Causal Modeling

Causality is a different concept than probability.

If rooster is crowing, there is high probability of sun rising.

But rooster crowing is not causing sun to rise.

How can we deduce Causality?

Statistics 101: Correlation is not causation !

Color of Left Eye is correlated with **Color of Right Eye**.

Color of Left Eye is correlated with **Color of Right Eye**.

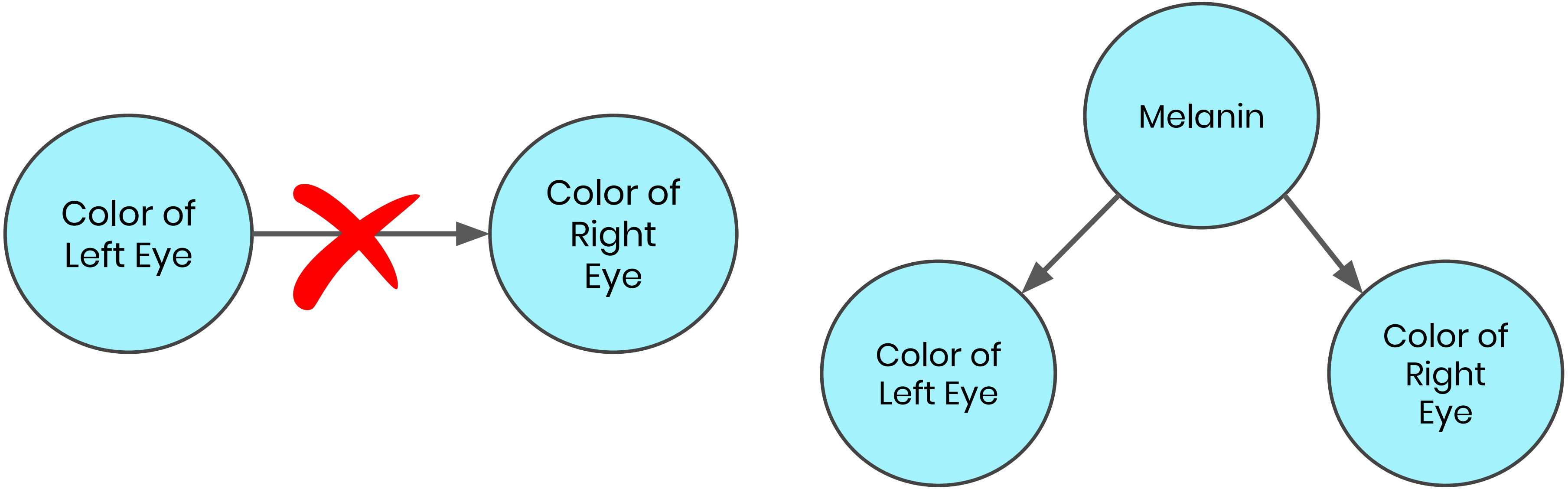
But is **Color of Left Eye** causing **Color of Right Eye**?

Color of Left Eye is correlated with **Color of Right Eye**.

But is **Color of Left Eye** causing **Color of Right Eye**?

If you change **Color of Left Eye**, does it change **Color of Right Eye**?

Statistics 101: Correlation is not causation !



Statistics 101: Correlation is not causation !

Statistics 101: Correlation is not causation !

BUT

Statistics 101: Correlation is not causation !

BUT

Causation and Correlation are correlated.

Statistics 101: Correlation is not causation !

BUT

Causation and Correlation are correlated.

Causation reveals itself in correlation.

Statistics 101: Correlation is not causation !

BUT

Causation and Correlation are correlated.

Causation reveals itself in correlation.

If A and B are correlated, A causes B or B causes A or they share a latent common cause.

Causal Inference

If there are only two variables , we have no way of telling which variable is cause and which is effect.

“Correlation is not causation”
This statement is true for single pairs of variables,

But when we have patterns of correlations, deducing causation is easier.

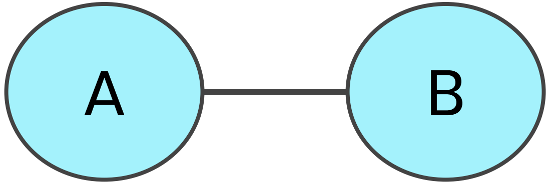
Causal Inference

Causal Inference uses *d*-separation : a criterion for deciding, from a given a causal graph, whether a set *X* of variables is independent of another set *Y*, given a third set *Z*. The idea is to associate "dependence" with "connectedness" (i.e., the existence of a connecting path) and "independence" with "unconnected-ness" or "separation".

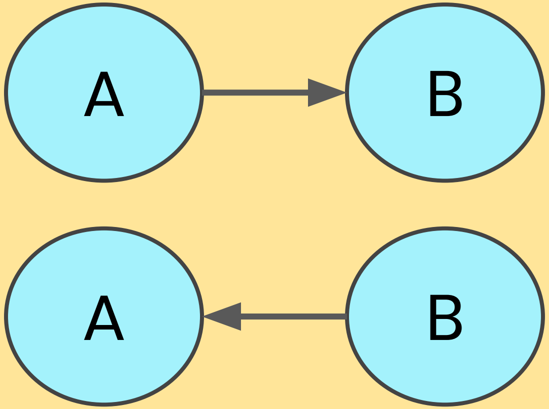
Causal Inference

Example 1: Two correlated variables

A and B are correlated (dependent) with each other.



We have no way of telling whether A causes B or B causes A.

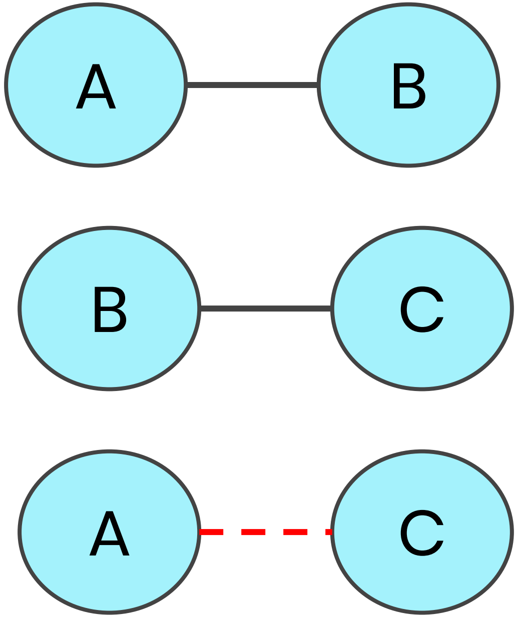


Example 2:

A and B are correlated

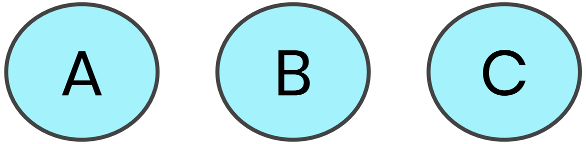
B and C are correlated

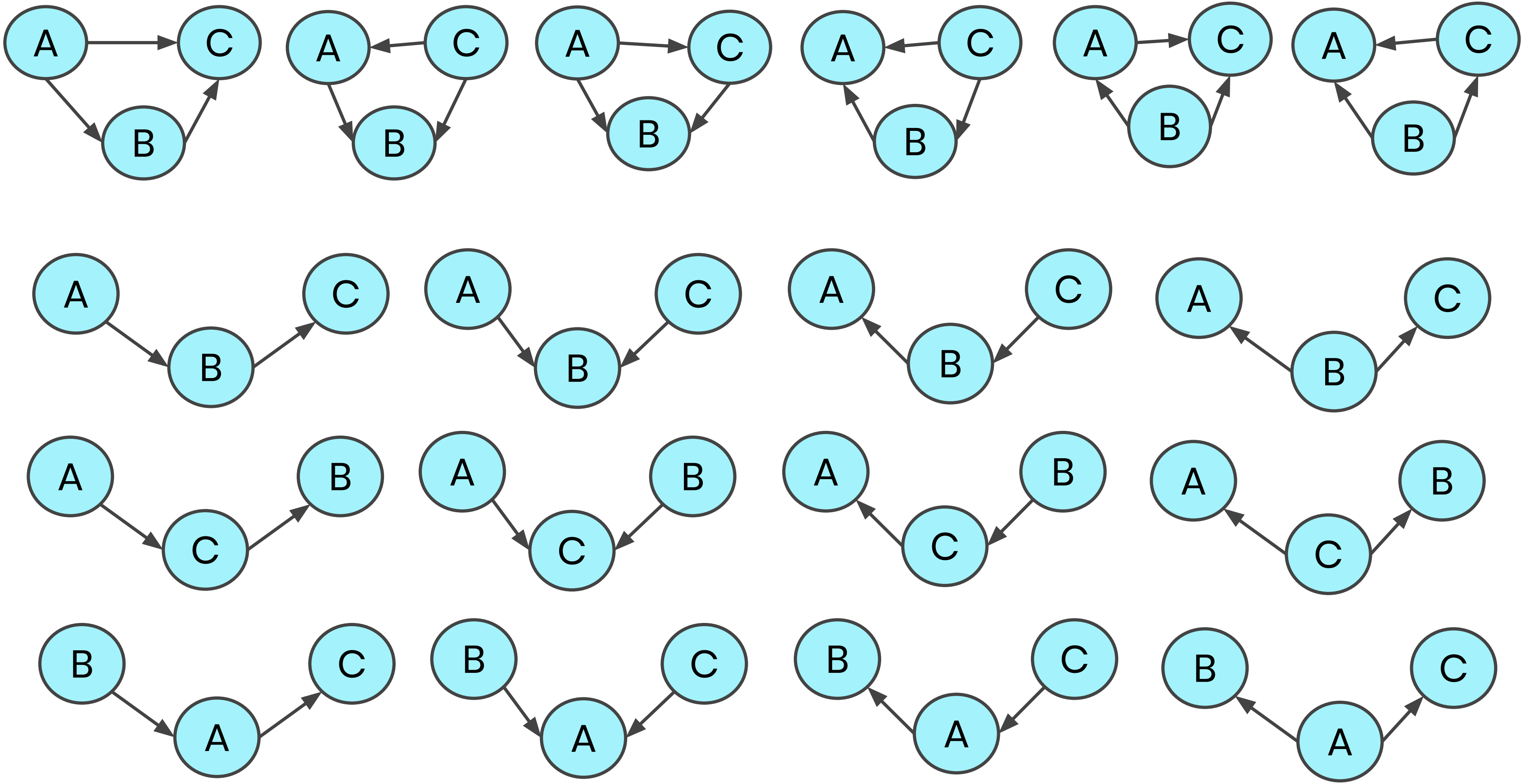
A and C are not correlated.



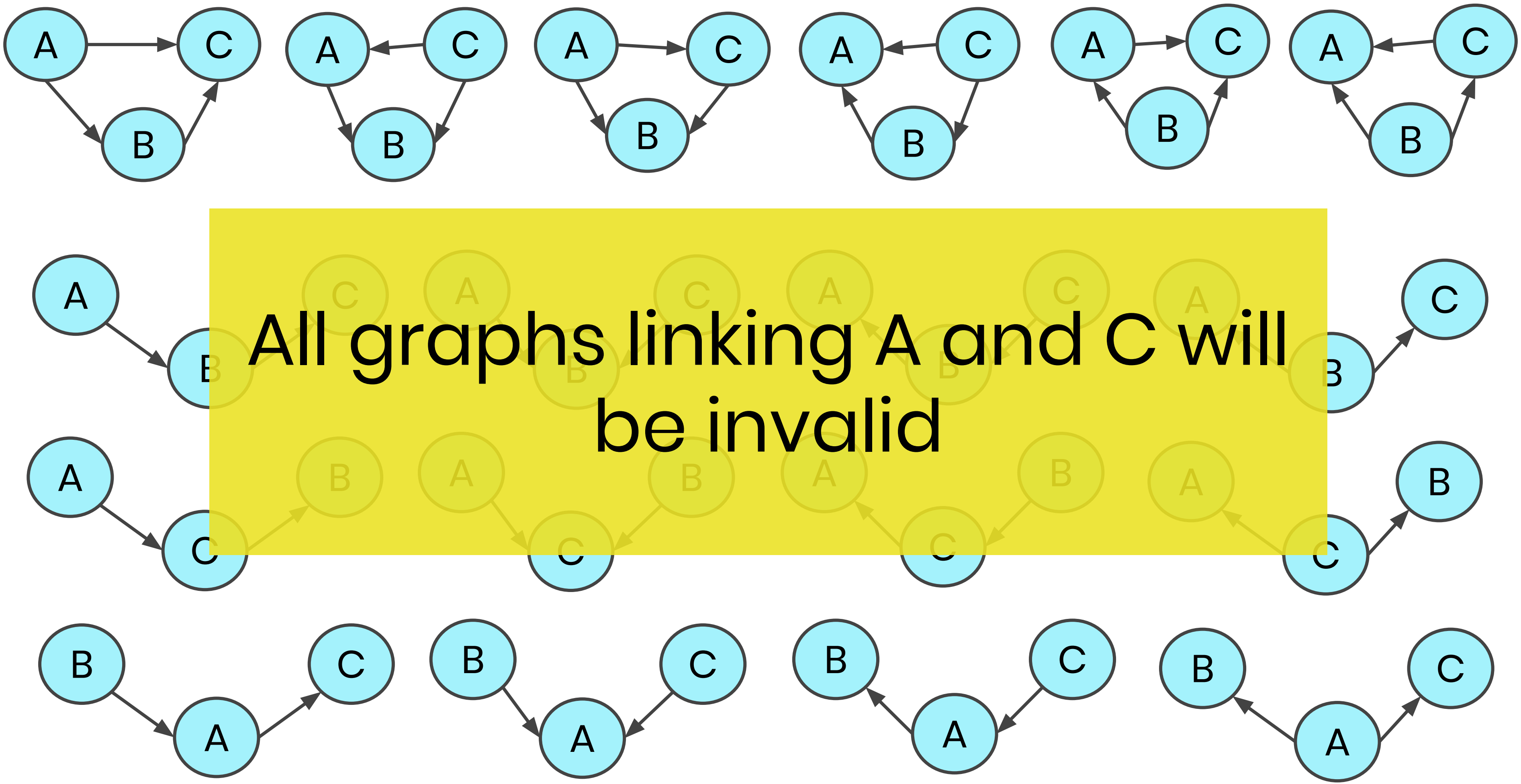
There are three variables and we don’t know yet how the causal graph would look like.

Let’s first create all possible Directed Acyclic Graph (DAG) from these three variables.

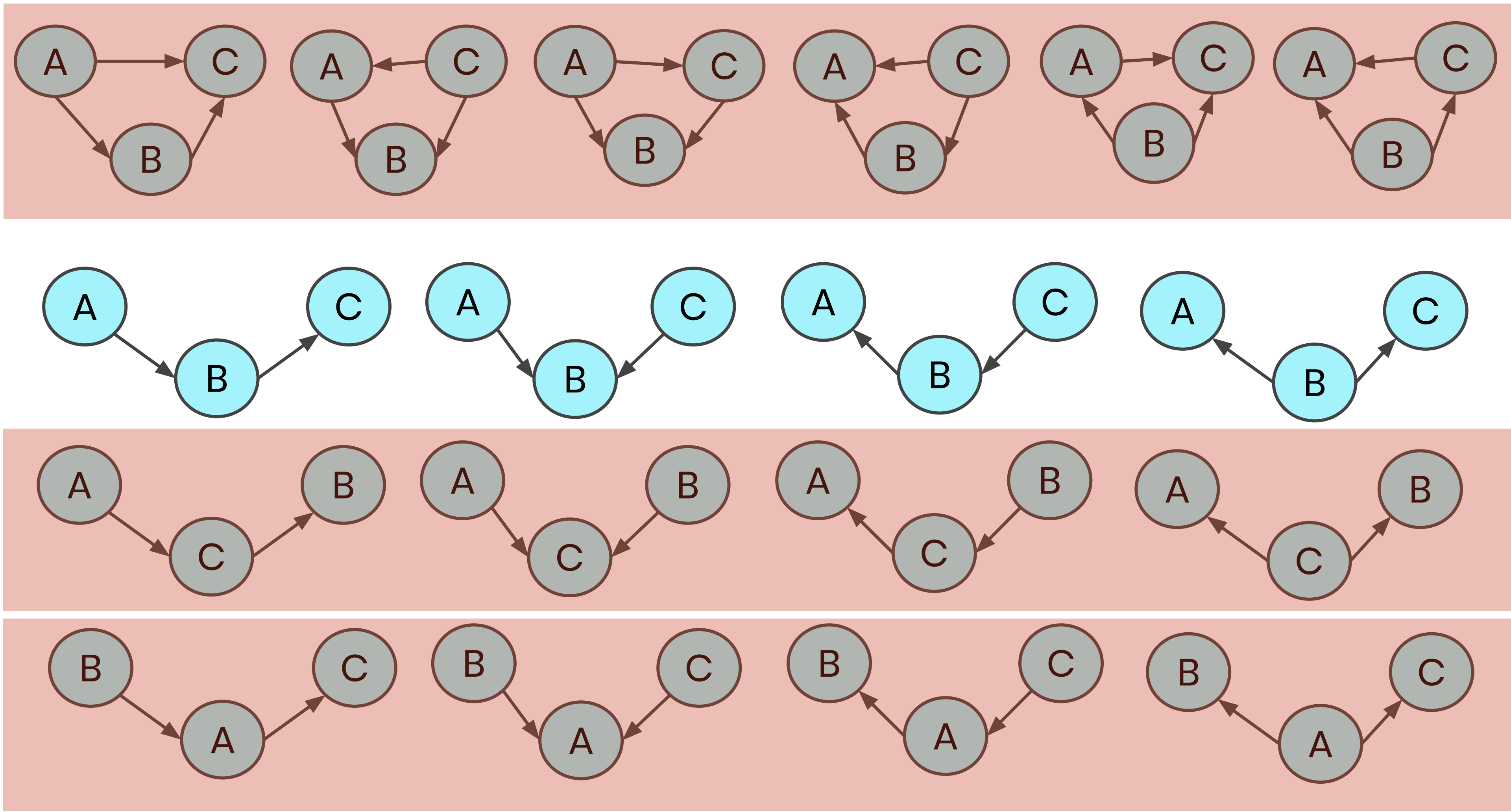




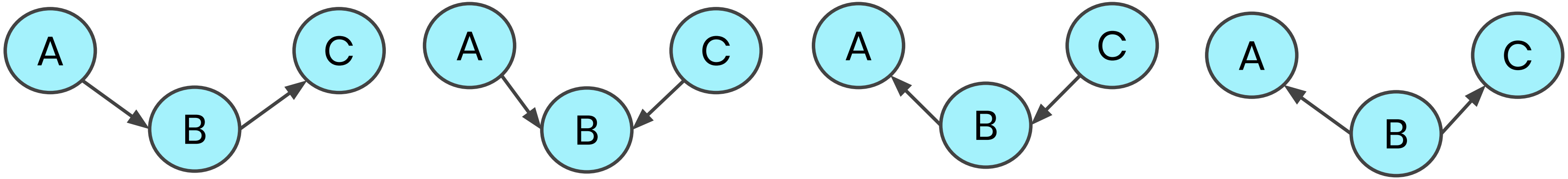
A C B
A C B
A C B
A C B
A C B
A C B
A B C
A B C
A B C
A B C
A C B
A C B
A C B
A C B
B A C
B A C
B A C
B A C



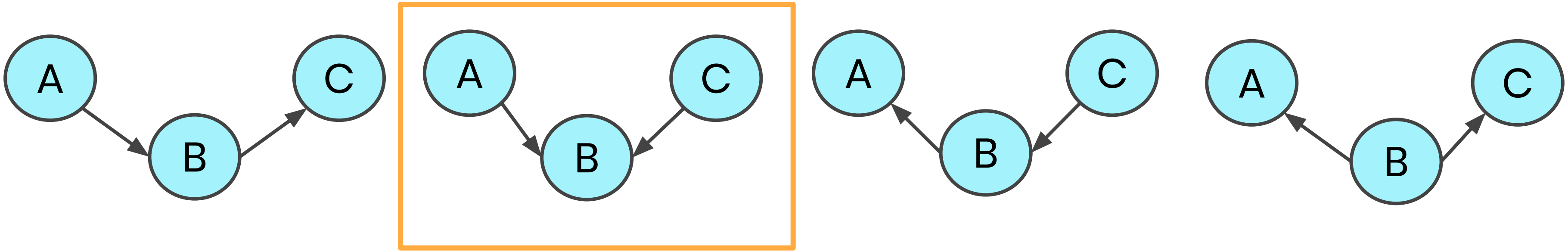
All graphs linking A and C will be invalid



There are four possible graphs



Using d-separation criterion, we select the second graph



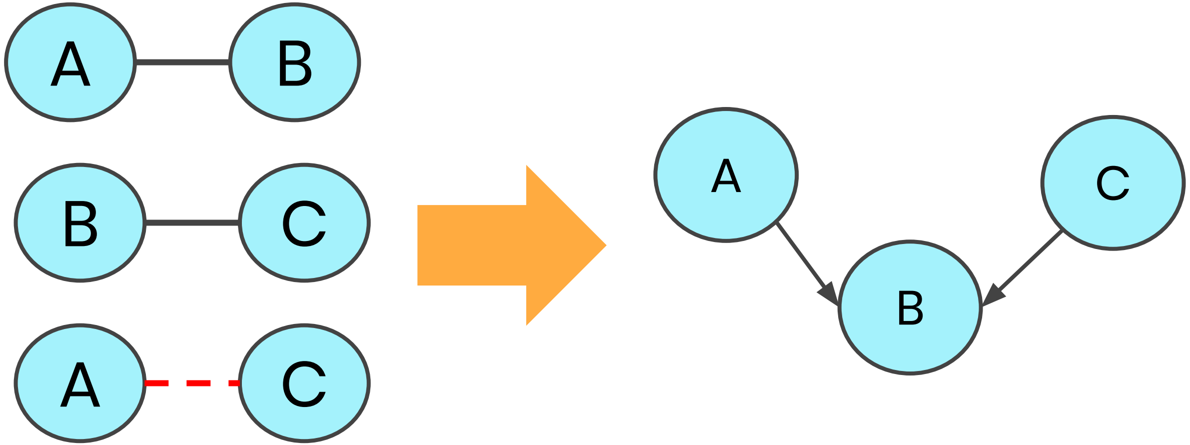
Causal Inference

Example 2:

A and B are correlated

B and C are correlated

A and C are not correlated.



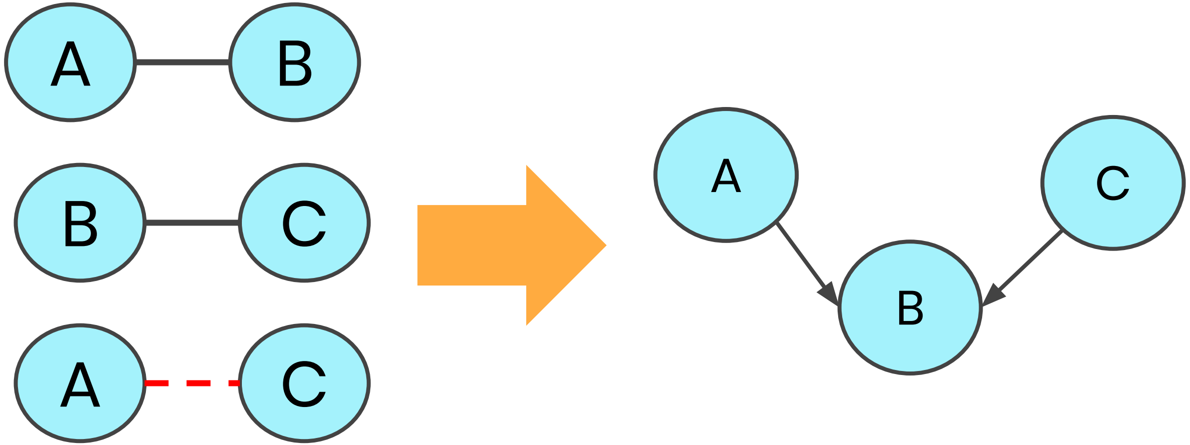
Causal Inference

Example 2:

A and B are correlated

B and C are correlated

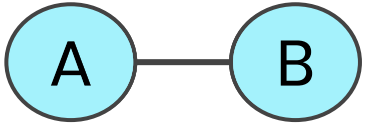
A and C are not correlated.



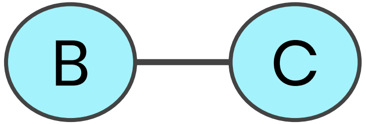
But, can we always get this clear inference d-separation ?

Example 3:

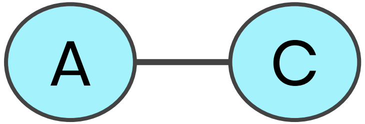
A and B are correlated



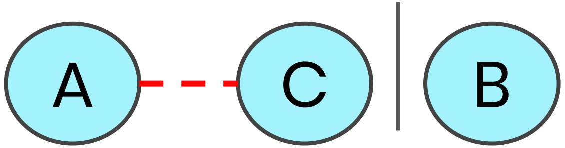
B and C are correlated



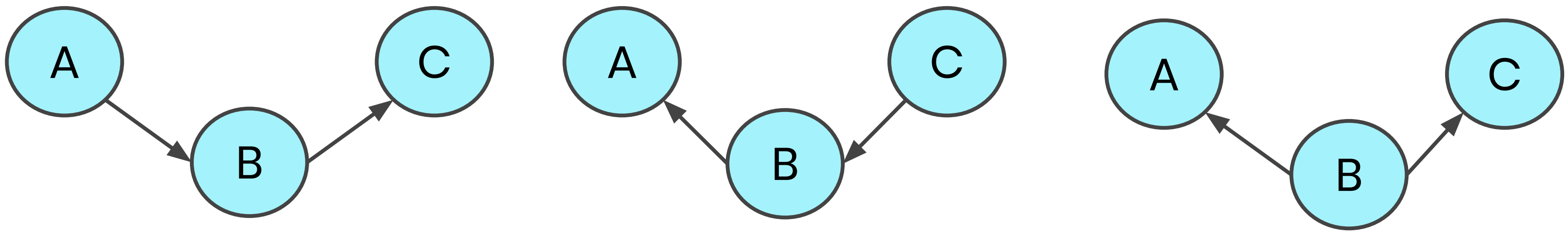
A and C are correlated



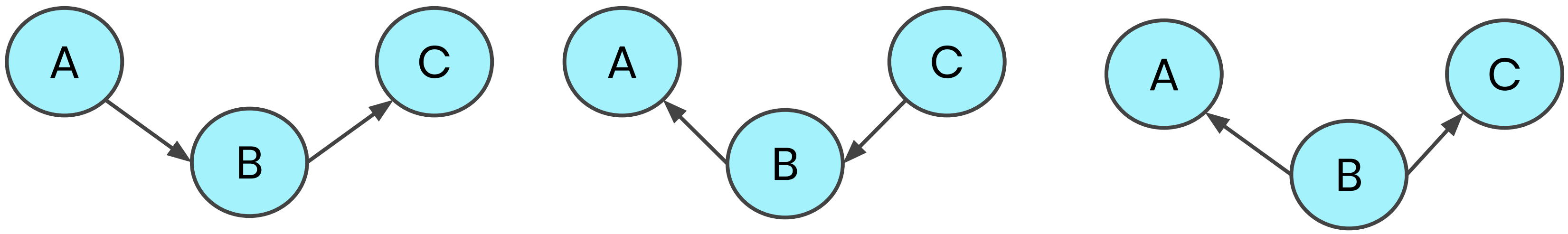
A and C are not correlated conditioned on B; A and C are conditionally independent on B



For this scenario, we have three possible Causal graphs



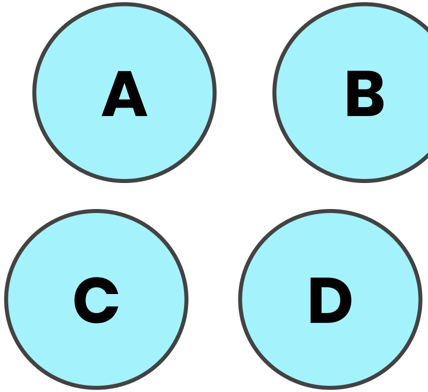
For this scenario, we have three possible Causal graphs



These are called
Markov Equivalent Structures

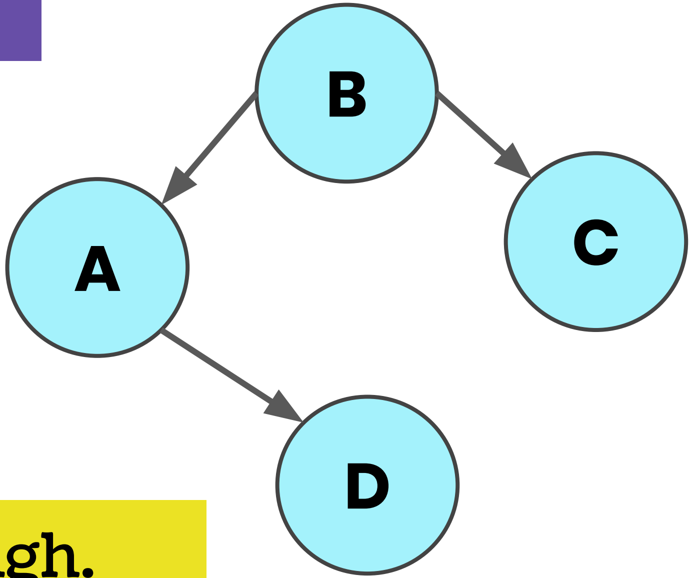
Data

(Variables and Joint Probability Distribution)

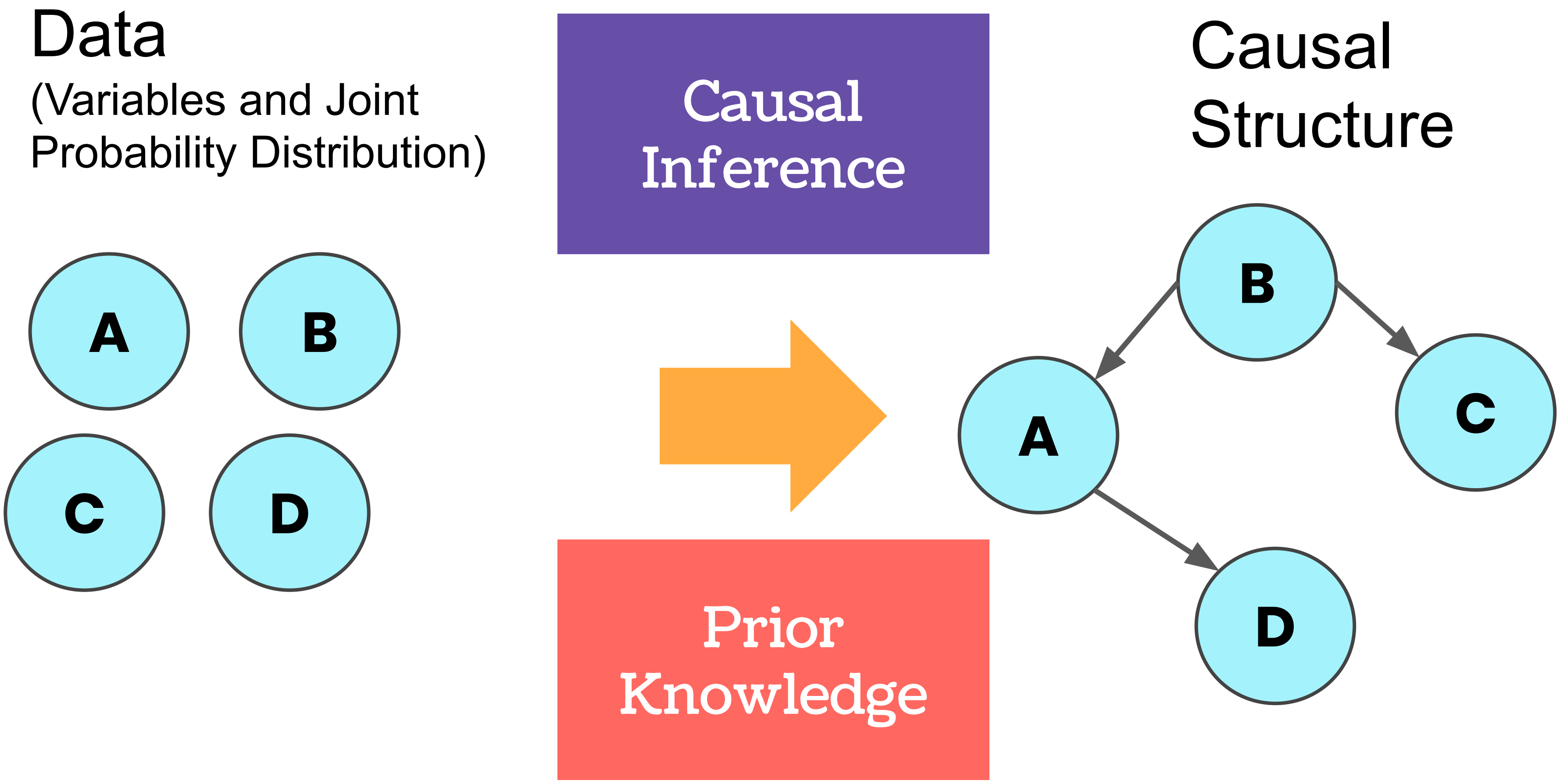


Causal Inference

Causal Structure

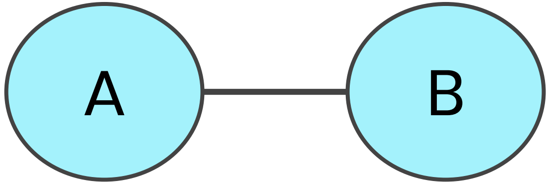


D-separation criterion is not enough. What next?

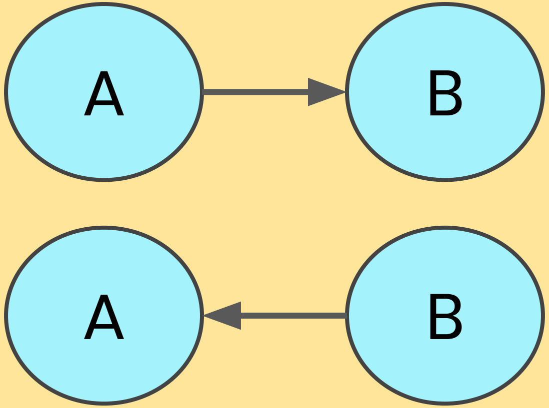

Data
(Variables and Joint
Probability Distribution)
A
B
C
D
Causal
Inference
Prior
Knowledge
Causal
Structure
B
A
C
D

Example 1: Two correlated variables

A and B are correlated (dependent) with each other.



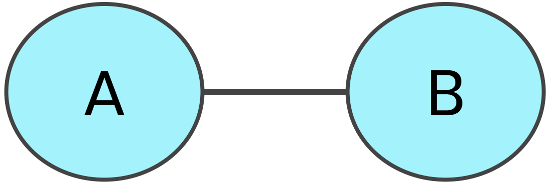
We have no way of telling whether A causes B or B causes A.



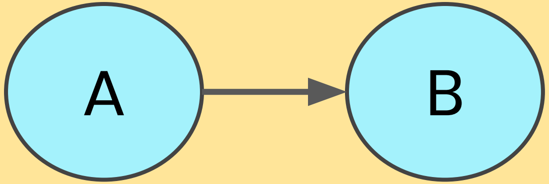
Causal Inference

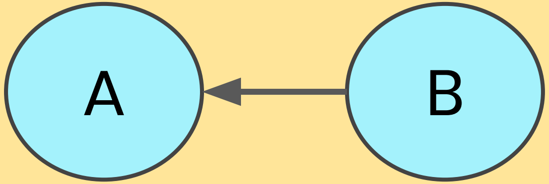
Example 1: Two correlated variables

A and B are correlated (dependent) with each other.



We have no way of telling whether A causes B or B causes A.

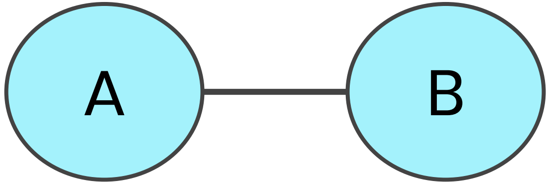




But what if we know that A happened before B ?

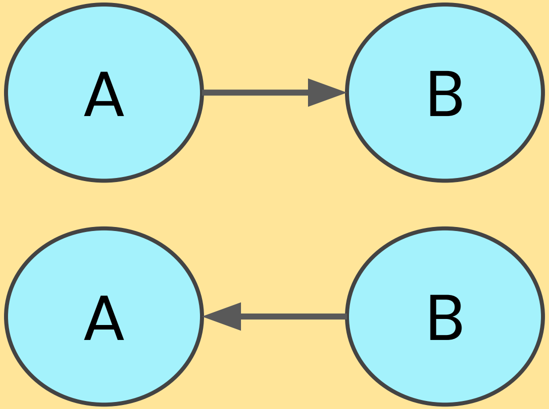
Example 1: Two correlated variables

A and B are correlated (dependent) with each other.



We have no way of telling whether A causes B or B causes A.

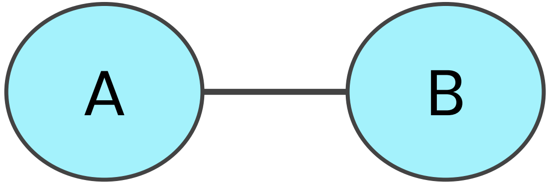
But what if we know that A happened before B ?



Causal Inference

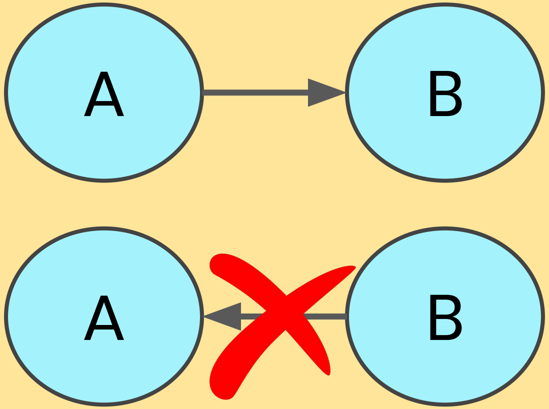
Example 1: Two correlated variables

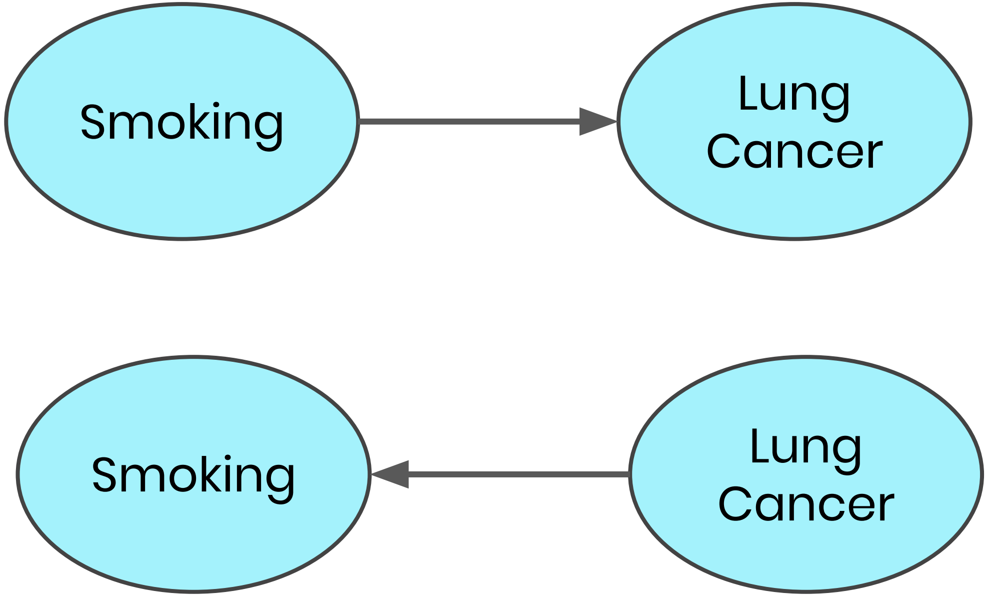
A and B are correlated (dependent) with each other.



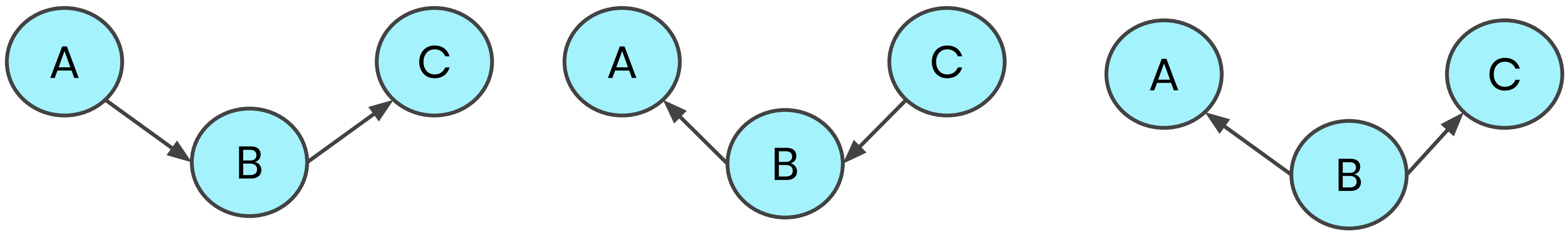
We have no way of telling whether A causes B or B causes A.

But what if we know that A happened before B ?

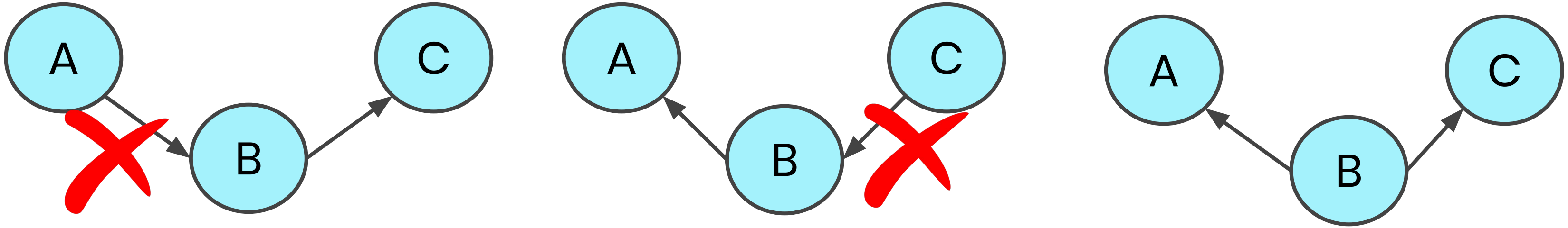



Smoking
Lung Cancer
Smoking
Lung Cancer

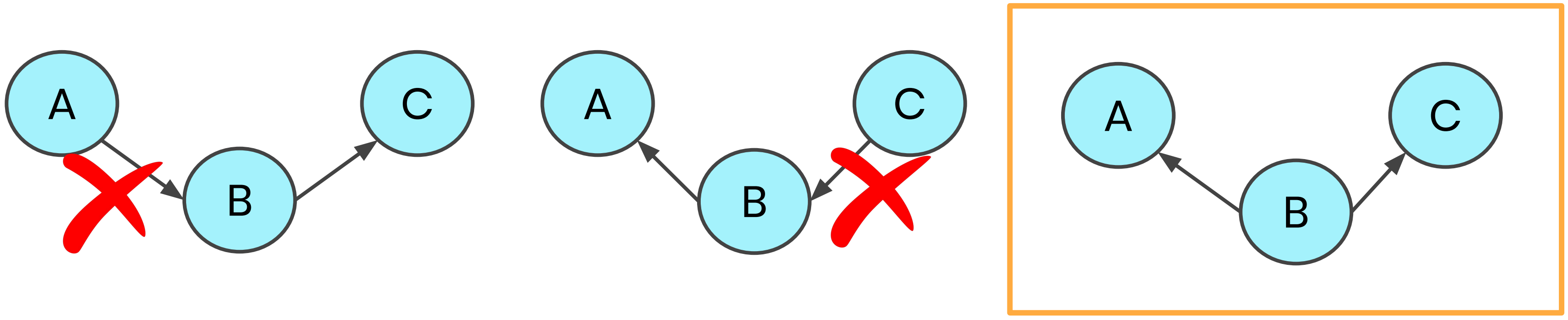
What if we know B happened before A and C?



What if we know B happened before A and C?

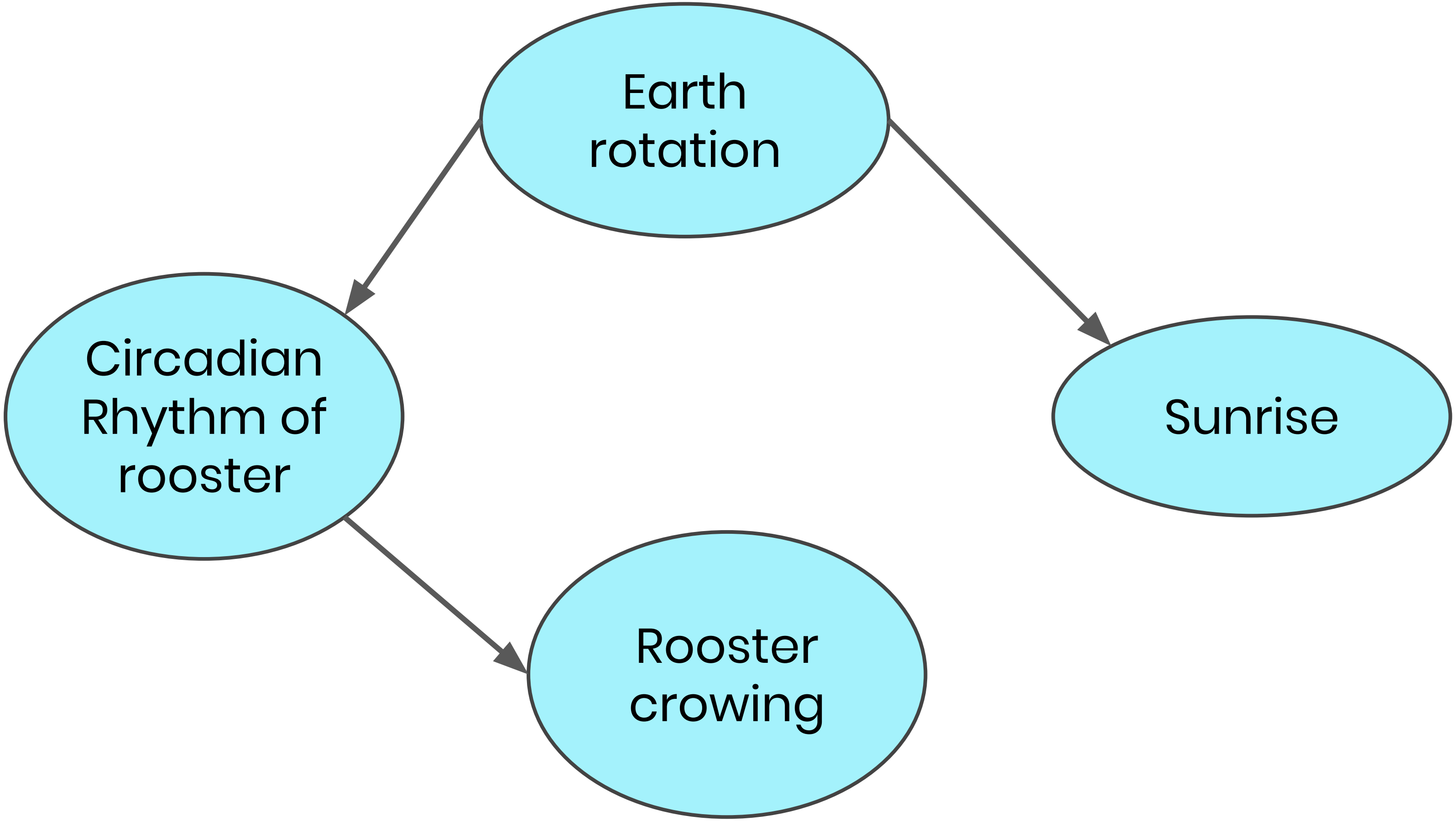


What if we know B happened before A and C?



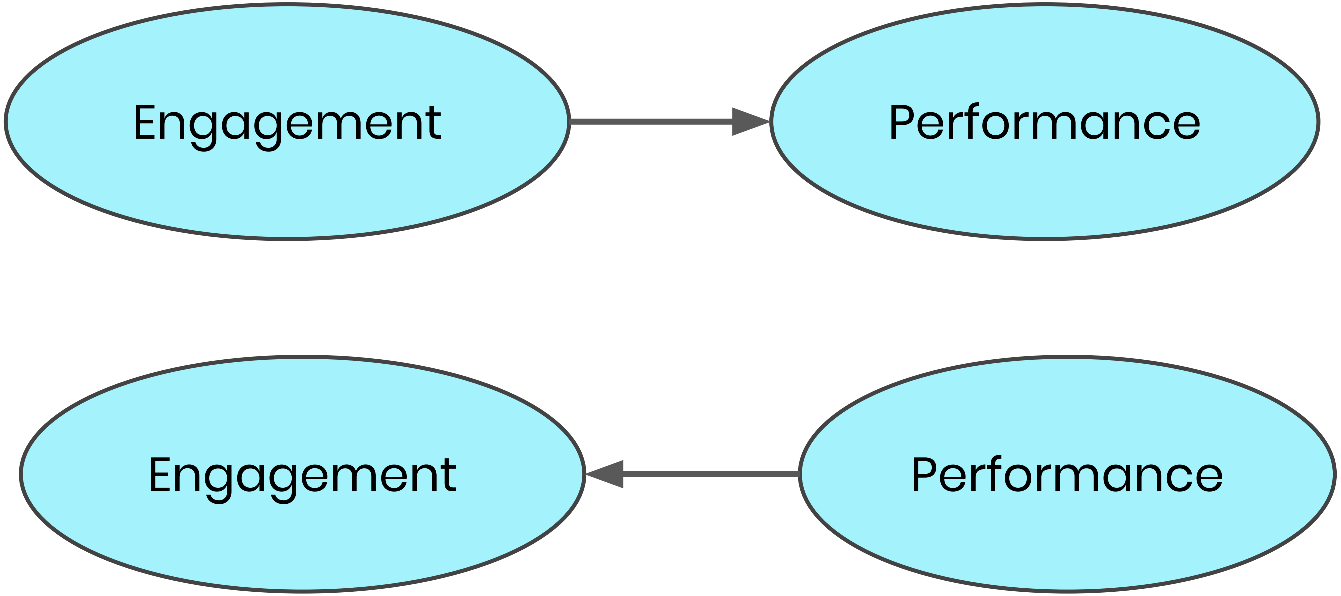
But can temporal precedence (happening earlier in time) always determine cause-effect?

Rooster crowing happens earlier than sunrise.



Earth rotation
Circadian Rhythm of rooster
Sunrise
Rooster crowing

And what if you cannot tell which variable occurred first?



Causal Modeling is not a perfect process.

It is still in relative infancy!

Why are Causal Modeling stronger than other statistical methods?

1. D-separation criterion / Causal Inference
2. Prior Knowledge
3. Causal Assumptions

Causal Assumptions

Statistical methods employ statistical assumption such as normality, independence, homoscedasticity, etc. On top of these statistical assumptions, causal modeling adds causal assumptions

1. Causal Markov assumption: A variable X is independent of every other variable (except X's effects) conditional on all of its direct causes.
2. Faithfulness: independencies within data is generated not by coincidence but by structure
3. Causal sufficiency: the set of measured variables M include all of the common causes of pairs in M

Limitations of Causal Modeling

1. Causal Assumptions
2. Markov Equivalent Structures

likeMath has both direct (likeMath→ %correct) and indirect (likeMath→ preTestScore→ %correct) effect on %correct. Based on this, we are considering two possible causal models

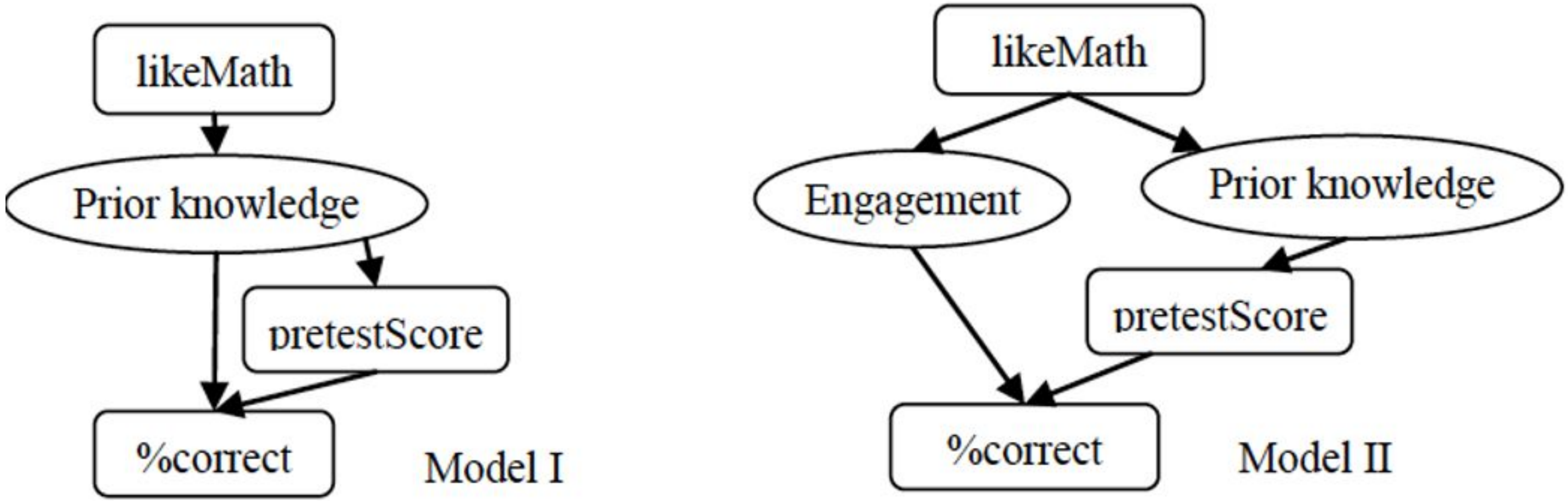
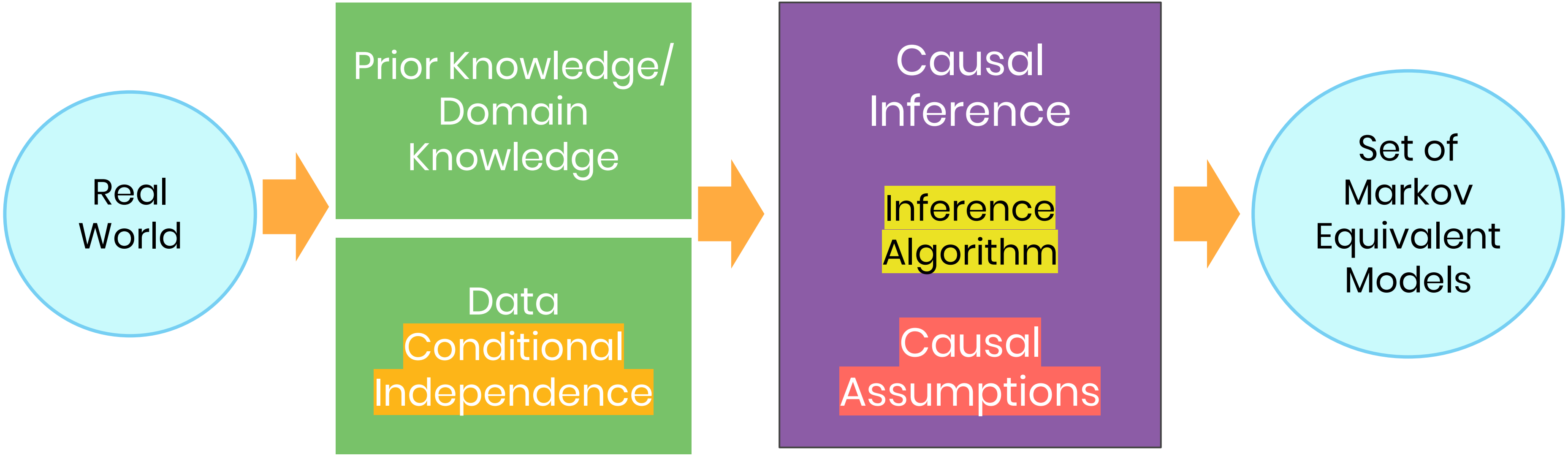
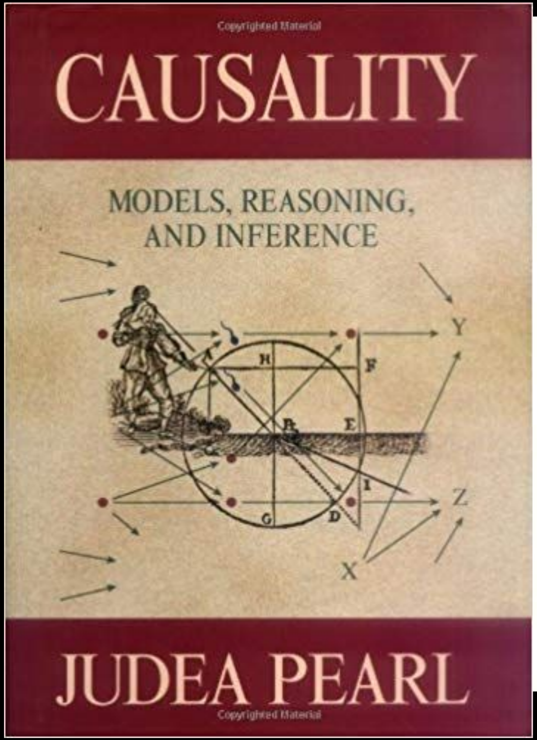
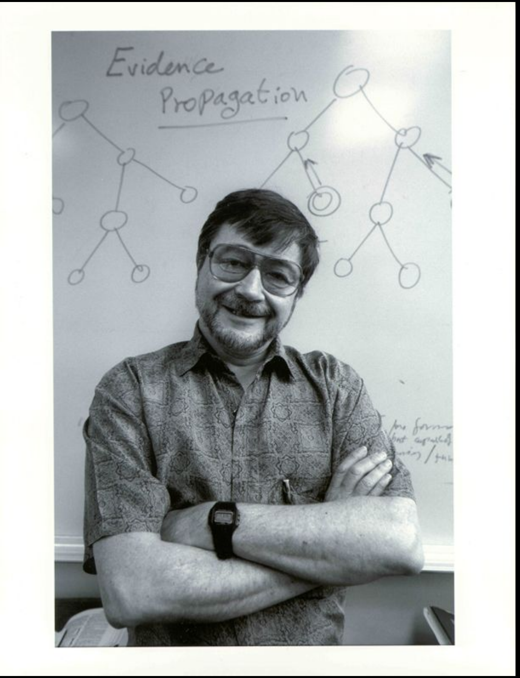


Figure 43 Two possible causal models linking LikeMath and %Correct

Causal Modeling Process



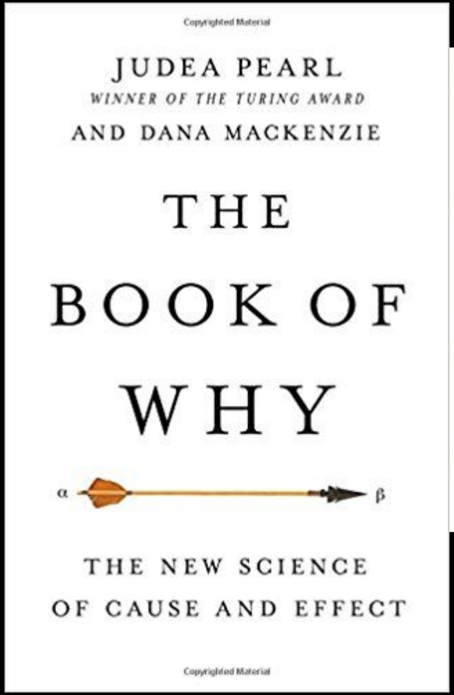

Copyrighted Material
CAUSALITY
MODELS, REASONING, AND INFERENCE
JUDEA PEARL
Copyrighted Material

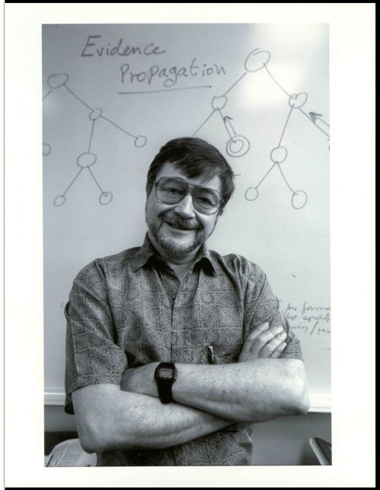

Evidence
Propagation

Open Challenge from Judea Pearl

"All the impressive achievements of deep learning amount to just curve fitting.

To Build Truly Intelligent Machines, Teach Them Cause and Effect"






THE
HOBBIT
THERE AND BACK AGAIN

Judea Pearl
3,638 Tweets

ooo Following

Judea Pearl
@yudapearl

bayes.cs.ucla.edu/WHY/ Joined July 2015

16 Following **26.3K** Followers

Followed by Sujaya Neupane, Naamii Nepal, and 10 others you follow

Tweets | **Tweets & replies** | **Media** | **Likes**

Pinned Tweet

Judea Pearl @yudapearl · Jun 28, 2018

Hi everybody, the intense discussion over The Book of Why drove me to add my two cents. I will not be able to comment on every tweet, but I will try to squeak where it makes a difference.

29 63 405

Judea Pearl @yudapearl · 19h

Seriously. What did the panelists think when the words "the ML community's approach to causal inference" were spoken? Did they take the word "approach" to mean "aspirations" or "buzz words" or "pay attention to" or (my favorite) "careful thinking" #Bookofwhy

1 9

Show this thread

Judea Pearl @yudapearl · 20h

Great thread! As an outside observer, the punchiest punch-line was: Contrast traditions that start with *the data generating process* versus those starting with *the algorithm*. The most perplexed line was "the ML community's *APPROACH* (???) to causal inference" #Bookofwhy

> **Nathan Kallus** @nathankallus · 21h
>
> @SusanMurphylab1 & @Susan_Athey: even if we shouldn't really care about classical confidence intervals, the decision makers at funding agencies, world bank, etc do care right now and so that's where we have to start. Maybe we can change that in the future...
>
> Show this thread

1 3 23

AI in Education: *80 minutes Tour*

- **AI in Education:** *City Tour (15 min)*
- **Intelligent Tutoring System (ITS):** *Restaurant Stop (30 mins)*
  - **Bayesian Modeling vs. Deep Learning:** *Duel Match (15 mins)*
  - **Causal Modeling:** *Detour (5 mins)*

- **AI in Education- Future Frontiers:** *Mountain View (10 mins)*
- **Holy Grail:** *Discussions (5 mins)*

Scope:
Mostly focus on 'AI in Education Research' not covering commercial/ industry products