Causal and Probabilistic Reasoning

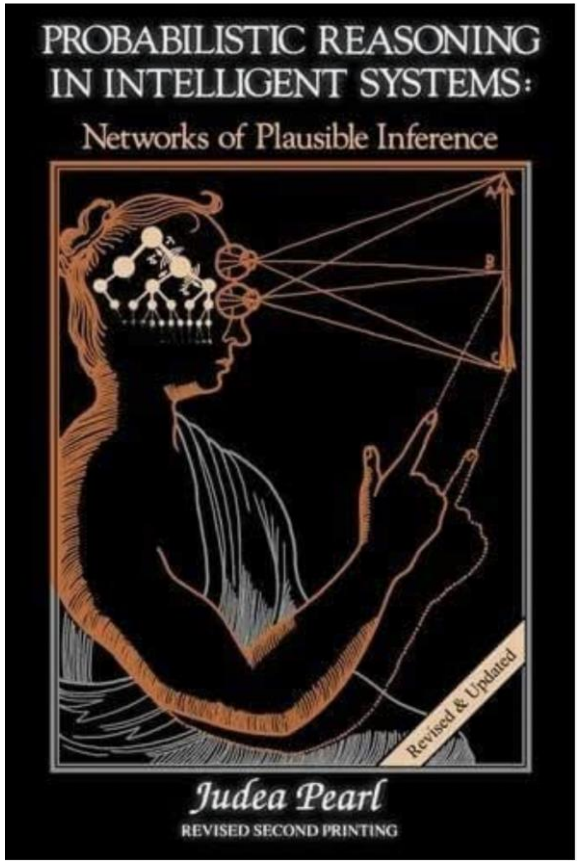
Slides Set 7:

Introduction to Causality

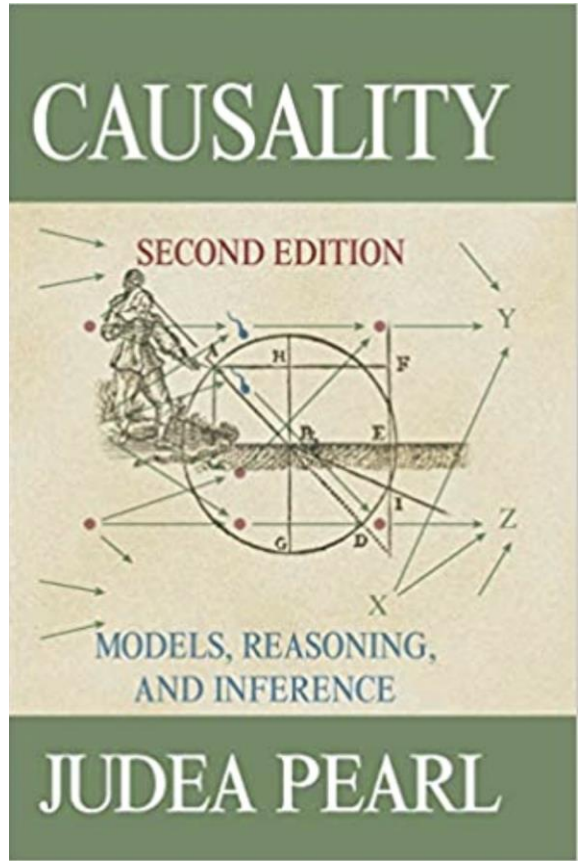
Rina Dechter

(**Causal Inference in Statistics, A primer, J. Pearl, M Glymur and N. Jewell Ch1, Why, ch1**

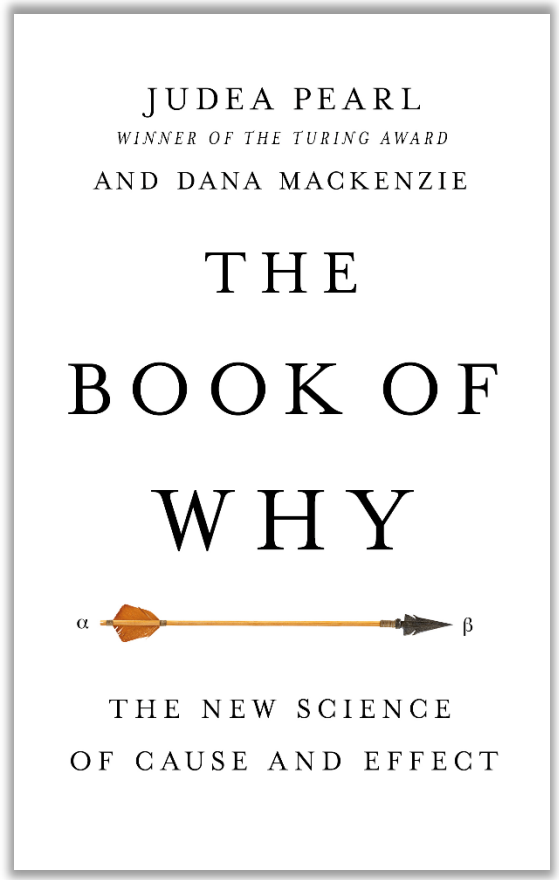
1987

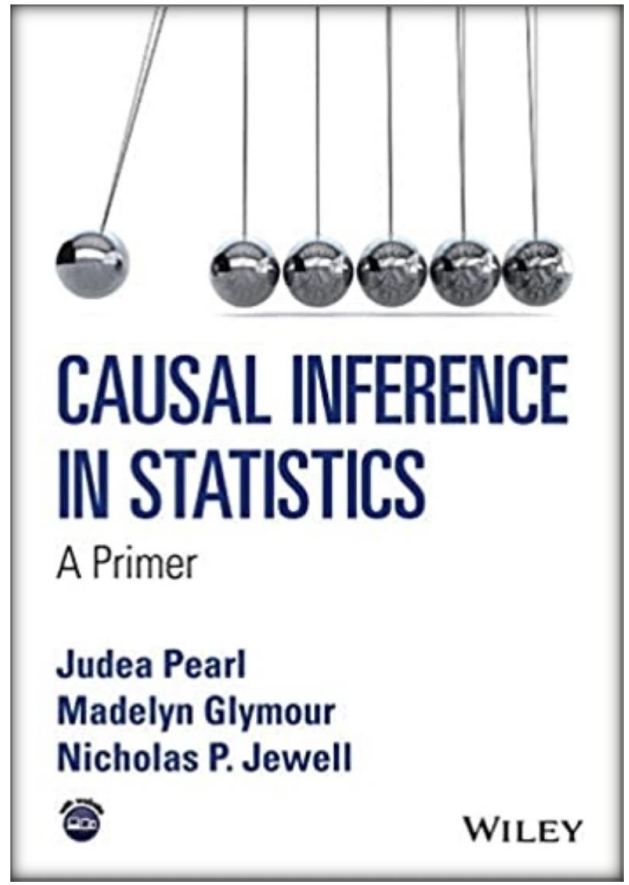


2009



2018



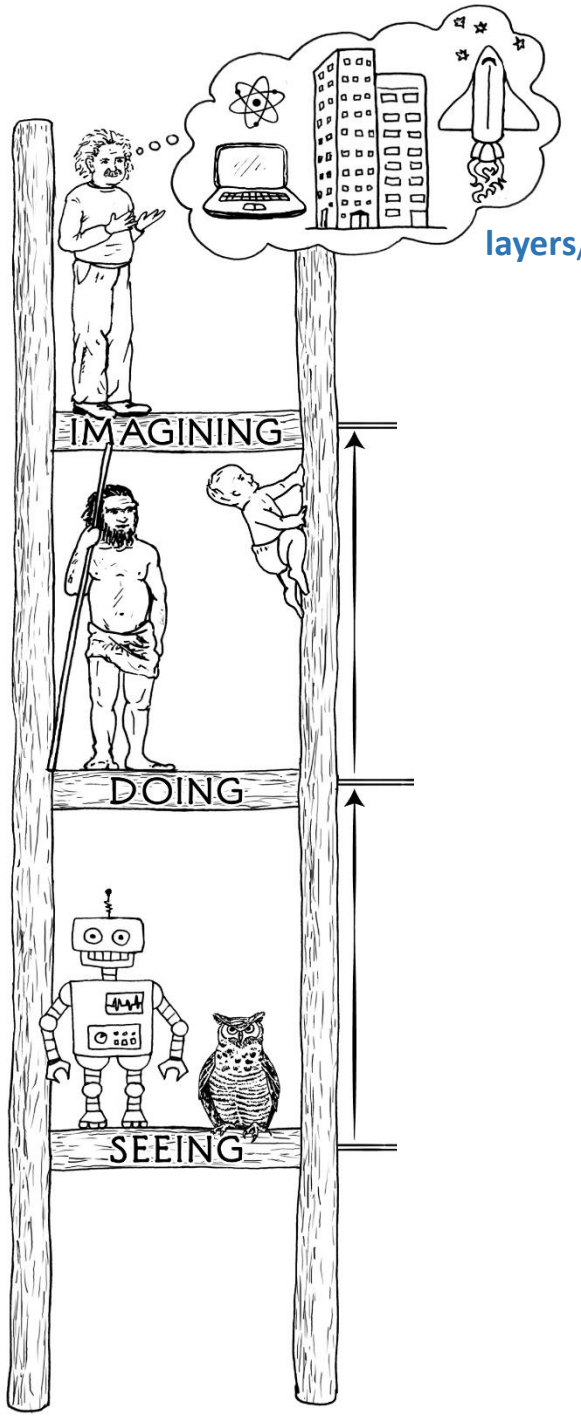


association

intervention & counterfactuals

http://bayes.cs.ucla.edu/PRIMER/

Why Causality?



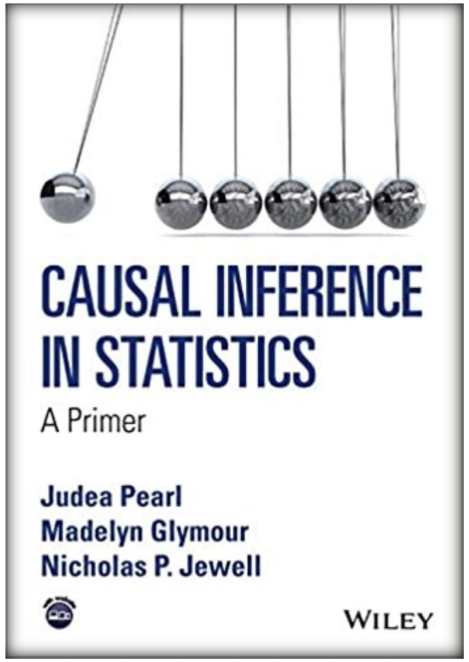
layers/rungs of the causal hierarchy

his course
econd part

this course
First part

Outline

- Simpson Paradox
- The causal Hierarchy
- Structural Causal Models
- Product form of Graphical models.



Outline

- Simpson Paradox
- The causal Hierarchy
- Structural Causal Models
- Product form of Graphical models.

The Simpson Paradox

- (Simpson 1951) a group of sick patients are given the option to try a new drug. Among those who took the drug, a lower percentage recover than among those who did not. However, when we partition by gender, we see that more men taking the drug recover than do men not taking the drug, and more women taking the drug recover than do women not taking the drug!
- We record the recovery rates of 700 patients who were given access to the drug. 350 patients chose to take the drug and 350 patients did not. We got:

	Drug	No drug
Men	81/87 recovered (**93%**)	234/270 recovered (**87%**)
Women	192/263 recovered (**73%**)	55/80 recovered (**69%**)
Combined	273/350 recovered (**78%**)	289/350 recovered (**83%**)

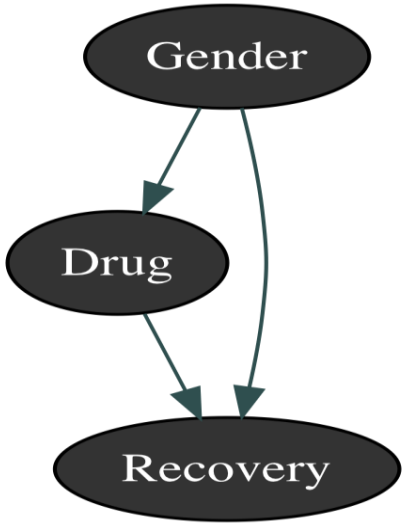
What is the right answer?

So, given the results of the study, should the doctor prescribe the drug for a man? For a woman? Or when gender is unknown?

The Simpson Paradox

- (Simpson 1951) a group of sick patients are given the option to try a new drug. Among those who took the drug, a lower percentage recover than among those who did not. However, when we partition by gender, we see that more men taking the drug recover than do men not taking the drug, and more women taking the drug recover than do women not taking the drug!
- We record the recovery rates of 700 patients who were given access to the drug. 350 patients chose to take the drug and 350 patients did not. We got:

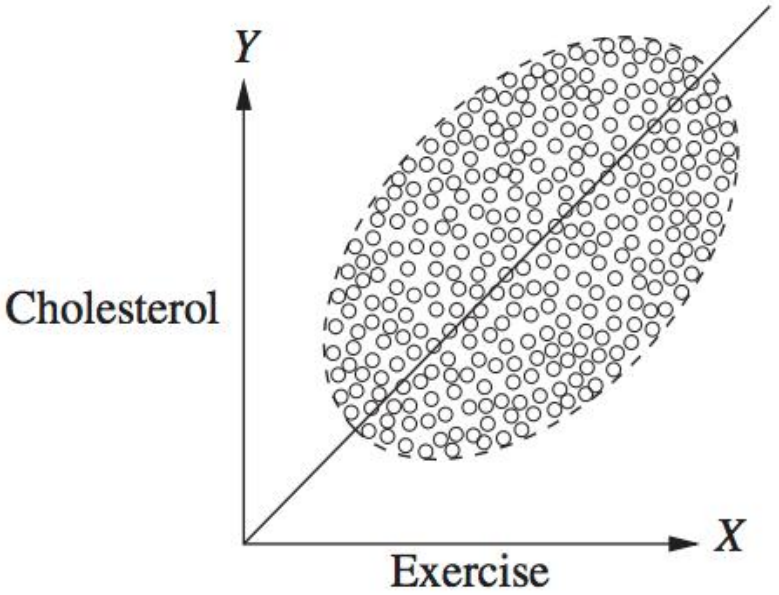
	Drug	No drug
Men	81/87 recovered (**93%**)	234/270 recovered (**87%**)
Women	192/263 recovered (**73%**)	55/80 recovered (**69%**)
Combined	273/350 recovered (**78%**)	289/350 recovered (**83%**)

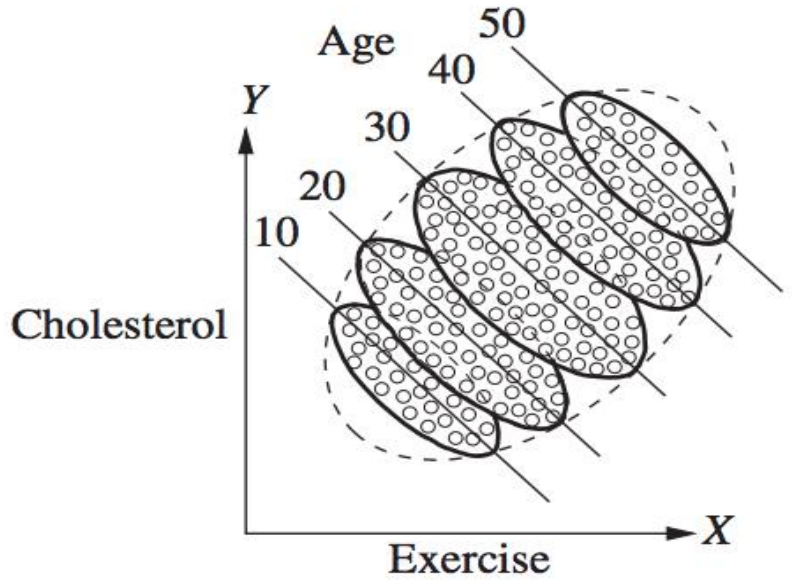


The answer cannot be found in the data!! We need to know the story of how the data was generated.

The Simpson Paradox

- The same phenomenon with continuous variables. Example: Impact of exercise on Cholesterol for different age groups:





- Because, Age is a common cause of both treatment (exercise) and outcome (cholesterol). So we should look at the age-segregated data in order to compare same-age people, and thereby eliminate the possibility that the high exercisers in each group we examine are more likely to have high cholesterol due to their age, and not due to exercising.

The Simpson Paradox

- Segregated data is not always the right way. What if we record blood (BP) pressure (post intervention) instead of gender?
- We know that drug lower blood pressure but also has a toxic effect.

- Would you recommend the drug to a patient?
- In the general population, the drug might improve recovery rates because of its effect on blood pressure. But in the subpopulations—the group of people whose post-treatment BP is high and the group whose post-treatment BP is low—we of course would not see that effect; we would only see the drug's toxic effect.
- In this case the aggregated data should be consulted.
- **Same data opposite conclusions!!!**

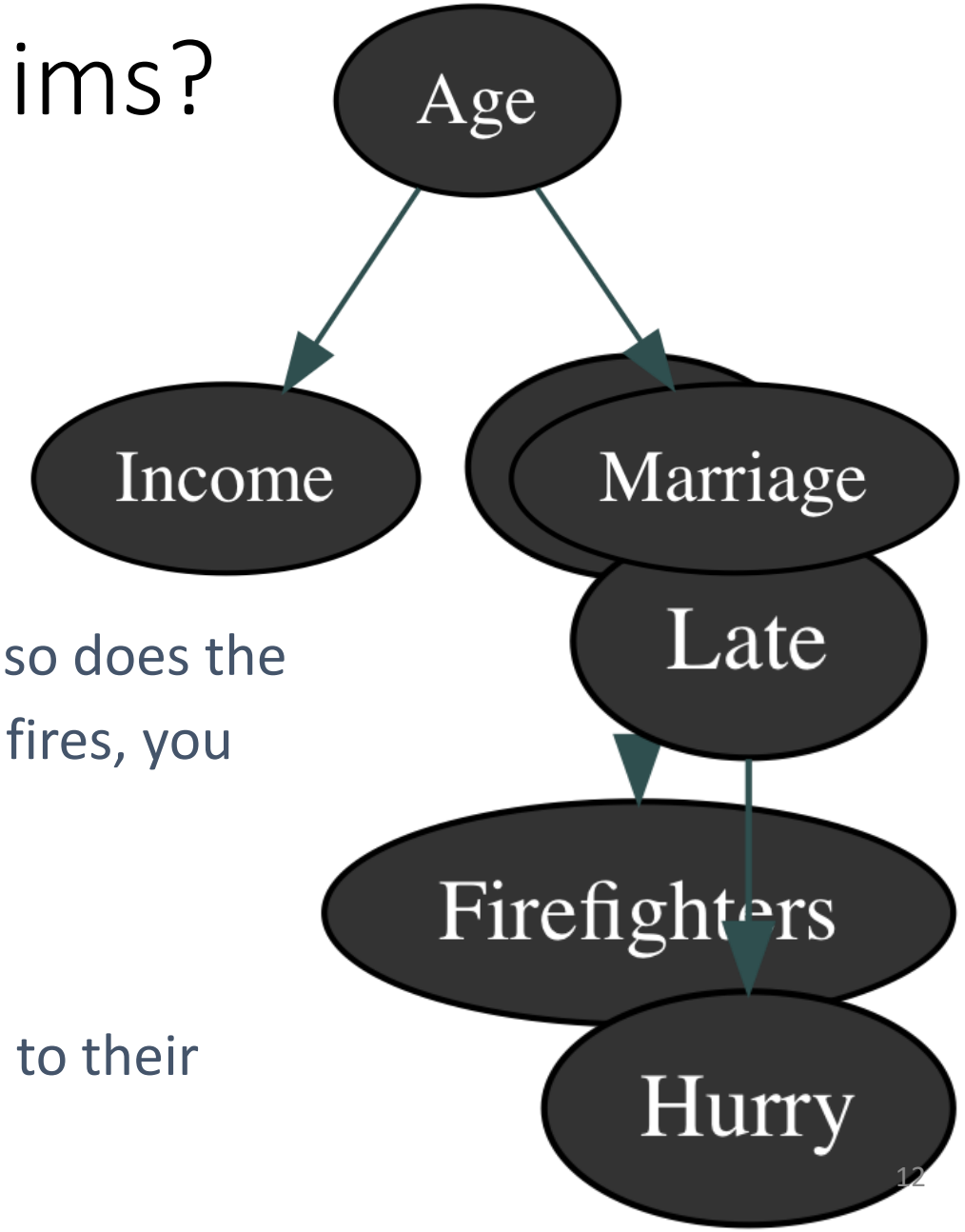
	No drug	Drug
Low BP	81/87 recovered (**93%**)	234/270 recovered (**87%**)
High BP	192/263 recovered (**73%**)	55/80 recovered (**69%**)
Combined	273/350 recovered (**78%**)	289/350 recovered (**83%**)

What is Wrong With These Claims?

- "Data show that income and marriage have a high positive correlation. Therefore, your earnings will increase if you get married."

- "Data shows that as the number of fires increase, so does the number of firefighters. Therefore, to cut down on fires, you should reduce the number of firefighters."

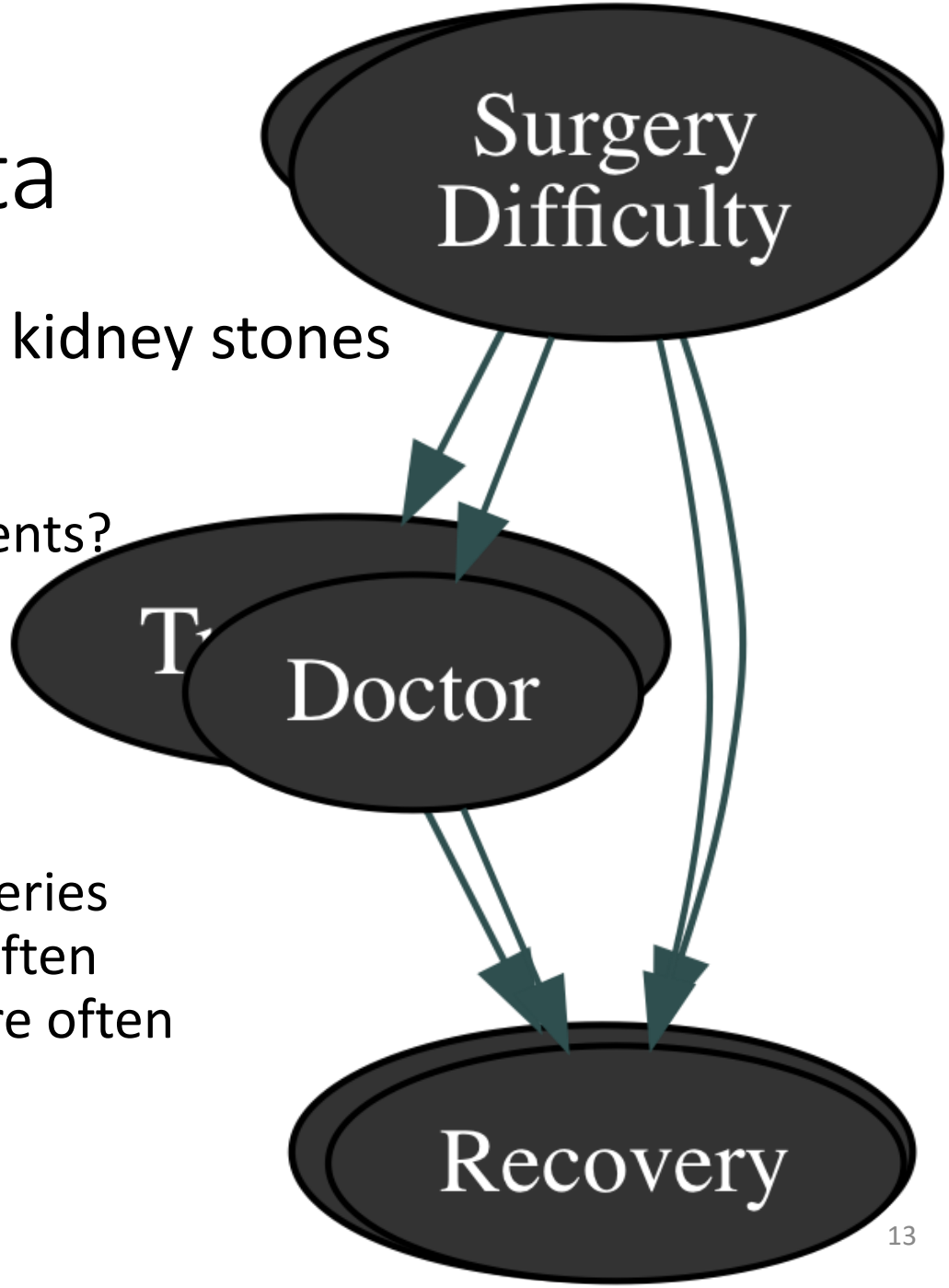
- "Data show that people who hurry tend to be late to their meetings. Don't hurry, or you'll be late."



Aggregated or Segregated Data

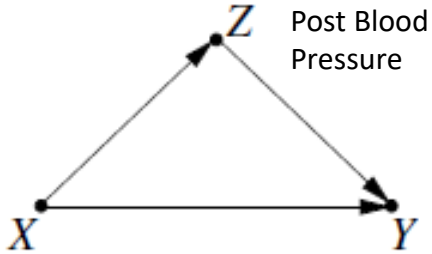
- Doctors choose between 2 treatments for kidney stones
  - Treatment A → large/severe stones
  - Treatment B → small stones
  - What might be reasons for different treatments?
  - Aggregate or segregated data?
  - What does DAG look like?

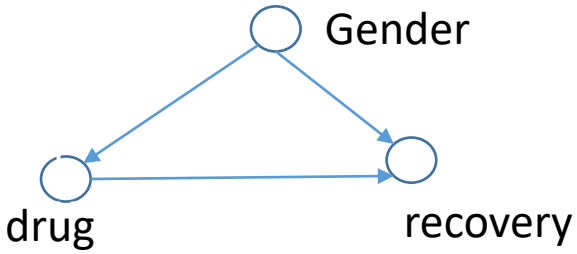
- 2 doctors perform 100 surgeries each
  - Some very difficult and some very easy surgeries
  - Doctor 1 performs easy surgeries far more often
  - Doctor 2 performs difficult surgeries far more often
  - What might be going on?
  - Aggregate or segregated data?
  - What does DAG look like?



The Simpson Paradox

- The fact that treatment affect BP and not the opposite was not in the data. Indeed in Statistics it is often stressed that “correlation is not causation”, so there is no statistical method that can determine the causal story from the data alone. Therefore, there is no statistical method that can aid in the decision.





- We can make causal assumptions because we know that drug cannot affect gender. “treatment does not cause sex” cannot be expressed in the data.
- So, what do we do? How can we make causal assumptions and make causal inferences?

The Simpson Paradox SCM (Structural Causal Model)

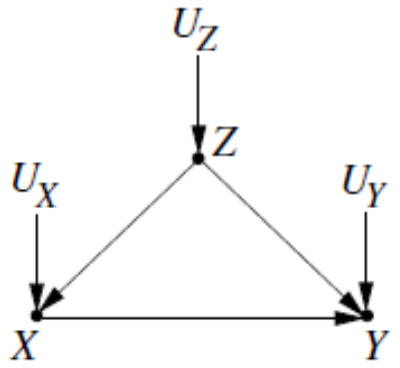


Figure 3.3: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

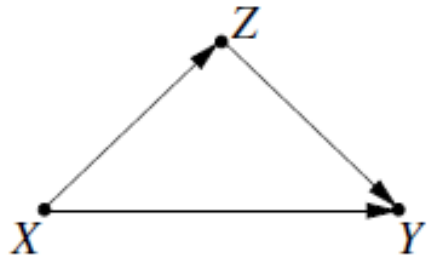


Figure 3.5: A graphical model representing the effects of a new drug, with X representing drug usage, Y representing recovery, and Z representing blood pressure (measured at the end of the study). Exogenous variables are not shown in the graph, implying that they are mutually independent

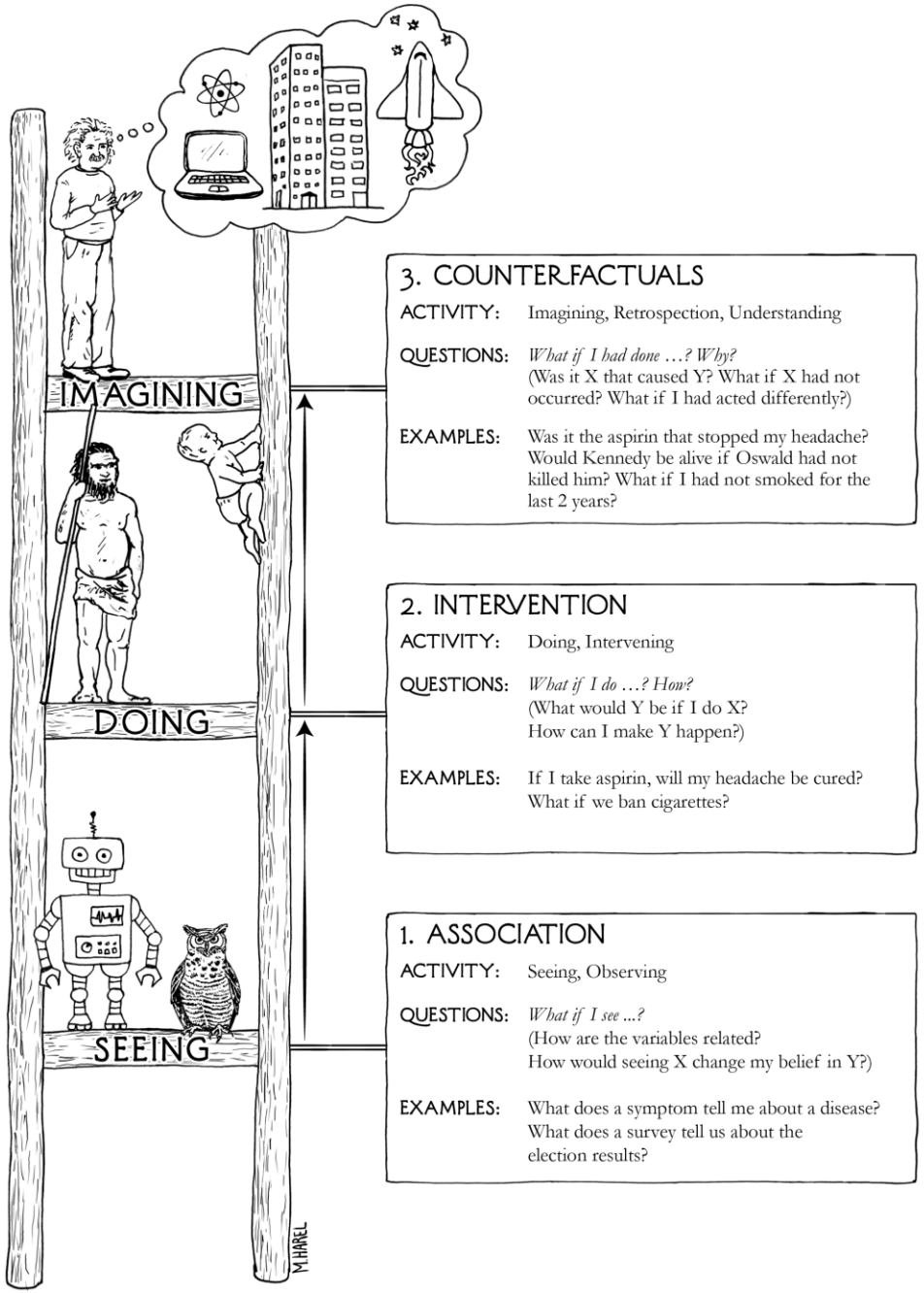
For Causal Inference We Need:

1. A working definition of “causation”
2. A method by which to formally articulate causal assumptions—that is, to create causal models
3. A method by which to link the structure of a causal model to features of data
4. A method by which to draw conclusions from the combination of causal assumptions embedded in a model and data.

Outline

- Simpson Paradox
- The causal Hierarchy
- Structural Causal Models
- Product form of Graphical models.

Ladder of Causation`



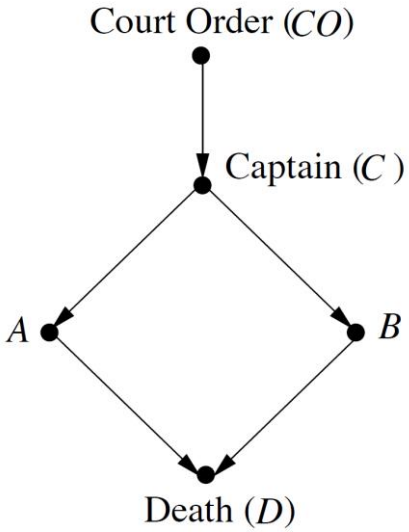
seeing, doing, and imagining.

- Most animals, learning machines are on the first rung, learning from association.
- Tool users, such as early humans, are on the second rung, if they act by planning and not merely by imitation. We can also use experiments to learn the effects of interventions, and presumably this is how babies acquire much of their causal knowledge.
- On the top rung, counterfactual learners can imagine worlds that do not exist and infer reasons for observed phenomena.

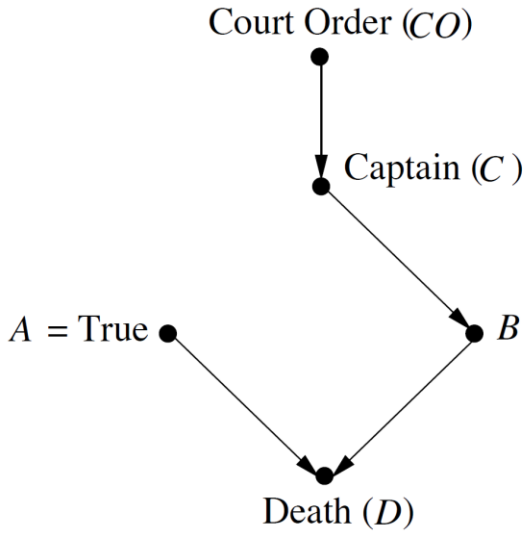
Darwiche 2017: "Human-Level Intelligence or Animal-Like Abilities?"

The Firing Squad

The story: Suppose that a prisoner is about to be executed. First, the court has to order the execution. The order goes to a captain, who signals the soldiers on the firing squad (A and B) to fire.



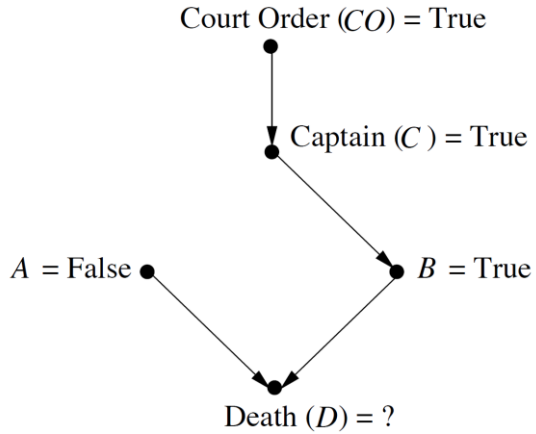
(CO, C, A, B, D) is a true/false variable

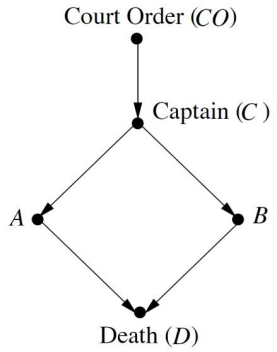


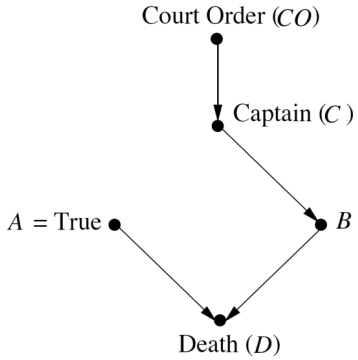
Ladder 1: If the prisoner is dead,
does that mean the court order was given?

Yes. Logic
Alternatively, suppose we find out that A fired.
What does that tell us about B? Yes.

Ladder 2: we can ask questions of intervention.
What if soldier A decides on his own initiative to fire, without waiting for the captain's command? Will the prisoner be dead or alive?







The Firing Squad, Counterfactuals

- **Ladder 3:** Suppose the prisoner is lying dead on the ground. Using level one implies that A shot, B shot, the captain gave the signal, and the court gave the order.
- If, contrary to fact, A had decided not to shoot, would the prisoner be alive?
- This question requires us to compare the real world with a fictitious and contradictory world where A didn't shoot.
- In the fictitious world, the arrow leading into A is erased and A is set to False, but the past history of A stays the same as it was in the real world.

In the firing squad example we ruled out uncertainties: maybe the captain gave his order a split second after rifleman A decided to shoot, maybe rifleman B's gun jammed, etc. To handle uncertainty we need information on how likely the alternatives are to occur.

Big picture - Pearl's Causal Hierarchy (PCH)

Level (Symbol)		Typical Activity	Typical Question	Examples
1 	Association $P(y \mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?

Big picture - Pearl's Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Question	Examples
1  Association $P(y \mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2  Intervention $P(y \mid do(x), c)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured?

Big picture - Pearl's Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Question	Examples
1  Association $P(y \mid x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2  Intervention $P(y \mid do(x), c)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured?
3  Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

Big picture - Pearl's Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Question	Examples
1  Association $P(y \mid x)$	Seeing ML - (Un)Supervised Deep Net, Bayes net, Hierarchical Model, DT	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
2  Intervention $P(y \mid do(x), c)$	Doing ML - Reinforcement Causal Bayes Net, MDP, POMDP, Planning	What if? What if I do X?	What if I take aspirin, will my headache be cured?
3  Counterfactual $P(y_x \mid x', y')$	Imagining, Retrospection Structural Causal Model	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

Outline

- Simpson Paradox
- The causal Hierarchy
- Structural Causal Models
- Product form of Graphical models.

Structural Causal Models (SCM), M

A structural causal model describes how nature assigns values to variables of interest.

- Two sets of variables, U and V and a set of functions **M: (U,V,F)**
- Each function assigns value to a variable in V based on the values of the other variables.
- We say that Variable X is a direct cause of Y if it appears in the function of Y.
- U are exogenous variables (external to the model. We do not explain how they are caused)..
- Variables in U have no parents.

SCM 1.5.1 (Salary Based on Education and Experience)

$$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$$
$$f_Z : Z = 2X + 3Y$$

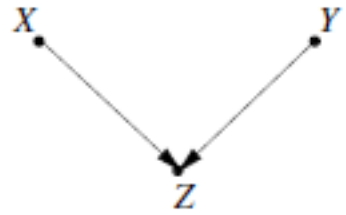
Z- salary, X – years in school, Y – years in the profession

Structural Causal Models (SCM), M

- Every SCM is associated with a graphical causal diagram.
- The graphical model G for a SCM M contains one node for each variable in M. If, in M, the function f_X for a variable X contains variable Y (i.e., if X depends on Y for its value), then, in G, there will be a directed edge from Y to X.
- We will deal primarily with SCMs that are acyclic graphs (DAGs).
- A graphical definition of causation: If, in a graphical model, a variable X is the child of another variable Y then Y is a direct cause of X; if X is a descendant of Y , then Y is a potential cause of X .

SCM 1.5.1 (Salary Based on Education and Experience)

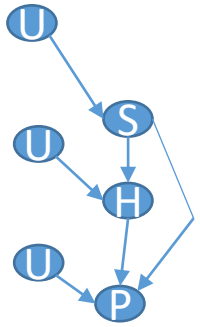
$$U = \{X, Y\}, \quad V = \{Z\}, \quad F = \{f_Z\}$$
$$f_Z : Z = 2X + 3Y$$



X, should be part of V?

X and Y are direct causes for Z

Structural Causal Models (SCM)



SCM 1.5.2 (Basketball Performance Based on Height and Sex)

$$
\begin{aligned}
V = \quad & \{\text{Height, Sex, Performance}\}, \quad U = \{U_1, U_2, U_3\}, \quad F = \{f1, f2\} \\
& \text{Sex} = U_1 \\
& \text{Height} = f_1(\text{Sex}, U_2) \\
& \text{Performance} = f_2(\text{Height, Sex}, U_3)
\end{aligned}
$$

U are unmeasured terms that we do not care to name. Random causes we do not care about.
U are sometime called error terms.

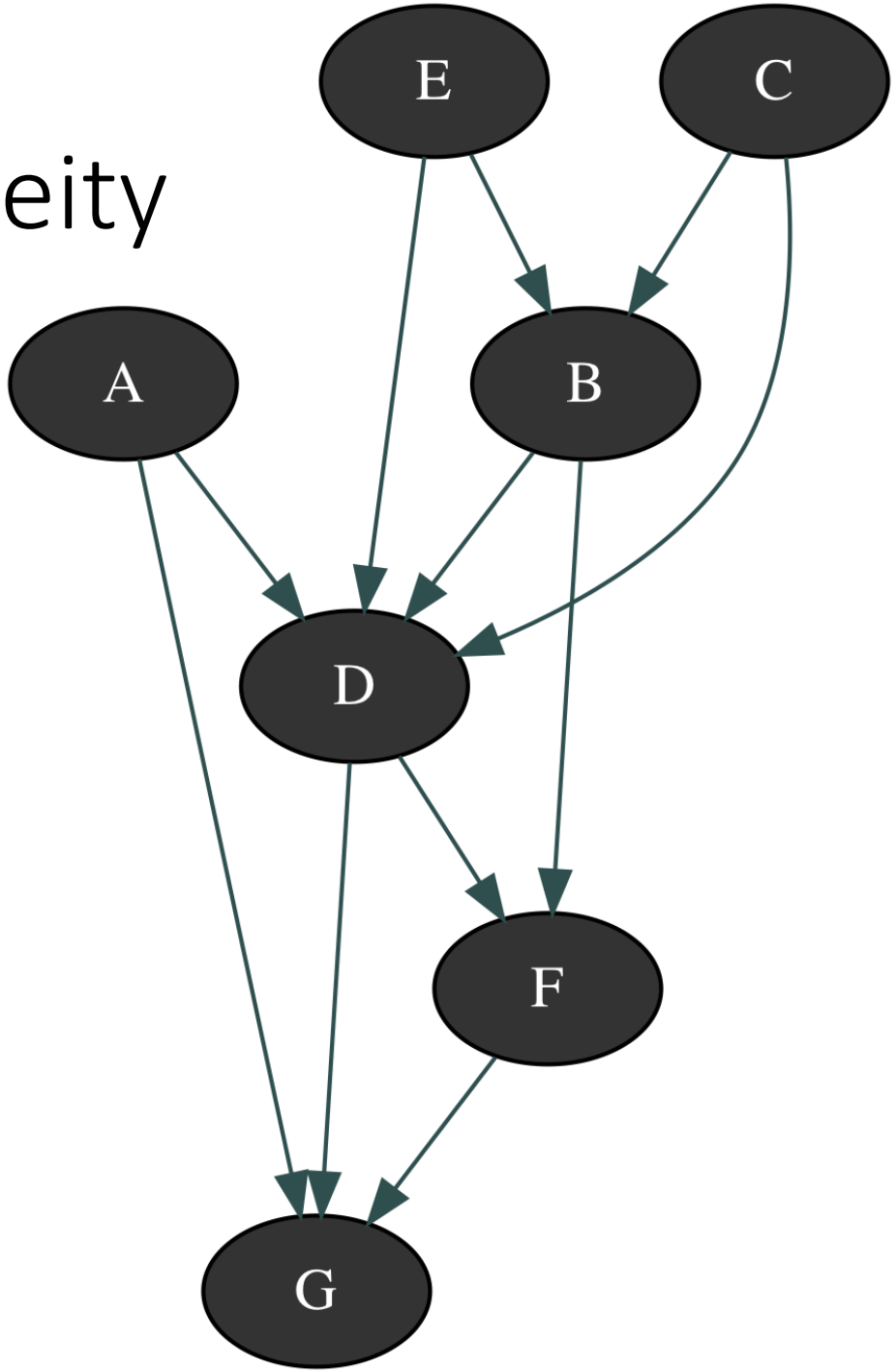
The graphical causal model provides lots of information about what is going on: X causes Y and Y causes Z

Structural Causal Model --Summary

- SCM (U,V,F)
  - Exogenous variables, U
    - External to model
      - No explanation for their cause
      - No parents
    - Root nodes
    - AKA error terms or omitted factors
  - Endogenous variables, V
    - Descendants of exogenous variable(s)
  - Functions for endogenous variables (set F)

More on Exogeneity/Endogeneity

- Exogenous variables?
  - {A, E, C}
- Endogenous variables?
  - {B, D, F, G}
- Functions and their inputs?
  - $f_A(?)$
    - No functions for exogenous variables
  - $f_B(C, E)$
  - $f_D(A, B, C, E)$
  - $f_F(B, D)$
  - $f_G(A, D, F)$



Outline

- Simpson Paradox
- The causal Hierarchy
- Structural Causal Models
- Product form of Graphical models.

Probability distibutions of SCM; Product Decomposition in Bayesian Networks

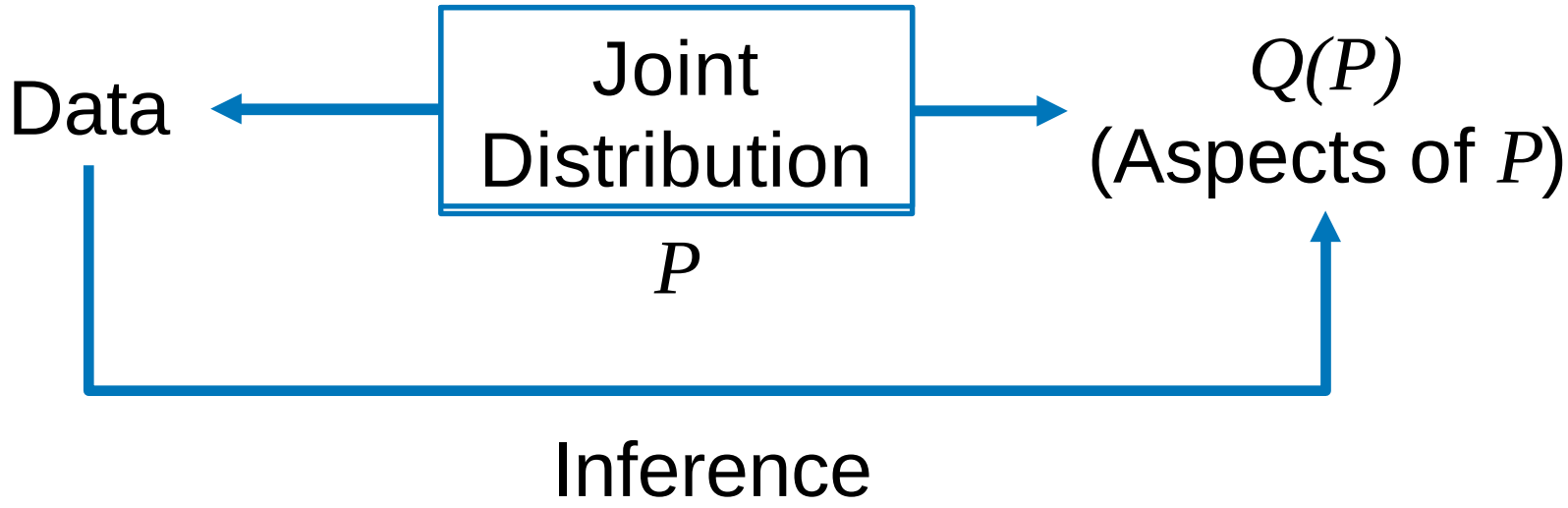
- How to calculate joint probability from SCM
- $P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i \mid pa_i)$
  - pa_i are the parents of x_i
- $X \rightarrow Y \leftarrow Z$, how to calculate $P(x, y, z)$?
  - $P(x, y, z) = P(X = x) \cdot P(Y = y \mid X = x, Z = z) \cdot P(Z = z)$
- Why is this important?
- $Y \rightarrow X \rightarrow Z$, X = eye color, Y = age (between 7 and 14), Z = hair color
  - How many rows of data do we need for joint probability?
    - $8 \cdot 7 \cdot 6$
  - How many rows of data do we need to be able to *calculate* joint probability?
    - $8 + 8 \cdot 7 + 6 \cdot 7$

Outline

- Structural Causal Models
- Product form of Markov SCM
- d-seperation
- Bayesian networks

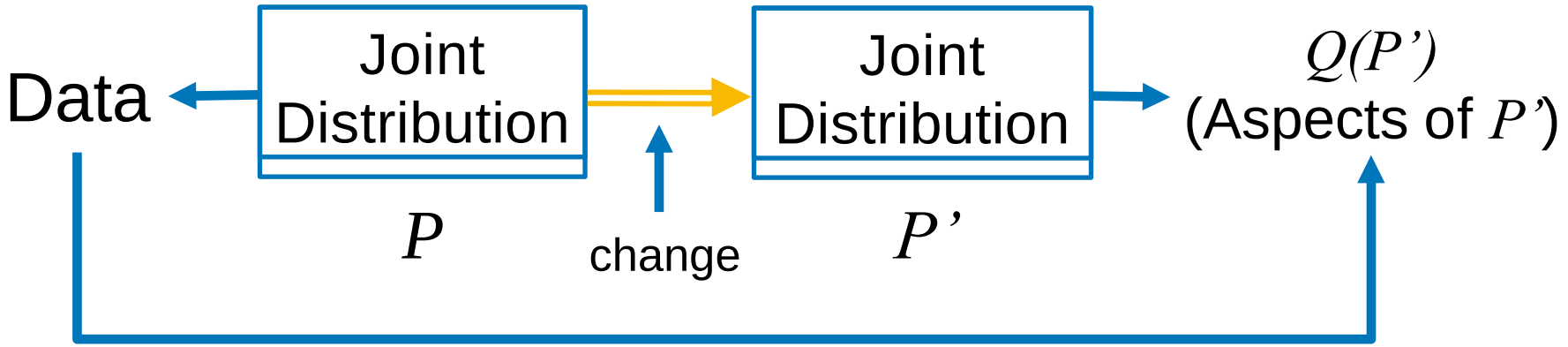
Traditional Stats-ML Inferential Paradigm

- **Approach:** Find a good representation for the data.



e.g., Infer whether customers who bought product A would also buy product B — or, compute $Q = P(B \mid A)$.

From Statistical to Causal Analysis

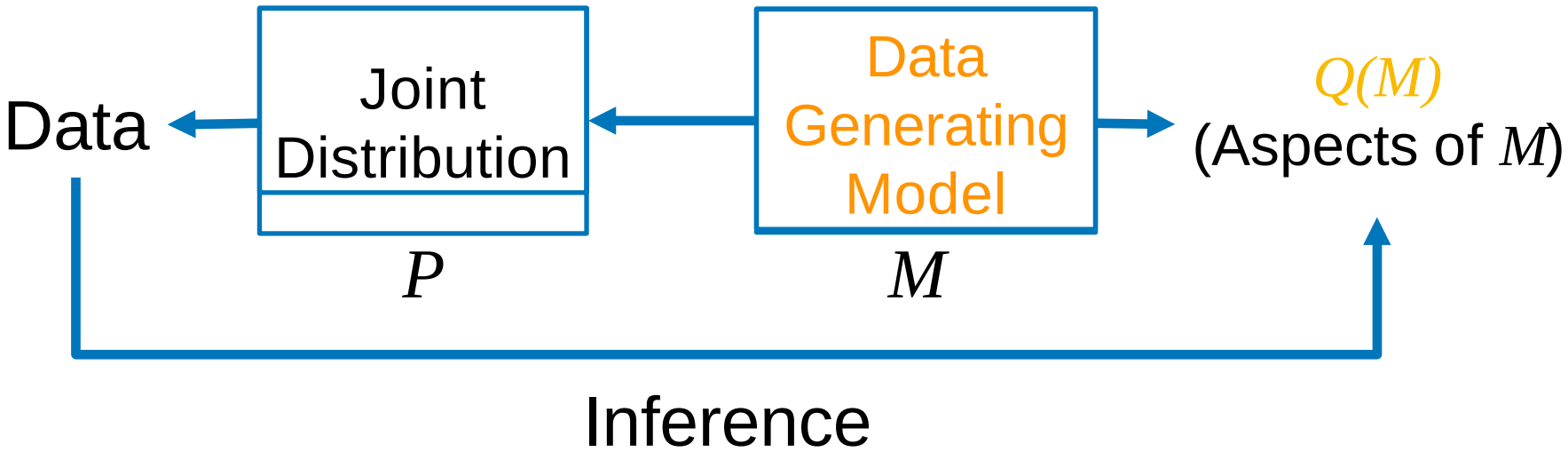


e.g., Estimate $P'(sales)$ if we double the price
Estimate $P'(cancer)$ if we ban smoking

Q: How does P (factual) changes to $P'(hypothetical)$?

Needed: New formalism to represent both P & P'.
P is tied to the data; P' is never observed, no data.

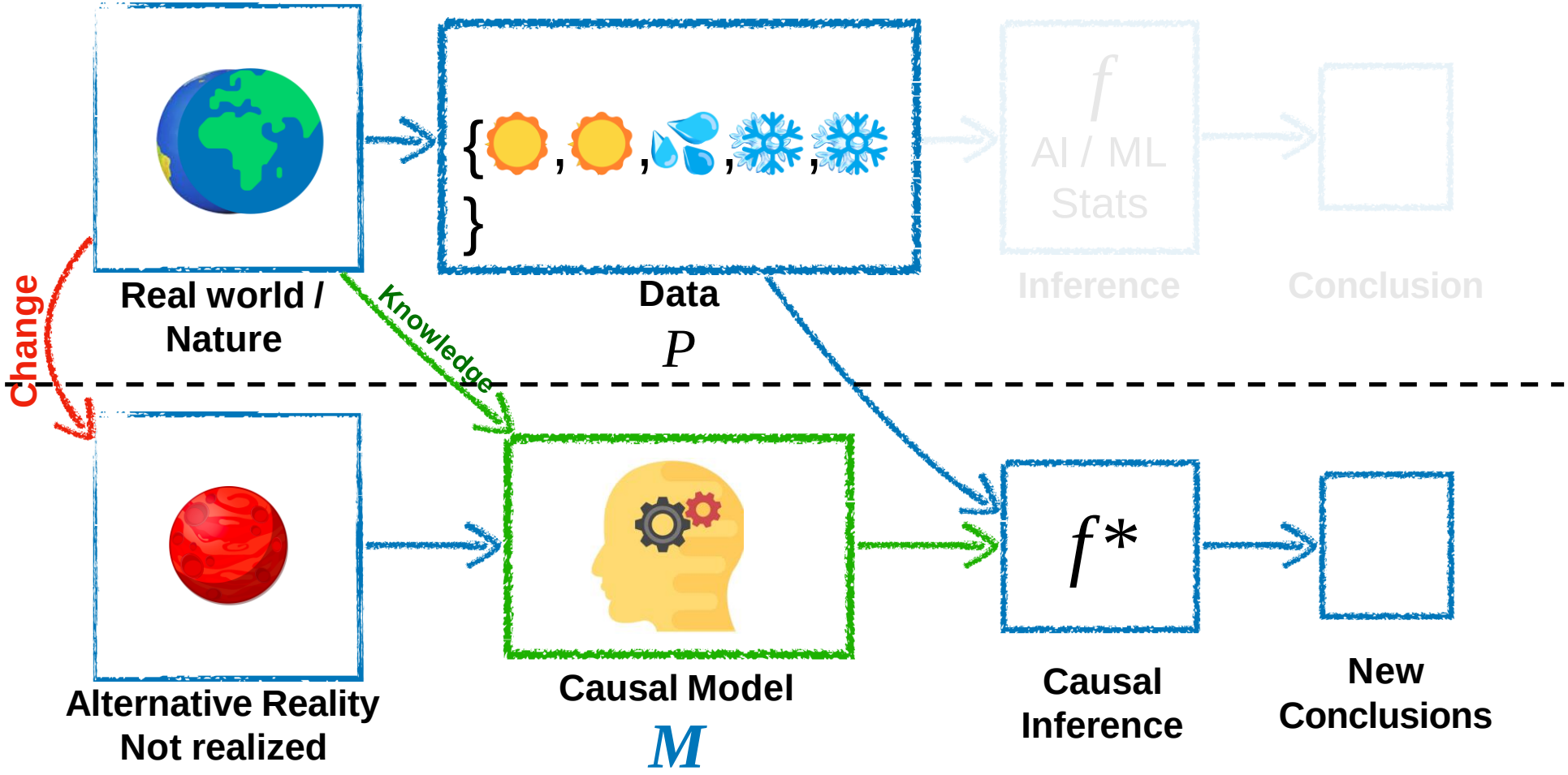
New Oracle -
The Structural Causal Model Paradigm



M – Invariant strategy (mechanism, recipe, law, protocol) by which Nature assigns values to variables in the analysis.

P - model of data, M - model of reality

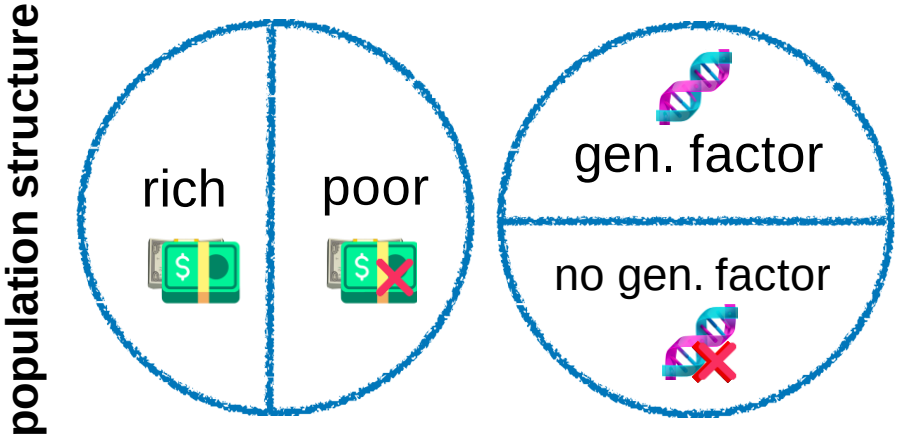
Back to the Big Picture



Modeling Reality with SCM

- The population of a certain city is falling ill from a contagious disease. There is a drug believed to help patients survive the infection.
- Unknown to the physicians, folks with good living conditions (rich) will always survive.

- While some people have a gene that naturally fights the disease and don't require treatment, they will develop an allergic reaction if treated, which is fatal under poor living conditions.



Being rich and having the genetic factor are independent events.

Reality (unknown to physicians):

rich = alive anyways
$poor_1$ = die anyways (no gene)
$poor_2$ = die iff take the drug (gene)

$\prod = rich \cup poor_1 \cup poor_2$

$P(rich) = P(poor)$
$P(poor_1) = P(poor_2)$

Modeling Reality in our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $=0$ for poor)

D ($D=1$ for taking the drug)

A ($A=1$ if person ends up alive)

Modeling Reality in Our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $=0$ for poor)
D ($D=1$ for taking the drug)
A ($A=1$ if person ends up alive)

Variables that are unobserved ($\mathbf{U}$):

U_g ($U_g=1$ has genetic factor, $=0$ o/w)
U_r (Other factors affecting Wealth)

Modeling Reality in Our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $=0$ for poor)
D ($D=1$ for taking the drug)
A ($A=1$ if person ends up alive)

How are the observed variables determined?

$R \leftarrow U_r$
$D \leftarrow R$
$A \leftarrow R \vee (U_g \wedge \neg D)$

Variables that are unobserved ($\mathbf{U}$):

U_g ($U_g=1$ has genetic factor, $=0$ o/w)
U_r (Other factors affecting Wealth)

Modeling Reality in our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $=0$ for poor)

D ($D=1$ for taking the drug)

A ($A=1$ if person ends up alive)

How are the observed variables determined?

$$R \leftarrow U_r$$
$$D \leftarrow R$$
$$A \leftarrow R \vee (U_g \wedge \neg D)$$

Variables that are unobserved ($\mathbf{U}$):

U_g ($U_g=1$ has genetic factor, $=0$ o/w)

U_r (Other factors affecting Wealth)

- Rich is always alive.
- Poor will survive only if they have the gene and don't take the drug.

Modeling Reality in our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $=0$ for poor)
D ($D=1$ for taking the drug)
A ($A=1$ if person ends up alive)

- How are the observed variables determined?

- $R \leftarrow U_r$
  $D \leftarrow R$
- $A \leftarrow R \vee (U_g \wedge \neg D)$

Variables that are unobserved ($\mathbf{U}$):

U_g ($U_g=1$ has genetic factor, $=0$ o/w)
U_r (Other factors affecting Wealth)

- What is the randomness over the unobserved vars:

- $P(U_g=1)=1/2, P(U_r=1)=1/2$

Modeling Reality in our Example

Variables we observe ($\mathbf{V}$):

R ($R=1$ for rich, $R=0$ for poor)

D ($D=1$ for taking the drug,)

A ($A=1$ if person en)

Variables that are unobserved ($\mathbf{U}$):

U_g ($U_g=1$ has genetic factor, 0 o/w)

U_r (Other factors affecting Wealth)

How are the observed variables determined?

$R = U_r$

$D = R$

$A = R \cdot (G \vee \neg D)$

What is the randomness over the hidden variab

$P(G=1)=1/2, P(U_r=1)=$

This is a fully specified Model of Reality!

It implies both P and P' (more soon).

This will be our new, almighty Oracle,
which is known as **Structural Causal Model**.

(Now, let's generalize this object…)

Outline

- Structural Causal Models
- Product form of Markov SCM
- d-seperation
- Bayesian networks

The New Oracle: Structural Causal Models

Definition: A structural causal model (SCM) M is a 4-tuple $<\boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, P(\boldsymbol{u})>$, where

- $\boldsymbol{V} = \{V_1,...,V_n\}$ are endogenous variables;
- $\boldsymbol{U} = \{U_1,...,U_m\}$ are exogenous variables;
- $\mathcal{F} = \{f_1,..., f_n\}$ are functions determining $\boldsymbol{V}$, $v_i \leftarrow f_i(pa_i, u_i), Pa_i \subset V_i, U_i \subset U;$
- $P(\boldsymbol{u})$ is a distribution over $\boldsymbol{U}$

Not regression!!

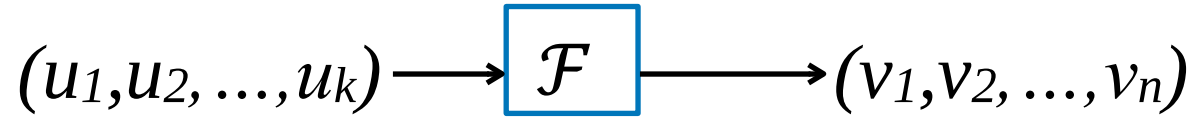
e.g. $y = \alpha + \beta X + U_Y$

Axiomatic Characterization:

(Galles-Pearl, 1998; Halpern, 1998).

1. SCM induces distribution $P(\boldsymbol{v})$

- $\mathcal{F}$ can be seen as a mapping from $\boldsymbol{U} \longrightarrow \boldsymbol{V}$



- When the input $\boldsymbol{U}$ is a set of random vars, then the output $\boldsymbol{V}$ also becomes a set of r.v's.
- P(v) is the layer 1 of the PCH, known as the observational (or passive) prob. distribution.
- Each event, person, observation, etc… corresponds to an instantiation of $\boldsymbol{U}=\boldsymbol{u}$.

1. SCM induces distribution $P(\mathbf{v})$

$$R \leftarrow U_r$$
$$D \leftarrow R$$
$$A \leftarrow R \vee (U_g \wedge \neg D)$$

Example: (Drug, Rich, Alive)

- Each citizen follows in one of four groups according to the unobservables in the model:

$$\mathcal{F} = \begin{cases} f_R : U_r \\ f_D : R \\ f_A : R \vee (U_g \wedge \neg D) \end{cases}$$

$$\mathcal{F}$$

$$(U_r=1, U_g=1) \longrightarrow (R=1, D=1, A=1)$$
$$(U_r=1, U_g=0) \longrightarrow (R=1, D=1, A=1)$$
$$(U_r=0, U_g=1) \longrightarrow (R=0, D=0, A=1)$$
$$(U_r=0, U_g=0) \longrightarrow (R=0, D=0, A=0)$$

1. SCM induces distribution $P(\boldsymbol{v})$

In our example:

- Events in the $\boldsymbol{U}$-space translate into events in the space of $\boldsymbol{V}$.

$$\mathcal{F} = \begin{cases} f_R : U_r \\ f_D : R \\ f_A : R \vee (U_g \wedge \neg D) \end{cases}$$

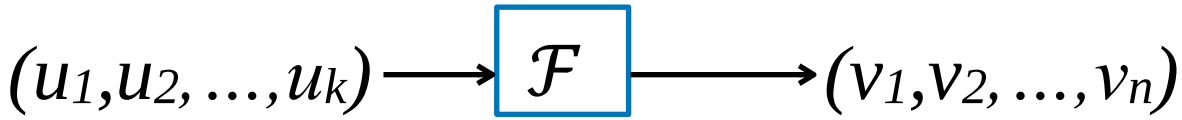
$P(\boldsymbol{u})$		$P(\boldsymbol{v})$
1/4	$(U_r=1, U_g=1) \longrightarrow (R=1, D=1, A=1)$	1/2
1/4	$(U_r=1, U_g=0) \longrightarrow (R=1, D=1, A=1)$	
1/4	$(U_r=0, U_g=1) \longrightarrow (R=0, D=0, A=1)$	1/4
1/4	$(U_r=0, U_g=0) \longrightarrow (R=0, D=0, A=0)$	1/4

1. SCM induces distribution $P(\boldsymbol{v})$

- [Def. 2, PCH chapter] An SCM $M = <\boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, P(\boldsymbol{u})>$ defines a joint probability distribution $P^M(V)$ *s.t. for each* $Y \subseteq V$:

$$P^M(y) = \sum_{u|Y(u)=y} P(u)$$

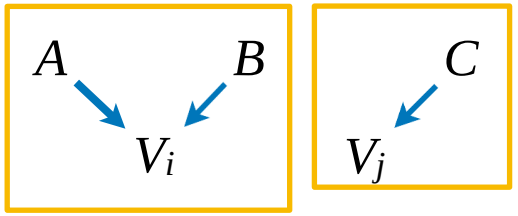
- $\mathcal{F}$ can be seen as a mapping from $\boldsymbol{U} \longrightarrow \boldsymbol{V}$



2. SCM ⟶ Causal Diagram

- Every SCM M induces a causal diagram
- Represented as a DAG where:
  - Each $V_i \in \boldsymbol{V}$ is a node,
  - There is $W \longrightarrow V_i$ if for $W \in Pa_i$,

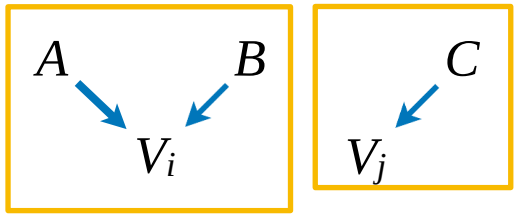
$$V_i \leftarrow f_i(A,B,U)$$
$$V_j \leftarrow f_j(C,U)$$

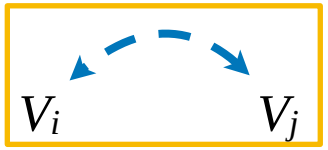


2. SCM ⟶ Causal Diagram

- Every SCM M induces a causal diagram
- Represented as a DAG where:
  - Each $V_i \in \boldsymbol{V}$ is a node,
  - There is $W \longrightarrow V_i$ if for $W \in Pa_i$,
  - There is $V_i \dashleftarrow\dashrightarrow V_j$ whenever $U_i \cap U_j \neq \emptyset$.

$$V_i \leftarrow f_i(A,B,U)$$
$$V_j \leftarrow f_j(C,U)$$

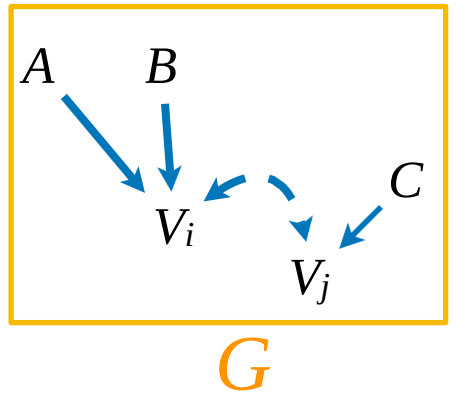




2. SCM ⟶ Causal Diagram

- Every SCM M induces a causal diagram
- Represented as a DAG where:
  - Each $V_i \in \boldsymbol{V}$ is a node,
  - There is $W \longrightarrow V_i$ if for $W \in Pa_i$,
  - There is $V_i \dashleftarrow\dashrightarrow V_j$ whenever $U_i \bigcap U_j \neq \varnothing$.

$$V_i \leftarrow f_i(A,B,U)$$
$$V_j \leftarrow f_j(C,U)$$



Causal Diagram — Definition (again)

- ***Causal Diagram*** [Def. 13, PCH chapter] — Consider an SCM $M = \langle \boldsymbol{V}, \boldsymbol{U}, \mathcal{F}, P(\boldsymbol{u}) \rangle$. Then G is said to be a causal diagram (of M) if constructed as follows:

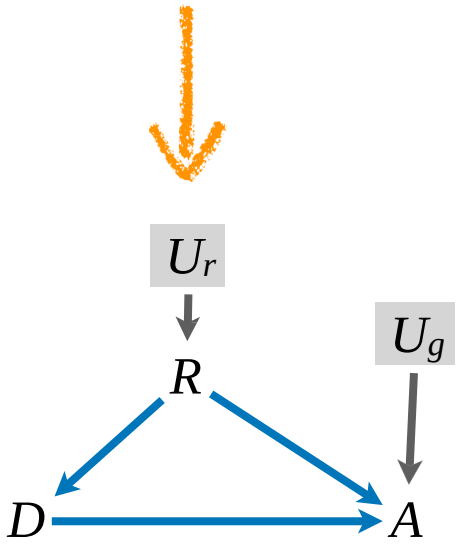
1. add vertex for every endogenous variable $\boldsymbol{V_i} \in \boldsymbol{V}$.
2. add edge ($\boldsymbol{V_j} \rightarrow \boldsymbol{V_i}$) for every $\boldsymbol{V_i}, \boldsymbol{V_j} \subset$V if $\boldsymbol{V_j}$ appears as argument of $\boldsymbol{f_i} \in \mathcal{F}$.
3. add a bidirected edge ($\boldsymbol{V_j} \dashleftarrow\!\dashrightarrow \boldsymbol{V_i}$) for every $\boldsymbol{V_i}, \boldsymbol{V_j} \subset$V if $\boldsymbol{U_i}, \boldsymbol{U_j} \subset \boldsymbol{U}$ are correlated or the corresponding functions $\boldsymbol{f_i}, \boldsymbol{f_j}$ share some $U \in \mathbf{U}$ as argument.

2. SCM ⟶ Causal Diagram

Recall our medical example:

- Endogenous (observed) variables $\boldsymbol{V}$:
  - R ($R=1$ for rich, $=0$ for poor)
  - D ($D=1$ for taking the drug, $D=0$ o/w)
  - A ($A=1$ if person ends up alive, $=0$ o/w)
- Exogenous (unobserved) Variables $\boldsymbol{U}$:
  - U_r (Wealthiness factors)
  - U_g ($=1$ has the genetic factor, $=0$ o/w)
- Distribution over $\boldsymbol{U}$: $P(U_r)=1/2, P(U_g)=1/2$

$$\mathcal{F} = \begin{cases} R \leftarrow U_r \\ D \leftarrow R \\ A \leftarrow R \vee (U_g \wedge \neg D) \end{cases}$$

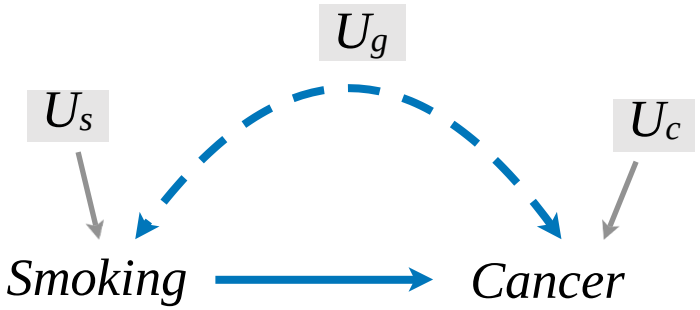


2. SCM ⟶ Causal Diagram

Another example:

- $\boldsymbol{V} = \{ Smoking, Cancer \}$
- $\boldsymbol{U} = \{ U_s, U_c, U_g \}$ unobserved factors
  - $\mathcal{F}$:

$$Smoking \leftarrow f_{Smoking}(U_s, U_g)$$
$$Cancer \leftarrow f_{Cancer}(Smoking, U_c, U_g)$$

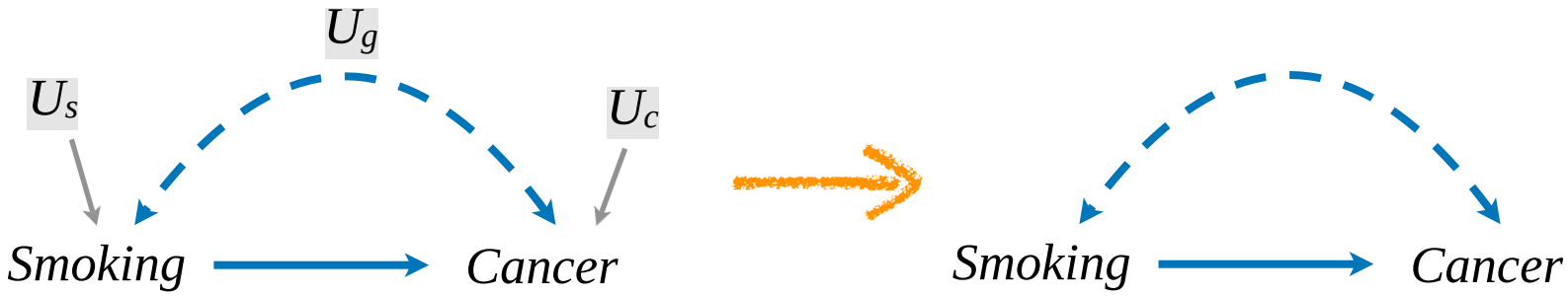


Remark 1. The mapping is just 1-way (i.e., from a SCM to a causal graph) since the graph itself is compatible with infinitely many SCMs with the same scope (the same functions signatures and compatible exogenous distributions).

Remark 2. This observation will be central to causal inference since, in most practical settings, researchers may know the scope of the functions, for example, but not the details about the underlying mechanisms.

Causal Diagrams

- Convention. The unobserved variables are left implicit in the graph.



Food for thought

Does the causal diagram give us any clues about the (in)dependence relations in the obs. distribution P(V)?

- Is T independent of W?
- Is W independent of T?
- Is Z independent of T?
- Is Z independent of X?
- Is Y independent of W?
- Is Y independent of W if we know the value of X?

Prop. 1

$$M = \begin{cases} Z \leftarrow f_Z(u_z) \\ X \leftarrow f_x(u_x) \\ W \leftarrow f_W(z, x, u_w) \\ Y \leftarrow f_Y(x, u_y) \\ T \leftarrow f_T(w, u_t) \end{cases}$$

Prop. 2