Causal Learning for AI Research

Brief Introduction & Research Trends

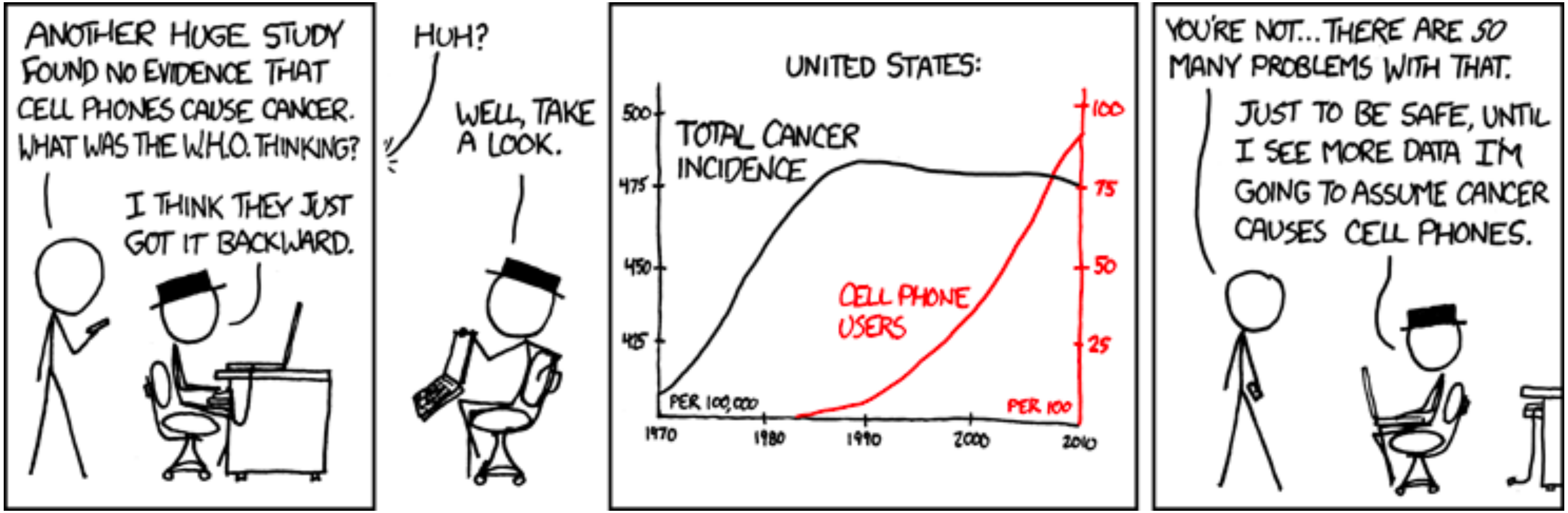
Sungbin Lim

kakaobrain

Preliminaries

Motivation & Notation

Causality ≠ Correlation



다른 대규모 연구에 따르면 핸드폰이 암을 유발하는건 근거가 없대. WHO 의견은 뭐였지?

그건 그들이 거꾸로 이해한거야

뭐?

이거 봐봐

전체 암 발생율 (미국)

핸드폰 사용자수

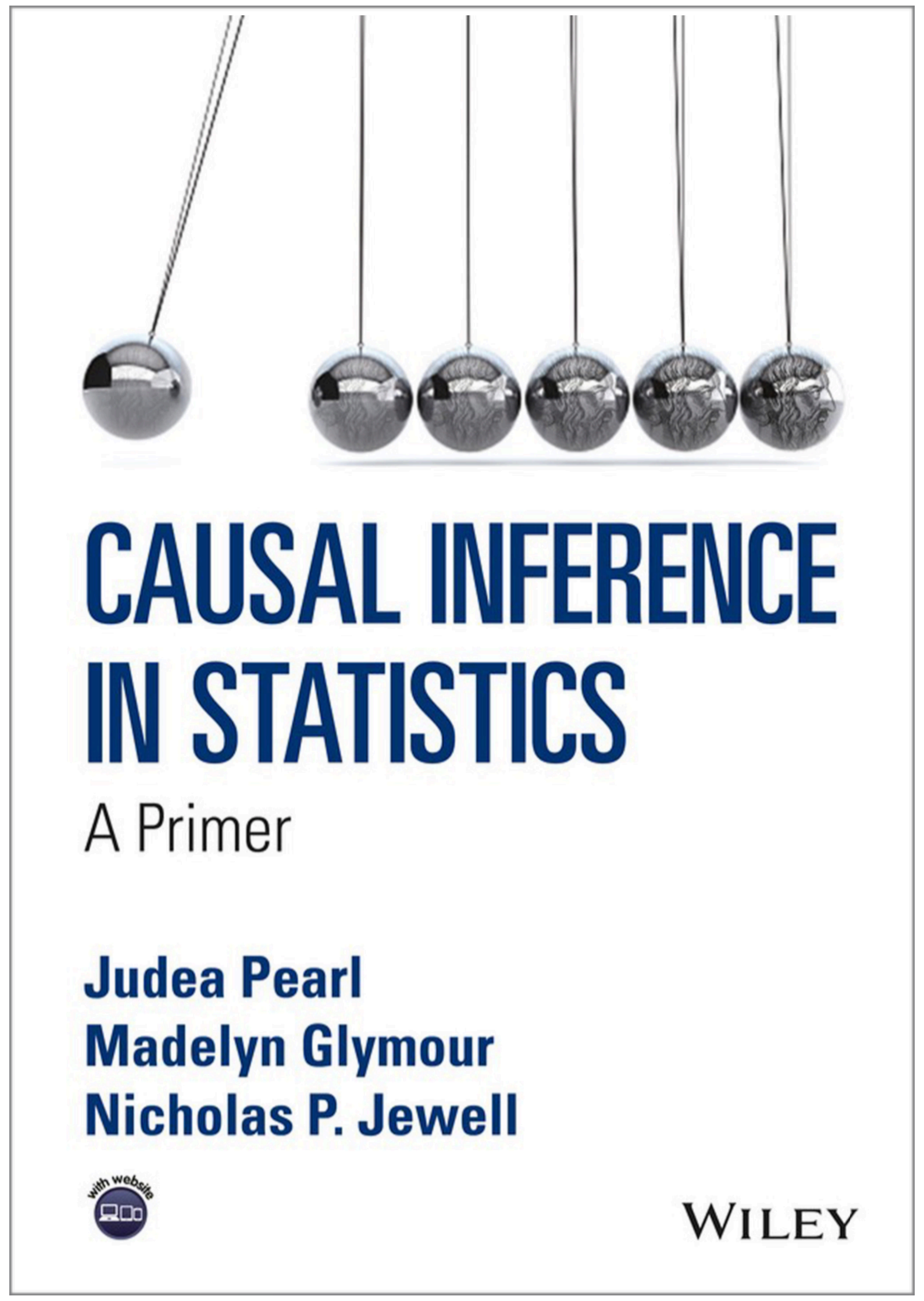
더 많은 데이터를 보기 전까진 암이 핸드폰을 유발한다고 가정하겠어.

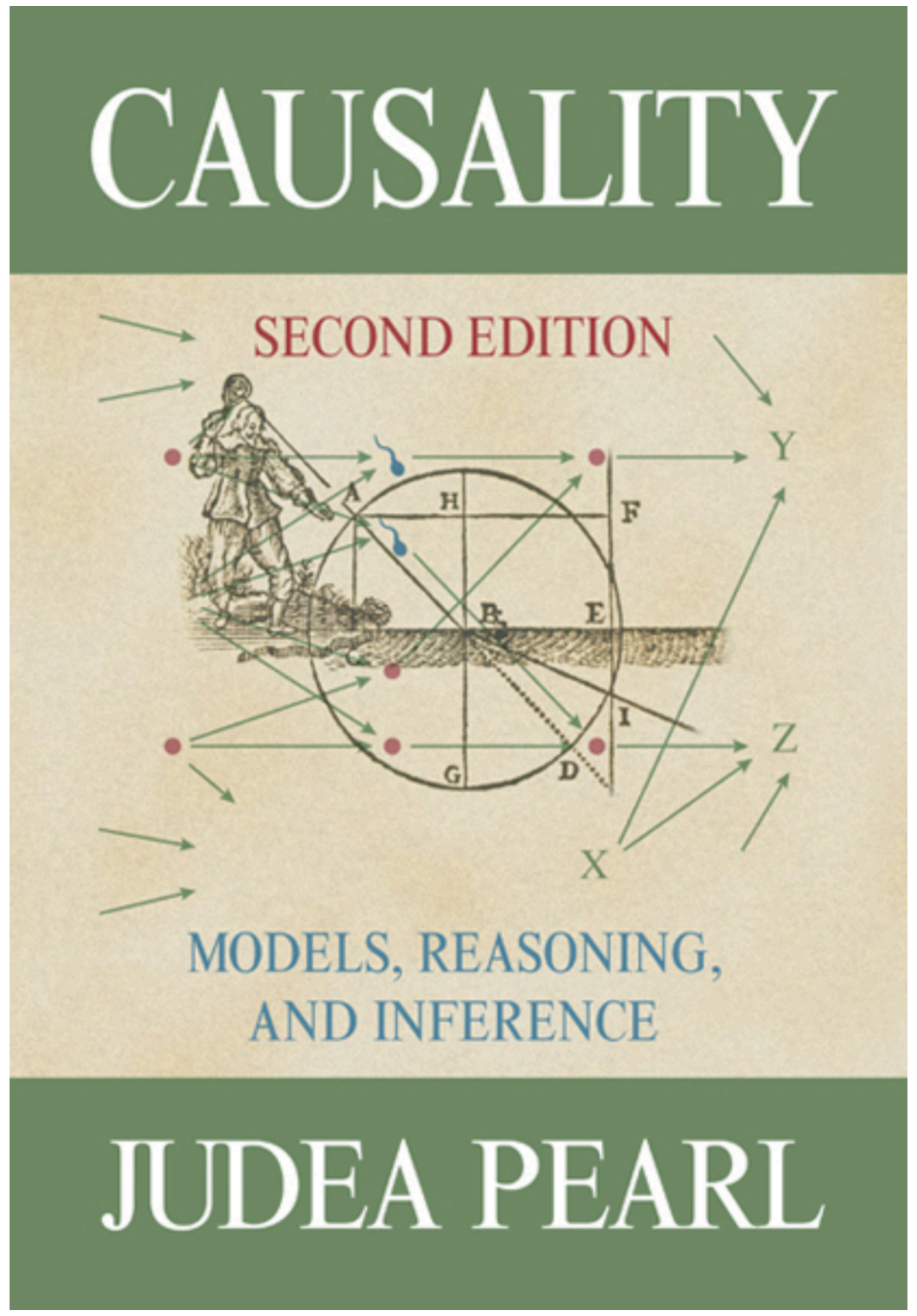
야.. 그건 좀 아니지 않냐?

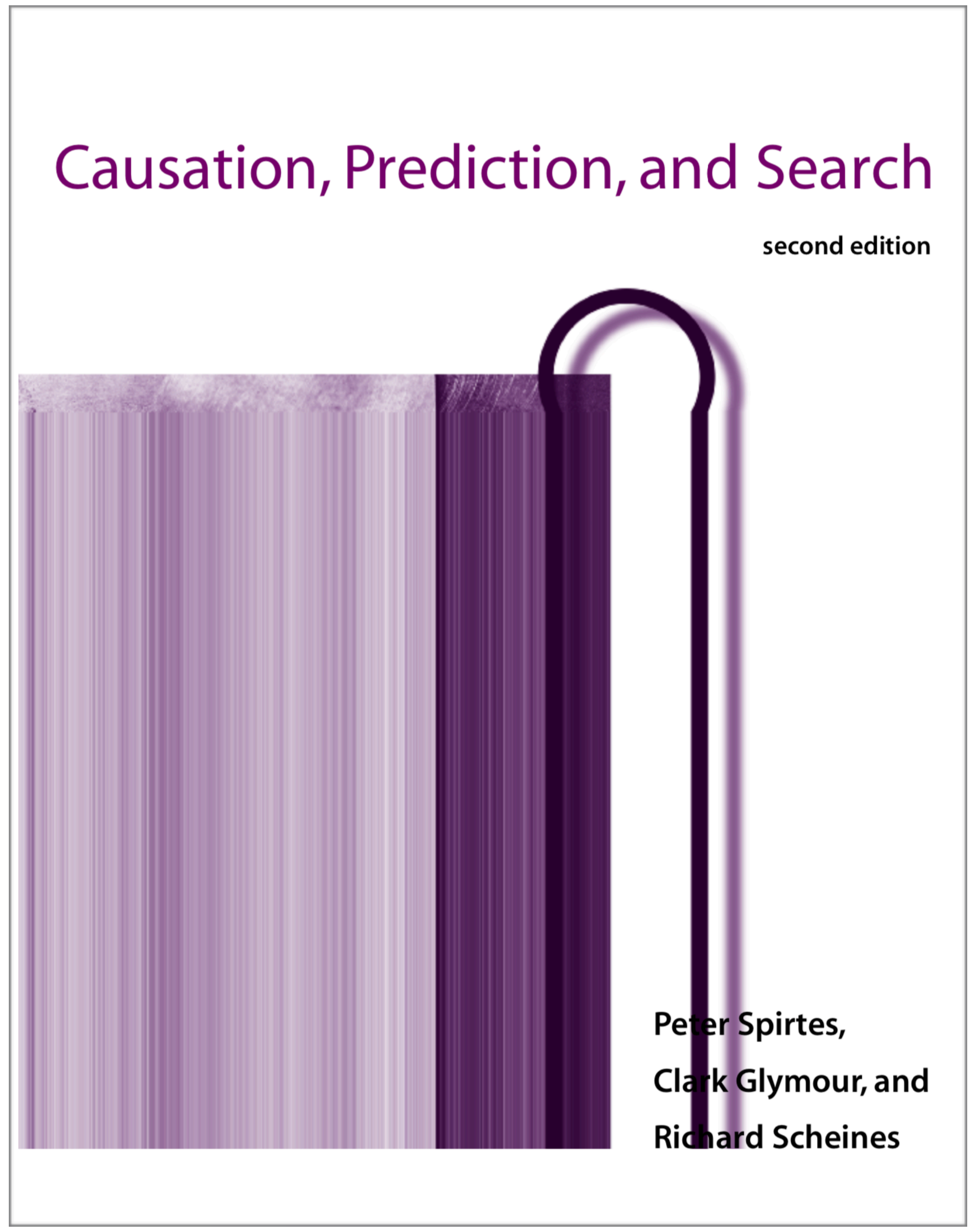
Terminology

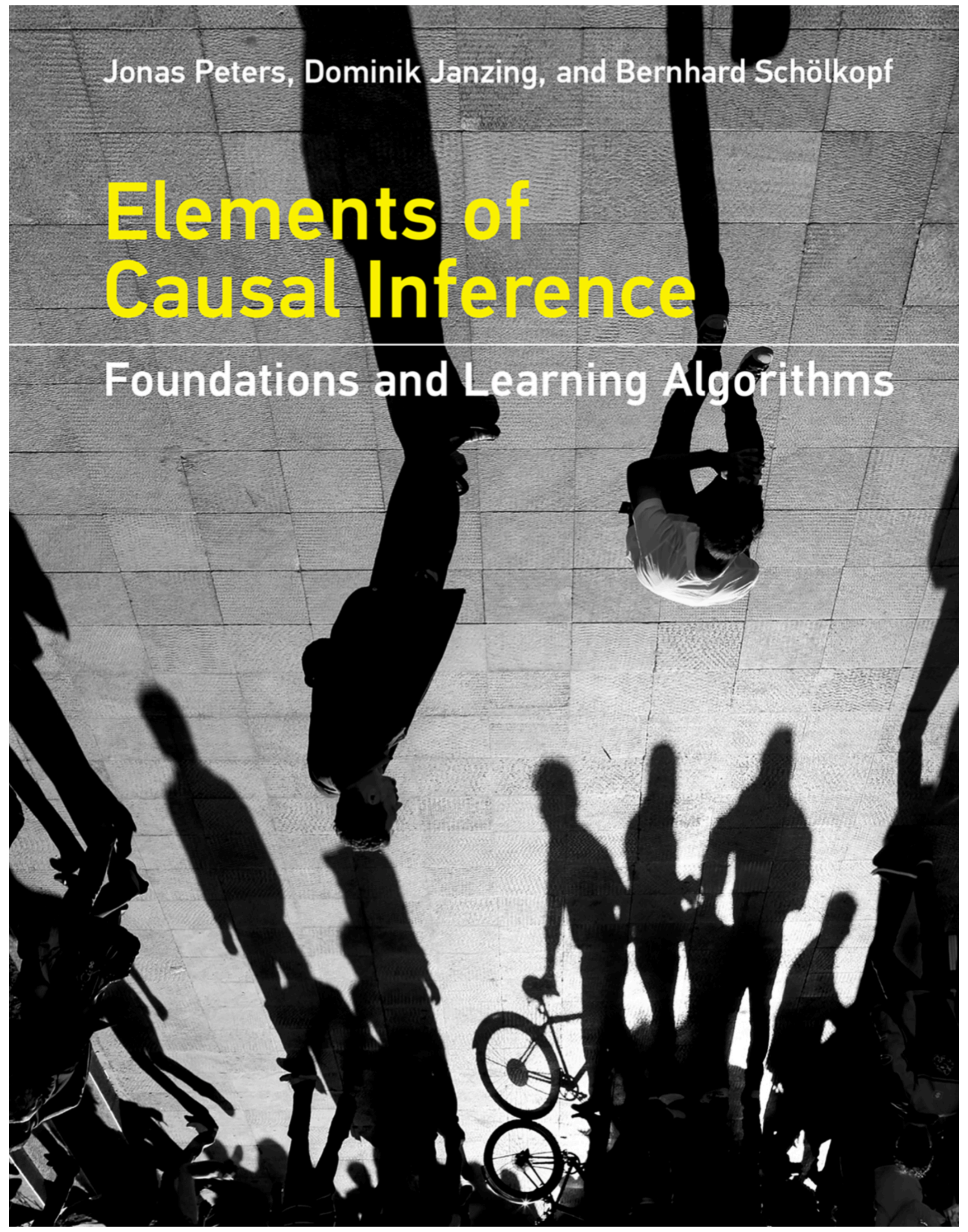
Statistics	Machine Learning	Meaning
Inference	Learning	using data to estimate parameters
Covariate	Feature	
DAG (directed acyclic graph)	Bayesian Network	multivariate distribution with given conditional independence relations
	...	

Textbooks


CAUSAL INFERENCE IN STATISTICS
A Primer
Judea Pearl
Madelyn Glymour
Nicholas P. Jewell
WILEY


CAUSALITY
SECOND EDITION
MODELS, REASONING, AND INFERENCE
JUDEA PEARL


Causation, Prediction, and Search
second edition
Peter Spirtes,
Clark Glymour, and
Richard Scheines


Jonas Peters, Dominik Janzing, and Bernhard Schölkopf
Elements of Causal Inference
Foundations and Learning Algorithms

Libraries for Causal Learning

Causal Discovery Toolbox

CausalML

DoWhy – A library for causal inference

August 21, 2018 | By Amit Sharma, Senior Researcher; Emre Kiciman, Senior Principal Researcher

Goal of Data Science $\stackrel{?}{=}$ Prediction

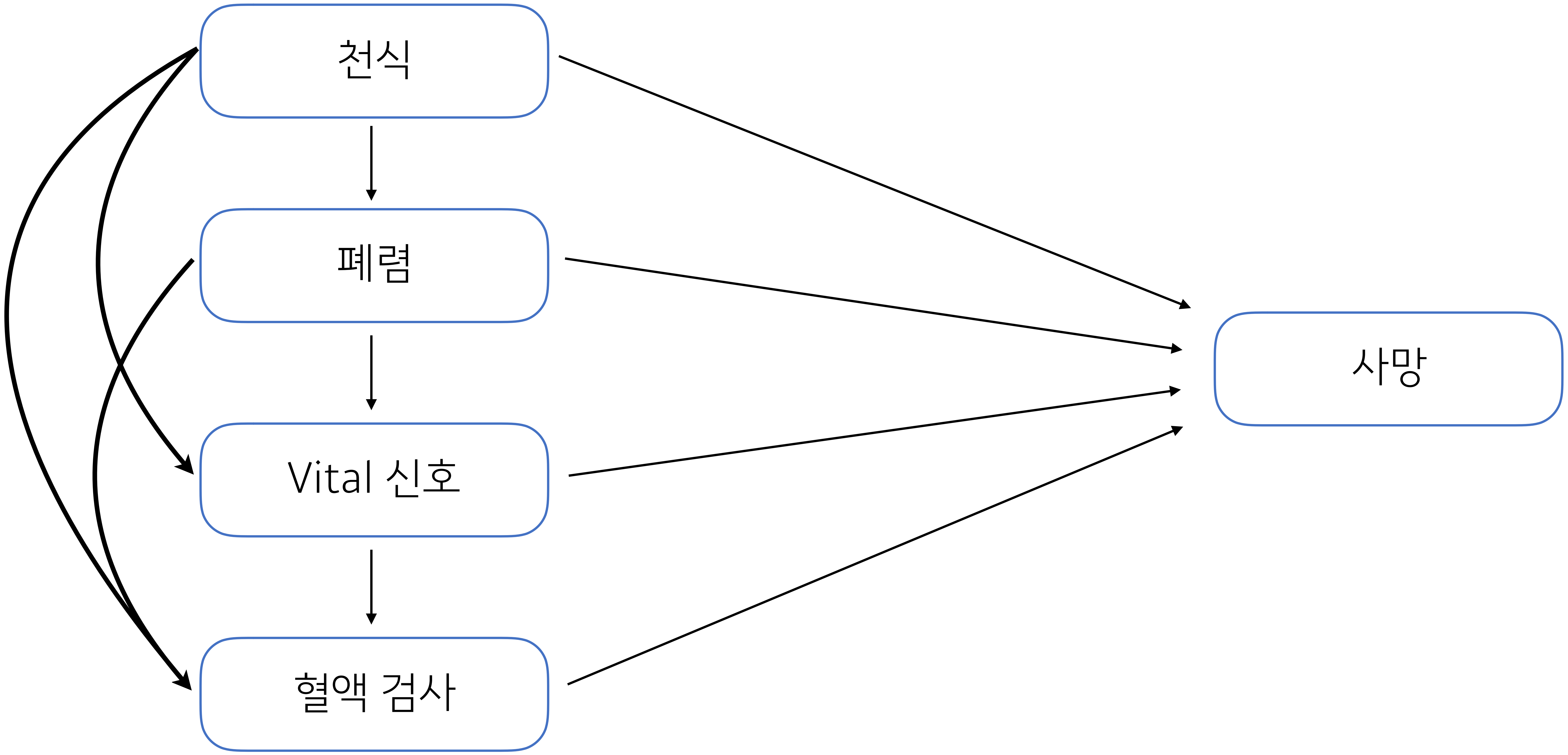
What is Causal Question?

- Beyond Association
    - **Intervention**
    - **Counterfactual**
- Applications
    - Medical AI
    - Econometrics
    - Science
    - Policy Evaluation

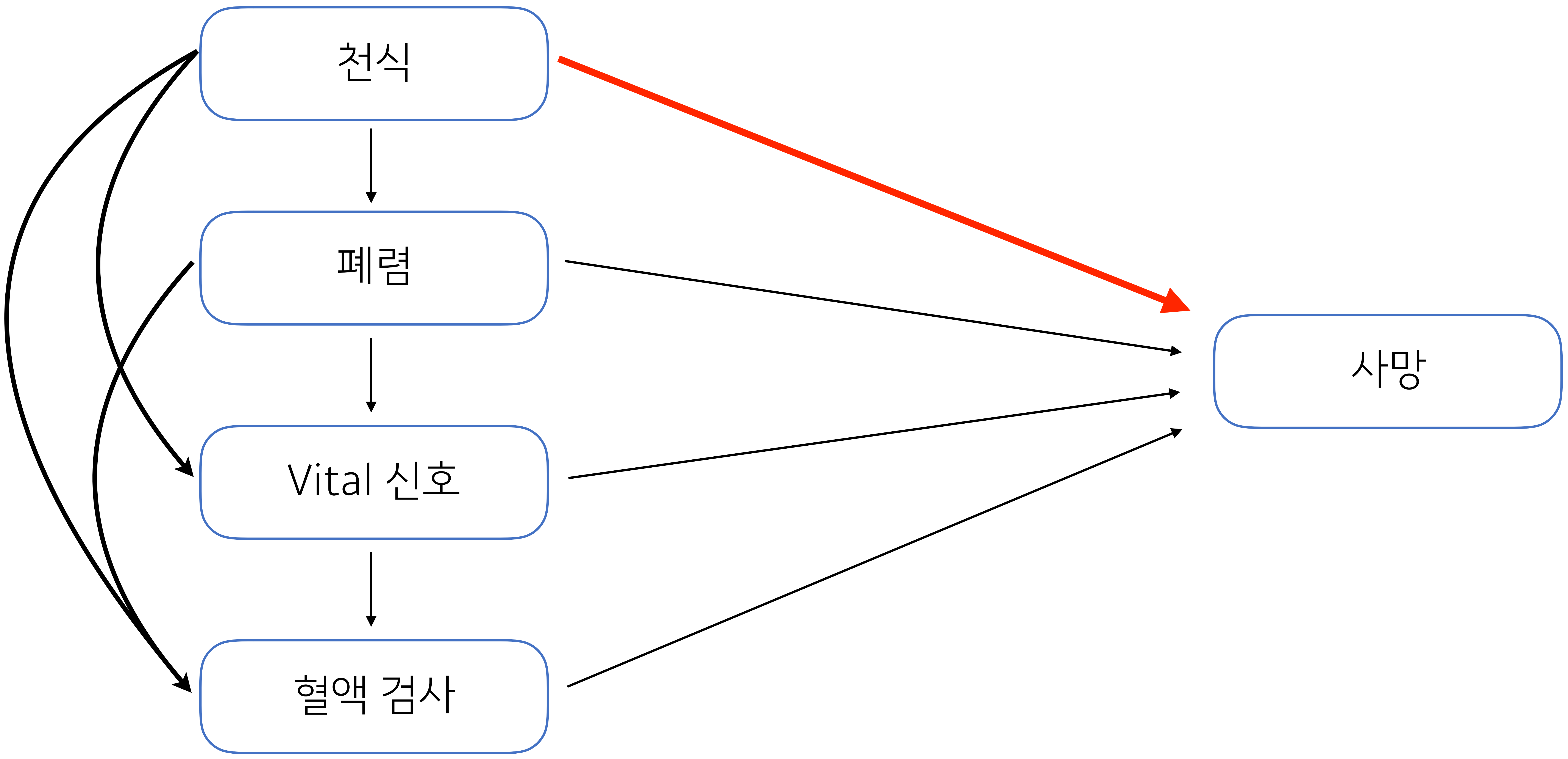
Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y \mid x)$	Seeing	What is? How would seeing *X* change my belief inY?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y \mid do(x), z)$	Doing, Intervening	What if? What if I do *X*?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x \mid x', y')$	Imagining, Retrospection	Why? Was it *X* that caused *Y*? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past two years?

The 7 tools of Causal Inference, Judea Pearl, (2019)

Why Causality?

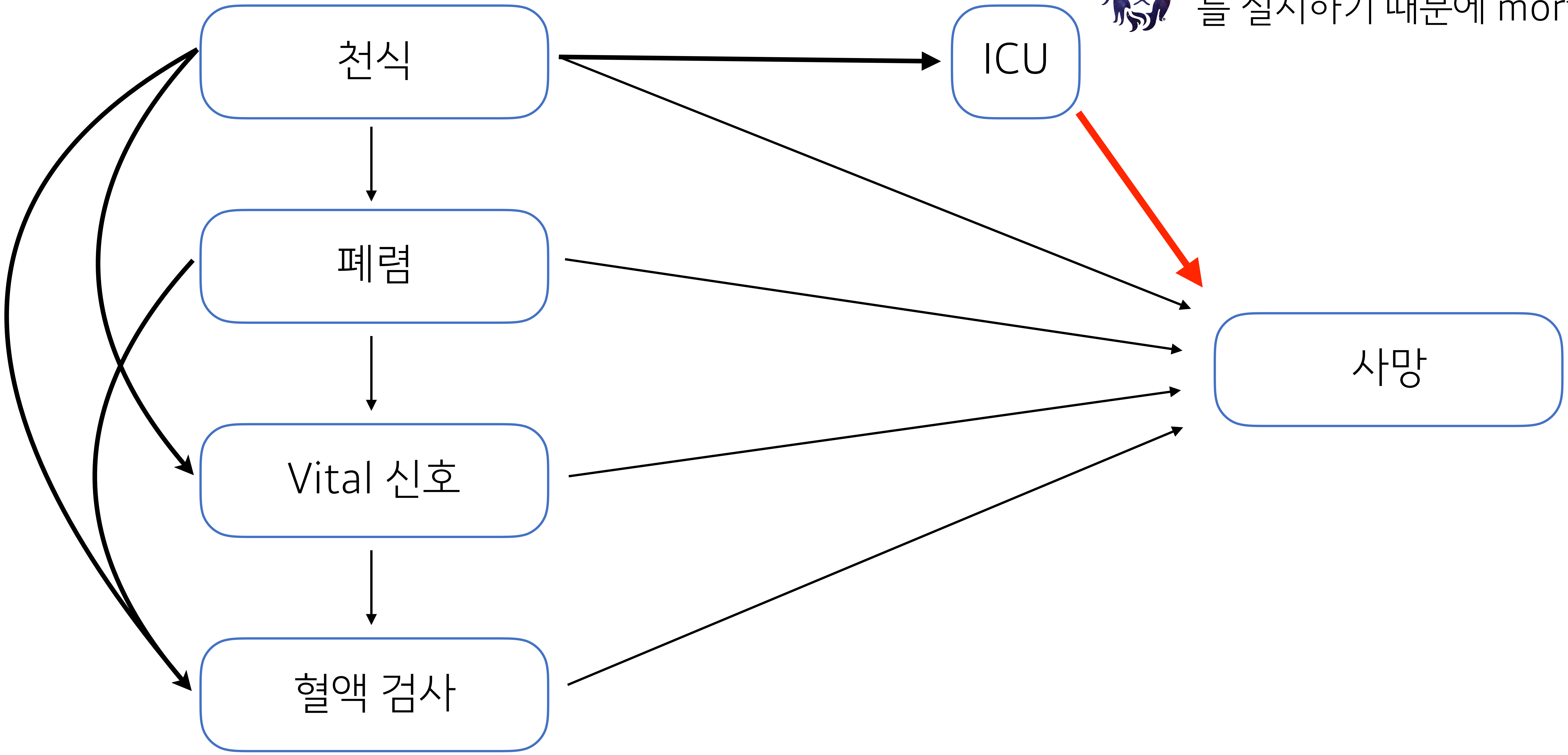


Why Causality?



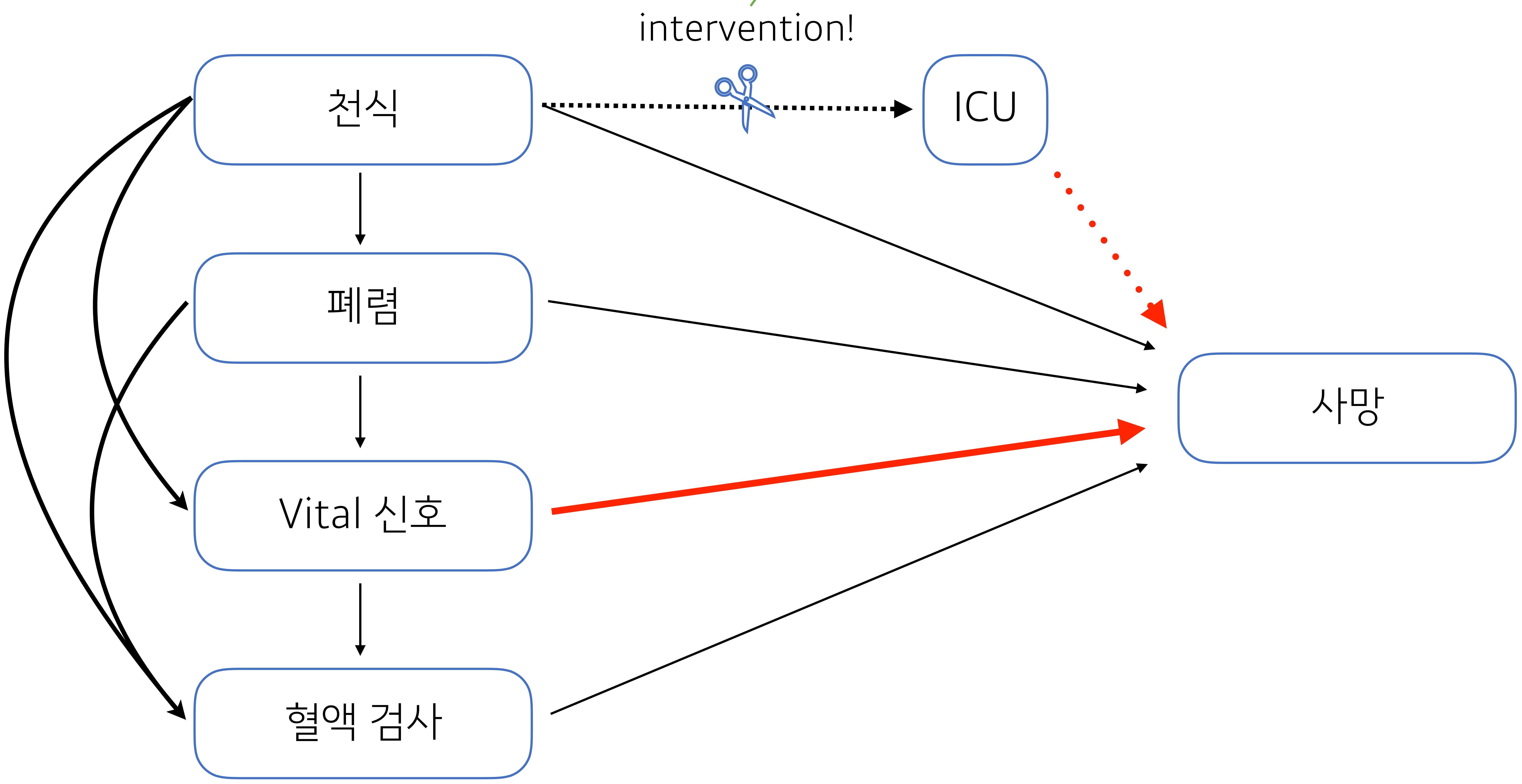
Why Causality?

test 데이터에선 ICU 에 가지 않고 home care 를 실시하기 때문에 mortality 는 더 커진다



Why Causality?

$P(\texttt{mortality}|\texttt{asthma}, \texttt{pneumonia}, do(\texttt{no ICU}))$

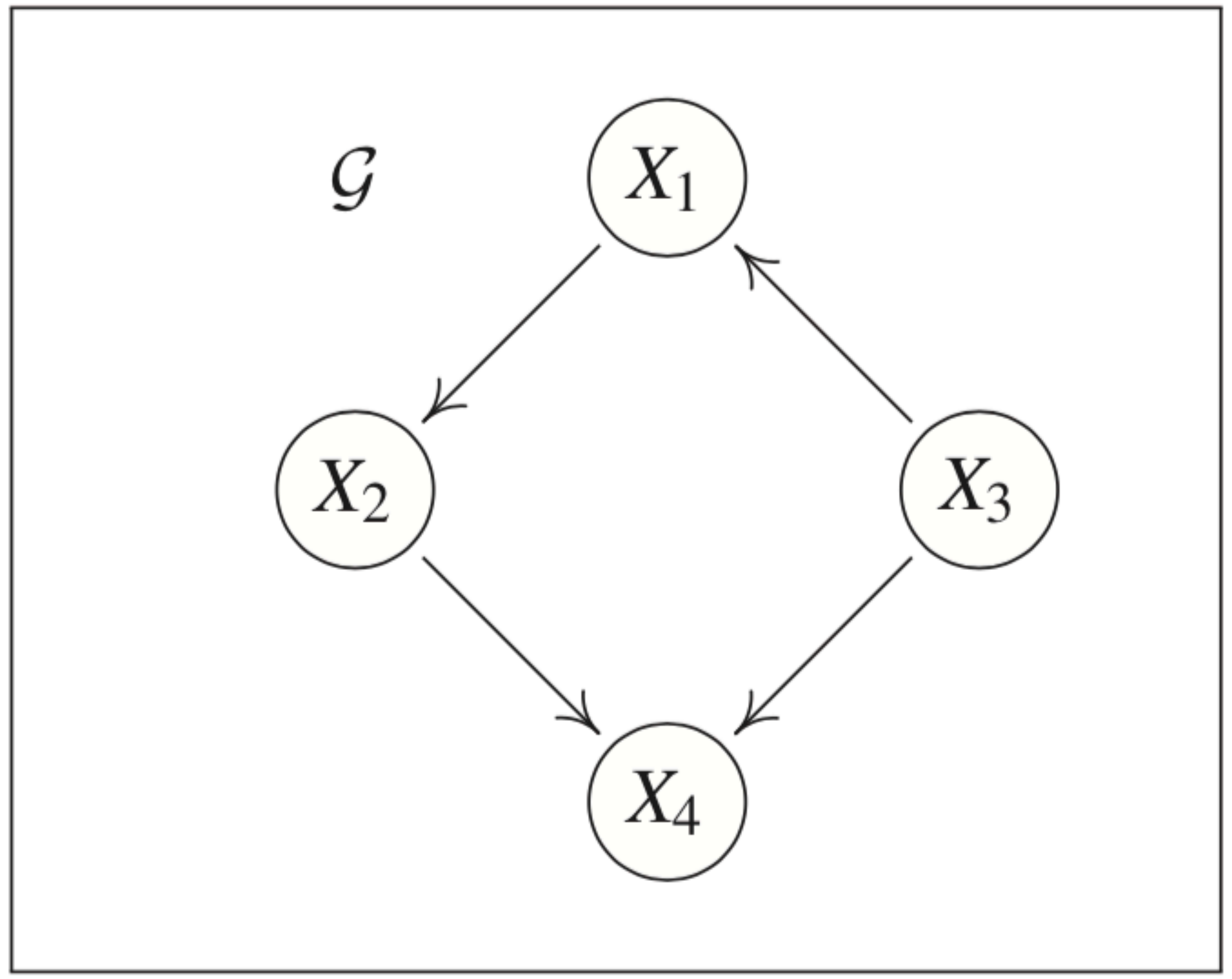


Structural Causal Models

- **SCM** (structural causal models)
  - social science, econometrics (Wright 1921)

$$X_1 := f_1(X_3, N_1)$$
$$X_2 := f_2(X_1, N_2)$$
$$X_3 := f_3(N_3)$$
$$X_4 := f_4(X_2, X_3, N_4)$$

- $N_1, \ldots, N_4$ jointly independent
- $\mathcal{G}$ is acyclic

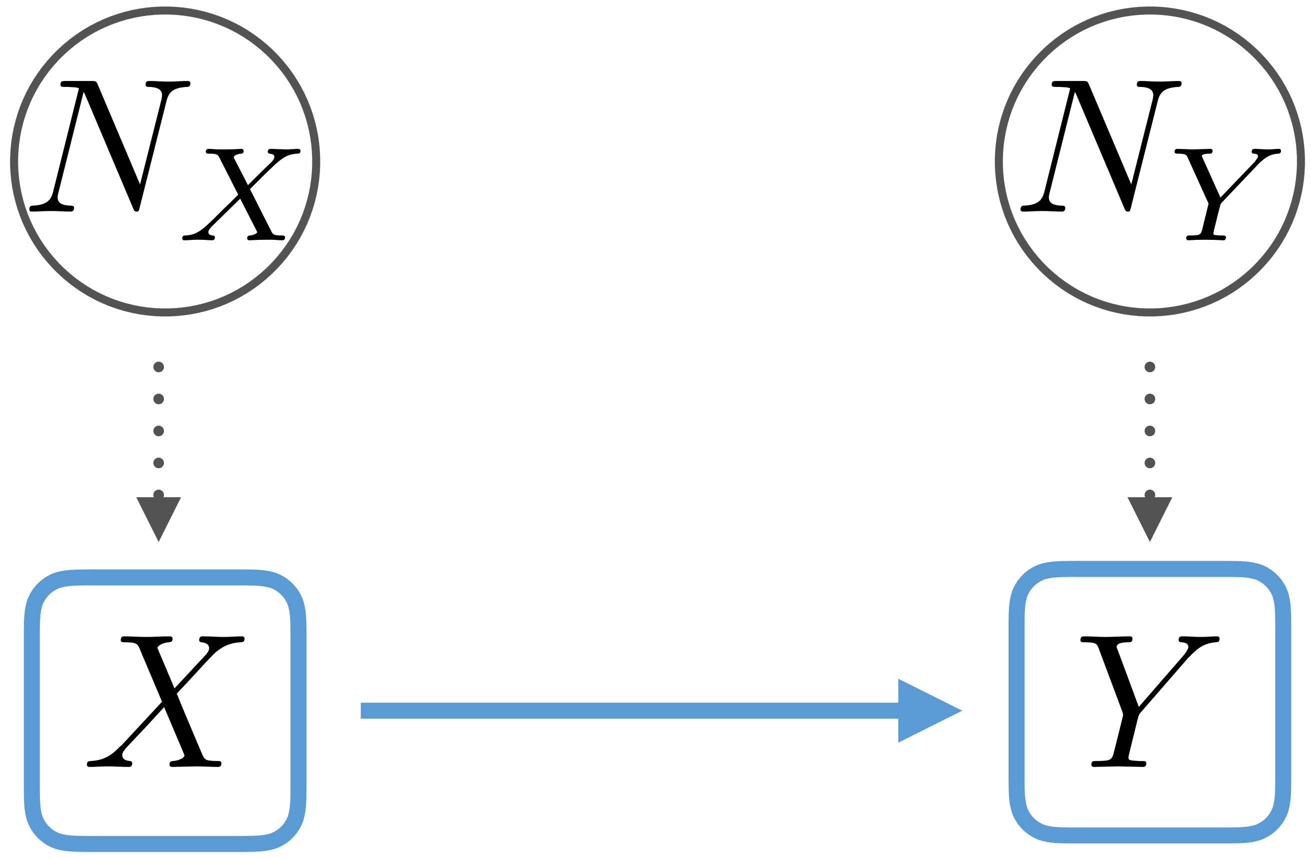


Formulation

$$X = f_X(N_X)$$
$$Y = f_Y(X, N_Y)$$

Structural Causal Model $\mathfrak{C}$

equivalent!



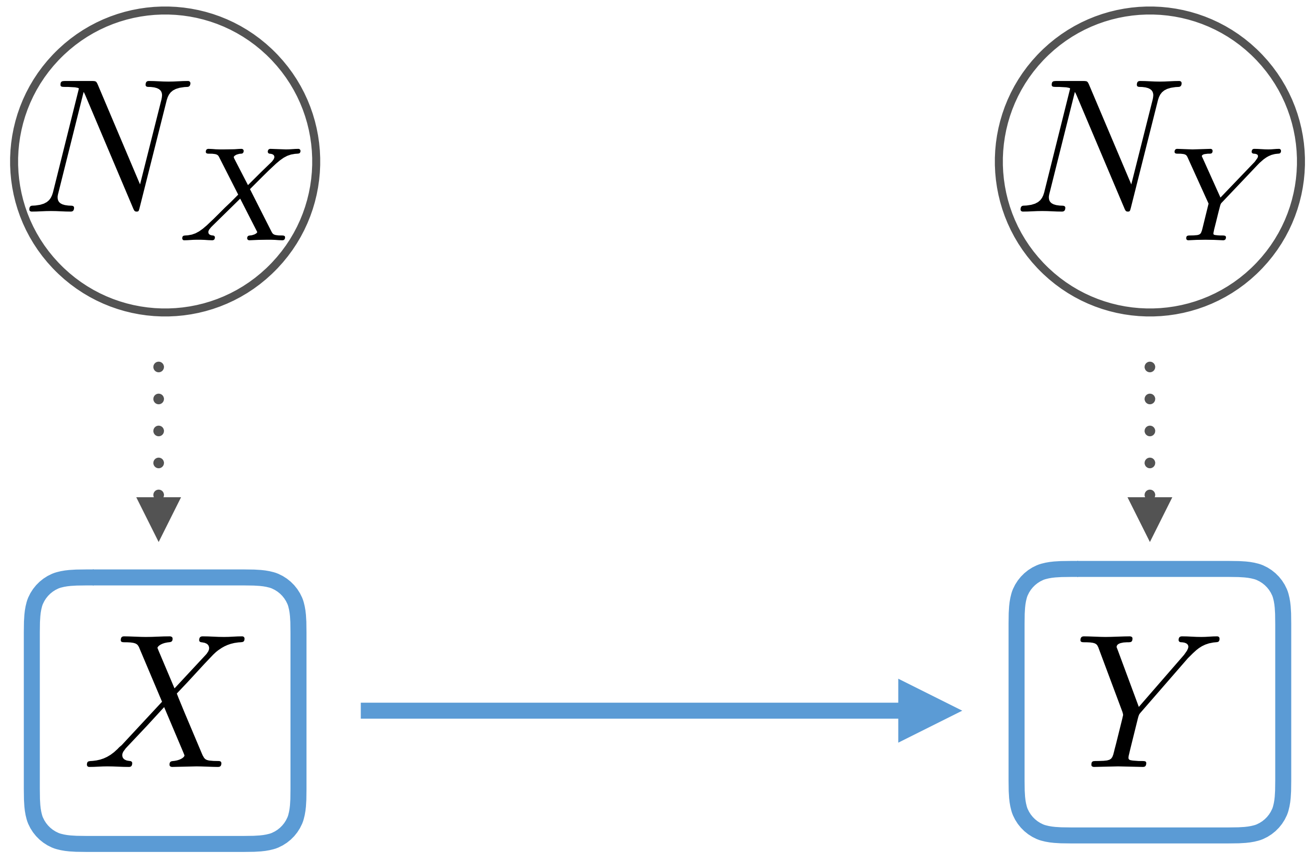
Graphical Causal Model $\mathcal{G}$

Formulation

$$X = f_X(N_X)$$
$$Y = f_Y(X, N_Y)$$

Structural Causal Model $\mathfrak{C}$

equivalent!



Graphical Causal Model $\mathcal{G}$

induce!

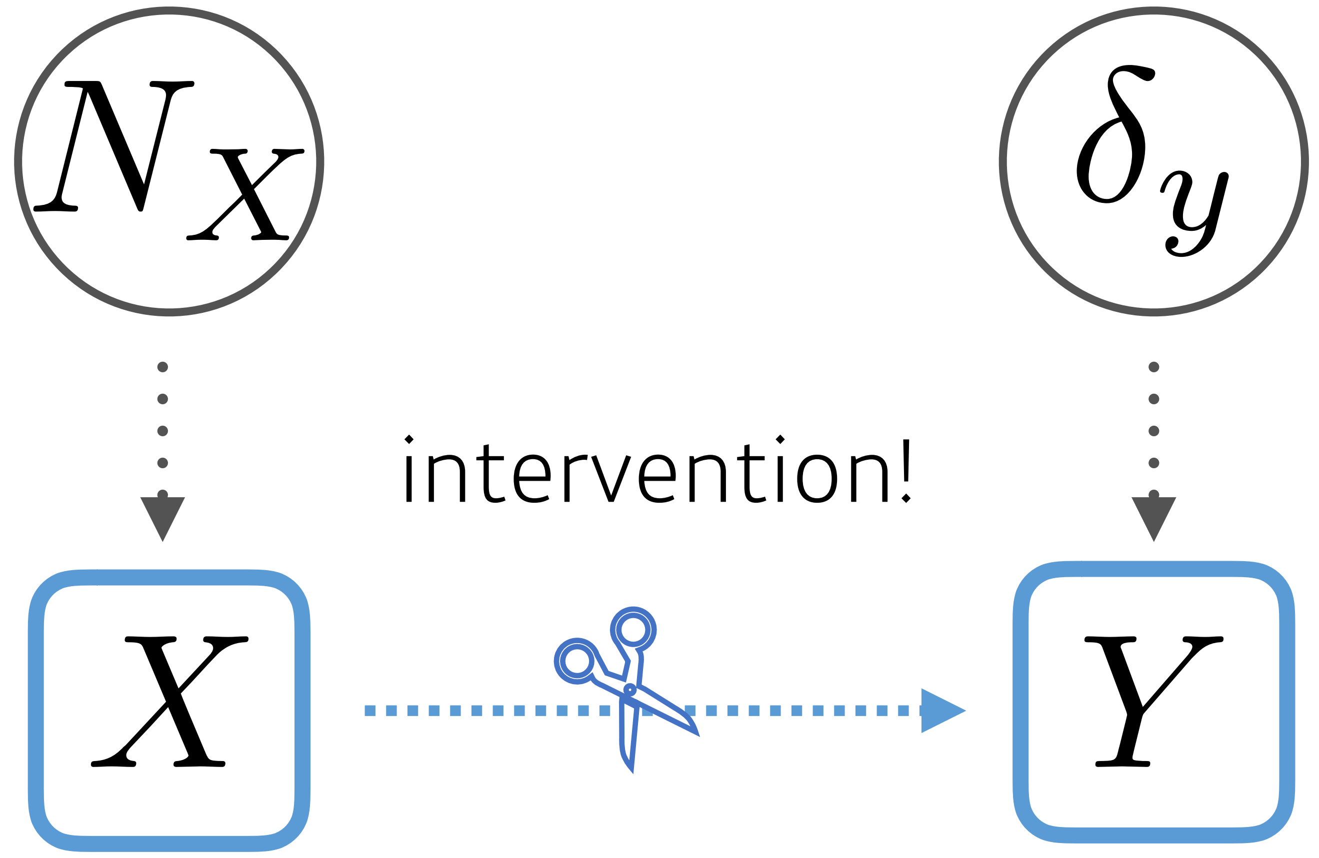
$$\mathbb{P}_{\mathfrak{C}}(x, y)$$

joint distribution

Effect of Intervention

$$X = f_X(N_X)$$
$$Y = y$$

Structural Causal Model $\mathfrak{C}_{\mathrm{do}(Y=y)}$



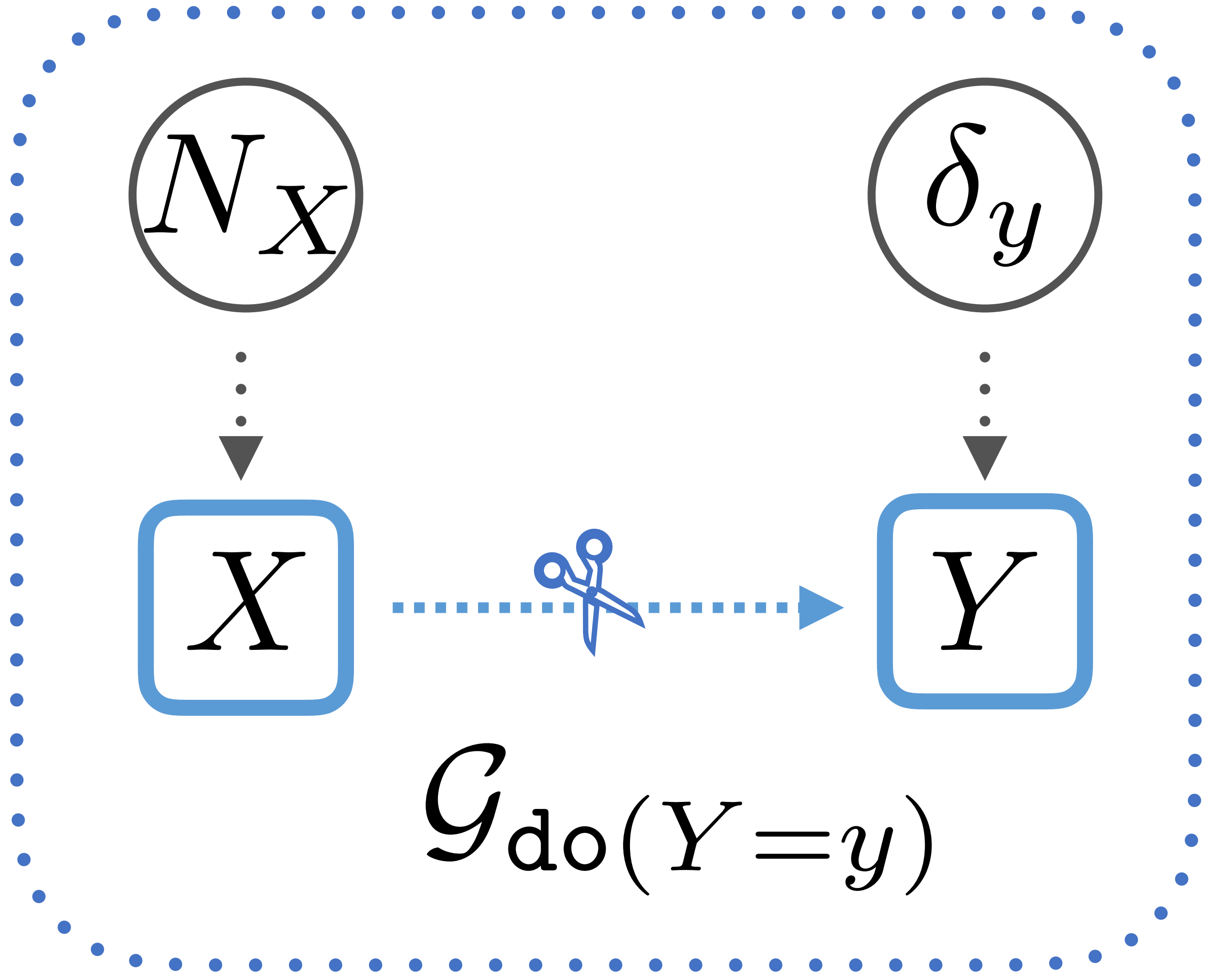
Graphical Causal Model $\mathcal{G}_{\mathrm{do}(Y=y)}$

$$\mathbb{P}_{\mathfrak{C}_{\mathrm{do}(Y=y)}}(x)$$

Effect of Intervention

$$\mathbb{P}(X = x \mid \mathtt{do}(Y = y)) = \mathbb{P}_{\mathfrak{C}_{\mathtt{do}(Y=y)}}(x) = \mathbb{P}_{\mathfrak{C}}(x)$$

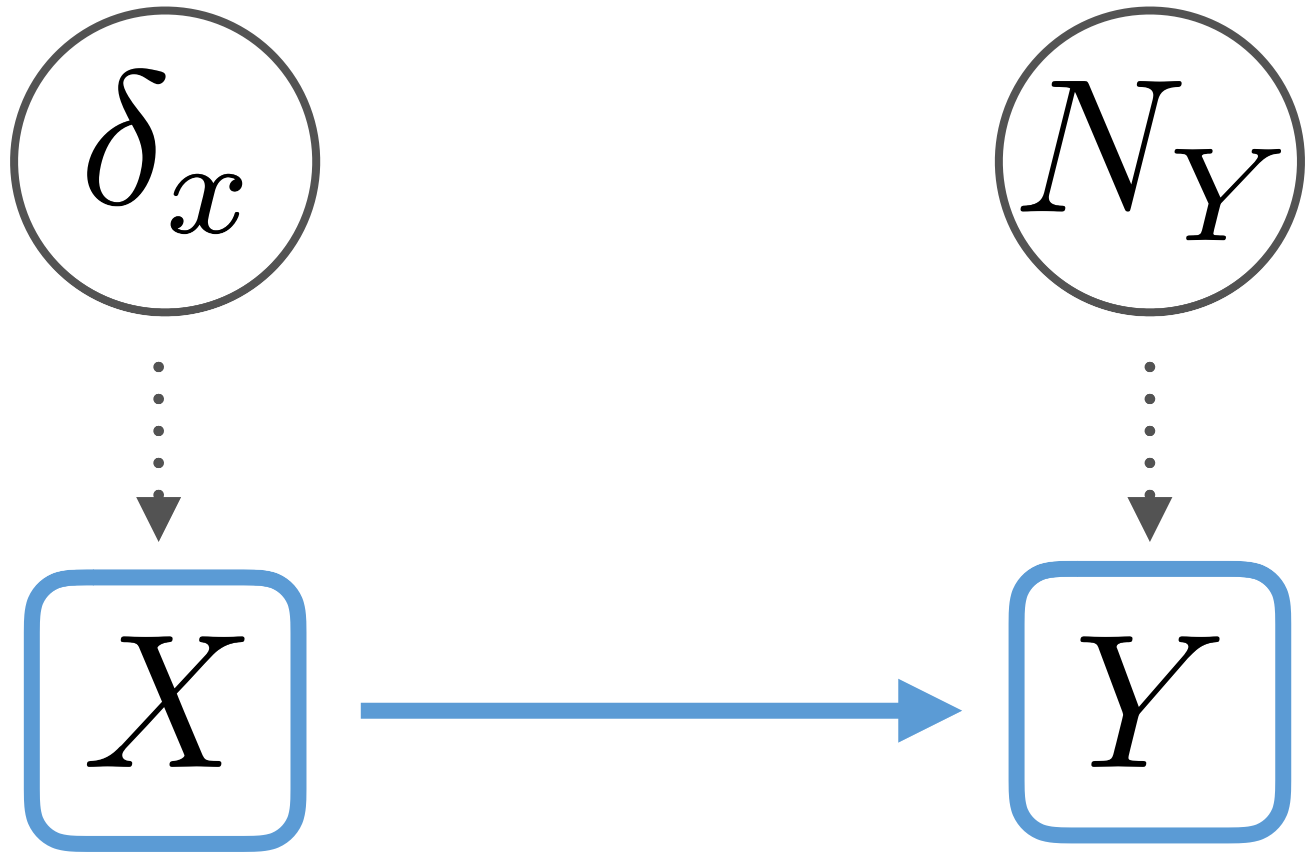
intervention 의 효과를 알고 싶다면 새로
도출한 SCM 의 joint 분포를 알아야 한다



Effect of Intervention

$$X = x$$
$$Y = f_Y(X, N_Y)$$

Structural Causal Model $\mathfrak{C}_{\mathtt{do}(X=x)}$



Graphical Causal Model $\mathcal{G}_{\mathtt{do}(X=x)}$

$$\mathbb{P}_{\mathfrak{C}_{\mathtt{do}(X=x)}}(y)$$

Effect of Intervention

$$
\begin{aligned}
X &= x \\
Y &= f_Y(X, N_Y)
\end{aligned}
$$

$$\longleftrightarrow$$

$$
\begin{tikzcd}
\delta_x \arrow[d, dotted] & N_Y \arrow[d, dotted] \\
\boxed{X} \arrow[r] & \boxed{Y}
\end{tikzcd}
$$

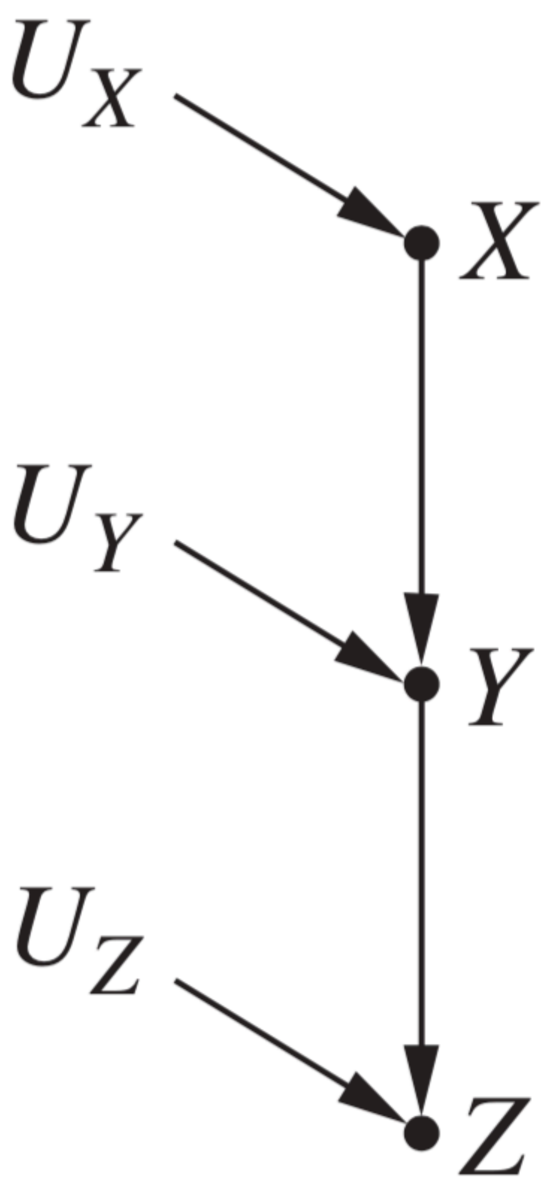
Structural Causal Model $\mathfrak{C}_{\mathrm{do}(X=x)}$

Graphical Causal Model $\mathcal{G}_{\mathrm{do}(X=x)}$

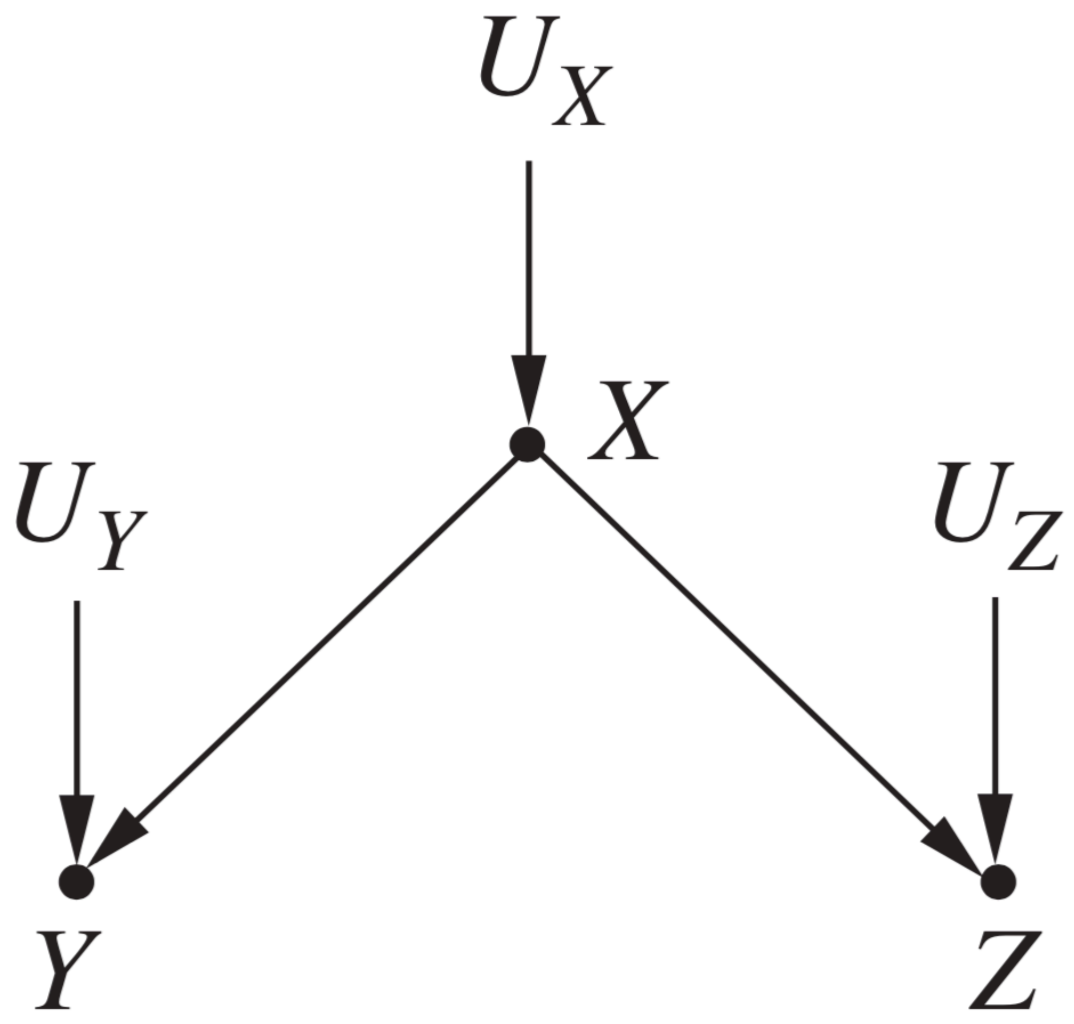
$$\mathbb{P}_{\mathfrak{C}_{\mathrm{do}(X=x)}}(y) = \mathbb{P}_{\mathfrak{C}}(y|x)$$

이 경우 intervention 이랑 conditioning 과 같다

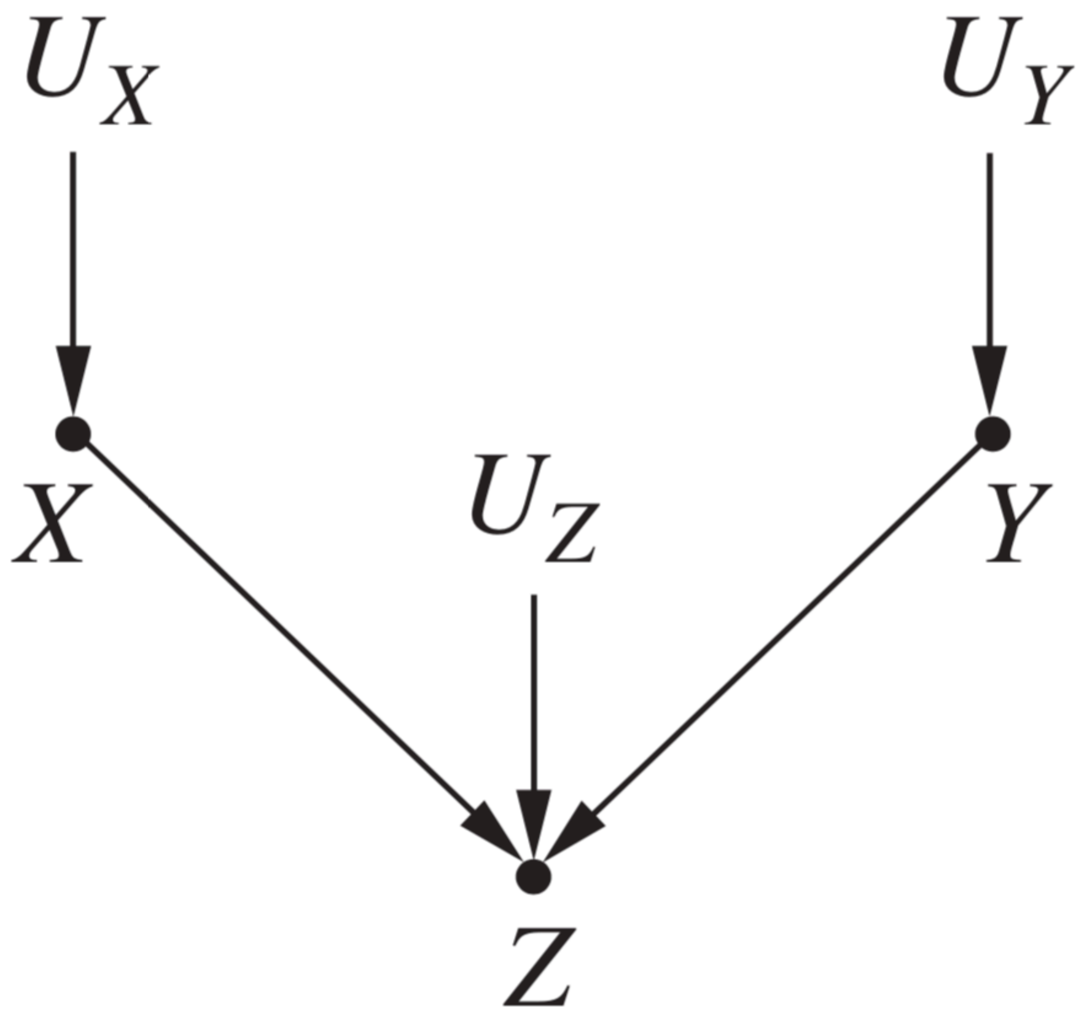
Conditional Independence



chain

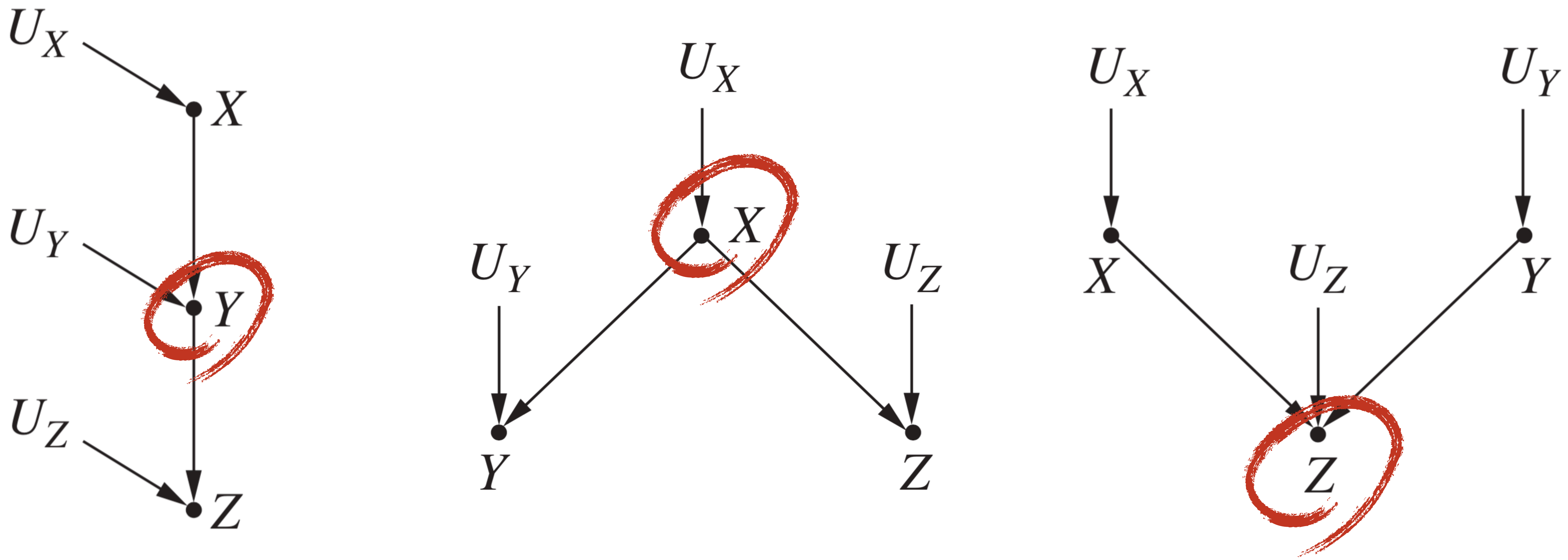


fork



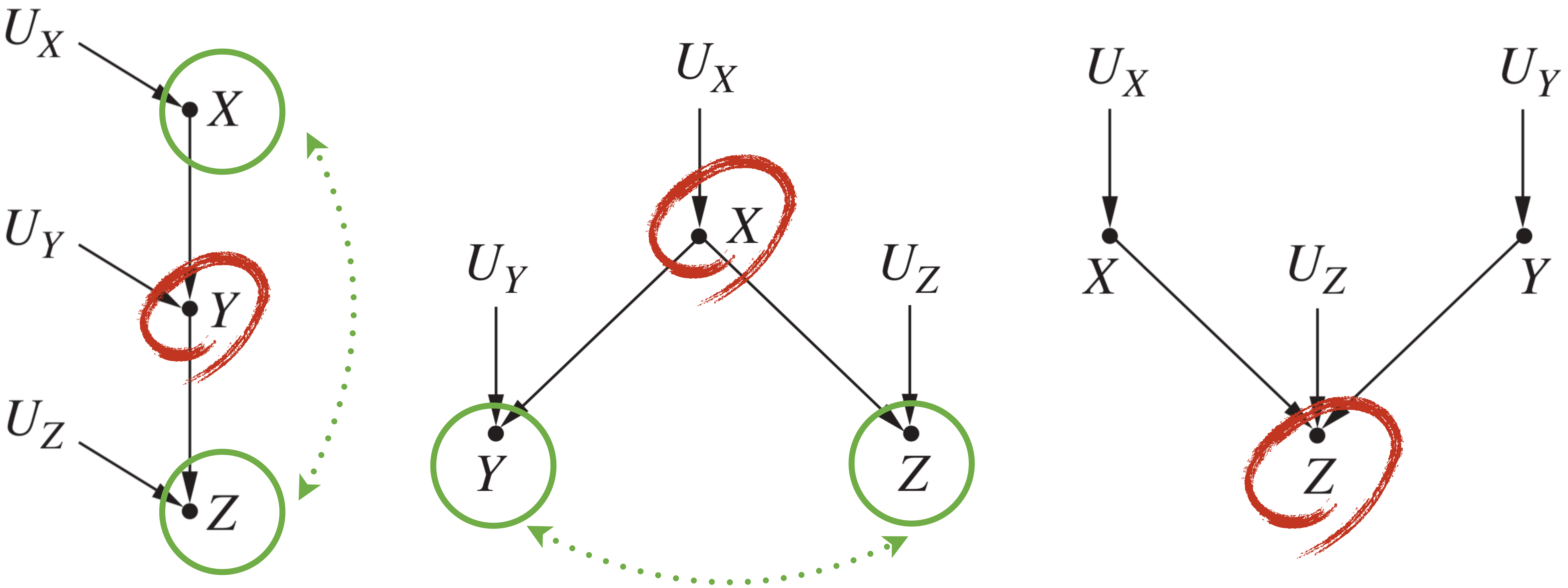
collider

Conditional Independence



chain fork collider

Conditional Independence



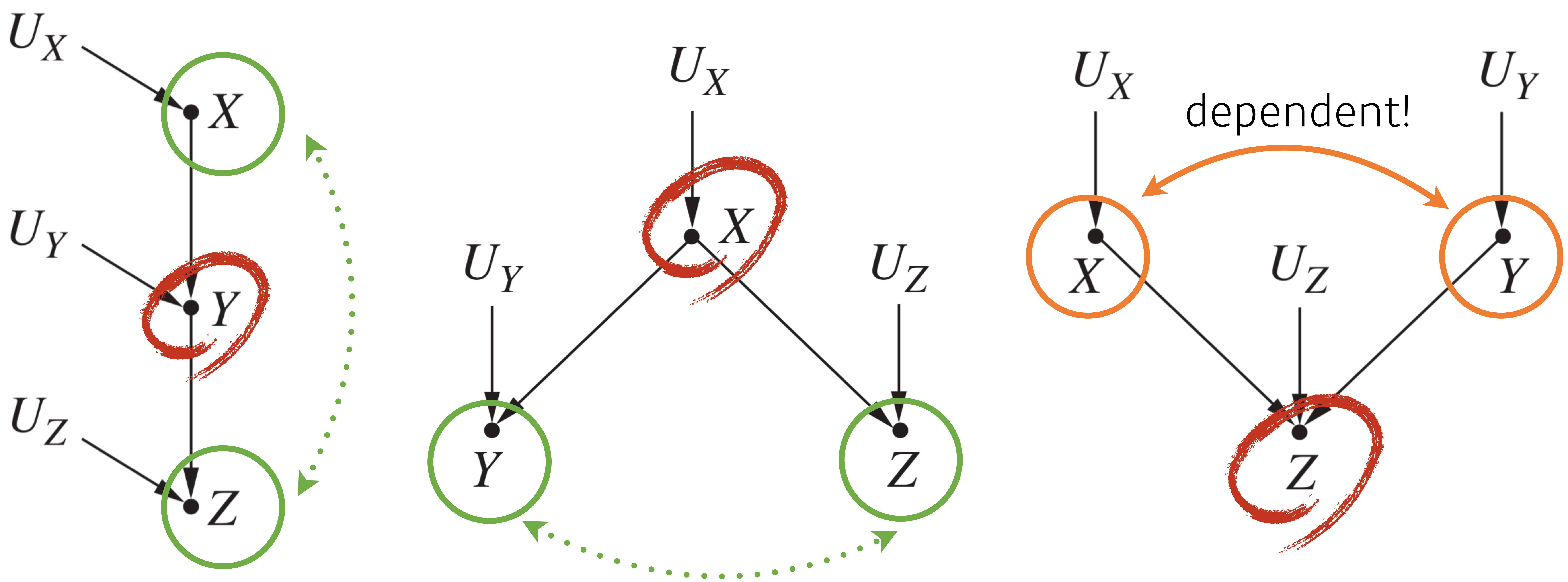
chain

fork

collider

$$X \perp Z | Y \qquad Y \perp Z | X$$

Conditional Independence

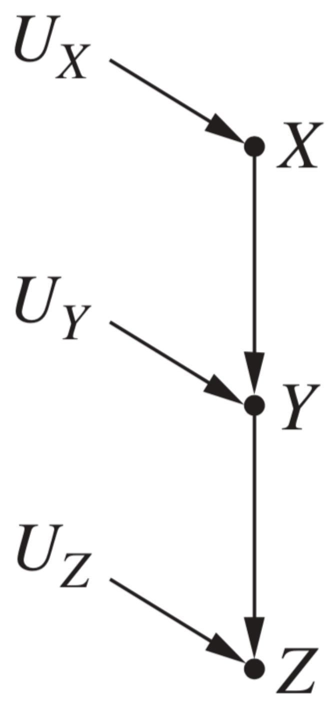


chain $X \perp Z | Y$

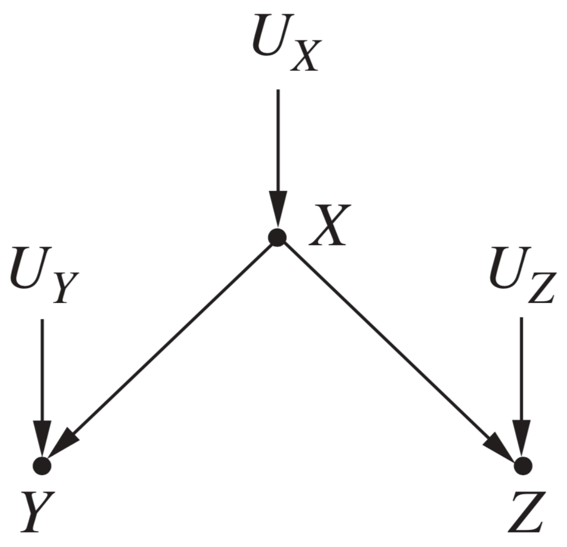
fork $Y \perp Z | X$

collider $X \not\perp Y | Z$

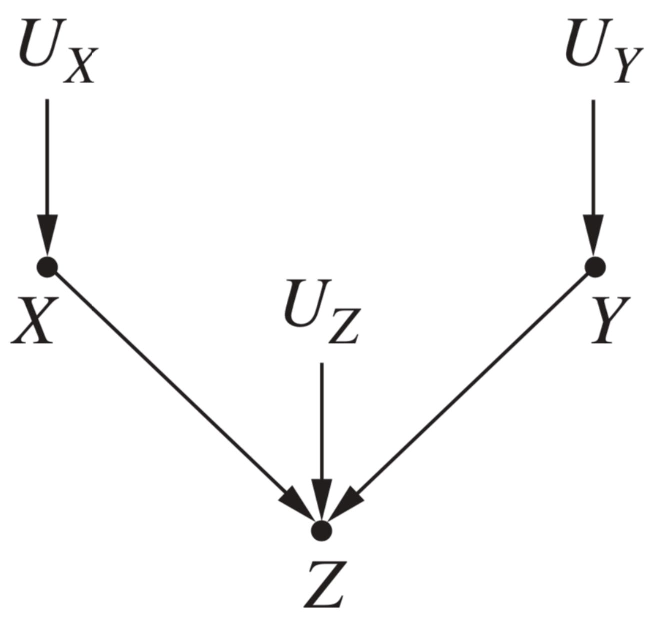
d–separation



chain



fork



collider

Definition 6.1 (Pearl's d-separation) *In a DAG $\mathcal{G}$, a path between nodes i_1 and i_m is **blocked by a set S** (with neither i_1 nor i_m in $\mathbf{S}$) whenever there is a node i_k, such that one of the following two possibilities holds:*

(i) $i_k \in \mathbf{S}$ *and*

$$\begin{aligned} & i_{k-1} \to i_k \to i_{k+1} \\ \text{or}\ & i_{k-1} \leftarrow i_k \leftarrow i_{k+1} \\ \text{or}\ & i_{k-1} \leftarrow i_k \to i_{k+1} \end{aligned}$$

(ii) neither i_k nor any of its descendants is in $\mathbf{S}$ and

$$i_{k-1} \to i_k \leftarrow i_{k+1}.$$

*Furthermore, in a DAG $\mathcal{G}$, we say that two disjoint subsets of vertices $\mathbf{A}$ and $\mathbf{B}$ are **d-separated** by a third (also disjoint) subset $\mathbf{S}$ if every path between nodes in $\mathbf{A}$ and $\mathbf{B}$ is blocked by $\mathbf{S}$. We then write*

$$\mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B} \,|\, \mathbf{S}.$$

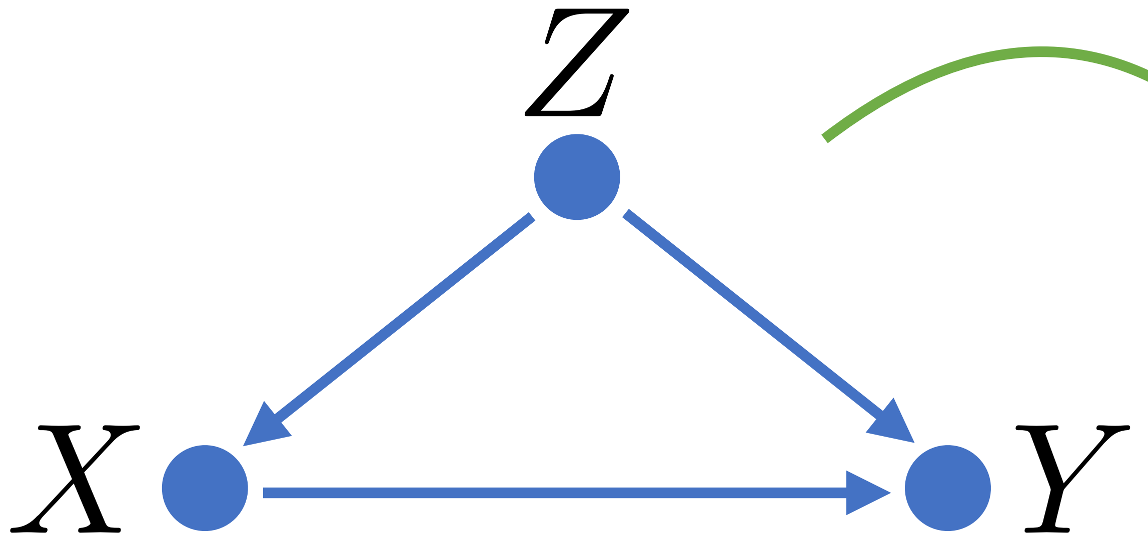
d-separation

- Suppose every d-separation condition in the model matches a conditional independence in the data
  - no further test can refute the model
  - hence it allows us to start with a data set, and reason back to a causal model
    - this is an exact goal of data-driven research

물론 현실은 combinatorial optimization 문제에 부딪힌다

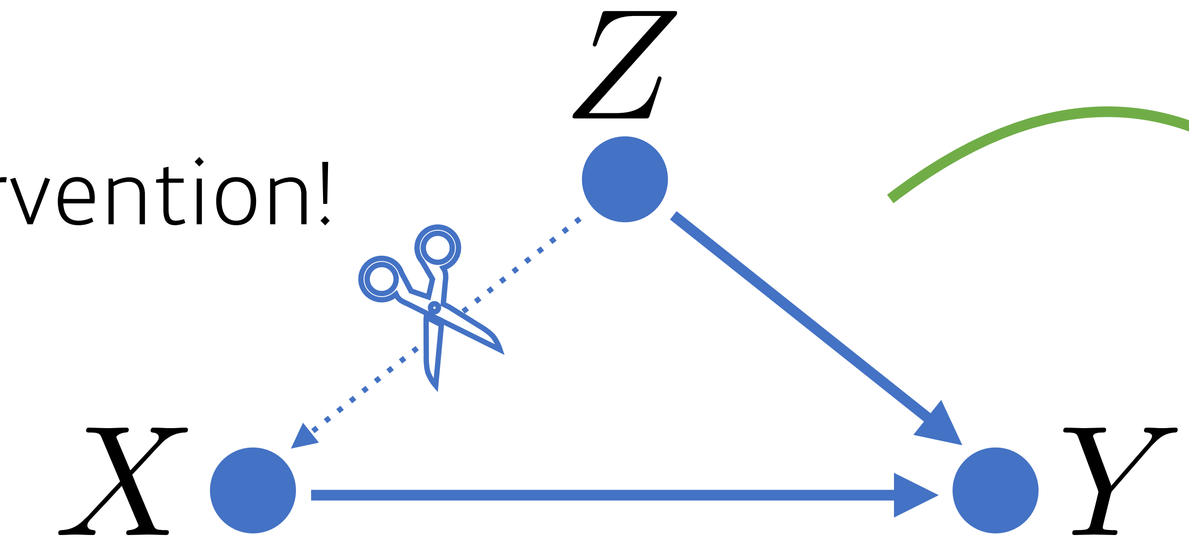
Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y=1|\texttt{do}(X=1)) - \mathbb{P}(Y=1|\texttt{do}(X=0))$



$$\mathbb{P}(Y=y, X=x) \longrightarrow \mathbb{P}(Y=y|X=x)$$

intervention!

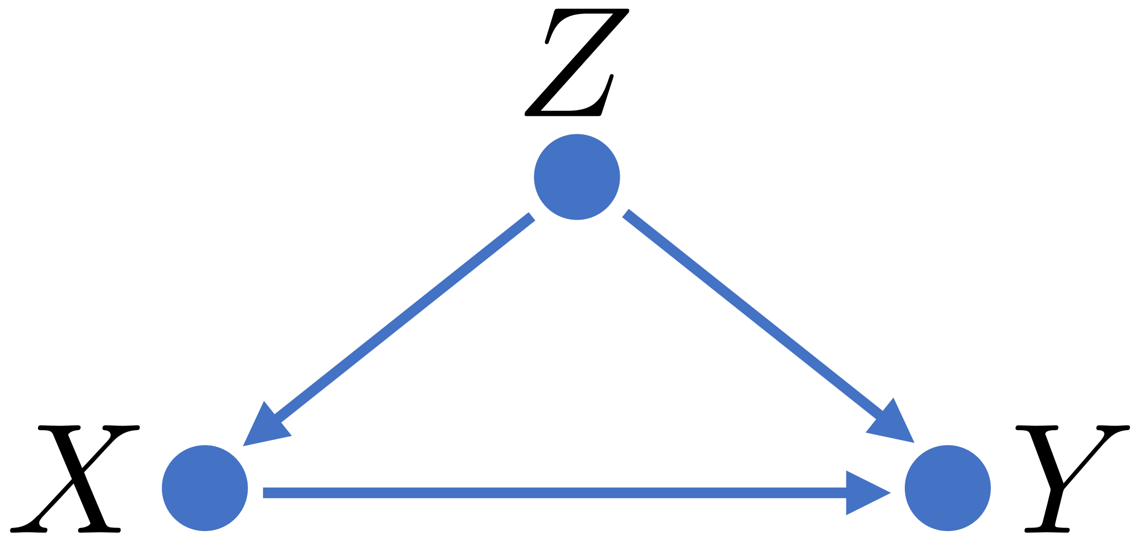


$$\mathbb{P}(Y=y|\texttt{do}(X=x)) :=$$

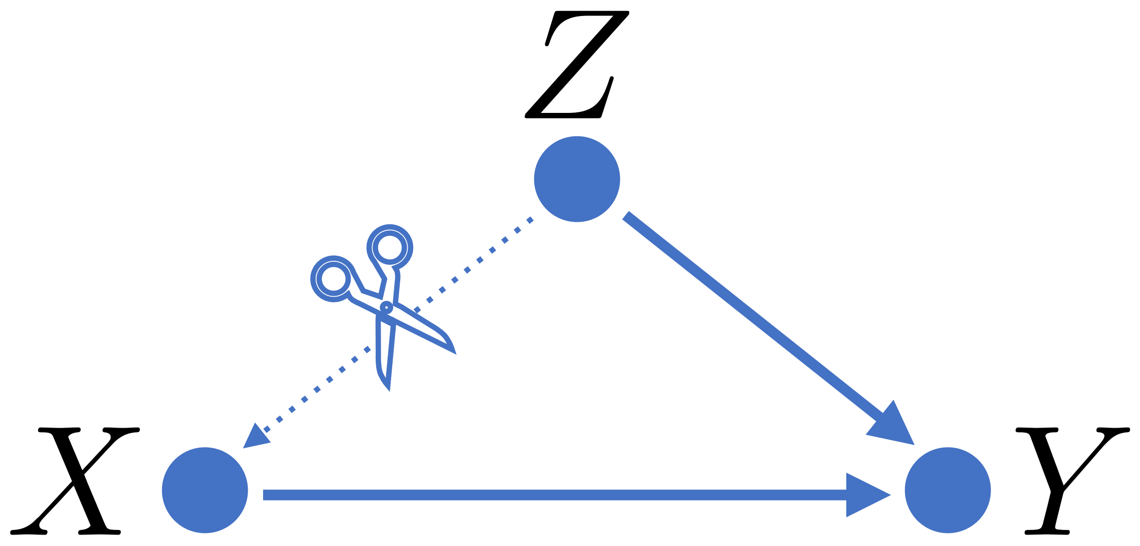
$$\widehat{\mathbb{P}}(Y=y, X=x) \longrightarrow \widehat{\mathbb{P}}(Y=y|X=x)$$

Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y=1|\mathtt{do}(X=1)) - \mathbb{P}(Y=1|\mathtt{do}(X=0))$



$$\mathbb{P}(Y=y|X=x)=\sum_z \mathbb{P}(Y=y|X=x,Z=z)\mathbb{P}(Z=z|X=x)$$

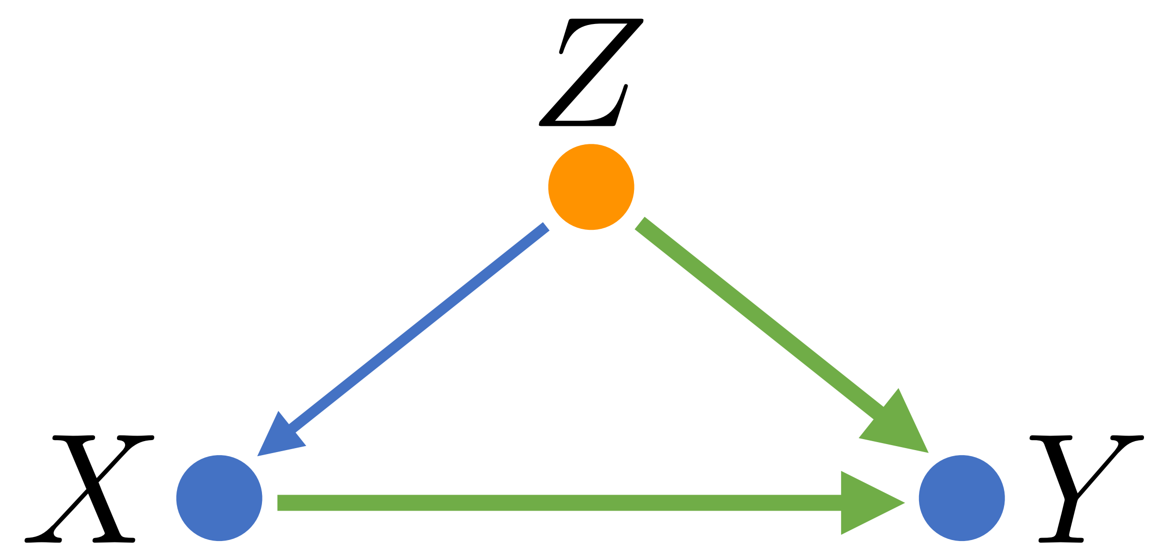


$$\widehat{\mathbb{P}}(Y=y|X=x)=\sum_z \widehat{\mathbb{P}}(Y=y|X=x,Z=z)\widehat{\mathbb{P}}(Z=z|X=x)$$

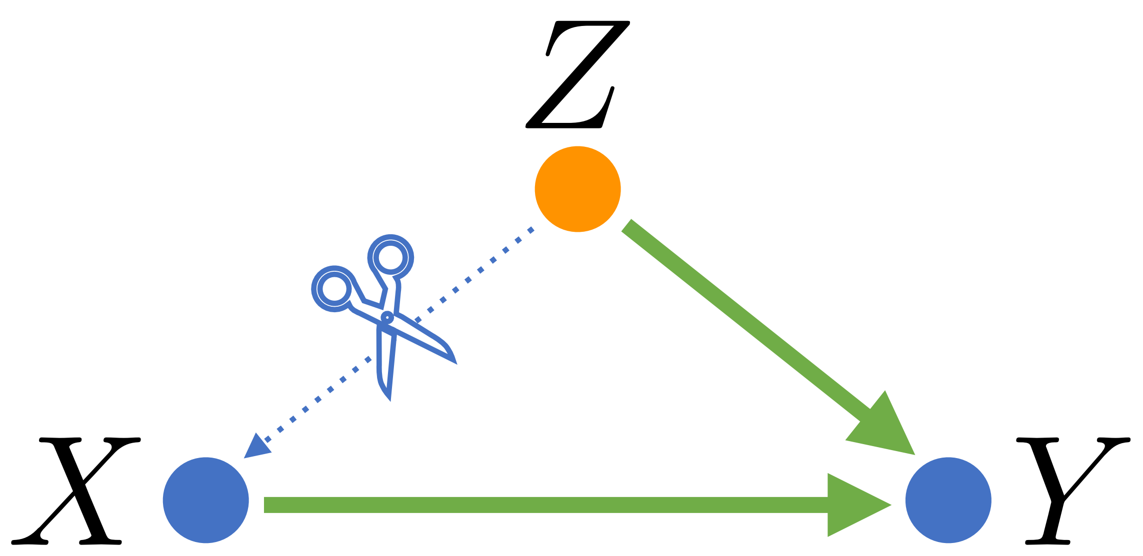
intervention 에 의해
Z 와 X 는 독립이 된다

Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y=1|\texttt{do}(X=1))-\mathbb{P}(Y=1|\texttt{do}(X=0))$



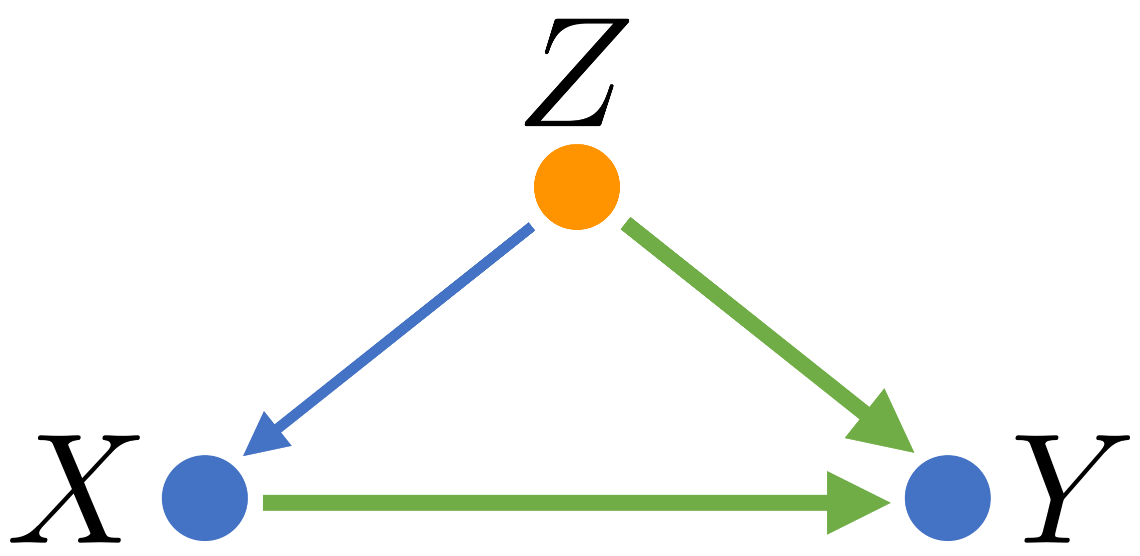
$$\mathbb{P}(Y=y|X=x)=\sum_z \mathbb{P}(Y=y|X=x,Z=z)\mathbb{P}(Z=z|X=x)$$



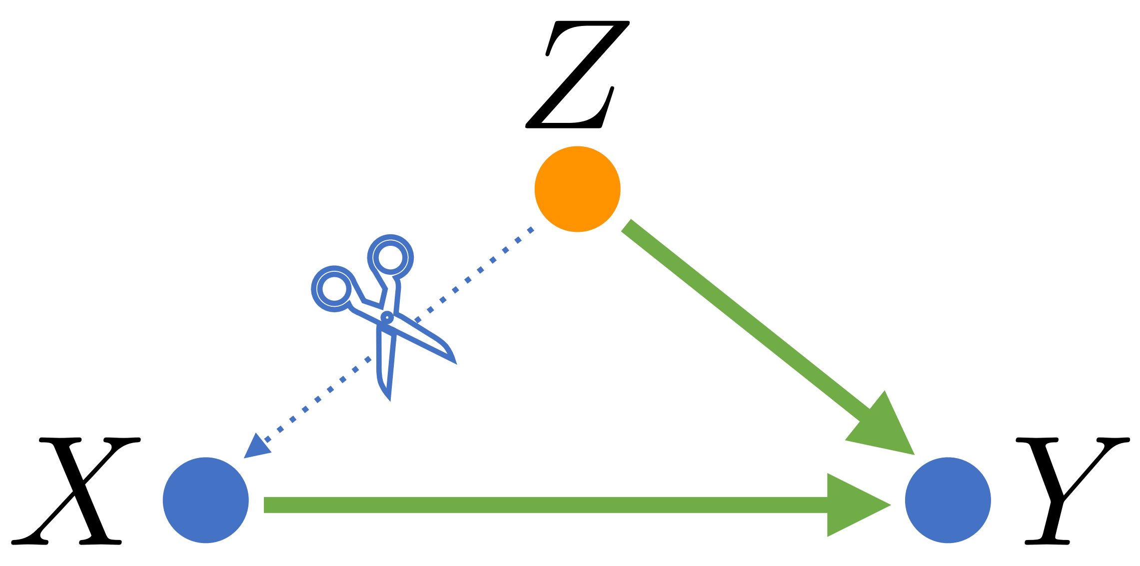
$$\widehat{\mathbb{P}}(Y=y|X=x)=\sum_z \widehat{\mathbb{P}}(Y=y|X=x,Z=z)\widehat{\mathbb{P}}(Z=z)$$

Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y=1|\mathtt{do}(X=1)) - \mathbb{P}(Y=1|\mathtt{do}(X=0))$



$$\mathbb{P}(Y=y|X=x) = \sum_z \mathbb{P}(Y=y|X=x, Z=z)\mathbb{P}(Z=z|X=x)$$



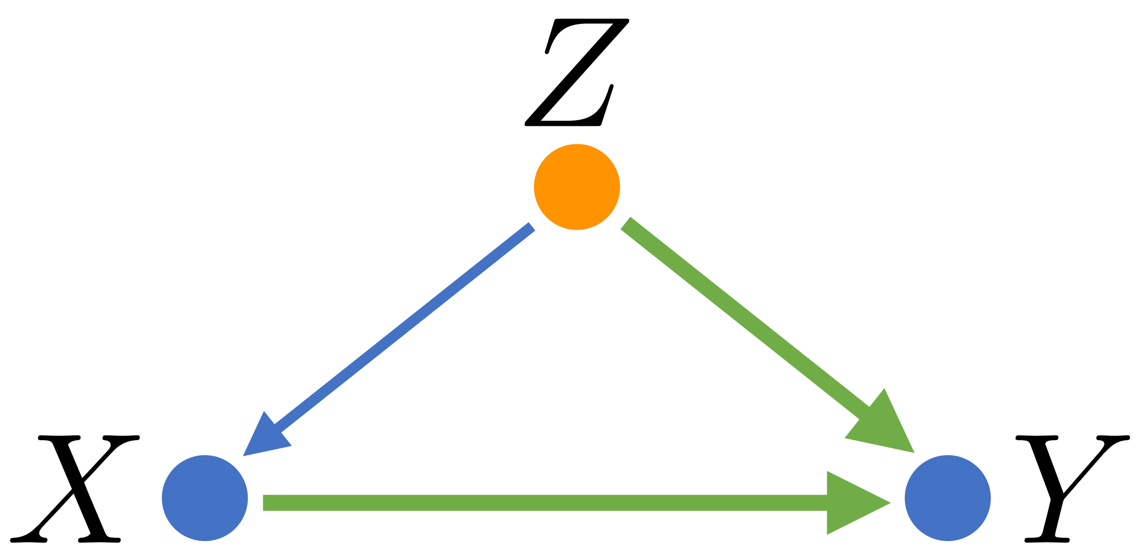
$$\widehat{\mathbb{P}}(Y=y|X=x) = \sum_z \widehat{\mathbb{P}}(Y=y|X=x, Z=z)\widehat{\mathbb{P}}(Z=z)$$

invariance relation!

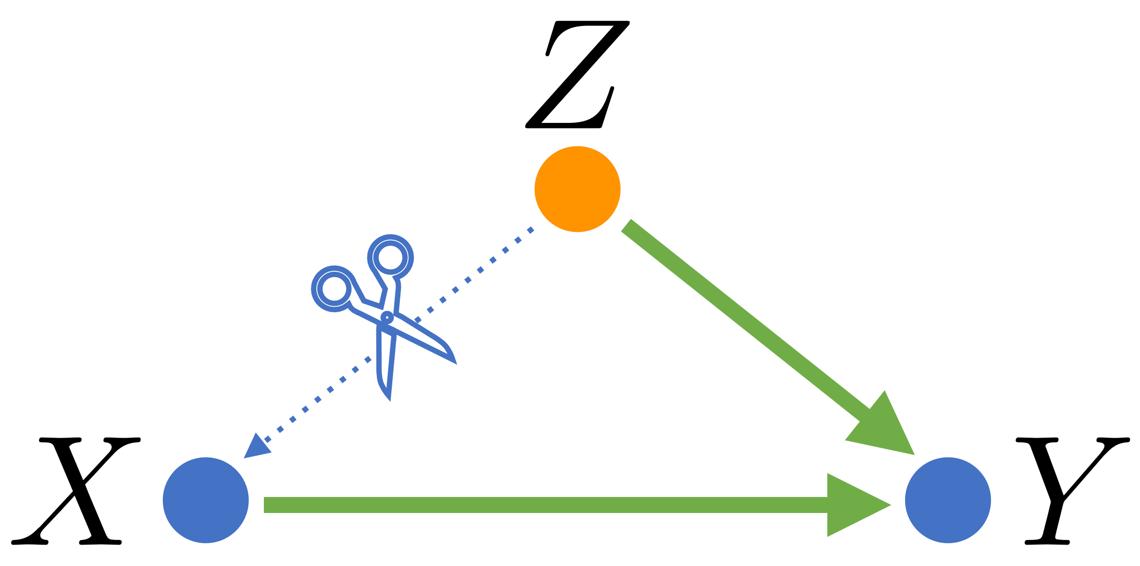
$$= \sum_z \mathbb{P}(Y=y|X=x, Z=z)\mathbb{P}(Z=z)$$

Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y=1|\mathtt{do}(X=1)) - \mathbb{P}(Y=1|\mathtt{do}(X=0))$



$$\mathbb{P}(Y=y|X=x) = \sum_z \mathbb{P}(Y=y|X=x, Z=z)\mathbb{P}(Z=z|X=x)$$



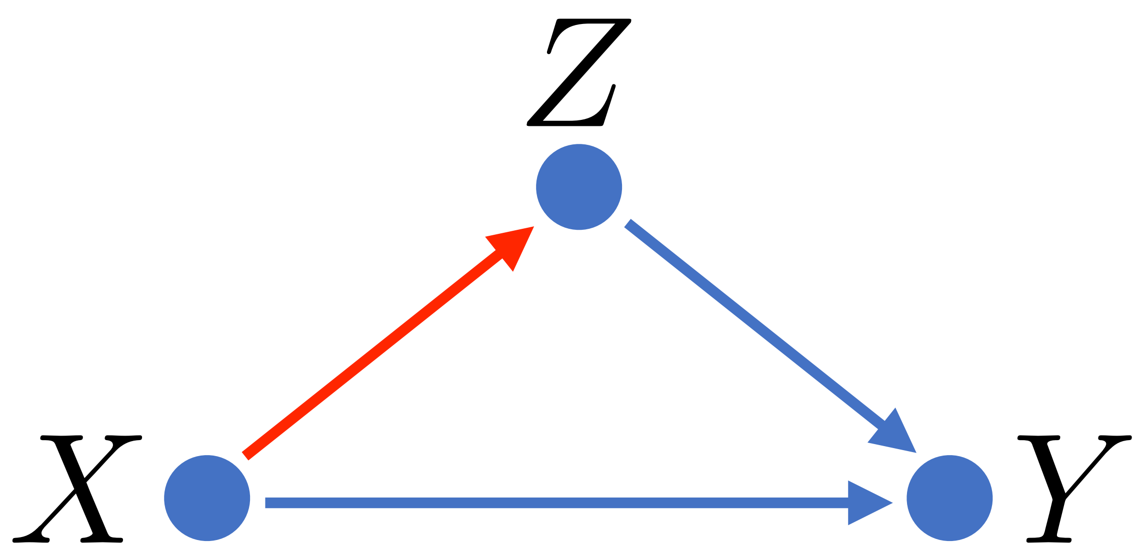
$$\widehat{\mathbb{P}}(Y=y|X=x) = \sum_z \widehat{\mathbb{P}}(Y=y|X=x, Z=z)\widehat{\mathbb{P}}(Z=z)$$

invariance relation!

$$= \sum_z \mathbb{P}(Y=y|X=x, Z=z)\mathbb{P}(Z=z)$$

Adjustment Formula

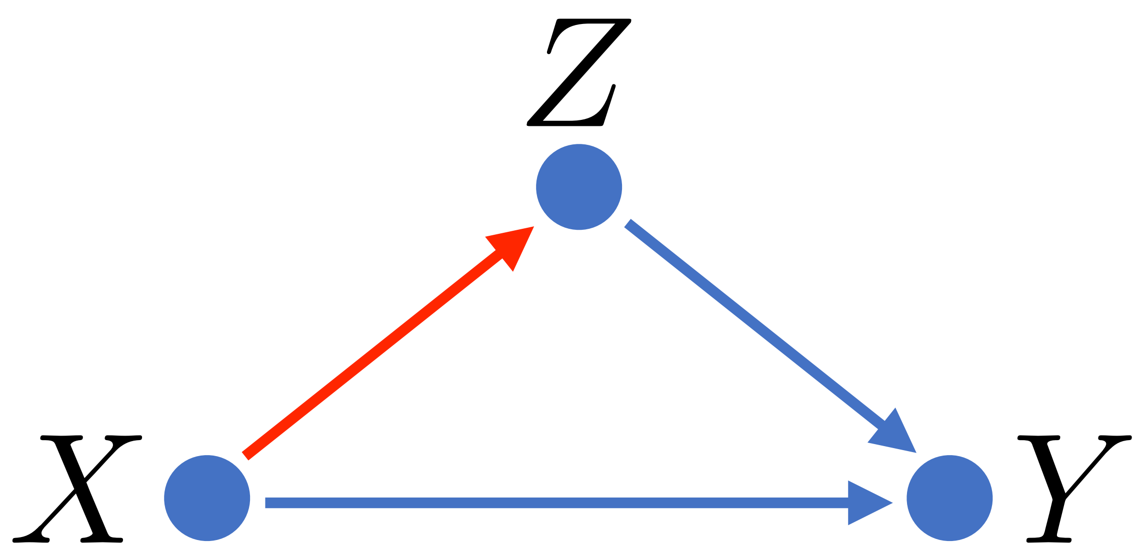
Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y = 1|\mathtt{do}(X = 1)) - \mathbb{P}(Y = 1|\mathtt{do}(X = 0))$



Quiz: 이 경우 intervention 의 효과는?

Adjustment Formula

Average **C**ausal **E**ffect **(ACE):** $\mathbb{P}(Y = 1|\mathtt{do}(X = 1)) - \mathbb{P}(Y = 1|\mathtt{do}(X = 0))$

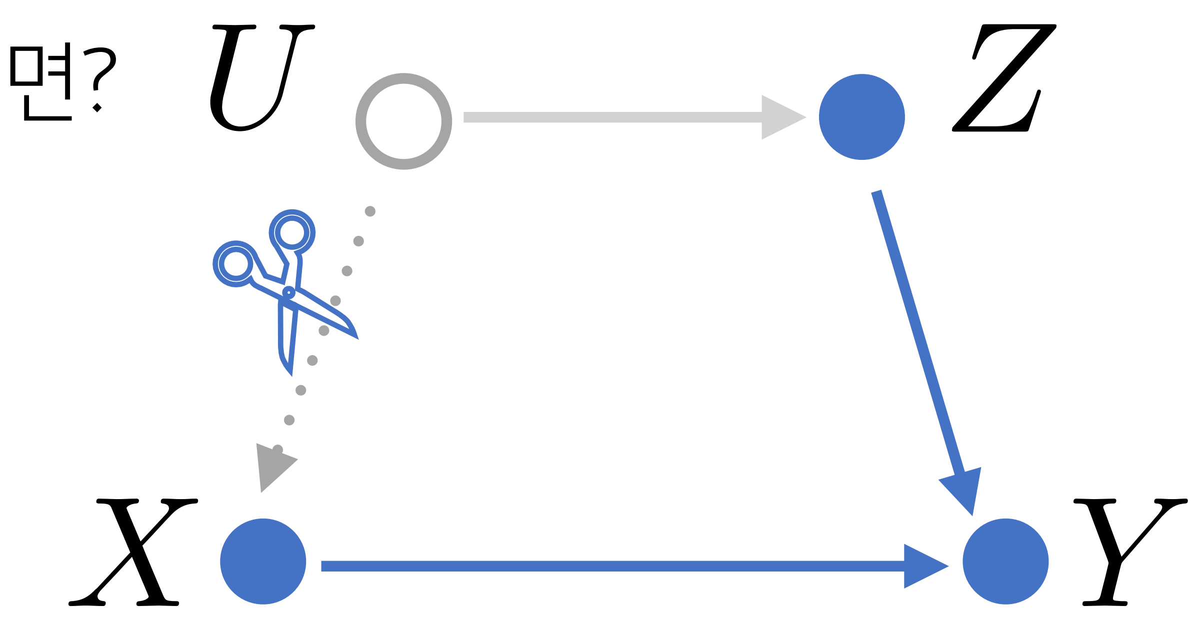


Answer: conditioning 이랑 같다!

$$\mathbb{P}(Y = y|\mathtt{do}(X = x)) = \mathbb{P}(Y = y|X = x)$$

Backdoor Criterion

U 는 측정 가능하지 않다면?

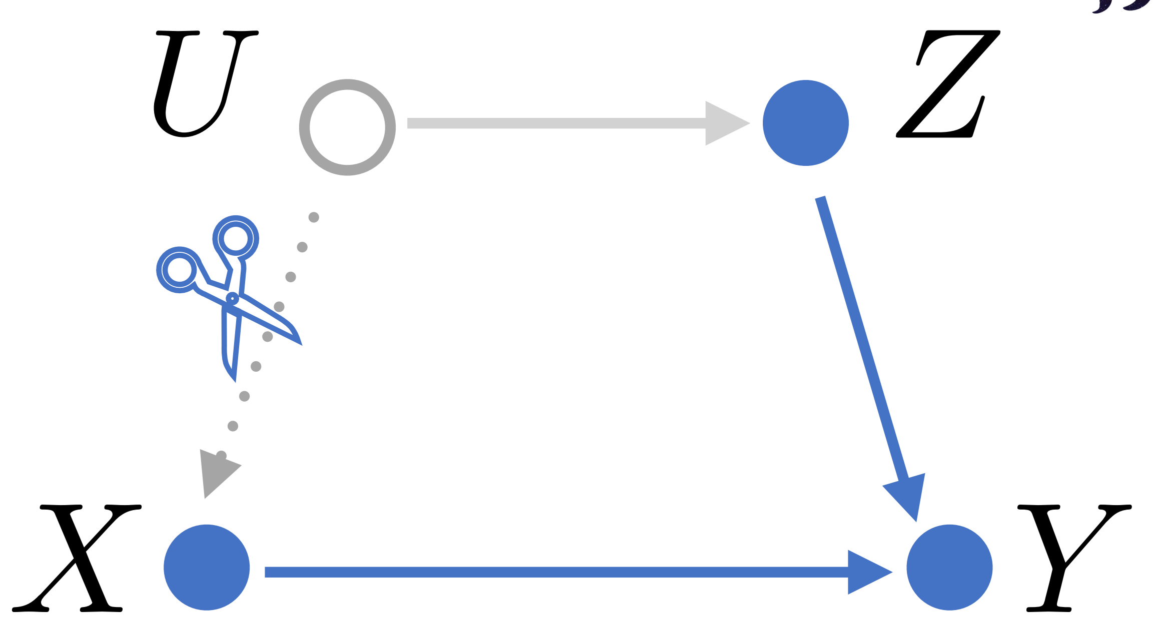


Quiz: 이 경우 intervention 의 효과는?

Backdoor Criterion

$X \leftarrow U \rightarrow Z \rightarrow Y$

Z 는 X 와 Y 사이에 X 를 effect 로 가지는 모든 path 를 block 하므로 adjust 할 수 있게 된다



Answer: Z 를 가지고 계산할 수 있다!

$$\mathbb{P}(Y = y | \texttt{do}(X = x)) = \sum_z \mathbb{P}(Y = y | X = x, Z = z)\mathbb{P}(Z = z)$$

Do-calculus

- 옆의 세 가지 rule 의 조합으로 **identifiable 인** 모든 intervention 은 joint 분포로 계산 가능하다
- d-separation 여부는 algebraic 하게 체크 가능
    - algorithm 존재 (많이 느림)

1. "Insertion/deletion of observations":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{z},\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X},\mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ have been removed.

2. "Action/observation exchange":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x},\mathbf{Z}=\mathbf{z})}(\mathbf{y}\,|\,\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{z},\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X},\mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ and outgoing edges from $\mathbf{Z}$ have been removed.

3. "Insertion/deletion of actions":

$$p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x},\mathbf{Z}=\mathbf{z})}(\mathbf{y}\,|\,\mathbf{w}) = p^{\mathfrak{C};do(\mathbf{X}:=\mathbf{x})}(\mathbf{y}\,|\,\mathbf{w})$$

   if $\mathbf{Y}$ and $\mathbf{Z}$ are d-separated by $\mathbf{X},\mathbf{W}$ in a graph where incoming edges in $\mathbf{X}$ and $\mathbf{Z}(\mathbf{W})$ have been removed. Here, $\mathbf{Z}(\mathbf{W})$ is the subset of nodes in $\mathbf{Z}$ that are not ancestors of any node in $\mathbf{W}$ in a graph that is obtained from $\mathcal{G}$ after removing all edges into $\mathbf{X}$.

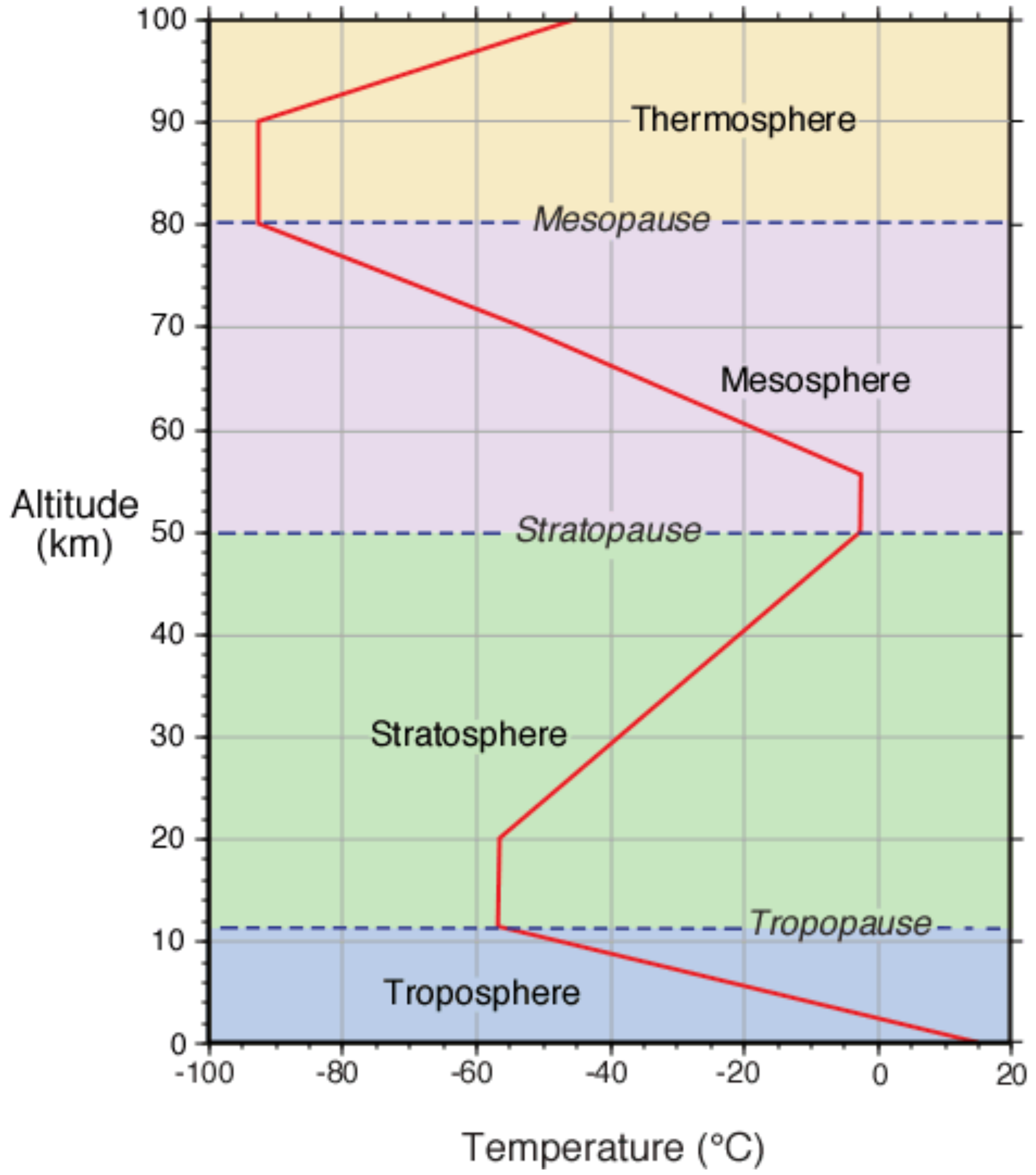
Identifiability

- **Identifiability for causal discovery**
  - Joint 분포만으로 causal direction 을 알아낼 수 있는가?
  - 적절한 **가정**을 추가하지 않으면 **불가능**

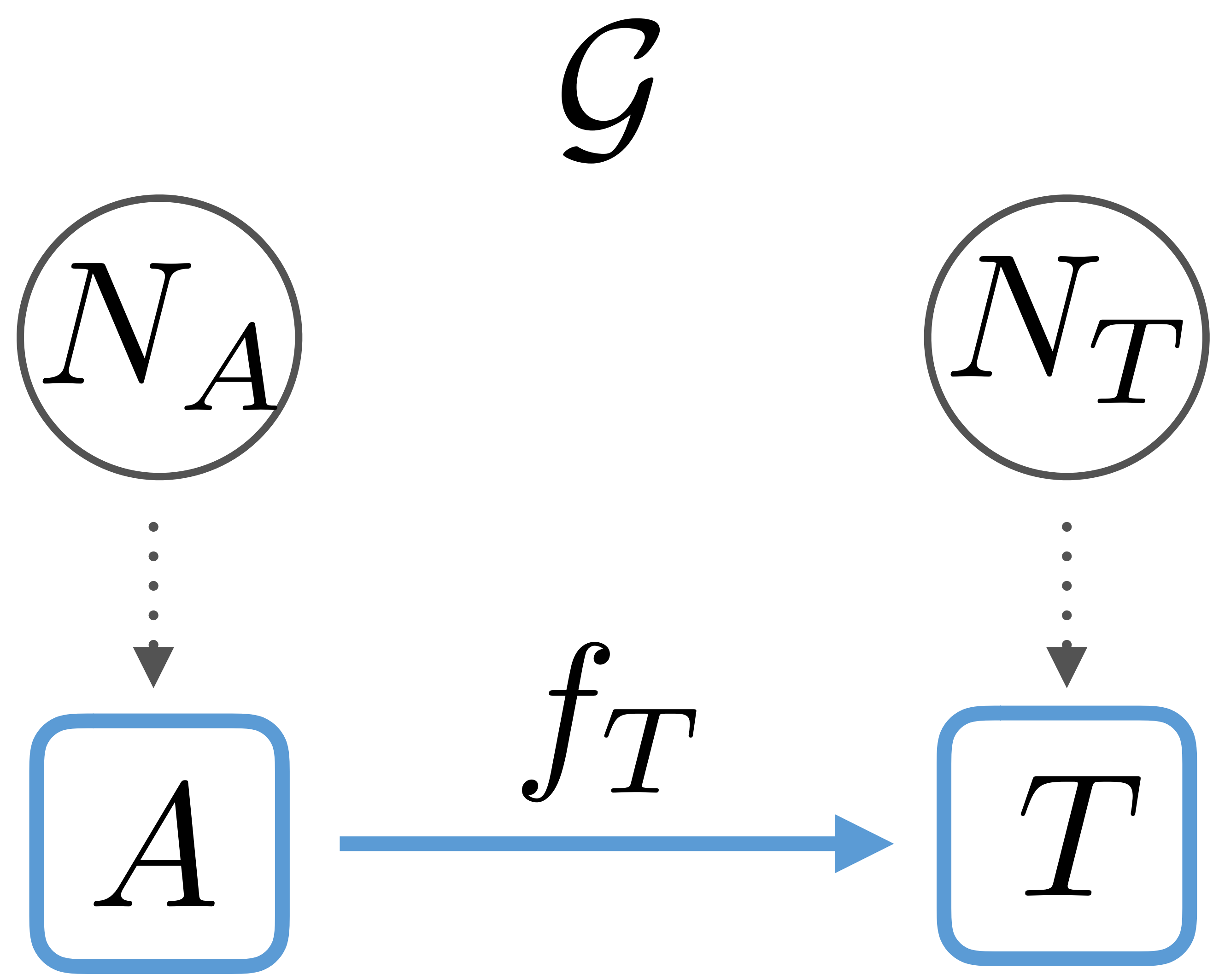
Proposition 4.1 (Non-uniqueness of graph structures) *For every joint distribution $P_{X,Y}$ of two real-valued variables, there is an SCM*

$$Y = f_Y(X, N_Y), \quad X \perp\!\!\!\perp N_Y,$$

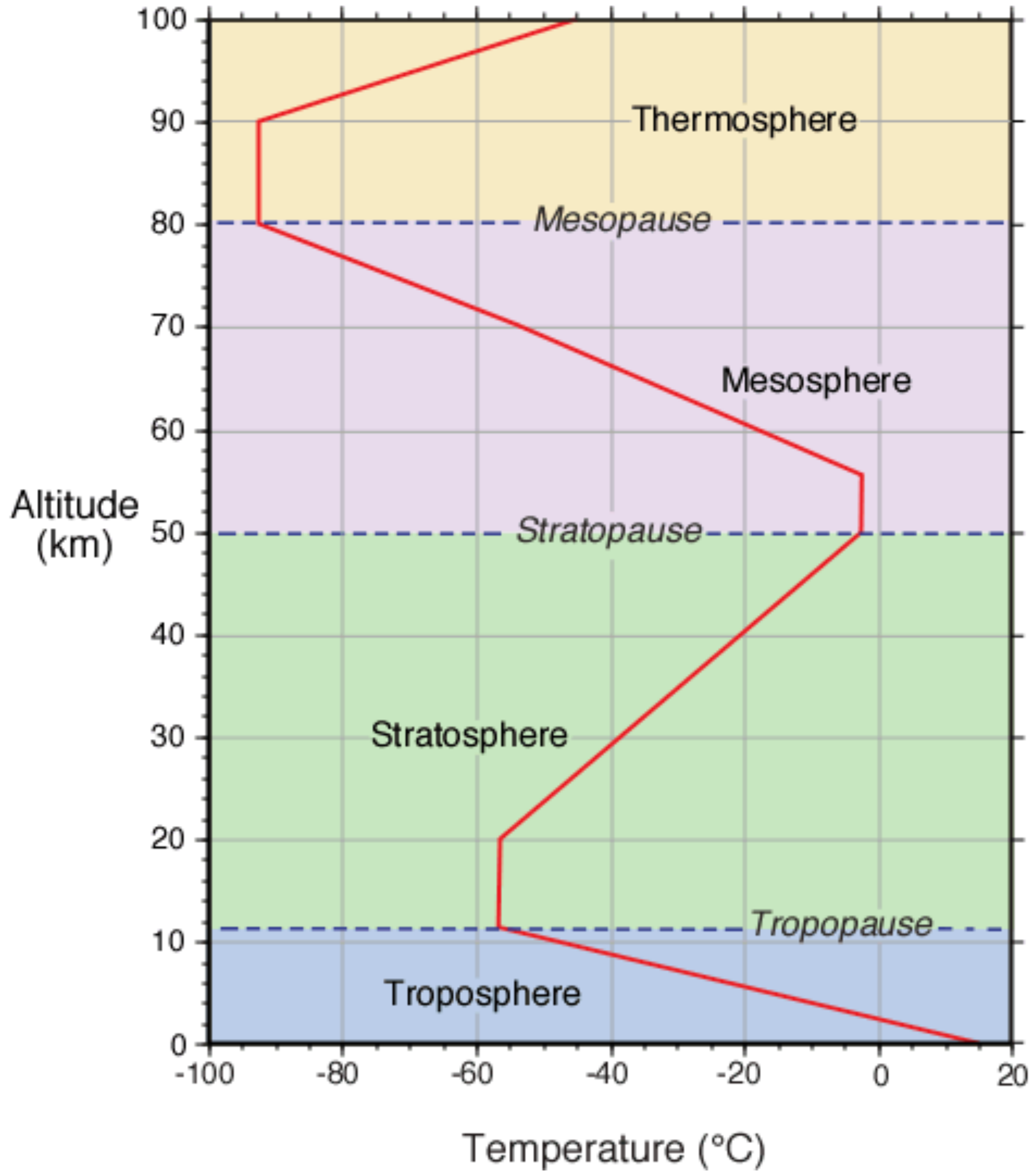
where f_Y is a measurable function and N_Y is a real-valued noise variable.

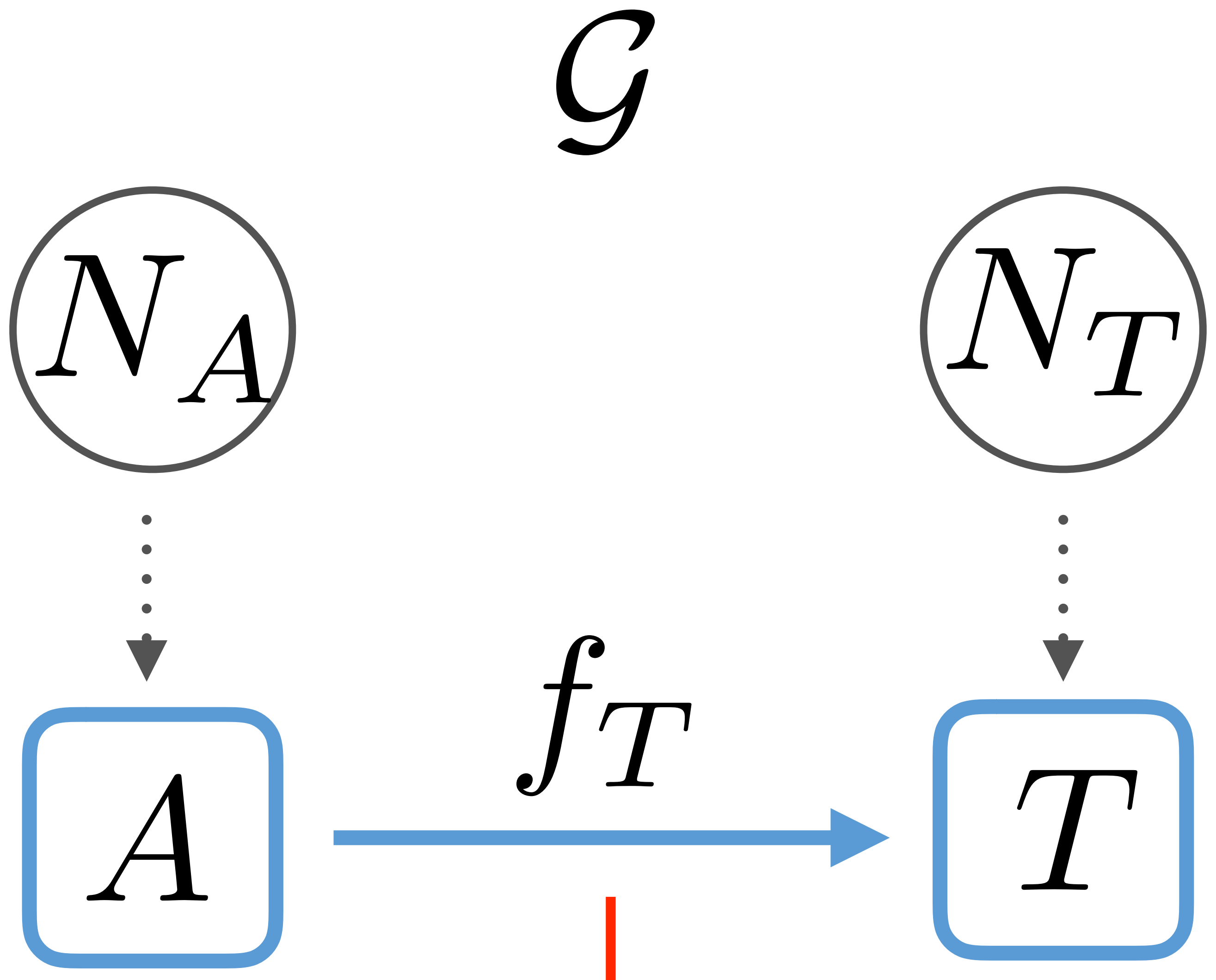


Altitude (km)
100
90
80
70
60
50
40
30
20
10
0
Thermosphere
Mesopause
Mesosphere
Stratopause
Stratosphere
Tropopause
Troposphere
-100
-80
-60
-40
-20
0
20
Temperature (°C)



$\mathcal{G}$
N_A
N_T
A
f_T
T



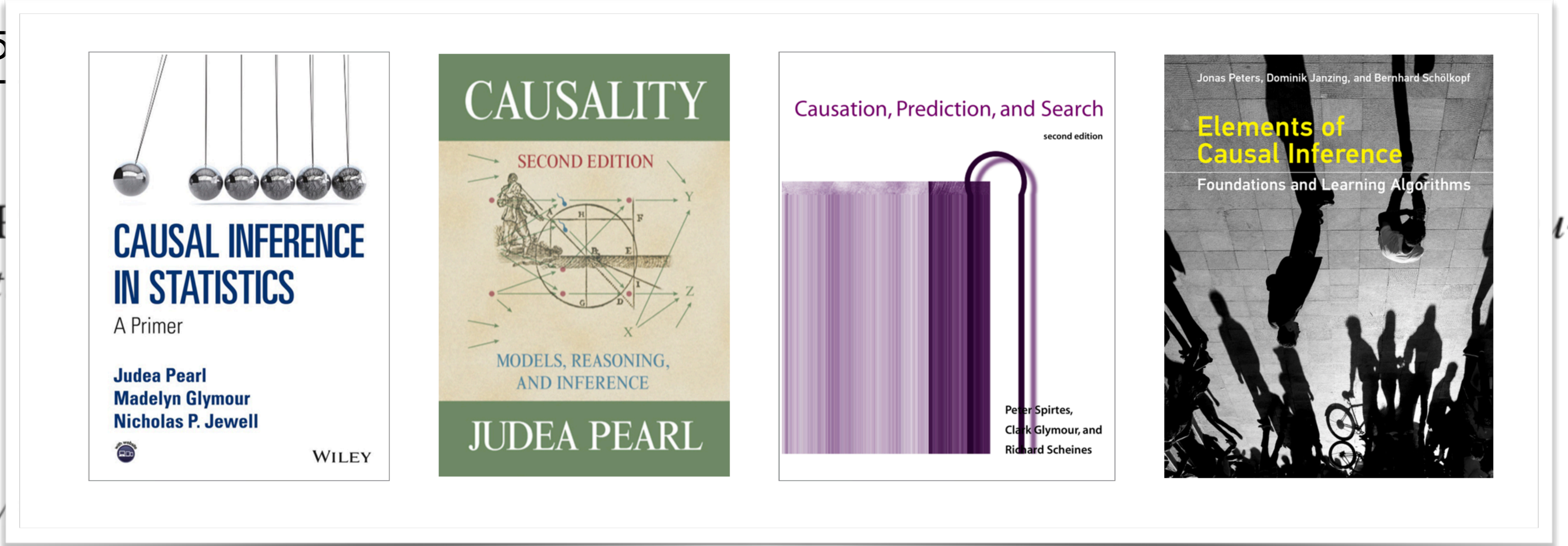


independence of mechanism

$$\mathbb{P}_{\texttt{japan}}(A,T) = \mathbb{P}(T|A)\mathbb{P}_{\texttt{japan}}(A)$$
$$\mathbb{P}_{\texttt{korea}}(A,T) = \mathbb{P}(T|A)\mathbb{P}_{\texttt{korea}}(A)$$

Identifiability

- **Identifiability for causal discovery**
  - Joint 분포만으로 causal direction 을 알아낼 수 있는가?
  - 적절



다같이 causality 의 세계에 빠져들어 보아요

Causal Learning for AI

Current Research & Applications

Machine Learning & Causality

- Most methods are developed under i.i.d hypothesis
  - good at prediction
  - poor at distribution change
- How to extract **knowledge** from data? → **causal discovery**
  - domain adaptation
  - eliminate confounding factors
  - meta learning (Y. Bengio)

Causality $\nrightarrow$ Better prediction

Example 1. *Consider the structural equation model [55]:*

$$\begin{aligned} X_1 &\leftarrow \text{Gaussian}(0, \sigma^2), \\ Y &\leftarrow X_1 + \text{Gaussian}(0, \sigma^2), \\ X_2 &\leftarrow Y + \text{Gaussian}(0, 1). \end{aligned}$$

causality 을 안다고 해서 predictive power
가 더 좋아진다고 얘기할 순 없다

As we formalize in Section 4, the set of all environments $\mathcal{E}_{\text{all}}$ contains all modifications of the structural equations for X_1 and X_2, and those varying the noise of Y within a finite range $[0, \sigma^2_{\text{MAX}}]$. For instance, $e \in \mathcal{E}_{\text{all}}$ may replace the equation of X_2 by $X_2^e \leftarrow 10^6$, or vary σ^2 within this finite range . To ease exposition consider:

$$\mathcal{E}_{\text{tr}} = \{\text{replace } \sigma^2 \text{ by } 10, \text{ replace } \sigma^2 \text{ by } 20\}.$$

Then, to predict Y from (X_1, X_2) using a least-squares predictor $\hat{Y}^e = X_1^e \hat{\alpha}_1 + X_2^e \hat{\alpha}_2$ for environment e, we can:

- regress from X_1^e, to obtain $\hat{\alpha}_1 = 1$ and $\hat{\alpha}_2 = 0$,
- regress from X_2^e, to obtain $\hat{\alpha}_1 = 0$ and $\hat{\alpha}_2 = \sigma(e)/(\sigma(e) + \frac{1}{2})$,
- regress from (X_1^e, X_2^e), to obtain $\hat{\alpha}_1 = 1/(\sigma(e) + 1)$ and $\hat{\alpha}_2 = \sigma(e)/(\sigma(e) + 1)$.

Invariant Risk Minimization, Martin Arjovsky et al., 2019

Causality → Stable Learning

Example 1. *Consider the structural equation model [55]:*

$$X_1 \leftarrow \text{Gaussian}(0, \sigma^2),$$
$$Y \leftarrow X_1 + \text{Gaussian}(0, \sigma^2),$$
$$X_2 \leftarrow Y + \text{Gaussian}(0, 1).$$

As we formalize in Section 4, the set of all environments $\mathcal{E}_{\text{all}}$ contains all modifications of the structural equations for X_1 and X_2, and those varying the noise of Y within a finite range $[0, \sigma^2_{\text{MAX}}]$. For instance, $e \in \mathcal{E}_{\text{all}}$ may replace the equation of X_2 by $X_2^e \leftarrow 10^6$, or vary σ^2 within this finite range . To ease exposition consider:

$$\mathcal{E}_{\text{tr}} = \{\text{replace } \sigma^2 \text{ by } 10, \ \text{replace } \sigma^2 \text{ by } 20\}.$$

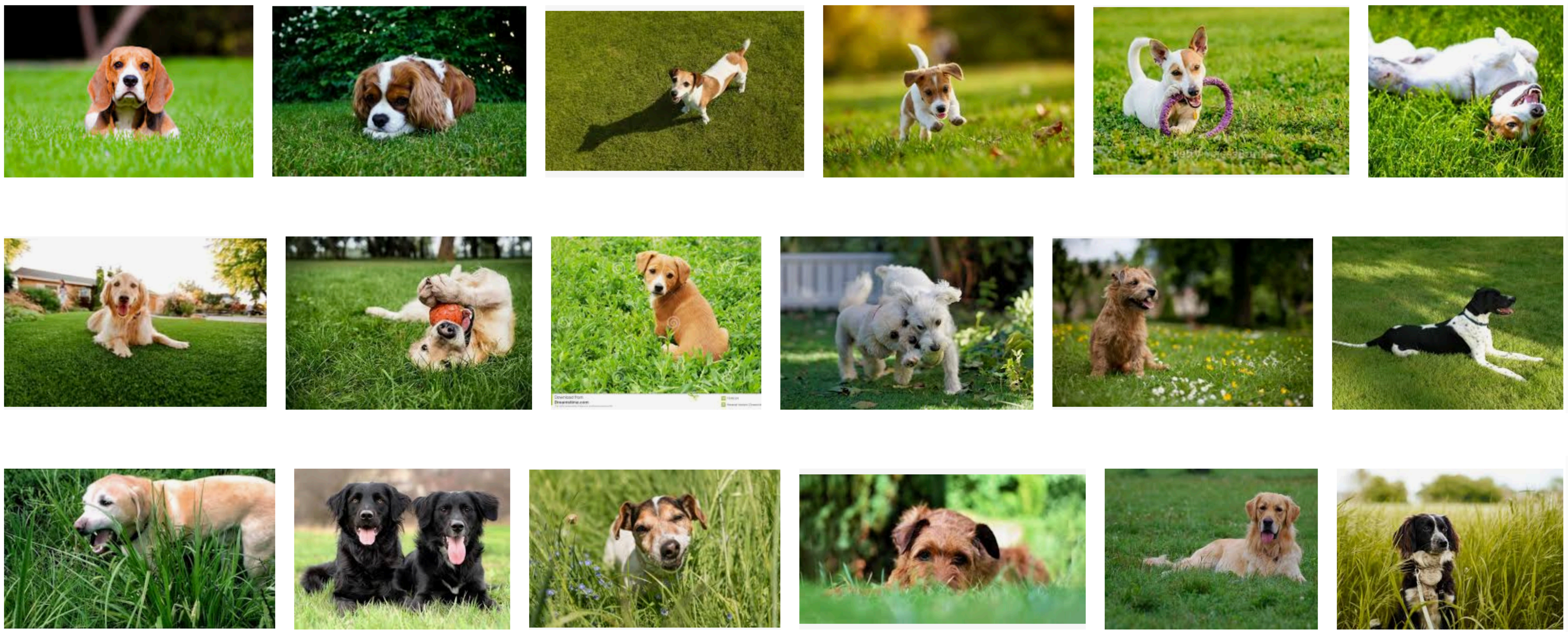
Then, to predict Y from (X_1, X_2) using a least-squares predictor $\hat{Y}^e = X_1^e \hat{\alpha}_1 + X_2^e \hat{\alpha}_2$ for environment e, we can:

- regress from X_1^e, to obtain $\boxed{\hat{\alpha}_1 = 1 \text{ and } \hat{\alpha}_2 = 0,}$
- regress from X_2^e, to obtain $\hat{\alpha}_1 = 0$ and $\hat{\alpha}_2 = \sigma(e)/(\sigma(e) + \frac{1}{2})$,
- regress from (X_1^e, X_2^e), to obtain $\boxed{\hat{\alpha}_1 = 1/(\sigma(e) + 1) \text{ and } \hat{\alpha}_2 = \sigma(e)/(\sigma(e) + 1).}$

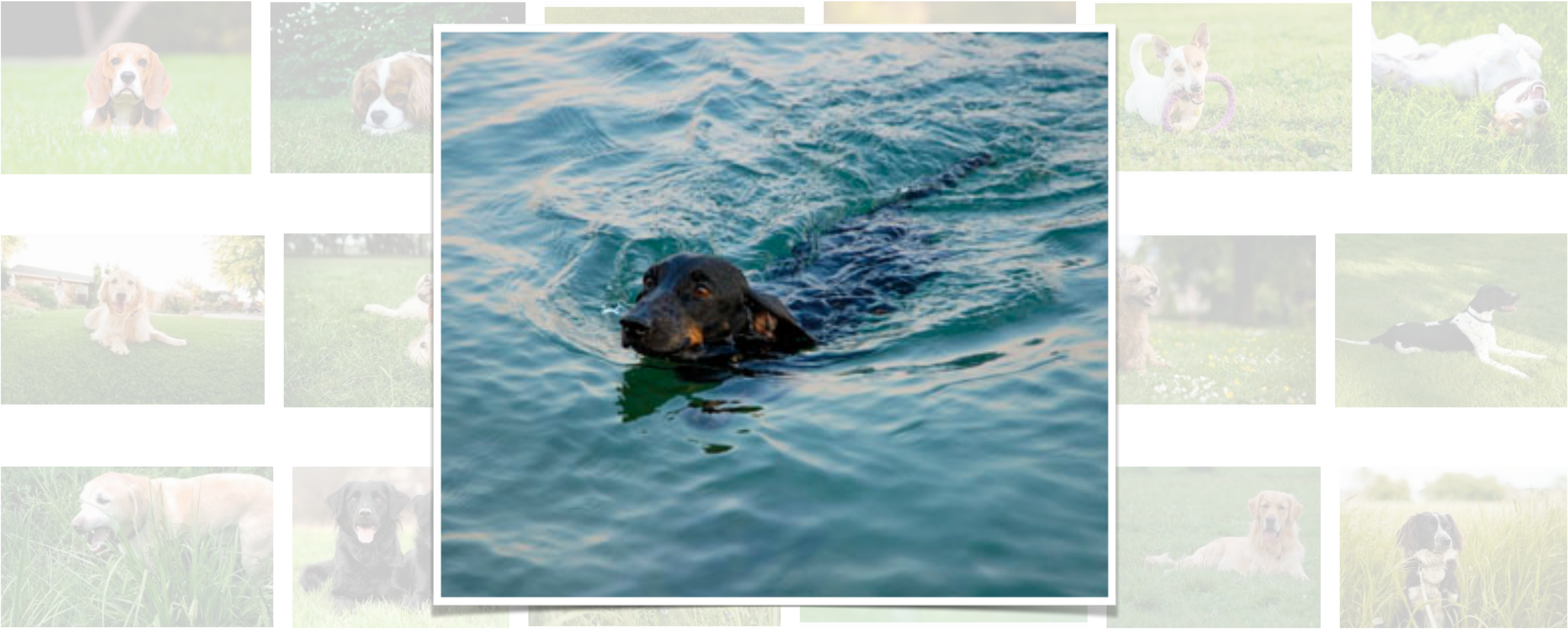
causality 을 알면 distributional change 에 robust 하게 대응할 수 있다

Invariant Risk Minimization, Martin Arjovsky et al., 2019

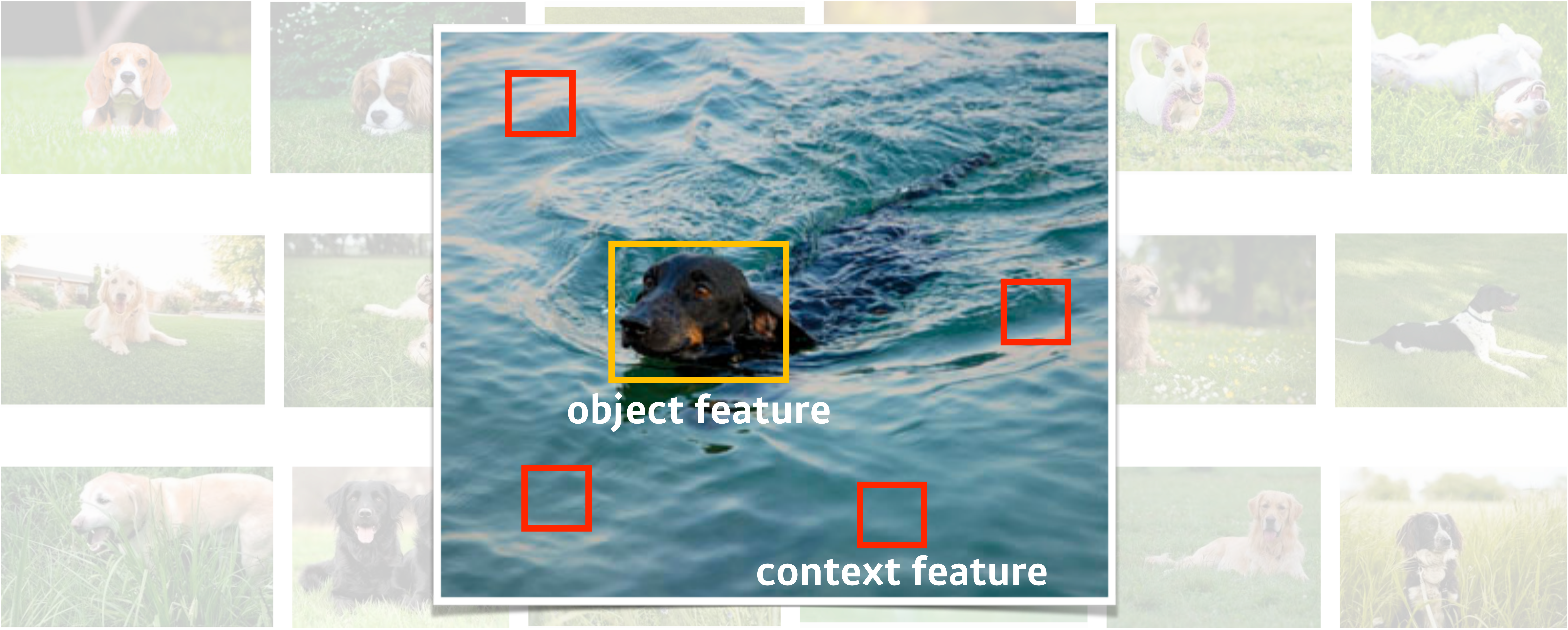
Causality for Stable Learning



Causality for Stable Learning



Causality for Stable Learning



Causal Discovery in Visual Domain

Discovering Causal Signals in Images

David Lopez-Paz
Facebook AI Research
dlp@fb.com

Robert Nishihara
UC Berkeley
rkn@eecs.berkeley.edu

Soumith Chintala
Facebook AI Research
soumith@fb.com

Bernhard Schölkopf
MPI for Intelligent Systems
bs@tue.mpg.de

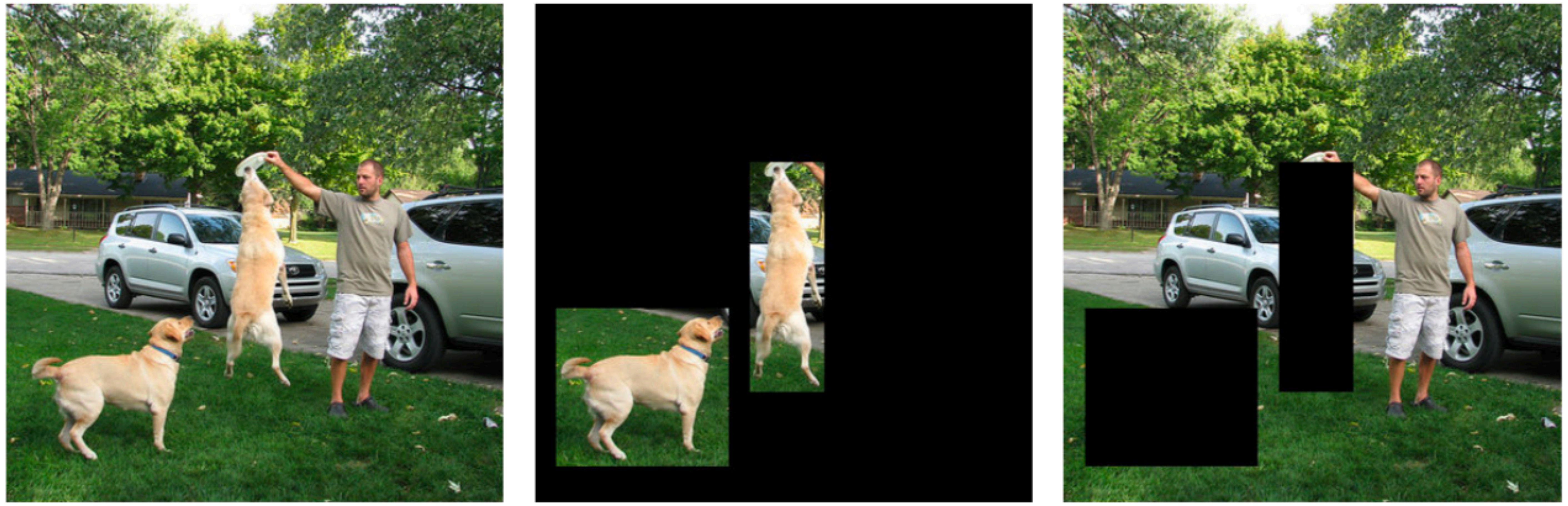
Léon Bottou
Facebook AI Research
leon@bottou.org

Discovering Causal Signals in Images, David Lopez-Paz et al., CVPR 2017

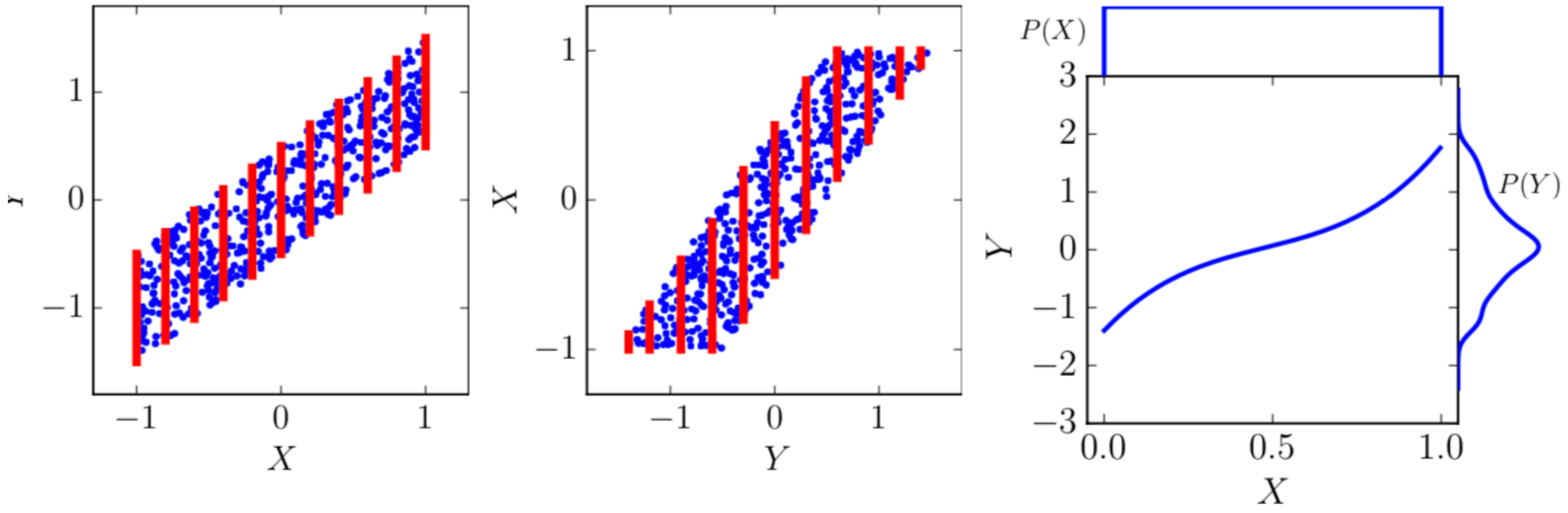
Causal Discovery in Visual Domain

Hypothesis 2. *There exists an observable statistical dependence between object features and anticausal features. The statistical dependence between context features and causal features is nonexistent or much weaker.*

여기서 causal feature 는 label 과 feature 간의 관계를 ANM 으로 모델링한 관계에서 추측



(a) x_j (b) x_j^o **object** feature (c) x_j^c **context** feature



(a) ANM $X \to Y$. (b) ANM $Y \to X$ (c) Monotonic $X \to Y$.

Discovering Causal Signals in Images, David Lopez-Paz et al., CVPR 2017

Meta Learning for Causal Discovery

LEARNING NEURAL CAUSAL MODELS FROM UNKNOWN INTERVENTIONS

Nan Rosemary Ke[* 1,2], Olexa Bilaniuk[* 1], Anirudh Goyal[1], Stefan Bauer[5], Hugo Larochelle[4], Chris Pal[1,2,3], Yoshua Bengio[1†]
[1] Mila, Université de Montréal
[2] Mila, Polytechnique Montréal
[3] Element AI
[4] Google AI
[5] Empirical Inference Group, Max Planck Institute for Intelligent Systems
[†]CIFAR Senior Fellow.
[*] **Authors contributed equally**
rosemary.nan.ke@gmail.com

Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Motivation

- Causal discovery provided interventions are unknown
  - not only learn the causal graph structure
    - also predict the intervention accurately
    - handle unknown interventions
    - model the effect of interventions
  - model the underlying causal structure
    - avoid an exponential search over all possible DAG

Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Learning causal structure

- parallel connections to **meta learning**
  - **inner loop**: fast adaptation to the distribution change
  - **outer loop**: learning the stationary meta-parameters

γ causal graph structure

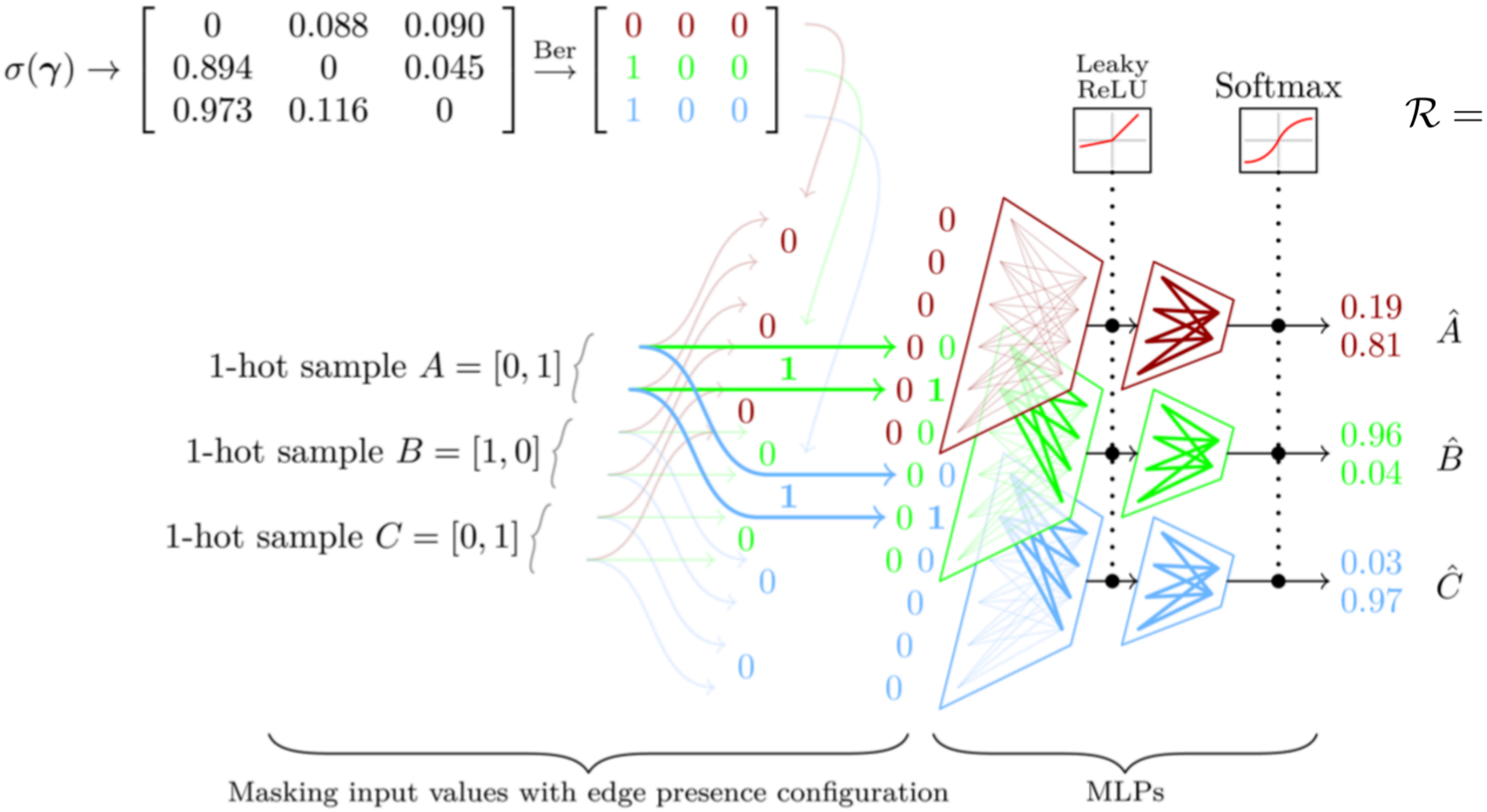
θ_{slow} meta-parameter

$\theta = \theta_{\text{fast}} + \theta_{\text{slow}}$

$$\mathcal{R} = -\mathbb{E}_{X \sim D_{\text{int}}}[\log \mathbb{E}_{C \sim \text{Ber}(\gamma)}[\prod_i \mathcal{L}_{C,i}(X; \theta_{\text{slow}})]]$$

Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Learning causal structure



$$\mathcal{R} = -\mathbb{E}_{X \sim D_{\text{int}}}[\log \mathbb{E}_{C \sim \text{Ber}(\gamma)}[\prod_i \mathcal{L}_{C,i}(X; \theta_{\text{slow}})]]$$

γ causal graph structure

θ_{slow} meta-parameter

θ_{fast} parameter

$\theta = \theta_{\text{fast}} + \theta_{\text{slow}}$

Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Learning causal structure

Figure 3: Earthquake: Learned edges at three different stages of training.

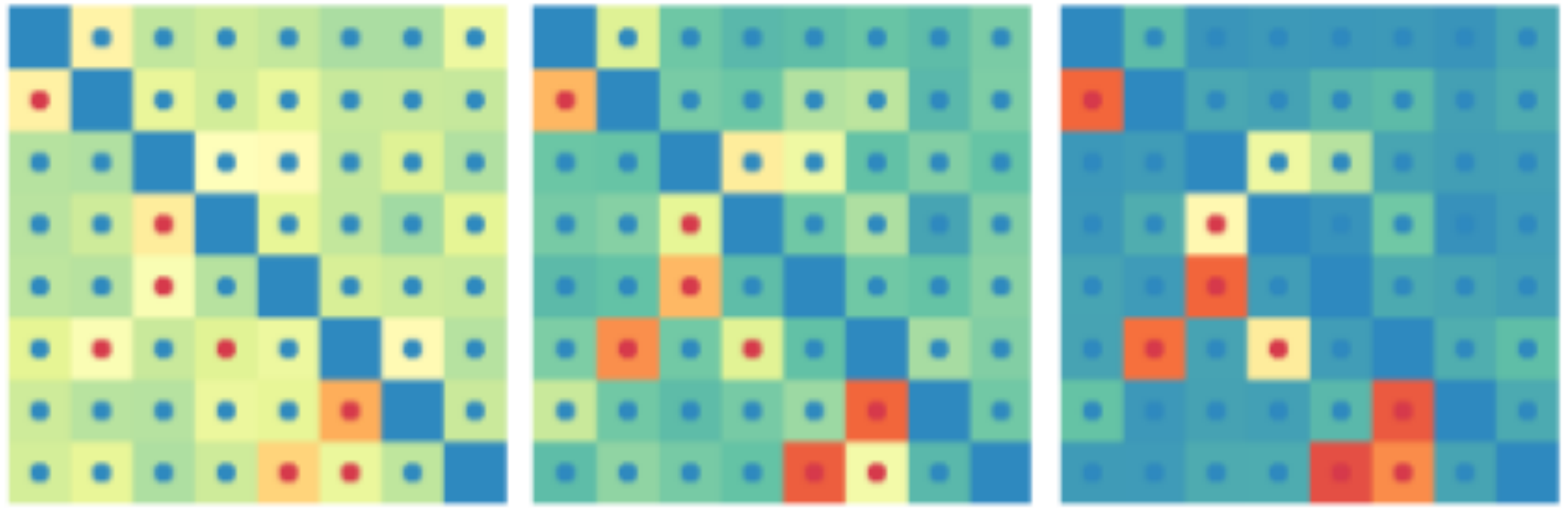


Figure 4: Asia: Learned edges at three different stages of training.

Table 3: **Intervention Prediction Accuracy:** (identify on which variable the intervention took place)

3 variables	4 variables	5 variables	8 variables
95 %	90 %	81 %	63 %

Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Learning causal structure

Figure 3: Earthquake: Learned edges at three different stages of training.

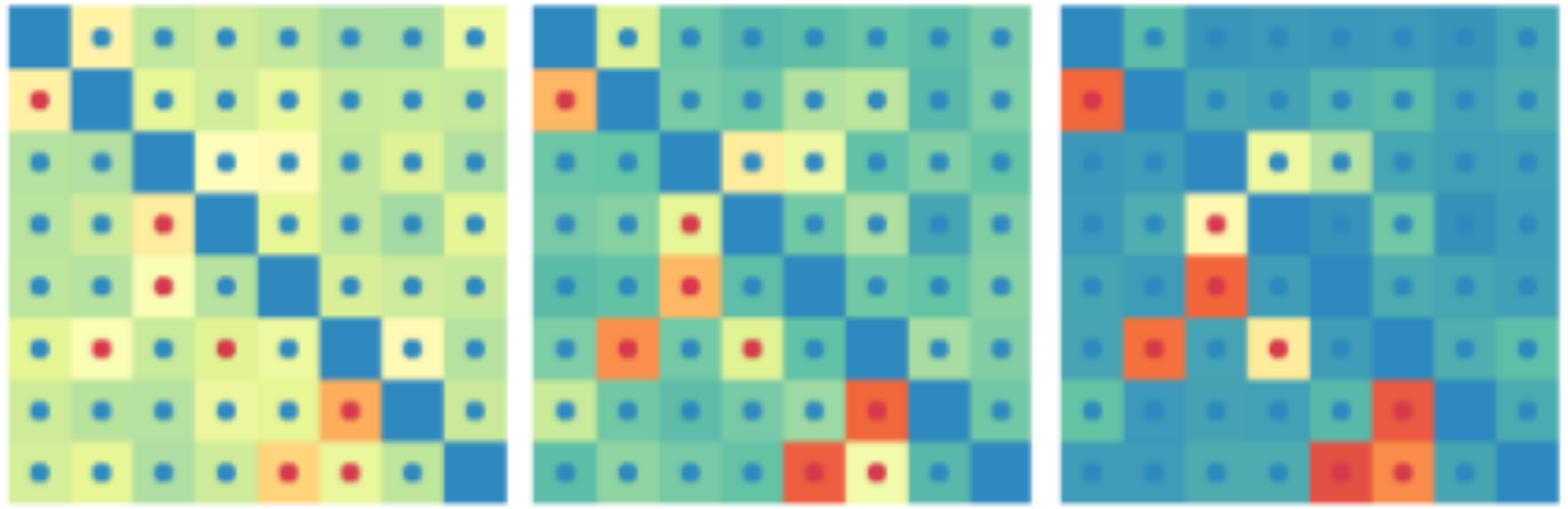


Figure 4: Asia: Learned edges at three different stages of training.

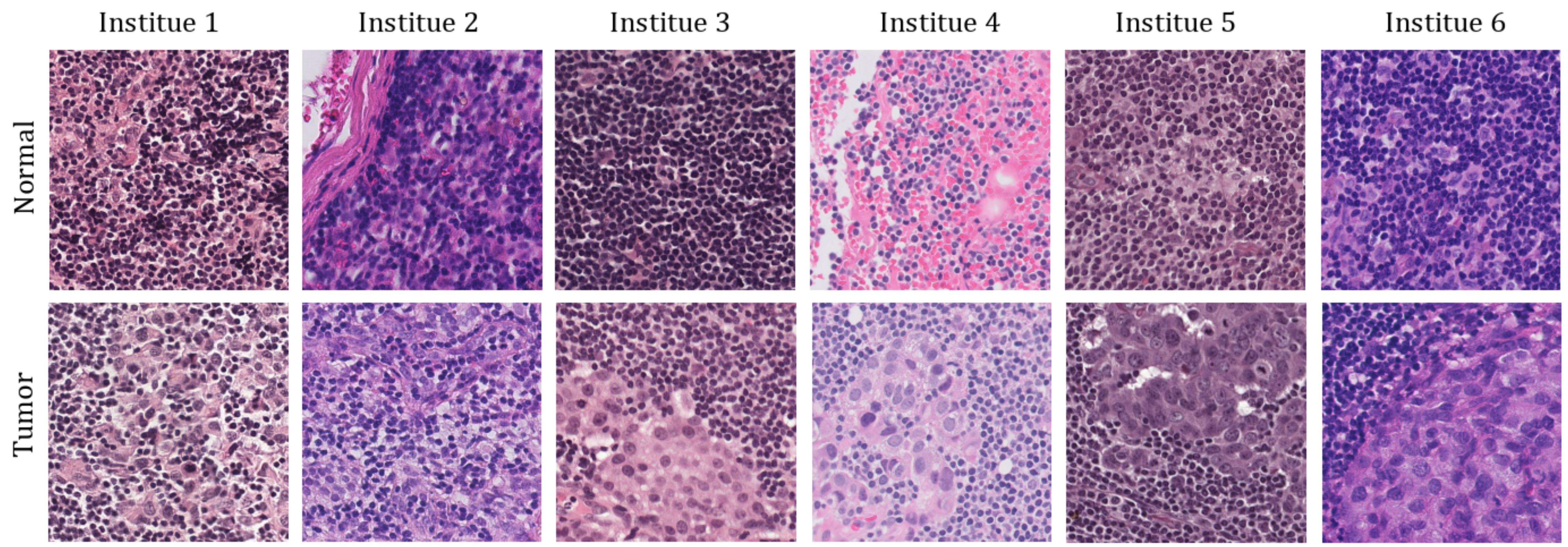
Table 3: **Intervention Prediction Accuracy:** (identify on which variable the intervention took place)

3 variables	4 variables	5 variables	8 variables
95 %	90 %	81 %	63 %

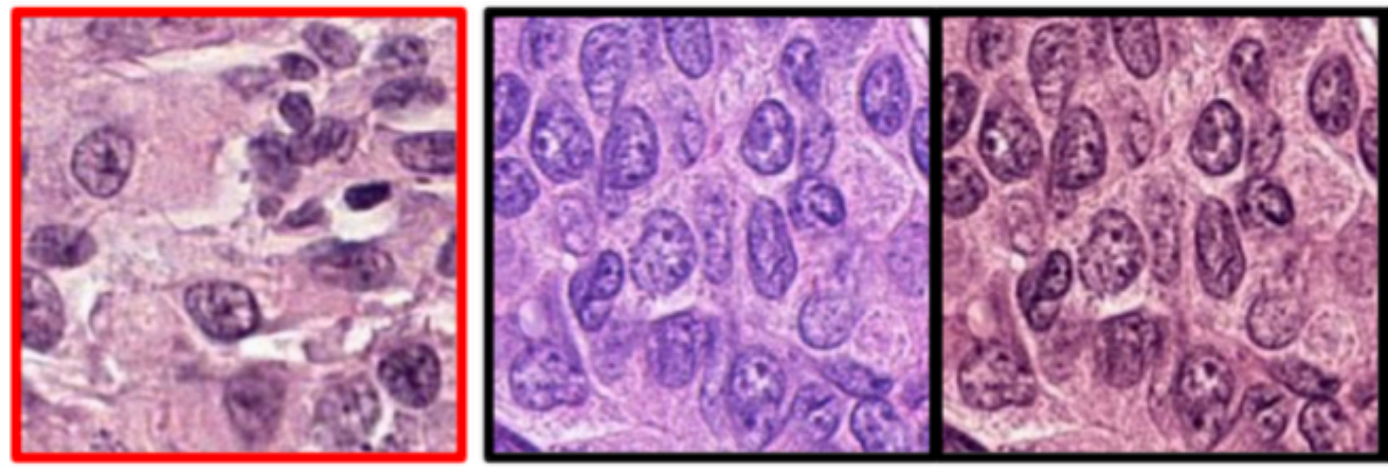
Learning Neural Causal Models From Unknown Interventions, Nan Rosemary Ke et al., 2019

Domain Adaptation in Medical AI

- **Stain-Style Transfer** using GAN
  - Feature is invariant w.r.t stain-color distribution



$$\mathbb{P}(Y = \text{cancer} | \mathcal{C} = \text{original}, X)$$
$$\rightarrow \mathbb{P}(Y = \text{cancer} | \text{do}(\mathcal{C}) = \text{target}, X)$$



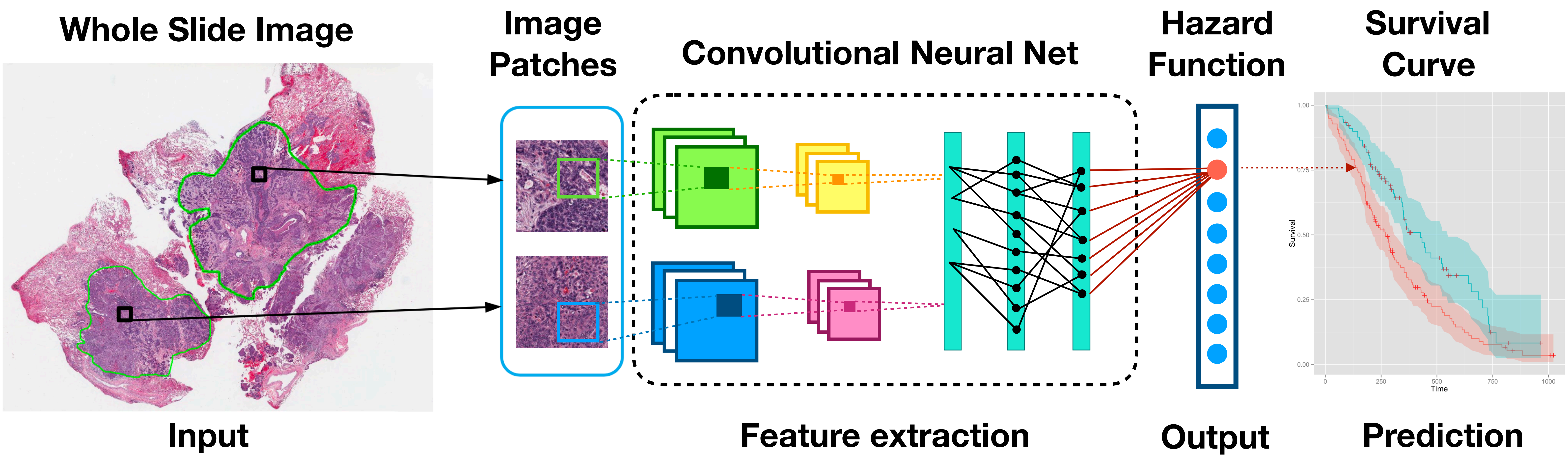
(a) Target (b) Original (c) SST

Neural Stain-Style Transfer Learning using GAN for Histopathological Images; (2017), **ACML**

Joint work with H. Cho (SNU), G. Choi (Yonsei Univ.), H. Min (KAIST)

Treatment Recommendation via ML

- Survival analysis for **treatment recommendation**
  - ex) chemotherapy vs radiotherapy
- Identification of patient subgroups in non-experimental data



Identification of Patient Subgroups who Can Benefit from Adjuvant Treatments; (2020)

Joint work with J. Kim (SMC), J. Son (SMC), S. Seo (SMC)

Causal Inference vs RL

Causal Inference

- no data acquisition for observational data
  - **medicine**
- estimation of future potential outcome
- find optimal treatment rule as sequence of action

Reinforcement Learning

- control over data acquisition
  - **simulation**
- delayed reward
- propagate backward to estimate value at an earlier state to find optimal rule

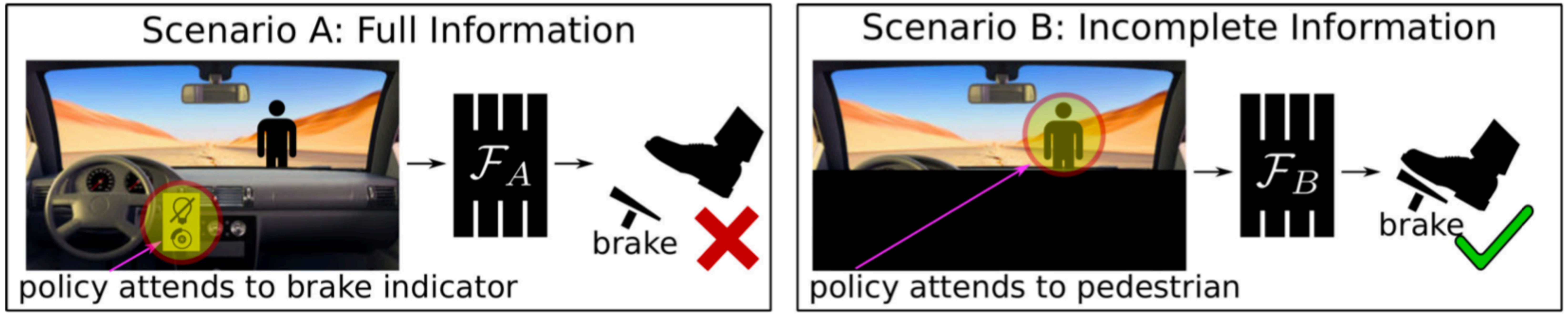
Causality for Imitation Learning

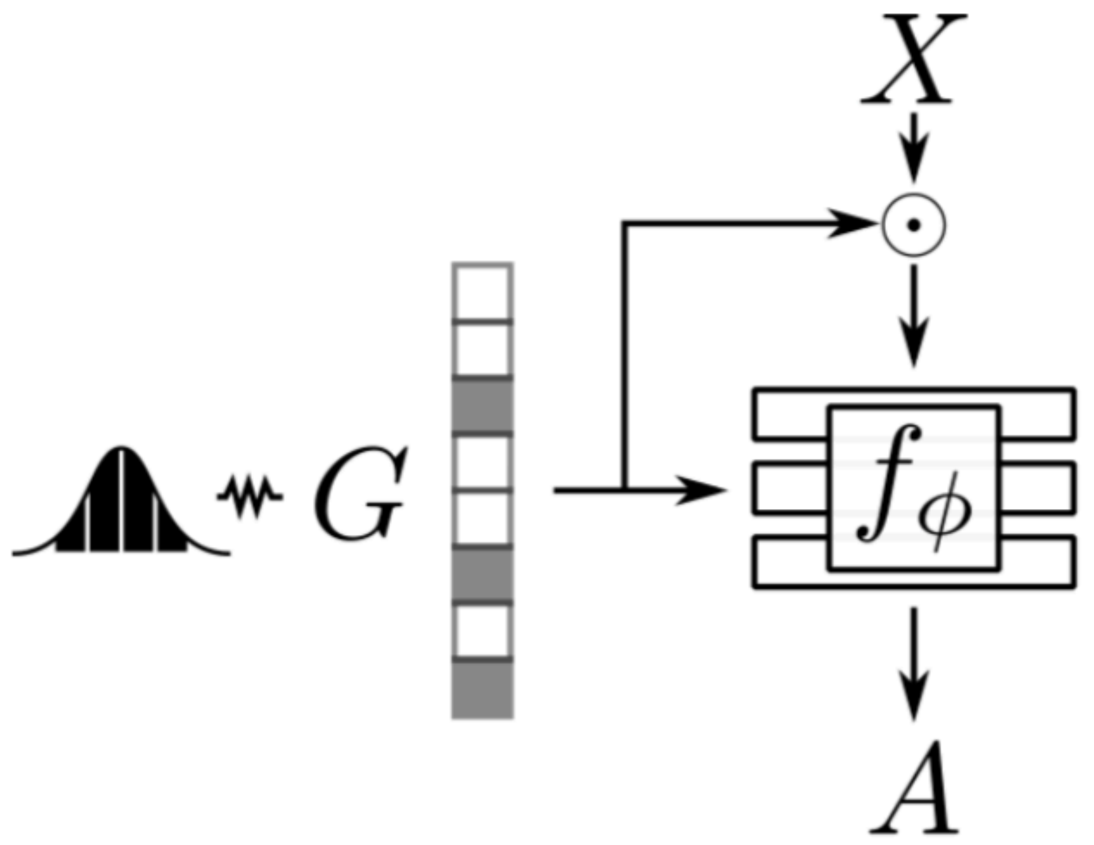
Causal Confusion in Imitation Learning

Pim de Haan[*1], **Dinesh Jayaraman**[†‡], **Sergey Levine**[†]
[*]Qualcomm AI Research, University of Amsterdam,
[†]Berkeley AI Research, [‡] Facebook AI Research

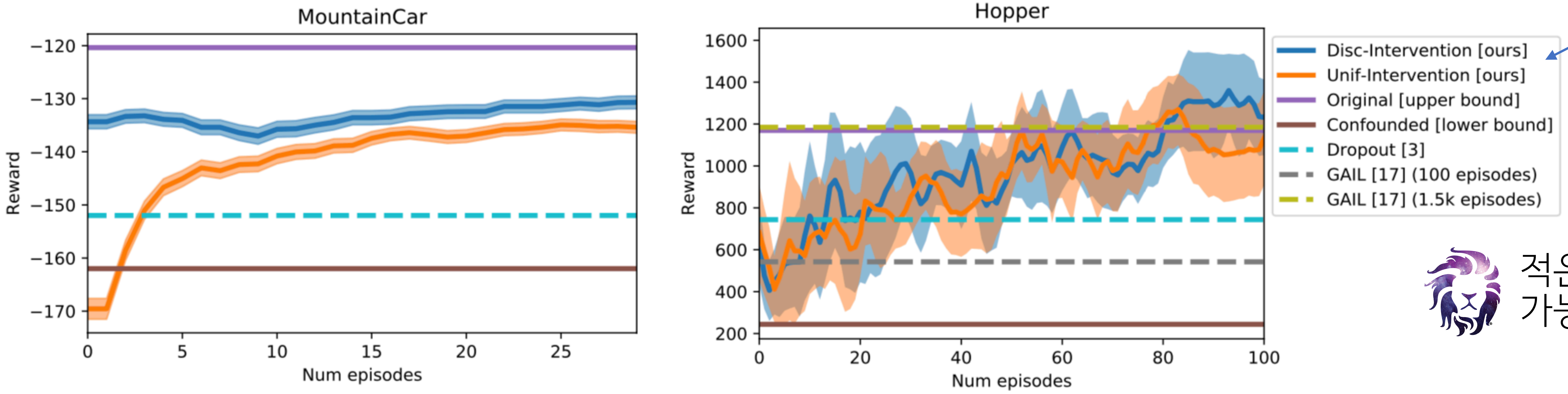
Causal Confusion in Imitation Learning, Pim de Haan et al., NIPS 2019

Causality for Imitation Learning





graph-parametrized policy

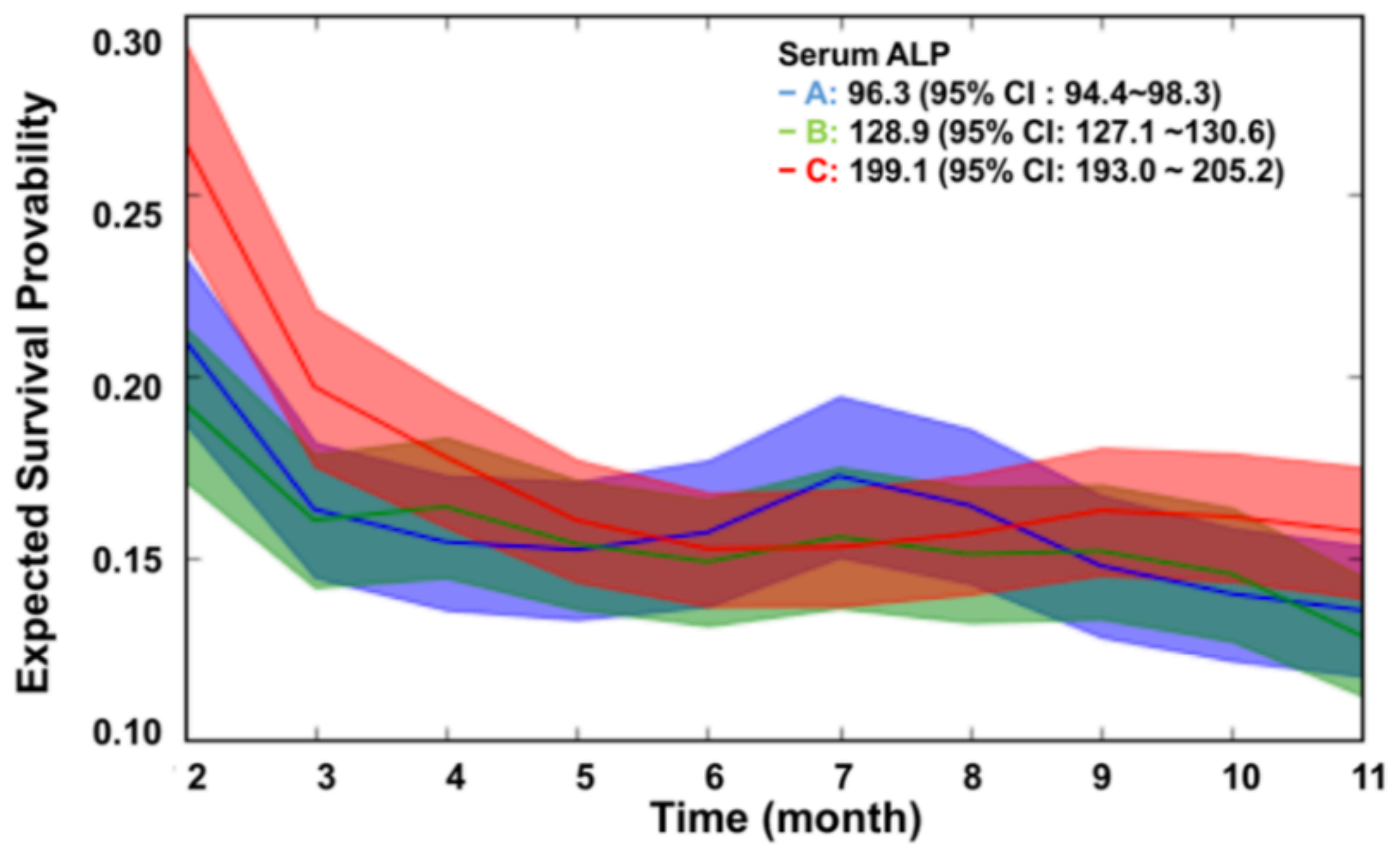


적은 episode 만으로 reward 달성이 가능하다 → imitation 을 잘 한다

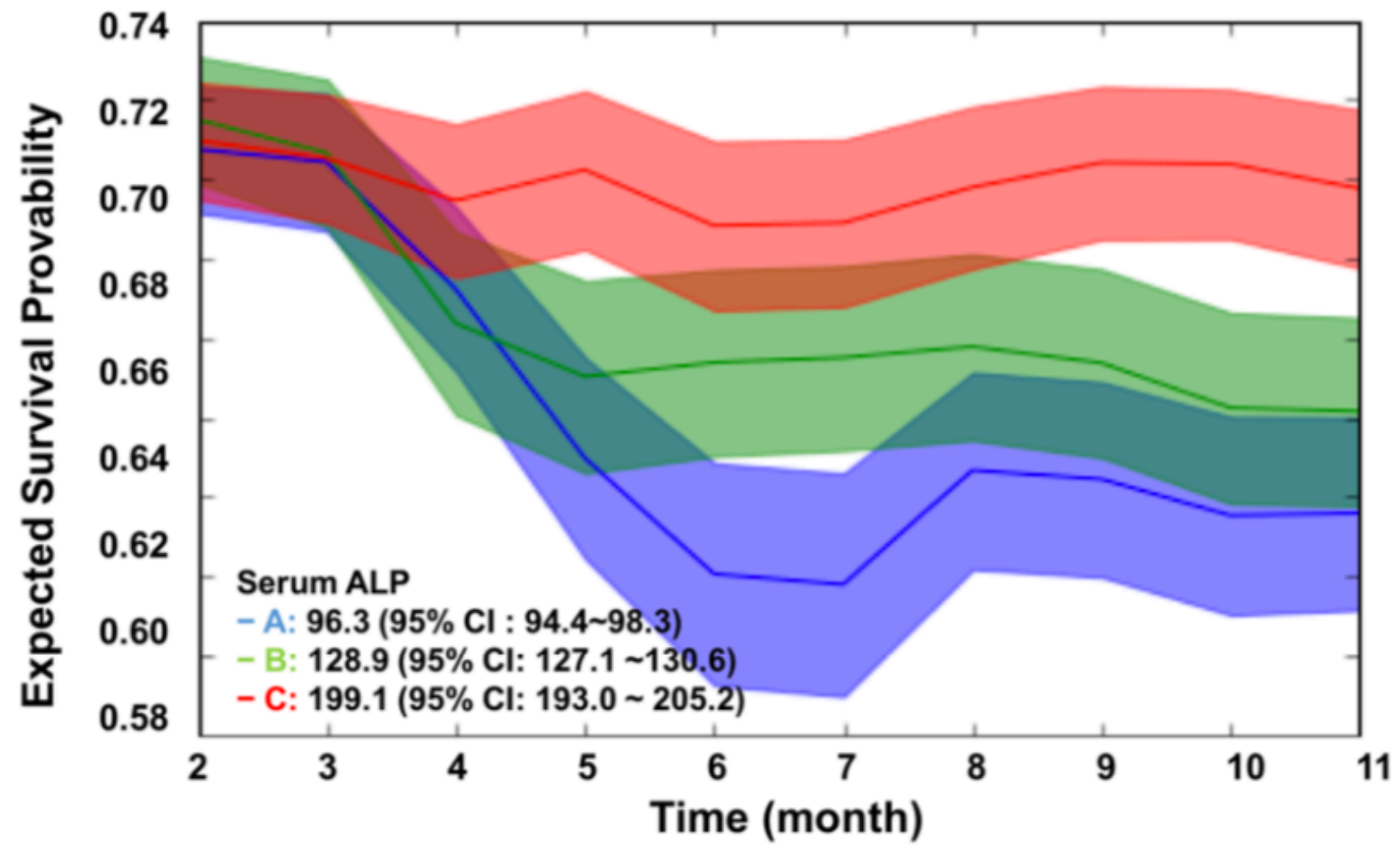
Causal Confusion in Imitation Learning, Pim de Haan et al., NIPS 2019

RL for Medication Control

- Optimal medication control considering cause-effect
  - Q-Learning guided treatment improves survival rate



(a) Simulation of No Treatment



(b) Simulation of Albumin Supplement

RL based Medication Control for end-stage Lung Cancer Patients; (2020)

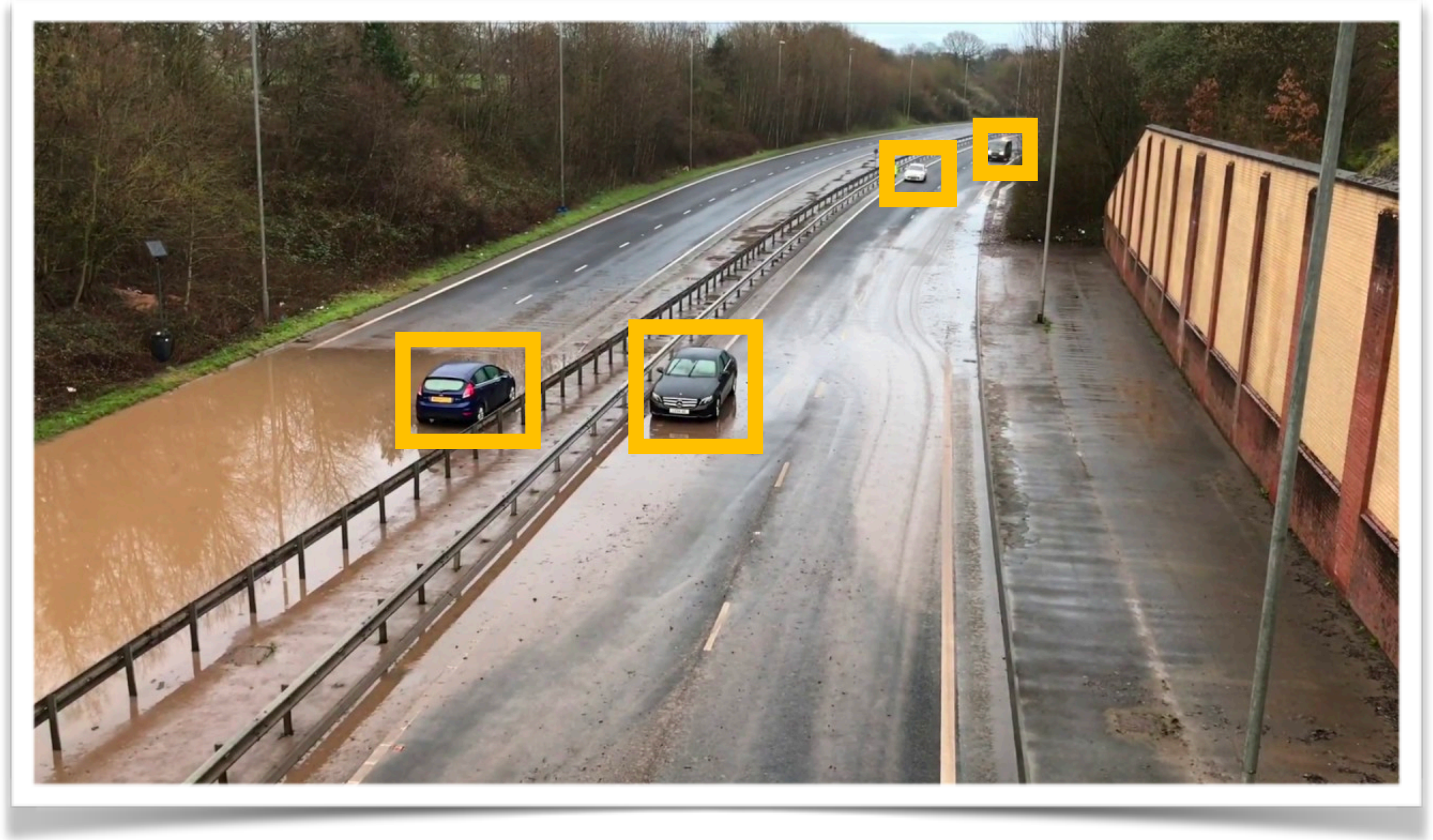
Joint work with J. Kim (SMC), J. Son (SMC), S. Seo (SMC)

Causal Reasoning with Deep Learning

classification (easy)

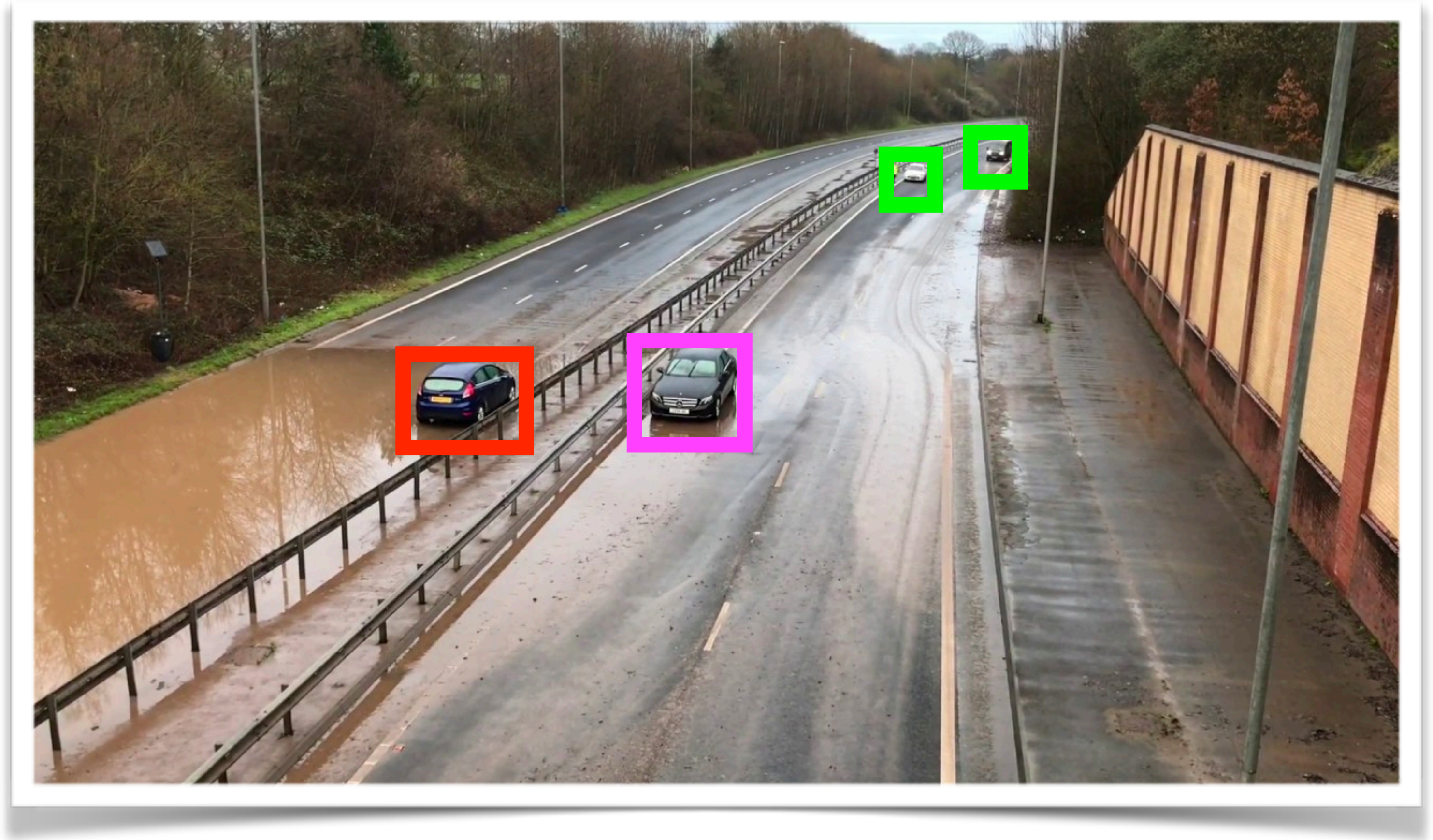
Q: Is there a car in this picture?
A: yes

Detection (moderate)



Q: How many cars in this picture?
A: 4

Uncertainty Estimation (hard)



Q: Is this road safe or danger?
A: danger (why? how?)

Causal Reasoning with Deep Learning

classification (easy)

Q: Is there a car in this pict
A: yes

Causal Reasoning (challenge)

Q: What will happen if more rain comes?

rtainty Estimation (hard)

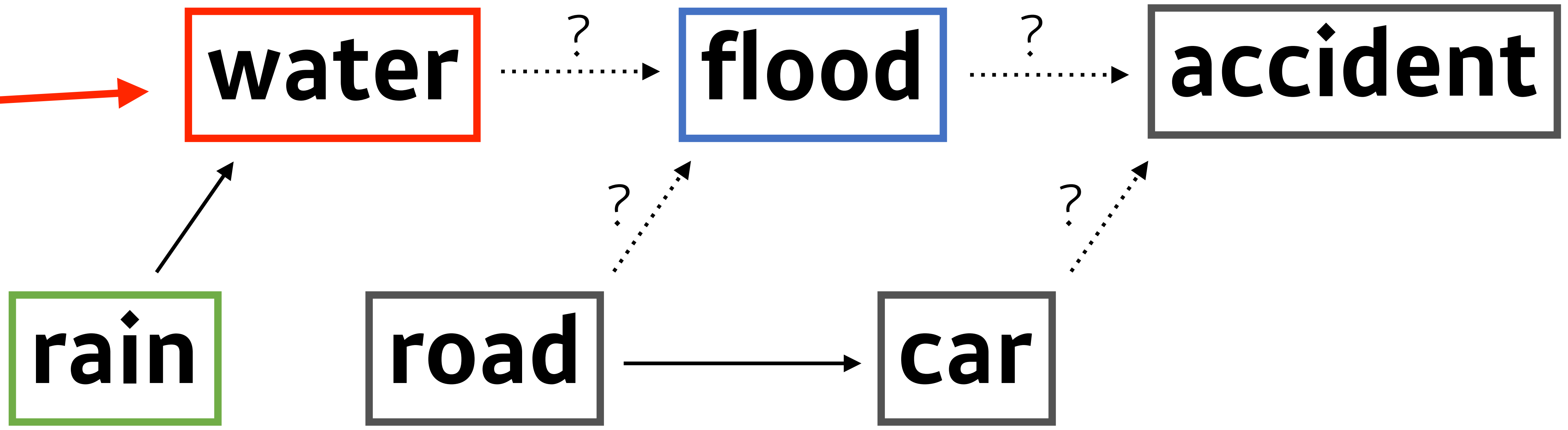
: Is this road safe or danger?
A: danger (why? how?)

Causal Reasoning with Deep Learning

Causal Reasoning (challenge)

Q: What will happen if more rain comes?

Causal Structural Graph



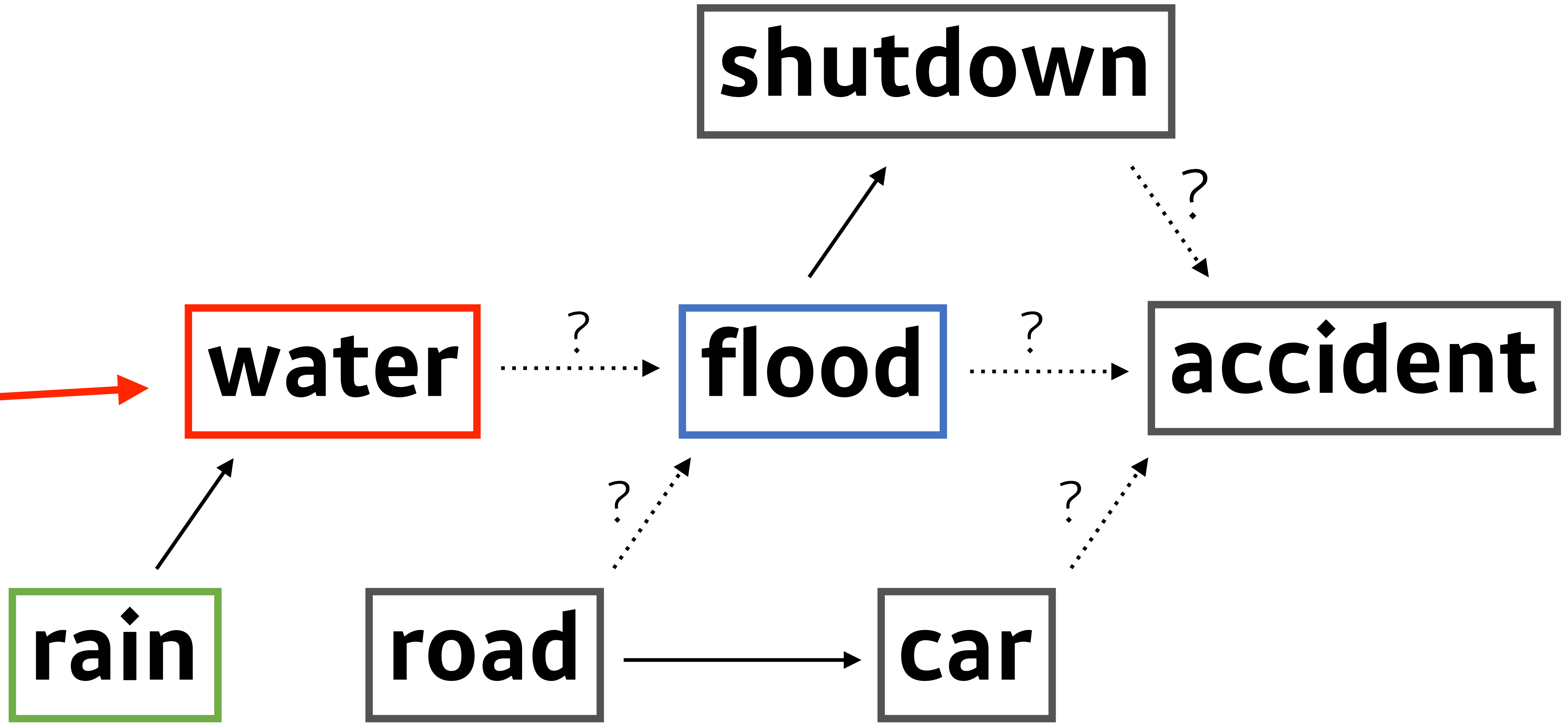
unobservable, deductive, graph-structured

Causal Reasoning with Deep Learning

Causal Reasoning (challenge)

Q: What will happen if more rain comes?

Causal Structural Graph



unobservable, deductive, graph-structured

Conclusion

- How to **connect** causality to machine learning?
  - stable learning under distribution change
  - meta learning for causal discovery
  - bias correction in reinforcement learning
  - causal reasoning in VQA
- Open problems
  - causality for explainability?
  - a new methodology for causal discovery

Q & A