rear) was impressed with the results. (*Source:* Drawing by Dakota Harr.)



Graphically Explaining Causes and Predictors

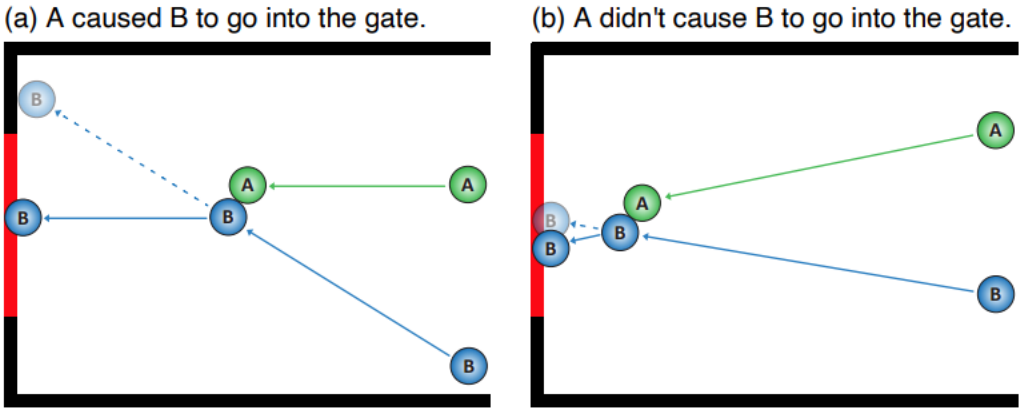


Pearlian Causality

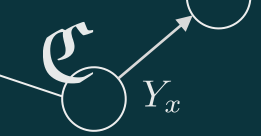
A success story

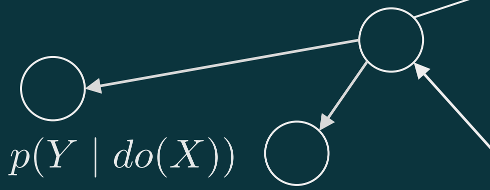
The formalization with most success in AI/ML so far.

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



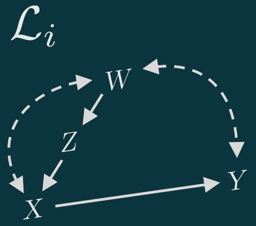
Gerstenberg. What would have happened?[...] PTRBAE 2022.





2 What?

Does Pearlian Causality look like?



“Some tens of thousands of years ago, humans began to realize that certain things cause other things and that tinkering with the former can change the latter... From this discovery came organized societies, then towns and cities, and eventually the science and technology-based civilization we enjoy today. All because we asked a simple question: Why?” – in The Book of Why

Philosophical Insight

1. Dichotomy between “reality” & what we observe (data)

2. We only have access to data

   e.g. J. Locke:

   “[when we observe data, we cannot] so much as guess, much less know, their manner of production”

Pearl's Solution

- Representation of Reality:

  Structural Causal Model (SCM)

- Data of Reality (implied by SCM):

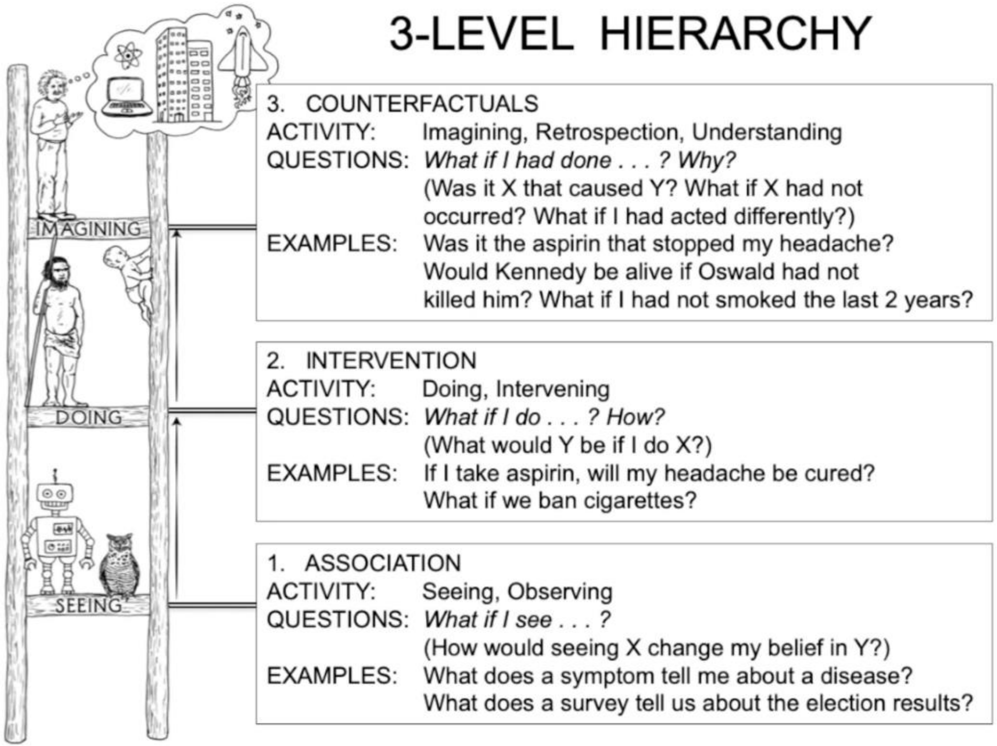
  Factual information

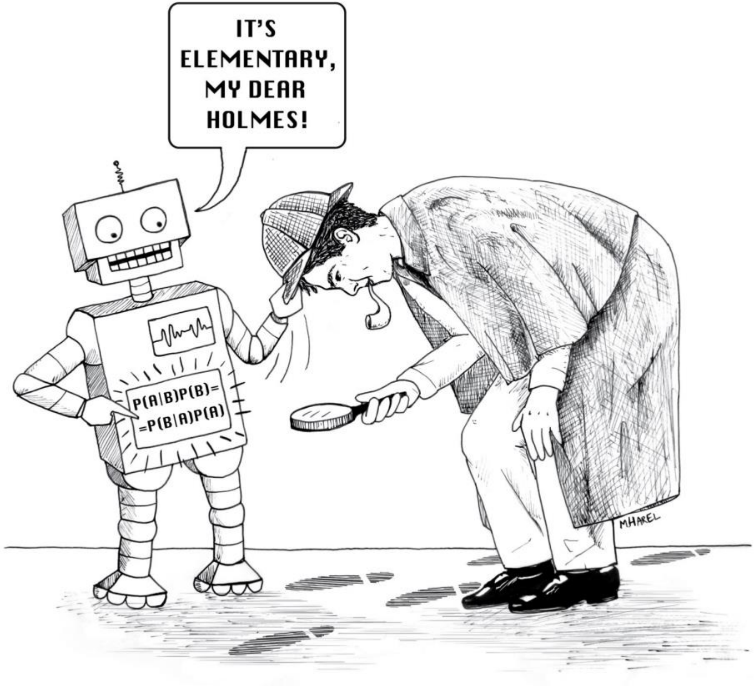
  Hypothetical information

  Retrospective information

Pearl's Causal Hierarchy (PCH)
("Ladder of Causation")

The Causal Hierarchy





Sherlock Holmes meets his modern counterpart, a robot equipped with a Bayesian network. In different ways both are tackling the question of how to infer causes from observations. The formula on the computer screen is Bayes's rule. (*Source:* Drawing by Maayan Harel.)

Pearl and Mackenzie, The Book of Why, 2017

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

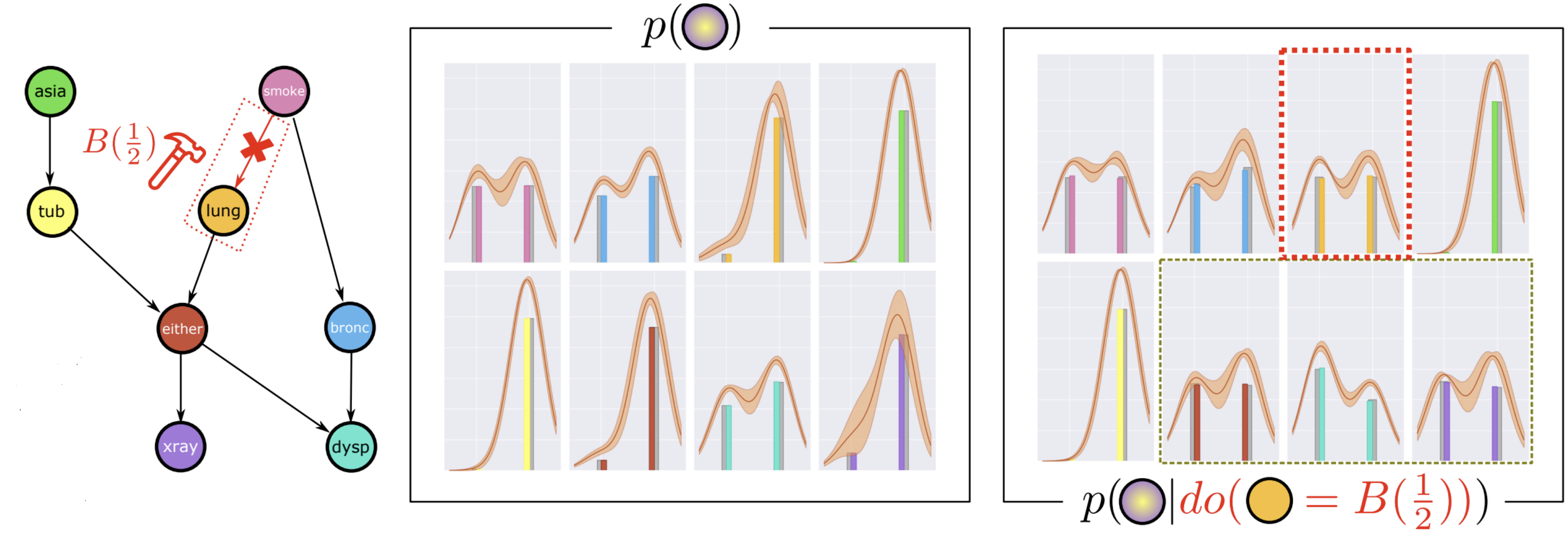
ACTIVITY: Doing, Intervening

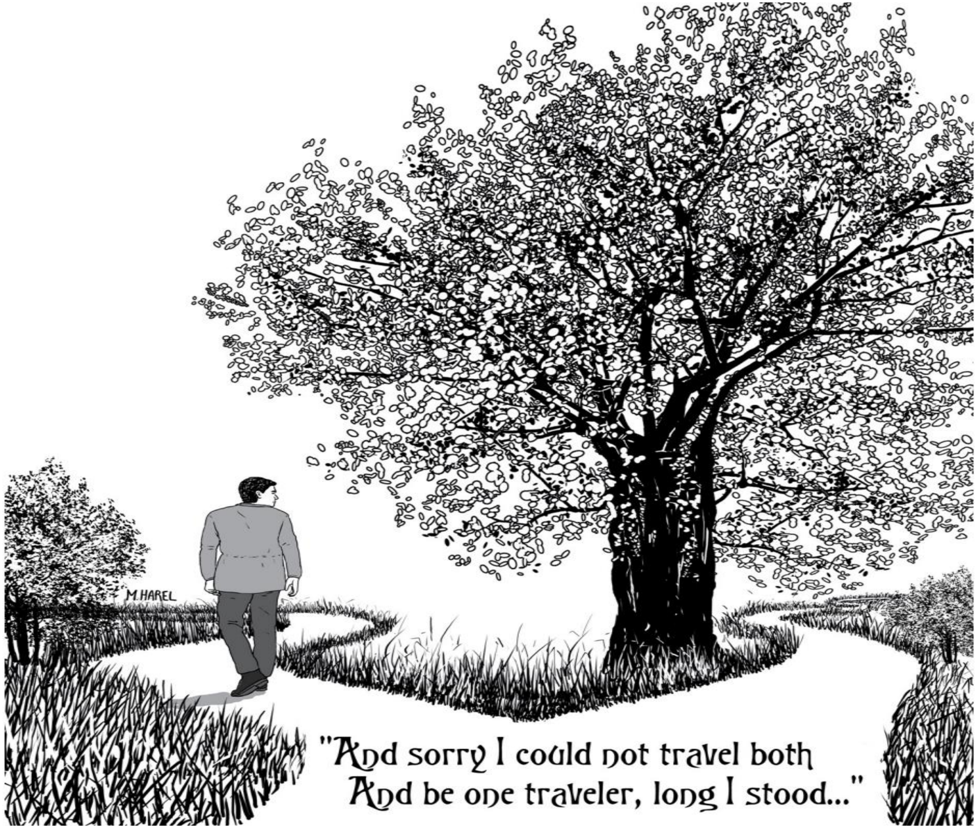
QUESTIONS: *What if I do …? How?*
(What would Y be if I do X?
How can I make Y happen?)

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

Pearl Causal Ladder

Level II





Robert Frost's famous lines show a poet's acute insight into counterfactuals. We cannot travel both roads, and yet our brains are equipped to judge what would have happened if we had taken the other path. Armed with this judgment, Frost ends the poem pleased with his choice, realizing that it "made all the difference." (*Source:* Drawing by Maayan Harel.)

Pearl and Mackenzie, The Book of Why, 2017

Pearl Causal Ladder

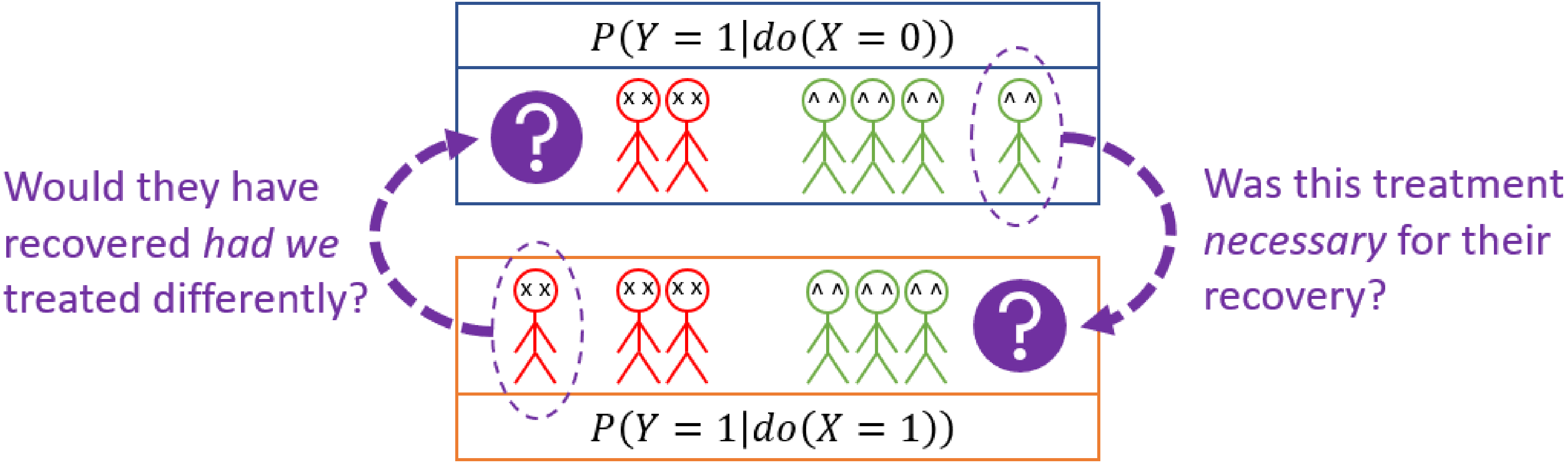
Level III

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 Causal Ladder

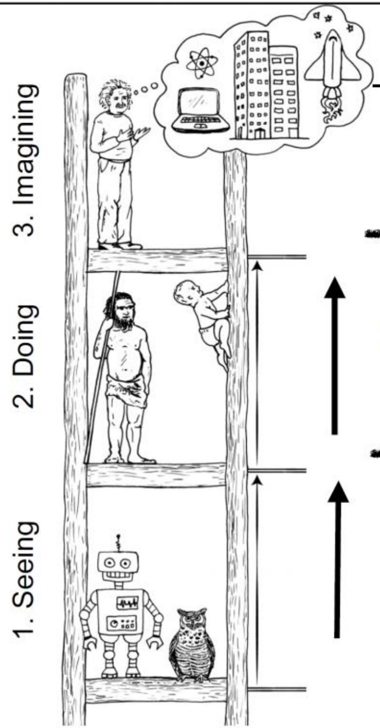
Level III



Credit: https://forns.lmu.build/classes/spring-2019/cmsi-498/lecture-5T.html

Pearl's Hierarchy

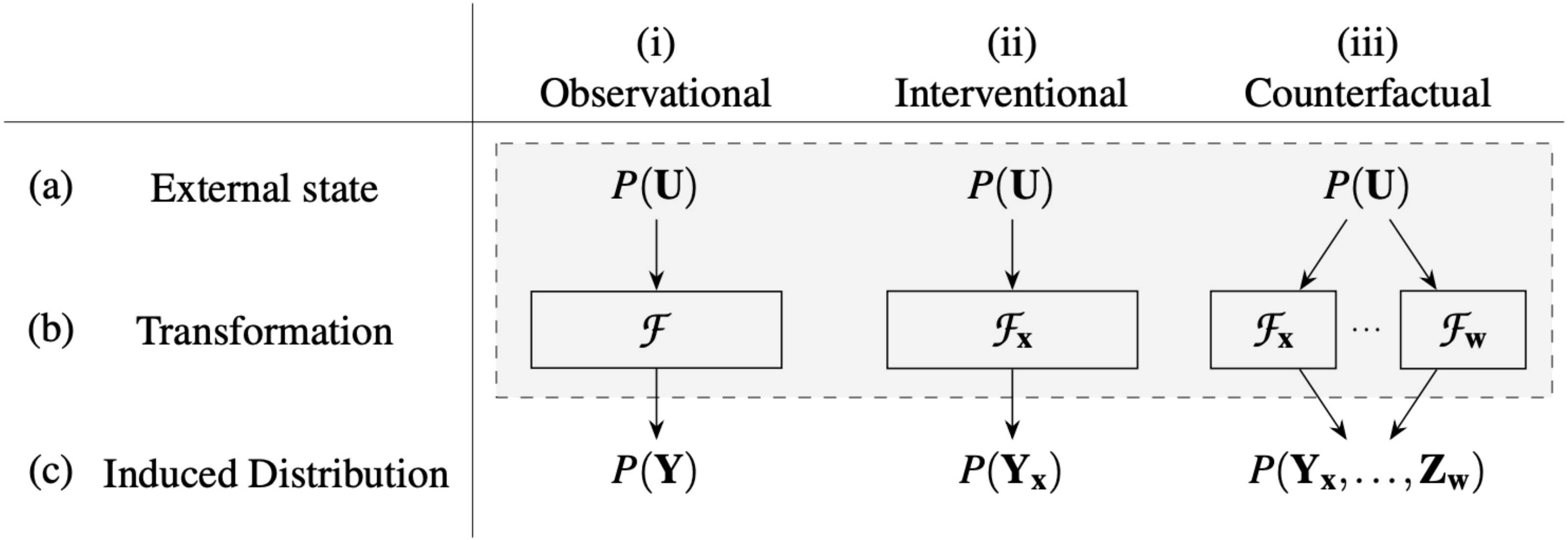
Ladder of Causation



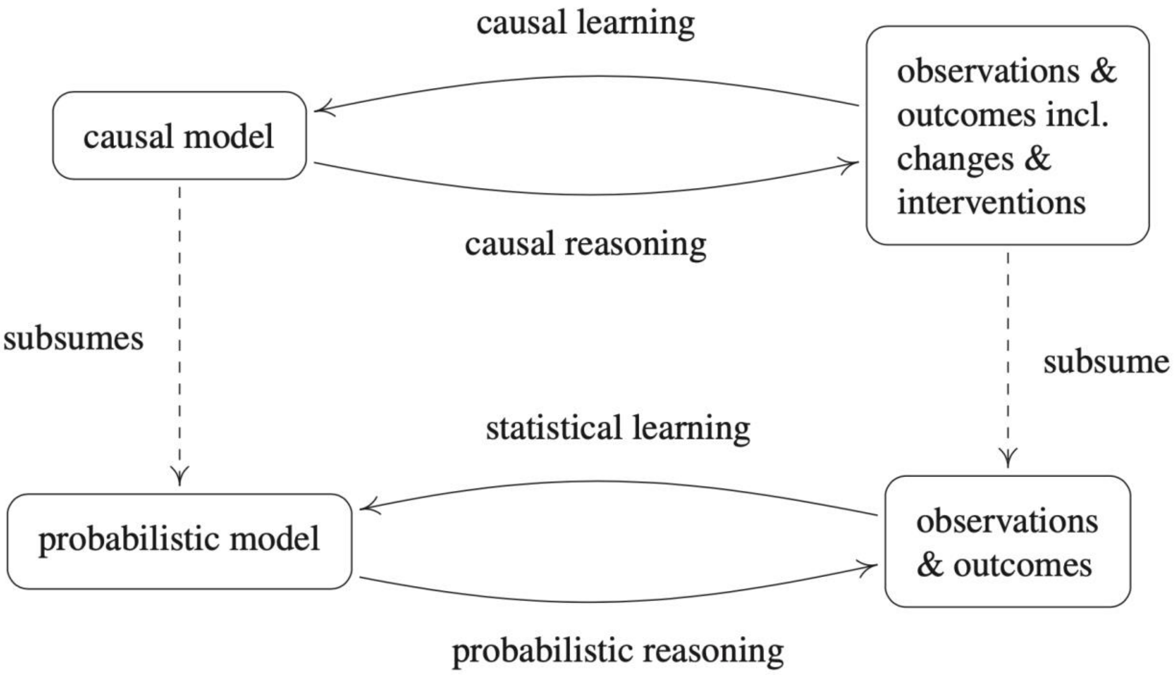
Layer	Task / Language	Typical Question	Examples
Counterfactual $P(y_x \mid x', y')$	Structural Causal Model	What if I had acted differently?	Was it the aspirin that stopped my headache?
Interventional $P(y \mid do(x), c)$	ML- Reinforcement (Causal Bayes Net)	What if I do X? What would Y be if I intervene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

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

Distinction Between the Rungs

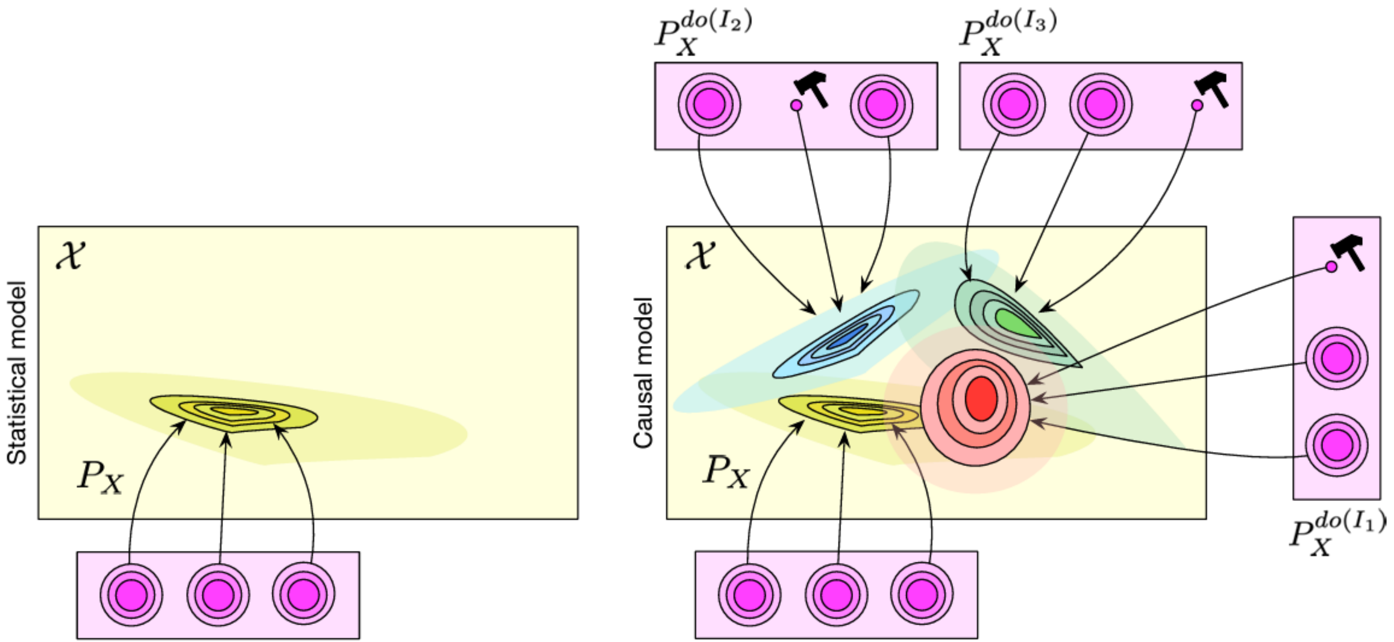


Causal versus Probabilistic Inference



Peters et al. 2017, Fig.1.1

Causal versus Statistical Models



Schölkopf et al., Towards Causal Representation Learning, Proceedings of the IEEE 2021

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.

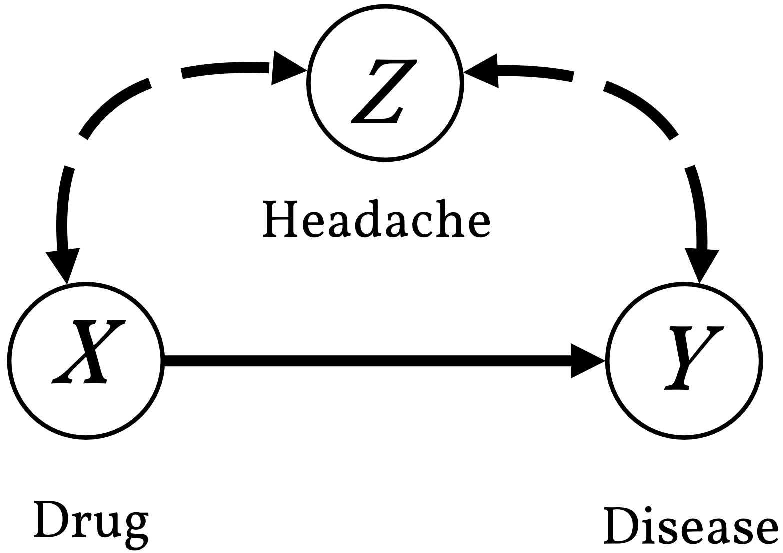
Assumption: $\mathcal{M}$ is recursive i.e., there are no feedback (cyclic) mechanisms

The Causal Graph

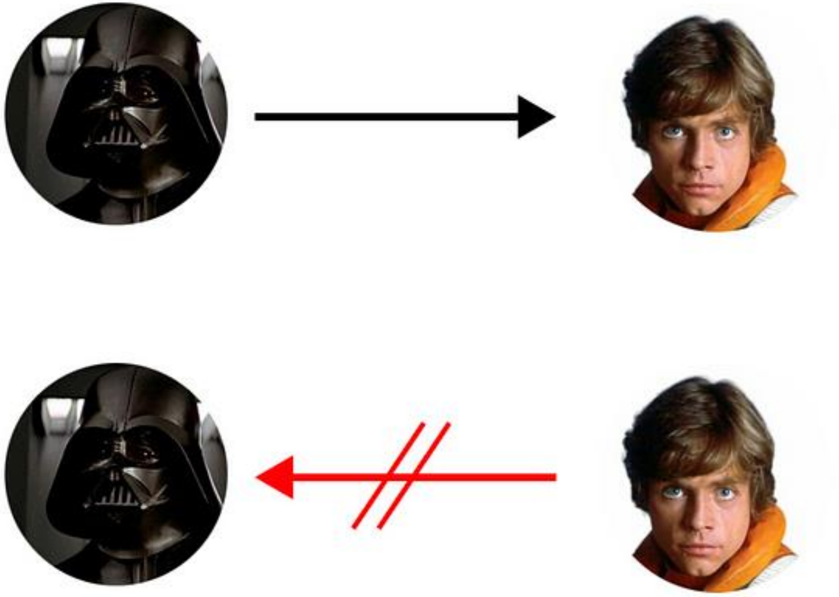
An induced property of the SCM

latent

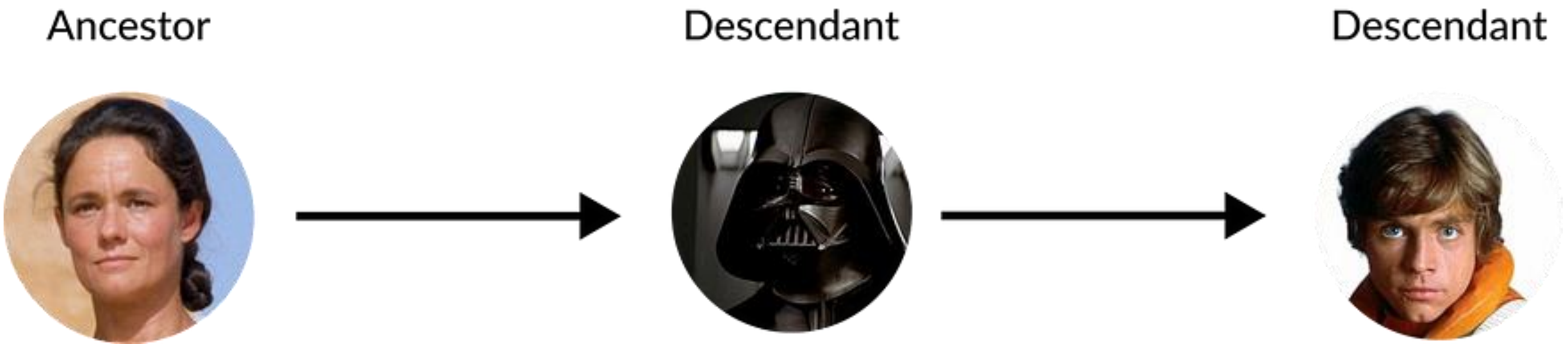
$$\mathcal{F} = \begin{cases} X = f_X(U_X, U_{XZ}) \\ Y = f_Y(X, U_Y, U_{YZ}) \\ Z = f_Z(U_Z) \end{cases}$$



A Causal Graph



https://medium.com/causality-in-data-science/what-are-causal-graphs-abdb50354c8a



https://medium.com/causality-in-data-science/what-are-causal-graphs-abdb50354c8a

The Causal Graph + Data

Structural Causal Model

$$A = U(0, 100)$$

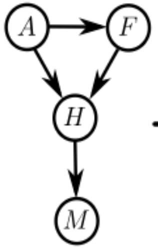
$$F = \frac{1}{2}A + \mathcal{N}(10, 10)$$

$$H = \frac{1}{100}(100 - A^2) + \frac{1}{2}F + \mathcal{N}(40, 30)$$

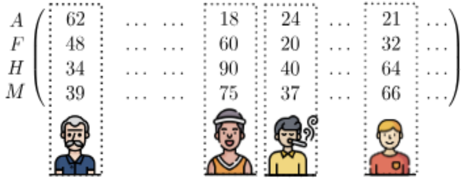
$$M = \frac{1}{2}H + \mathcal{N}(20, 10)$$

(A)ge
(F)ood Habits
(H)ealth
(M)obility

G



D

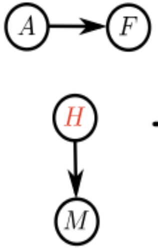


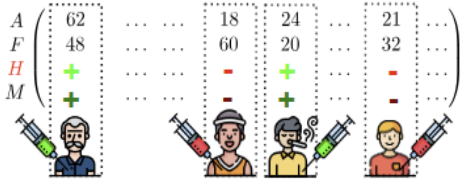
$$A = U(0, 100)$$

$$F = \frac{1}{2}A + \mathcal{N}(10, 10)$$

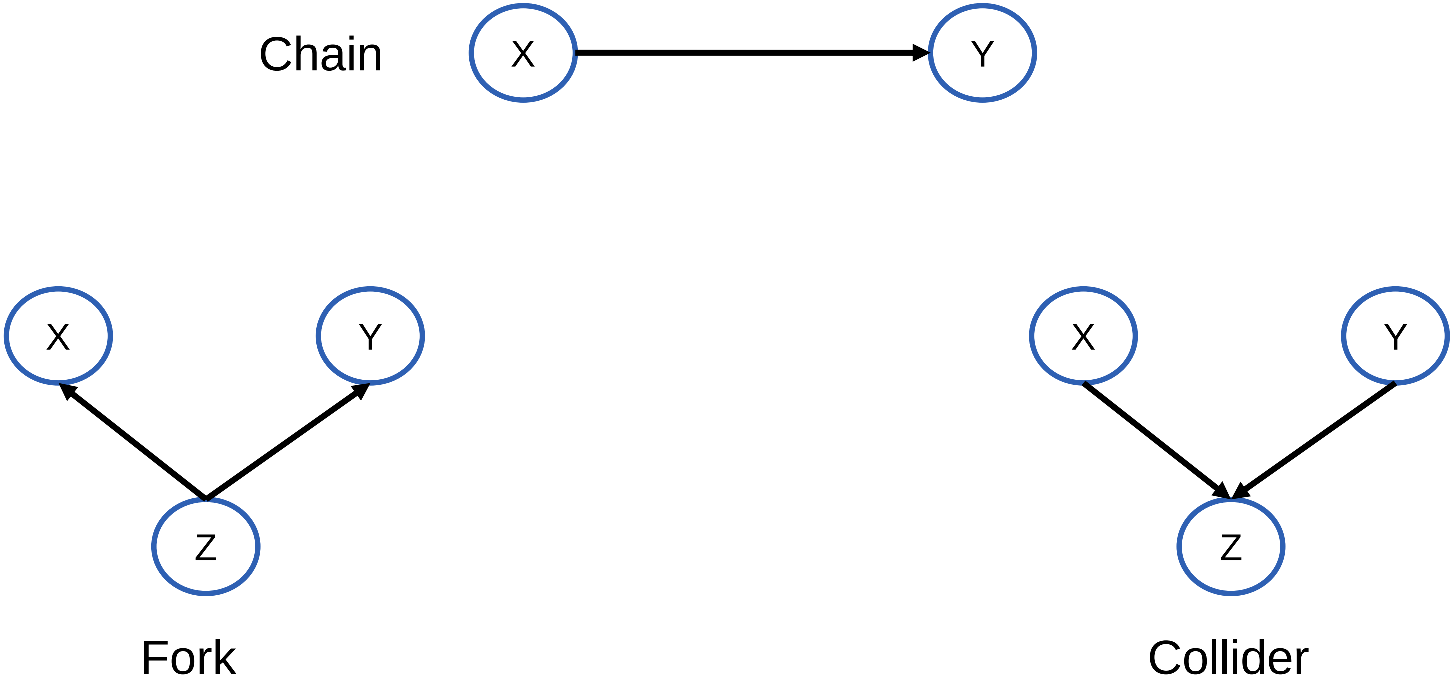
$$H = U(0, 100)$$

$$M = \frac{1}{2}H + \mathcal{N}(20, 10)$$





The 3 Basic Graphs: “gifts from the gods”



How Graphs Encode Independence

- A path p is said to be d-separated by a set of nodes Z iff.

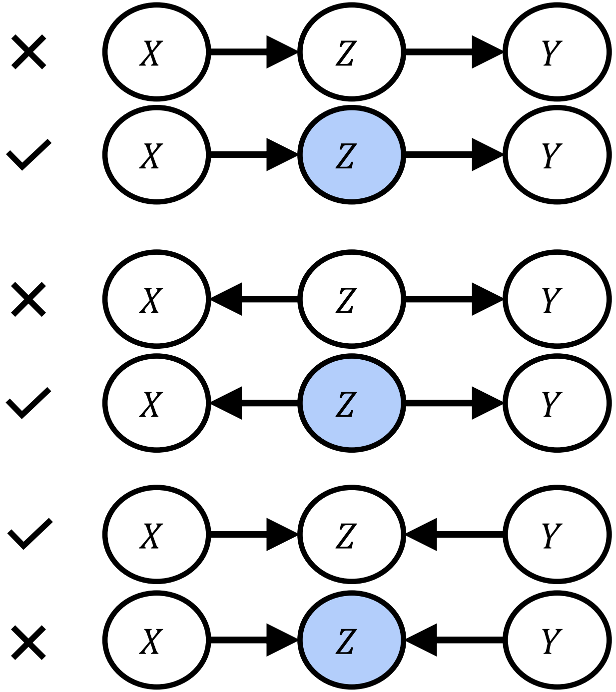
  1. p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node m is in Z, or

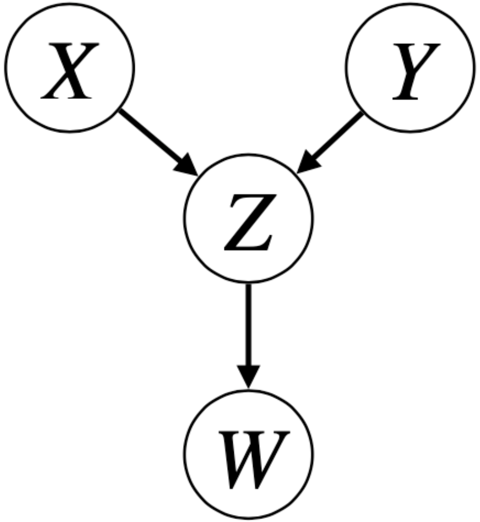
  2. p contains a collider $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and s.t. no descendant of m is in Z

  A set Z is said to d-separate X from Y if and only if Z blocks every path from a node in X to a node in Y

D-separation: Intuition

- Chain: $X \perp_G Y \mid Z$
- Fork: $X \perp_G Y \mid Z$
- Collider: $X \perp_G Y \mid Z$





- X is Diet

  Y is Physical Activity

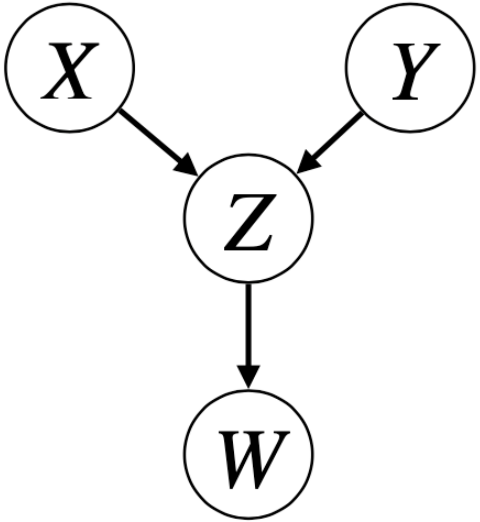
  Z is Obesity

  W is Risk of Heart Disease

- Independences $I(G)$:

  $X \perp_G W \mid Z,$

  $Y \perp_G W \mid Z, \qquad X \perp_G Y$



- For example, knowing that someone has a good diet but are obese let’s us conclude that that someone lacks physical activity
  (up to correctness of G)

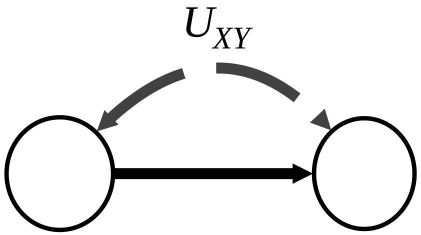
Graphical Representation of an SCM

Structural Causal Model
(SCM)

Graphical Causal Model
(Causal Diagram)

Observational

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

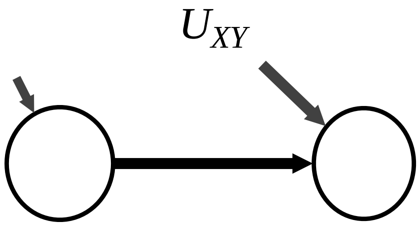


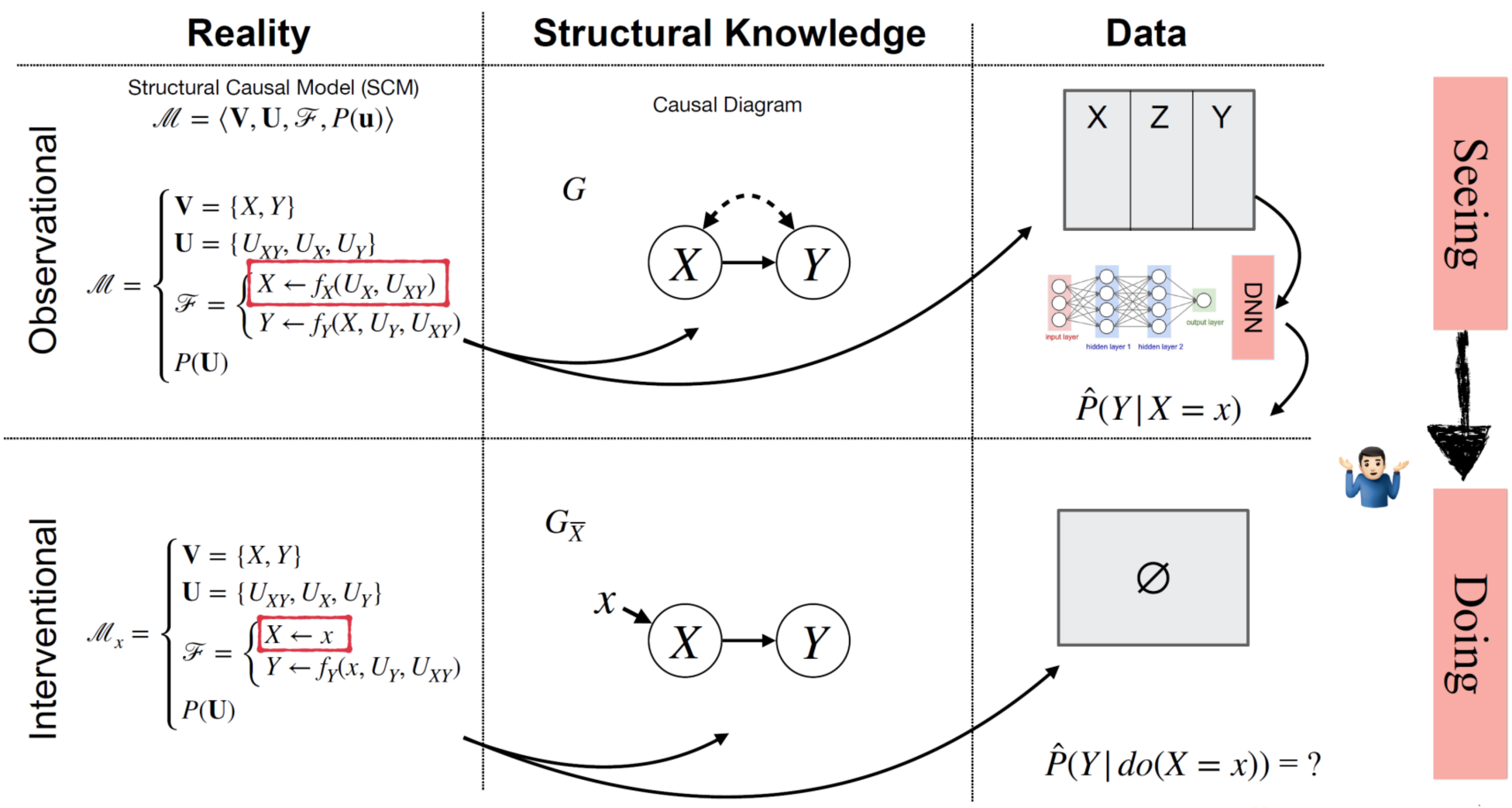
$do(X = x)$

$do(X = x)$

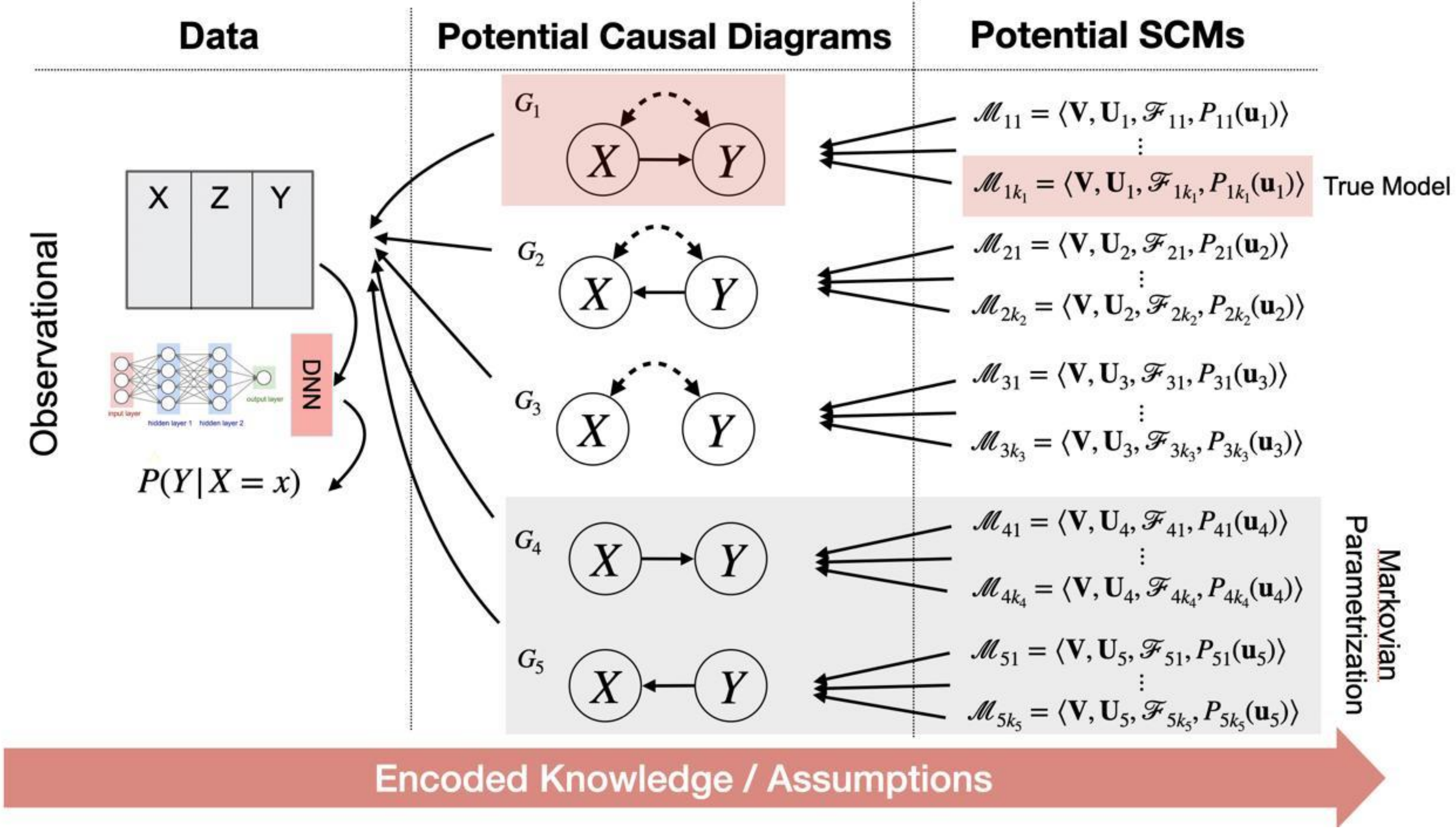
Interventional
$do(X = x)$

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

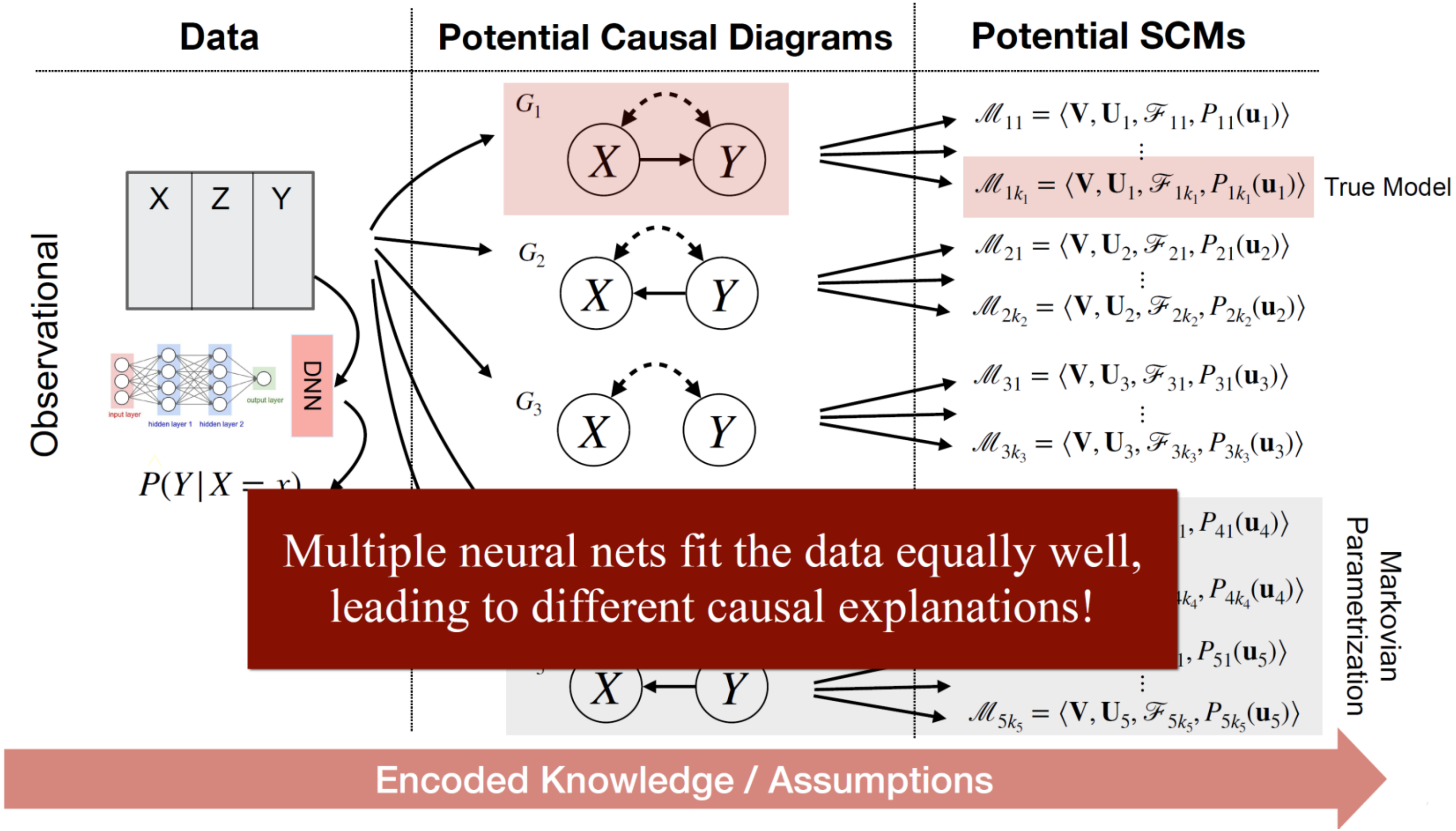




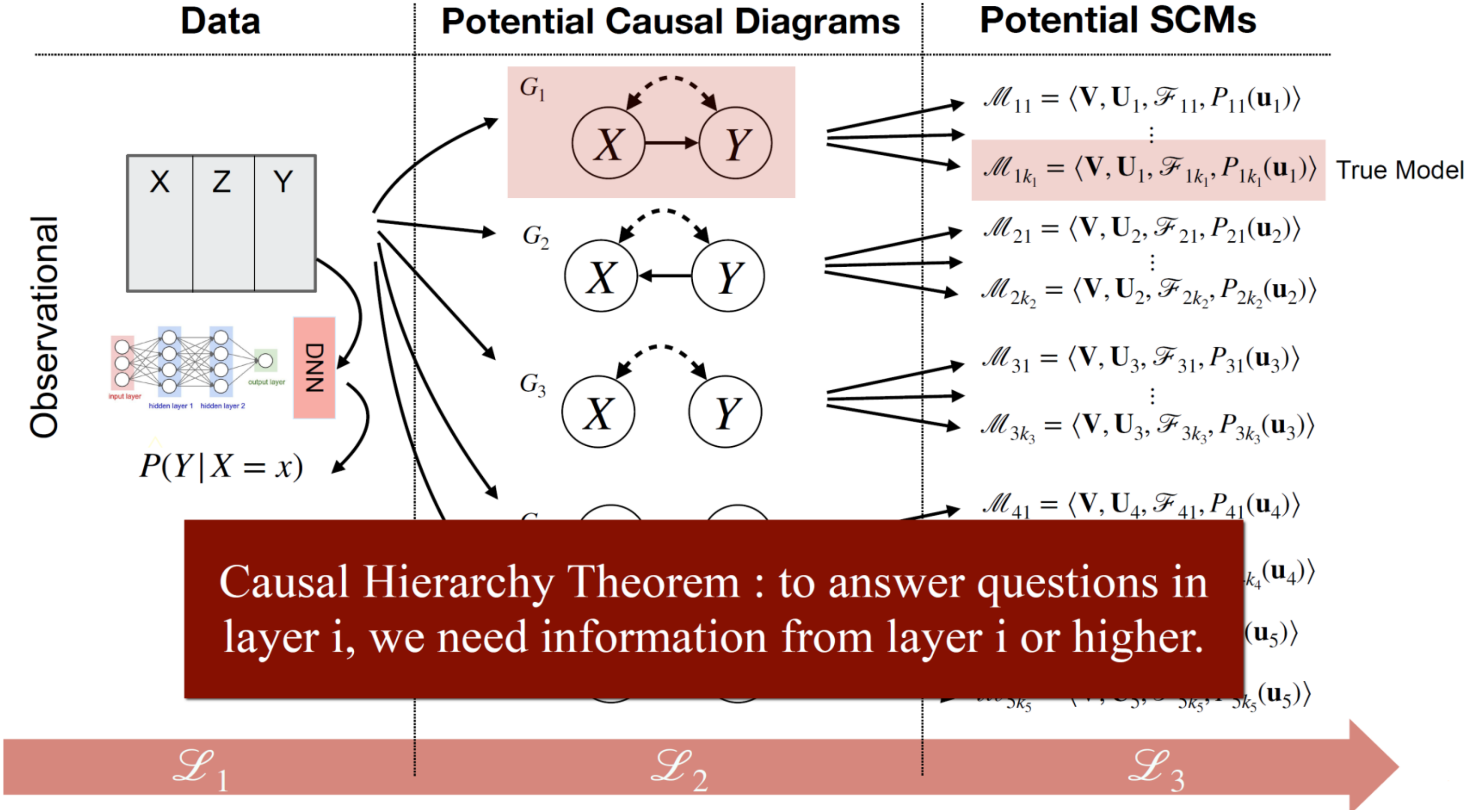
Reality
Structural Knowledge
Data
Structural Causal Model (SCM)
$\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$
Observational
$\mathcal{M} = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} \boxed{X \leftarrow f_X(U_X, U_{XY})} \\ Y \leftarrow f_Y(X, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$
Causal Diagram
G
X
Y
X
Z
Y
input layer
hidden layer 1
hidden layer 2
output layer
DNN
$\hat{P}(Y \mid X = x)$
Seeing
Interventional
$\mathcal{M}_x = \begin{cases} \mathbf{V} = \{X, Y\} \\ \mathbf{U} = \{U_{XY}, U_X, U_Y\} \\ \mathcal{F} = \begin{cases} \boxed{X \leftarrow x} \\ Y \leftarrow f_Y(x, U_Y, U_{XY}) \end{cases} \\ P(\mathbf{U}) \end{cases}$
$G_{\overline{X}}$
x
X
Y
$\varnothing$
$\hat{P}(Y \mid do(X = x)) = ?$
Doing



Data
Potential Causal Diagrams
Potential SCMs
Observational
X
Z
Y
DNN
input layer
hidden layer 1
hidden layer 2
output layer
$P(Y|X=x)$
G_1
G_2
G_3
G_4
G_5
$\mathcal{M}_{11} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{11}, P_{11}(\mathbf{u}_1) \rangle$
$\mathcal{M}_{1k_1} = \langle \mathbf{V}, \mathbf{U}_1, \mathcal{F}_{1k_1}, P_{1k_1}(\mathbf{u}_1) \rangle$
True Model
$\mathcal{M}_{21} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{21}, P_{21}(\mathbf{u}_2) \rangle$
$\mathcal{M}_{2k_2} = \langle \mathbf{V}, \mathbf{U}_2, \mathcal{F}_{2k_2}, P_{2k_2}(\mathbf{u}_2) \rangle$
$\mathcal{M}_{31} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{31}, P_{31}(\mathbf{u}_3) \rangle$
$\mathcal{M}_{3k_3} = \langle \mathbf{V}, \mathbf{U}_3, \mathcal{F}_{3k_3}, P_{3k_3}(\mathbf{u}_3) \rangle$
$\mathcal{M}_{41} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{41}, P_{41}(\mathbf{u}_4) \rangle$
$\mathcal{M}_{4k_4} = \langle \mathbf{V}, \mathbf{U}_4, \mathcal{F}_{4k_4}, P_{4k_4}(\mathbf{u}_4) \rangle$
$\mathcal{M}_{51} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{51}, P_{51}(\mathbf{u}_5) \rangle$
$\mathcal{M}_{5k_5} = \langle \mathbf{V}, \mathbf{U}_5, \mathcal{F}_{5k_5}, P_{5k_5}(\mathbf{u}_5) \rangle$
Markovian Parametrization
Encoded Knowledge / Assumptions



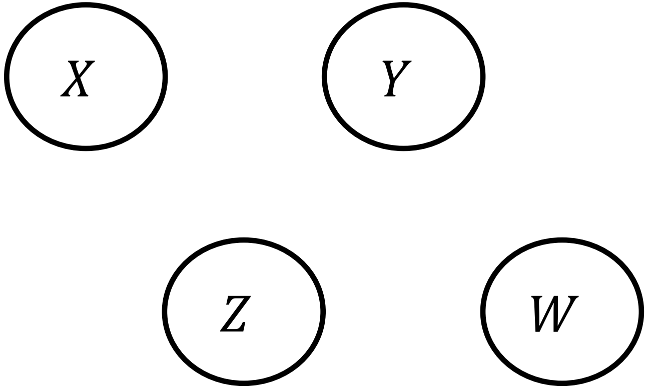
Multiple neural nets fit the data equally well, leading to different causal explanations!



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

Working with DAGs is Hard

- Number of nodes is 4
- Number of possible DAGs is **543**



Super-Exponential Growth

- The number of possible DAGs depending on the number d of nodes:

d	Number of DAGs with d nodes
1	1
2	3
3	25
4	543
5	29281
6	3781503
7	1138779265
8	783702329343
9	1213442454842881
10	4175098976430598143
11	31603459396418917607425
12	521939651343829405020504063
13	18676600744432035186664816926721
14	1439428141044398334941790719839535103
15	237725265553410354992180218286376719253505
16	83756670773733320287699303047996412235223138303
17	62707921196923889899446452602494921906963551482675201
18	99421195322159515895228914592354524516555026878588305014783
19	332771901227107591736177573311261125883583076258421902583546773505

The length of the numbers grows faster than any linear term

Counterfactuals

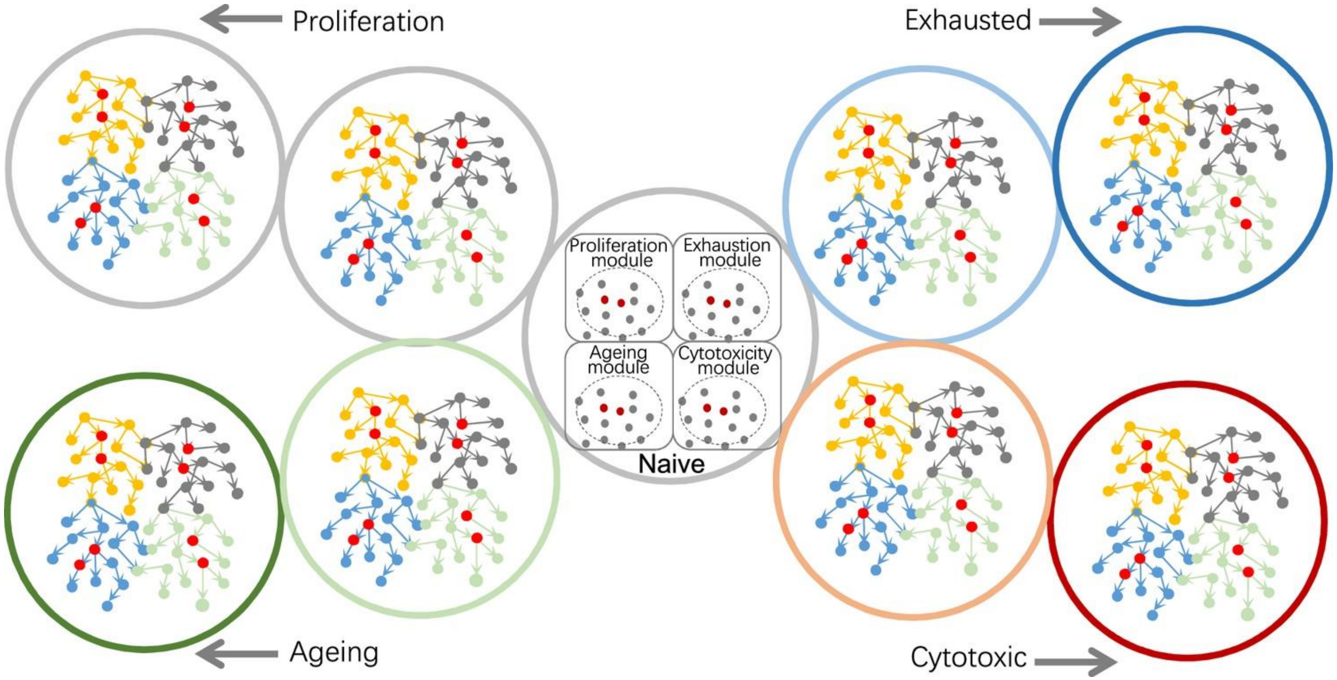
A 3-step procedure

Abduction : Update belief in exogeneous variables given evidence

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

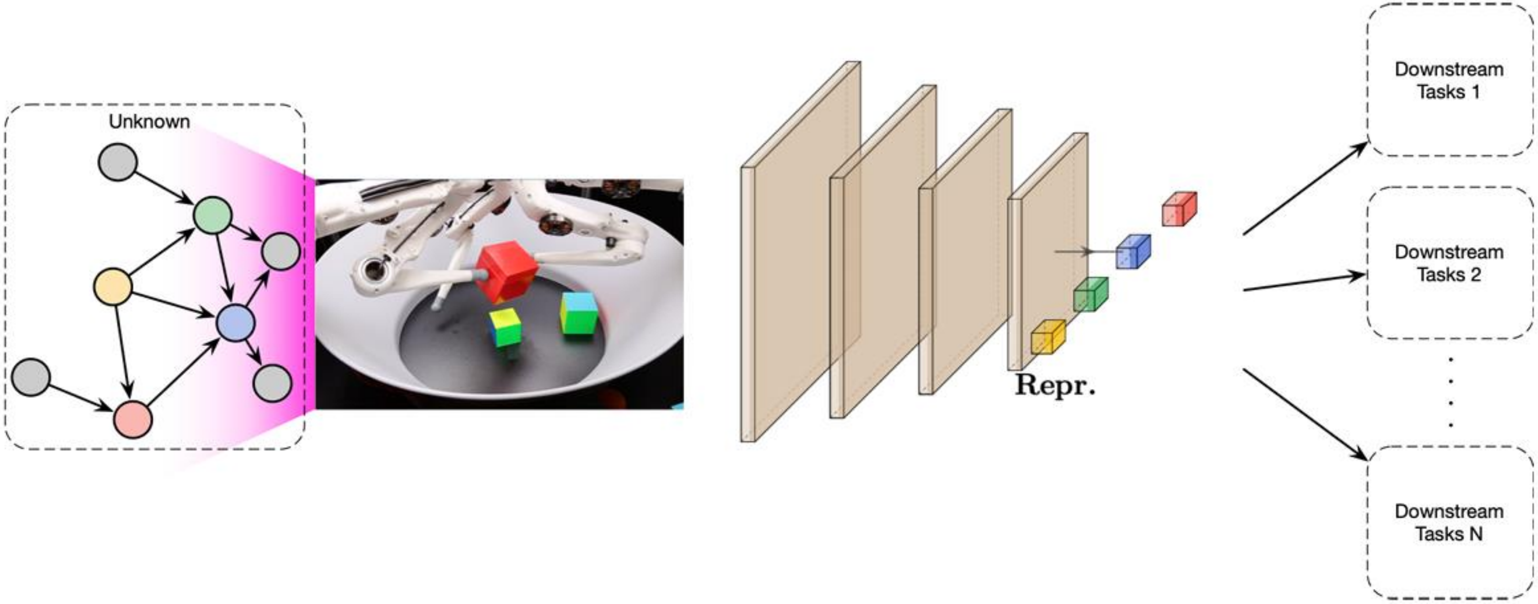
Prediction : Look at variable of interest $P(Y = y)$

Problems to be solved



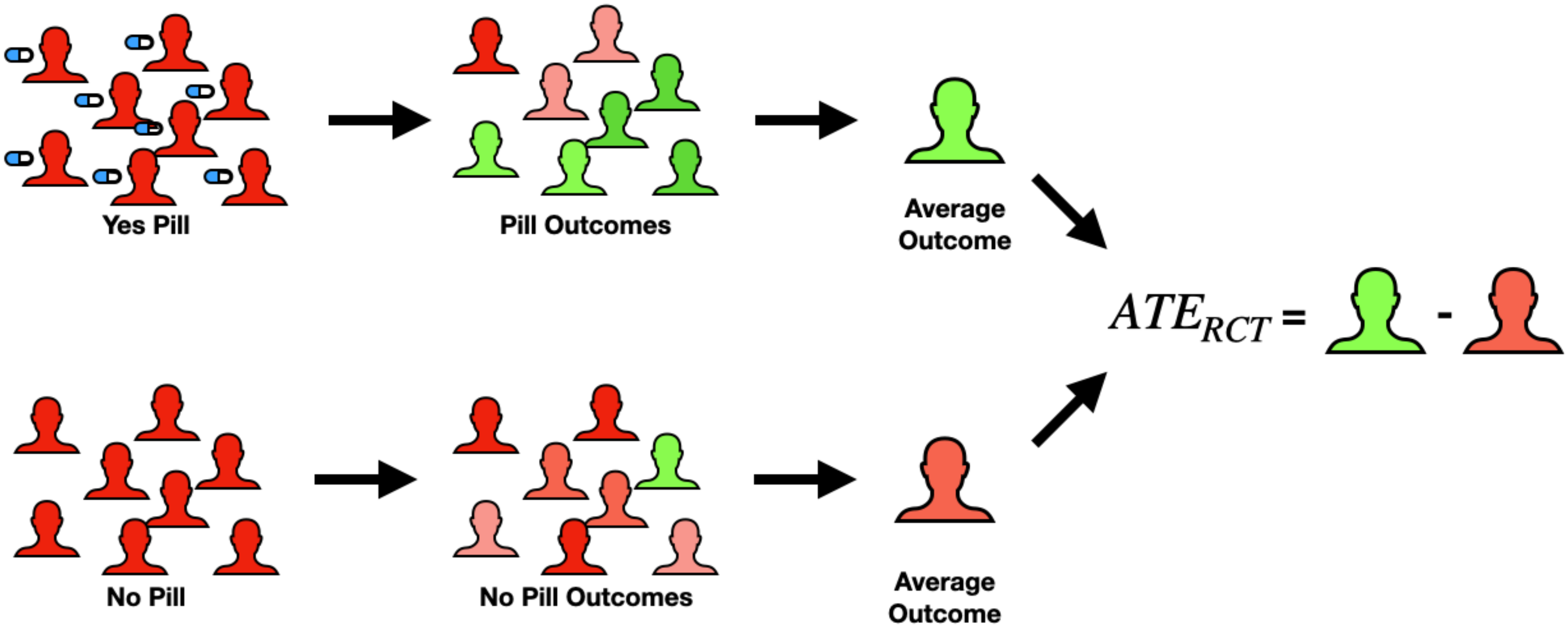
Wen et al., Applying causal discovery to single-cell analyses using CausalCell, eLife 2023

Problems to be solved



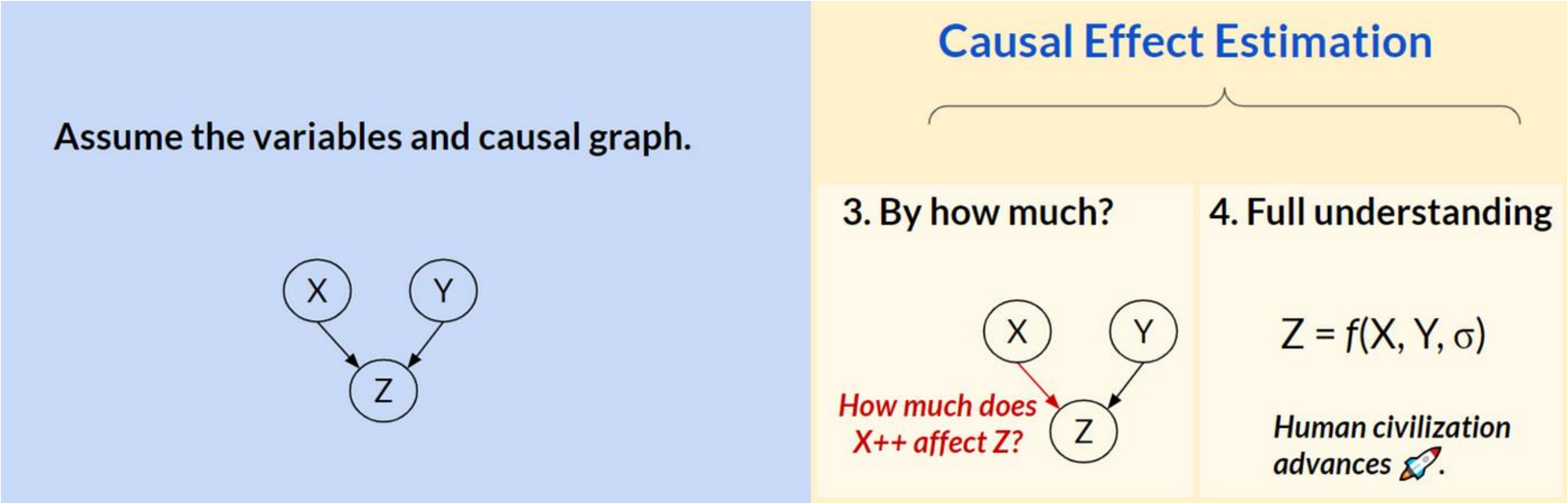
Schölkopf et al., Towards Causal Representation Learning, Proceedings of the IEEE 2021

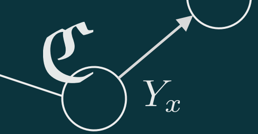
Problems to be solved

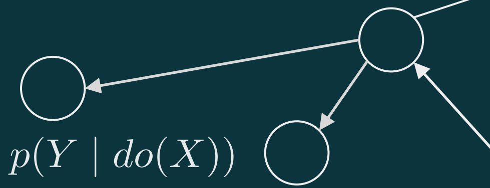


https://towardsdatascience.com/causal-effects-f30f962ffff3

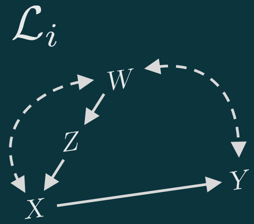
Problems to be solved



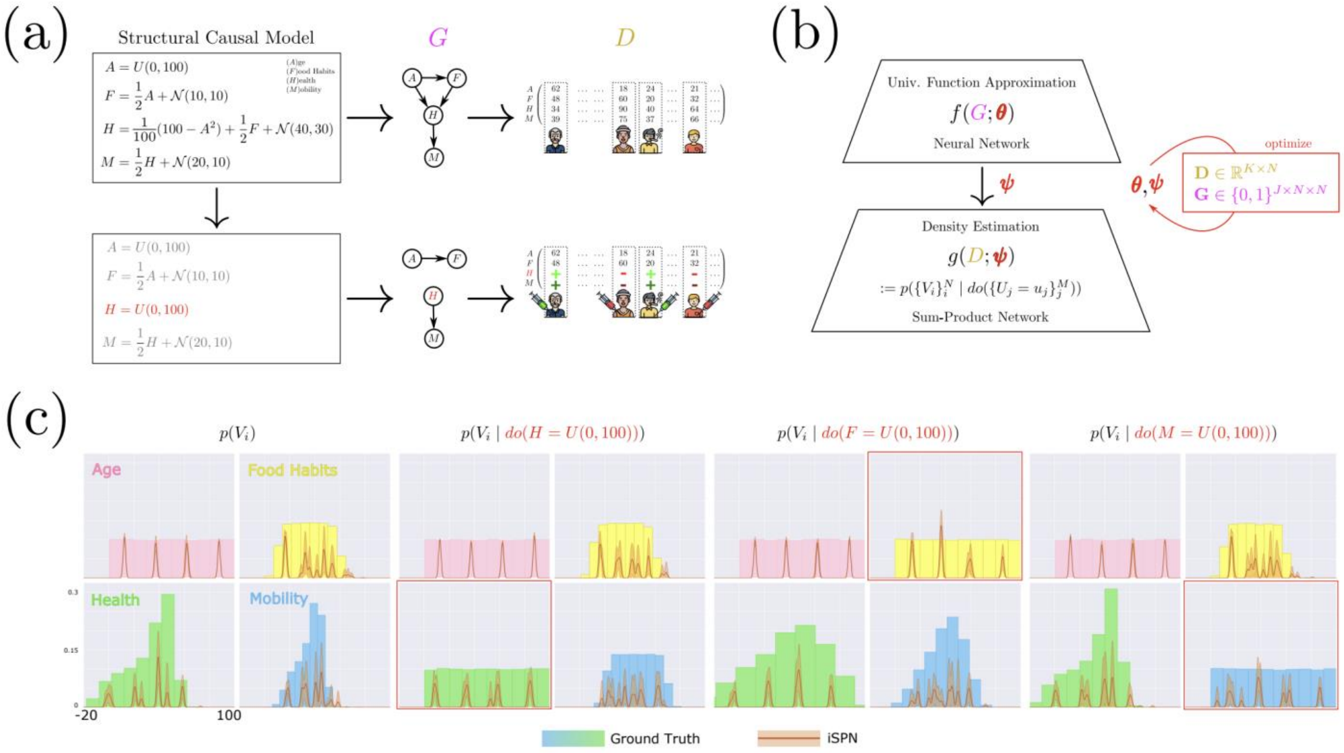




3 Machine Learning for Causality: A Flavour

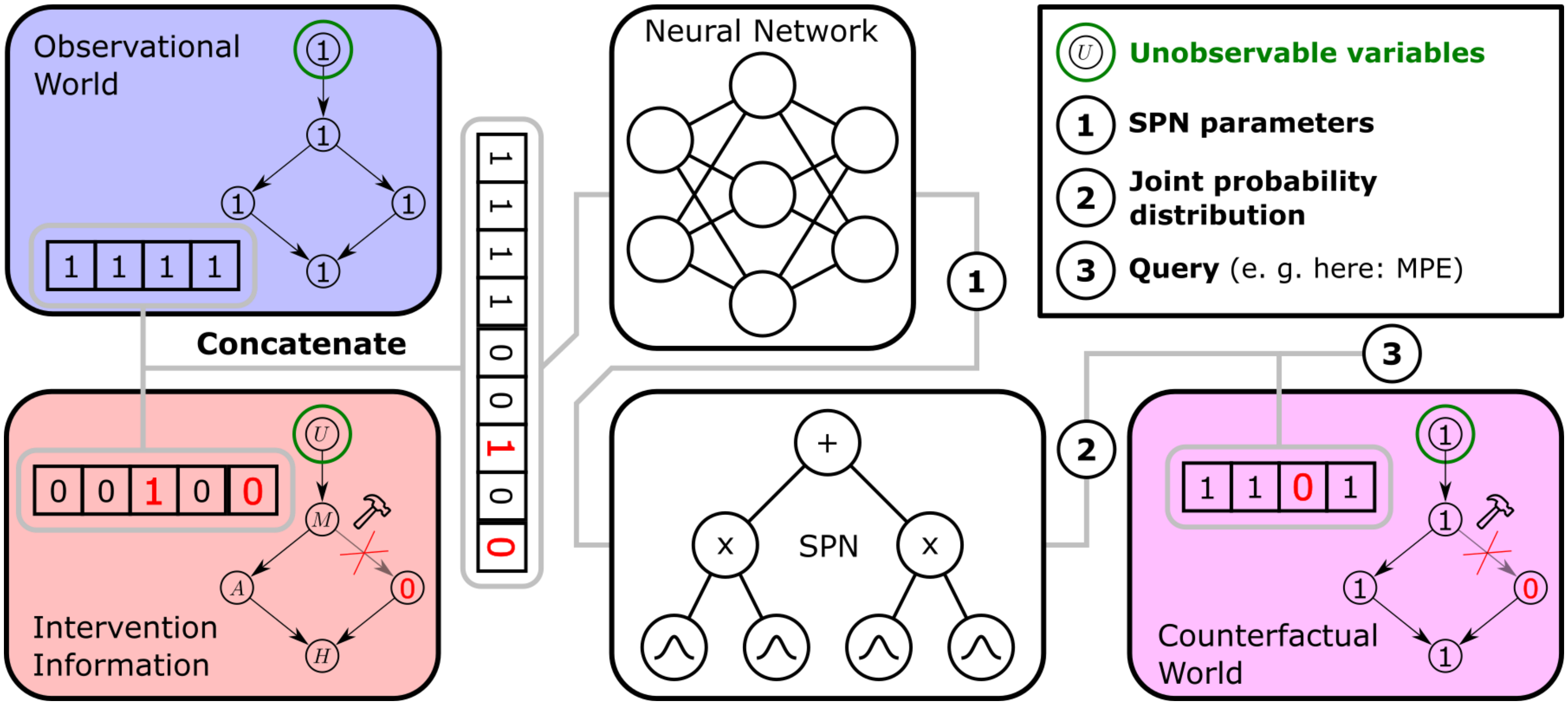


Probabilistic Circuits + Causality



Zečević et al., Interventional SPN, NeurIPS 2021

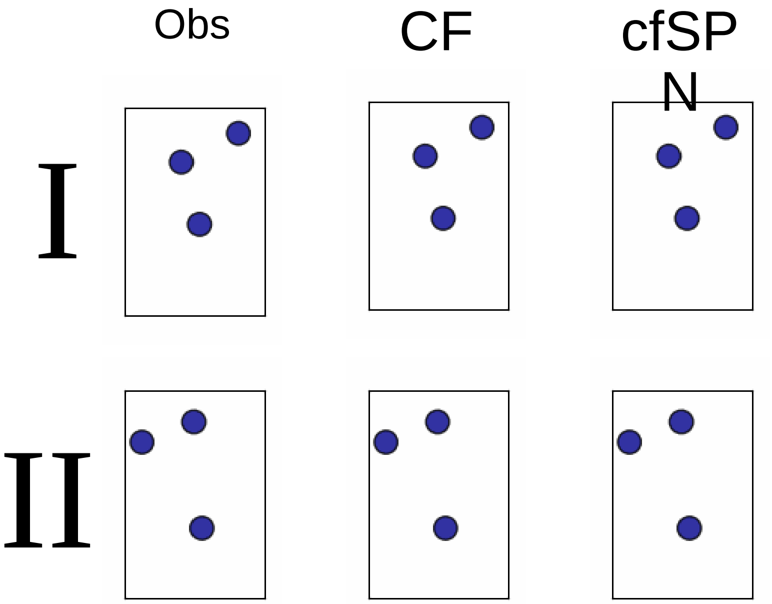
Counterfactual Sum-Product Networks



Busch et al., Computing Counterfactuals using SPNs, Fundamental Challenges in Causality Colloquim 2023; IJAR TBS

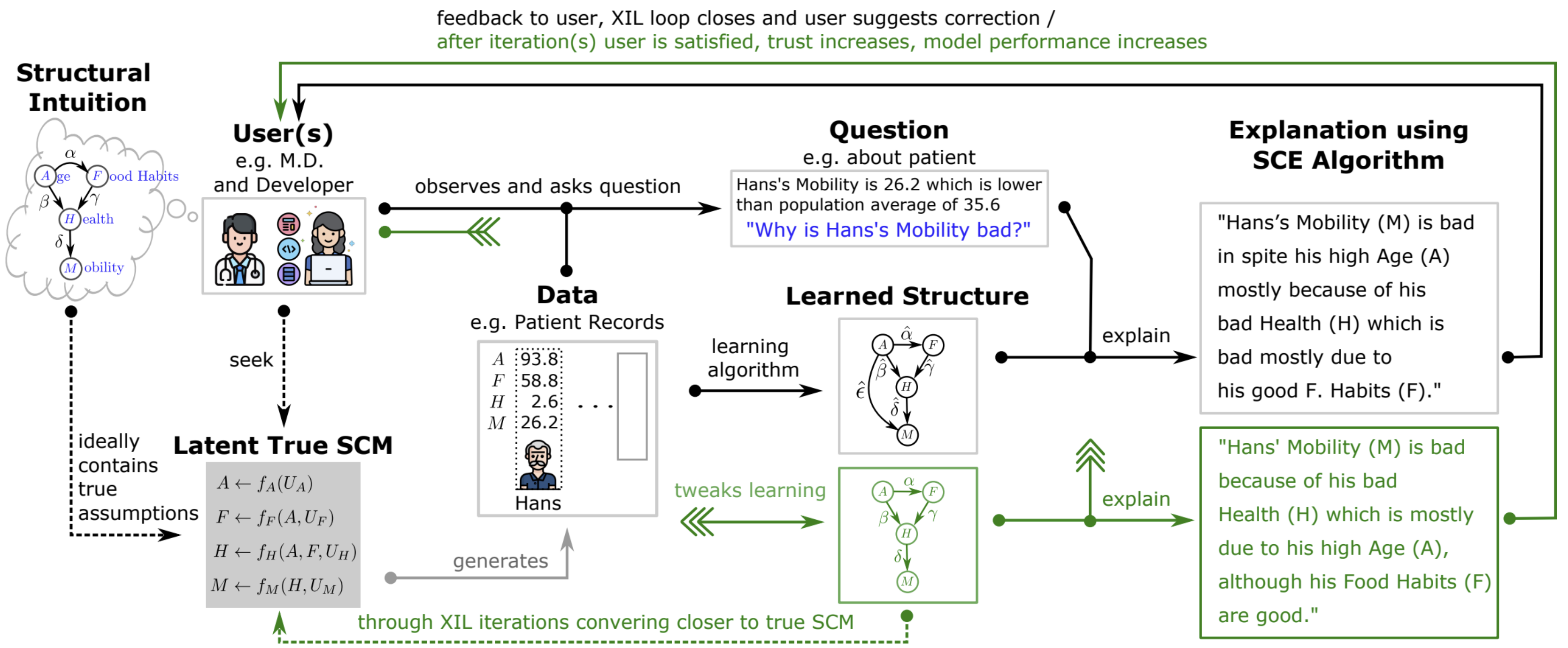
Experiment: Particle Collision

- A difficult problem: Particle simulation with gravity and collisions
- Goal: cfSPN prediction should match true counterfactual simulation (CF)
- **I**: Move the bottom particle to the right after some timesteps
- **II**: Change the velocity of the top particle to slightly upwards at the start



Based on the code repository for particle simulation: https://github.com/ineporozhnii/particles_in_a_box

Logic + Causality = Explanations



Zečević et al. 2022, Explanations of SCMs

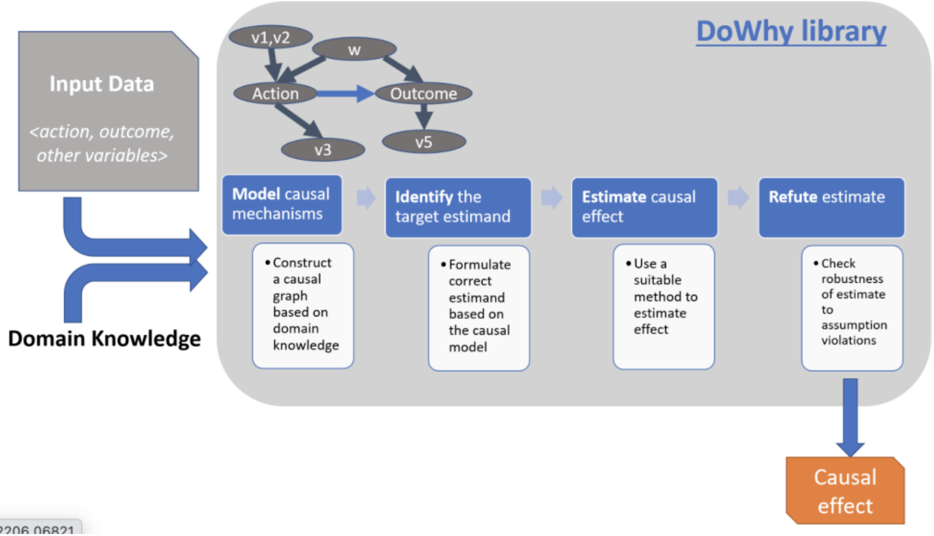
Free Code Libraries

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

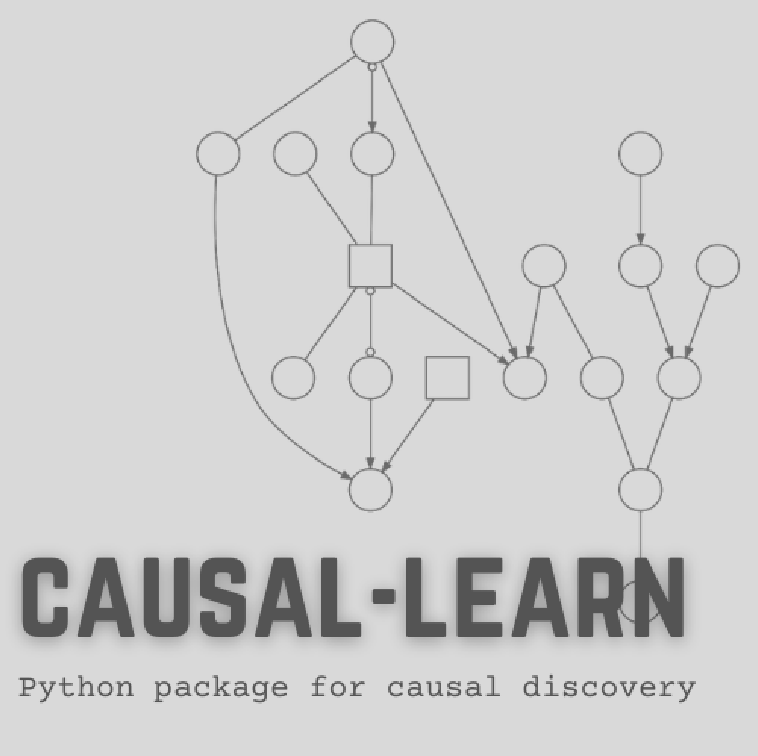
Case Studies using DoWhy: Hotel booking cancellations | Effect of customer loyalty programs | Optimizing article headlines | Effect of home visits on infant health (IHDP) | Causes of customer churn/attrition



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

Free Code Libraries

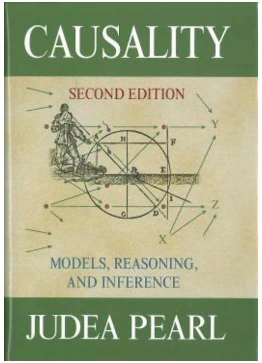
Do it for you

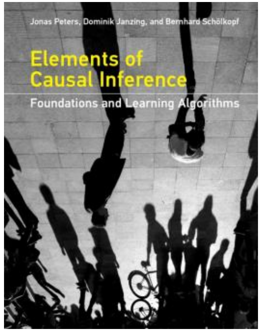


DoWhy, https://github.com/py-why/causal-learn

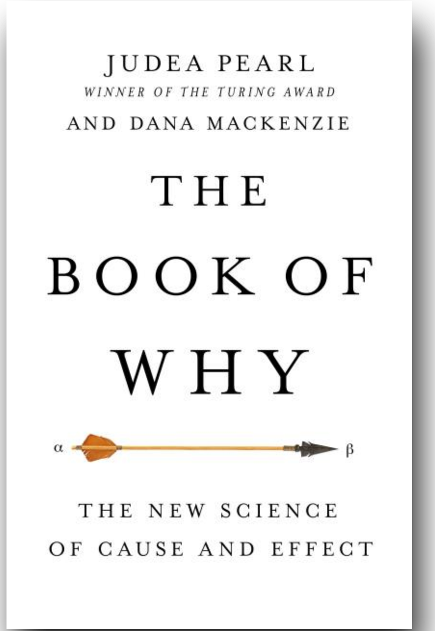
Pointers to Causal Inference References

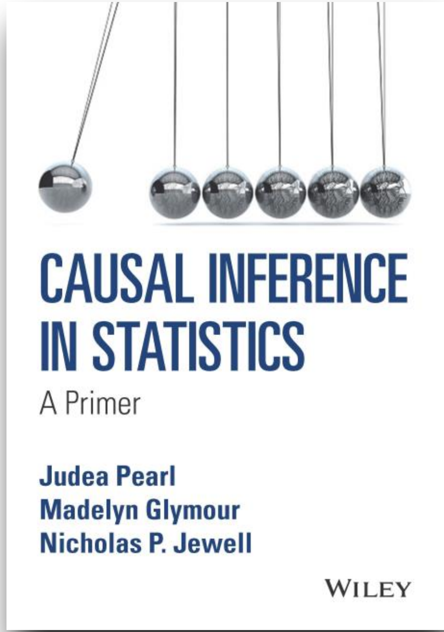
- ❑ Judea Pearl, “**Causality**”, Cambridge University Press, 2009.
- ❑ Peters et al., “**Elements of Causal Inference**”, MIT Press, 2017.
- ❑ Elias Bareinboim Lecture “**Causal Data Science**”, 2019.
  https://www.youtube.com/watch?v=dUsokjG4DHc
- ❑ Brady Neal’s Free Online Course ”**Introduction to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



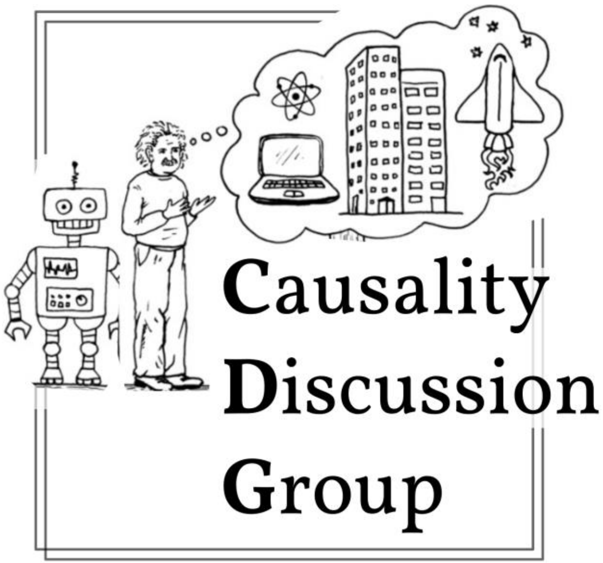


Causality Theory by Judea Pearl





Every Week with Paper Authors → Discuss LIVE



530 members on Slack

Join the community via discuss.causality.link

Past Sessions: [Password: Causa1ity, Direct Access Link]

- Session 01.03.2023 I **Deep Counterfactual Estimation with Categorical Background Variables** I Discussant: Edward De Brouwer
- Session 22.02.2023 I **Information-Theoretic Causal Discovery and Intervention Detection over Multiple Environments** I Discussant: Osman Ali Mian
- Session 08.02.2023 I **CLEAR: Generative Counterfactual Explanations on Graphs** I Discussants: Jing Ma, Ruocheng Guo
- Session 01.02.2023 I **Causal Transformer for Estimating Counterfactual Outcomes** I Discussant: Valentyn Melnychuk
- Session 25.01.2023 I **Abstracting Causal Models** I Discussant: Sander Beckers
- Session 18.01.2023 I **Desiderata for Representation Learning: A Causal Perspective** I Discussant: Yixin Wang
- Session 11.01.2023 I **Causal Feature Selection via Orthogonal Search** I Discussant: Ashkan Soleymani
- Session 14.11.2022 I **Rewind 2022** I Final session of 2022 to simply rewind on what we experienced throughout the year
- Session 07.12.2022 I **Causal Inference Through the Structural Causal Marginal Problem** I Discussant: Luigi Gresele
- Session 30.11.2022 I **Selecting Data Augmentation for Simulating Interventions** I Discussant: Maximilian Ilse
- Session 23.11.2022 I **On Disentangled Representations Learned from Correlated Data** I Discussant: Frederik Träuble
- Session 16.11.2022 I **Causal Curiosity: RL Agents Discovering Self-supervised Experiments for Causal Repr. Learning** I Discussant: Sumedh Sontakke
- Session 09.11.2022 I **Causal Machine Learning: A Survey and Open Problems** I Discussants: Jean Kaddour, Aengus Lynch
- Session 02.11.2022 I **A Critical Look at the Consistency of Causal Estimation with Deep Latent Variable Models** I Discussant: Severi Rissanen
- Session 26.10.2022 I **Nonlinear Invariant Risk Minimization: A Causal Approach** I Discussant: Chaochao Lu
- Session 19.10.2022 I **CausalVAE: Disentangled Representation Learning via Neural Structural Causal Models** I Discussant: Mengyue Yang
- Session 12.10.2022 I **Weakly Supervised Causal Representation Learning** I Discussant: Johann Brehmer
- Session 05.10.2022 I **Towards Causal Representation Learning** I Discussant: Anirudh Goyal
- Session 21.09.2022 I **Selection Collider Bias in Large Language Models** I Discussant: Emily McMilin
- Session 14.09.2022 I **The Causal-Neural Connection: Expressiveness, Learnability, and Inference** I Discussants: Kai-Zhan Lee, Kevin Xia
- Session 07.09.2022 I **Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style** I Discussant: Julius von Kügelgen

⋮

50+ Sessions Completed and All Recorded

Genealogy of Causality

Access via genealogy.causality.link

Name	Institution	Supervisor	Location	Previous Positions
UCLA				
Judea Pearl	UCLA	?	US	Rutgers, Technion, New
Wesley Salmon	UCLA	Hans Reichenbach	US	?
Hans Reichenbach	UCLA	Paul Hensel, Max Noeth	US	Berlin, Istanbul, Erlange
John Hopkins				
Ilya Shpitser	John Hopkins		US	UCLA, Judea Pearl
Oregon State University				
Karthika Mohan	Oregon State University	Judea Pearl	US	
CMU				
Kun Zhang	CMU		Pittsburgh, US	MPI Tübingen
Clark Glymour	CMU	Wesley Salmon	Pittsburgh, US	
Peter Spirtes	CMU		Pittsburgh, US	
ETH Zürich				
Peter Bühlmann	ETH		Zürich	?
Marloes Maathuis	ETH		Zürich	?
Nicolai Meinshausen	ETH			
LMU Munich				
Stephan Hartmann	LMU		Munich, Germany	
MPI Tübingen				
Bernhard Schölkopf	MPI Tübingen	Vladimir Vapnik	Tübingen, Germ	TU Berlin
Ulrike von Luxburg	MPI Tübingen		Tübingen, Germany	
Michel Besserve				

Genealogy of Causality

Access via genealogy.causality.link

Name	Institution	Supervisor	Location	Previous Position
UCLA				

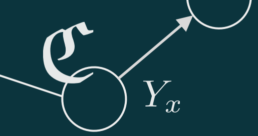
To Extend This!

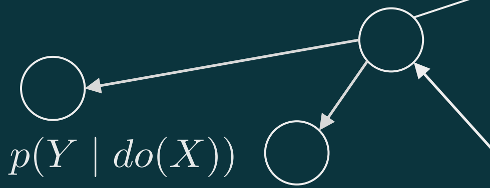
Join the community via discuss.causality.link

Access the genealogy via genealogy.causality.link

Genealogy of Causality

Name			ion	Previous Positions
UCLA				
Judea Pearl	[illegible]		[illegible]	Rutgers, Technion, New
Wesley Salmon	UCLA	Hans Reichenbach	US	?
Hans Reichenbach	UCLA	Paul Hensel, Max Noeth	US	Berlin, Istanbul, Erlange
John Hopkins				
Ilya Shpitser	John Hopkins		US	UCLA, Judea Pearl
Oregon State University				
Karthika Mohan	Oregon State University	Judea Pearl	US	
CMU				
Kun Zhang	CMU		Pittsburgh, US	MPI Tübingen
Clark Glymour	CMU	Wesley Salmon	Pittsburgh, US	
Peter Spirtes	CMU		Pittsburgh, US	
ETH Zürich				
Peter Bühlmann	ETH		Zürich	?
Marloes Maathuis	ETH		Zürich	?
Nicolai Meinshausen	ETH			
LMU Munich				
Stephan Hartmann	LMU		Munich, Germany	
MPI Tübingen				
Bernhard Schölkopf	MPI Tübingen	Vladimir Vapnik	Tübingen, Germ	TU Berlin
Ulrike von Luxburg	MPI Tübingen		Tübingen, Germany	
Michel Besserve				





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)