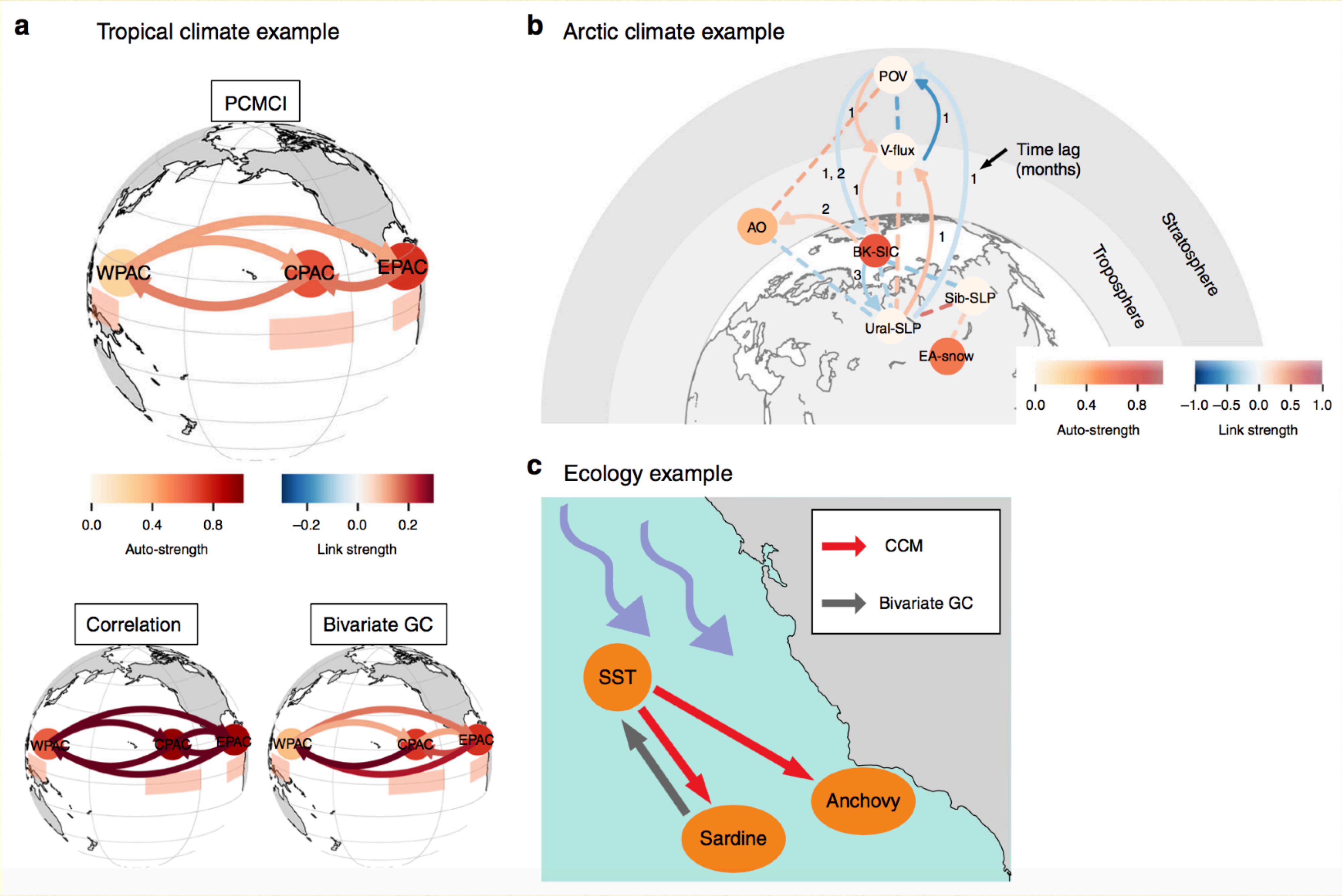
CAUSAL INFERENCE

Matt J. Kusner

University College London

WHY DO WE CARE?



Challenges

Process:

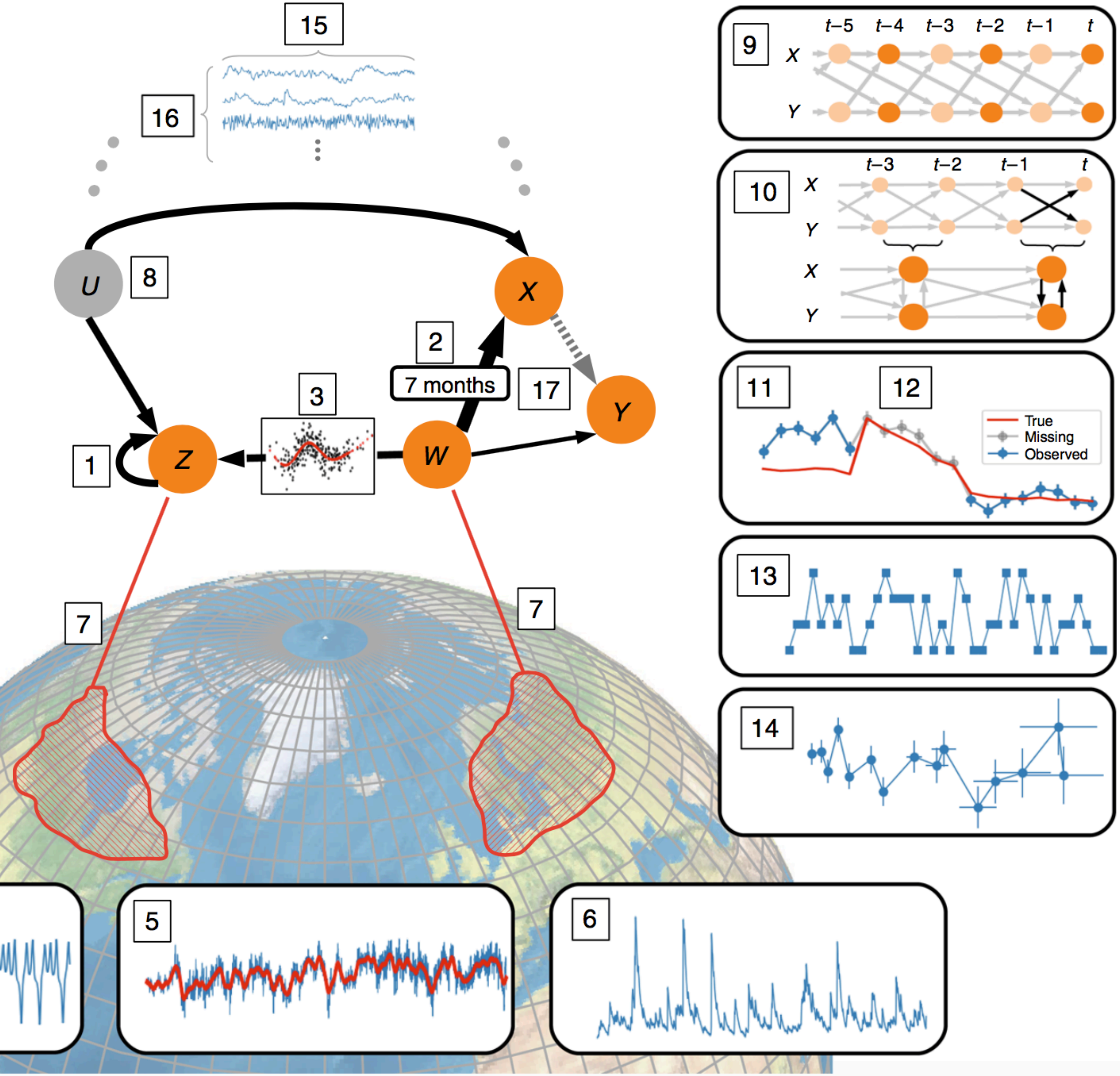
1 Autocorrelation
2 Time delays
3 Nonlinear dependencies
4 Chaotic state-dependence
5 Different time scales
6 Noise distributions

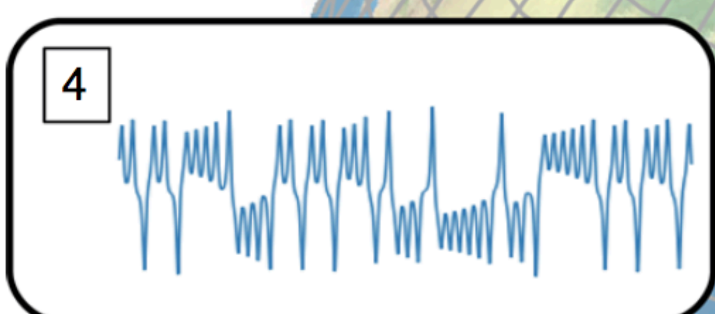
Data:

7 Variable extraction
8 Unobserved variables
9 Time subsampling
10 Time aggregation
11 Measurement errors
12 Selection bias
13 Discrete data
14 Dating uncertainties

Computational/statistical:

15 Sample size
16 High dimensionality
17 Uncertainty estimation





1. WHAT IS CAUSALITY?

2. HOW DO WE ESTIMATE **CAUSAL EFFECTS (INTERVENTIONS)**?

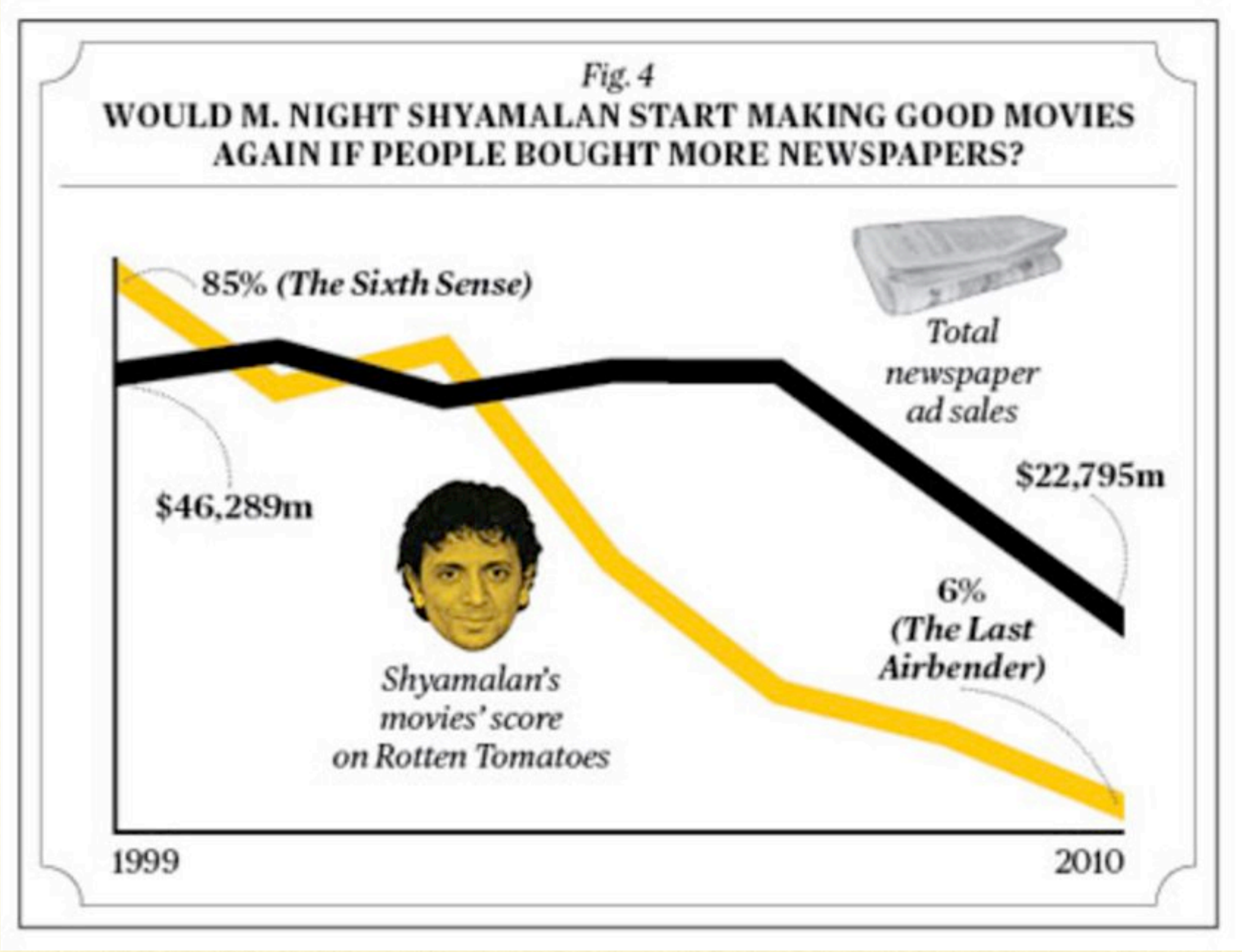
3. HOW DO WE OBTAIN A **CAUSAL GRAPH**?

4. ANOTHER CAUSAL QUANTITY: **COUNTERFACTUALS**

WHAT IS CAUSALITY?

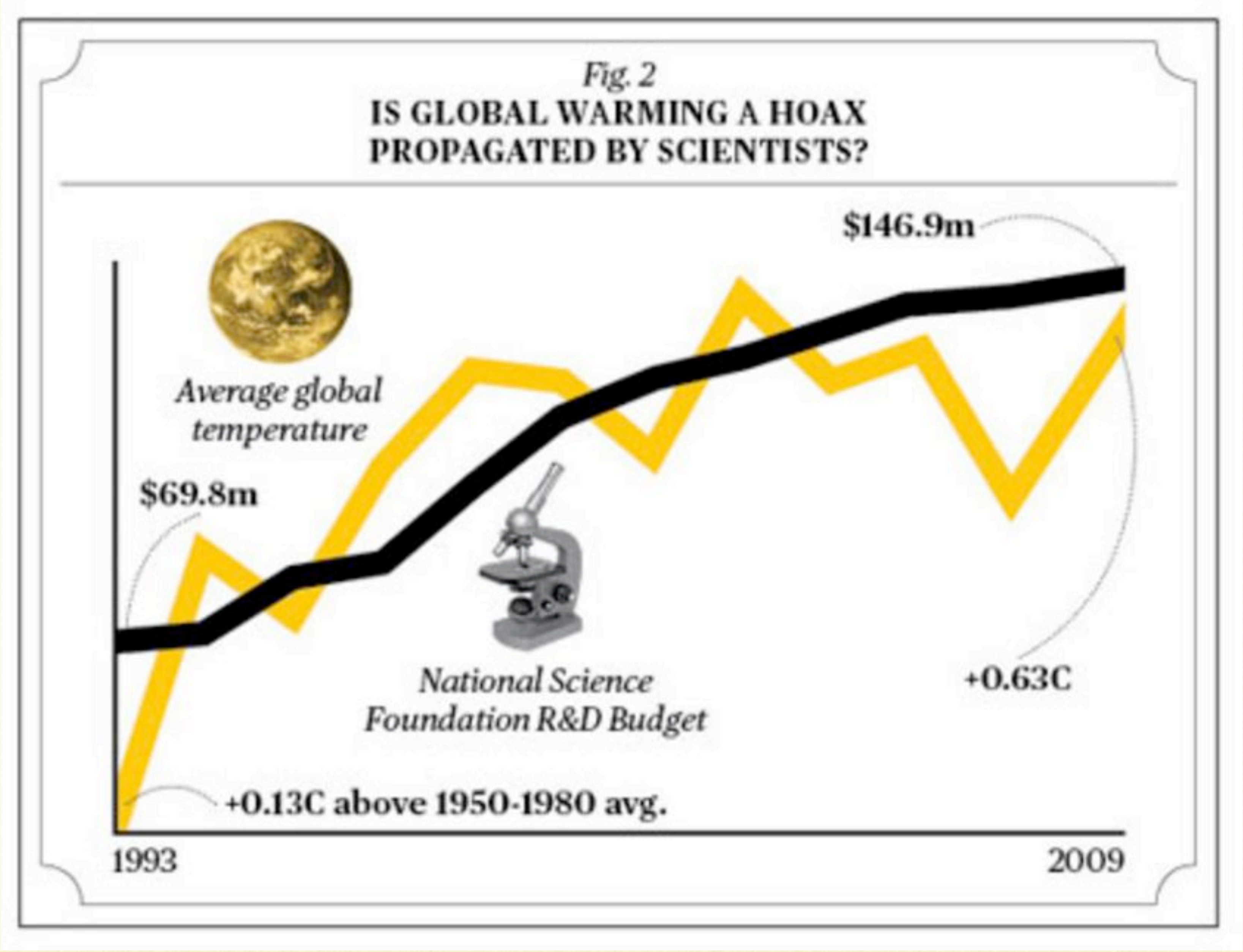
WHAT IS CAUSALITY?

Probably not this…

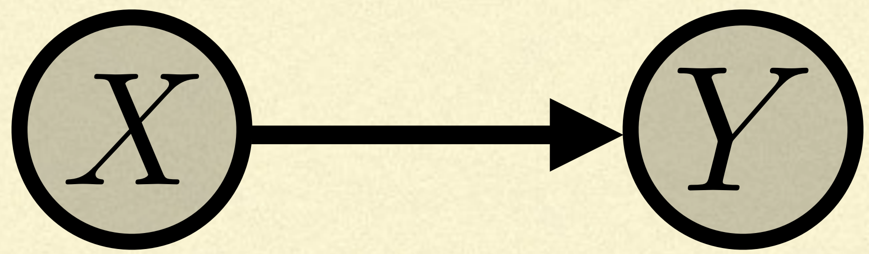


WHAT IS CAUSALITY?

or this…



WHAT IS CAUSALITY?



X **causes** Y

intuitively:

$$\mathbb{P}(Y \mid X \text{ set to be } x) \neq \mathbb{P}(Y \mid X \text{ set to be } x')$$

CONDITIONING VS. INTERVENING

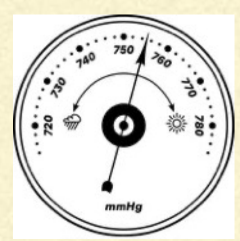
conditioning		**intervening**
$\mathbb{P}(Y \mid X = x)$	vs.	$\mathbb{P}(Y \mid X \text{ set to be } x)$

CONDITIONING VS. INTERVENING

conditioning | **intervening**

$\mathbb{P}(Y \mid X = x)$ vs. $\mathbb{P}(Y \mid X \text{ set to be } x)$

Y:  (rain)

X: 

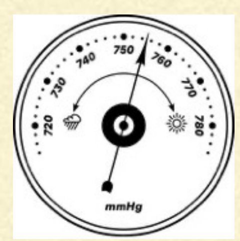
(barometer reading)

CONDITIONING VS. INTERVENING

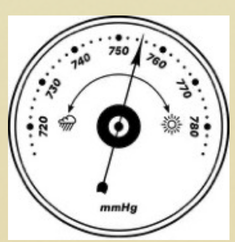
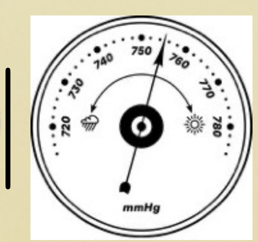
conditioning **intervening**

$\mathbb{P}(Y \mid X = x)$ **vs.** $\mathbb{P}(Y \mid X \text{ set to be } x)$

Y:  (rain)

X:  (barometer reading)

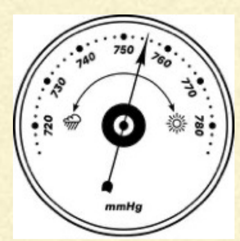
conditioning

$\mathbb{P}($  $|$  $= \text{high}) > \mathbb{P}($  $|$  $= \text{low})$

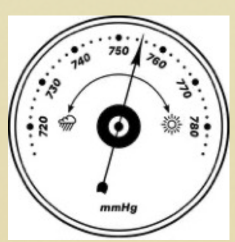
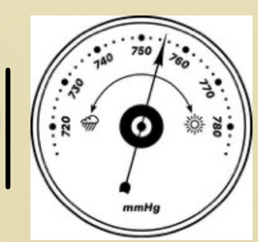
CONDITIONING VS. INTERVENING

conditioning $\mathbb{P}(Y \mid X = x)$ vs. **intervening** $\mathbb{P}(Y \mid X \text{ set to be } x)$

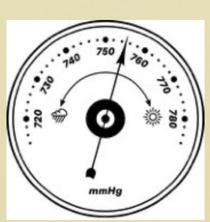
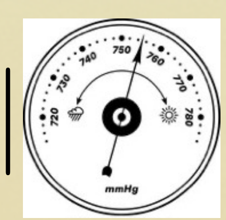
Y:  (rain)

X:  (barometer reading)

conditioning

$\mathbb{P}($  $|$  $= \text{ high}) > \mathbb{P}($  $|$  $= \text{ low})$

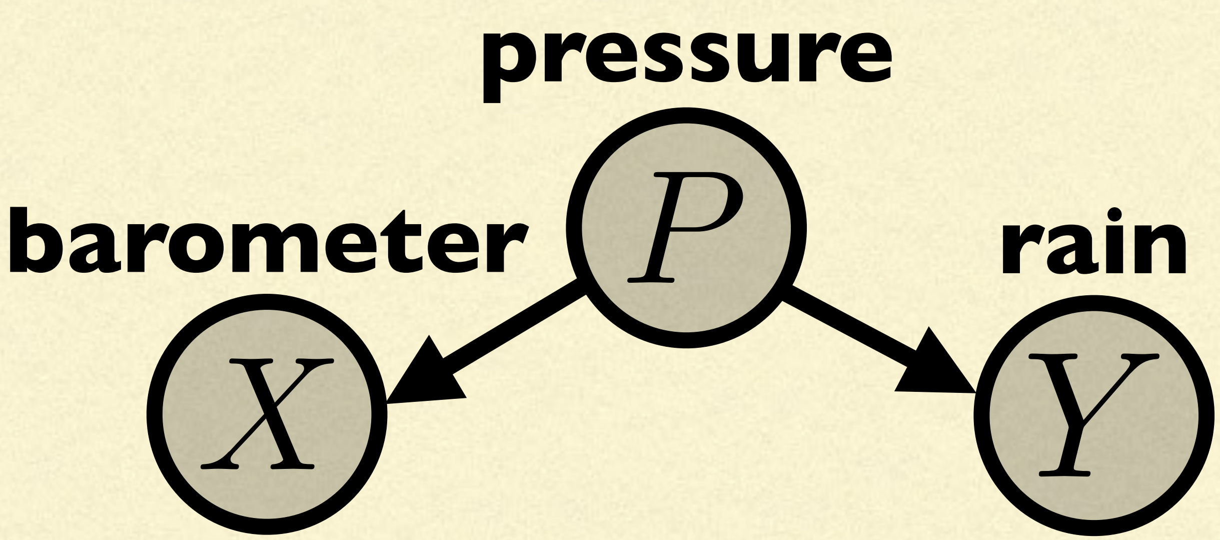
intervening

$\mathbb{P}($  $|$  $\text{set to be high}) = \mathbb{P}($  $|$  $\text{set to be low})$

HOW DOES THIS ARISE?

HOW DOES THIS ARISE?

Confounding

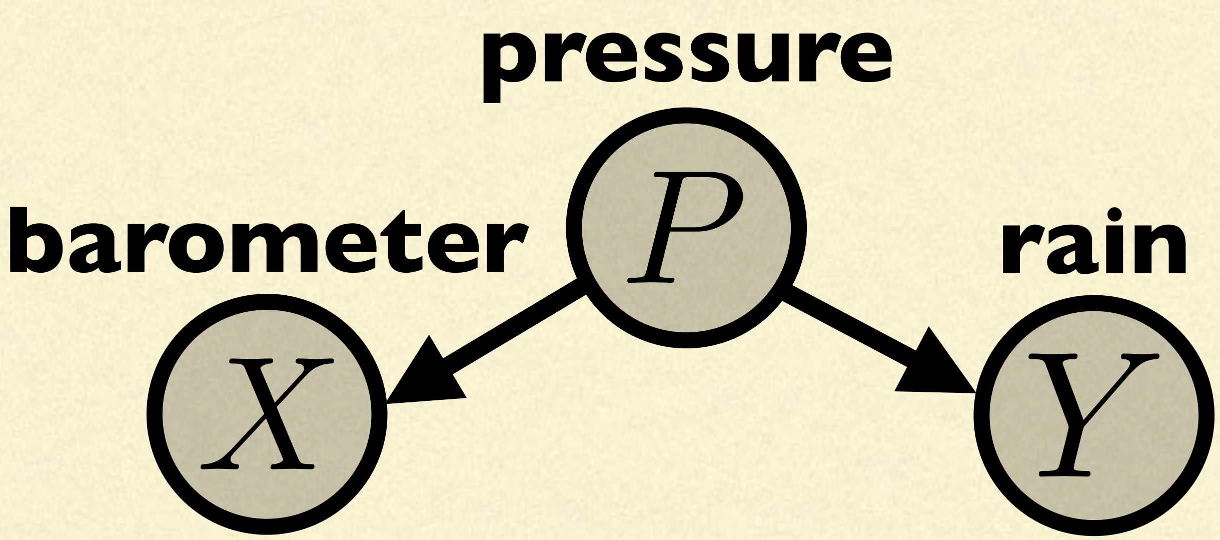


conditioning **intervening**

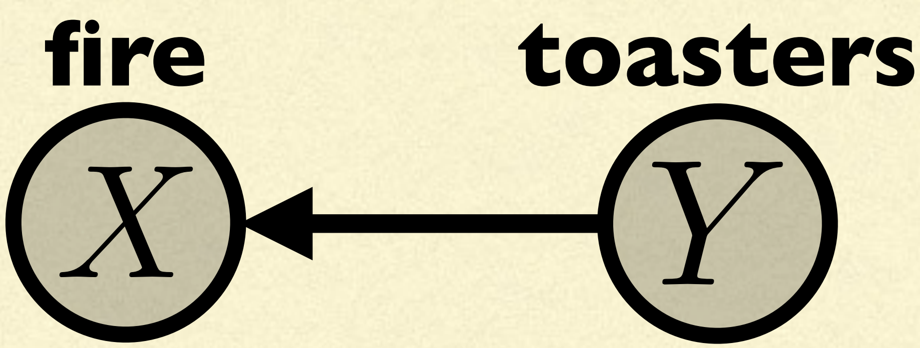
$$\mathbb{P}(Y \mid X = x) \neq \mathbb{P}(Y \mid X \text{ set to be } x)$$

HOW DOES THIS ARISE?

Confounding



Conditioning **against** the causal direction



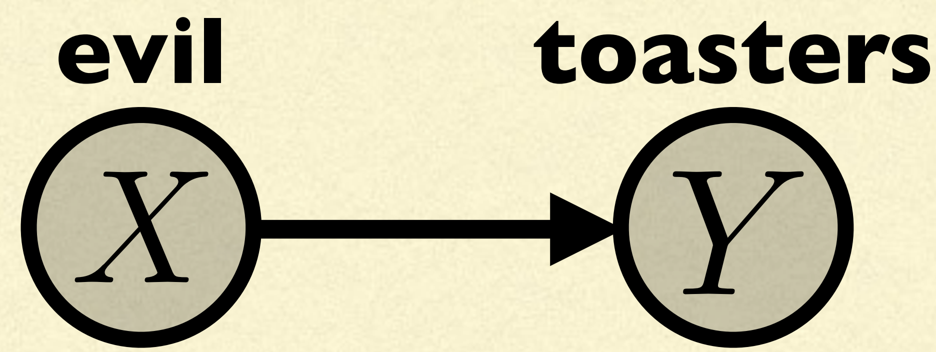
conditioning **intervening**

$$\mathbb{P}(Y \mid X = x) \neq \mathbb{P}(Y \mid X \text{ set to be } x)$$

IN SOME CASES CONDITIONING AND INTERVENING AGREE

Independence

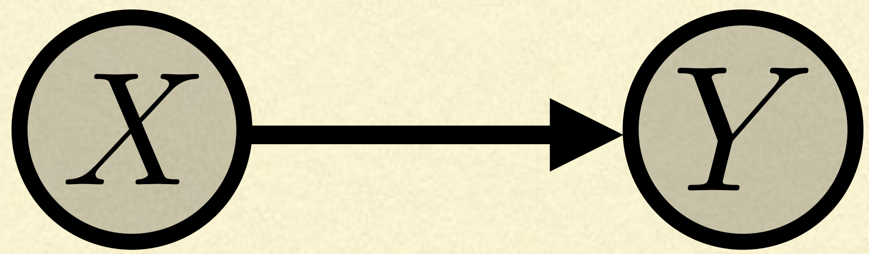
Conditioning **following** the causal direction



conditioning **intervening**

$$\mathbb{P}(Y \mid X = x) \;=\; \mathbb{P}(Y \mid X \text{ set to be } x)$$

WHAT IS CAUSALITY?

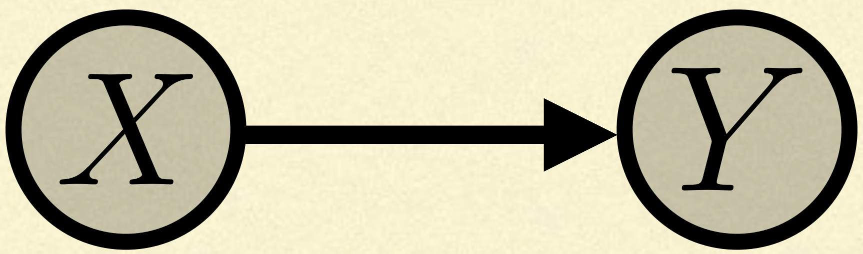


X causes Y

intuitively:

$$\mathbb{P}(Y \mid X \text{ set to be } x) \neq \mathbb{P}(Y \mid X \text{ set to be } x')$$

WHAT IS CAUSALITY?



X causes Y

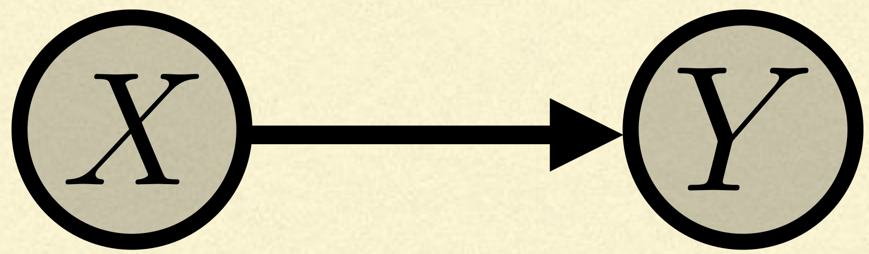
intuitively:

$$\mathbb{P}(Y \mid X \text{ set to be } x) \neq \mathbb{P}(Y \mid X \text{ set to be } x')$$

formally:

$$Y \sim \mathbb{P}(Y \mid f(X))$$

WHAT IS CAUSALITY?



X causes Y

intuitively:

$$\mathbb{P}(Y \mid X \text{ set to be } x) \neq \mathbb{P}(Y \mid X \text{ set to be } x')$$

formally:

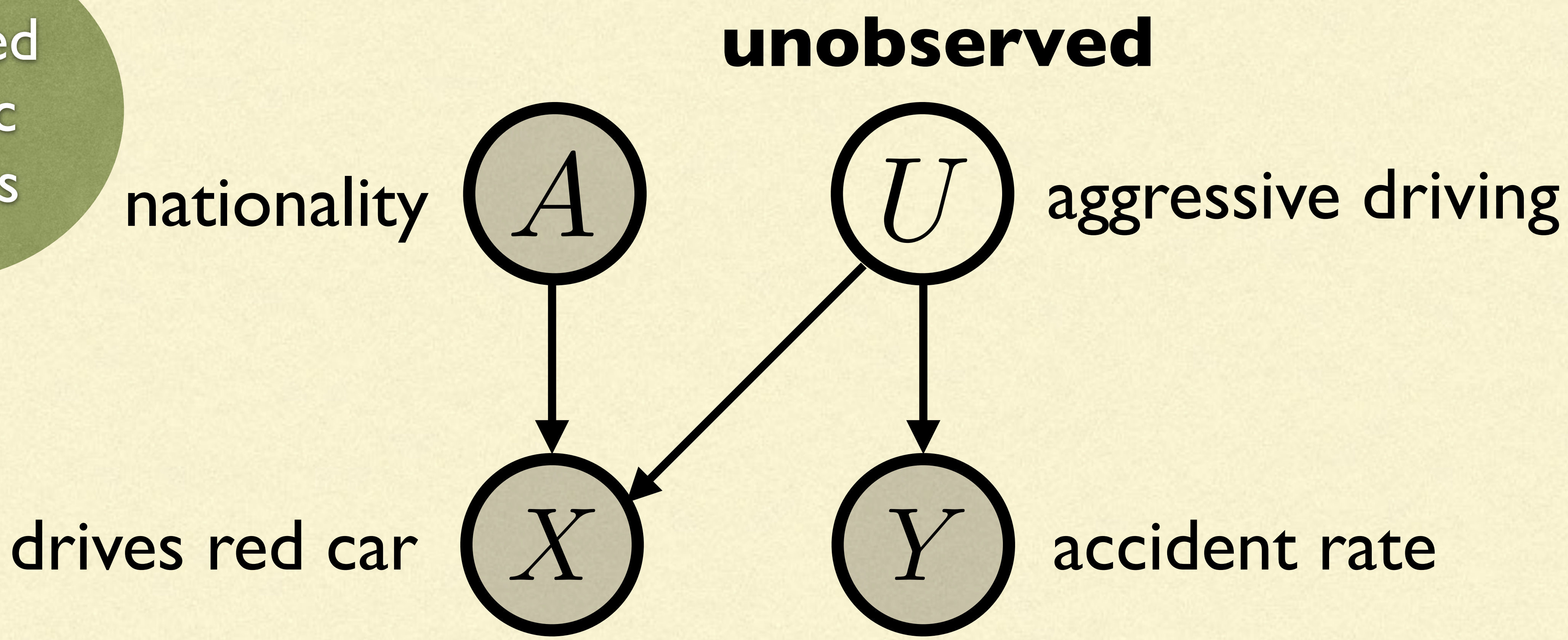
$$Y \sim \mathbb{P}(Y \mid f(X))$$

structural equation model

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

PROBABILISTIC GRAPHICAL MODELS

Directed Acyclic Graphs

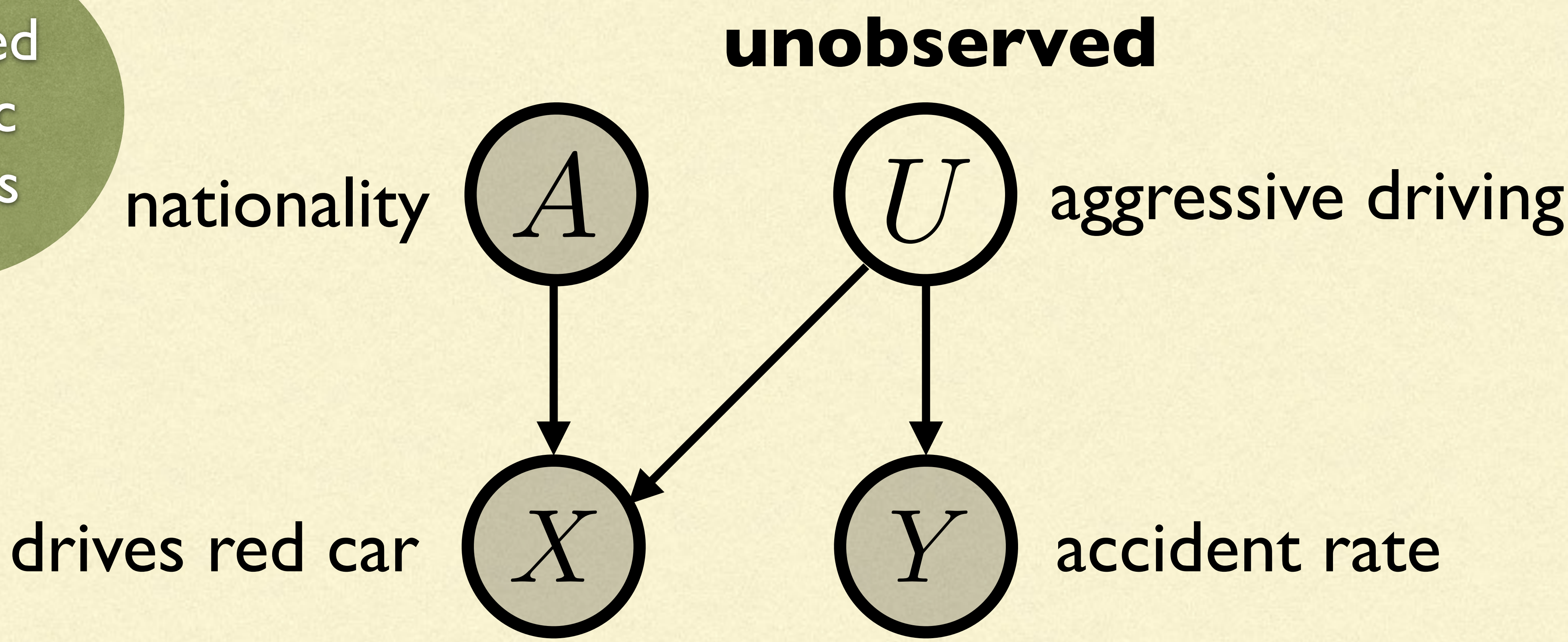


$$X = \alpha A + \beta U \qquad Y = \gamma U$$

PROBABILISTIC CAUSAL MODELS

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

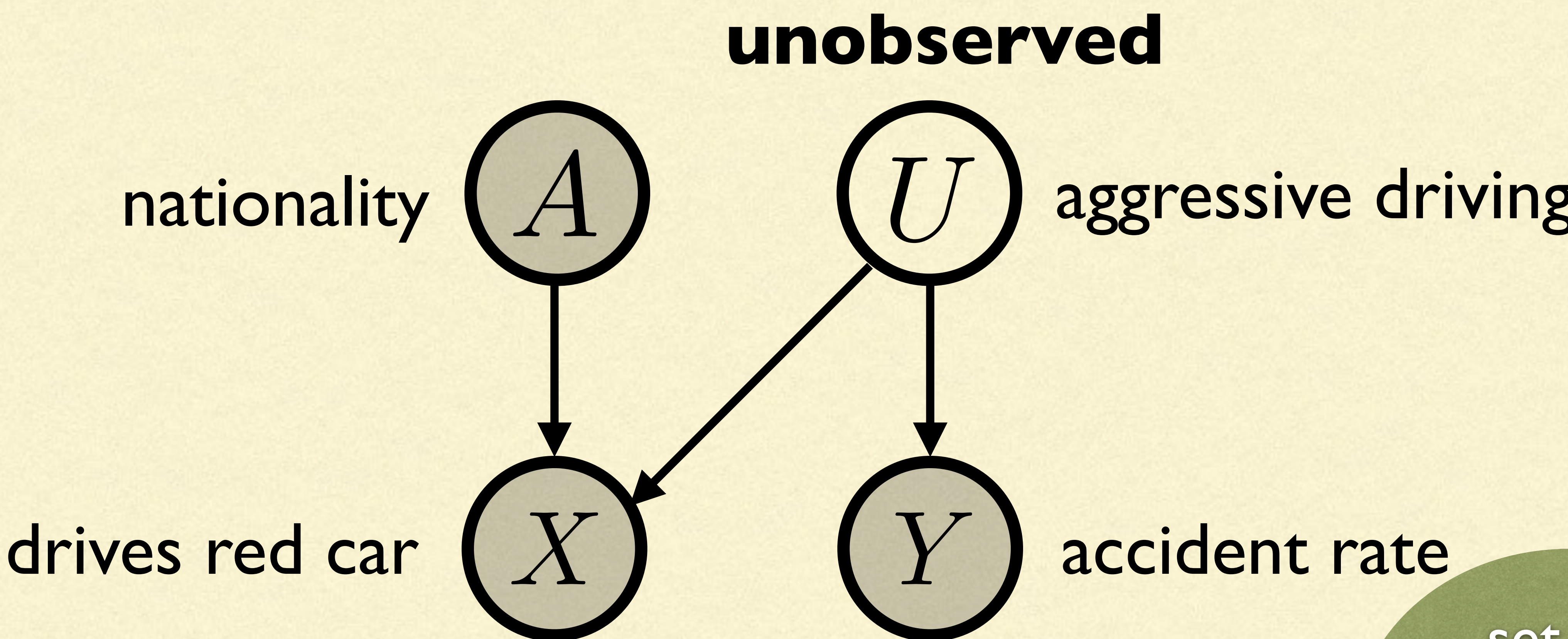
Directed Acyclic Graphs



$$X = \alpha A + \beta U \qquad Y = \gamma U$$

FORMALIZING INTERVENTIONS

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

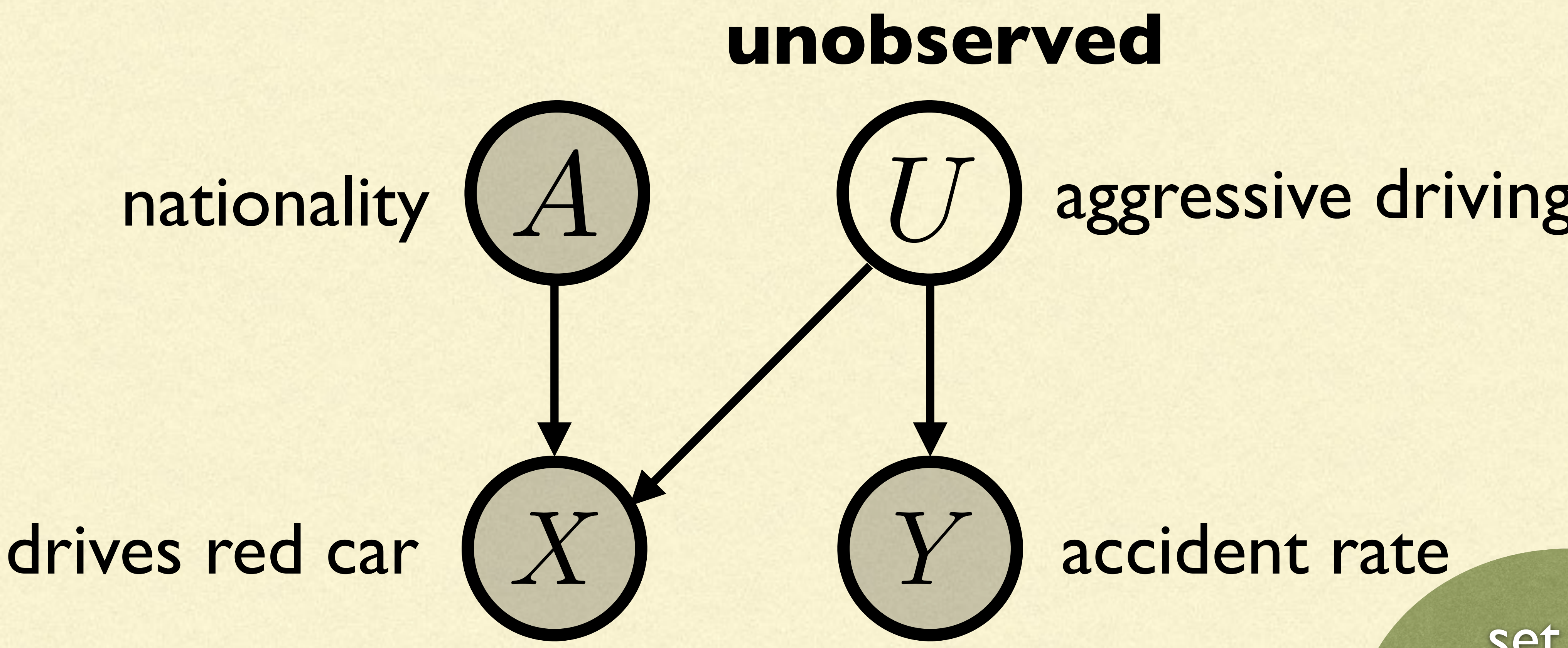


$$X \text{ set to be } x \equiv do(X = x)$$

set for everyone in the population

FORMALIZING INTERVENTIONS

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]



$$X \text{ set to be } x \equiv do(X = x)$$

set for everyone in the population

How to estimate: $\mathbb{P}(Y \mid do(X = x))$

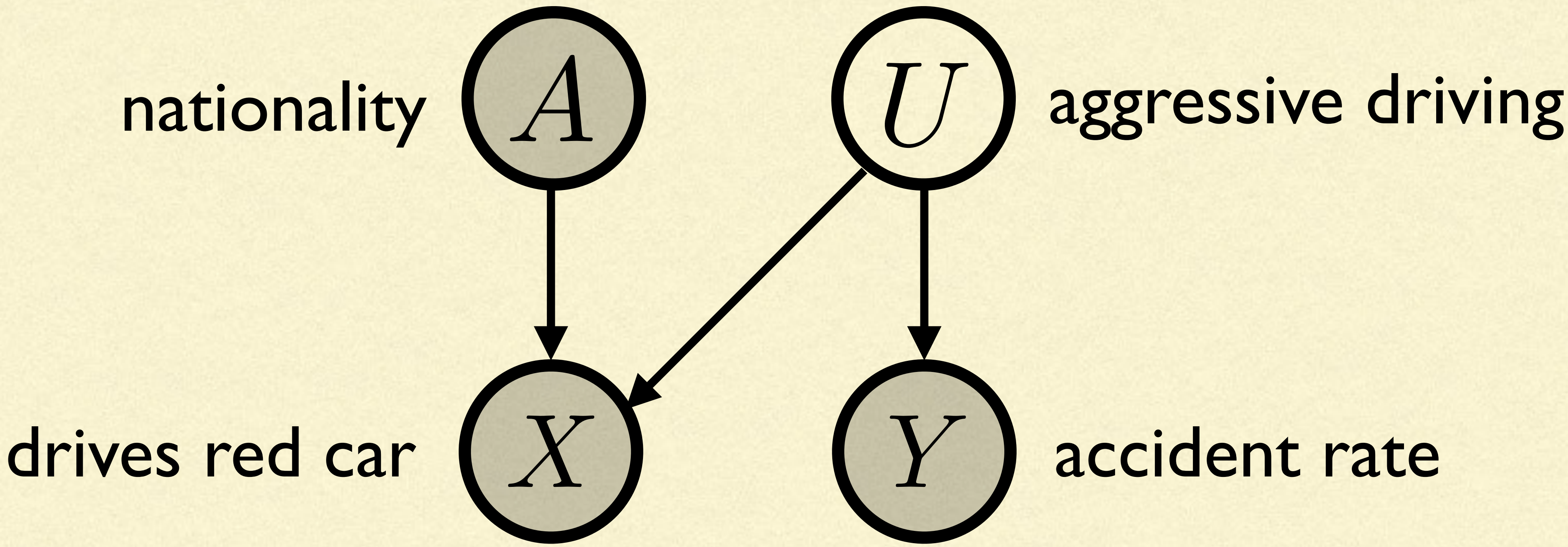
1. WHAT IS CAUSALITY?

2. HOW DO WE ESTIMATE **CAUSAL EFFECTS (INTERVENTIONS)**?

3. HOW DO WE OBTAIN A **CAUSAL GRAPH**?

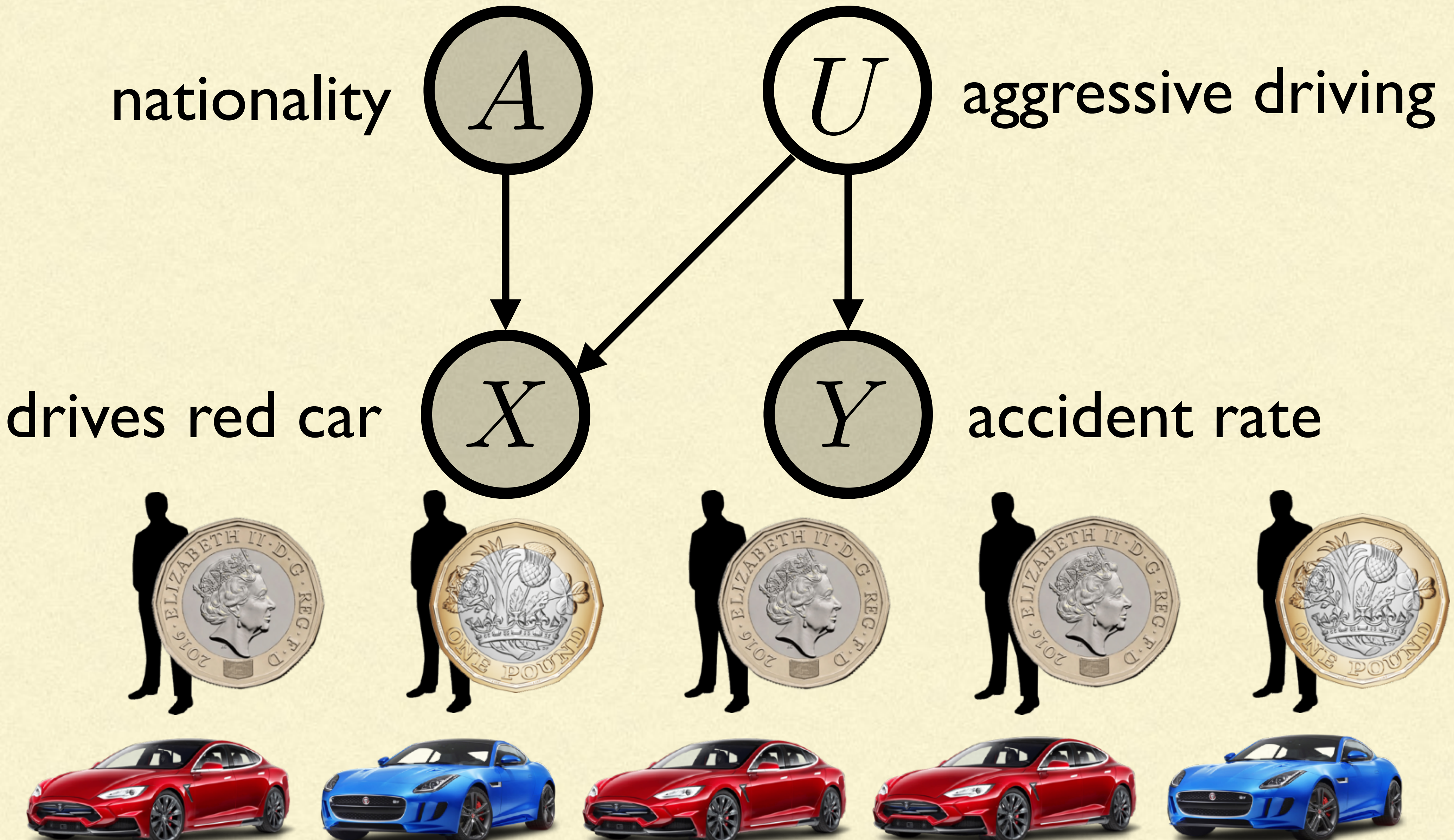
4. ANOTHER CAUSAL QUANTITY: **COUNTERFACTUALS**

THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



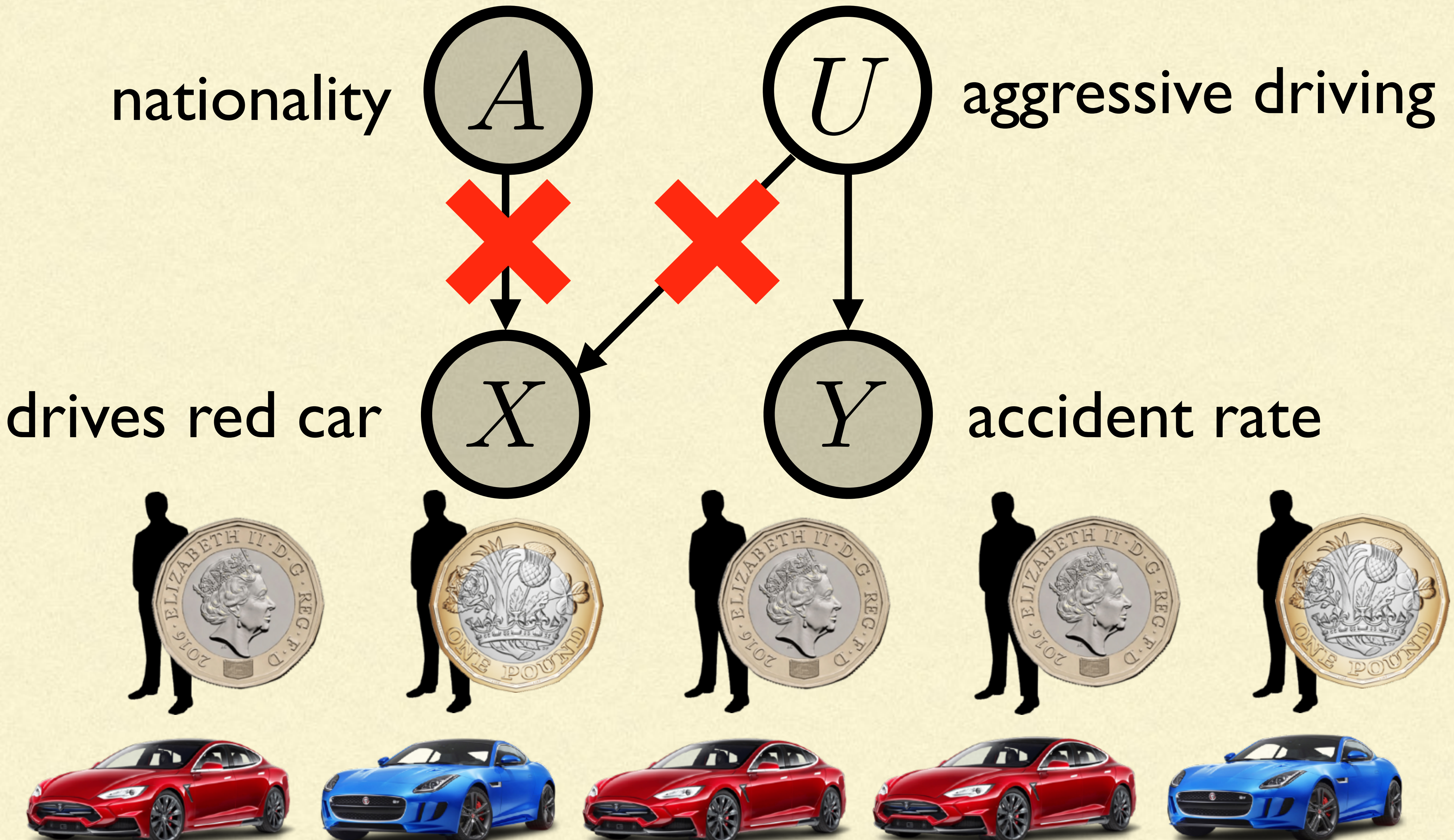
THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS

Be Oprah



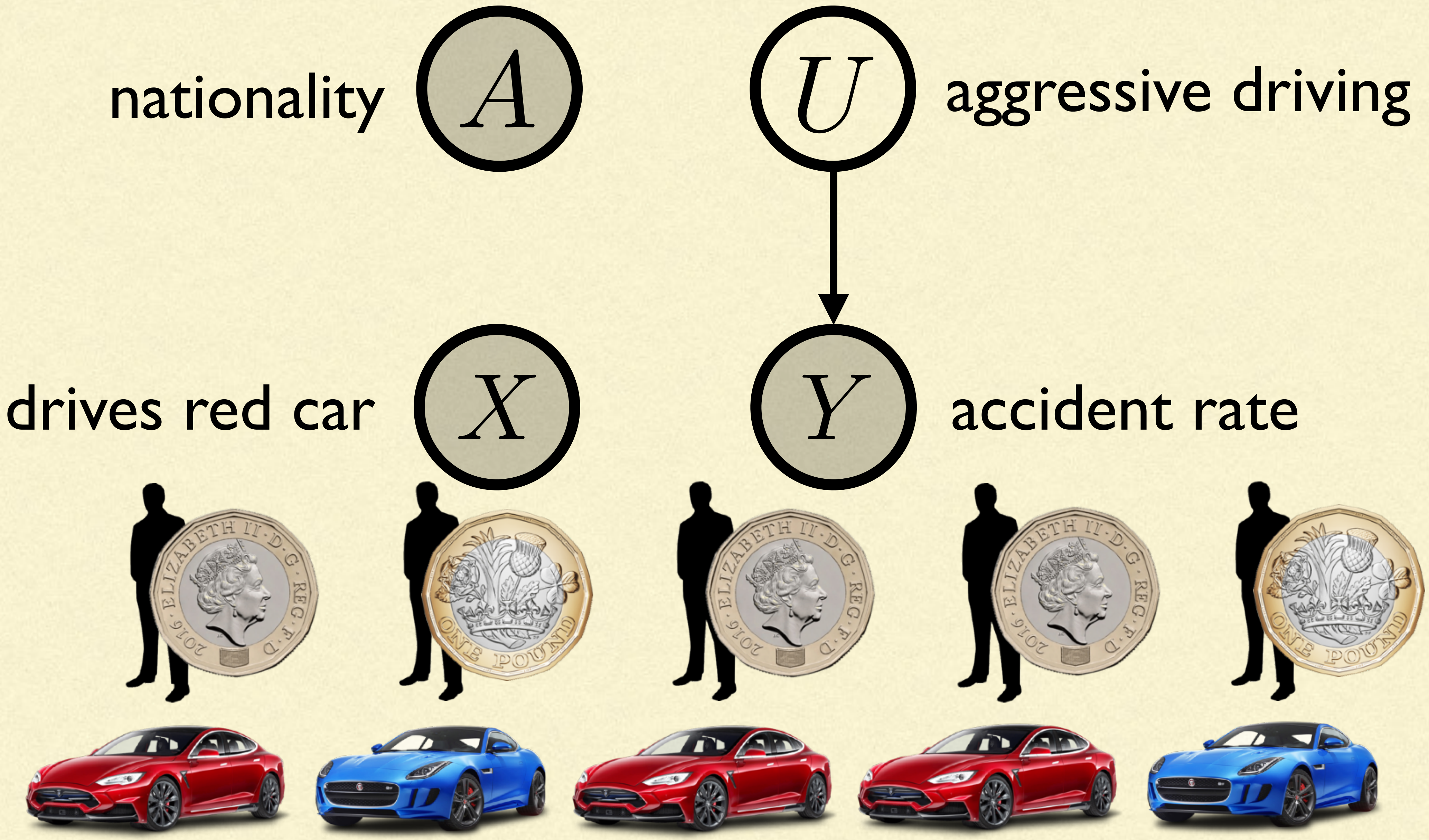
THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS

Be Oprah

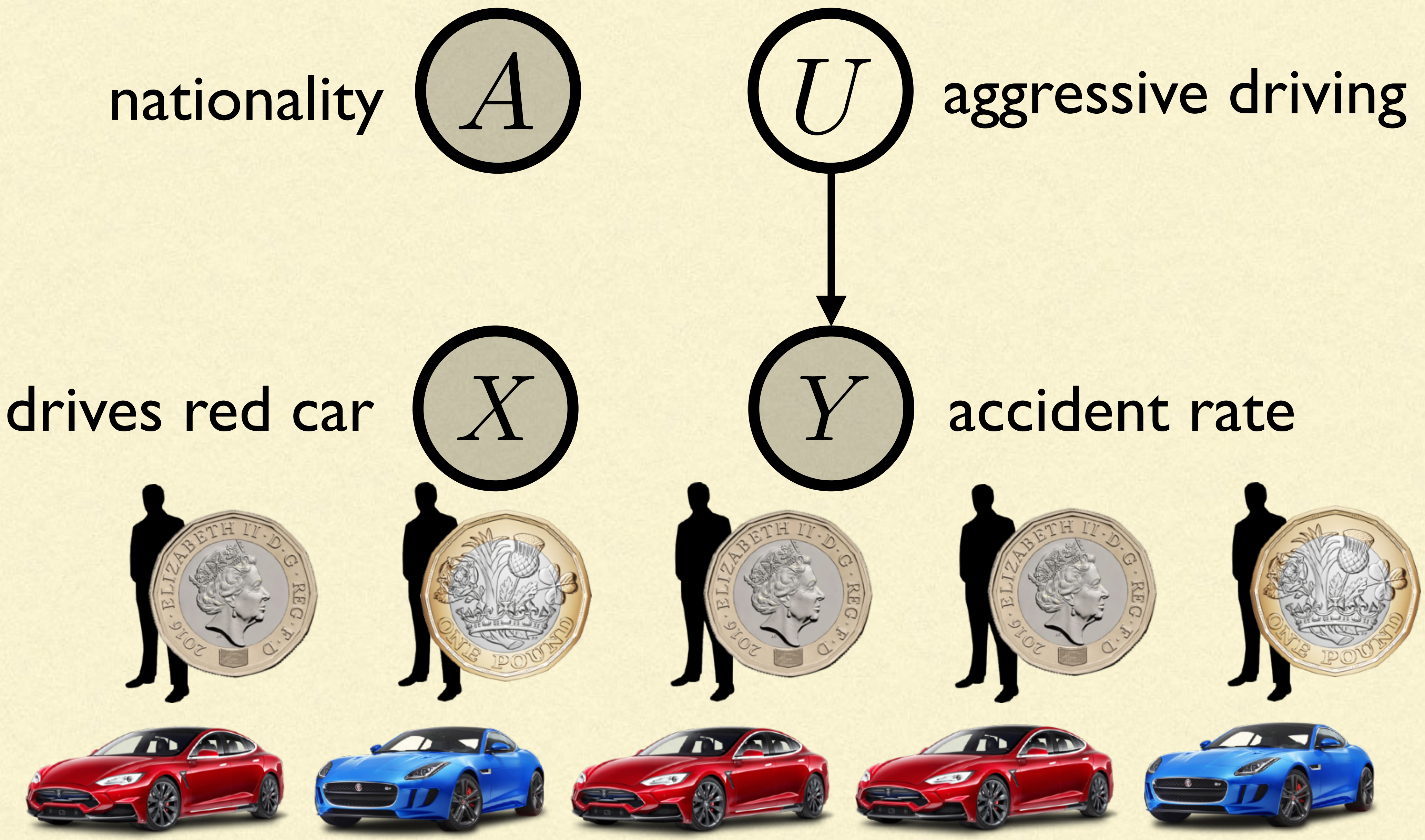


THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS

Be Oprah

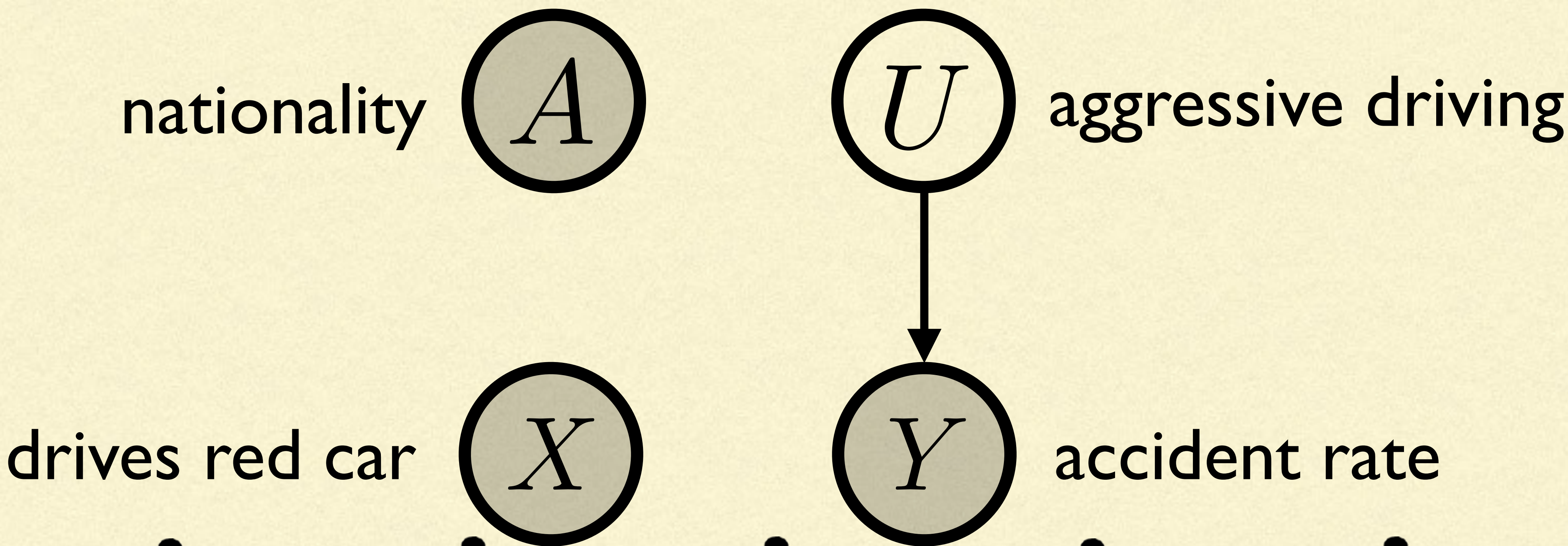


THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



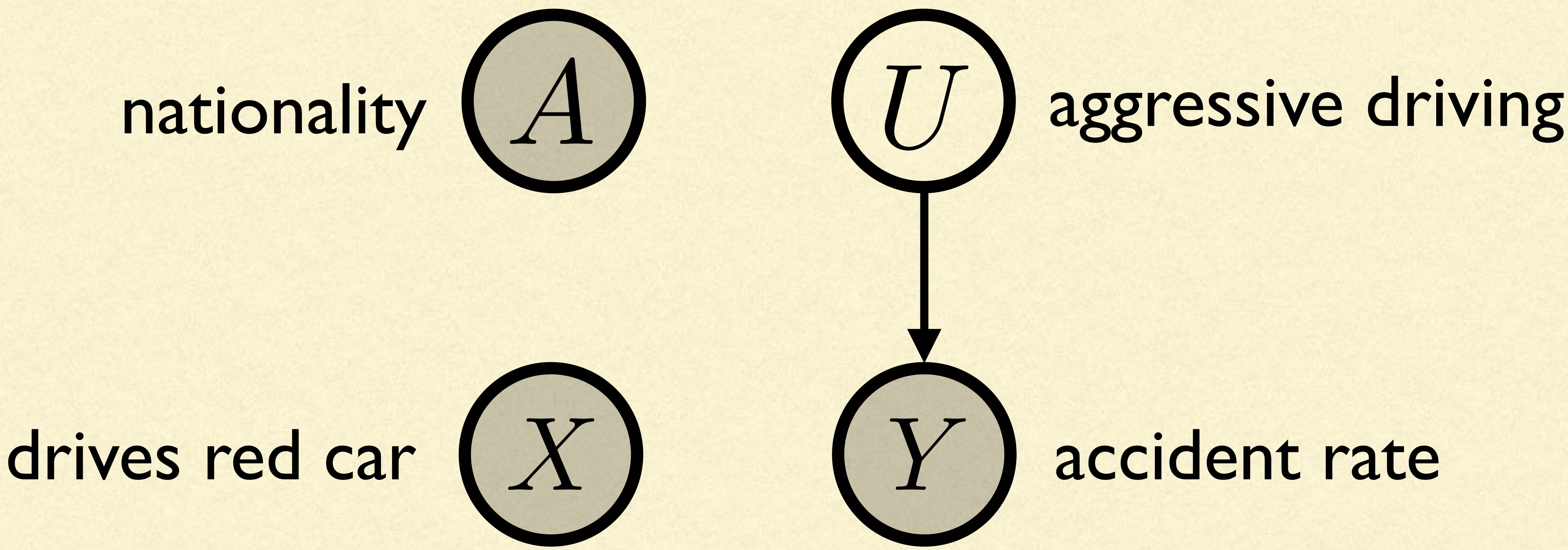
$$\mathbb{P}(Y \mid do(X = x))$$

THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



$$\mathbb{P}(Y \mid do(X = x)) = \mathbb{P}_{\mathrm{m}}(Y \mid X = x)$$

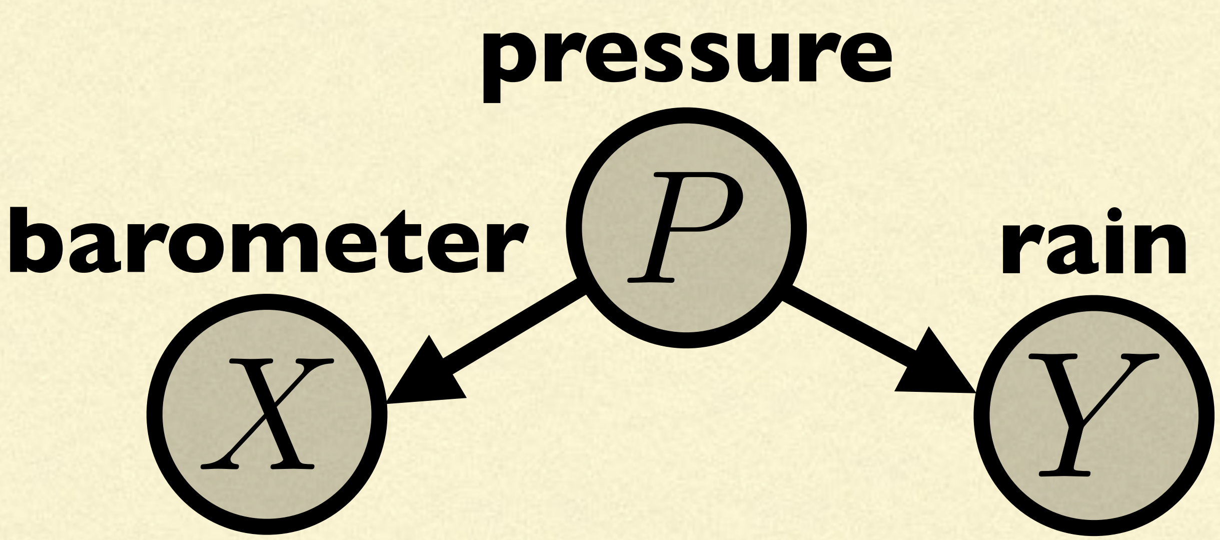
THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



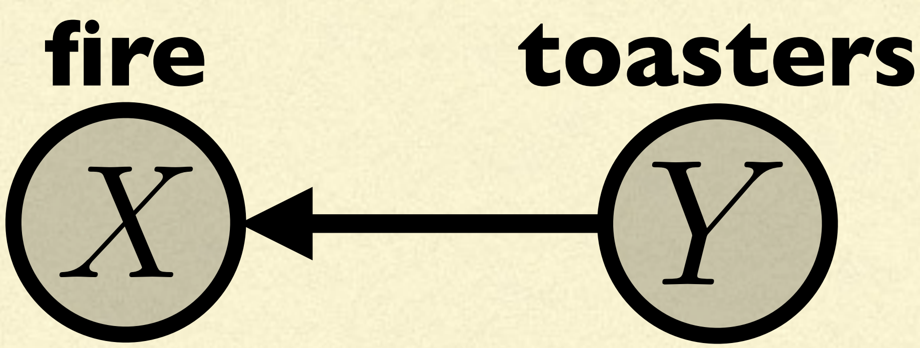
$$\mathbb{P}(Y \mid do(X = x)) = \mathbb{P}_{\mathrm{m}}(Y \mid X = x)$$

WHEN CONDITIONING AND INTERVENING ARE DIFFERENT

Confounding



Conditioning **against** the causal direction

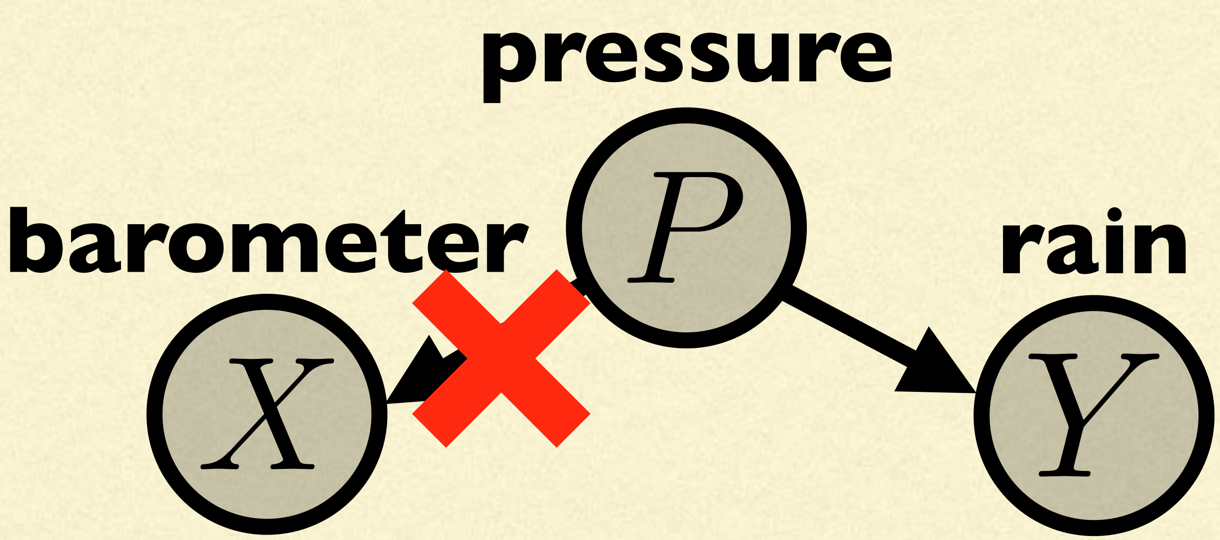


conditioning **intervening**

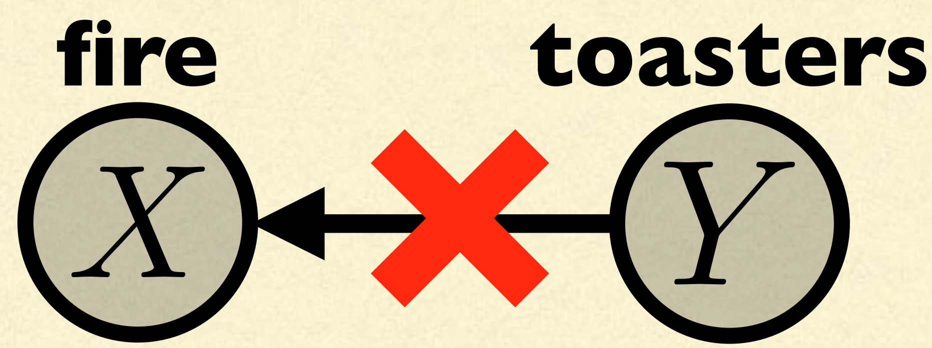
$$\mathbb{P}(Y \mid X = x) \neq \mathbb{P}(Y \mid X \text{ set to be } x)$$

WHEN CONDITIONING AND INTERVENING ARE DIFFERENT

Confounding



Conditioning **against** the causal direction

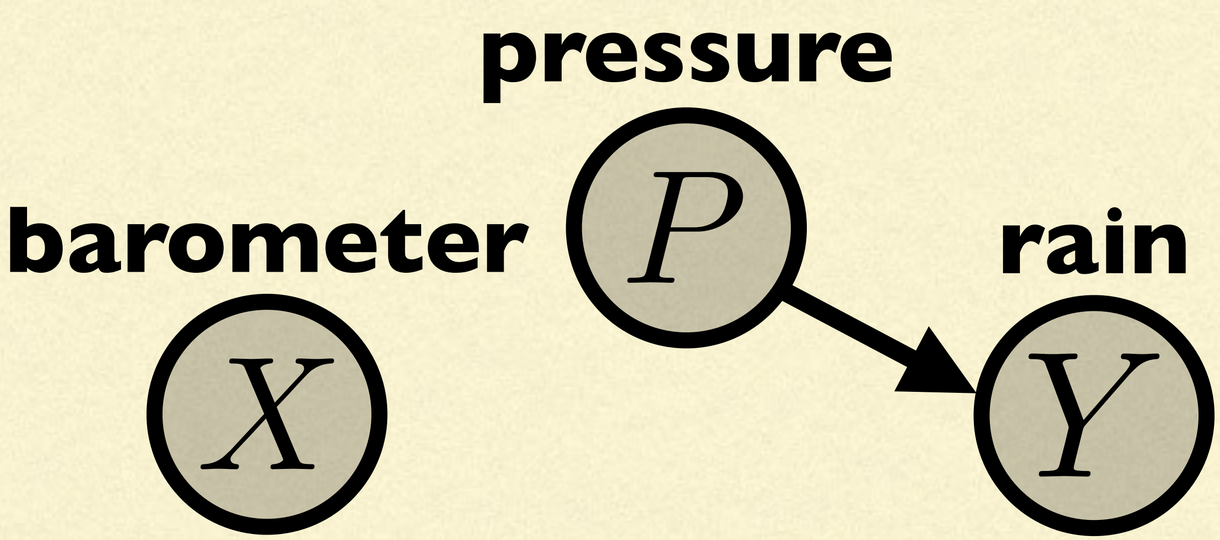


conditioning **intervening**

$$\mathbb{P}(Y \mid X = x) \neq \mathbb{P}(Y \mid X \text{ set to be } x)$$

RANDOMIZING MAKES THEM EQUIVALENT!

Confounding



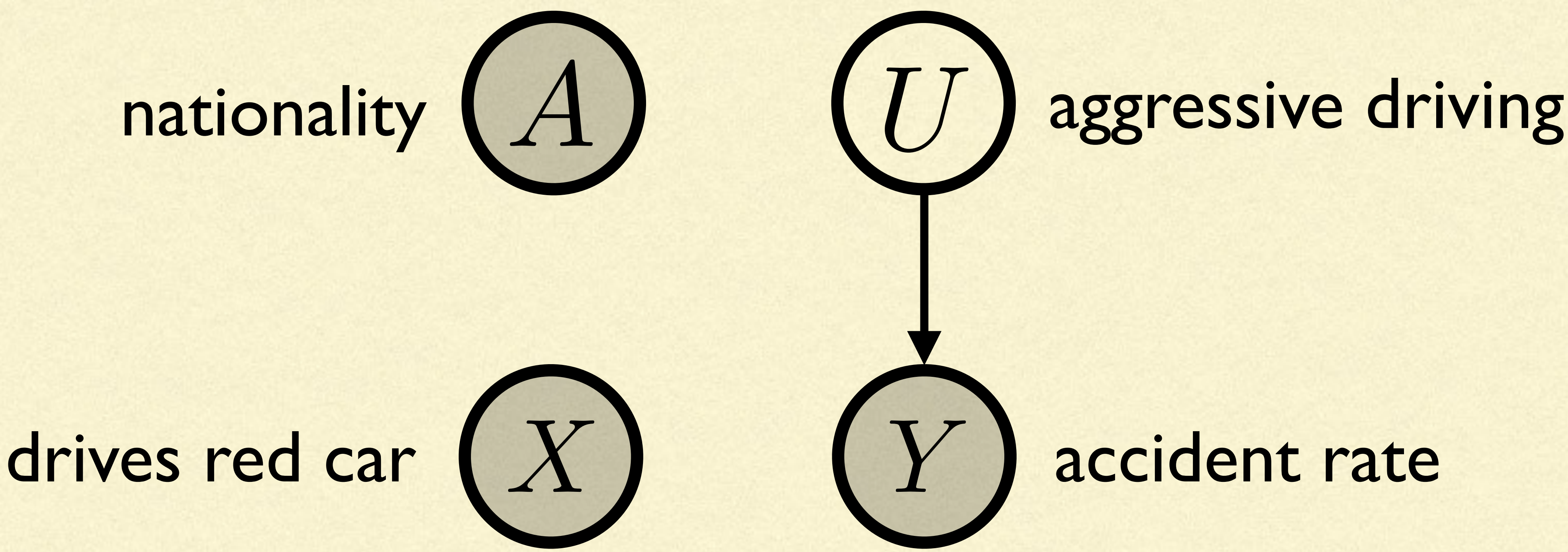
Conditioning **against** the causal direction

conditioning

intervening

$$\mathbb{P}_{\mathrm{m}}(Y \mid X = x) \;=\; \mathbb{P}(Y \mid X \text{ set to be } x)$$

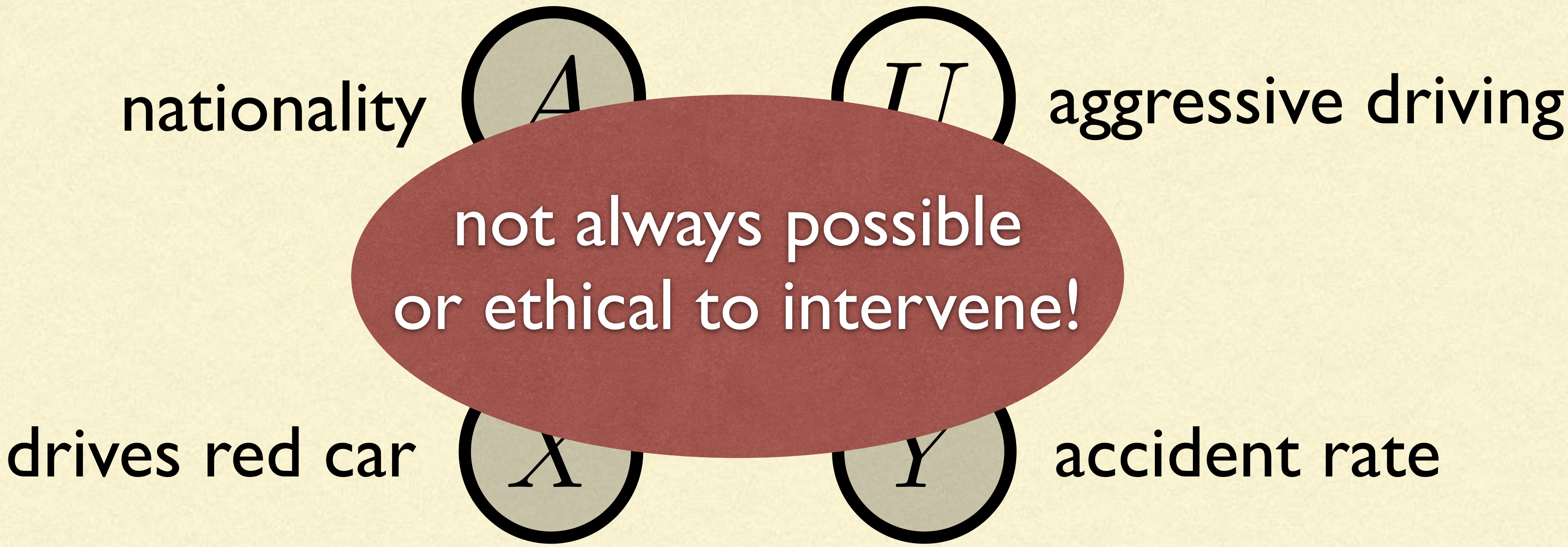
THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



$$\mathbb{P}(Y \mid do(X = x)) = \mathbb{P}_{\mathrm{m}}(Y \mid X = x)$$

modified model

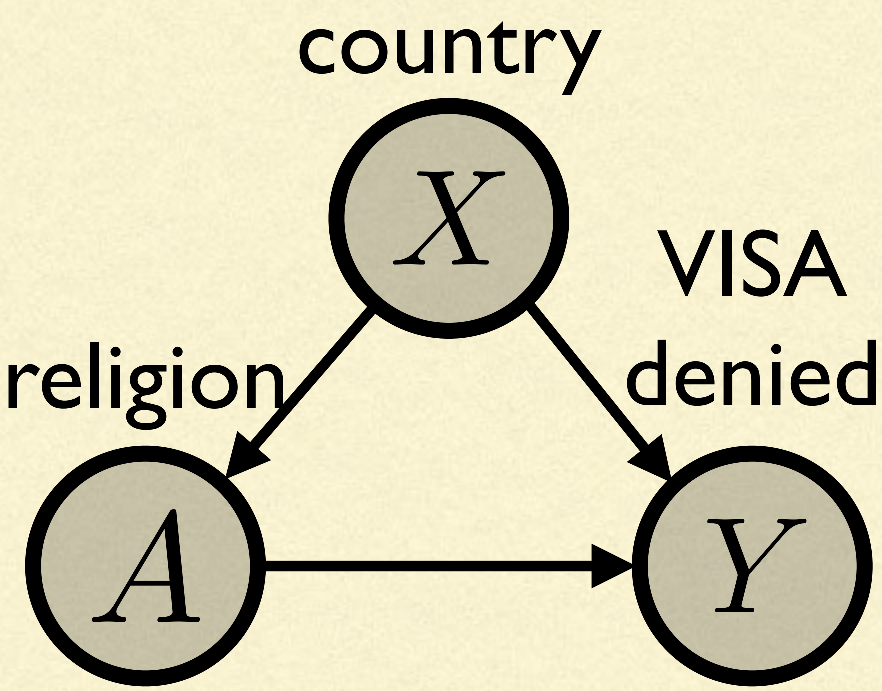
THE GOLD STANDARD: RANDOMIZED CONTROL TRIALS



$$\mathbb{P}(Y \mid do(X = x)) = \mathbb{P}_{\mathrm{m}}(Y \mid X = x)$$

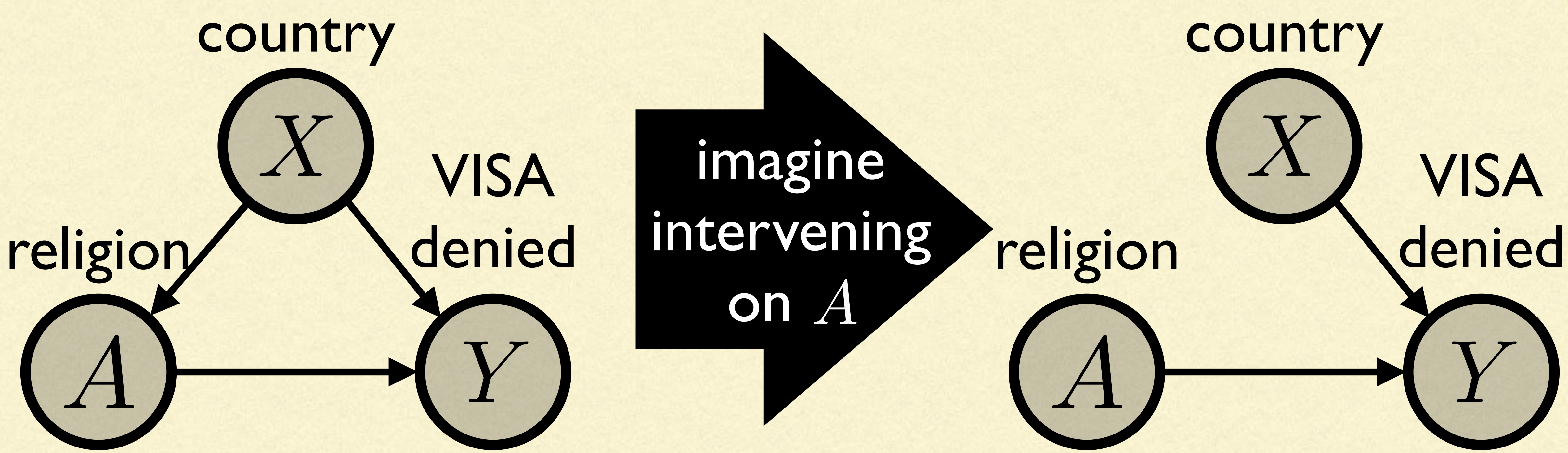
modified
model

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



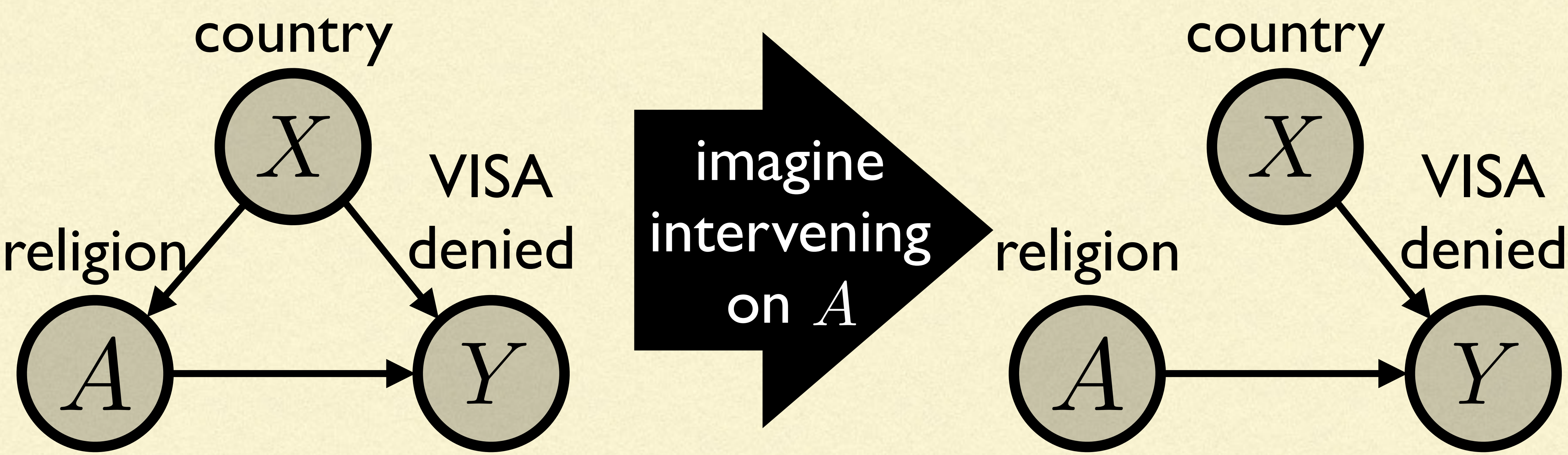
$$\mathbb{P}(Y \mid do(A = a))$$?

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



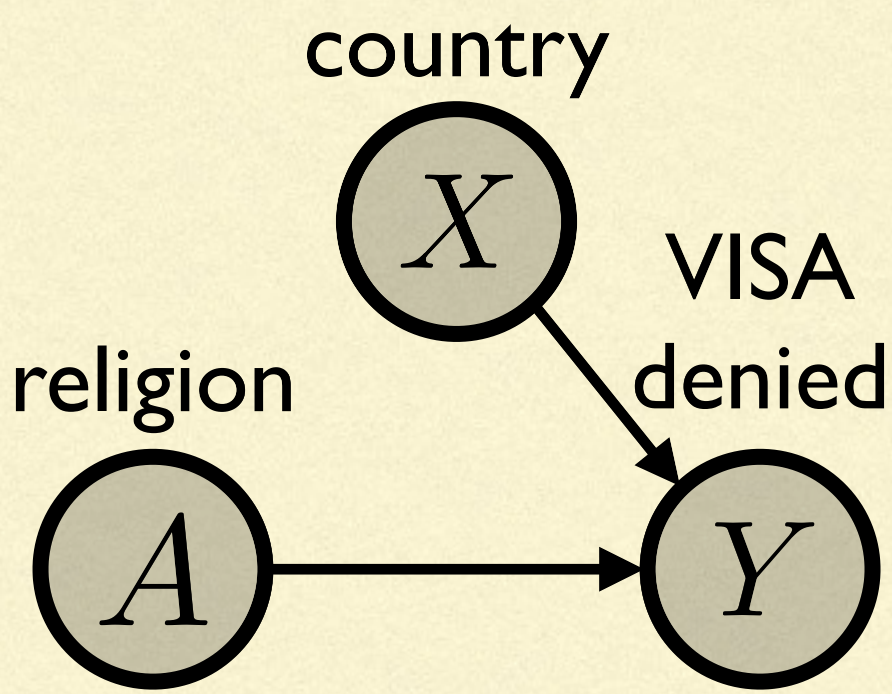
$$\mathbb{P}(Y \mid do(A = a))$$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



$$\mathbb{P}(Y \mid do(A = a)) \quad = \quad \mathbb{P}_{\mathrm{m}}(Y \mid A = a)$$

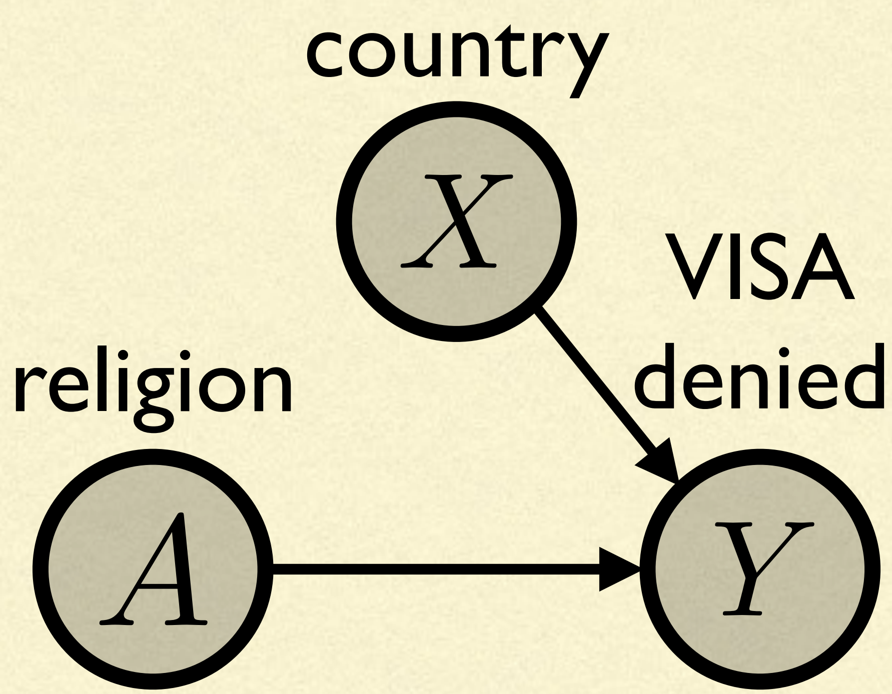
ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



$$\mathbb{P}_{\mathrm{m}}(Y \mid A = a)$$

$$= \sum_{x} \mathbb{P}_{\mathrm{m}}(Y \mid A = a, X = x)\mathbb{P}_{\mathrm{m}}(X = x \mid A = a)$$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH

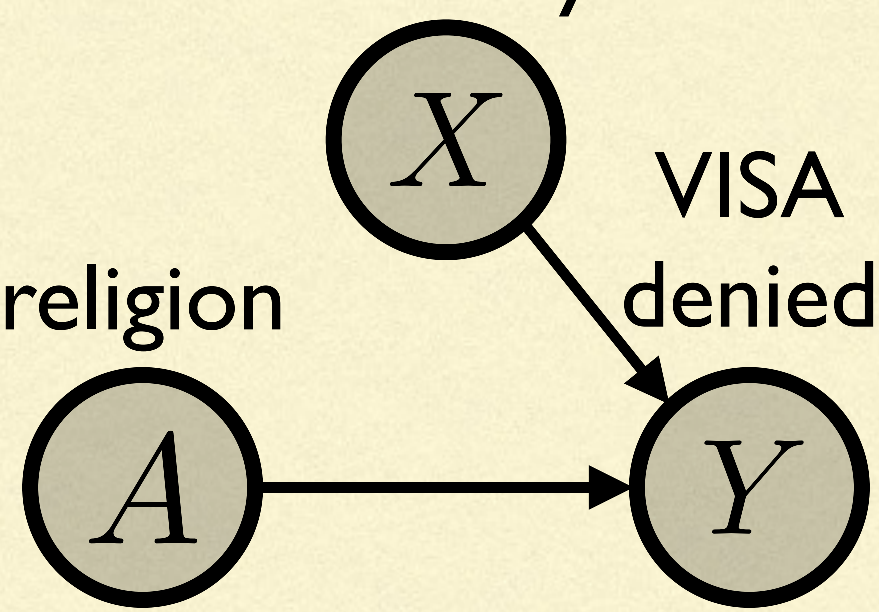


$$\mathbb{P}_{\mathrm{m}}(Y \mid A = a)$$

$$X \perp A$$

$$= \sum_x \mathbb{P}_{\mathrm{m}}(Y \mid A = a, X = x)\mathbb{P}_{\mathrm{m}}(X = x \mid A = a)$$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH

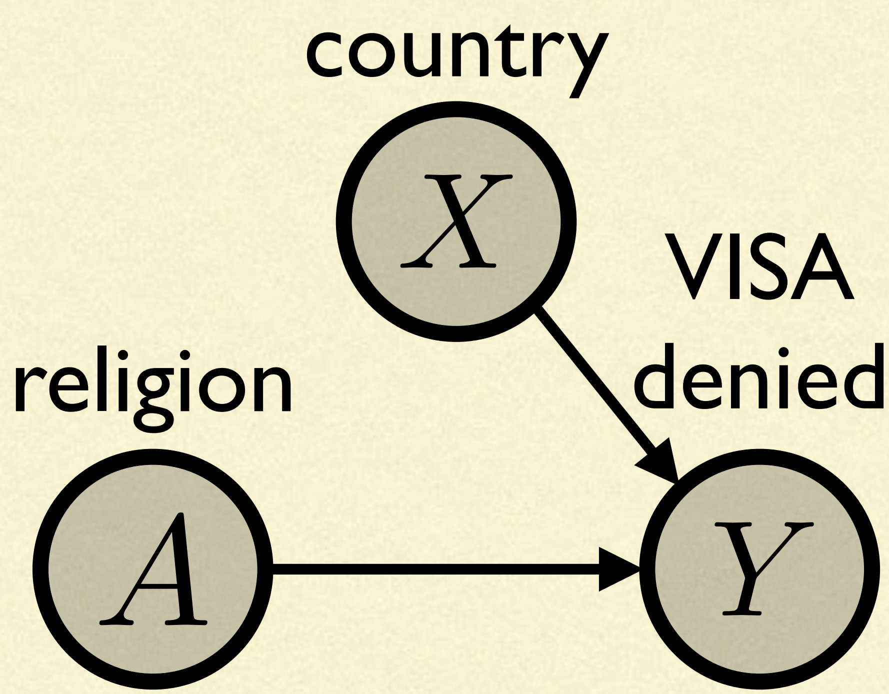


$$\mathbb{P}_{\mathrm{m}}(Y \mid A = a)$$

$$= \sum_x \mathbb{P}_{\mathrm{m}}(Y \mid A = a, X = x)\mathbb{P}_{\mathrm{m}}(X = x)$$

$X \perp\!\!\!\perp A$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



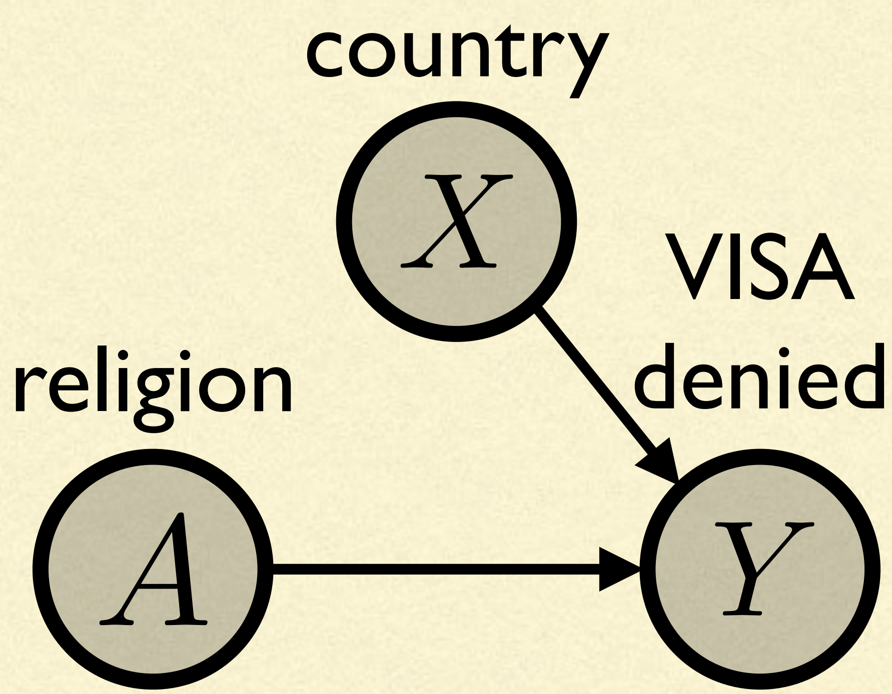
$$= \mathbb{P}(Y \mid A = a, X = x)$$

$$= a)$$

$$= \mathbb{P}(X = x)$$

$$= \sum_x \mathbb{P}_{\mathrm{m}}(Y \mid A = a, X = x)\mathbb{P}_{\mathrm{m}}(X = x)$$

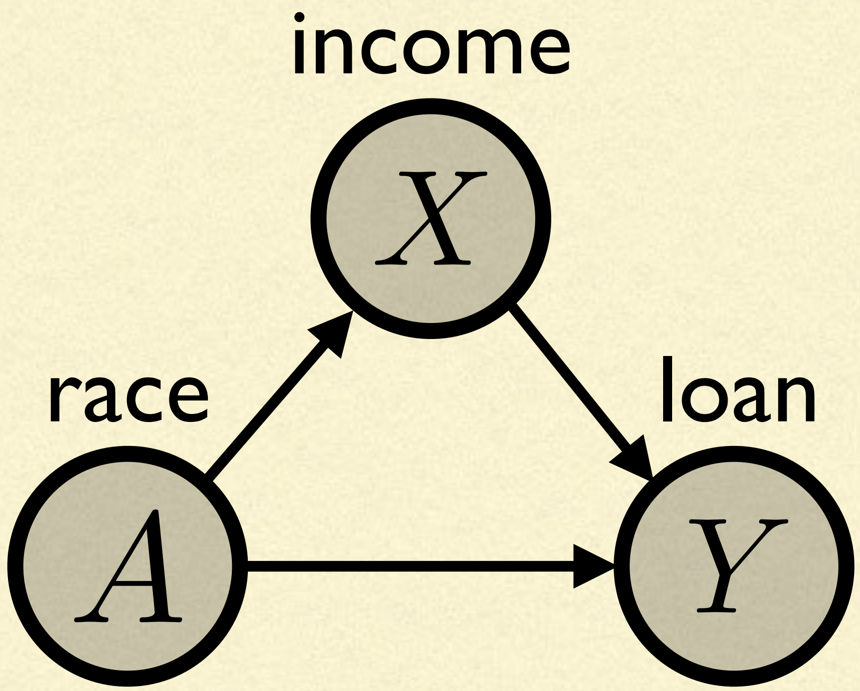
ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



$$\mathbb{P}_{\mathrm{m}}(Y \mid A = a)$$

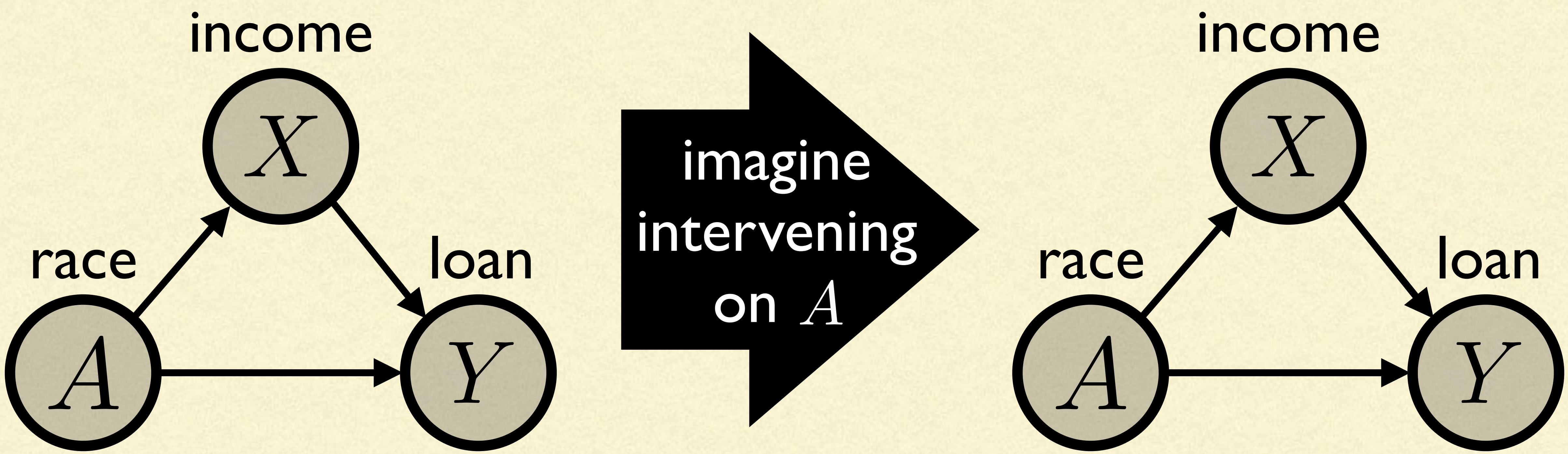
$$= \sum_x \mathbb{P}(Y \mid A = a, X = x)\mathbb{P}(X = x)$$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



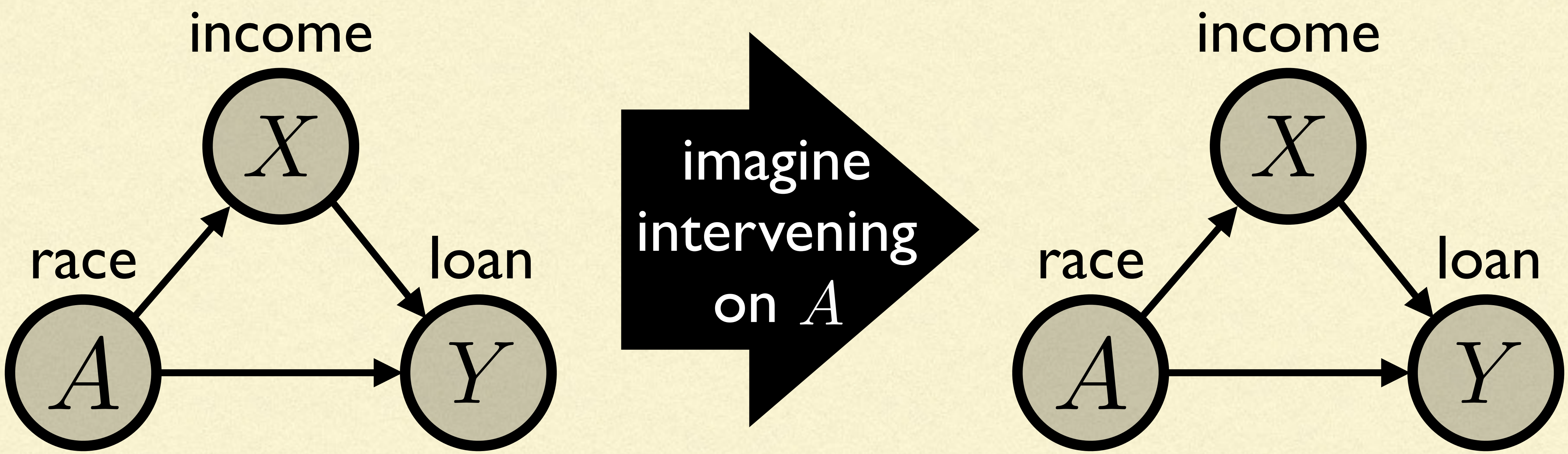
$\mathbb{P}(Y \mid do(A = a))$?

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



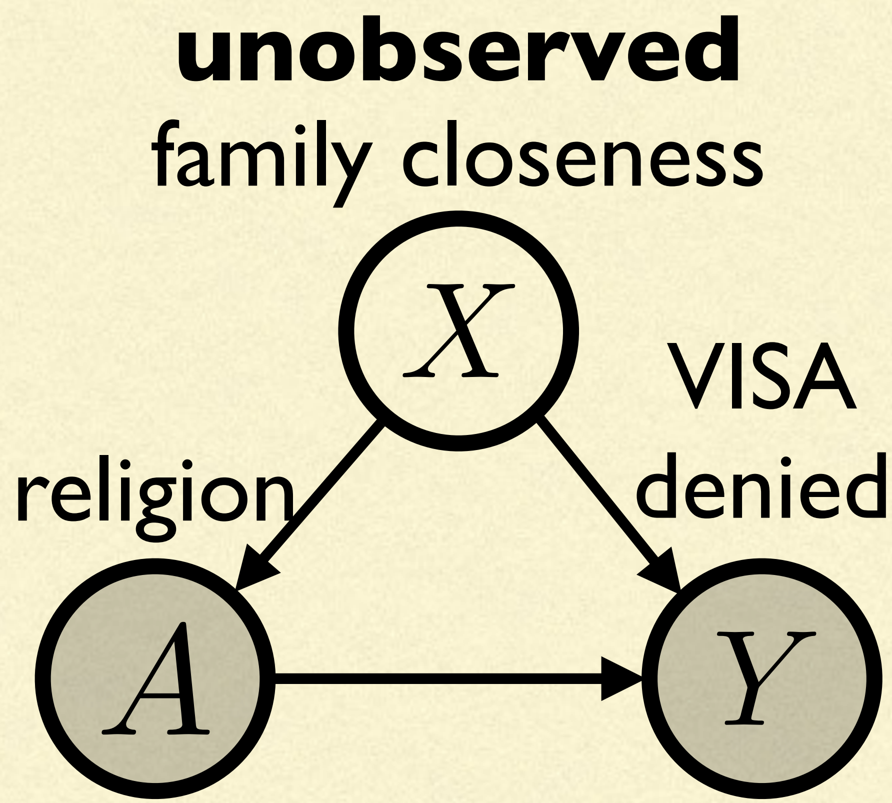
$\mathbb{P}(Y \mid do(A = a))$?

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



$$\mathbb{P}(Y \mid do(A = a)) \quad = \quad \mathbb{P}(Y \mid A = a)$$

ESTIMATING CAUSALITY WITH OBSERVATIONS & A GRAPH



$$\mathbb{P}(Y \mid do(A = a))$$?

1. WHAT IS CAUSALITY?

2. HOW DO WE ESTIMATE **CAUSAL EFFECTS (INTERVENTIONS)**?

3. HOW DO WE OBTAIN A **CAUSAL GRAPH**?

4. ANOTHER CAUSAL QUANTITY: **COUNTERFACTUALS**

HOW DO WE GET A GRAPH?

Way 1: Conditional Independence Testing

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

HOW DO WE GET A GRAPH? (WAY 1)

1. Markov assumption

[Spirtes et al., 2000; Pearl, 2000]

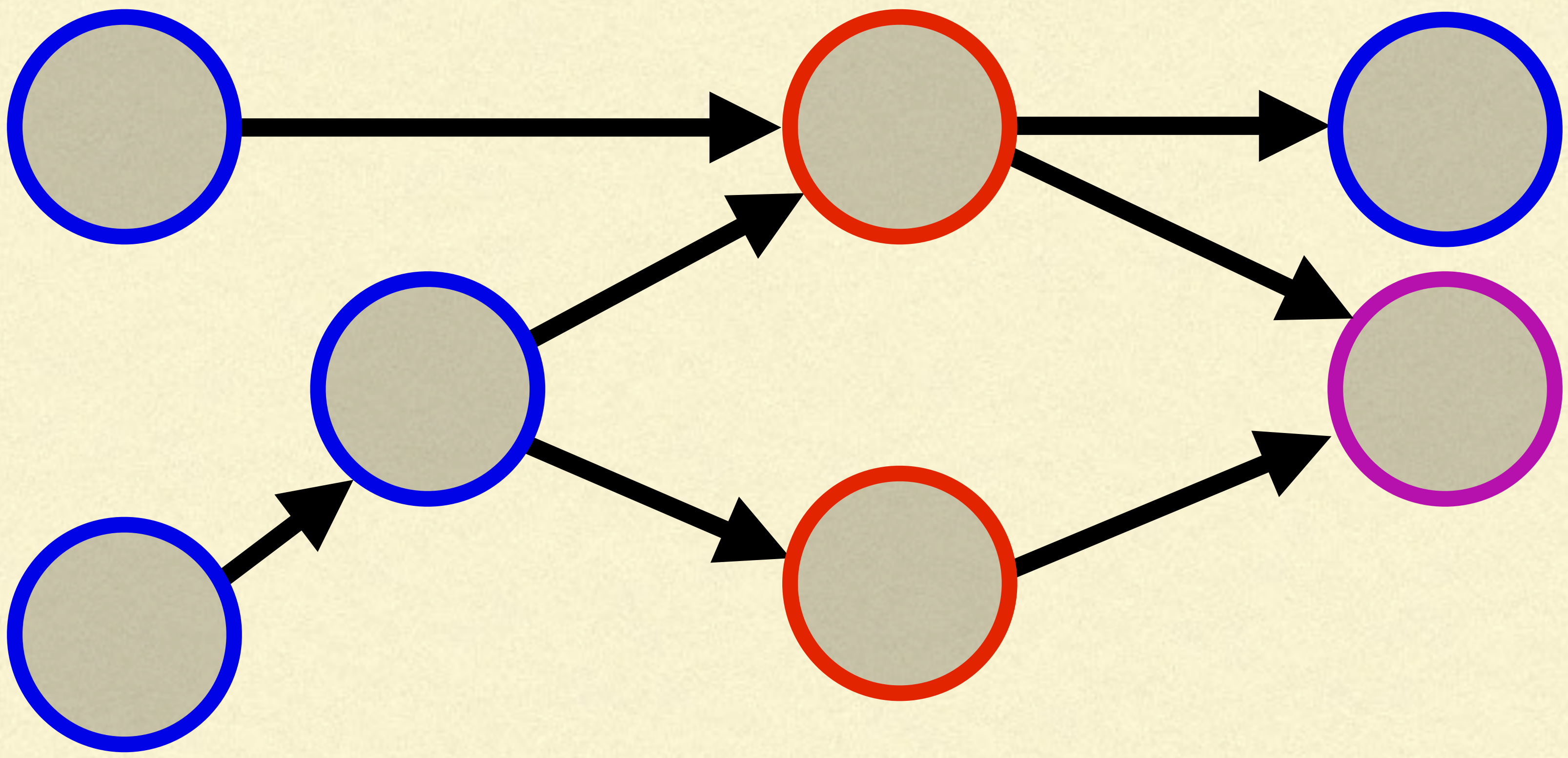
a **variable** in a causal graph is **independent of its ancestors** given its **direct parents**

HOW DO WE GET A GRAPH? (WAY 1)

1. Markov assumption

[Spirtes et al., 2000; Pearl, 2000]

a **variable** in a causal graph is **independent of its non-descendants** given its **direct parents**

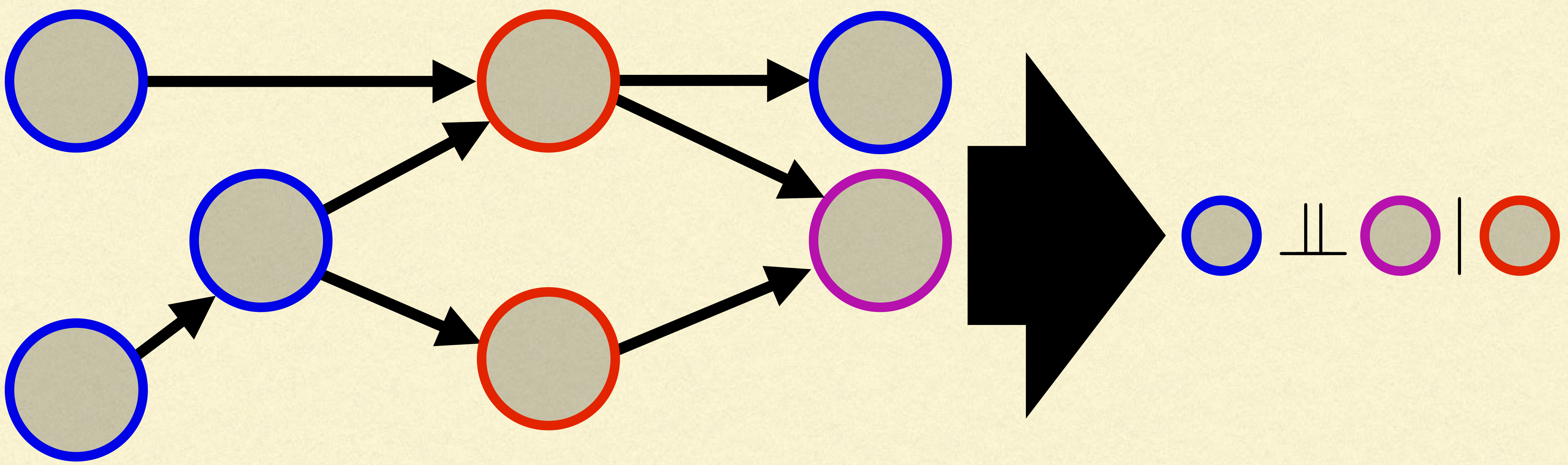


HOW DO WE GET A GRAPH? (WAY 1)

1. Markov assumption

[Spirtes et al., 2000; Pearl, 2000]

a **variable** in a causal graph is **independent of its non-descendants** given its **direct parents**



HOW DO WE GET A GRAPH? (WAY 1)

2. Faithfulness assumption

[Spirtes et al., 2000; Pearl, 2000]

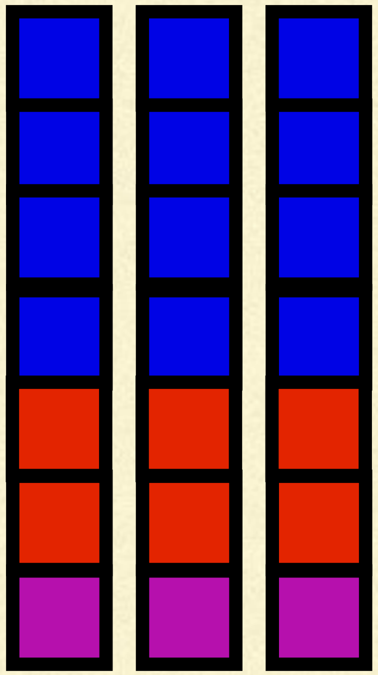
any independences in the data exist **if and only if** they exist **in the causal model**

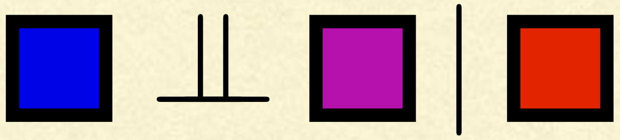
HOW DO WE GET A GRAPH? (WAY 1)

2. Faithfulness assumption

[Spirtes et al., 2000; Pearl, 2000]

any independences in the data exist **if and only if** they exist **in the causal model**



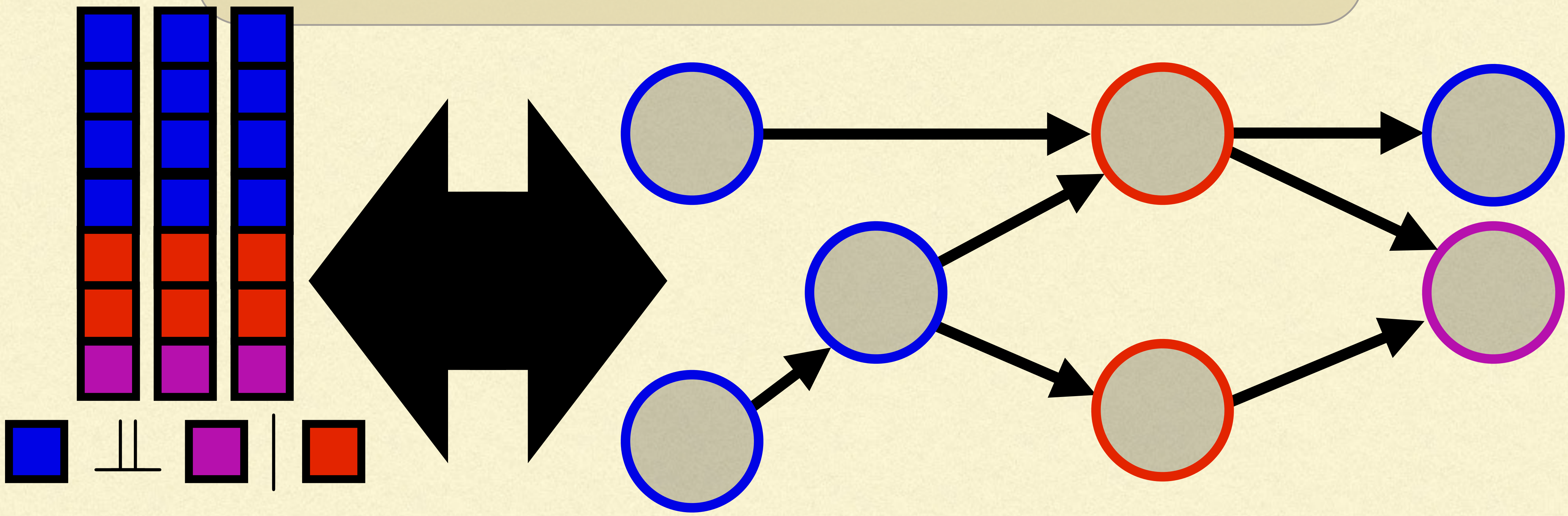


HOW DO WE GET A GRAPH? (WAY 1)

2. Faithfulness assumption

[Spirtes et al., 2000; Pearl, 2000]

any independences in the data exist **if and only if** they exist **in the causal model**



D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent

D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent

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

D-SEPARATION

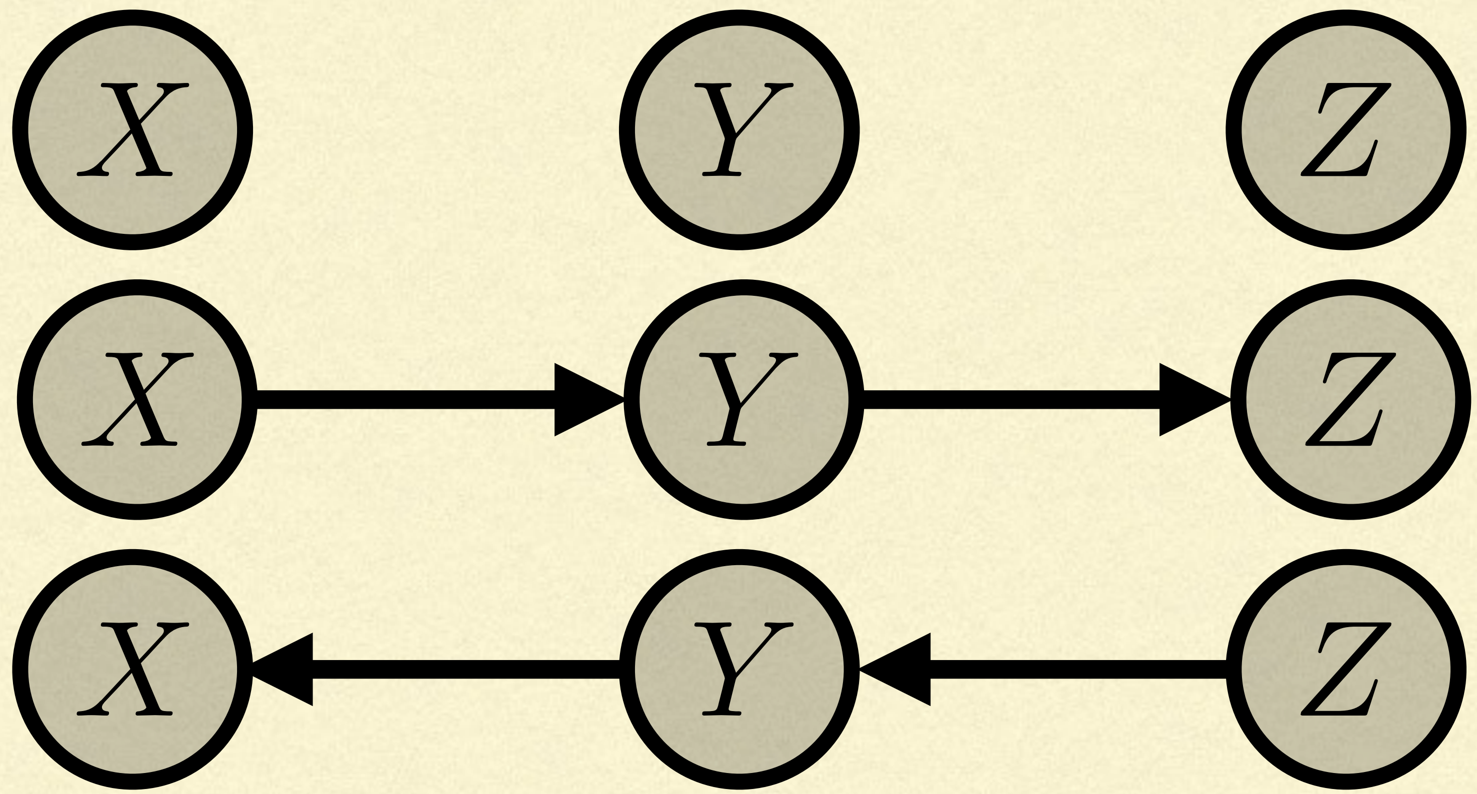
[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent

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

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

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



D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent

$X \perp\!\!\!\perp Z \mid Y$ $\qquad X \quad Y \quad Z$

$X \perp\!\!\!\perp Z \mid Y$ $\qquad X \rightarrow Y \rightarrow Z$

$X \perp\!\!\!\perp Z \mid Y$ $\qquad X \leftarrow Y \leftarrow Z$

$X \perp\!\!\!\perp Z \mid Y$ $\qquad X \leftarrow Y \rightarrow Z$

D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes for any pair of variables, whether they are (conditionally) independent

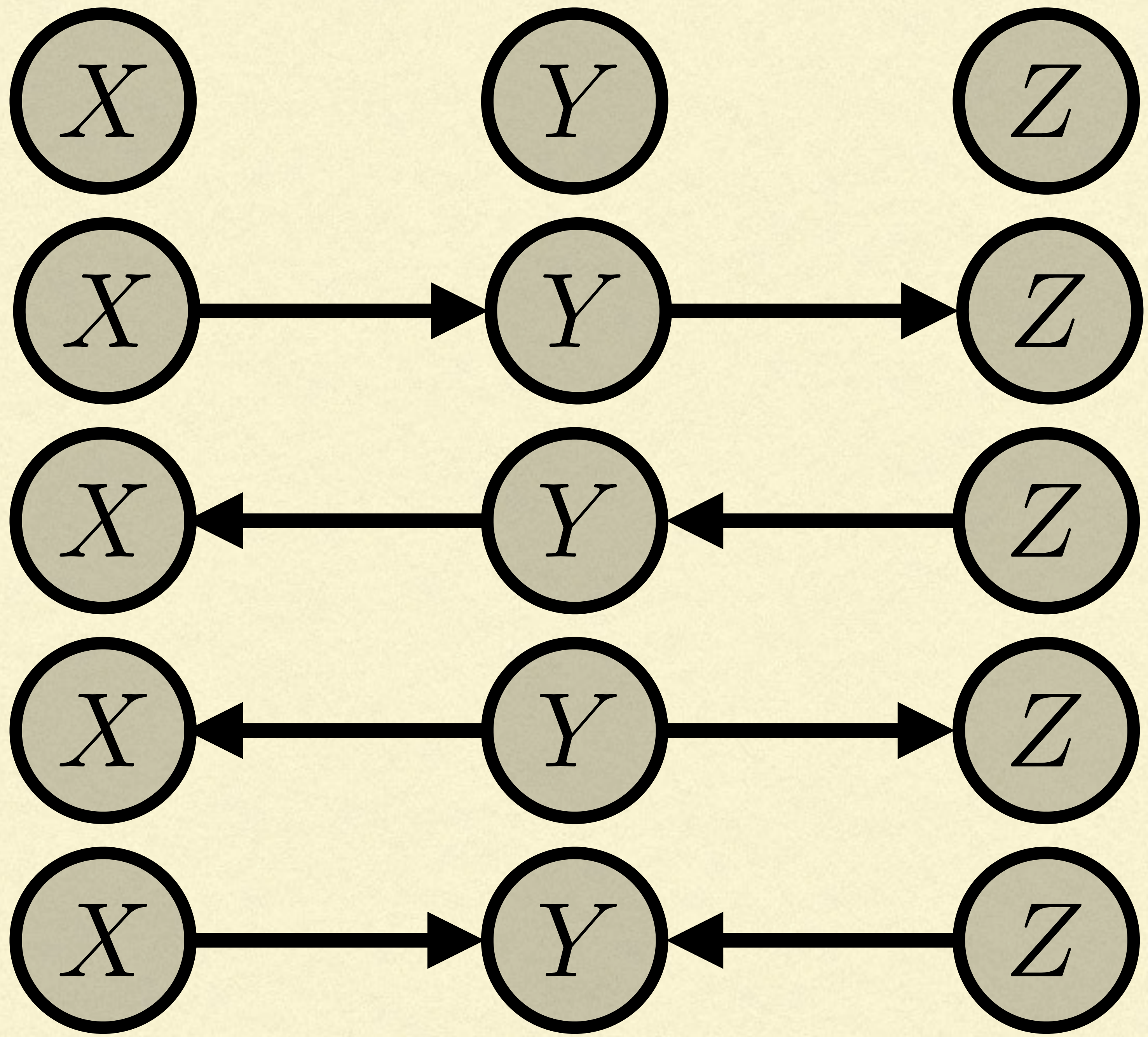
$$X \perp\!\!\!\perp Z \mid Y$$

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

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

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

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

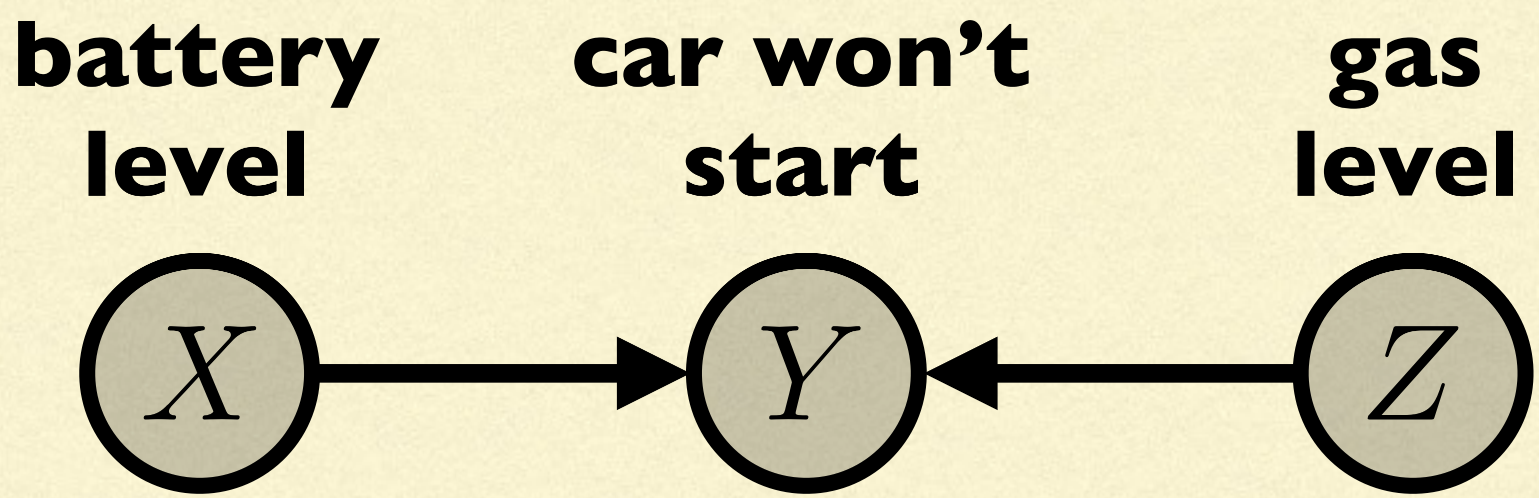


D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes for any pair of variables, whether they are (conditionally) independent

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

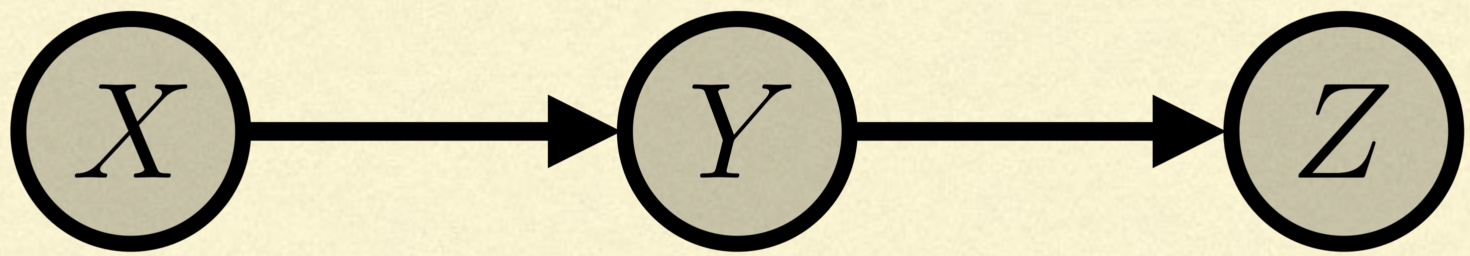


D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent

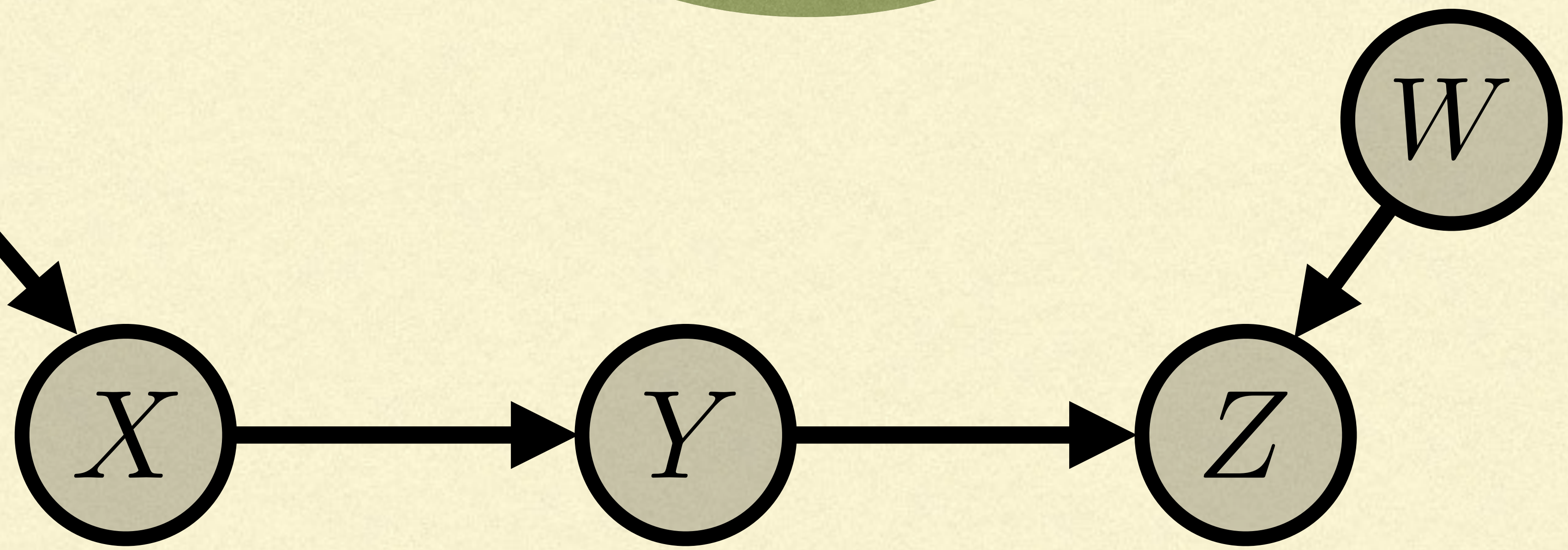
$$X \perp\!\!\!\perp Z \mid Y$$



D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes
for any pair of
variables, whether they
are (conditionally)
independent



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

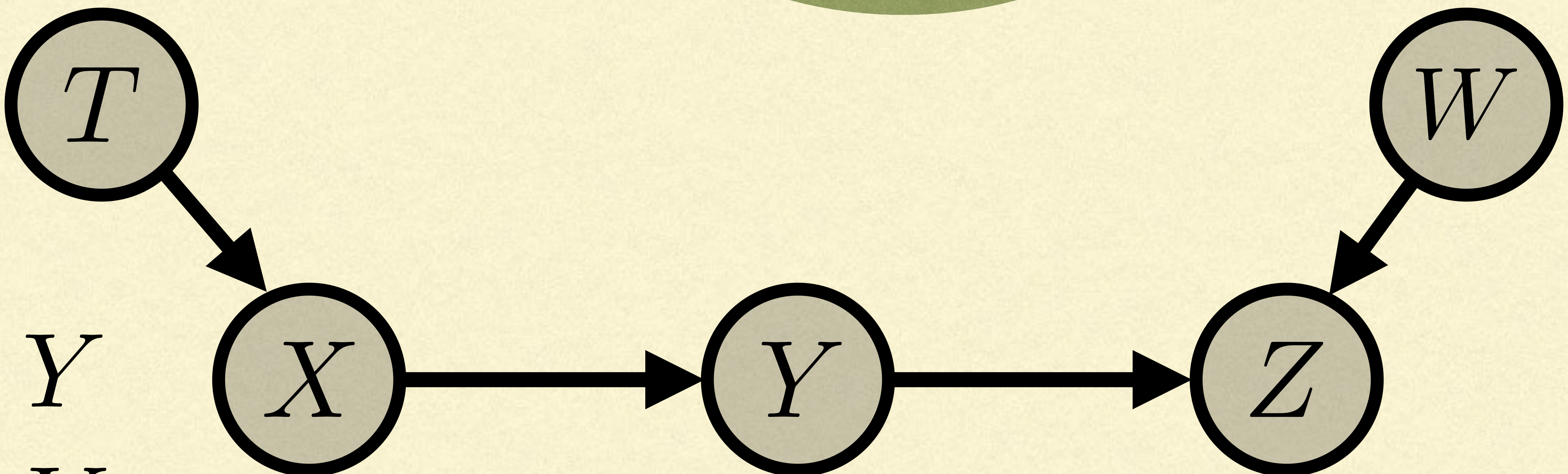
D-SEPARATION

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

describes for any pair of variables, whether they are (conditionally) independent

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

$T \perp\!\!\!\perp W \mid Y$

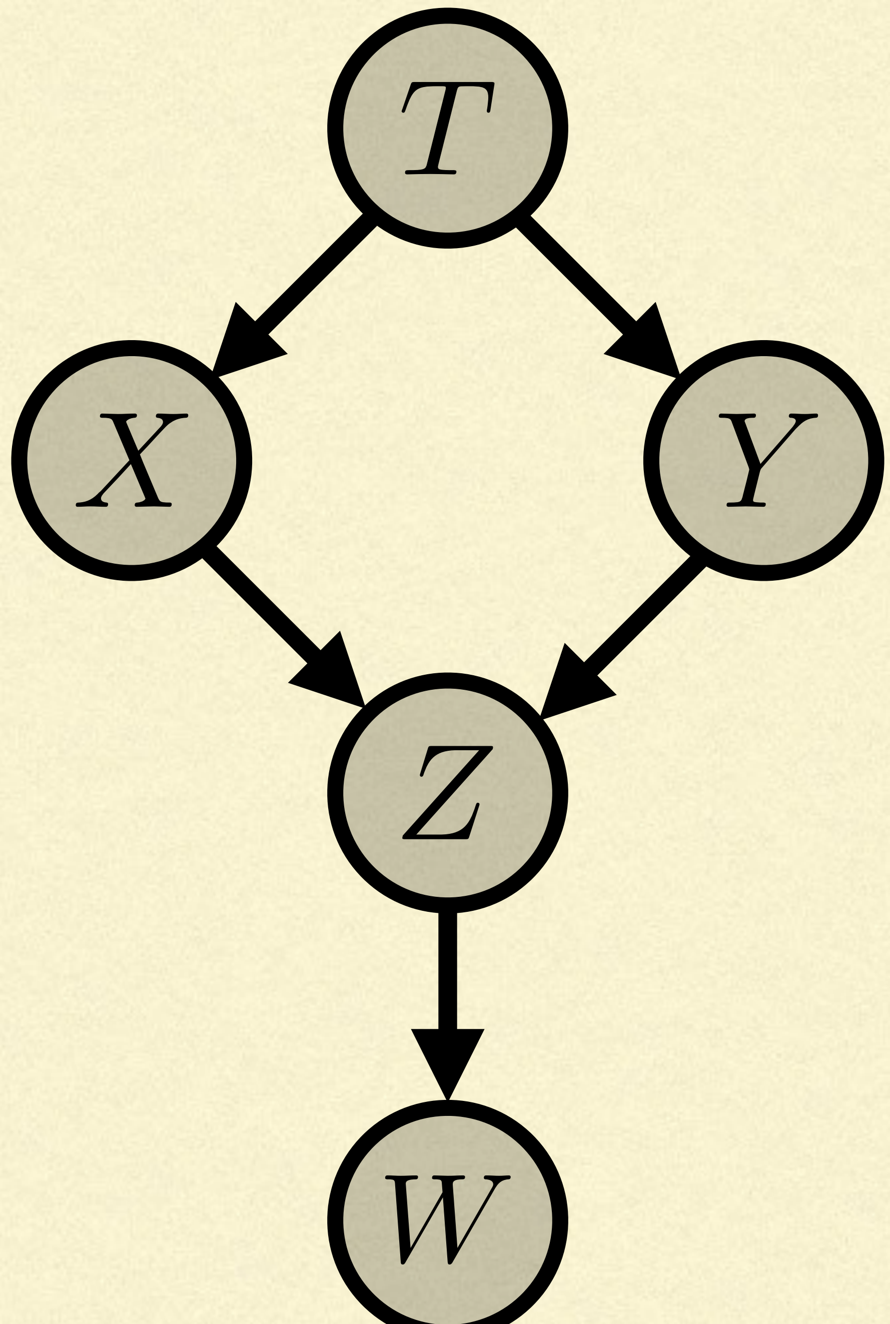


Y "blocks" these paths!

PC ALGORITHM

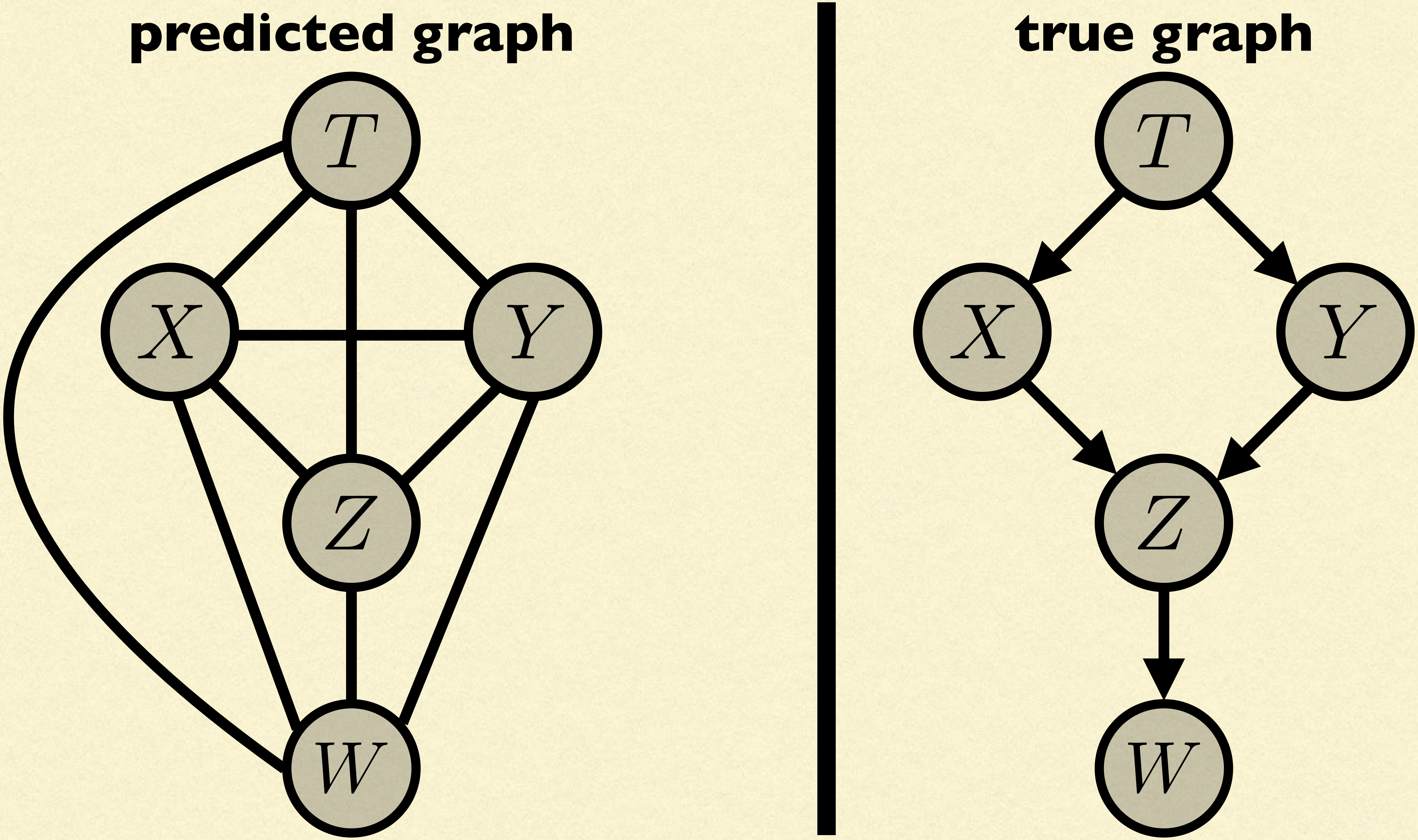
[Spirtes et al. 2001]

true graph



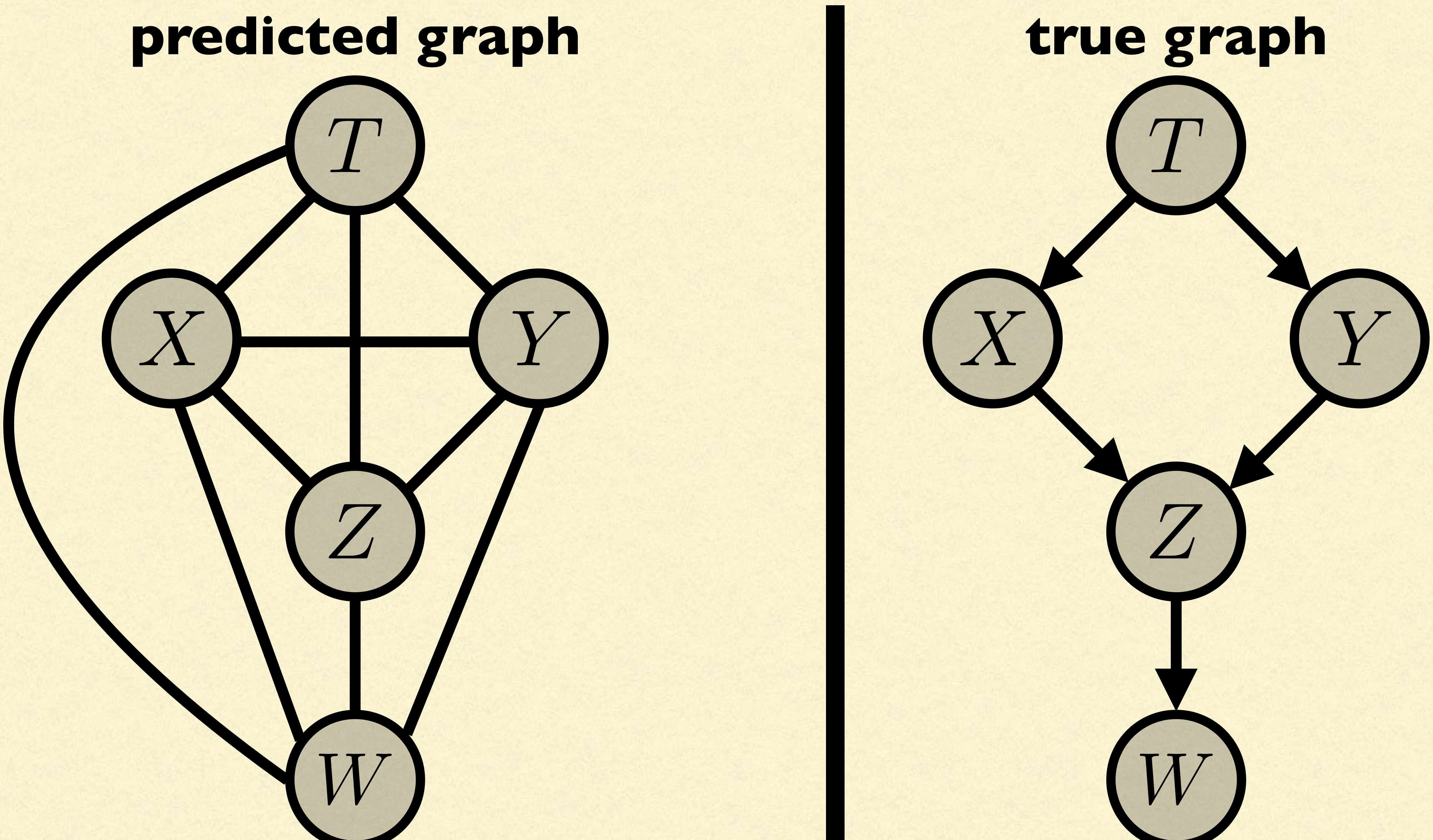
PC ALGORITHM

[Spirtes et al. 2001]



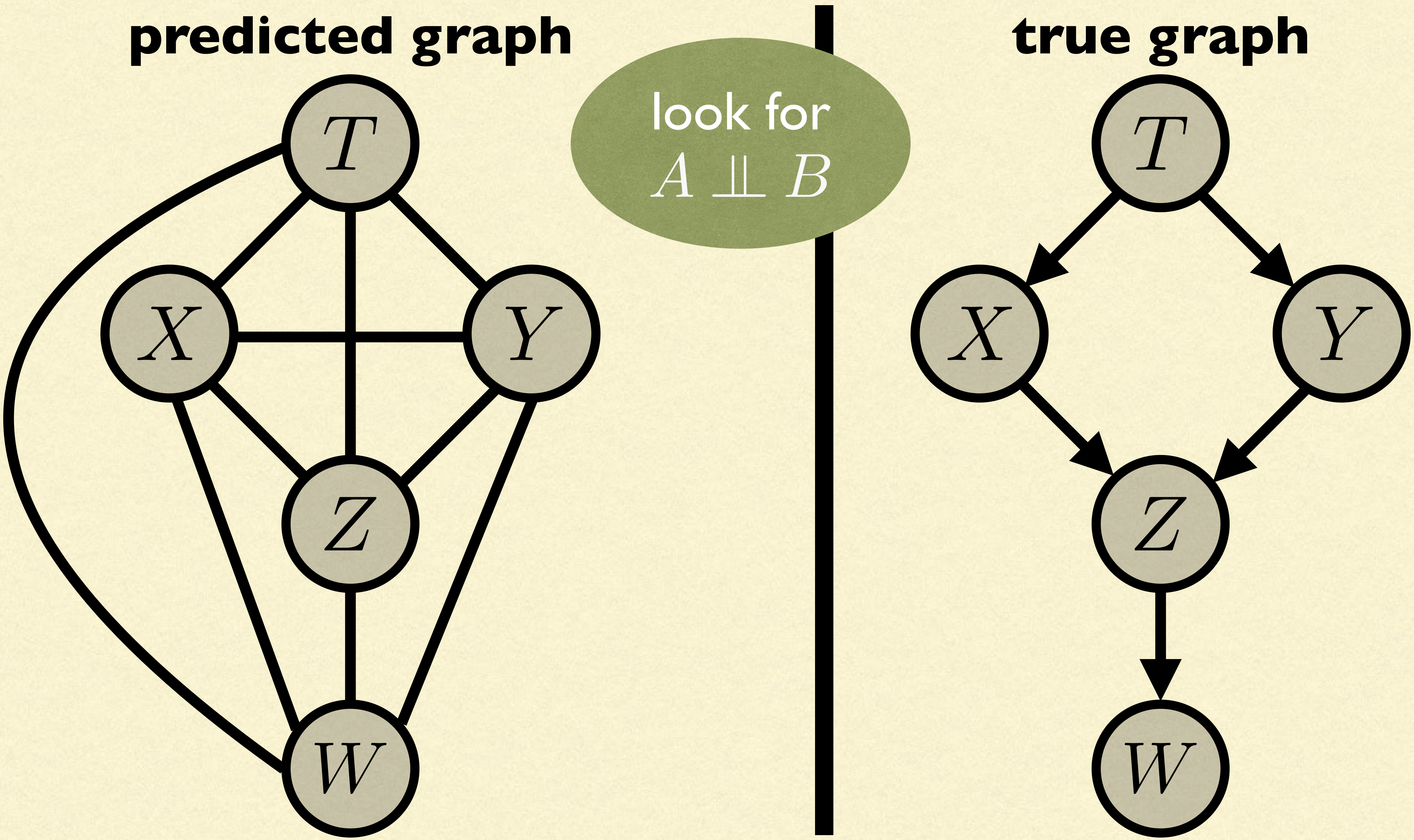
STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]



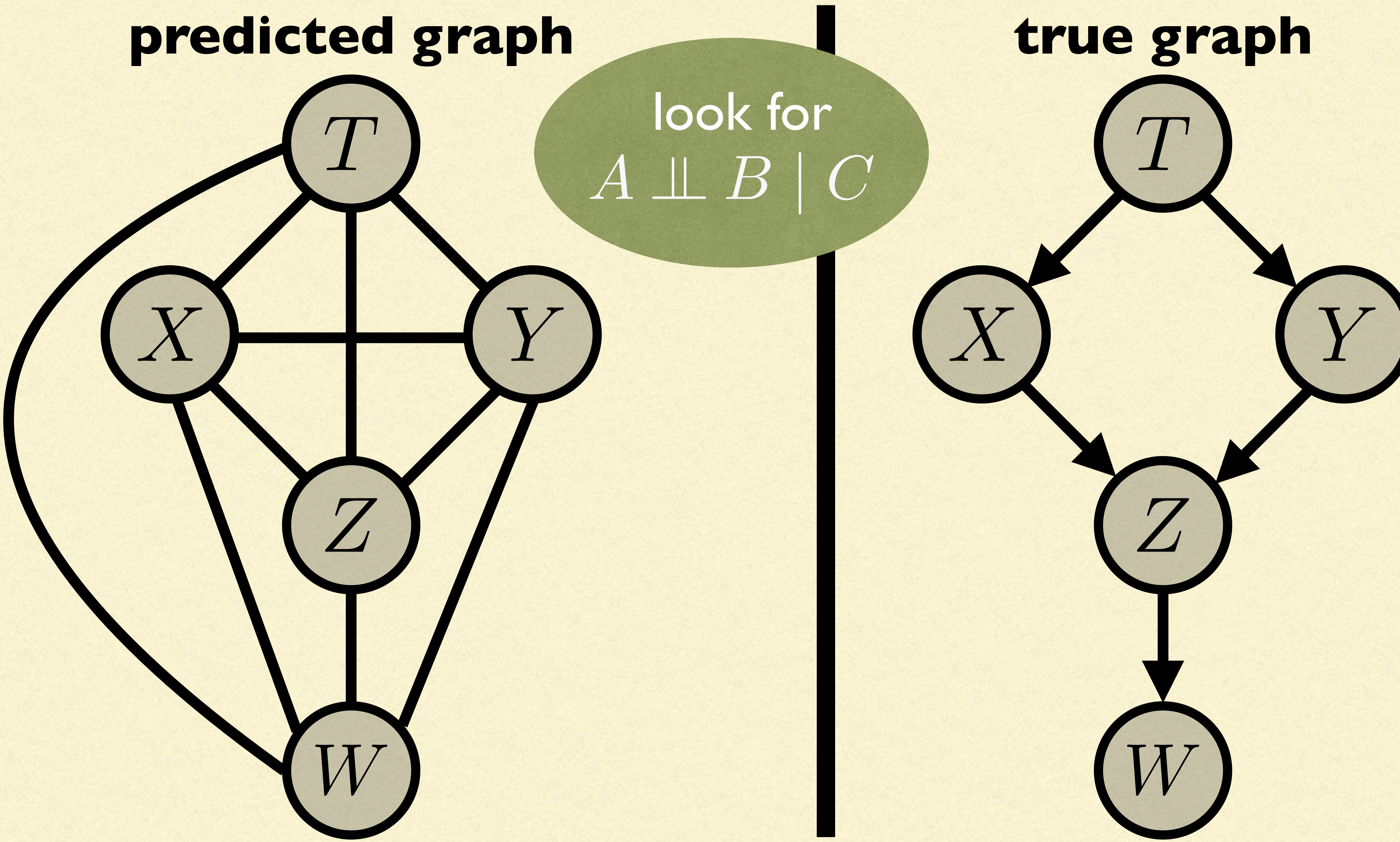
STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]



STEP 1: FIND ADJACENCIES

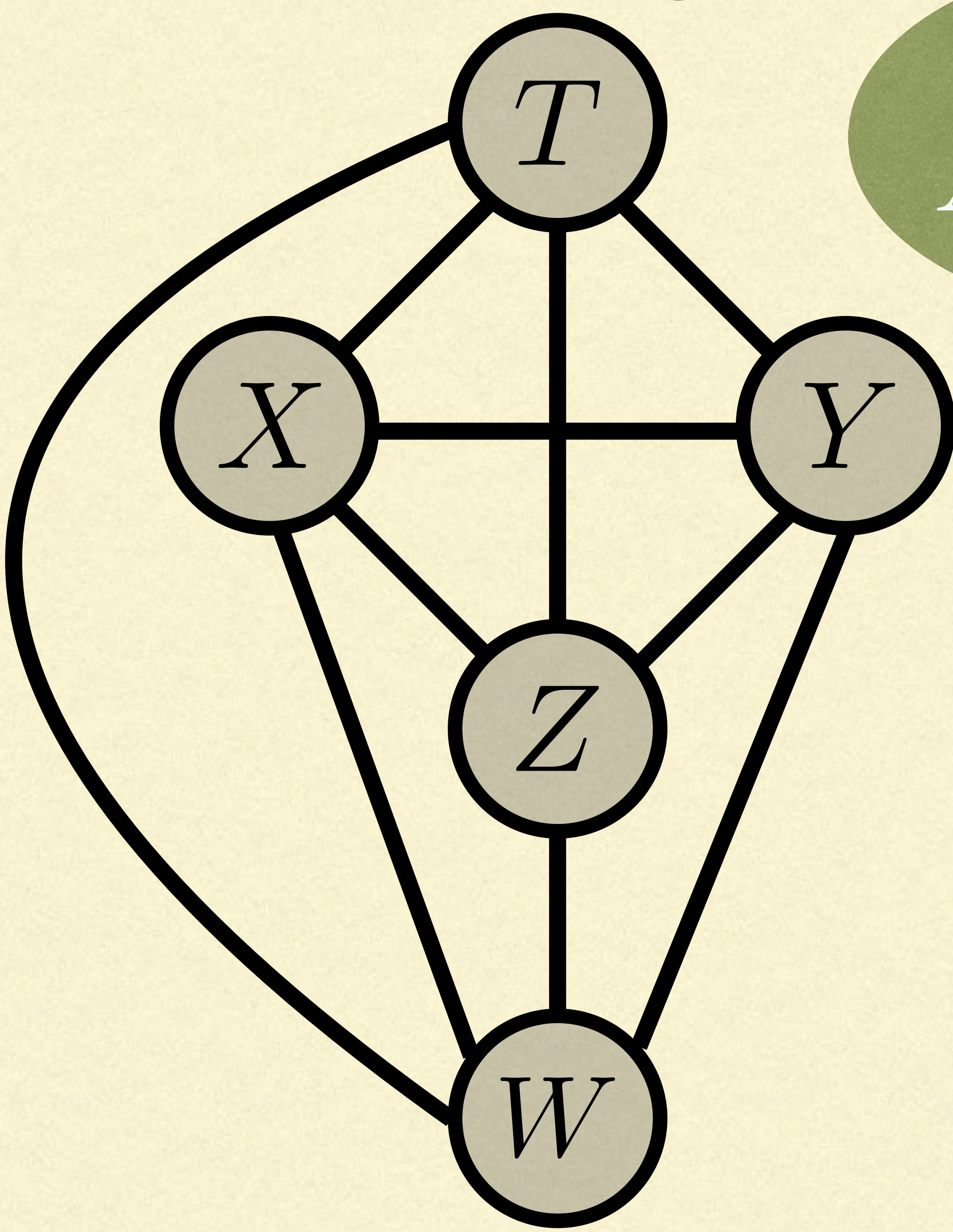
[Spirtes et al. 2001]



STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]

predicted graph

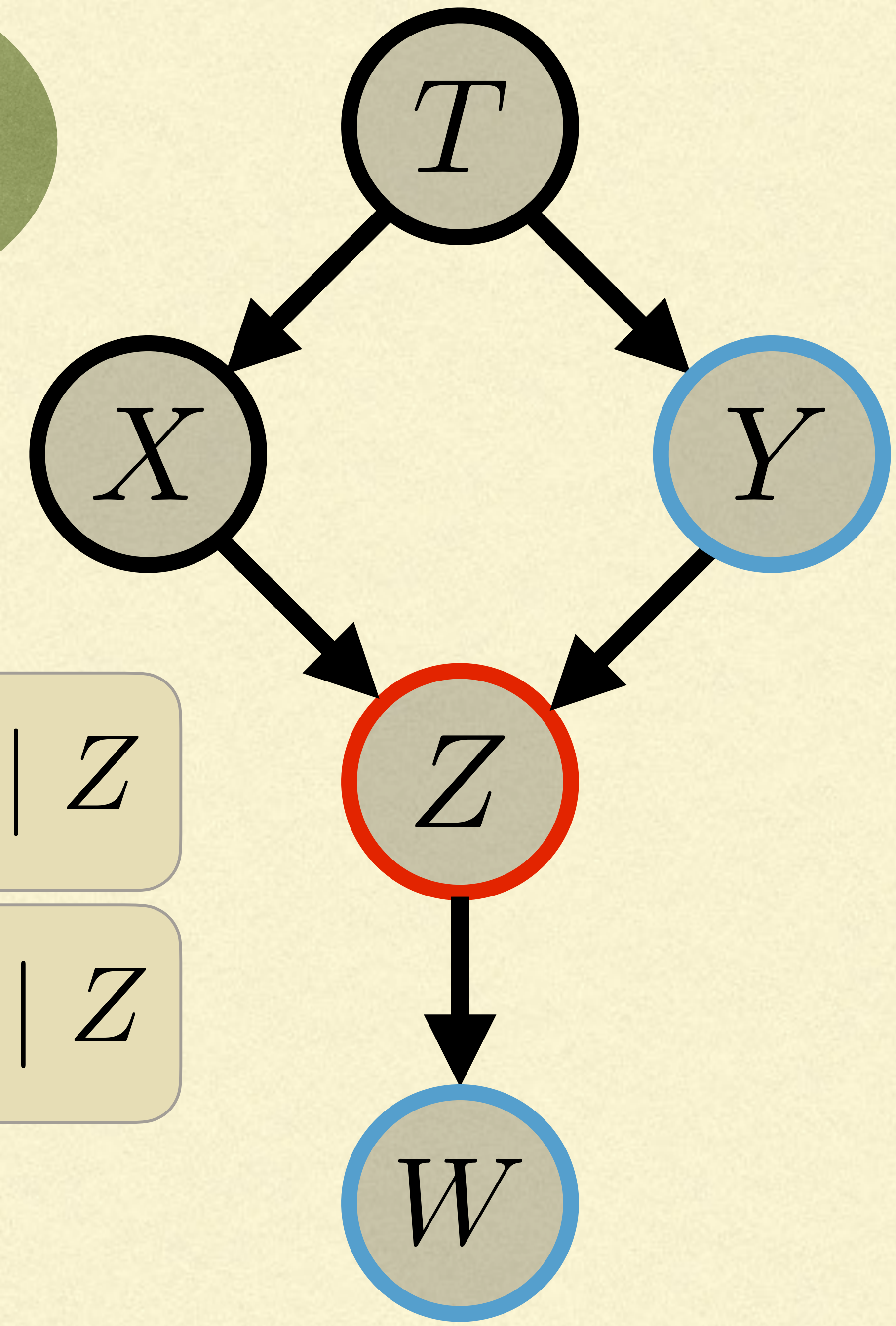


look for $A \perp\!\!\!\perp B \mid C$

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

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

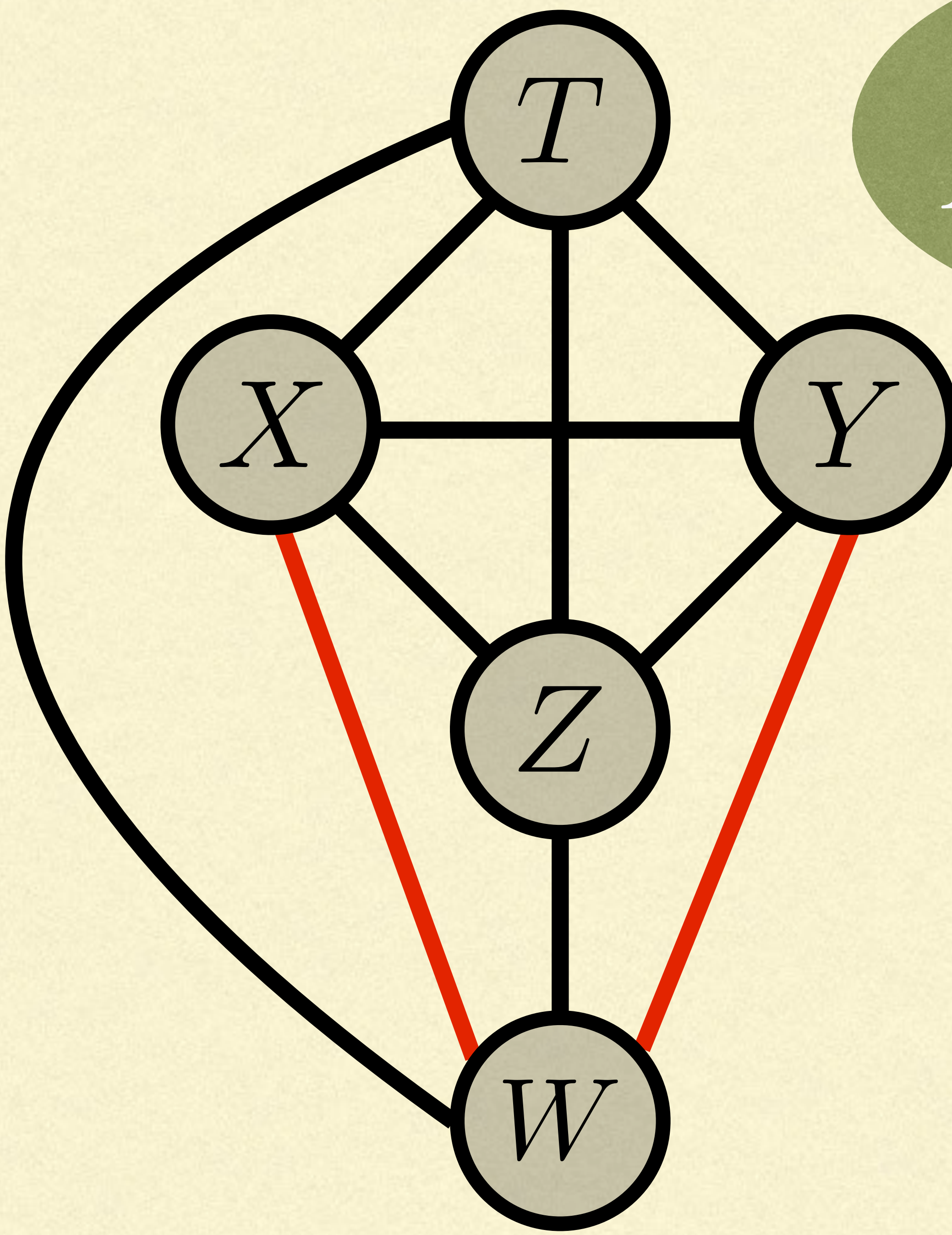
true graph



STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]

predicted graph

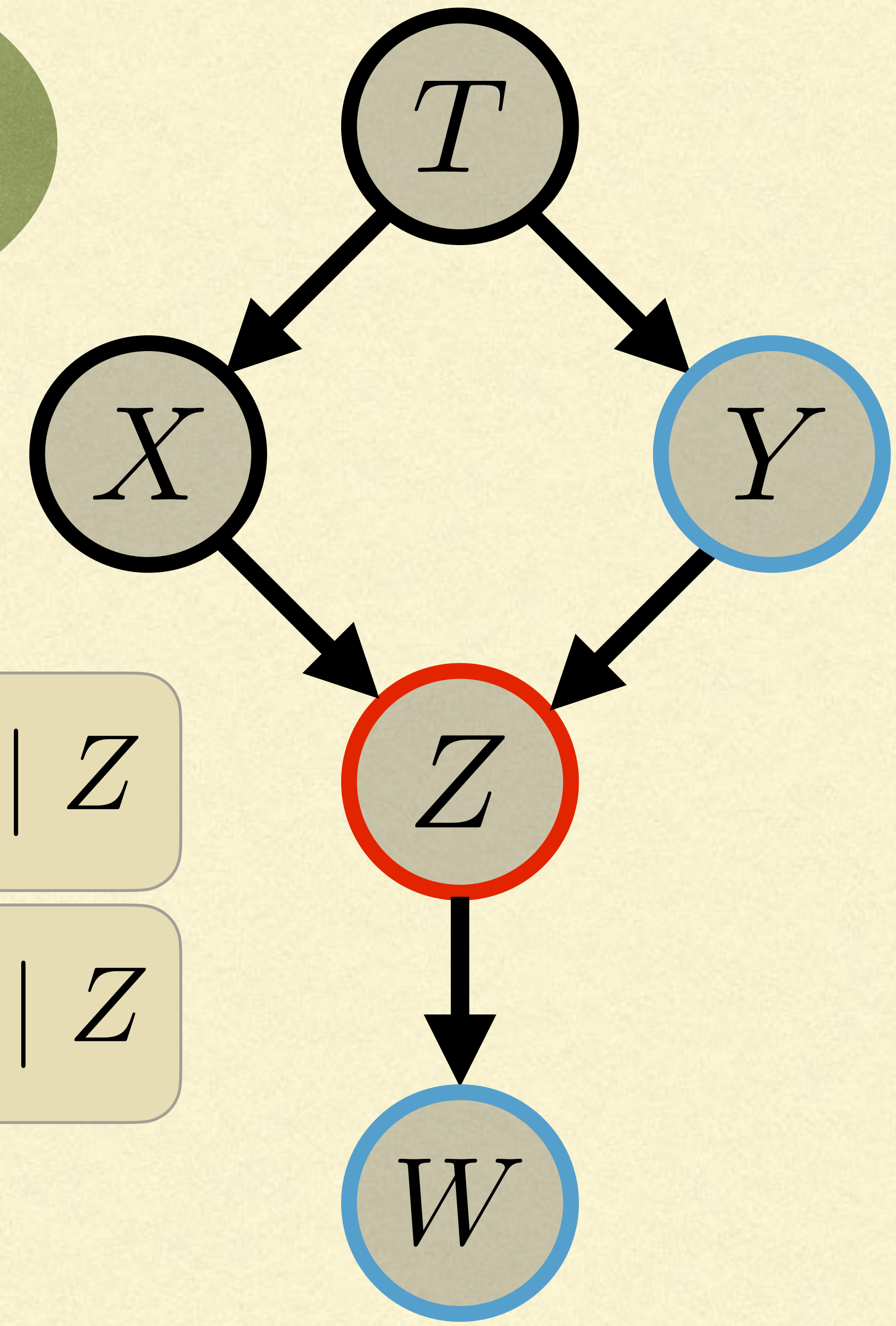


look for
$A \perp\!\!\!\perp B \mid C$

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

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

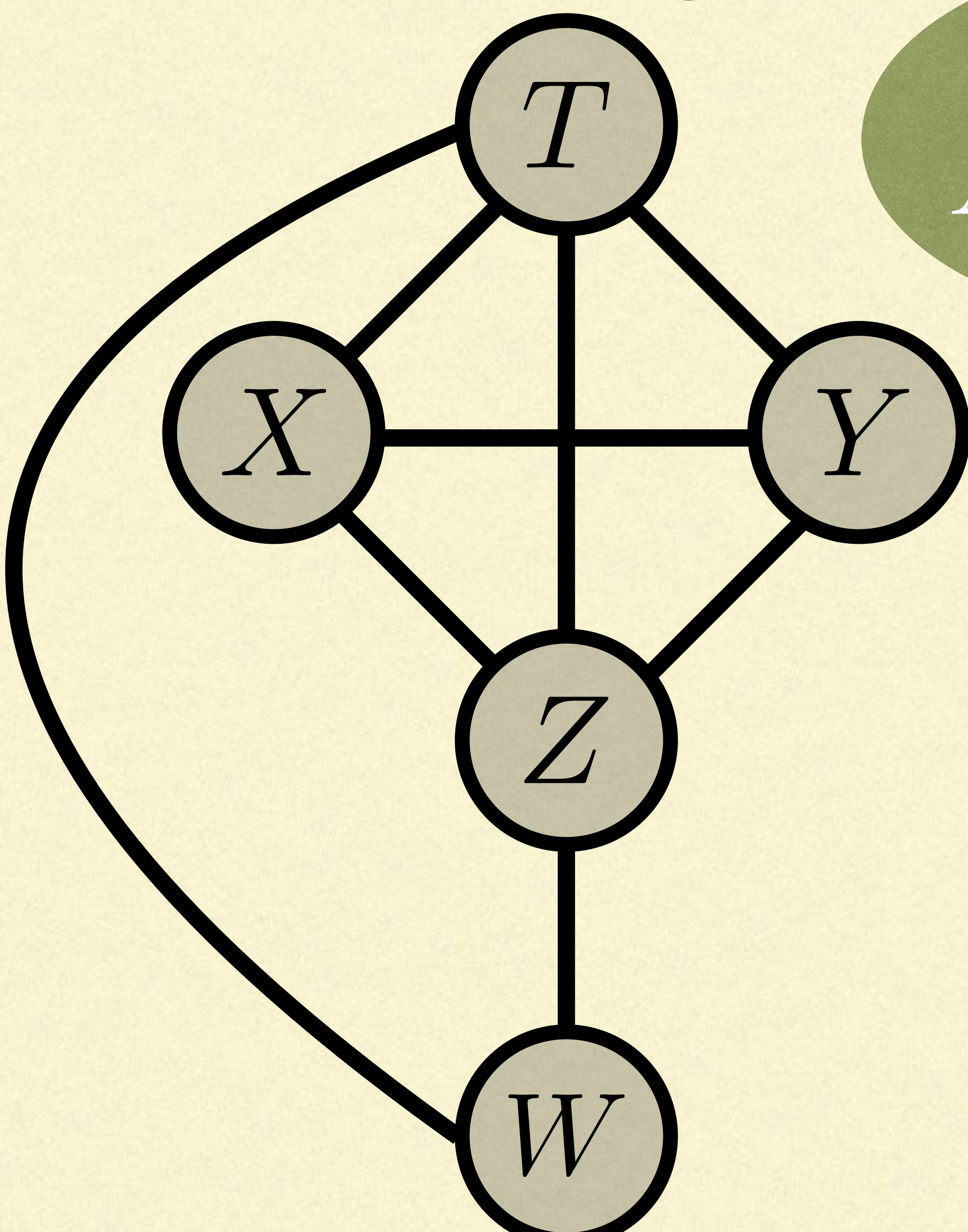
true graph



STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]

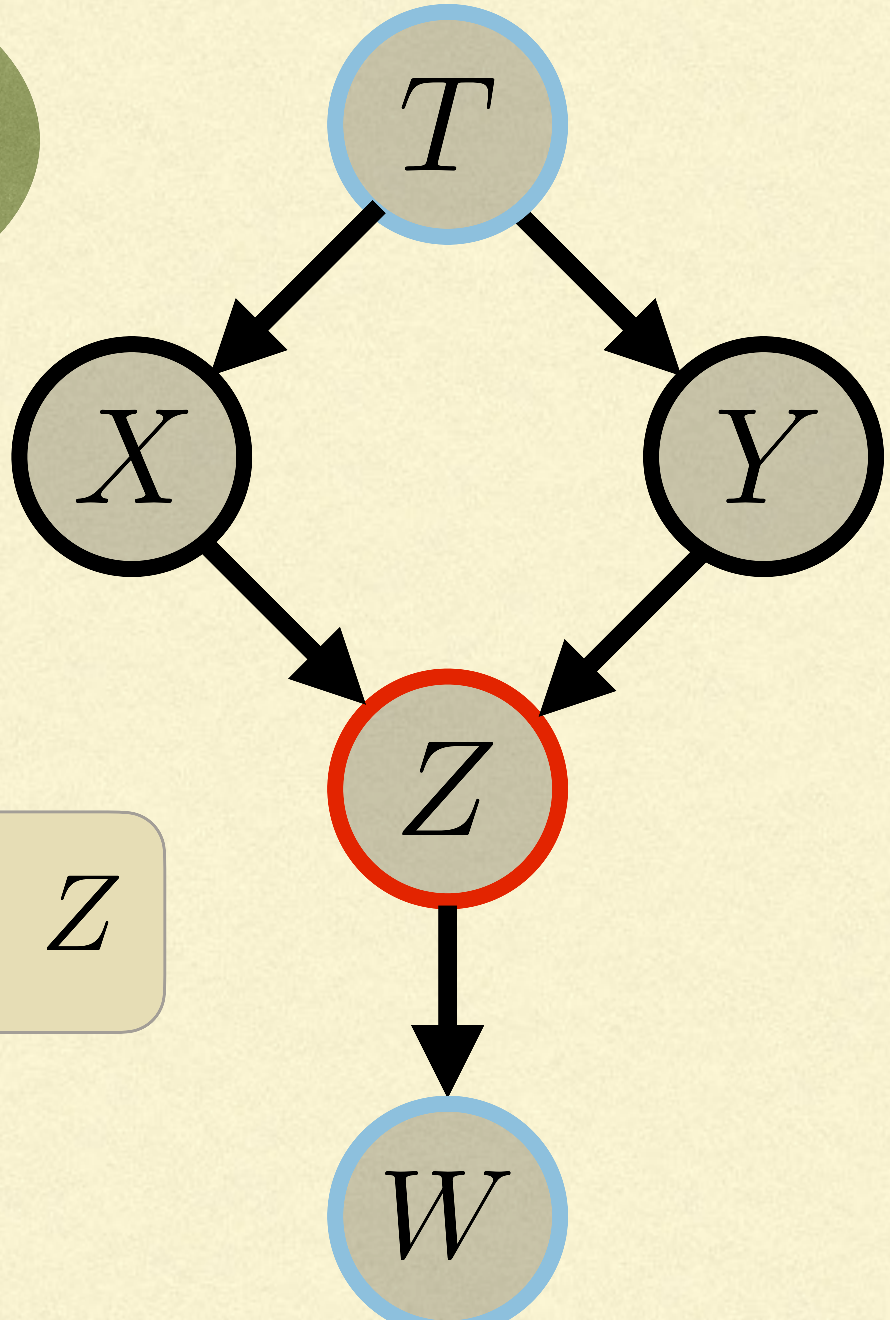
predicted graph



look for
$A \perp\!\!\!\perp B \mid C$

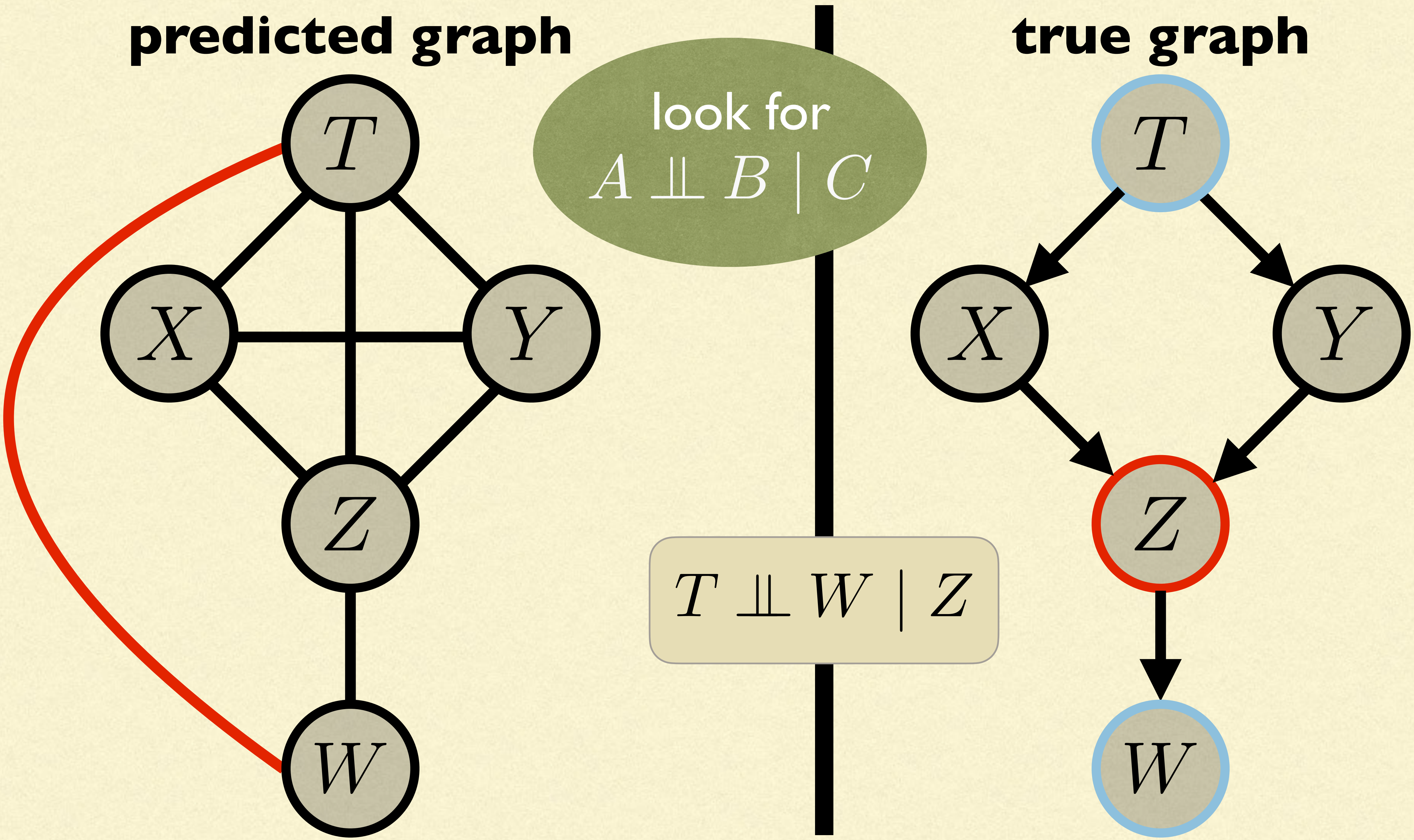
$T \perp\!\!\!\perp W \mid Z$

true graph



STEP 1: FIND ADJACENCIES

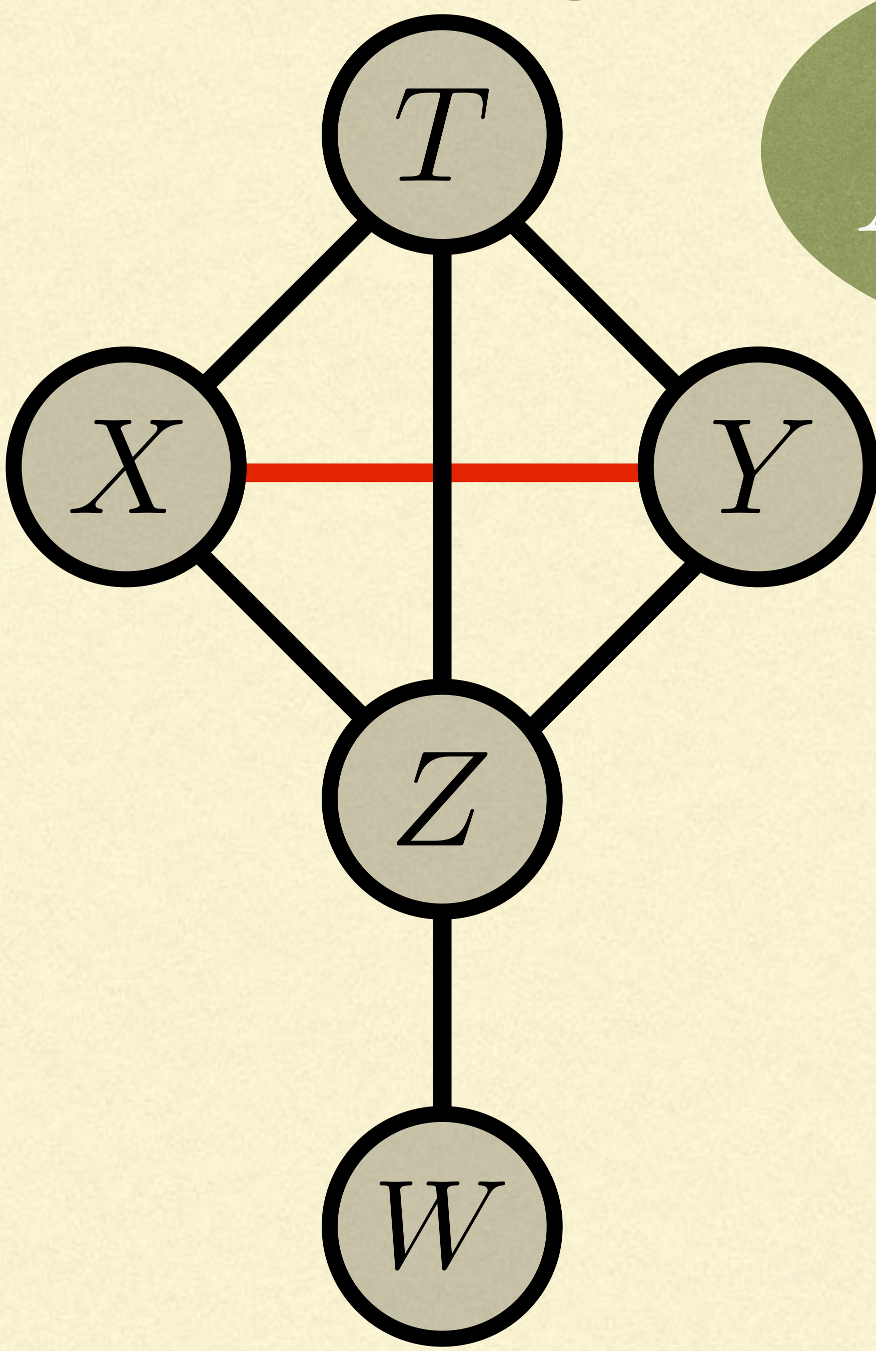
[Spirtes et al. 2001]



STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]

predicted graph

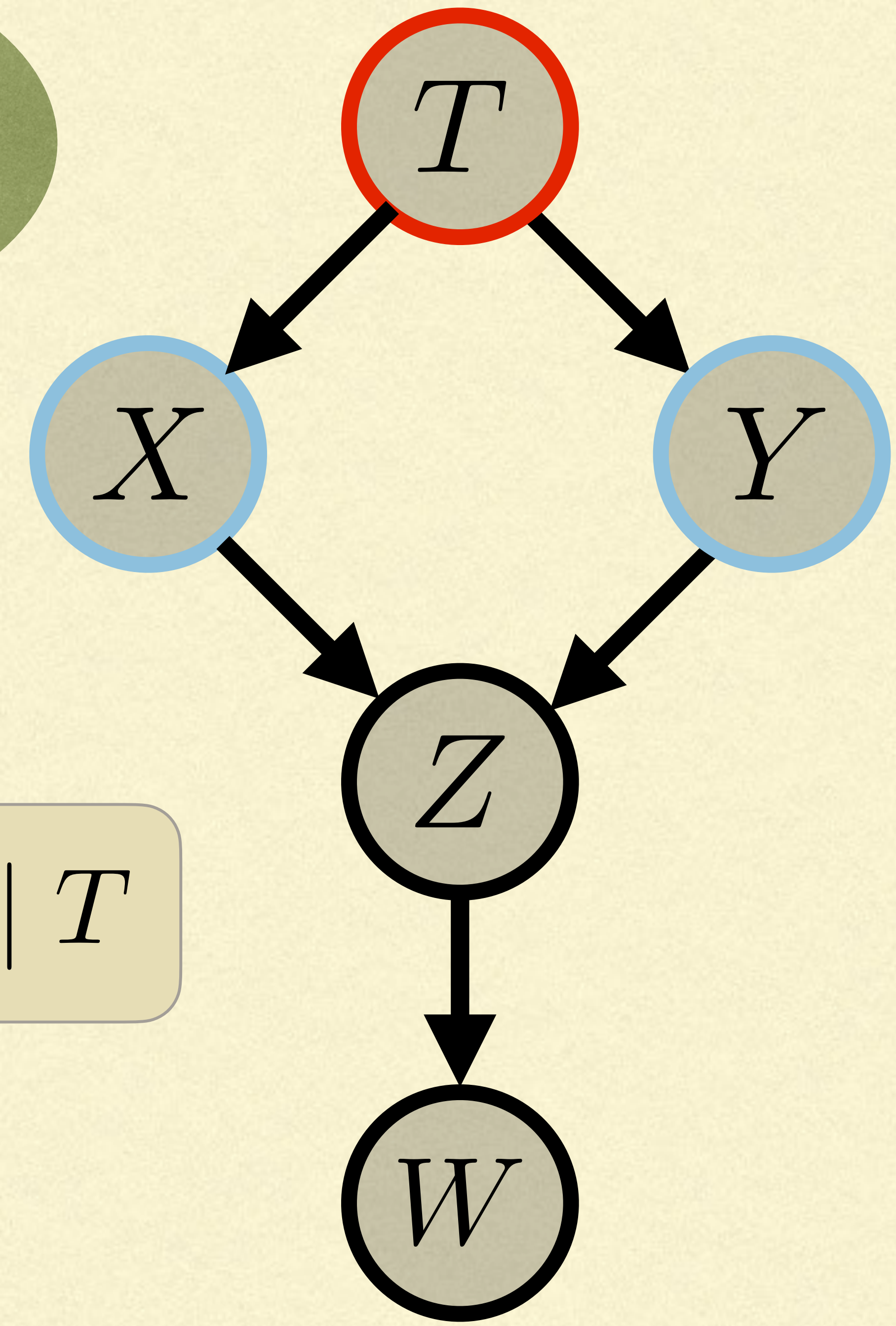


look for

$A \perp\!\!\!\perp B \mid C$

$X \perp\!\!\!\perp Y \mid T$

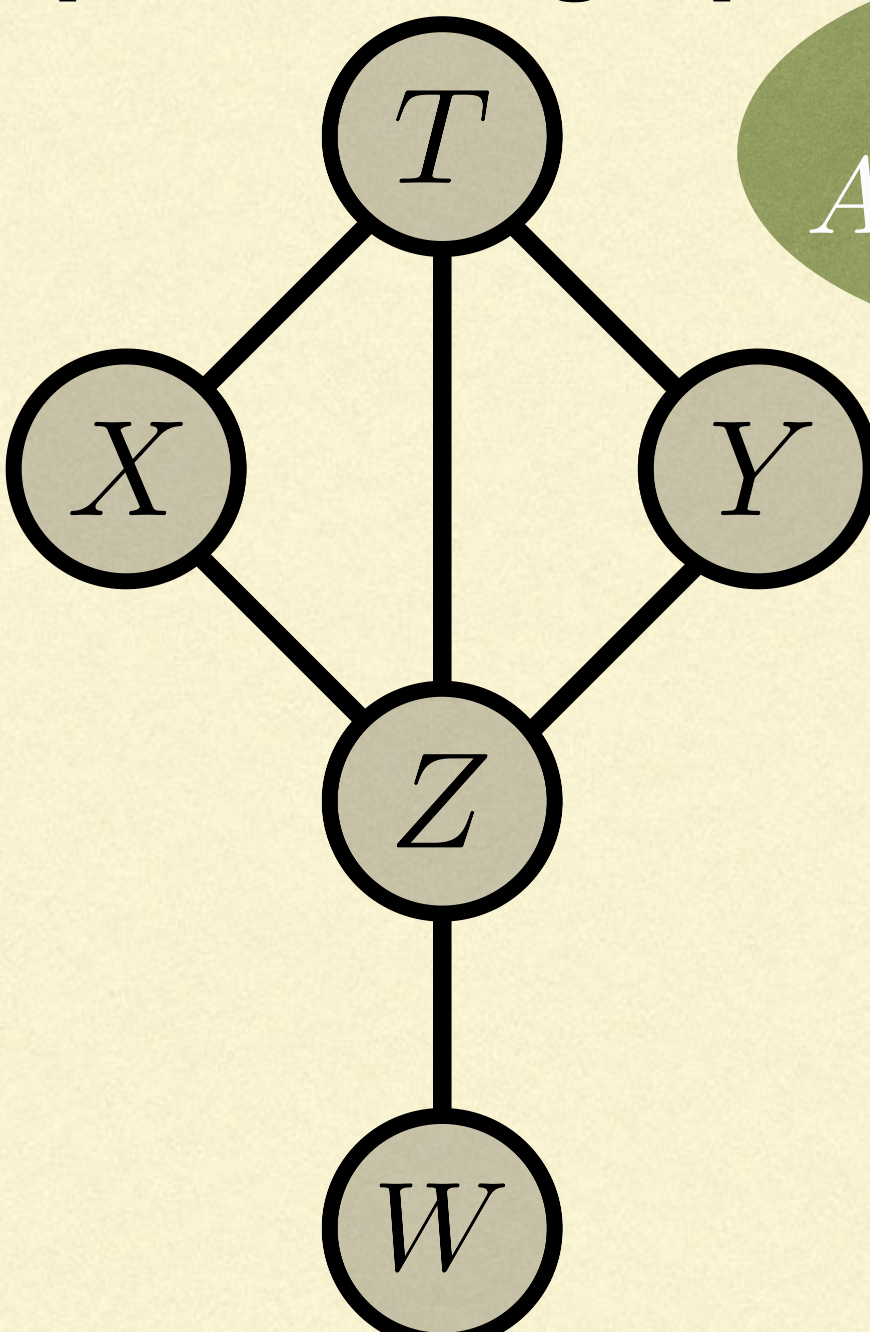
true graph



STEP 1: FIND ADJACENCIES

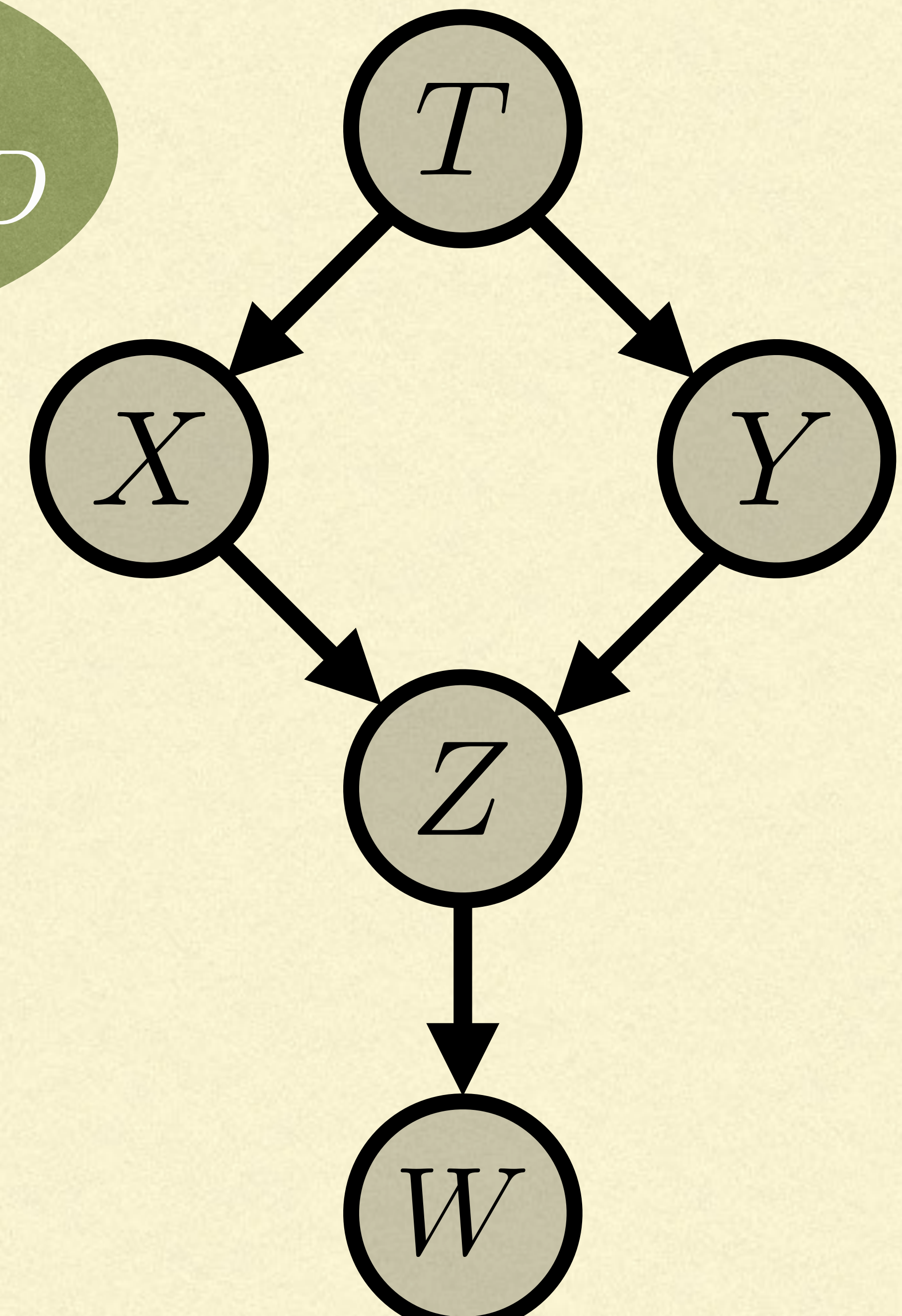
[Spirtes et al. 2001]

predicted graph



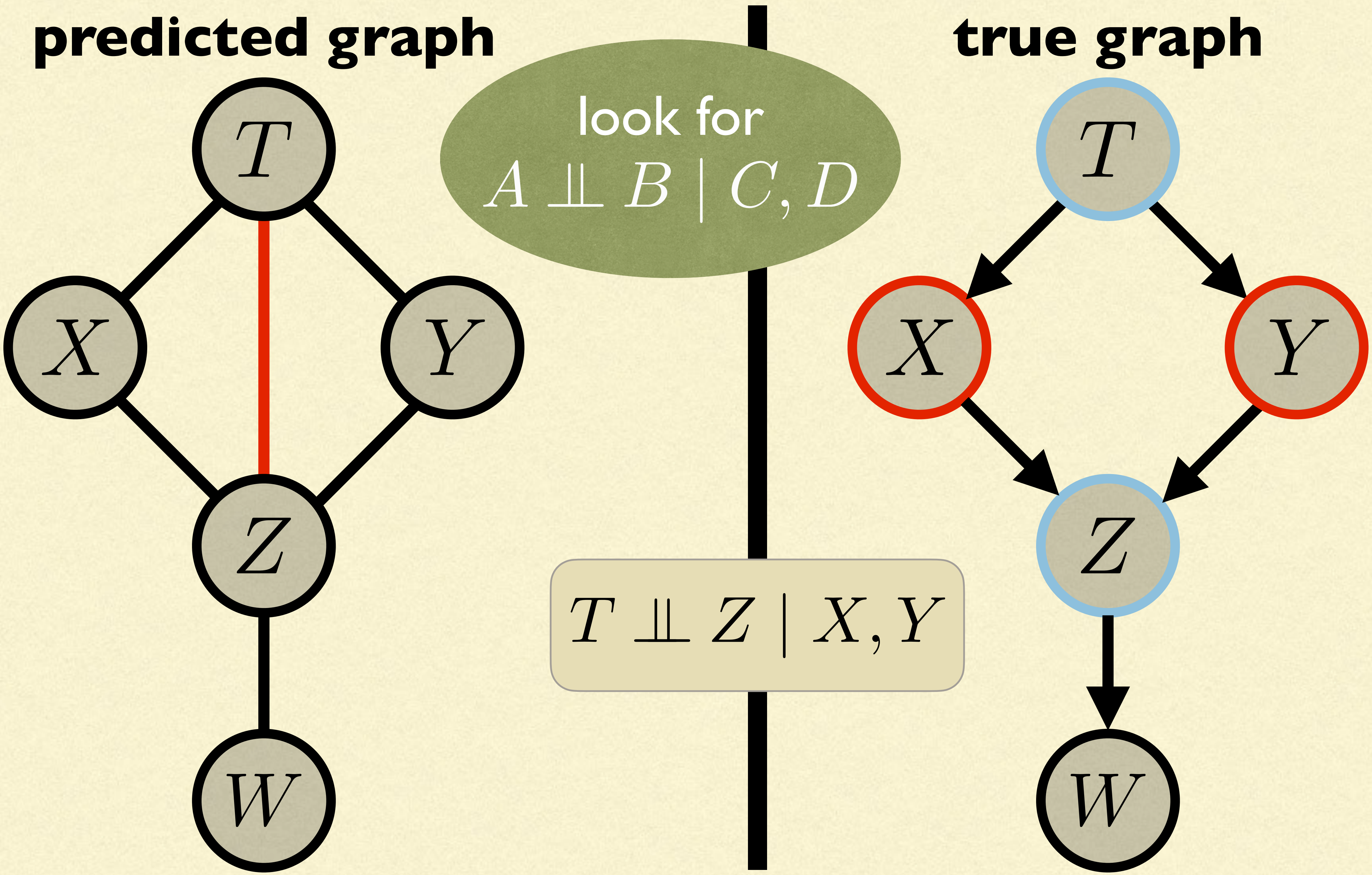
look for
$A \perp\!\!\!\perp B \mid C, D$

true graph



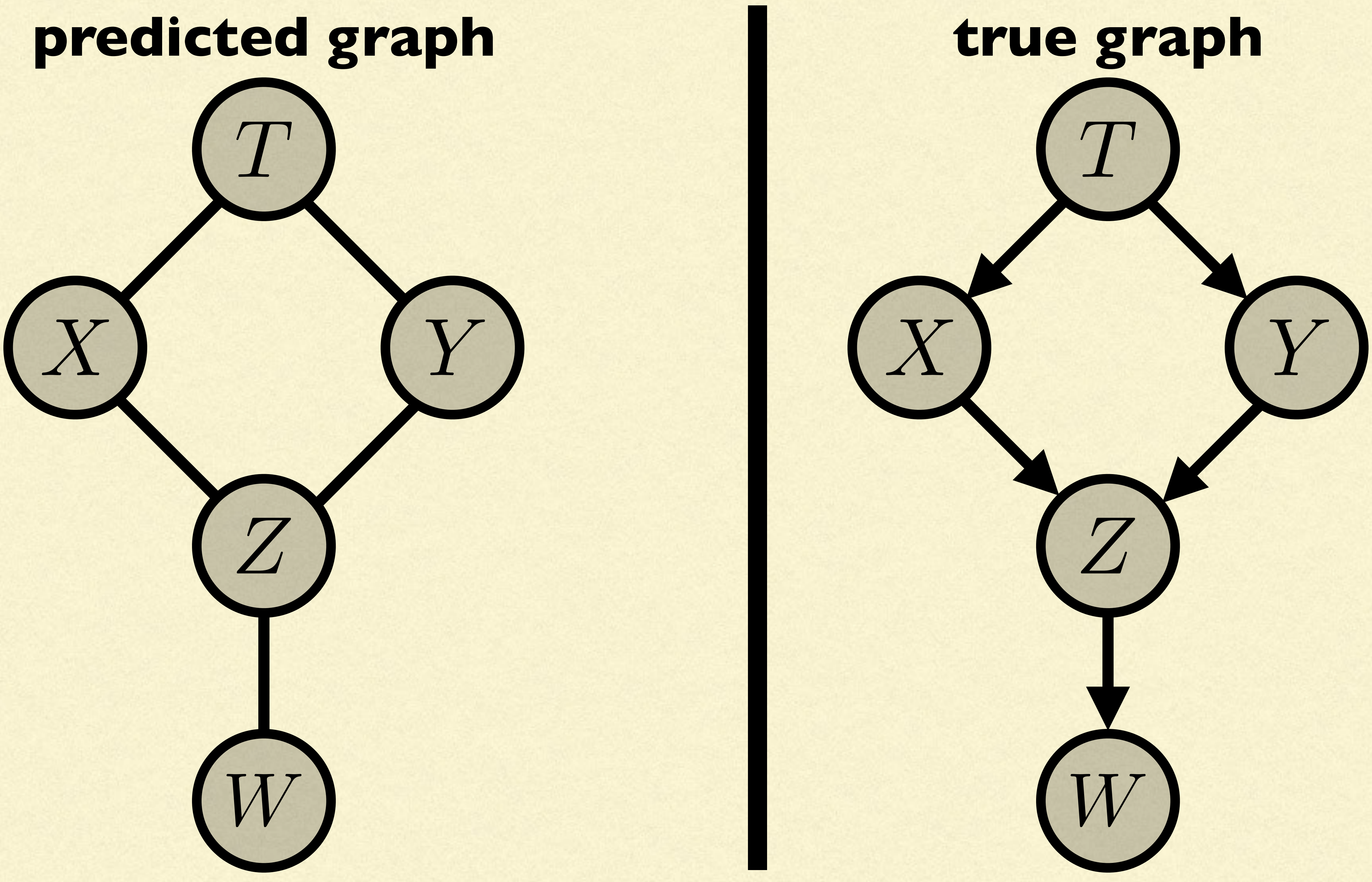
STEP 1: FIND ADJACENCIES

[Spirtes et al. 2001]



STEP 2: EDGE ORIENTATION

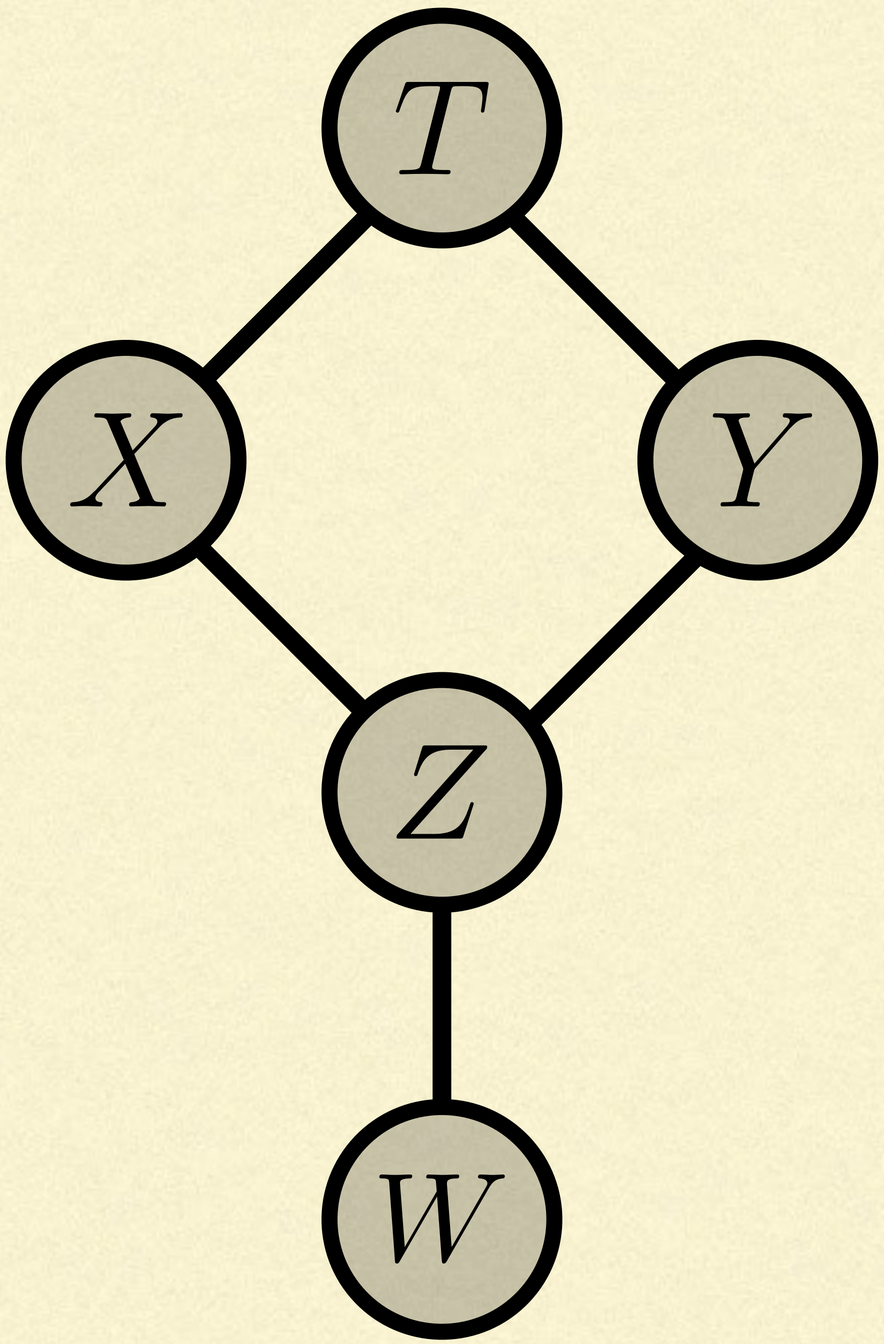
[Spirtes et al. 2001]



STEP 2: EDGE ORIENTATION

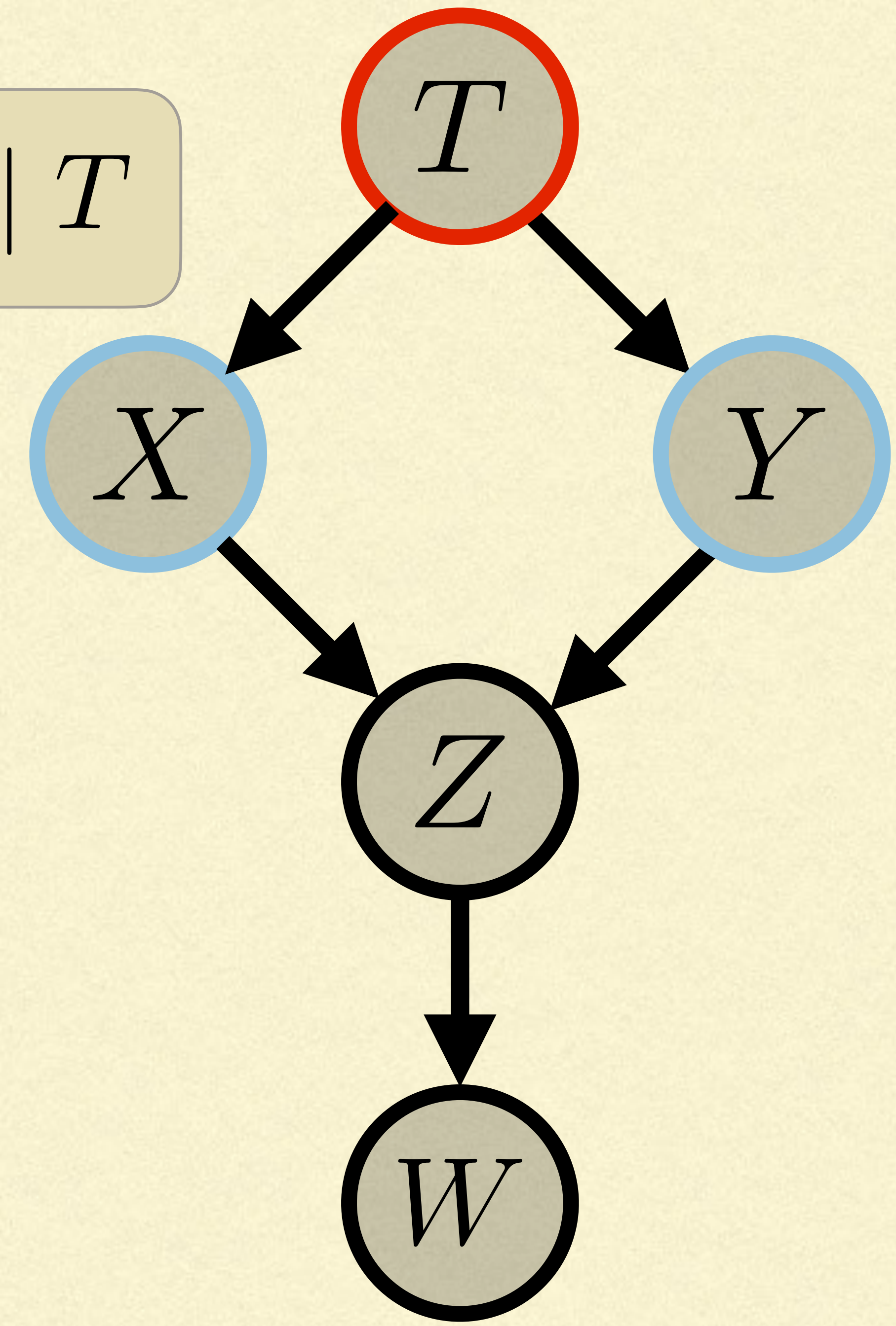
[Spirtes et al. 2001]

predicted graph



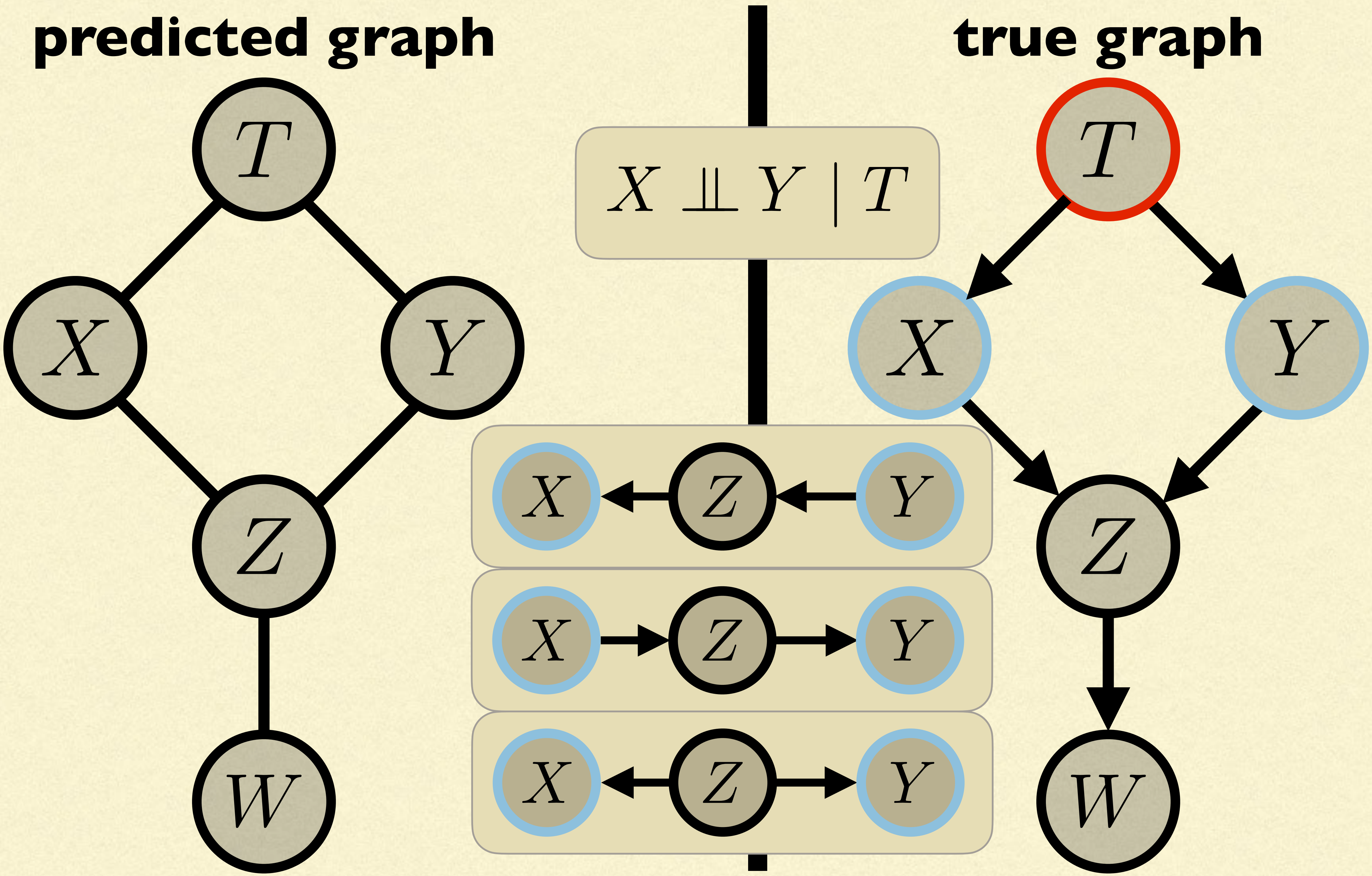
$$X \perp\!\!\!\perp Y \mid T$$

true graph



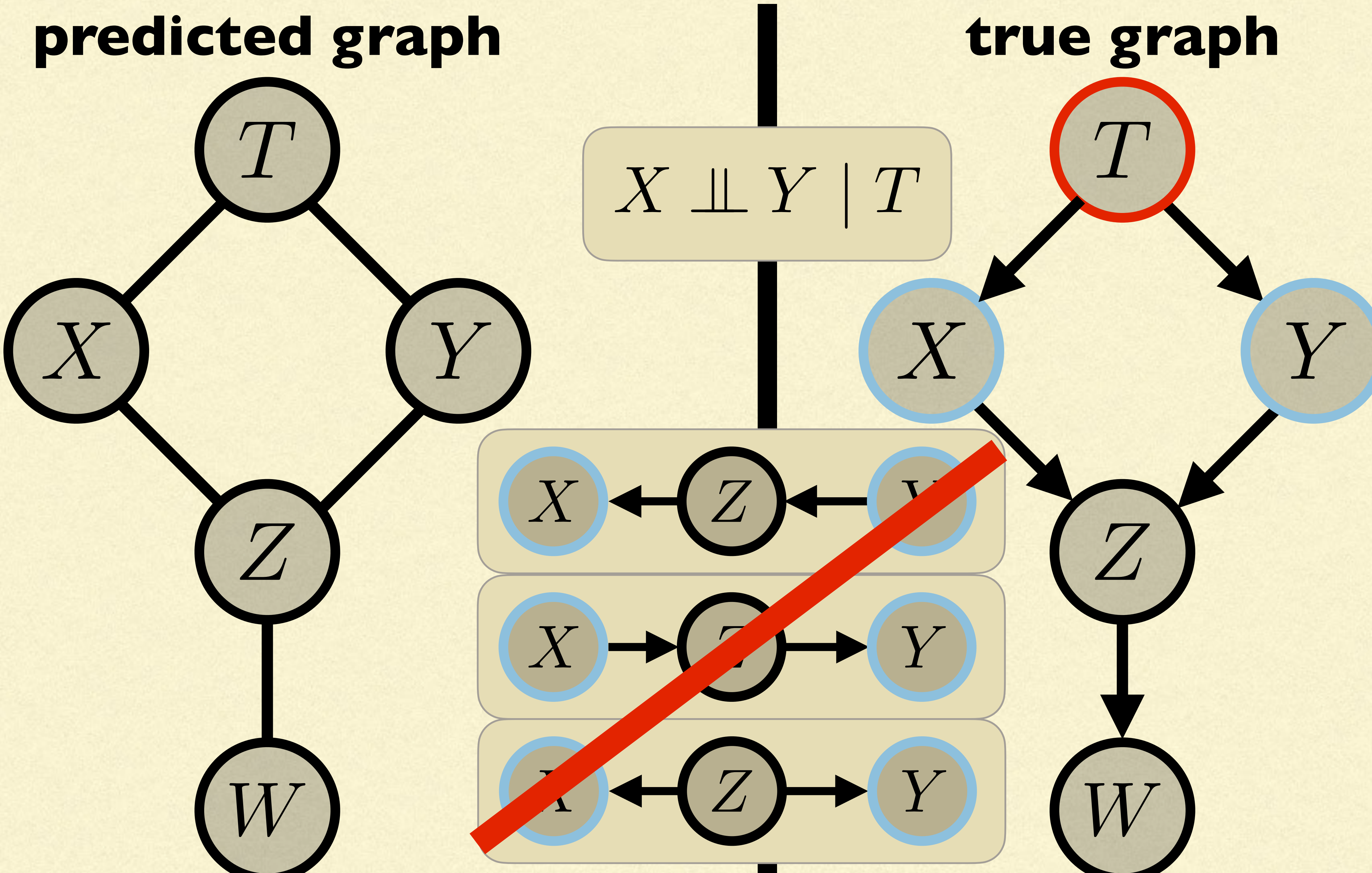
STEP 2: EDGE ORIENTATION

[Spirtes et al. 2001]



STEP 2: EDGE ORIENTATION

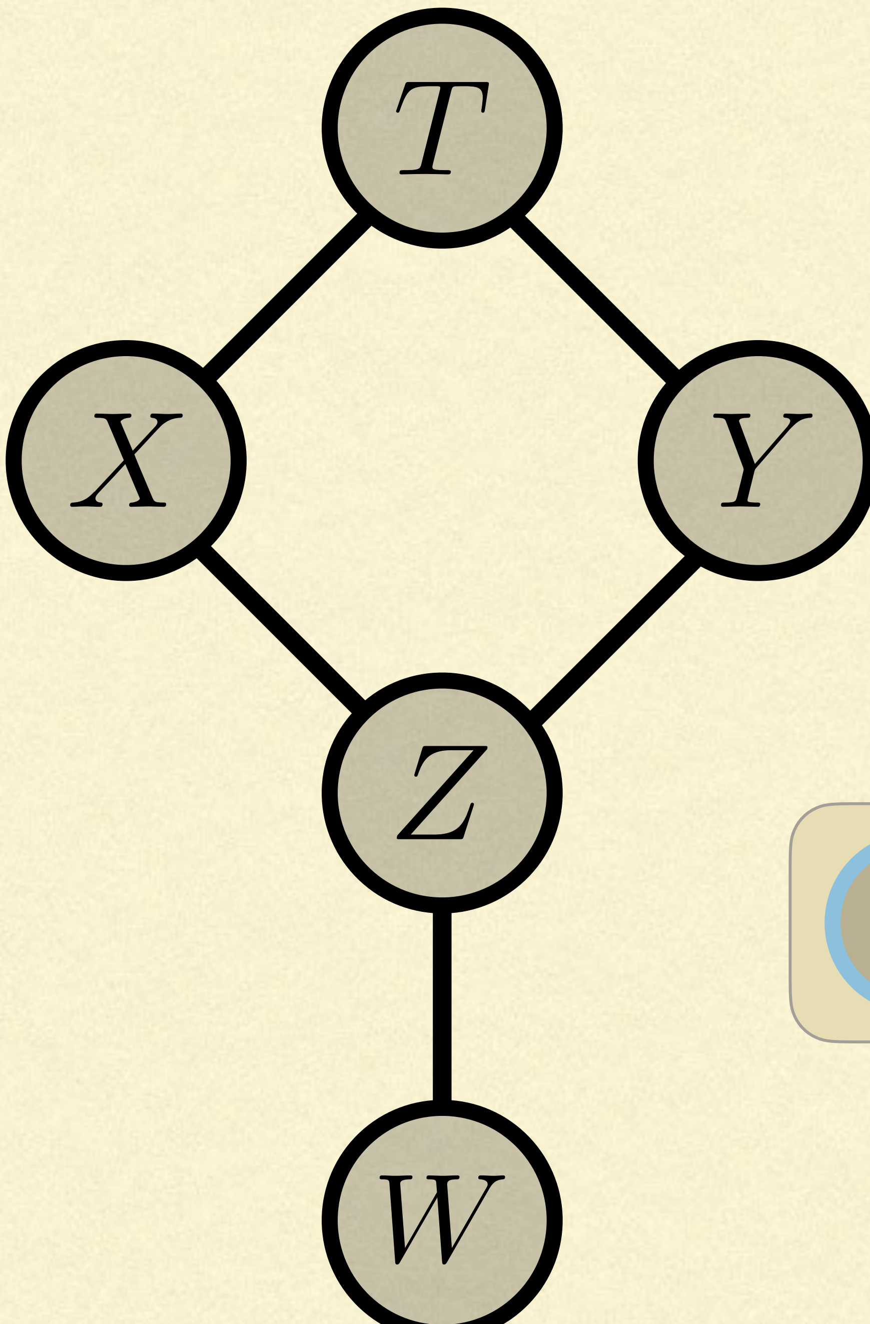
[Spirtes et al. 2001]



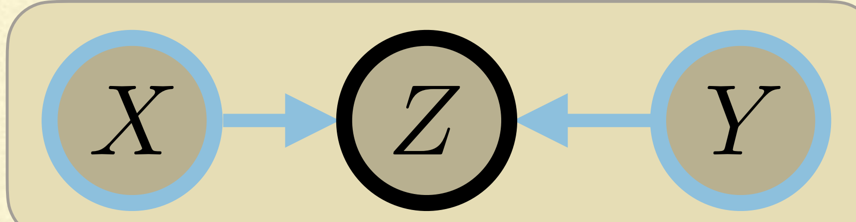
STEP 2: EDGE ORIENTATION

[Spirtes et al. 2001]

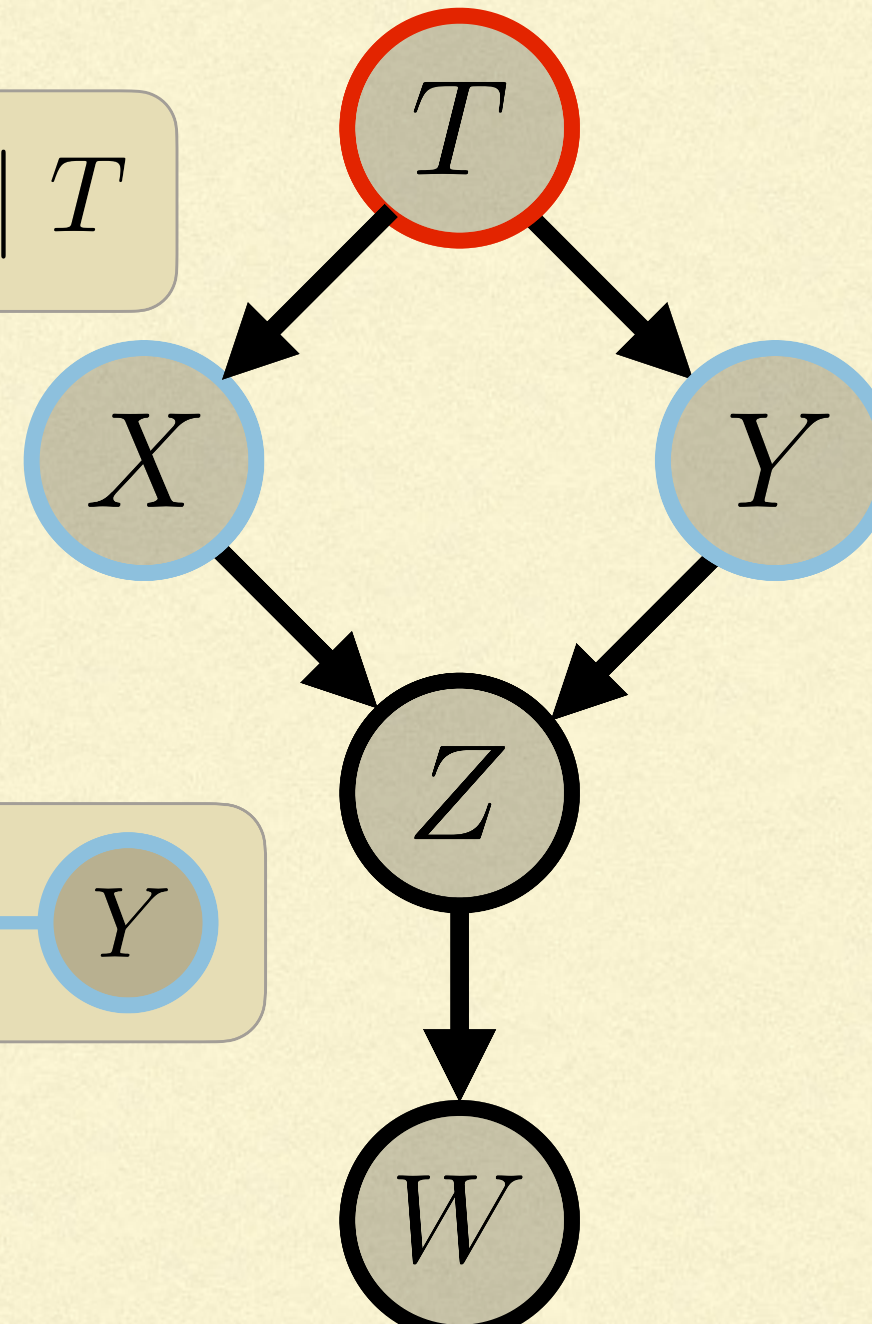
predicted graph



$$X \perp\!\!\!\perp Y \mid T$$



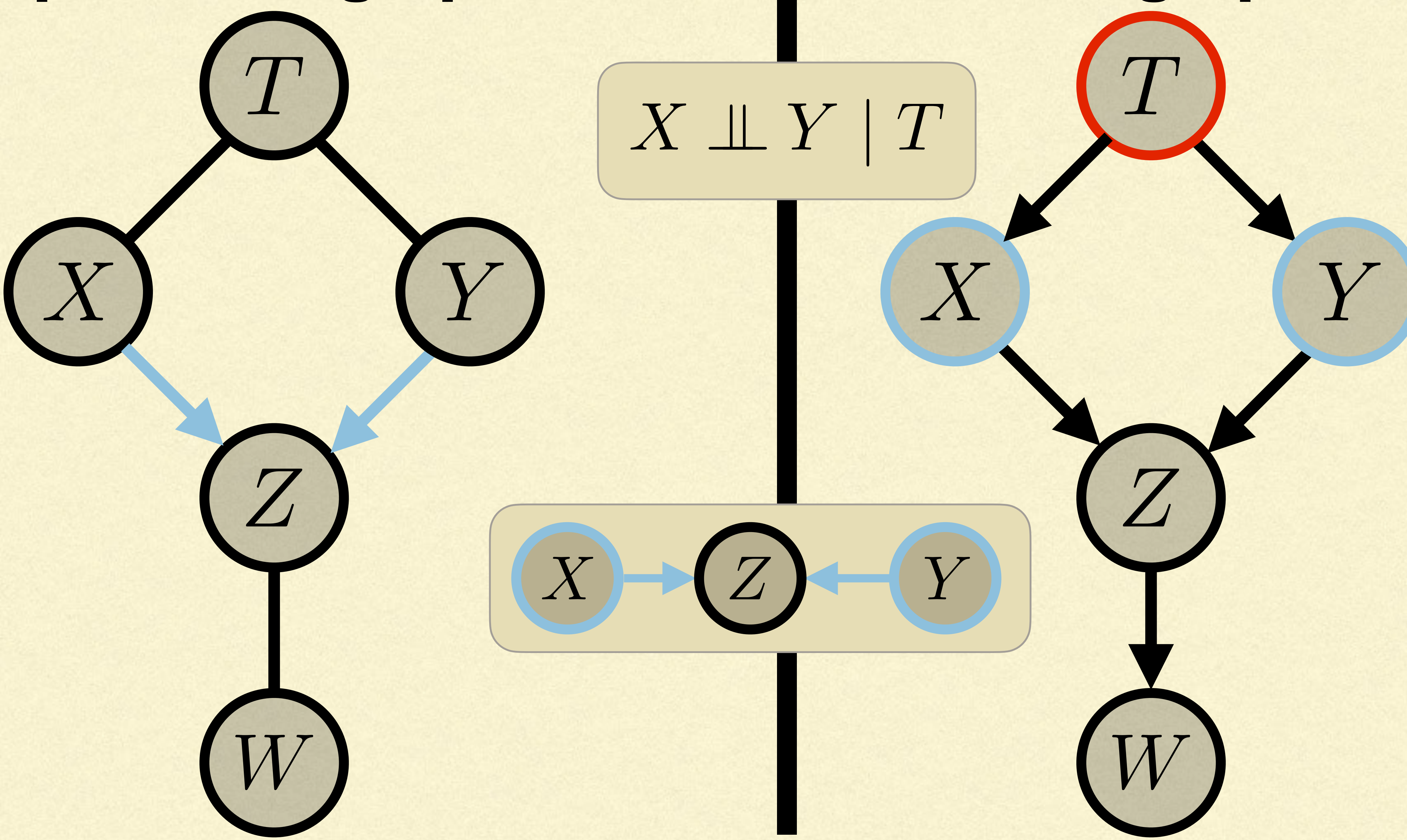
true graph



STEP 2: EDGE ORIENTATION

[Spirtes et al. 2001]

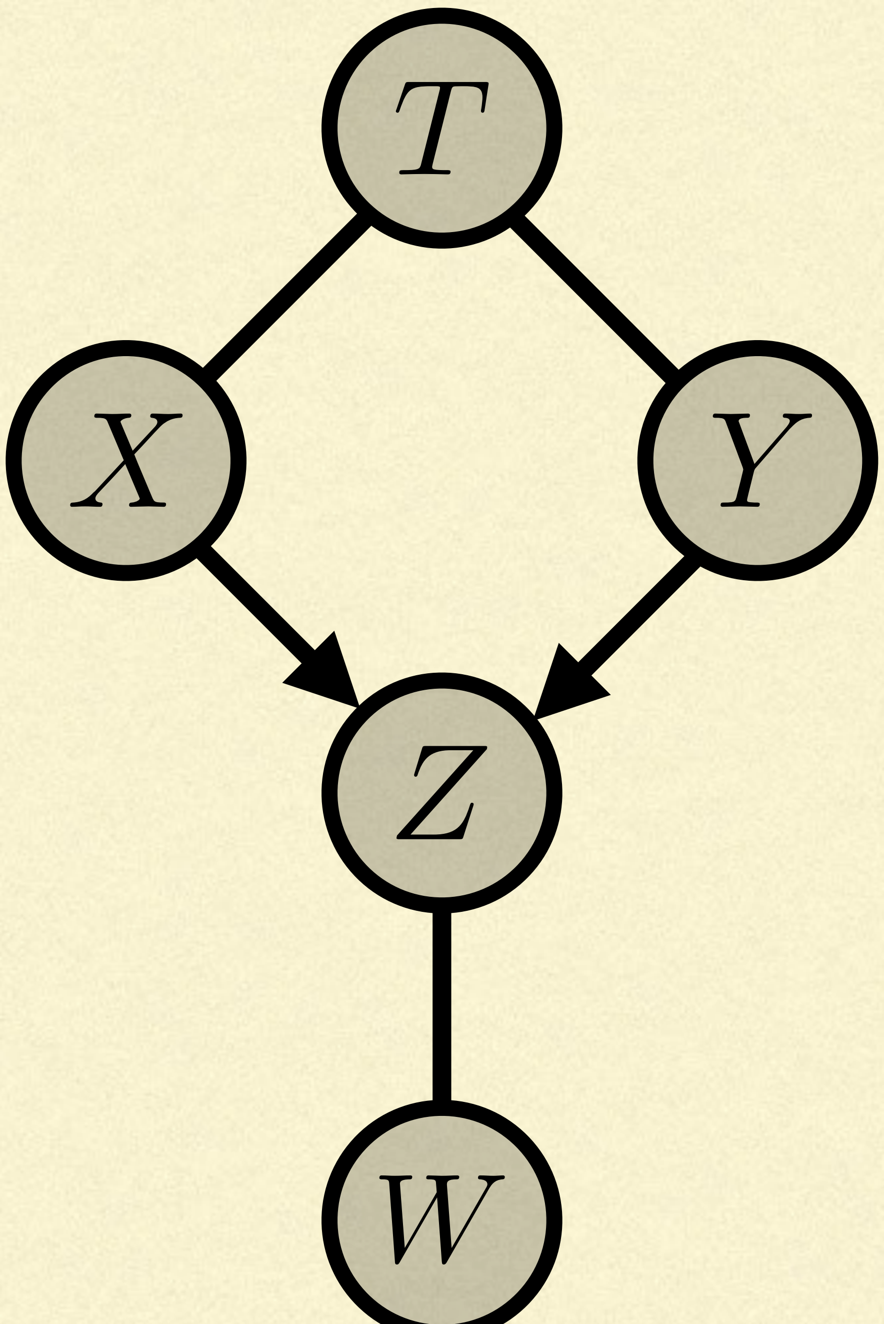
predicted graph **true graph**



STEP 2: EDGE ORIENTATION

[Spirtes et al. 2001]

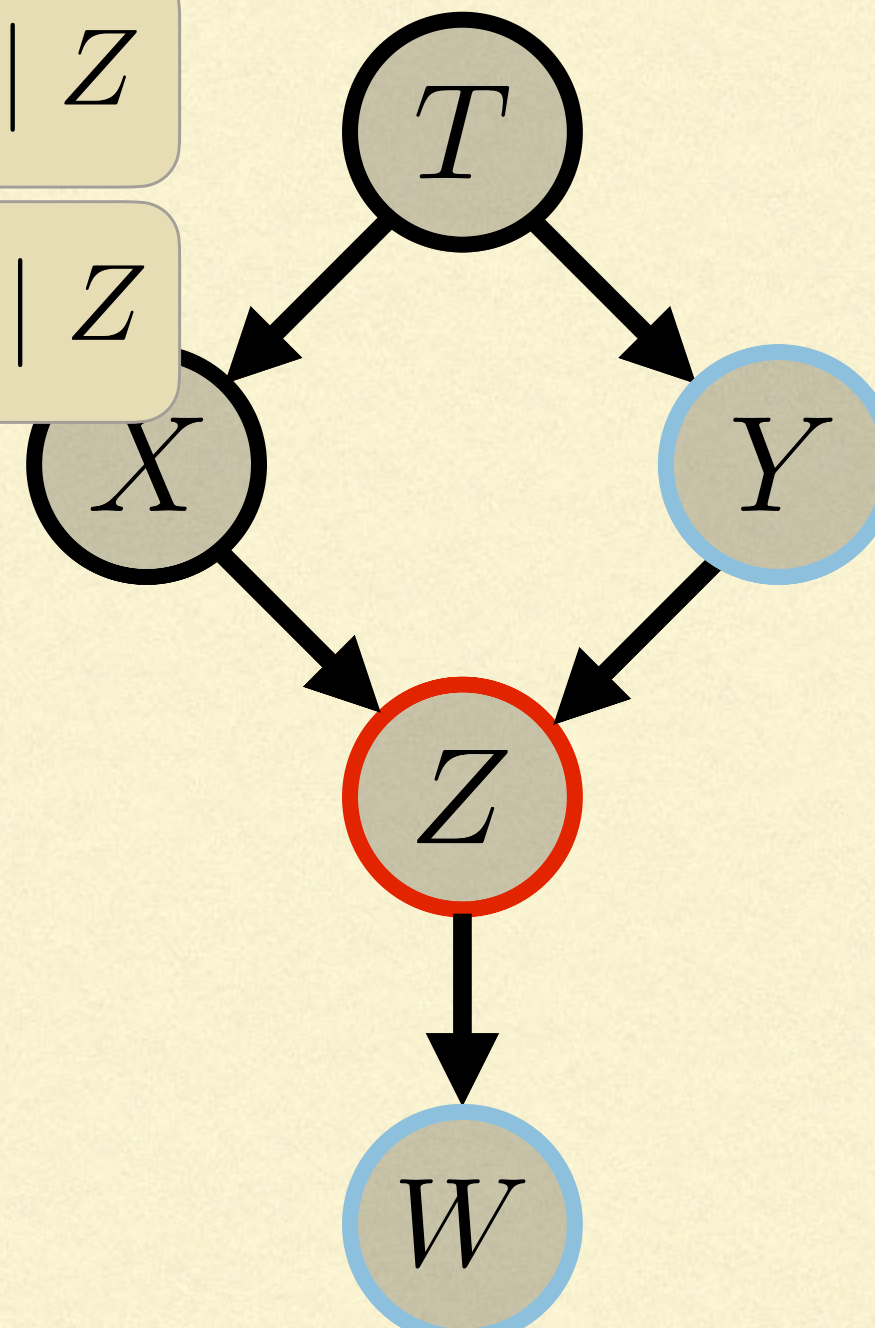
predicted graph



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

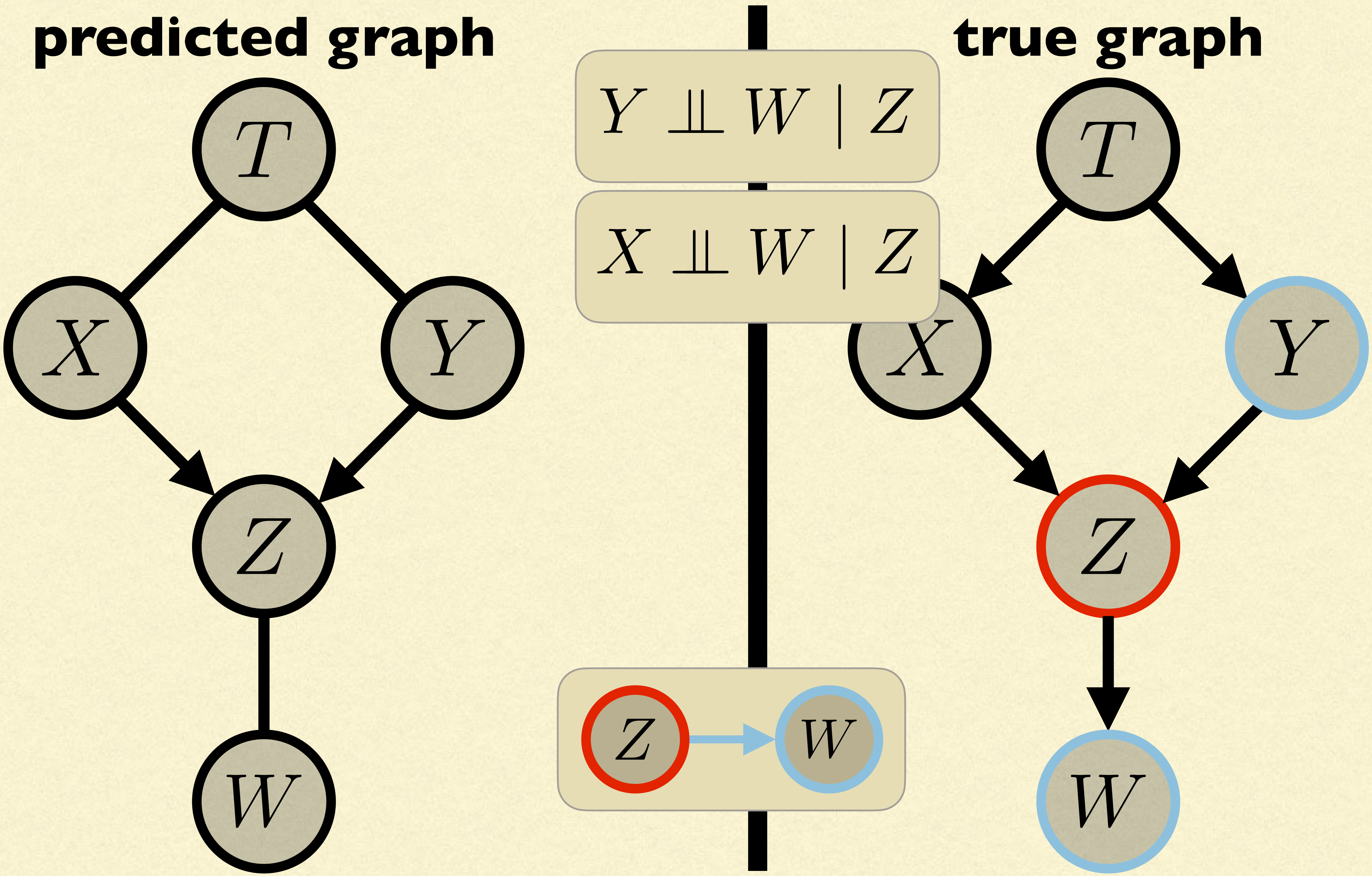
$$X \perp\!\!\!\perp W \mid Z$$

true graph



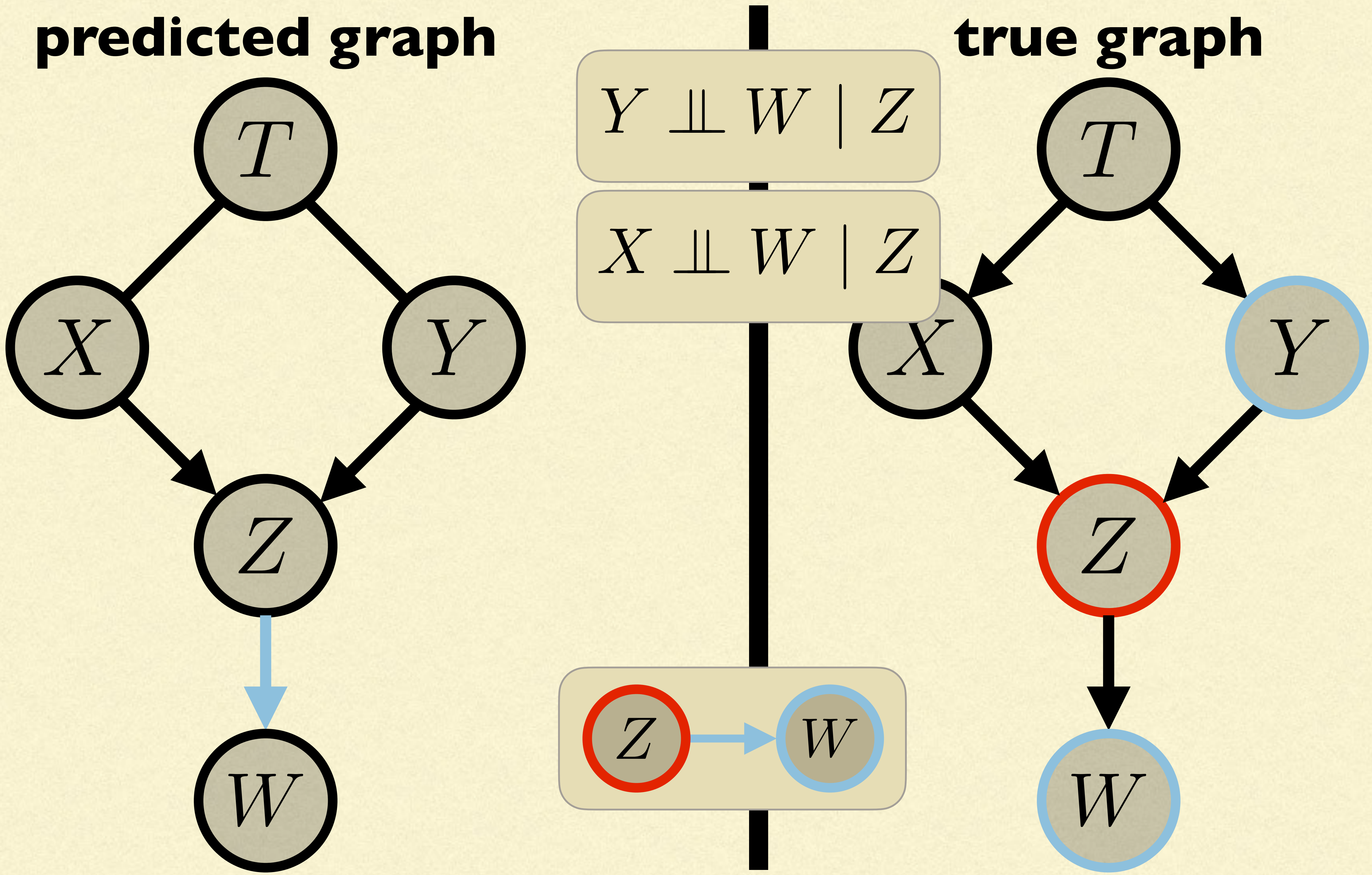
STEP 2: EDGE ORIENTATION

[Spirtes et al. 2001]



STEP 2: EDGE ORIENTATION

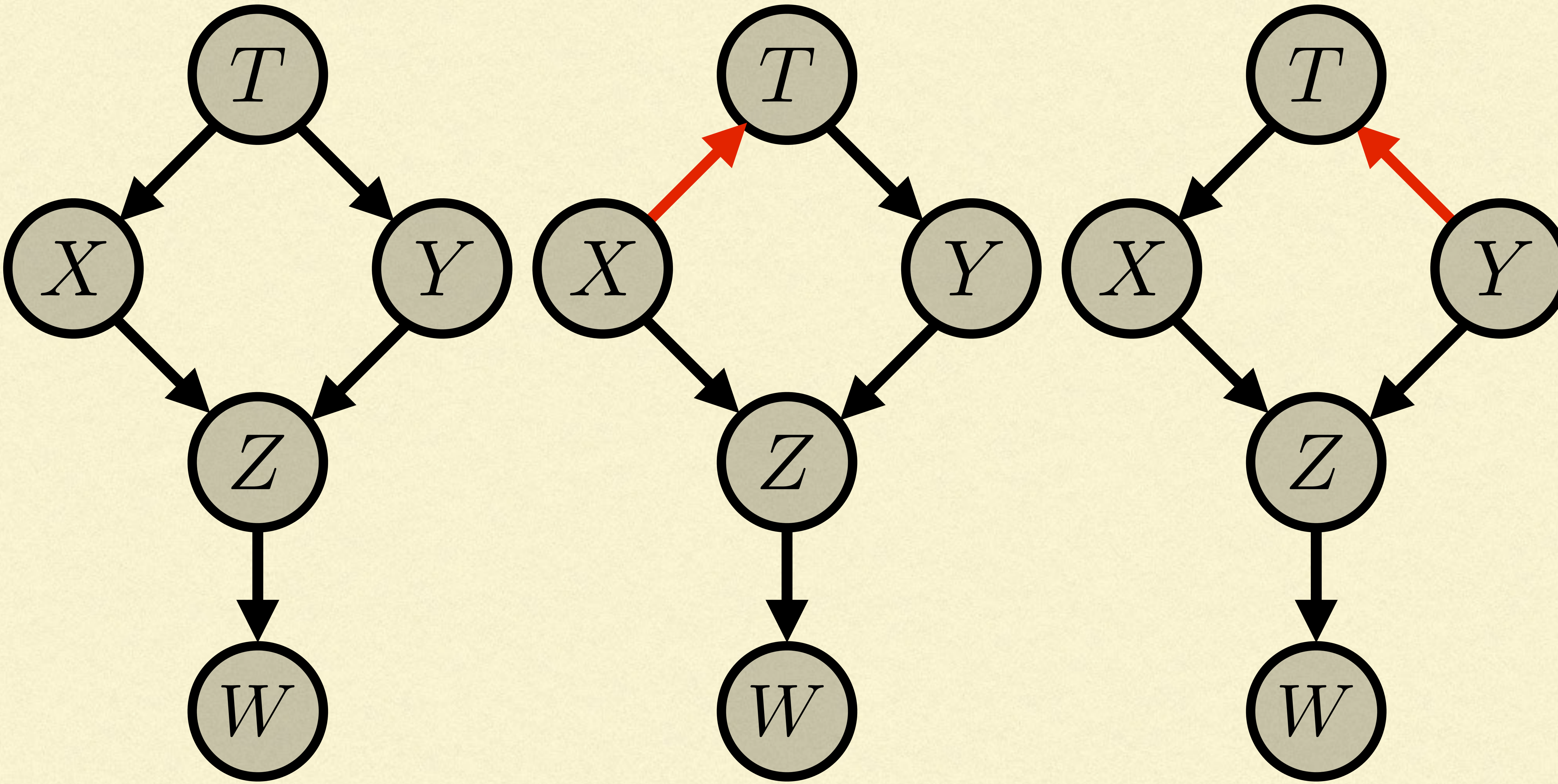
[Spirtes et al. 2001]



POSSIBLE GRAPHS

[Spirtes et al. 2001]

predicted graphs



HOW DO WE GET A GRAPH?

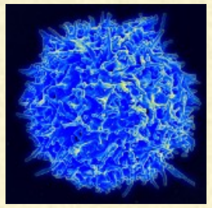
Way 2: Additive Noise Models

[Hoyer, Janzing, Mooij, Peters, Schölkopf, 2009]

ANM: ADDITIVE NOISE MODEL

[Hoyer, Janzing, Mooij, Peters, Schölkopf, 2009]

X: 

Y: 

assumption:

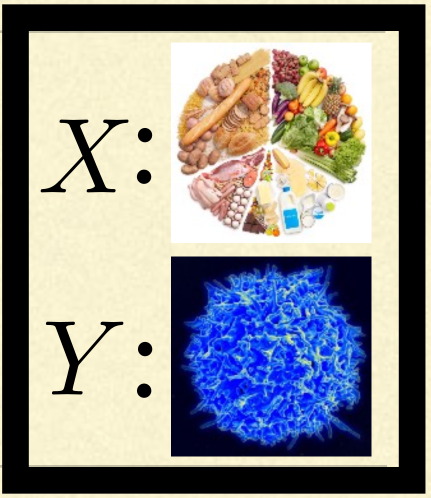
$$X \to Y$$

$$Y = f(X) + N_Y \qquad N_Y \perp\!\!\!\perp X$$

noise

ANM: ADDITIVE NOISE MODEL

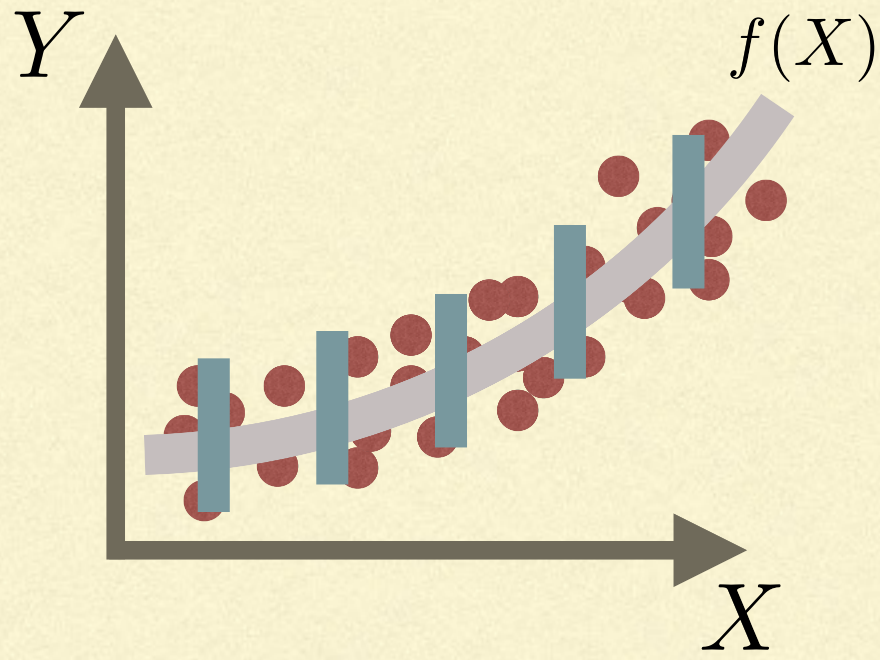
[Hoyer, Janzing, Mooij, Peters, Schölkopf, 2009]



assumption:

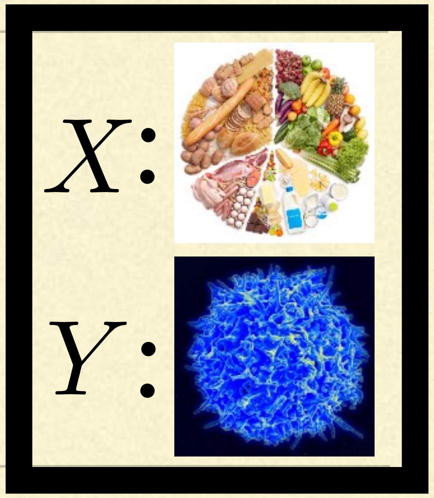
$$X \to Y$$

$$Y = f(X) + N_Y \qquad N_Y \perp\!\!\!\perp X$$



ANM: ADDITIVE NOISE MODEL

[Hoyer, Janzing, Mooij, Peters, Schölkopf, 2009]



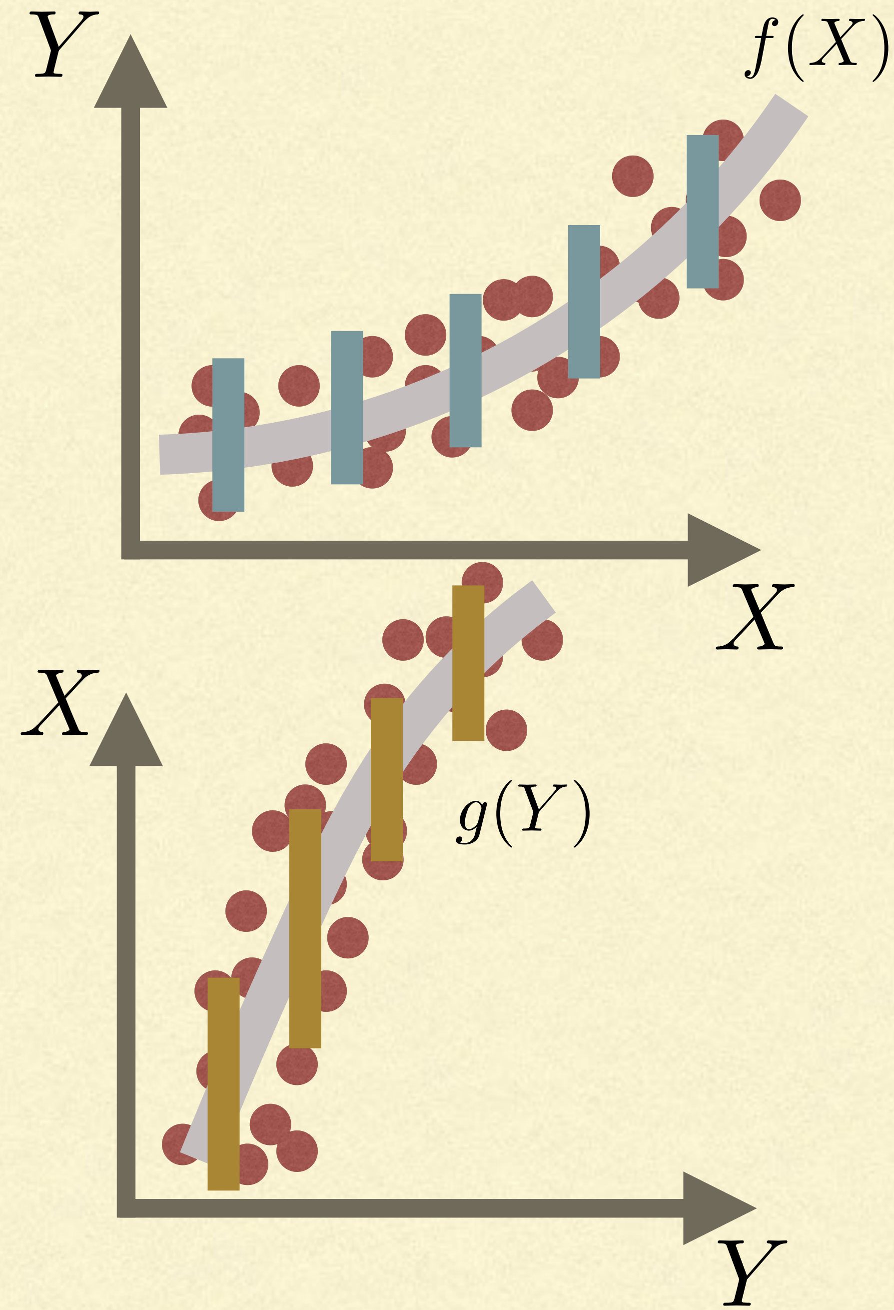
assumption:

$$X \to Y$$

$$Y = f(X) + N_Y \qquad N_Y \perp\!\!\!\perp X$$

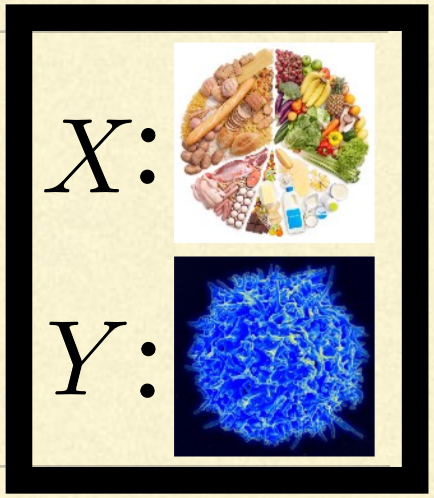
$$Y \to X$$

$$X = g(Y) + N_X \qquad N_X \perp\!\!\!\perp Y$$



ANM: ADDITIVE NOISE MODEL

[Hoyer, Janzing, Mooij, Peters, Schölkopf, 2009]



assumption:

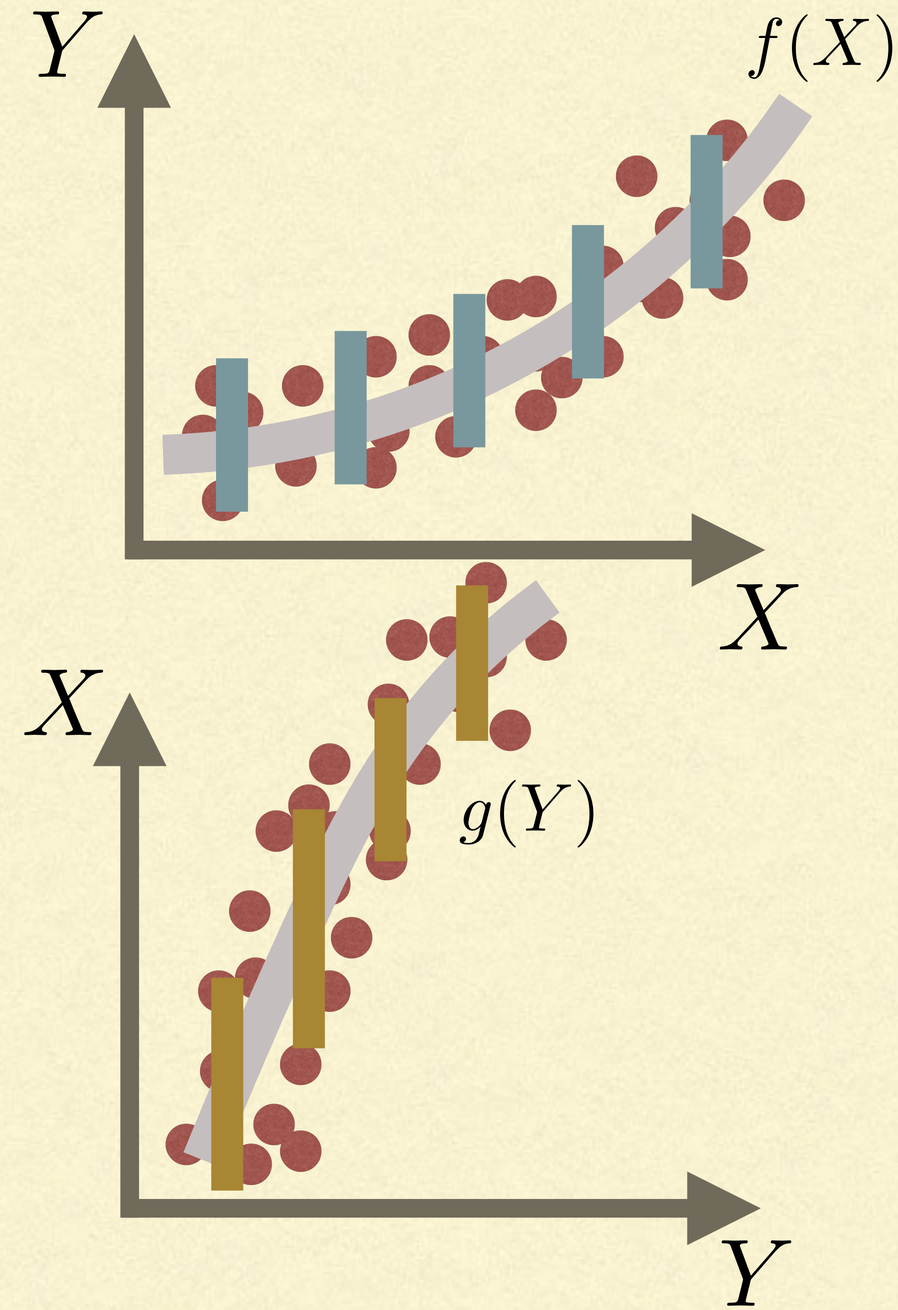
$$X \to Y$$

$$Y = f(X) + N_Y \qquad N_Y \perp\!\!\!\perp X$$

non-linear

$$Y \to X$$

$$X = g(Y) + N_X \qquad N_X \perp\!\!\!\perp Y$$

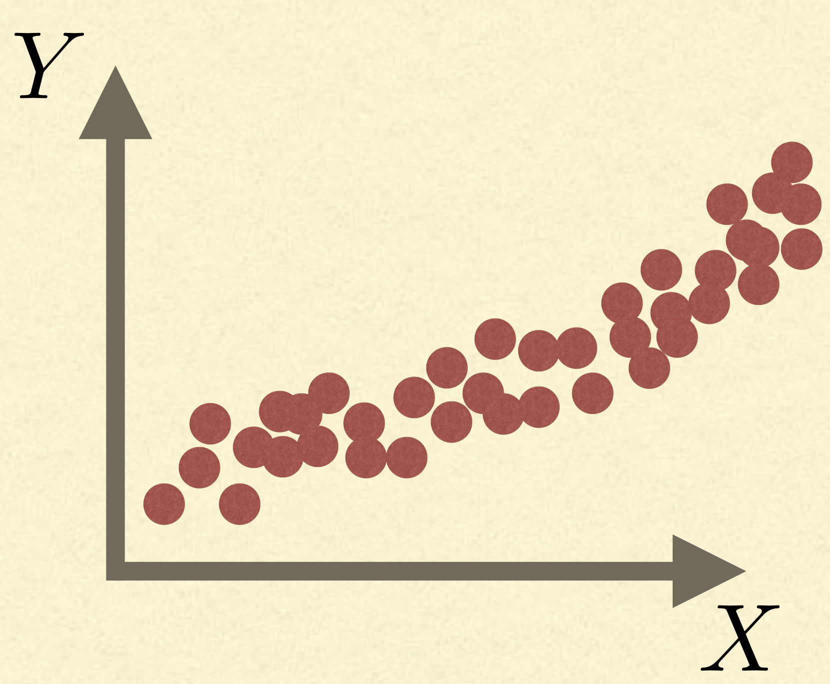


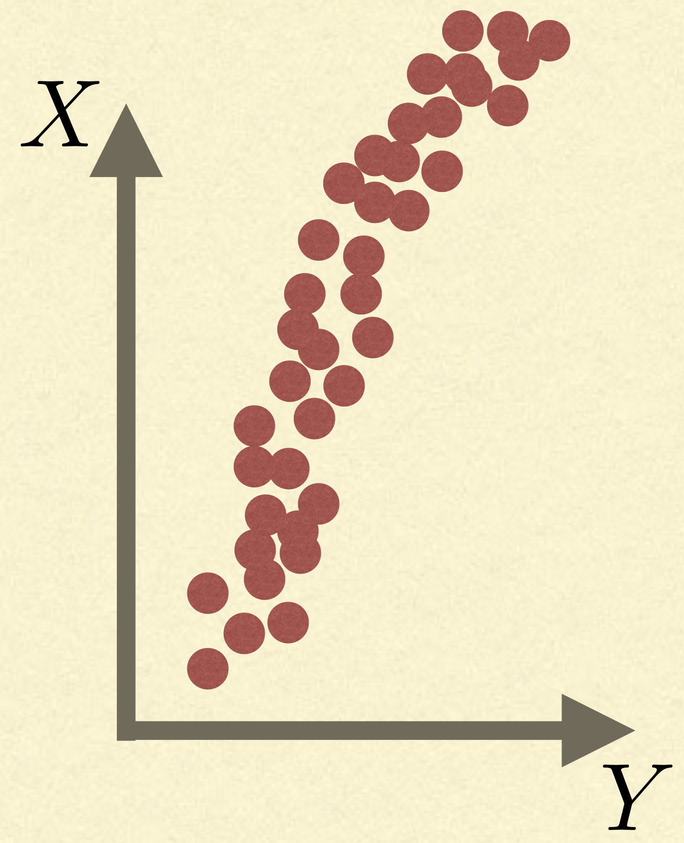
ANM: 1. COLLECT DATA

$X \rightarrow Y$

● training data

$Y \rightarrow X$



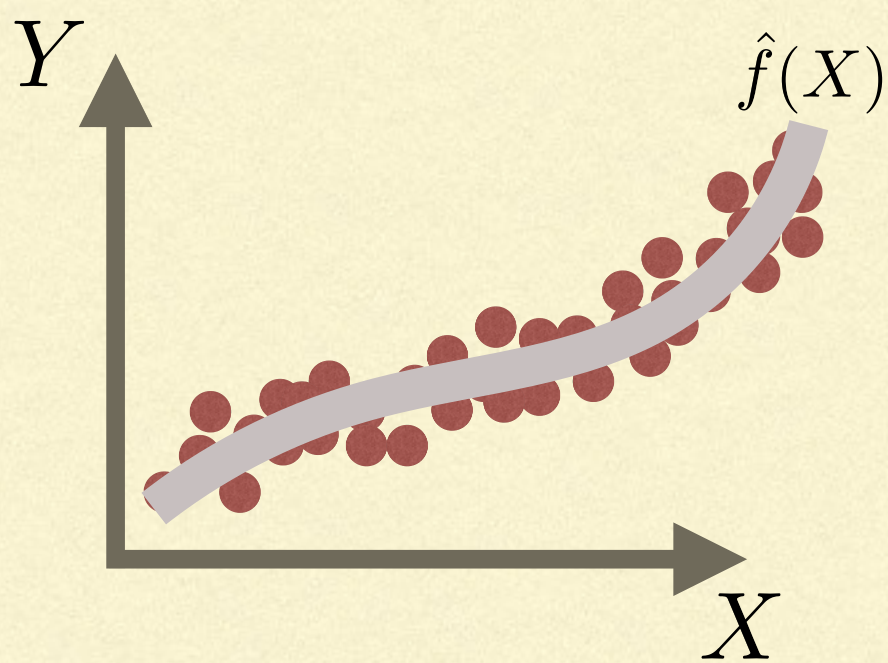


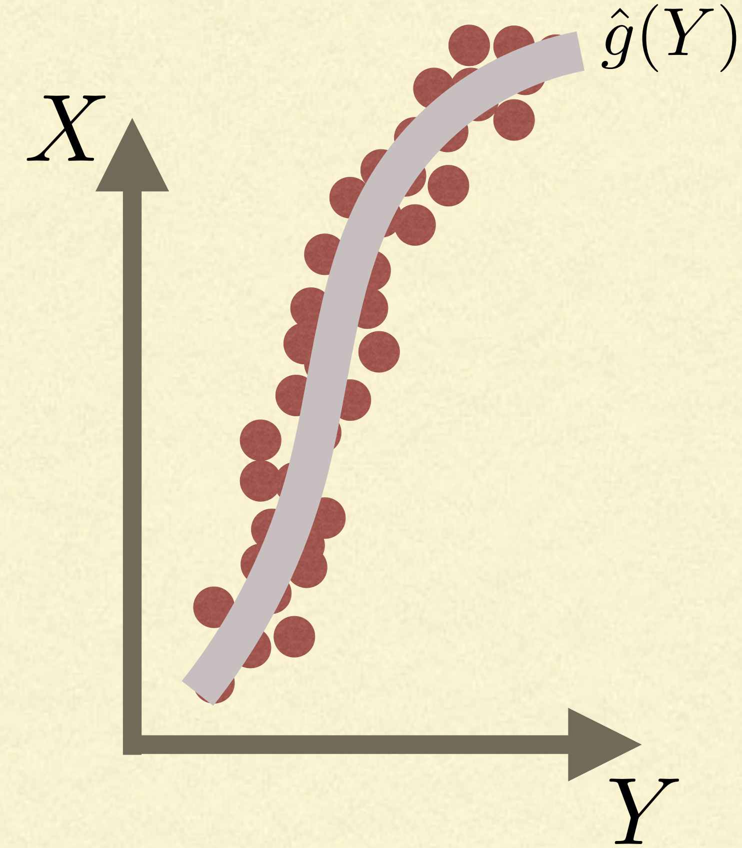
ANM: 2. LEARN MAPPING

$X \rightarrow Y$

● training data

$Y \rightarrow X$



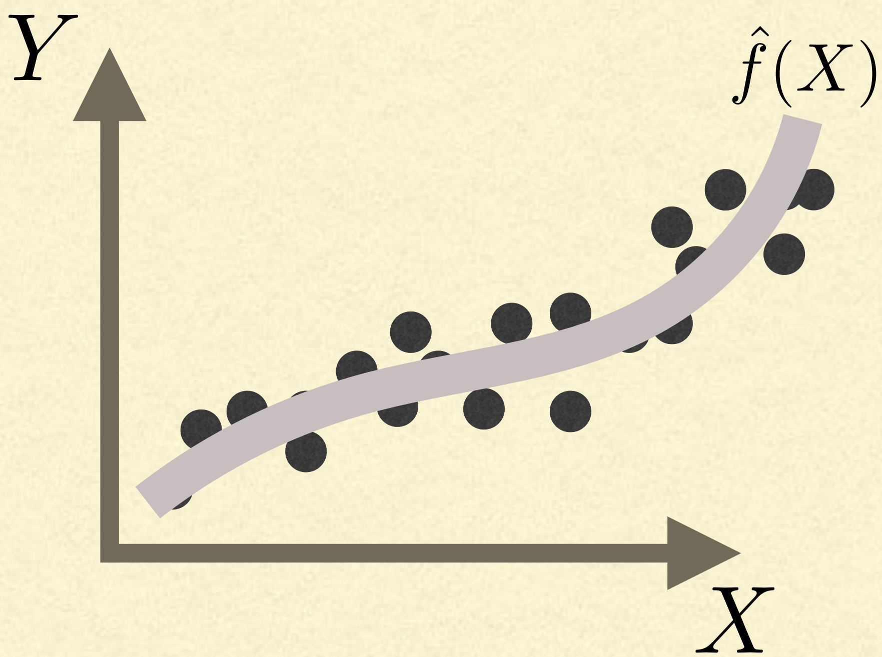


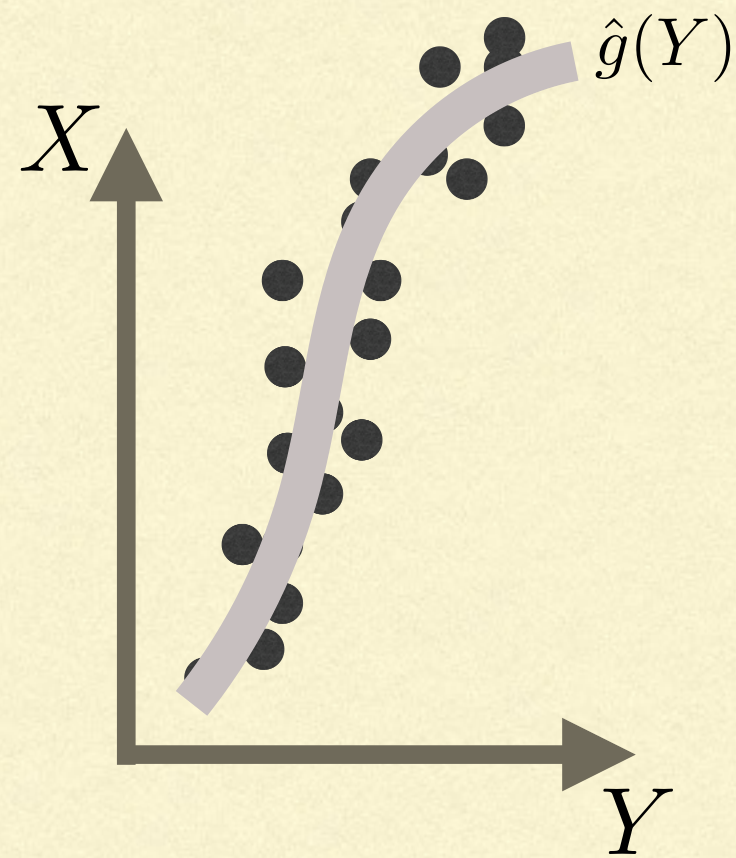
ANM: 3. COMPUTE RESIDUALS

$X \to Y$

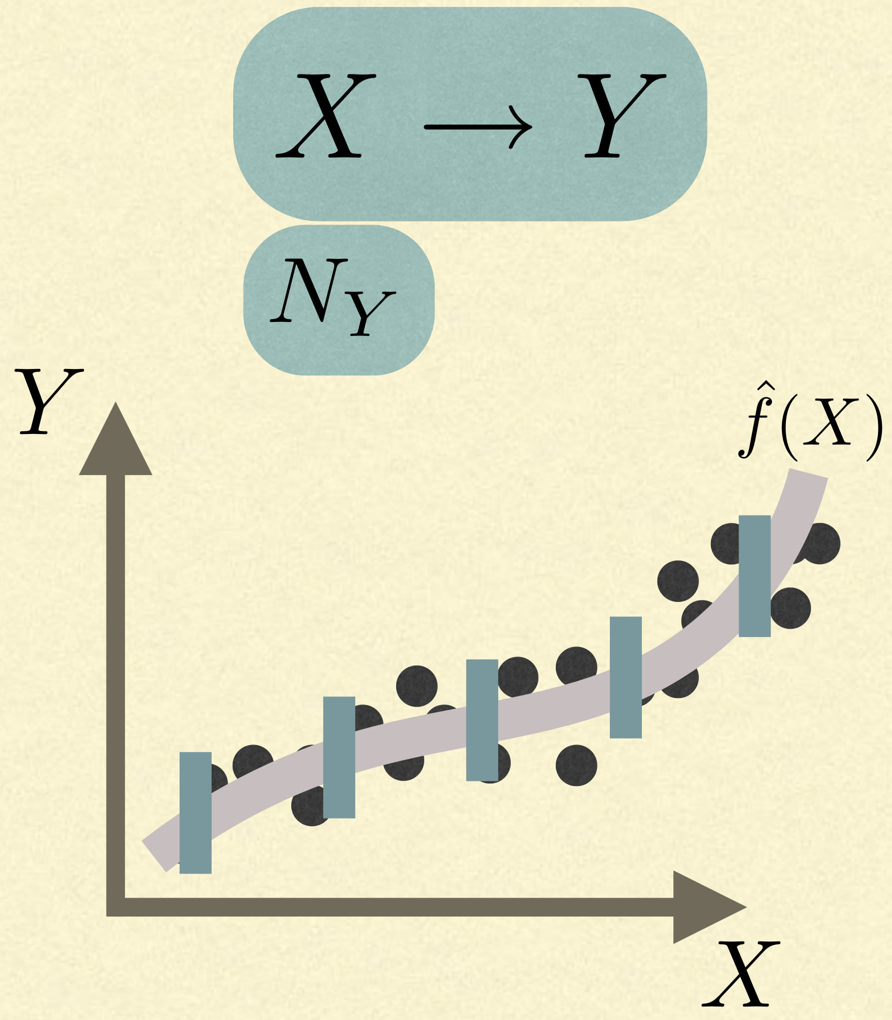
training data
test data

$Y \to X$

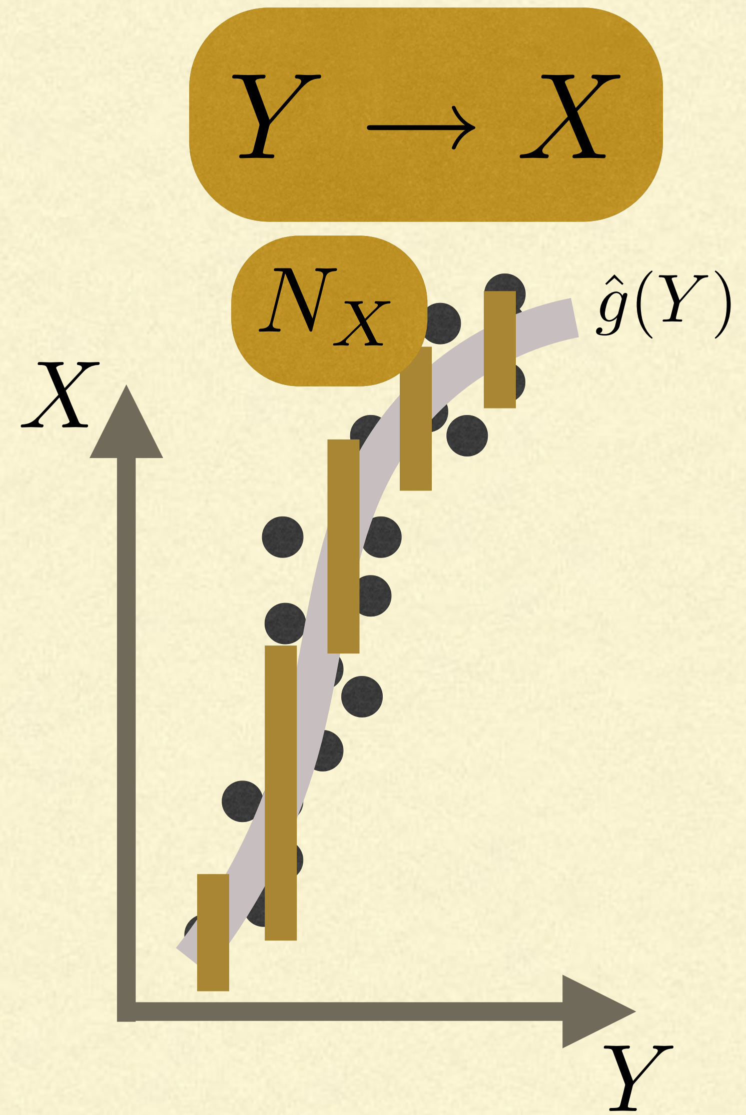




ANM: 3. COMPUTE RESIDUALS



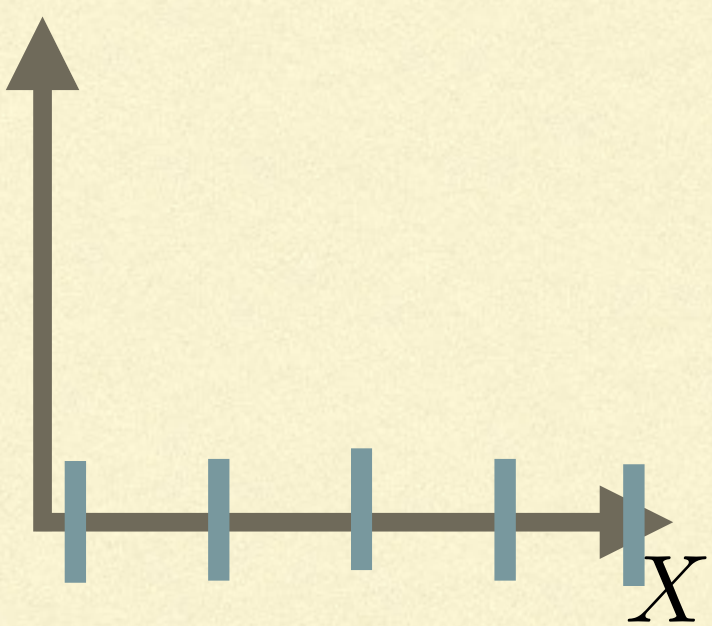
training data
test data



ANM: 4. DEPENDENCE TEST

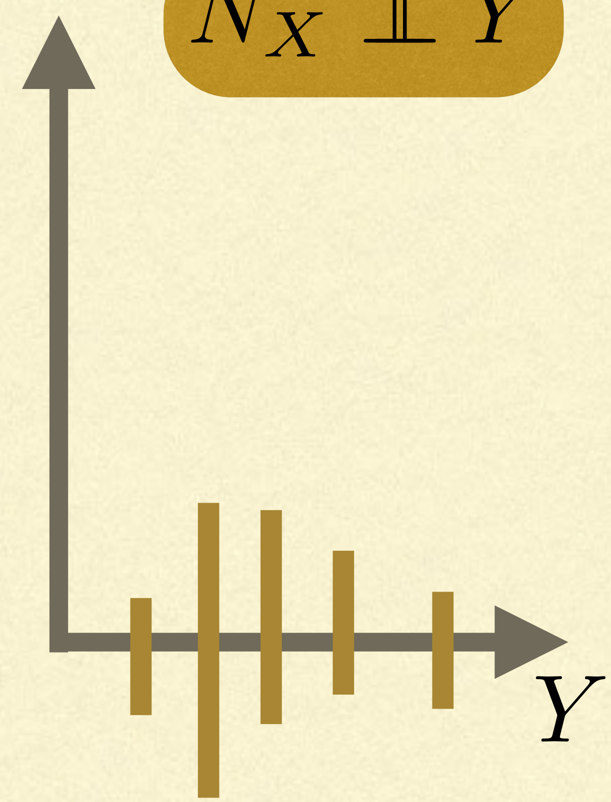
$X \rightarrow Y$

$N_Y \perp\!\!\!\perp X$



$Y \rightarrow X$

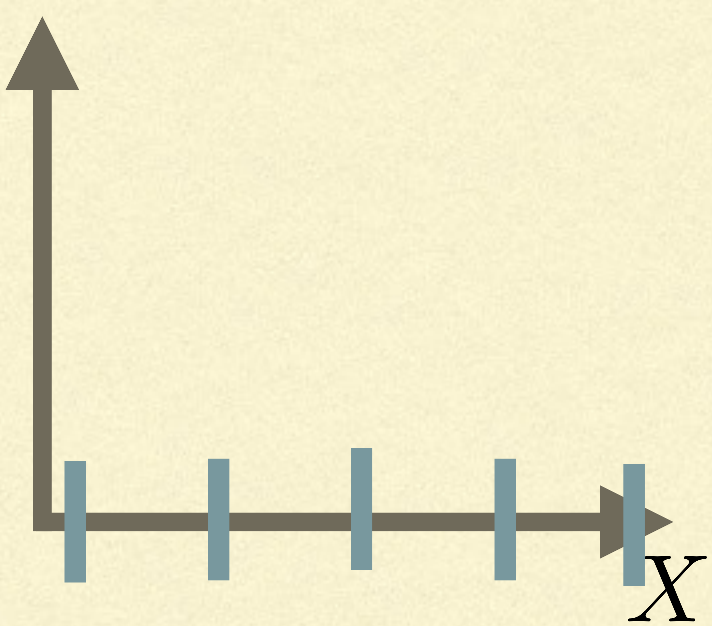
$N_X \perp\!\!\!\perp Y$



ANM: 4. DEPENDENCE TEST

$X \to Y$

$N_Y \perp\!\!\!\perp X$



Dependence Scores

Rank Correlation

- Spearman's rho
  [Spearman, 1904]
- Kendall's tau
  [Kendall, 1970]

Kernelized

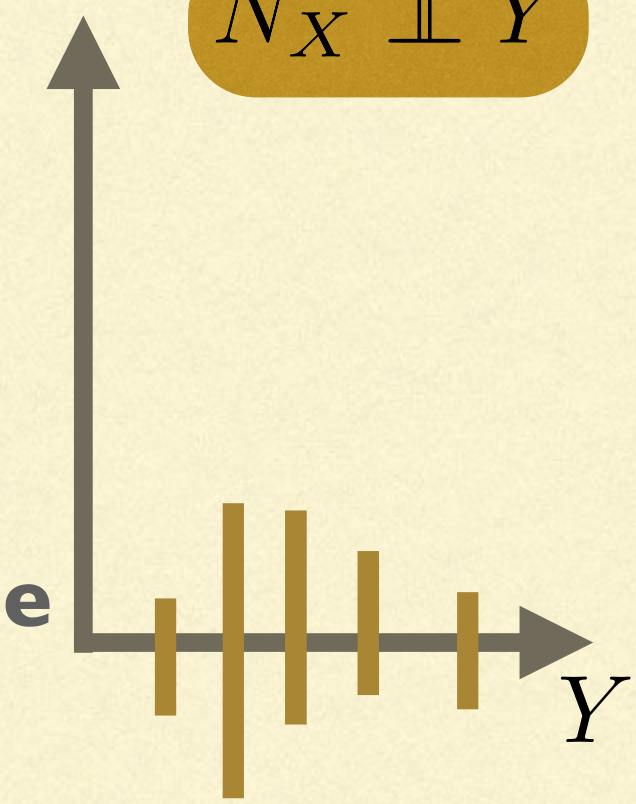
- Hilbert-Schmidt Independence Criterion
  [Gretton et al., 2005]

If Gaussian Noise

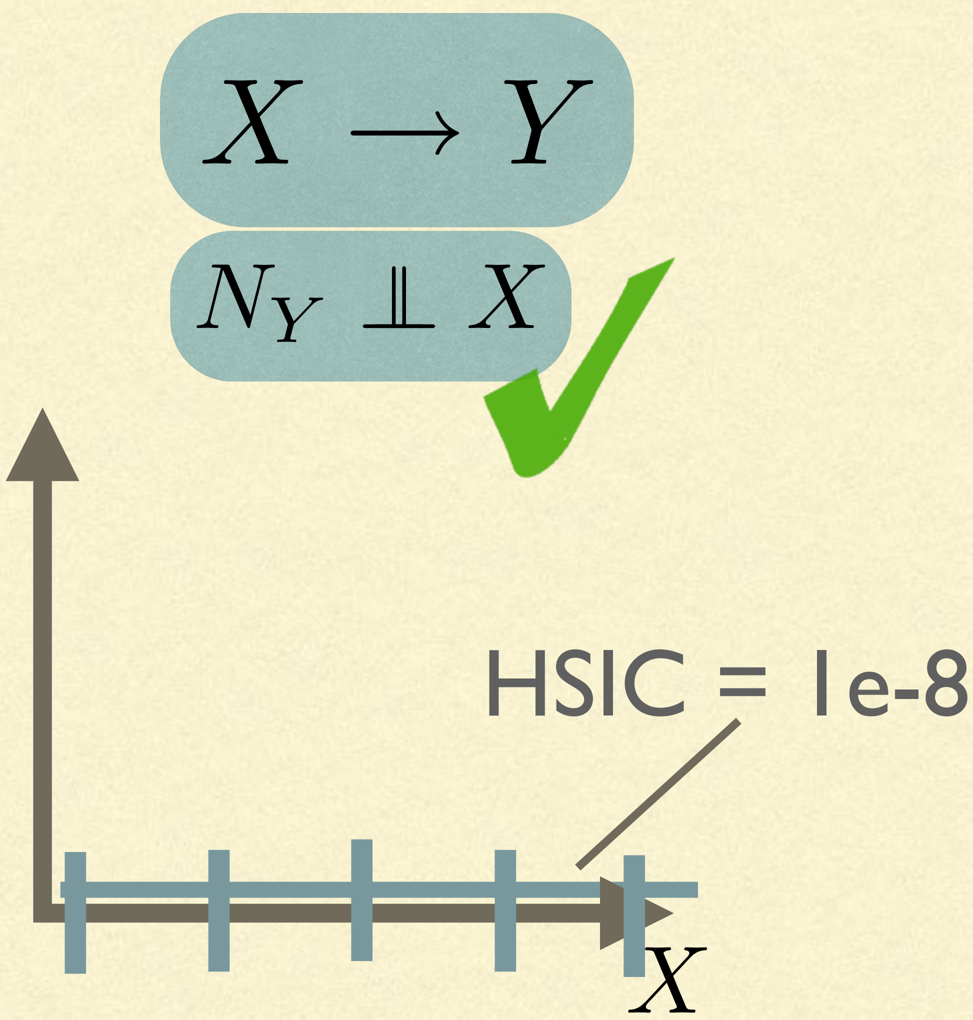
- Log-Variance
  [Buhlmann et al., 2014]

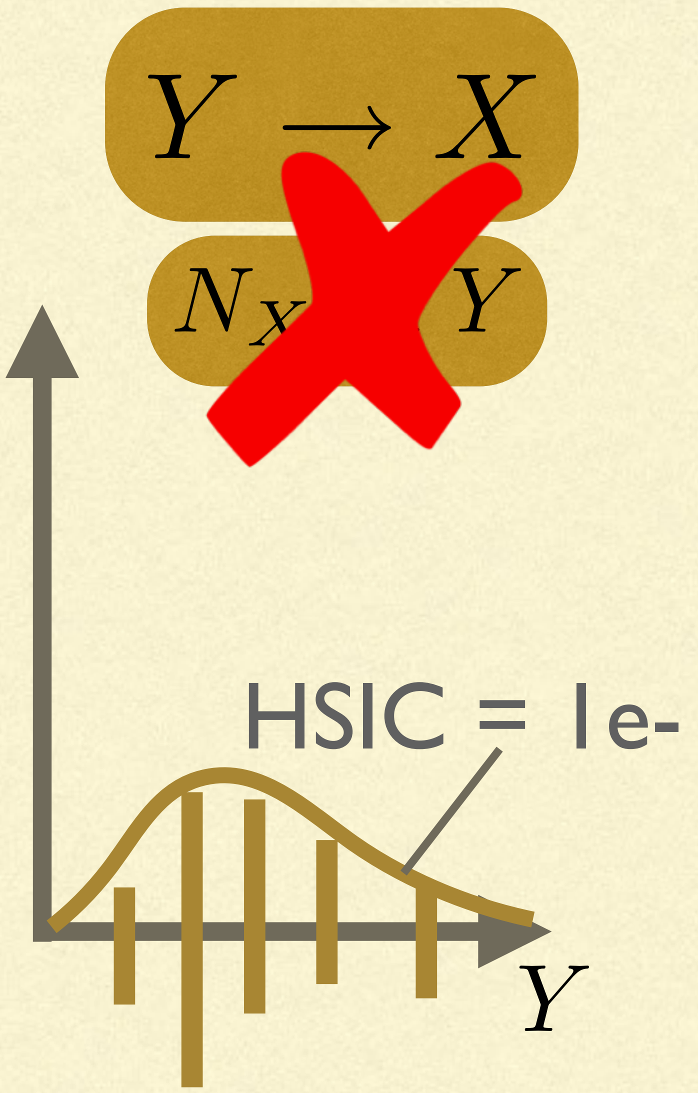
$Y \to X$

$N_X \perp\!\!\!\perp Y$



ANM: 5. CAUSAL DIRECTION



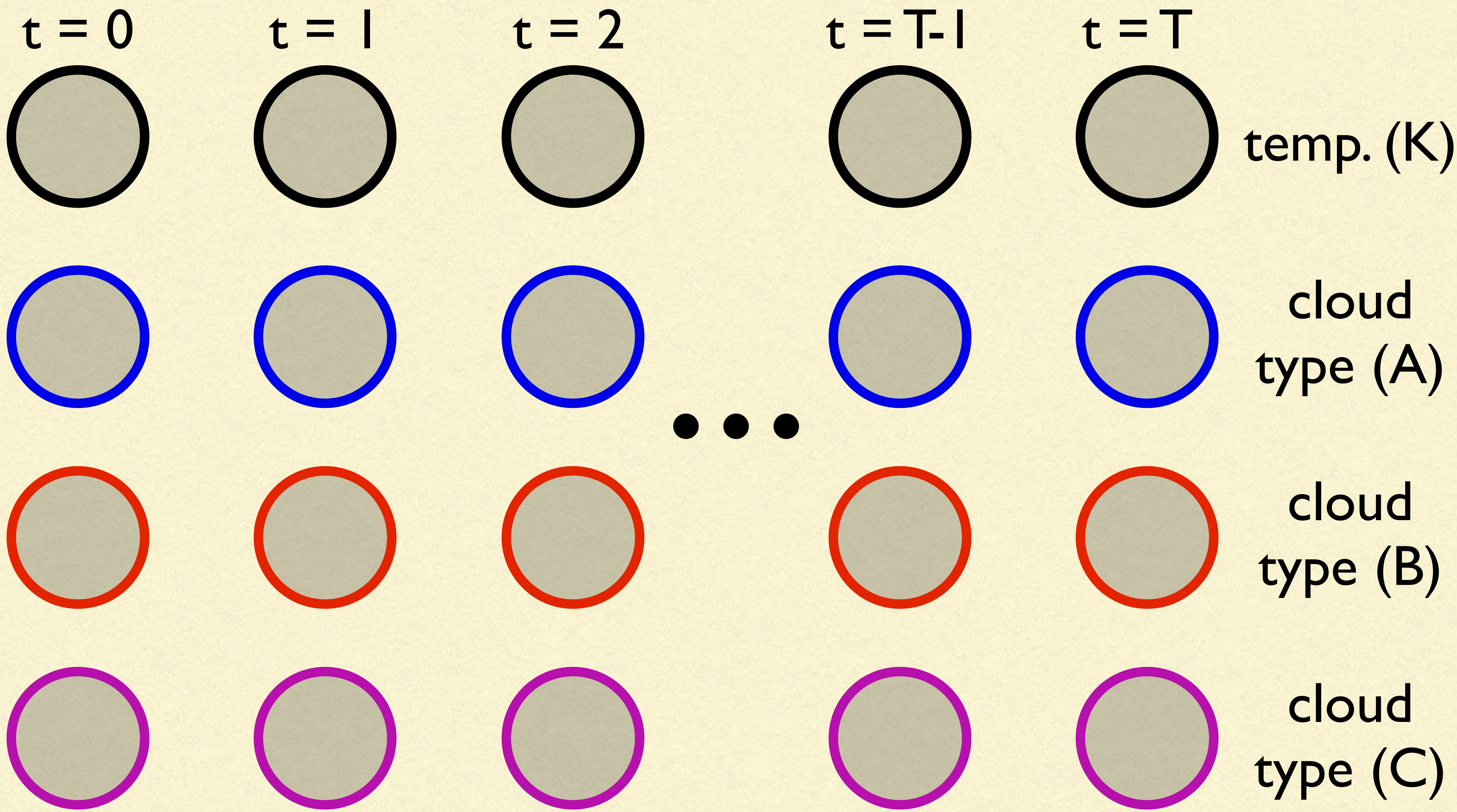


HOW DO WE GET A GRAPH? (IN A TIME-SERIES)

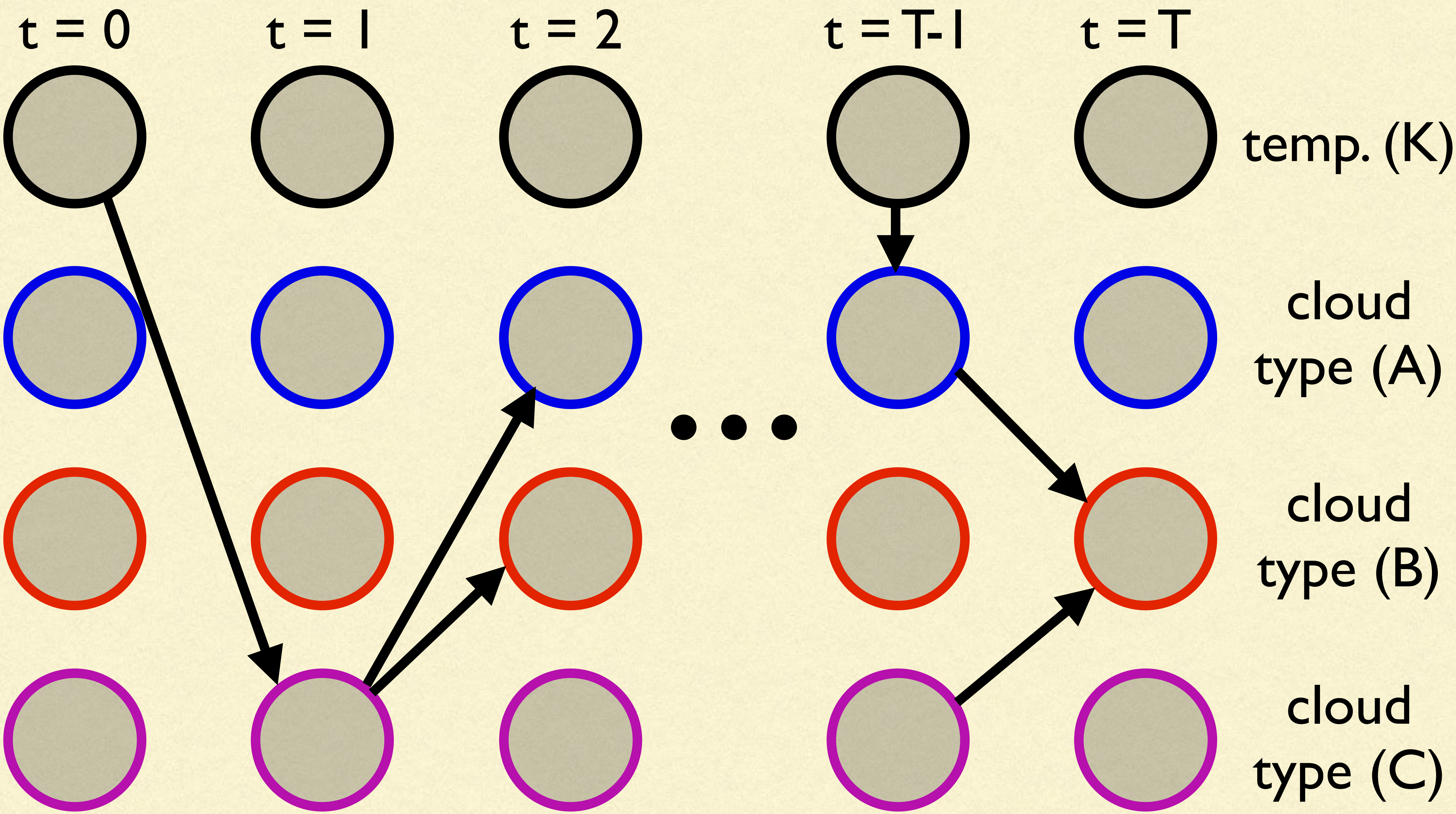
Way 3: Granger Causality

[Granger, 1969]

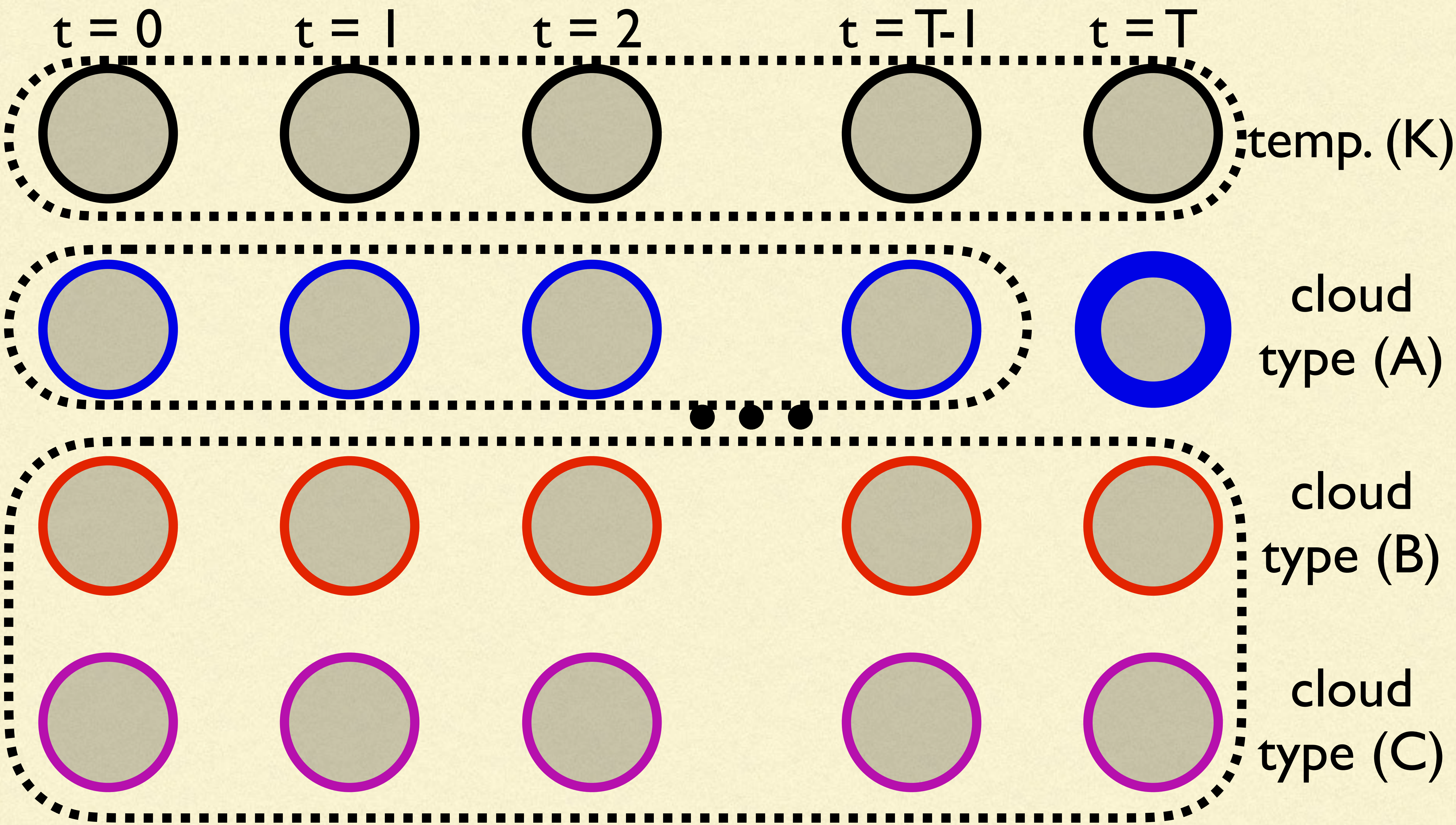
GRANGER CAUSALITY



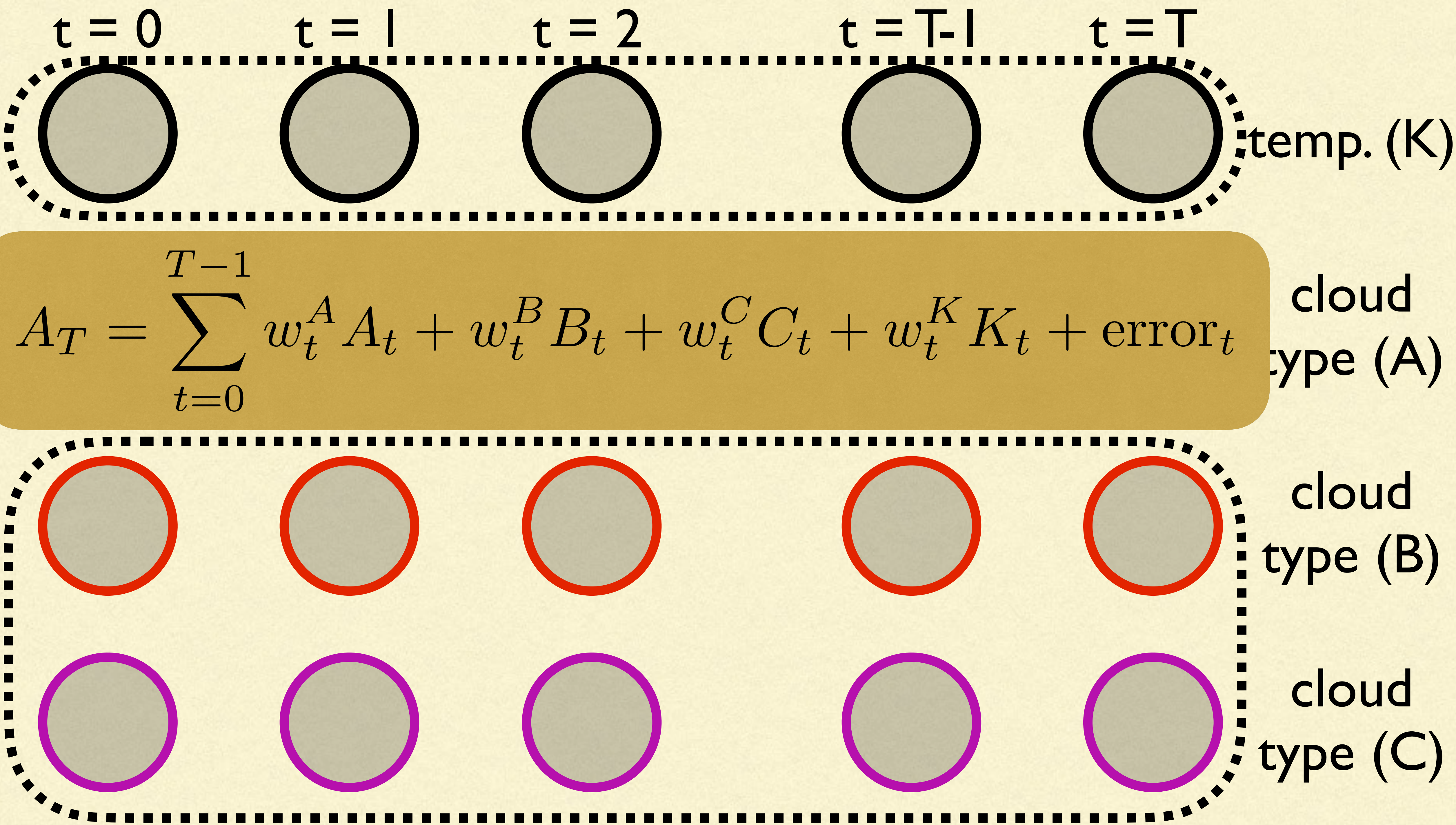
GRANGER CAUSALITY



GRANGER CAUSALITY

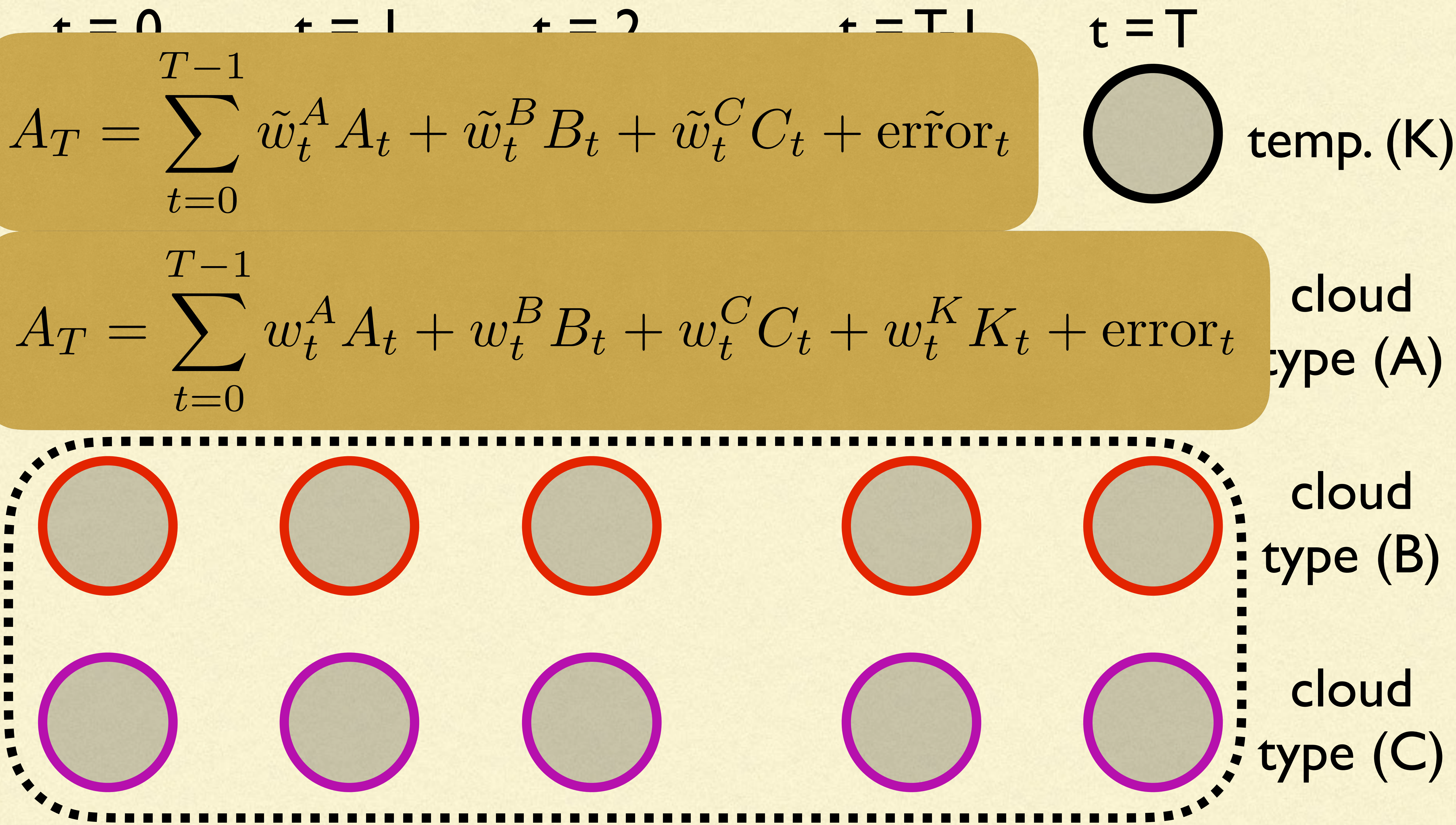


GRANGER CAUSALITY



$$A_T = \sum_{t=0}^{T-1} w_t^A A_t + w_t^B B_t + w_t^C C_t + w_t^K K_t + \mathrm{error}_t$$

GRANGER CAUSALITY



$$A_T = \sum_{t=0}^{T-1} \tilde{w}_t^A A_t + \tilde{w}_t^B B_t + \tilde{w}_t^C C_t + \tilde{\text{error}}_t$$

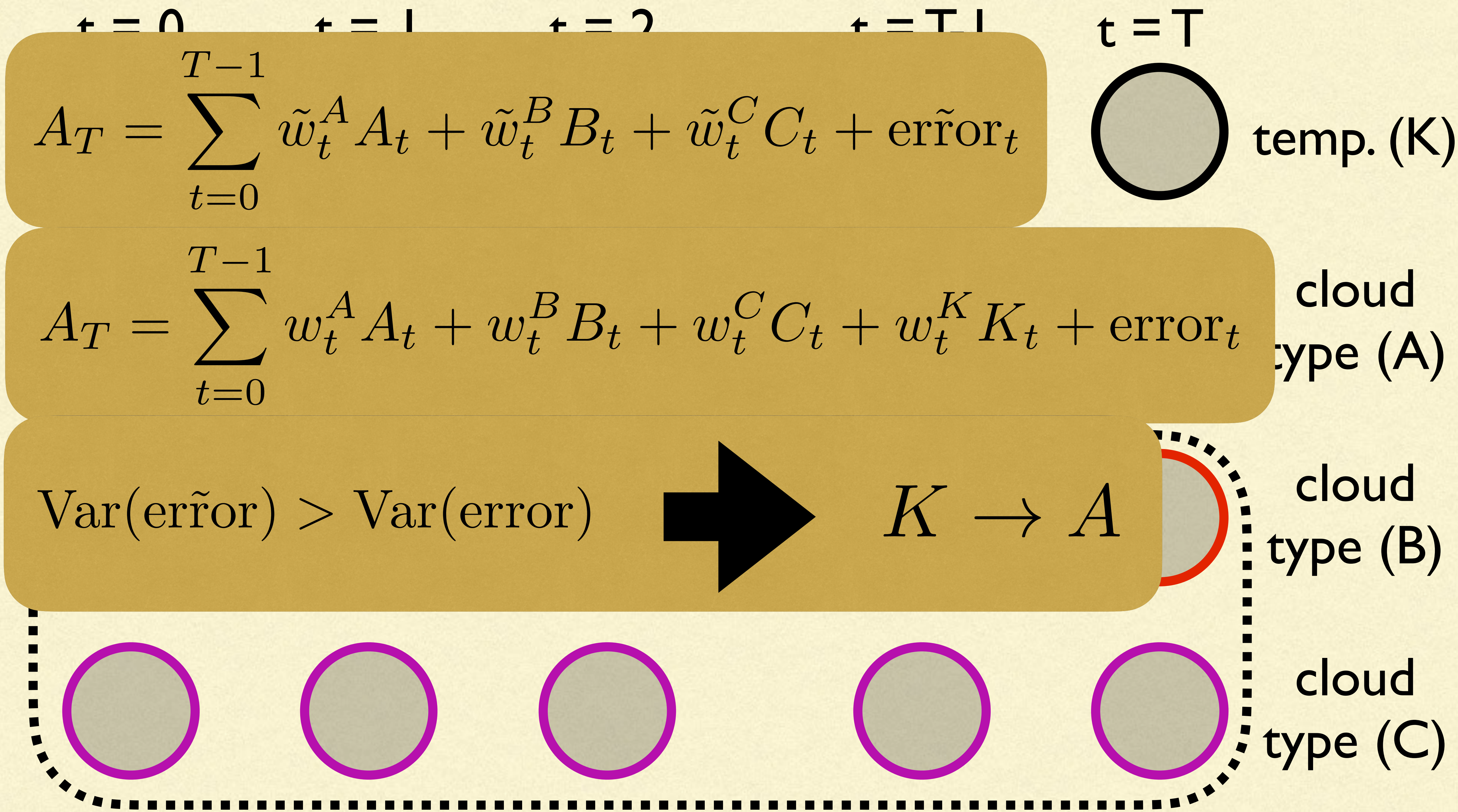
$$A_T = \sum_{t=0}^{T-1} w_t^A A_t + w_t^B B_t + w_t^C C_t + w_t^K K_t + \text{error}_t$$

GRANGER CAUSALITY

$$A_T = \sum_{t=0}^{T-1} \tilde{w}_t^A A_t + \tilde{w}_t^B B_t + \tilde{w}_t^C C_t + \tilde{\text{error}}_t$$

$$A_T = \sum_{t=0}^{T-1} w_t^A A_t + w_t^B B_t + w_t^C C_t + w_t^K K_t + \text{error}_t$$

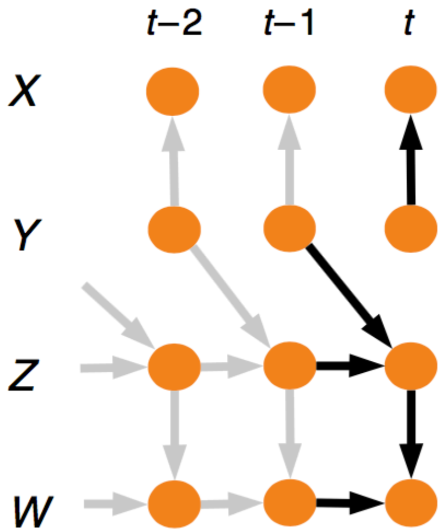
$$\text{Var}(\tilde{\text{error}}) > \text{Var}(\text{error}) \quad \Rightarrow \quad K \to A$$



HOW DO THESE COMPARE?

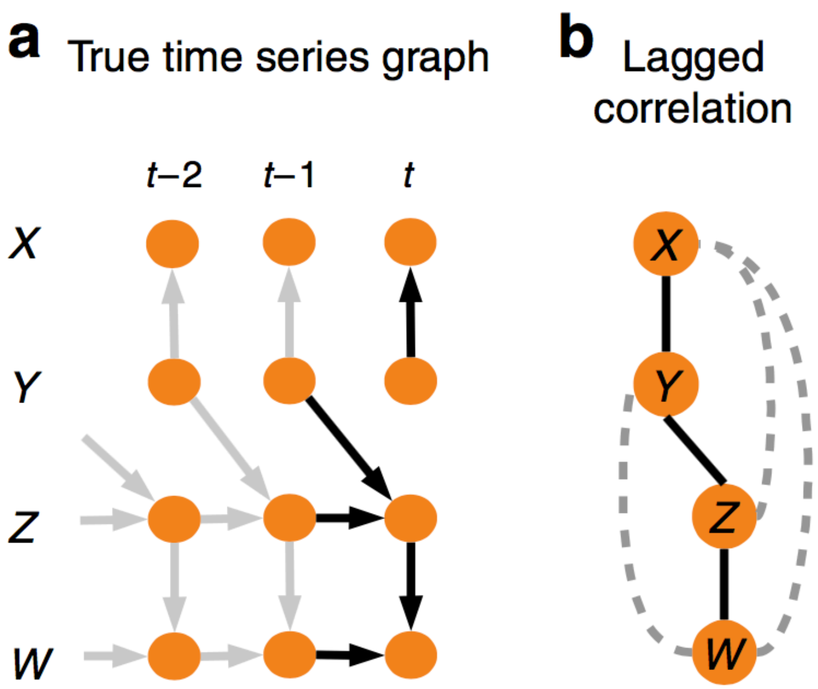
[RUNGE ET AL., 2019]

a True time series graph



$$X_t = aY_t + E_t^X$$
$$Y_t = E_t^Y$$
$$Z_t = bZ_{t-1} + cY_{t-1} + E_t^Z$$
$$W_t = dW_{t-1} + eZ_t + E_t^W,$$

HOW DO THESE COMPARE?
[RUNGE ET AL., 2019]



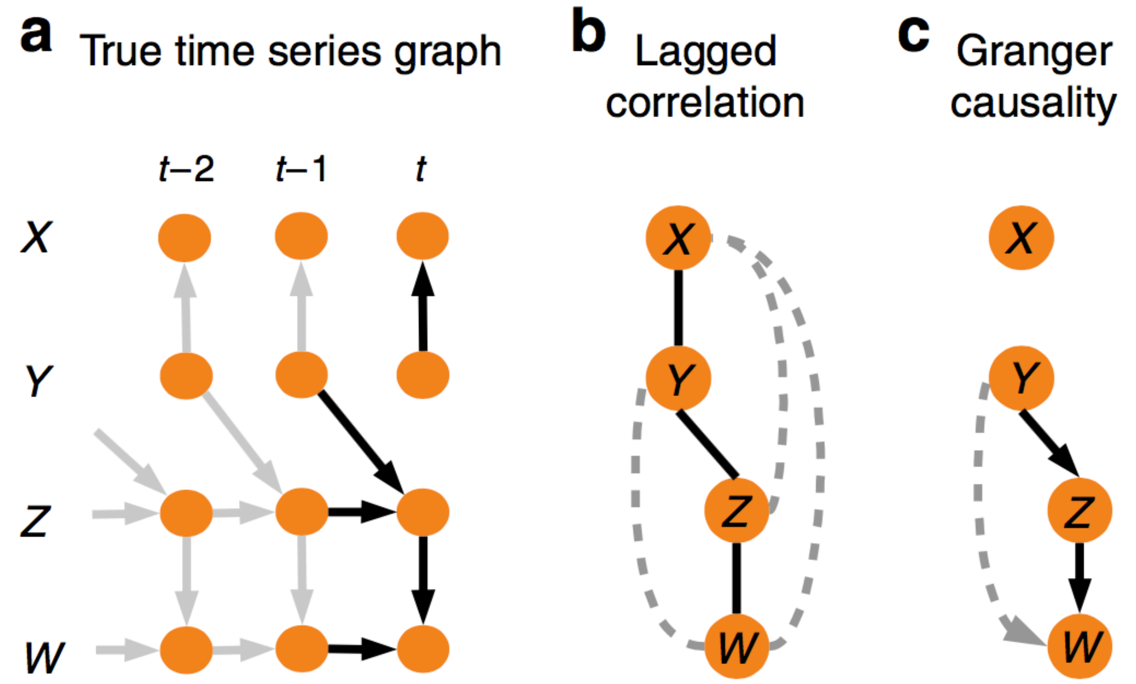
$$X_t = aY_t + E_t^X$$
$$Y_t = E_t^Y$$
$$Z_t = bZ_{t-1} + cY_{t-1} + E_t^Z$$
$$W_t = dW_{t-1} + eZ_t + E_t^W,$$

HOW DO THESE COMPARE?

[RUNGE ET AL., 2019]

Way 3

[Granger, 1969]



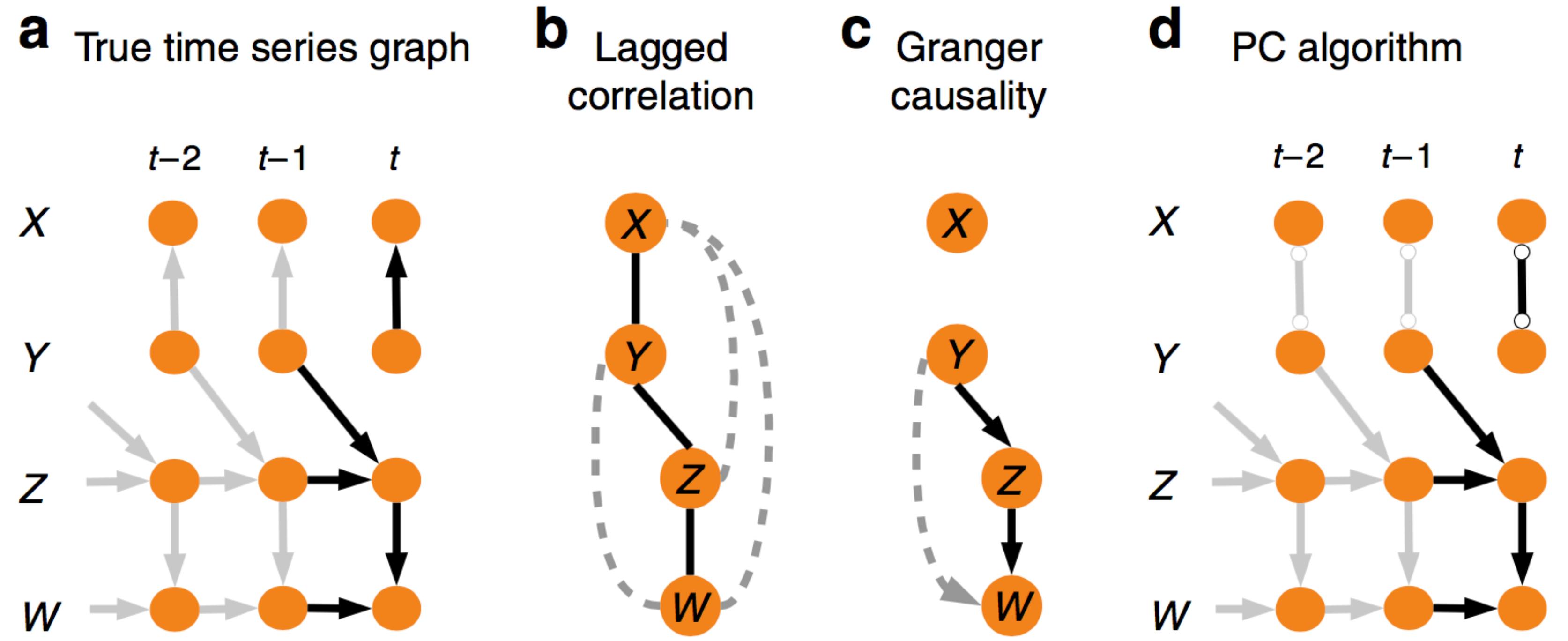
$$X_t = aY_t + E_t^X$$
$$Y_t = E_t^Y$$
$$Z_t = bZ_{t-1} + cY_{t-1} + E_t^Z$$
$$W_t = dW_{t-1} + eZ_t + E_t^W,$$

HOW DO THESE COMPARE?

[RUNGE ET AL., 2019]

Way 3 [Granger, 1969]

Way 1 [Pearl, 2000]



$$
\begin{aligned}
X_t &= aY_t + E_t^X \\
Y_t &= E_t^Y \\
Z_t &= bZ_{t-1} + cY_{t-1} + E_t^Z \\
W_t &= dW_{t-1} + eZ_t + E_t^W,
\end{aligned}
$$

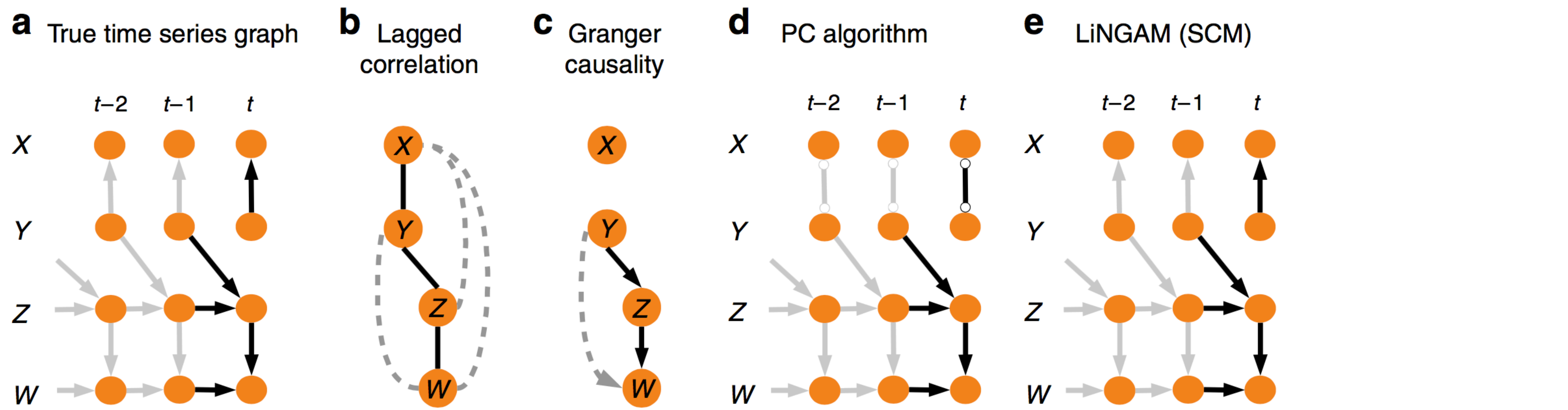
HOW DO THESE COMPARE?

[RUNGE ET AL., 2019]

Way 3 [Granger, 1969]

Way 1 [Pearl, 2000]

Way 2 [Hoyer et al., 2009]



$$X_t = aY_t + E_t^X$$
$$Y_t = E_t^Y$$
$$Z_t = bZ_{t-1} + cY_{t-1} + E_t^Z$$
$$W_t = dW_{t-1} + eZ_t + E_t^W,$$

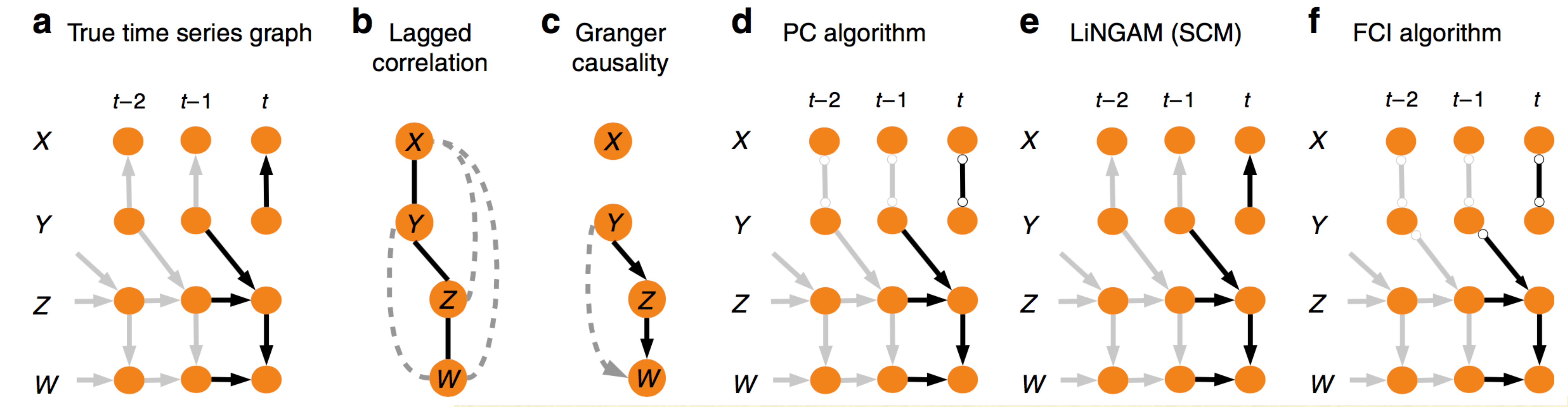
HOW DO THESE COMPARE?

[RUNGE ET AL., 2019]

Way 3
[Granger, 1969]

Way 1
[Pearl, 2000]

Way 2
[Hoyer et al., 2009]



$$
\begin{aligned}
X_t &= aY_t + E_t^X \\
Y_t &= E_t^Y \\
Z_t &= bZ_{t-1} + cY_{t-1} + E_t^Z \\
W_t &= dW_{t-1} + eZ_t + E_t^W,
\end{aligned}
$$

1. WHAT IS CAUSALITY?

2. HOW DO WE ESTIMATE **CAUSAL EFFECTS (INTERVENTIONS)**?

3. HOW DO WE OBTAIN A **CAUSAL GRAPH**?

4. ANOTHER CAUSAL QUANTITY: **COUNTERFACTUALS**

COUNTERFACTUALS

Driving home, you hit a fork:

COUNTERFACTUALS

Driving home, you hit a fork:

option 1:
highway

option 2:
country
road

COUNTERFACTUALS

Driving home, you hit a fork:

option 1:
highway

option 2:
country
road

COUNTERFACTUALS

While driving, you get stuck behind this:

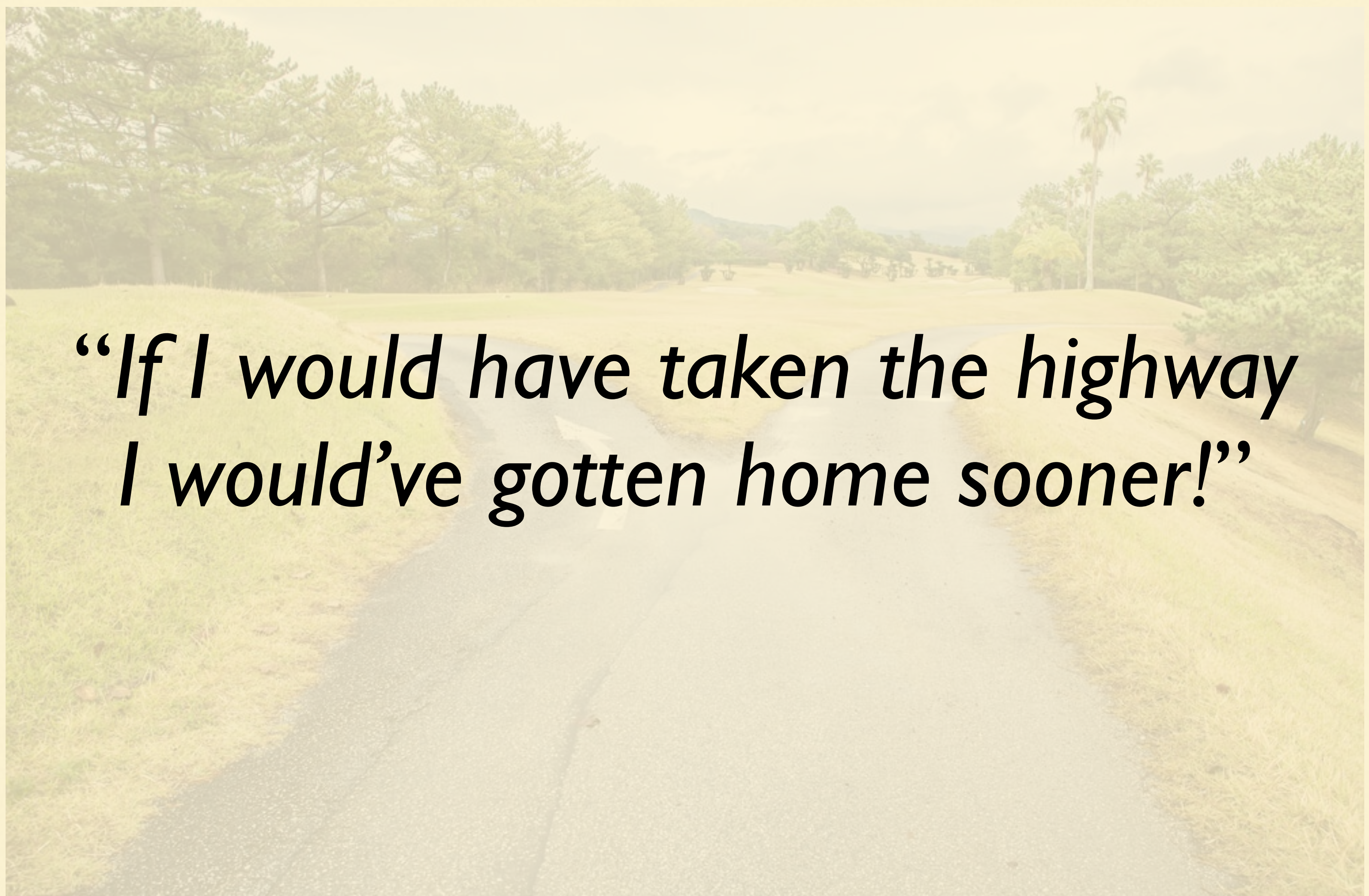
option 2:
country
road

it takes 1 hour to get home…

COUNTERFACTUALS

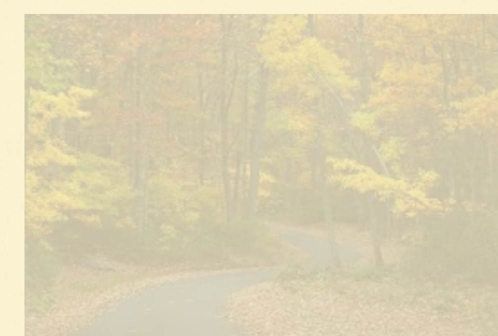
Driving home, you hit a fork:

option 1:
highway



"If I would have taken the highway I would've gotten home sooner!"

option 2:
country road



COUNTERFACTUALS

"If I would have taken the highway
I would've gotten home sooner!"

Can we use **interventions** to formalize this?

COUNTERFACTUALS

"If I would have taken the highway I would've gotten home sooner!"

Can we use **interventions** to formalize this?

$$\mathbb{E}[\text{Time} \mid do(\text{Road} = \text{highway}), \text{Time} = 1 \text{ hour}]$$

????!!??

COUNTERFACTUALS

"If I would have taken the highway I would've gotten home sooner!"

~~Can we use **interventions** to formalize this?~~

$$\underbrace{\mathbb{E}[\text{Time}}_{\textbf{hypothetical}\text{ driving time}} \mid do(\text{Road} = \text{highway}), \underbrace{\text{Time} = 1 \text{ hour}}_{\textbf{actual}\text{ driving time}}]$$

COUNTERFACTUALS

*"If I would have taken the highway
I would've gotten home sooner!"*

~~Can we use **interventions** to formalize this?~~

$$\mathbb{E}[\text{Time} \mid do(\text{Road} = \text{highway}), \text{Time} = 1 \text{ hour}]$$

hypothetical
driving time
$T_{R=\text{highway}}$

actual
driving time
T

COUNTERFACTUALS

"If I would have taken the highway
I would've gotten home sooner!"

~~Can we use **interventions** to formalize this?~~

$$\mathbb{E}[\text{Time} \mid do(\text{Road} = \text{highway}), \text{Time} = 1 \text{ hour}]$$

hypothetical
driving time

actual
driving time

$$\mathbb{E}[T_{R=\text{highway}} \mid R = \text{country}, T = 1 \text{ hour}]$$

counterfactual

COUNTERFACTUALS

we get to condition on the actual choice we made!

option 1:
highway

option 2:
country road

COUNTERFACTUALS

Driving home, you hit a fork:

option 1:
highway

option 2:
country
road

interventions ask about the impact of the decision before it happens

How do you compute them?

COUNTERFACTUALS

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y
male	poor			

classifier

$$\hat{Y}(G, L)$$

3-STEP PROCEDURE

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y
male	poor			

classifier

$$\hat{Y}(G, L)$$

1. COMPUTE UNOBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y
male	poor			

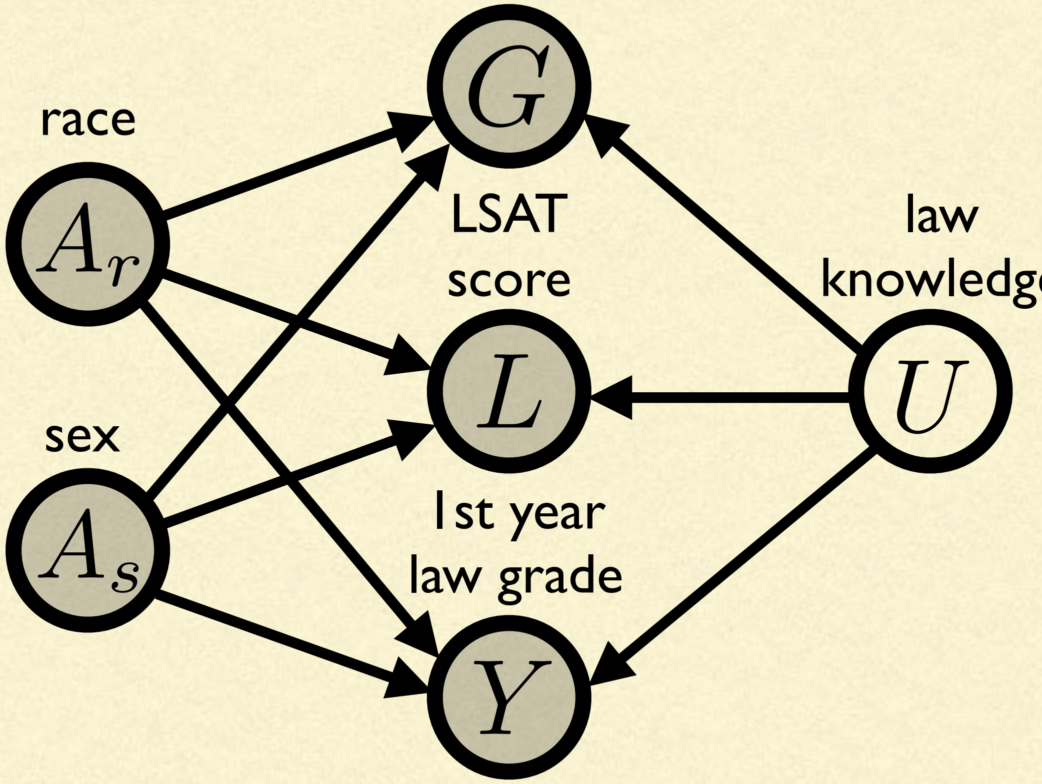
classifier

$$\hat{Y}(G, L)$$

1. COMPUTE UNOBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y
male	poor			



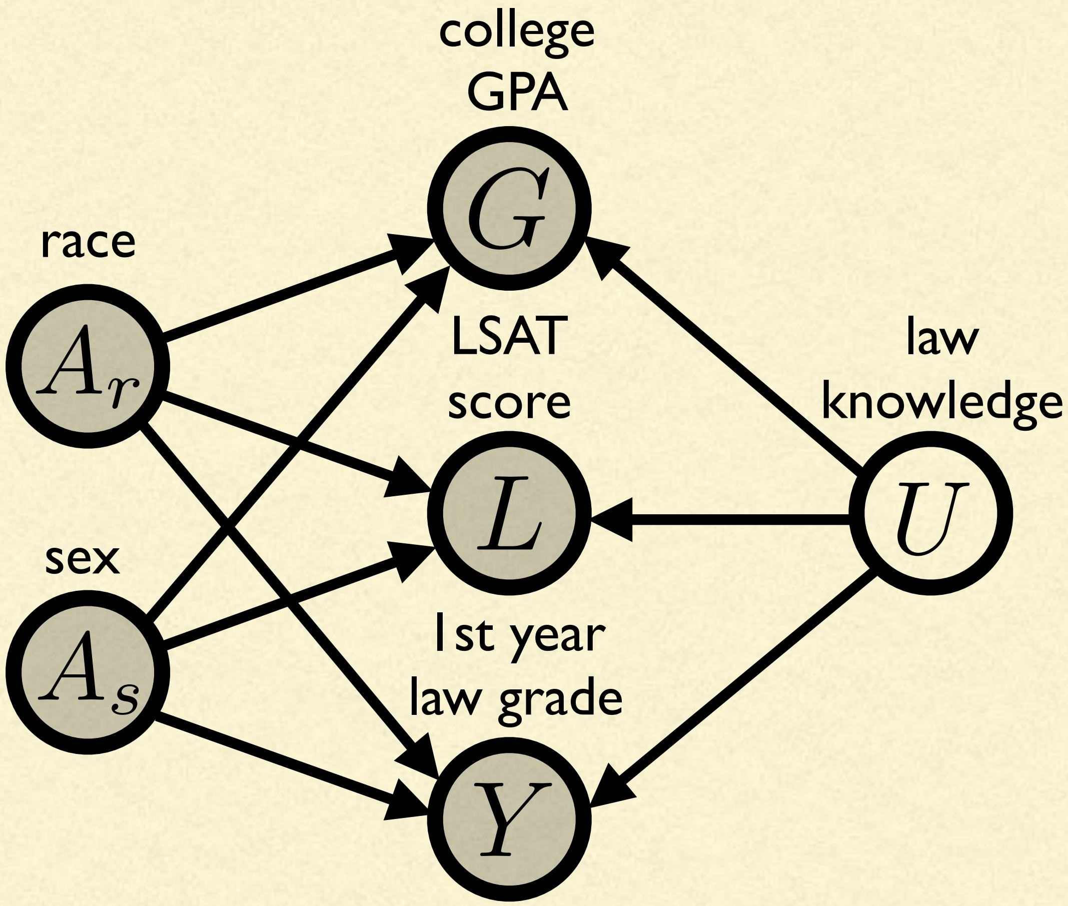
classifier

$$\hat{Y}(G, L)$$

1. COMPUTE UNOBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y	U
male	poor				



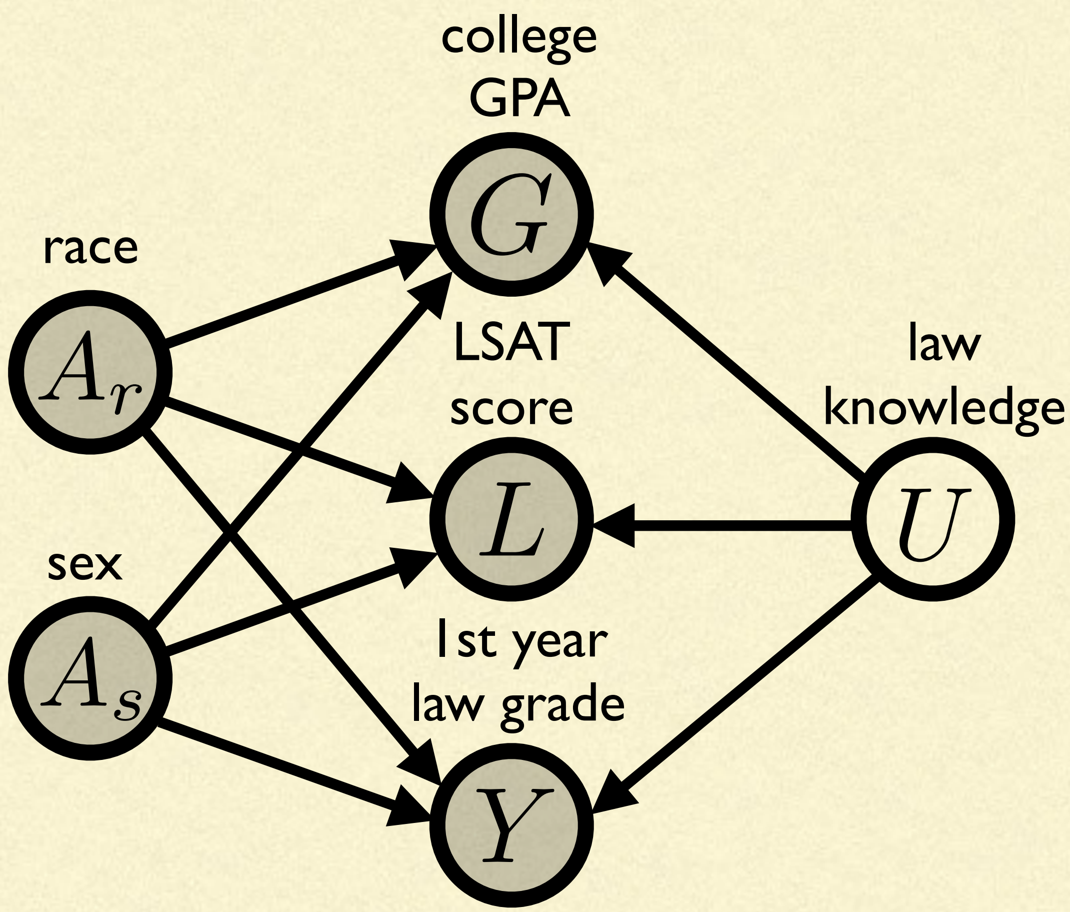
classifier

$$\hat{Y}(G, L)$$

2. CHANGE A

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	A_r	G	L	Y	U
male	poor				



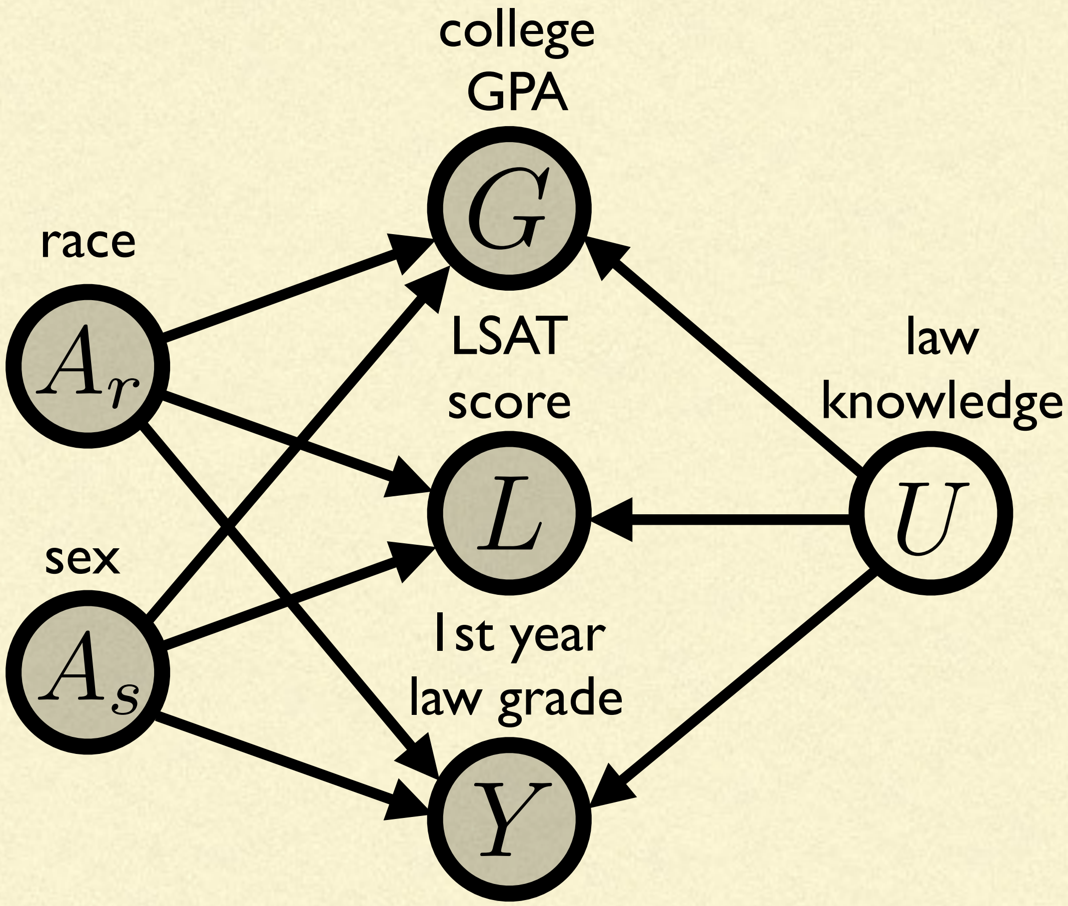
classifier

$$\hat{Y}(G, L)$$

2. CHANGE A

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	$A_r \leftarrow a'$	G	L	Y	U
male	rich				



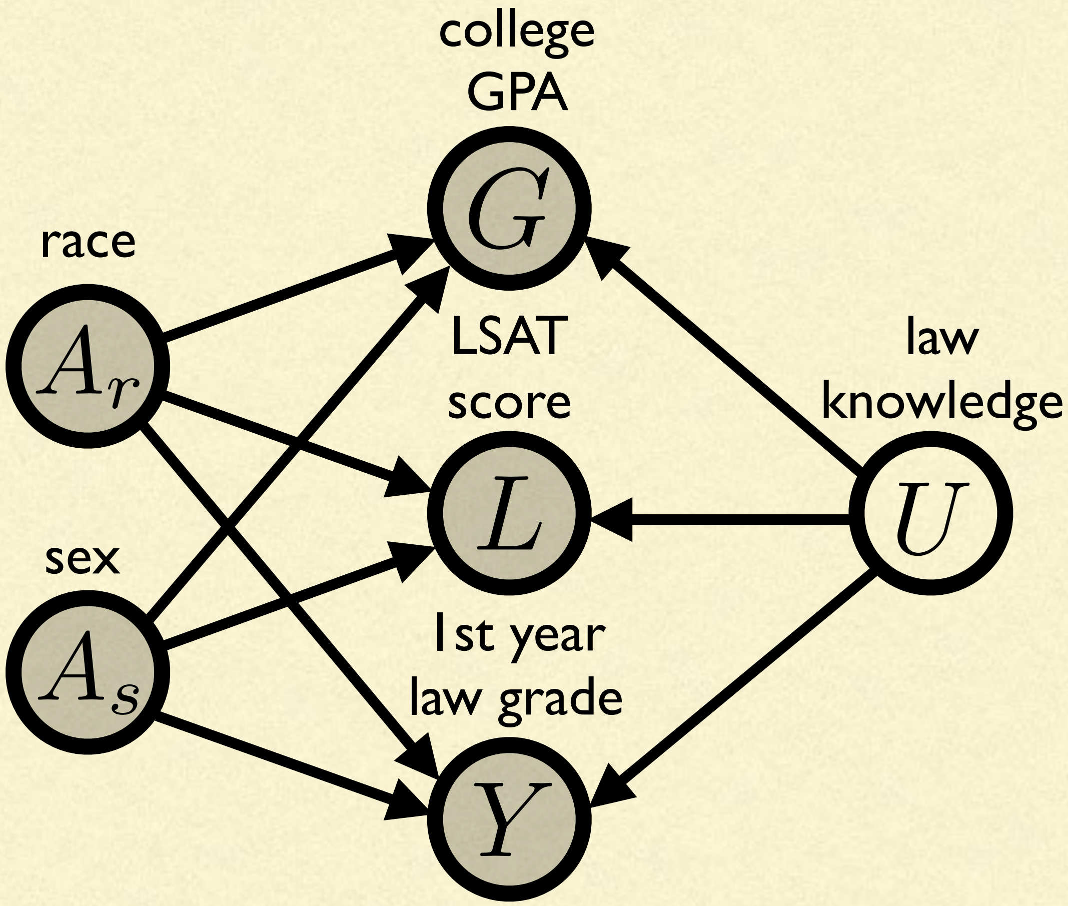
classifier

$$\hat{Y}(G, L)$$

3. RECOMPUTE OBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

A_s	$A_r \leftarrow a'$	G	L	Y	U
male	rich				



classifier

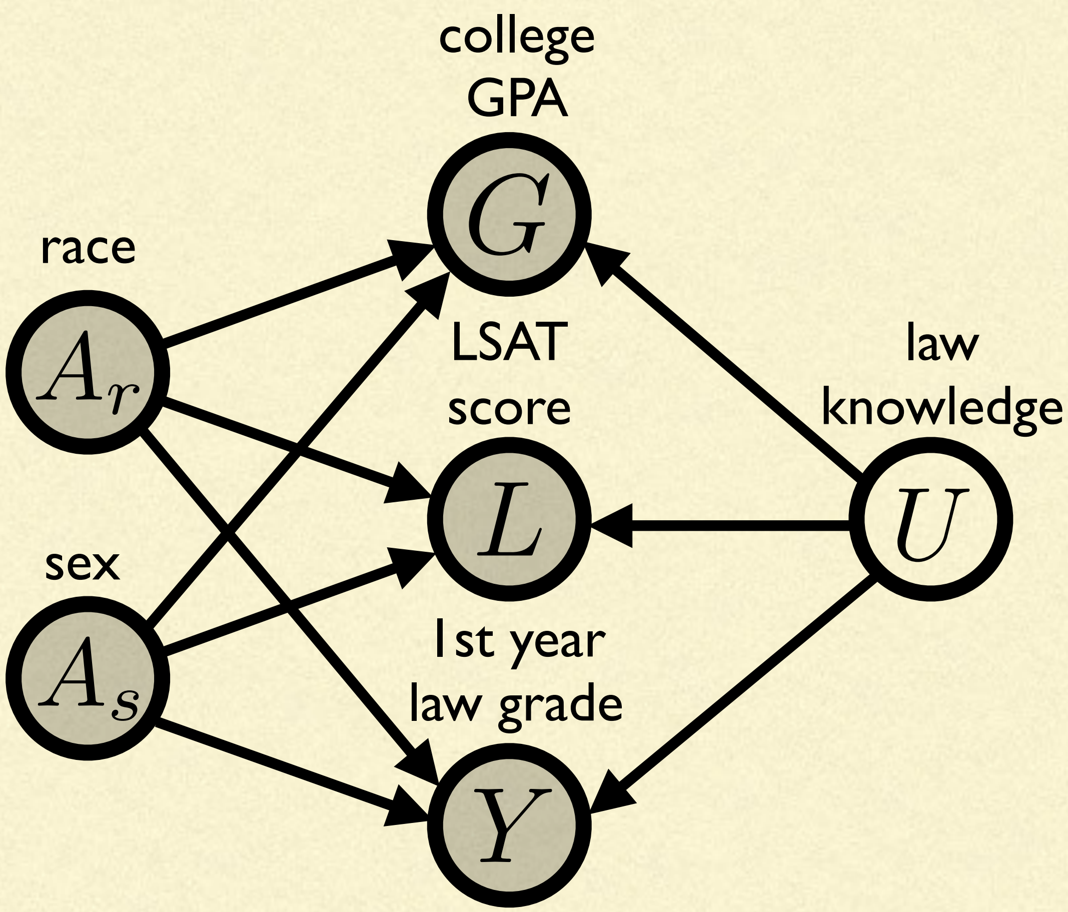
$$\hat{Y}(G, L)$$

3. RECOMPUTE OBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

$$\frac{A_s \quad A_r \leftarrow a' \; G_{A_r \leftarrow a'} \quad L_{A_r \leftarrow a'} \; Y_{A_r \leftarrow a'} \; U}{}$$

male rich



classifier

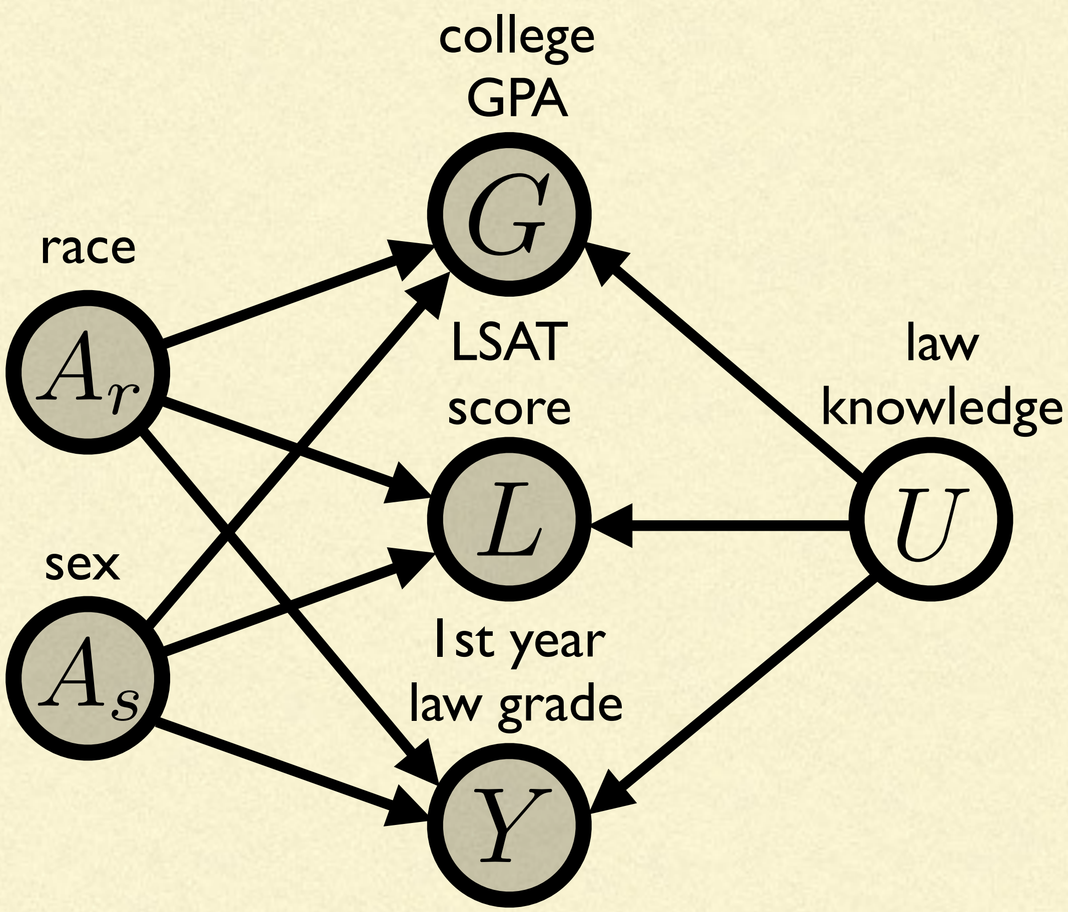
$$\hat{Y}(G, L)$$

3. RECOMPUTE OBSERVED VARIABLES IN CAUSAL MODEL

[Pearl, 2000; Pearl, 2009; Pearl et al., 2016]

$$A_s \quad A_r \leftarrow a' \; G_{A_r \leftarrow a'} \quad L_{A_r \leftarrow a'} \; Y_{A_r \leftarrow a'} \; U$$

male rich



classifier

$$\hat{Y}(G, L)$$

TAKE-AWAYS

- Causality allows us to quantify how quantities change when others change!

intervening

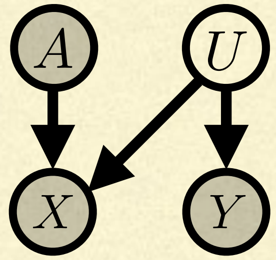
$$\boxed{\mathbb{P}(Y \mid X \text{ set to be } x)}$$

TAKE-AWAYS

- Causality allows us to quantify how quantities change when others change!
- Idea: Use **directed acyclic graphs (DAGs)** to describe network of causes and effects

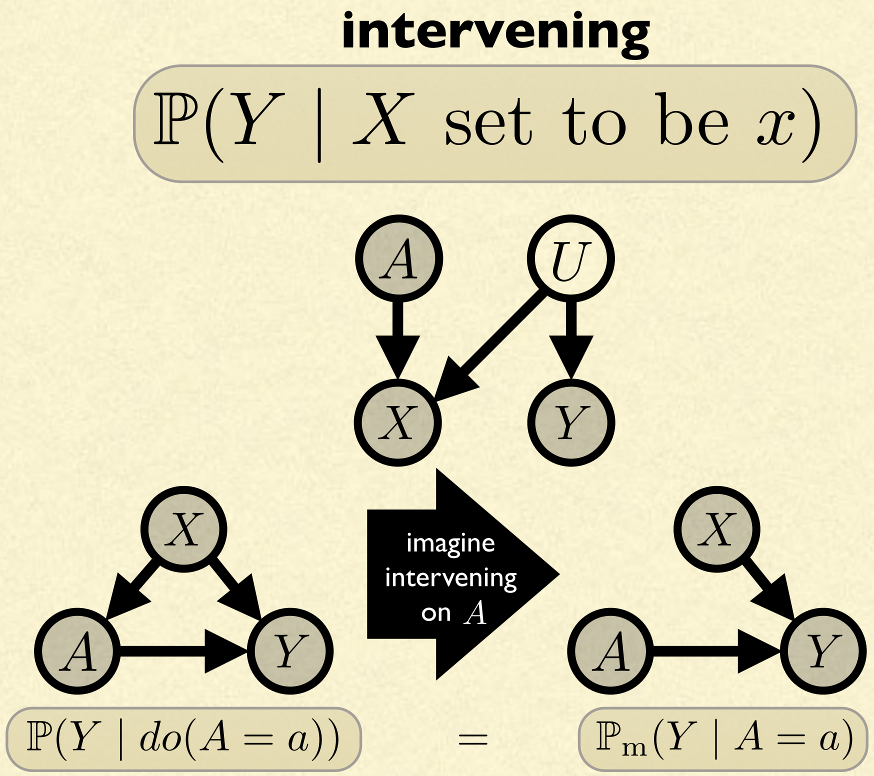
intervening

$\mathbb{P}(Y \mid X \text{ set to be } x)$



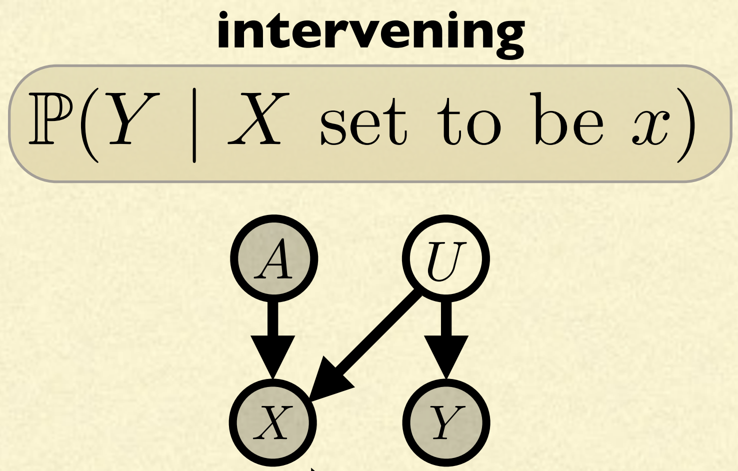
TAKE-AWAYS

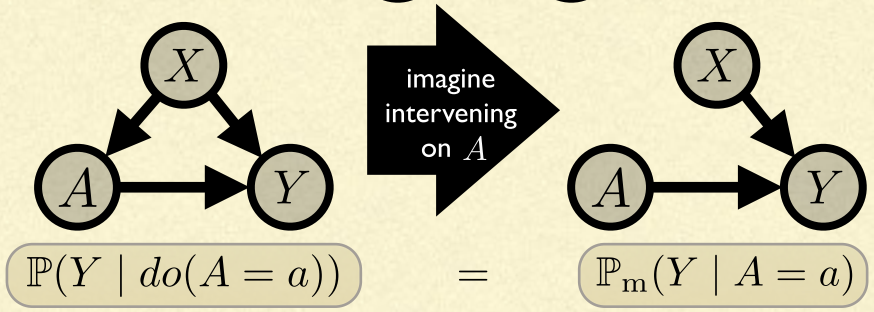
- Causality allows us to quantify how quantities change when others change!
- Idea: Use **directed acyclic graphs (DAGs)** to describe network of causes and effects
- We can estimate **how changing some variables affects others**! (sometimes we need randomized control trials)

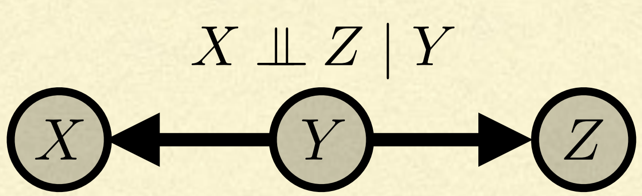


TAKE-AWAYS

- Causality allows us to quantify how quantities change when others change!
- Idea: Use **directed acyclic graphs (DAGs)** to describe network of causes and effects
- We can estimate **how changing some variables affects others**! (sometimes we need randomized control trials)
- We can discover causal graphs **but we always need additional assumptions**!







TAKE-AWAYS

- Causality allows us to quantify how quantities change when others change!
- Idea: Use **directed acyclic graphs (DAGs)** to describe network of causes and effects
- We can estimate **how changing some variables affects others**! (sometimes we need randomized control trials)
- We can discover causal graphs **but we always need additional assumptions**!
- Causal models also allow us to express **counterfactuals**: "what if I had done X?"