Structural Causal Models and Potential Outcome Models

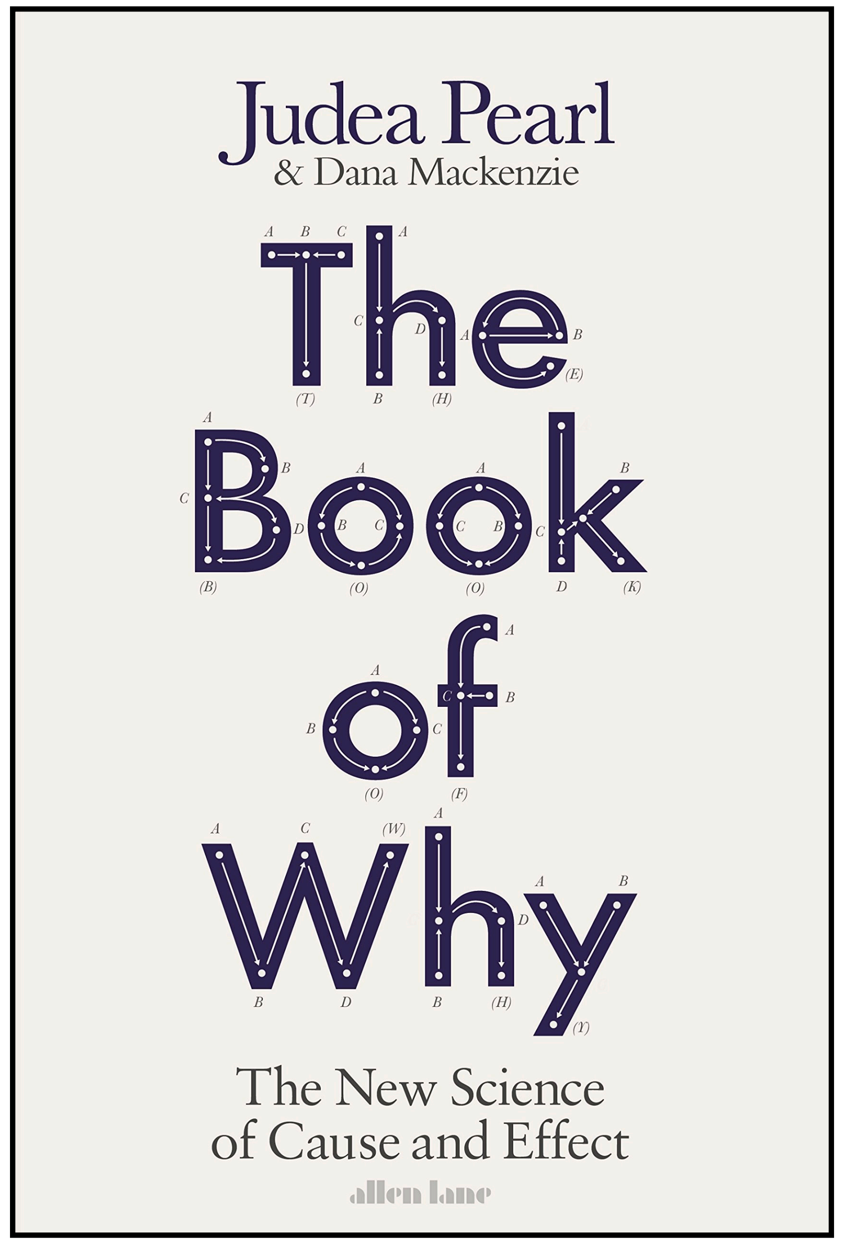
Graphical versus Symbolic Analysis in Causal Inference

Chaochao Lu

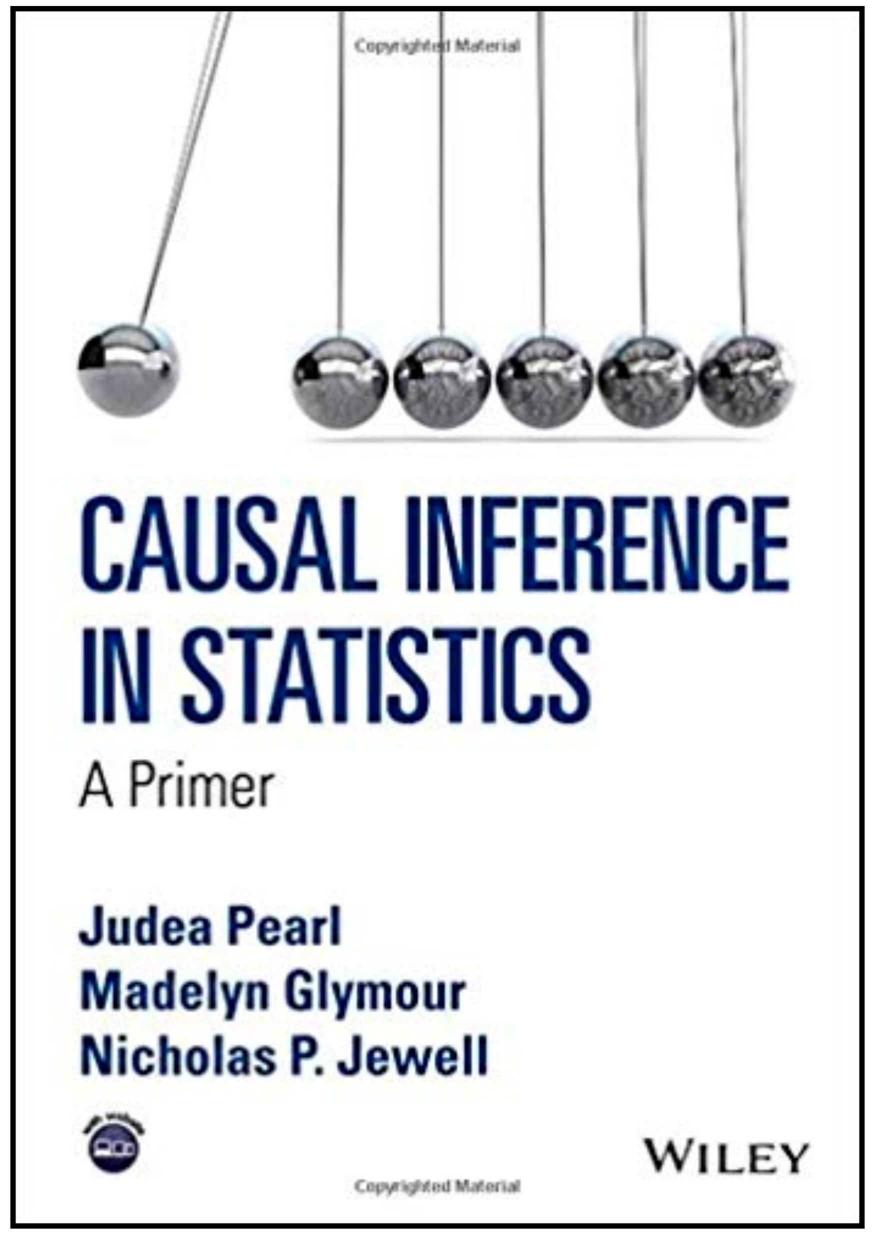
University of Cambridge & Max Planck Institute for Intelligent Systems

21 March 2021

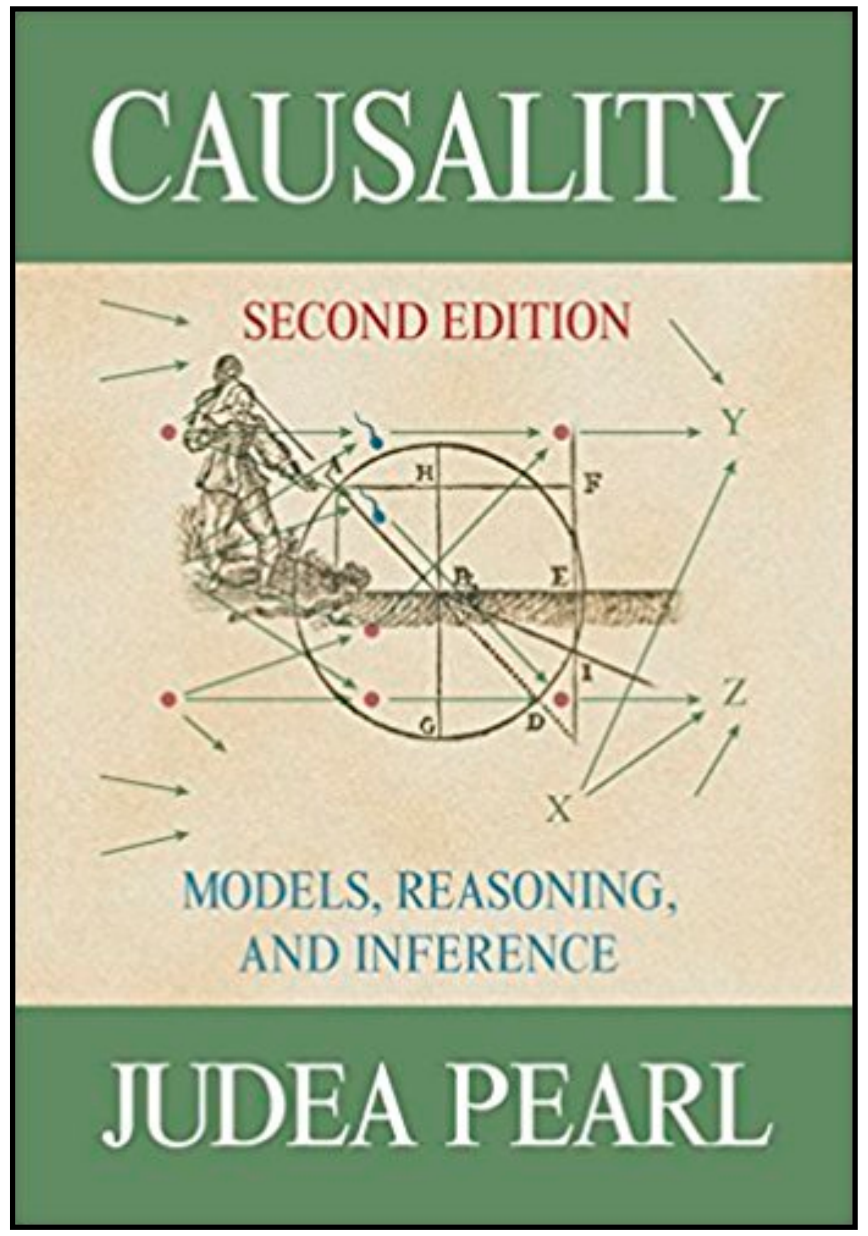
Disclaimer



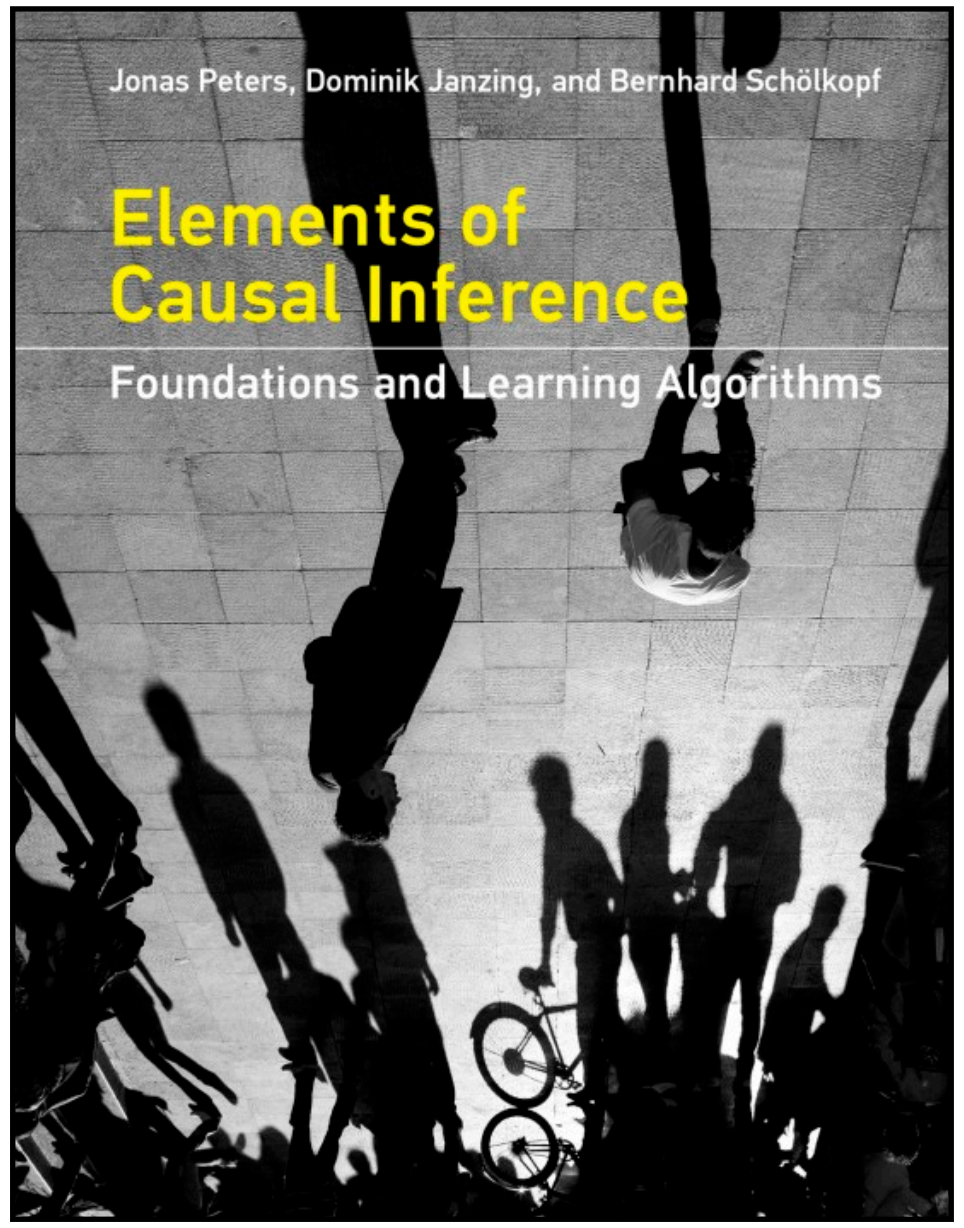
WHY



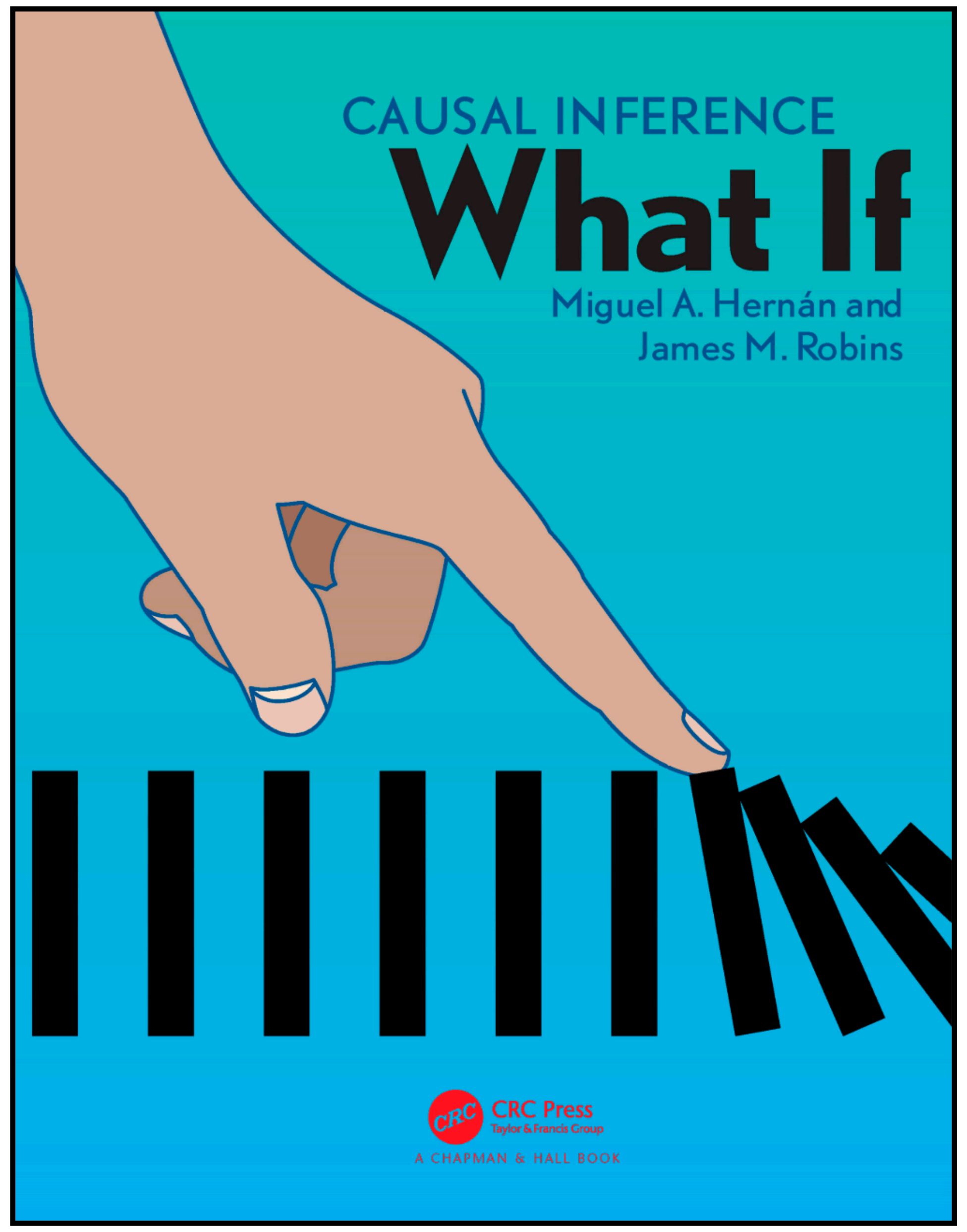
PRIMER



CAUSALITY

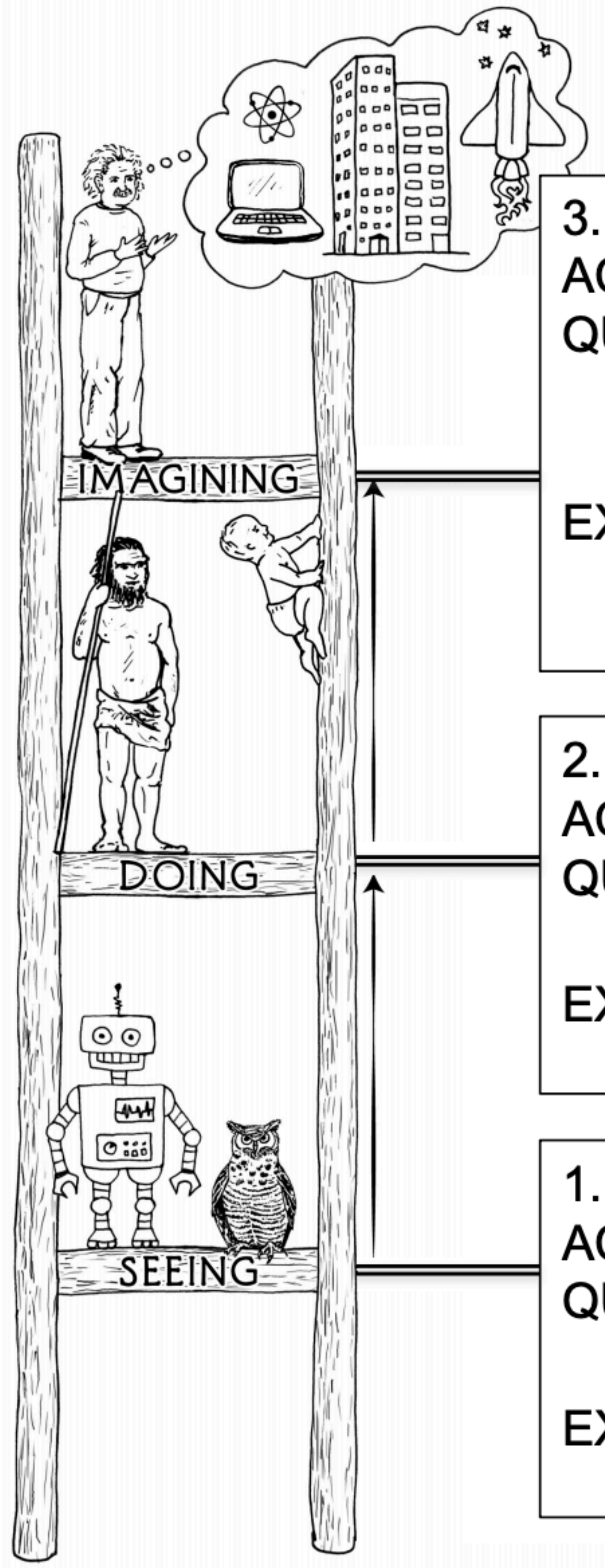


ELEMENTS



WHATIF

The Ladder of Causation



3-LEVEL HIERARCHY

3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done . . . ? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache?
Would Kennedy be alive if Oswald had not killed him? What if I had not smoked the last 2 years?

$$P(Y_{X'}|X)$$

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do . . . ? How?*
(What would Y be if I do X?)

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

$$P(Y|do(X)), P(Y_x), P(Y(x))$$

1. ASSOCIATION

ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see . . . ?*
(How would seeing X change my belief in Y?)

EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

$$P(Y|X)$$

Structural Causal Models

Structural Causal Model

Definition: A structural causal model is a 4-tuple $<V, U, F, P(u)>$, where

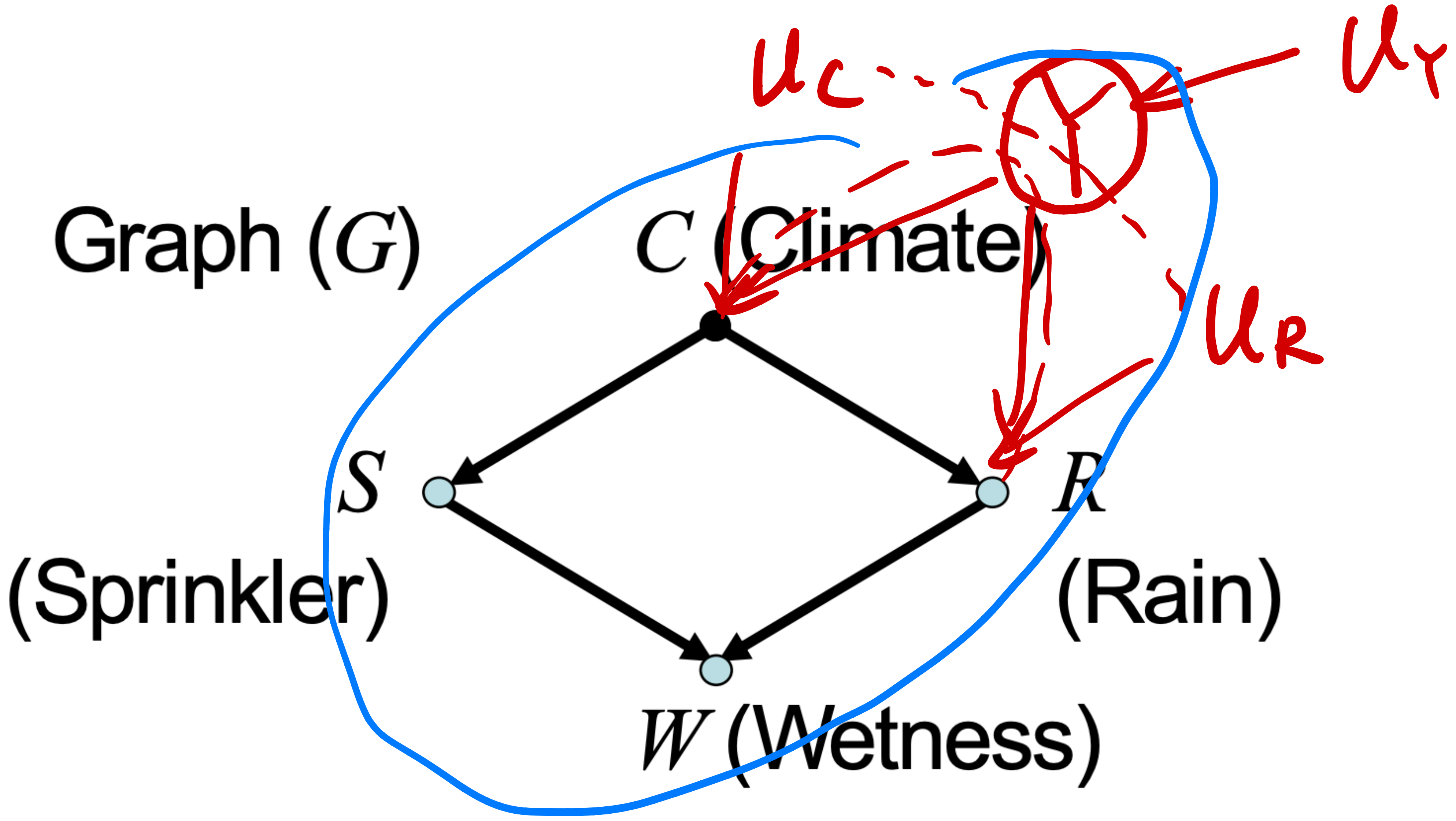
- $V = \{V_1, \ldots, V_n\}$ are endogenous variables
- $U = \{U_1, \ldots, U_m\}$ are background variables
- $F = \{f_1, \ldots, f_n\}$ are functions determining V,
  $v_i = f_i(v, u)$ e.g., $y = \alpha + \beta x + u_Y$
- $P(u)$ is a distribution over U

$U_1 \perp\!\!\!\perp U_2 \perp\!\!\!\perp U_3 \ldots$

$P(u)$ and F induce a distribution $P(v)$ over observable variables

Graphical Representation

Graph (G)



Model (M)

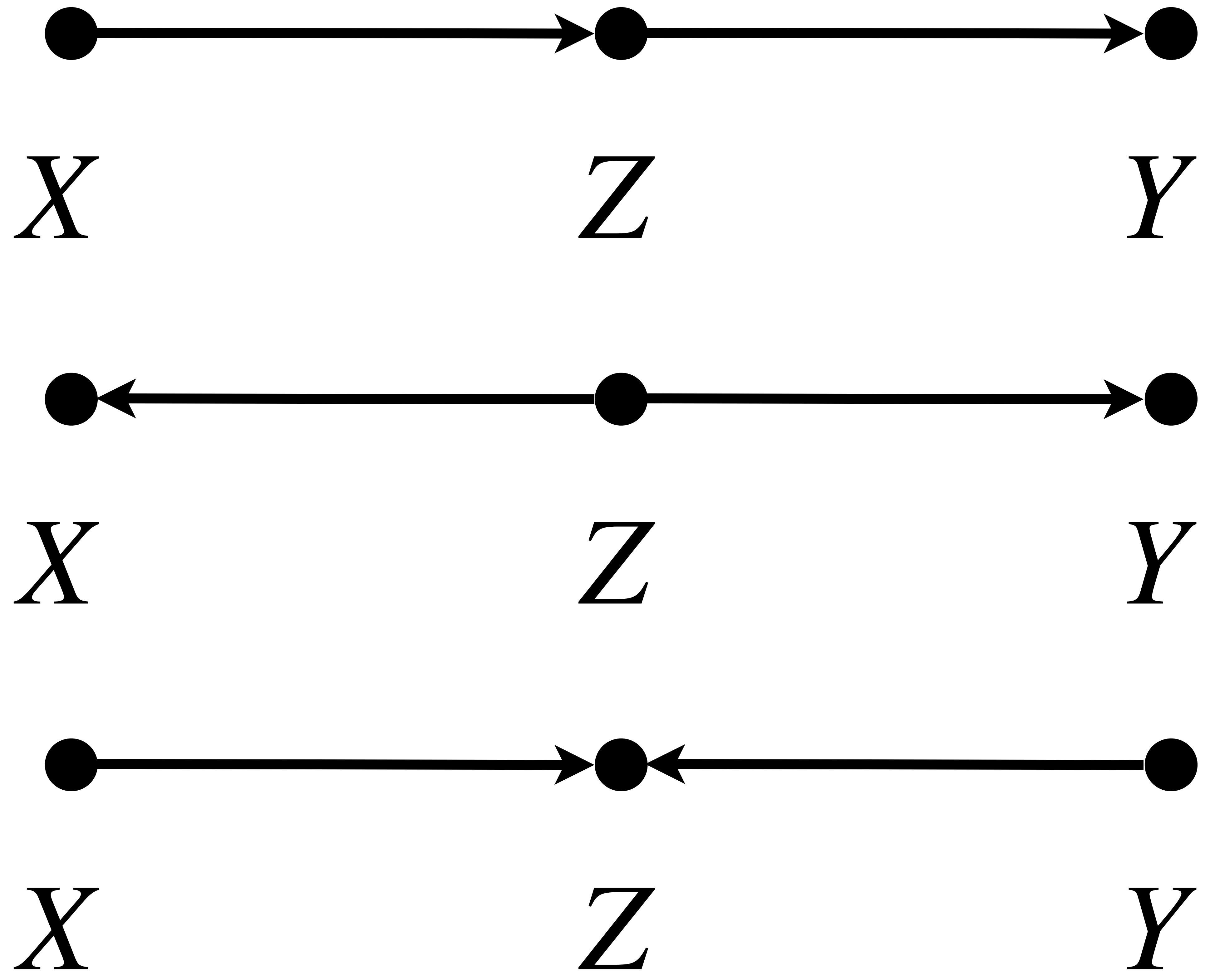
$$C = f_C(U_C)$$

$$S = f_S(C, U_S)$$

$$R = f_R(C, U_R)$$

$$W = f_W(S, R, U_W)$$

Three Building Blocks



$X \not\perp\!\!\!\perp Y$

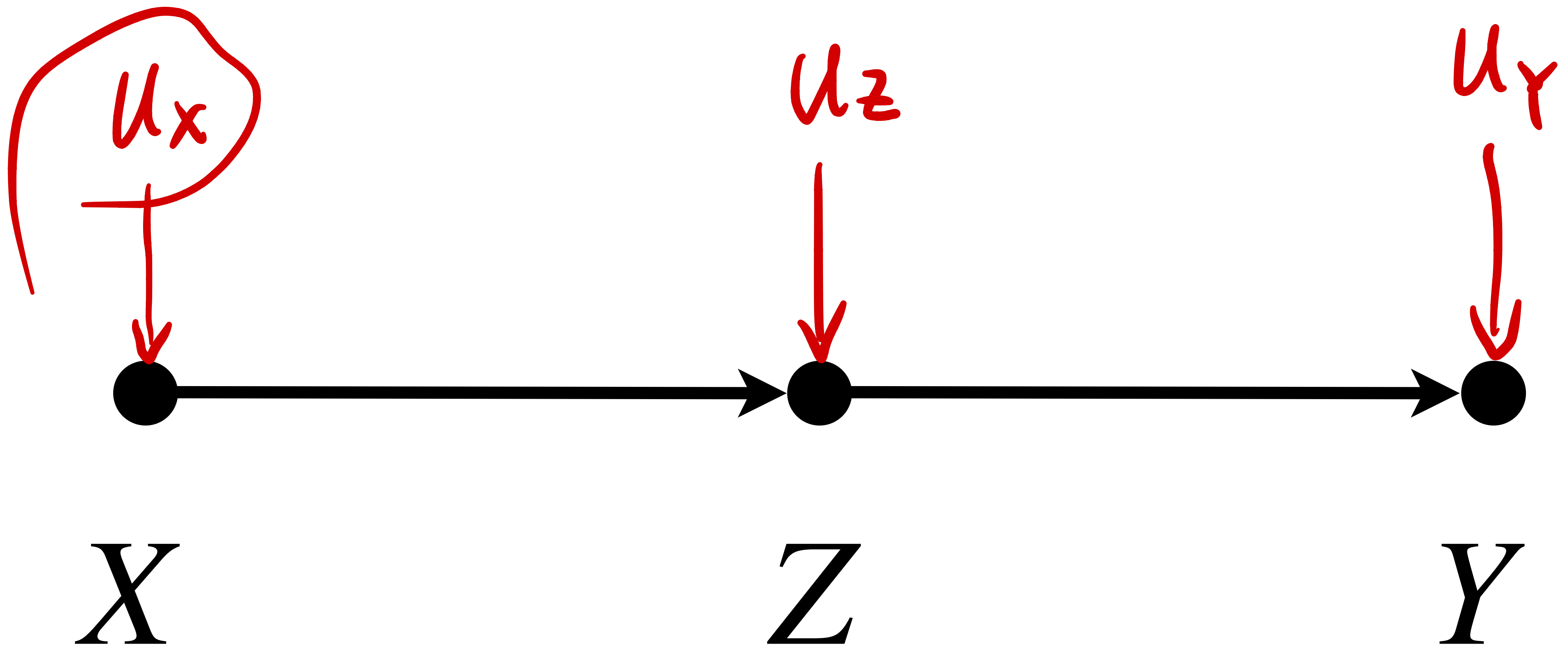
$X \perp\!\!\!\perp Y \mid Z$

$X \perp\!\!\!\perp Y$

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

(PRIMER, CH2)

Chain



$$U_X \perp\!\!\!\perp U_Y \mid Z$$

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

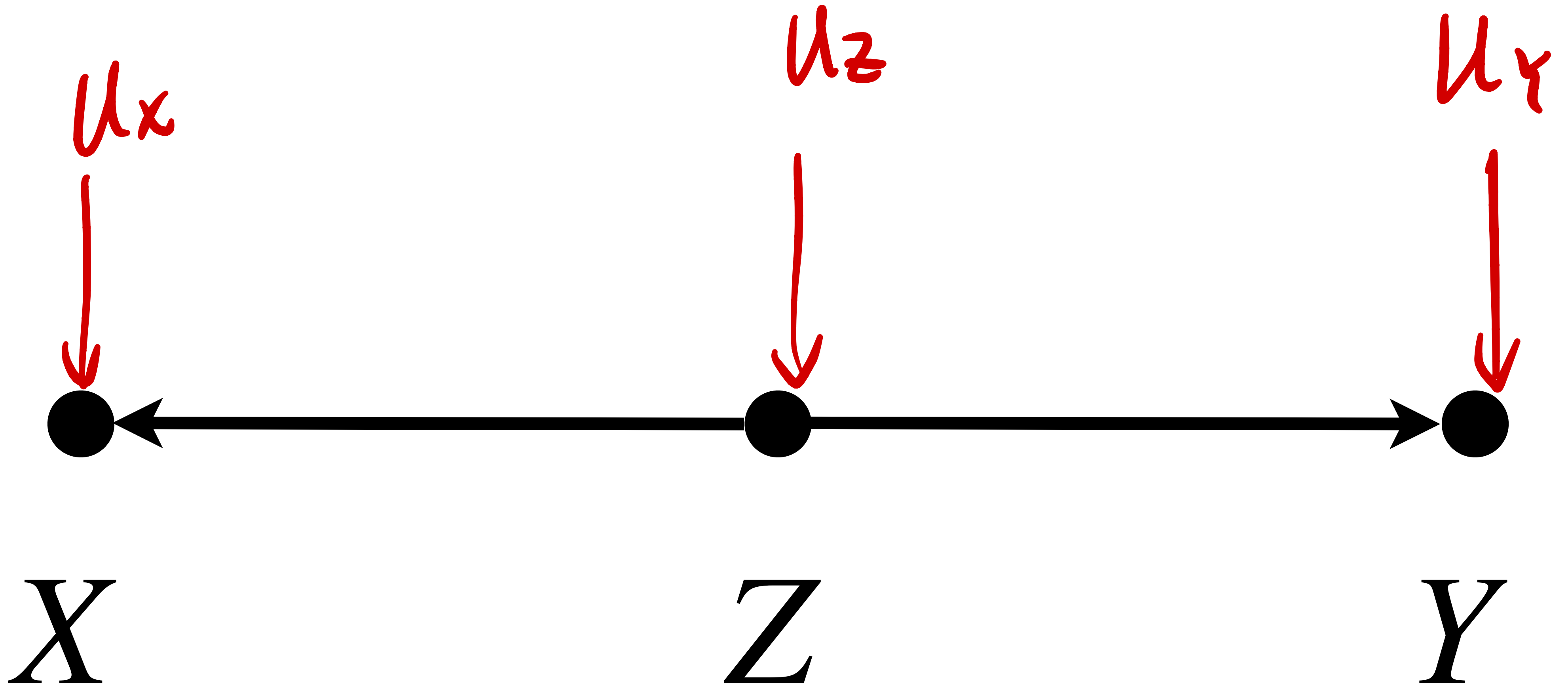
$$\underline{\underline{Y}} = f_Y(Z, \underline{\underline{U_Y}})$$

$$\underline{Z = f_Z(X, U_Z)}$$

$$\underline{\underline{Y}} = g(\underline{\underline{X}}, U_Z, U_Y)$$

$$X \not\perp\!\!\!\perp Y$$

Fork



$$Y = f_Y(Z, U_Y)$$

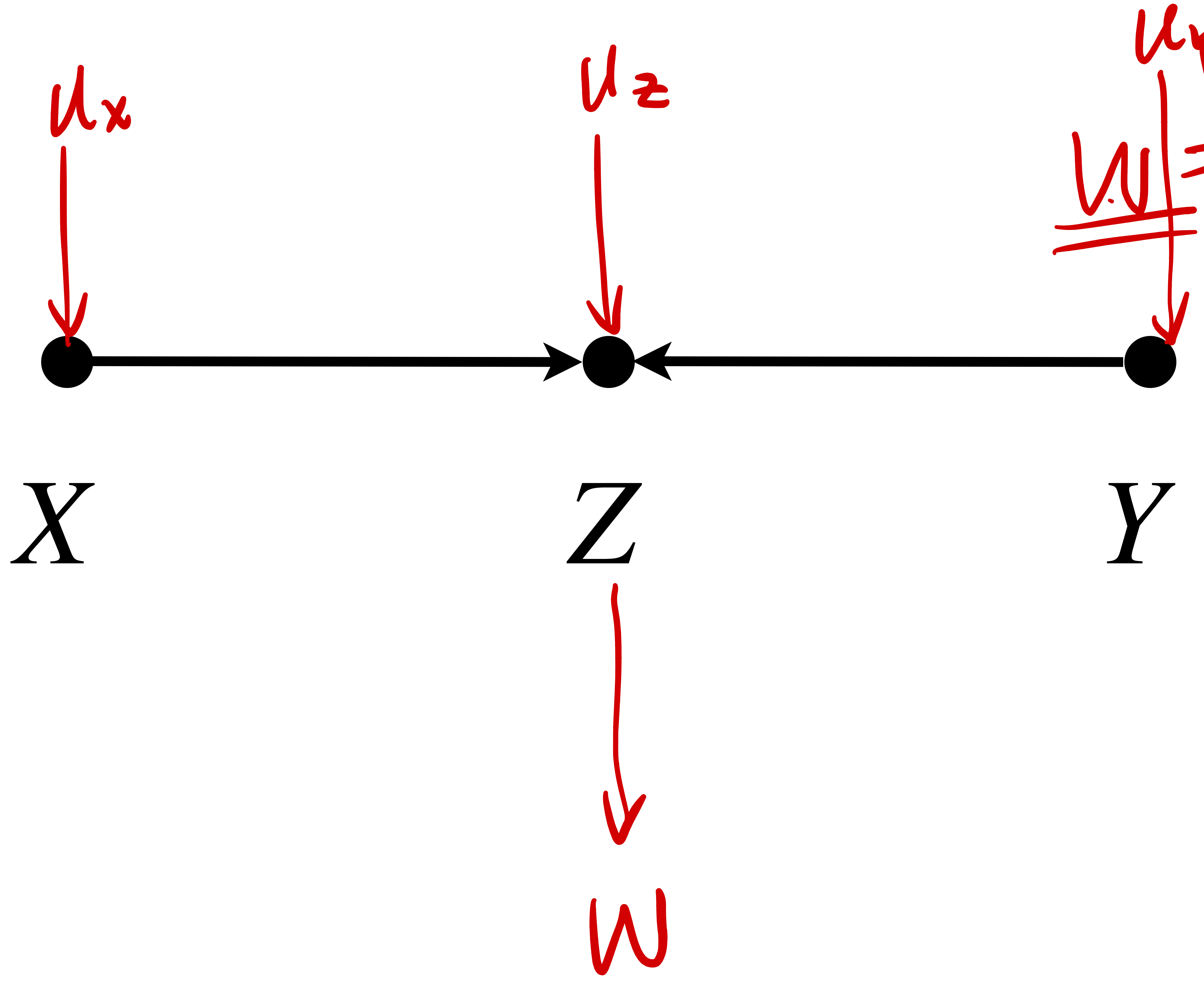
$$X = f_X(Z, U_X)$$

$$Z = f_z^{-1}(X, U_Z)$$

$$X \not\perp Y$$

(PRIMER, CH2)

Collider



$Z \quad U_{W'} \quad U_W$

$$\underline{W} = f(\underline{Z}) = f_Z(\underline{X, Y, U_Z})$$

$$Y = \underline{U_Y}$$

$$X = \underline{U_X} \Rightarrow X \perp\!\!\!\perp Y$$

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

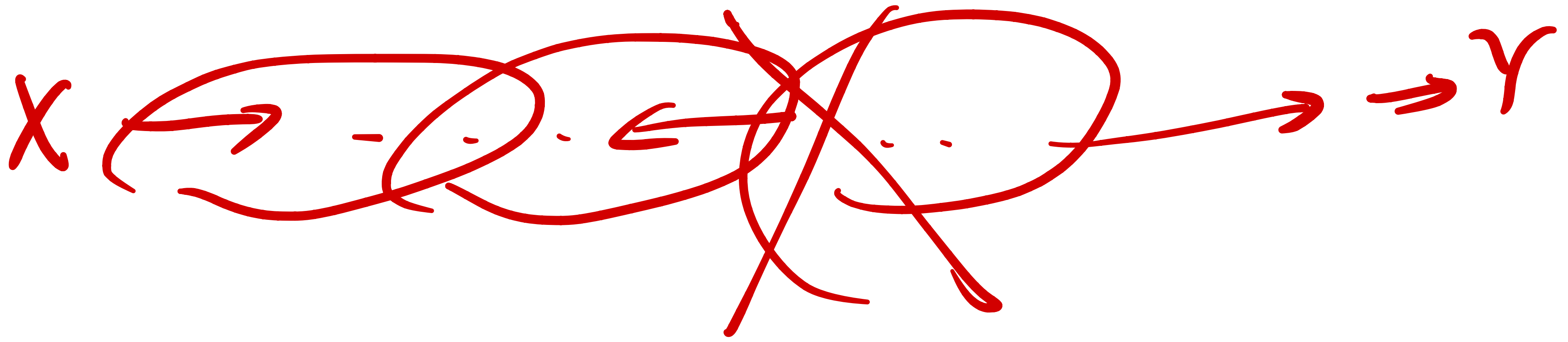
X ''

d-Separation

Definition 2.4.1 (d-separation) *A path p is blocked by a set of nodes Z if and only if*

1. *p contains a chain of nodes $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that the middle node B is in Z (i.e., B is conditioned on), or*
2. *p contains a collider $A \rightarrow B \leftarrow C$ such that the collision node B is not in Z, and no descendant of B is in Z.*

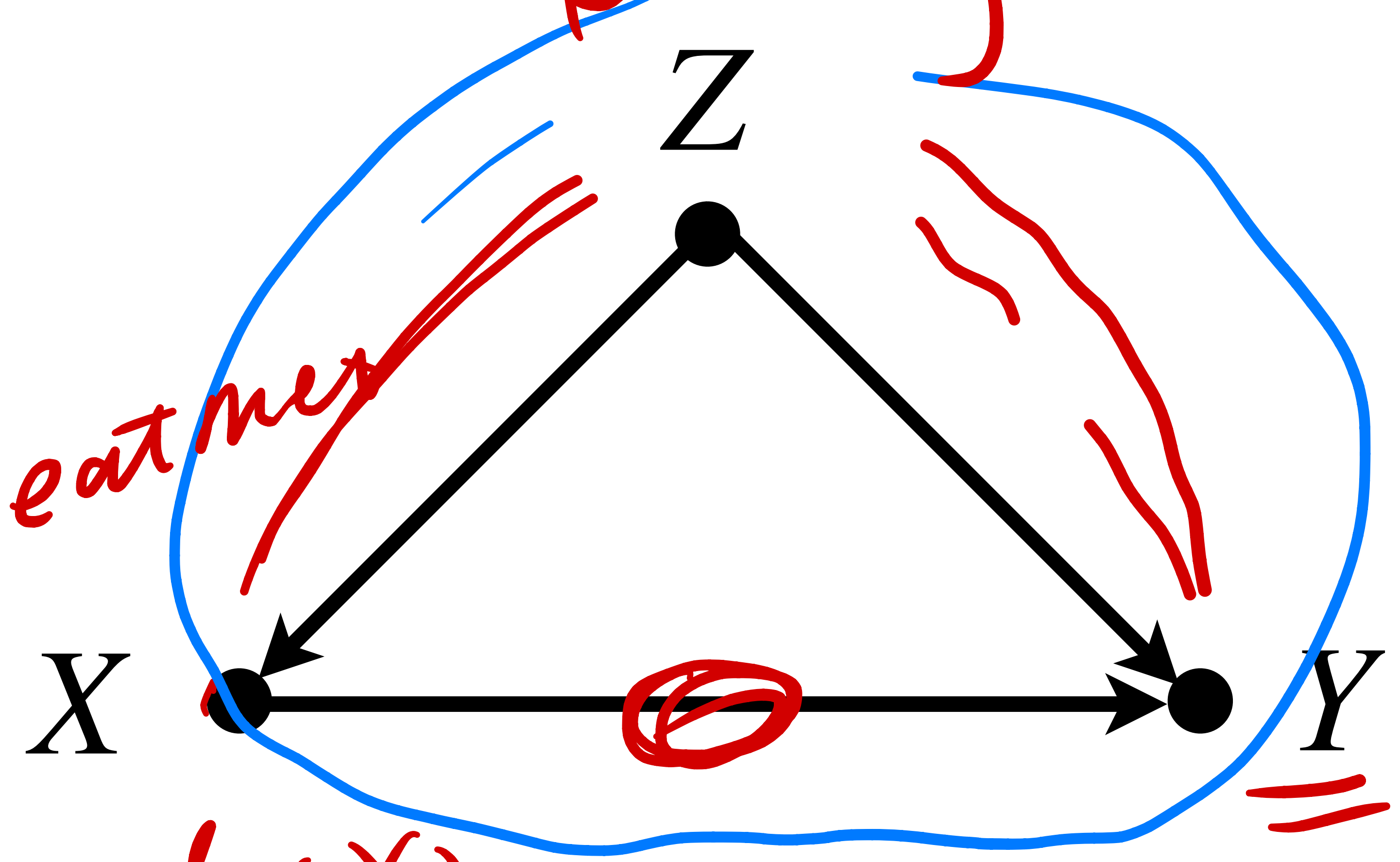
If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.



(PRIMER, CH2)

Back-Door Criterion

Definition 3.3.1 (The Backdoor Criterion) *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the* backdoor criterion *relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*



$$p(Y \mid do(X)) = \sum_z \boxed{p(Y \mid X, z)}\, p(z)$$

"Sure Thing" Principle

$$p(Y \mid do(X)) = \sum_z p(Y \mid do(X), z) \cdot p(z \mid do(X))$$

$$= \sum_z p(Y \mid X, z)\, p(z)$$

$X \longrightarrow Y$

Y

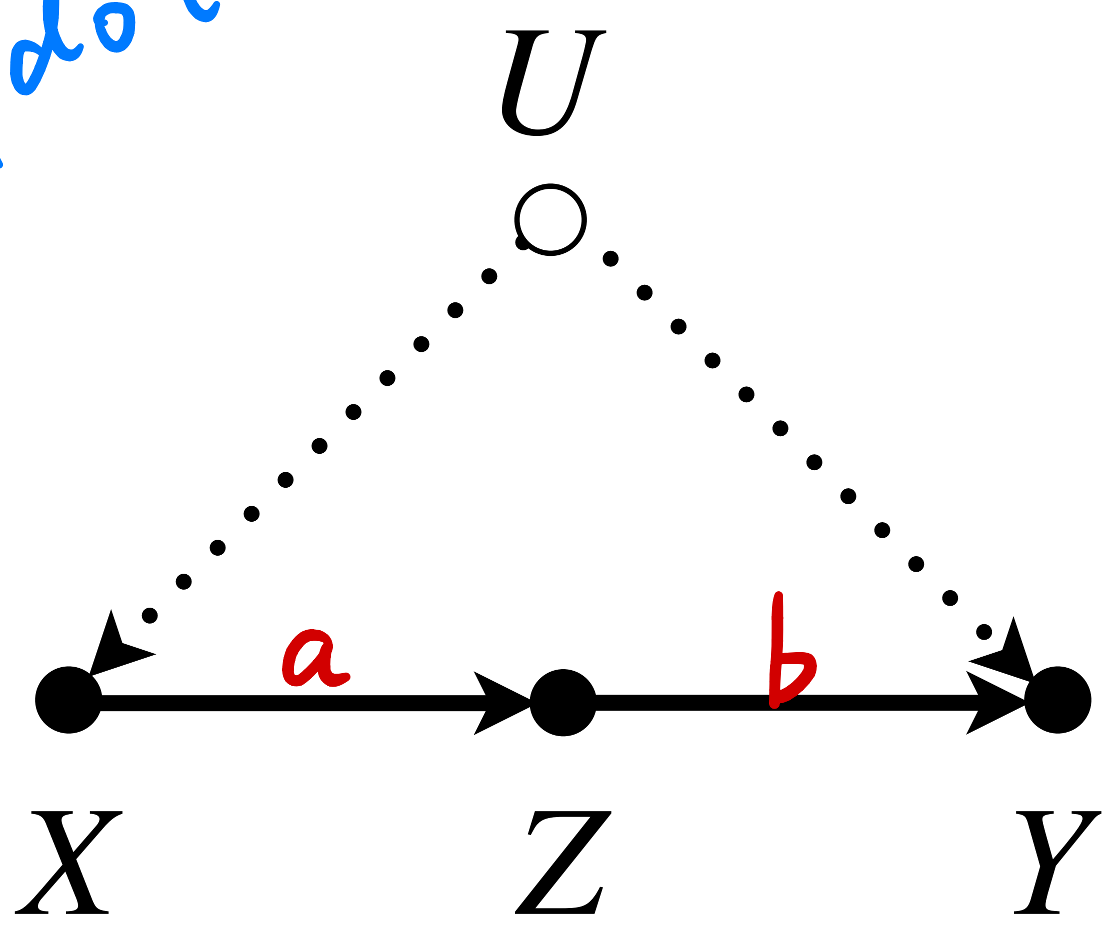
Age

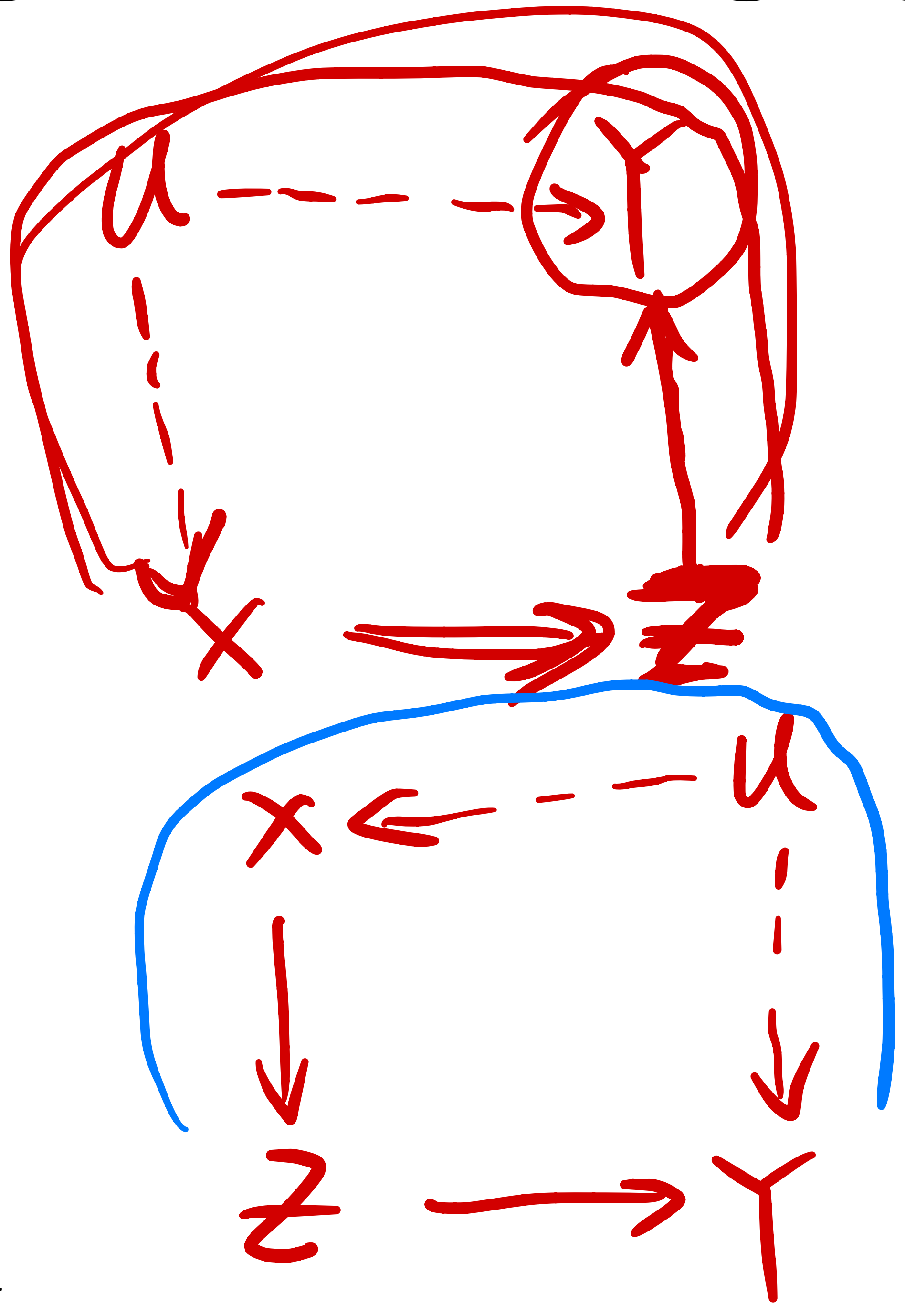
X

(PRIMER, CH3)

Front-Door Criterion

$E(Y|do(x)) = cd + ab$





path-coefficient

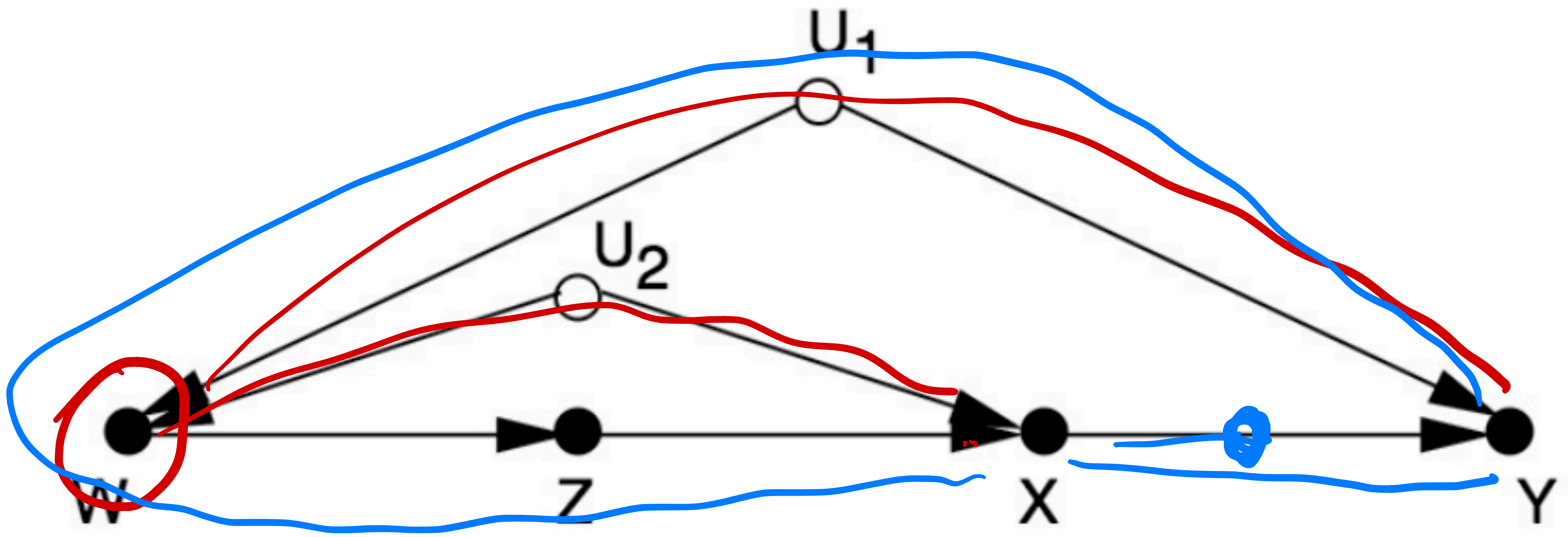
$p(z|do(x)) = p(z|x)$

$p(Y|do(z)) = \underbrace{\sum_x p(Y|z,x)\,p(x)}_{\text{back-door}}$

$E(Y|do(X)) = ab$

$= p(z|do(x)) \cdot p(Y|do(z)) = \sum_z p(z|x) \sum_{x'} p(y|x',z)\,p(x')$

Front-Door or Back-Door?



$$P(Y \mid do(X))?$$

(WHY, CH7)

do-Calculus

Rule 1 (*Insertion/deletion of observations*):

$$P(y \mid \hat{x}, z, w) = P(y \mid \hat{x}, w) \quad \textit{if } (Y \perp\!\!\!\perp Z) \big| X, W)_{G_{\overline{X}}}.$$

Rule 2 (*Action/observation exchange*):

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, z, w) \quad \textit{if } (Y \perp\!\!\!\perp Z) \big| X, W)_{G_{\overline{X}\underline{Z}}}.$$

Rule 3 (*Insertion/deletion of actions*):

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \textit{ if } (Y \perp\!\!\!\perp Z \big| X, W)_{G_{\overline{X}, \overline{Z(W)}}},$$

where Z(W) is the set of Z-nodes that are not ancestors of any W-node in $G_{\overline{X}}$.

(CAUSALITY, CH3)

Insertion/deletion of observations

$\hat{x} = do(x)$

Rule 1 (*Insertion/deletion of observations*):

$$P(y \mid \cancel{\hat{x}}, z, w) = P(y \mid \cancel{\hat{x}}, w) \quad \text{if } \underline{(Y \perp\!\!\!\perp Z)} \mid X, W)_{G_{\overline{X}}}.$$

$\underline{p(y \mid z, w) = p(y \mid w)}$

Action/observation exchange

back-door

Rule 2 (*Action/observation exchange*):

$$P(y \mid \cancel{\hat{x}}, \hat{z}, w) = P(y \mid \cancel{\hat{x}}, z, w) \quad \text{if } (Y \perp\!\!\!\perp Z) \mid X, W)_{G_{\overline{X}\underline{Z}}}.$$

$$P(y \mid \hat{z}) = \sum_w P(y \mid z, w)\, P(w)$$

Insertion/deletion of actions

Rule 3 (*Insertion/deletion of actions*):

$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}},$$

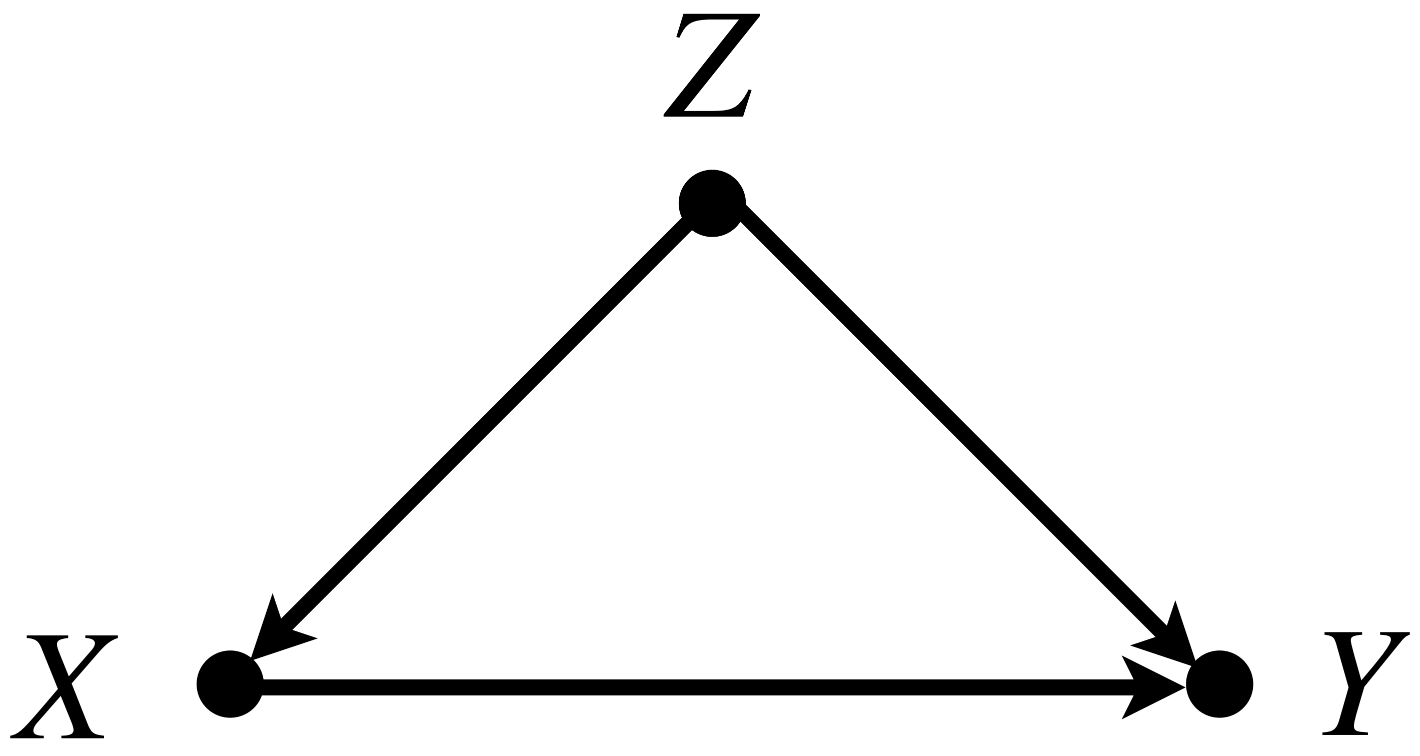
where $Z(W)$ *is the set of* Z*-nodes that are not ancestors of any* W*-node in* $G_{\overline{X}}$.

$$x \longrightarrow \longrightarrow \underline{x} \longrightarrow \longrightarrow Y$$

$$p(y \mid do(x)) = p(y)$$

(WHY, CH7)

Revisit Back-Door Criterion

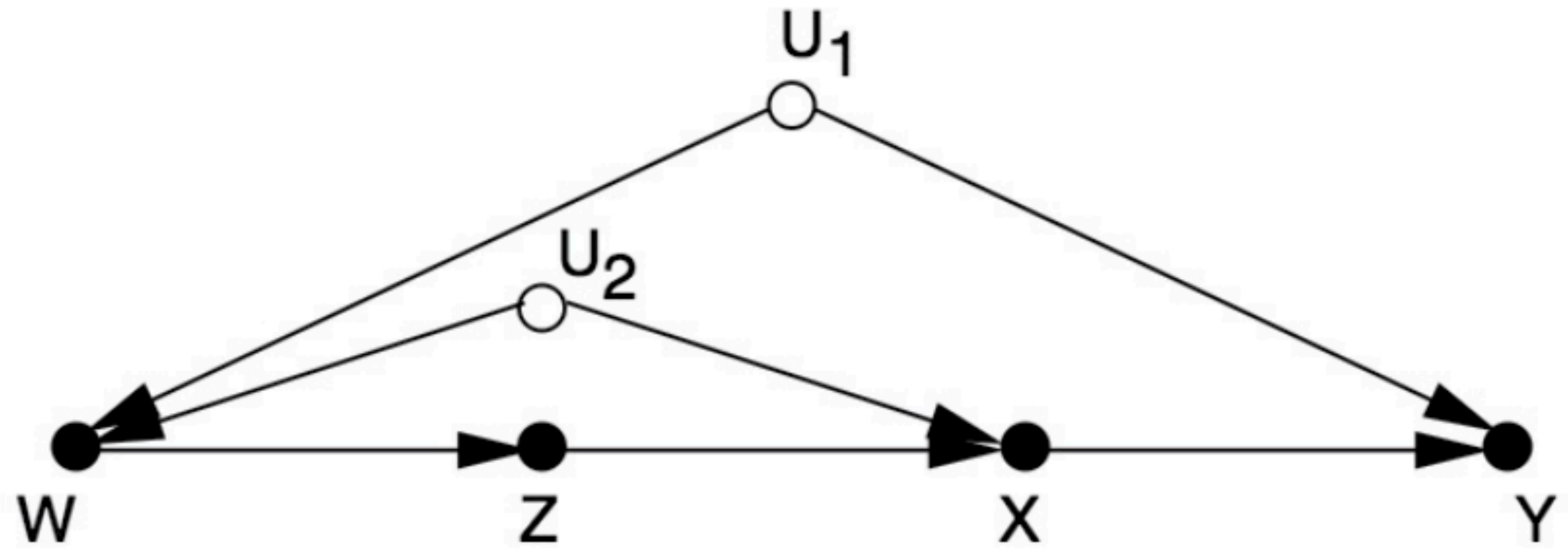


$$p(Y \mid do(X))$$

$$= \sum_z \underbrace{p(Y \mid do(X), z)}_{\Downarrow \text{Rule 2}} \; \underbrace{p(z \mid do(X))}_{\Downarrow \text{Rule 3}}$$

$$= \sum_z p(Y \mid x, z)\, p(z)$$

Let's *do*-Calculus!



$$P(Y \mid do(X))\,!$$

(WHY, CH7)

Counterfactuals

These three steps can be generalized to any causal model M as follows. Given evidence e, to compute the probability of $Y = y$ under the hypothetical condition $X = x$ (where X is a subset of variables), apply the following three steps to M.

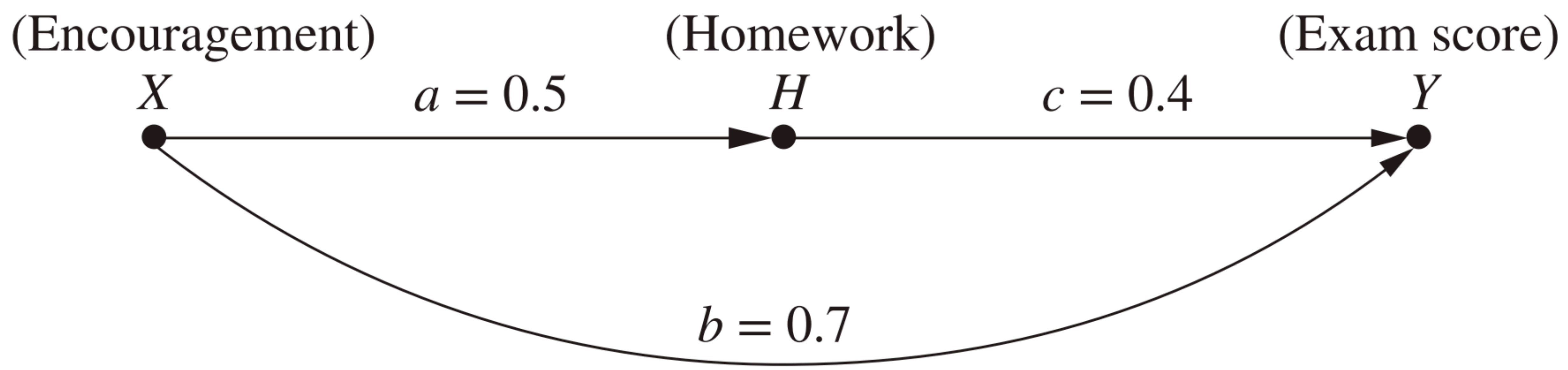
Step 1 **(*abduction*):** Update the probability $P(u)$ to obtain $P(u \mid e)$.

Step 2 **(*action*):** Replace the equations corresponding to variables in set X by the equations $X = x$.

Step 3 **(*prediction*):** Use the modified model to compute the probability of $Y = y$.

(CAUSALITY, CH3)

A Toy Example



$$X = U_X$$
$$H = a \cdot X + U_H$$
$$Y = b \cdot X + c \cdot H + U_Y$$

Let us consider a student named Joe, for whom we measure X = 0.5, H = 1, and Y = 1.5. Suppose we wish to answer the following query: What would Joe's score have been had he doubled his study time?

$E[Y_{H=2} \mid X=0.5, H=1, Y=1.5]$

$$U_X = 0.5,$$
$$U_H = 1 - 0.5 \cdot 0.5 = 0.75, \text{ and}$$
$$U_Y = 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75.$$

$$\begin{aligned} Y_{H=2}(U_X = 0.5, U_H = 0.75, U_Y = 0.75) &= 0.5 \cdot 0.7 + 2.0 \cdot 0.4 + 0.75 \\ &= 1.90 \end{aligned}$$

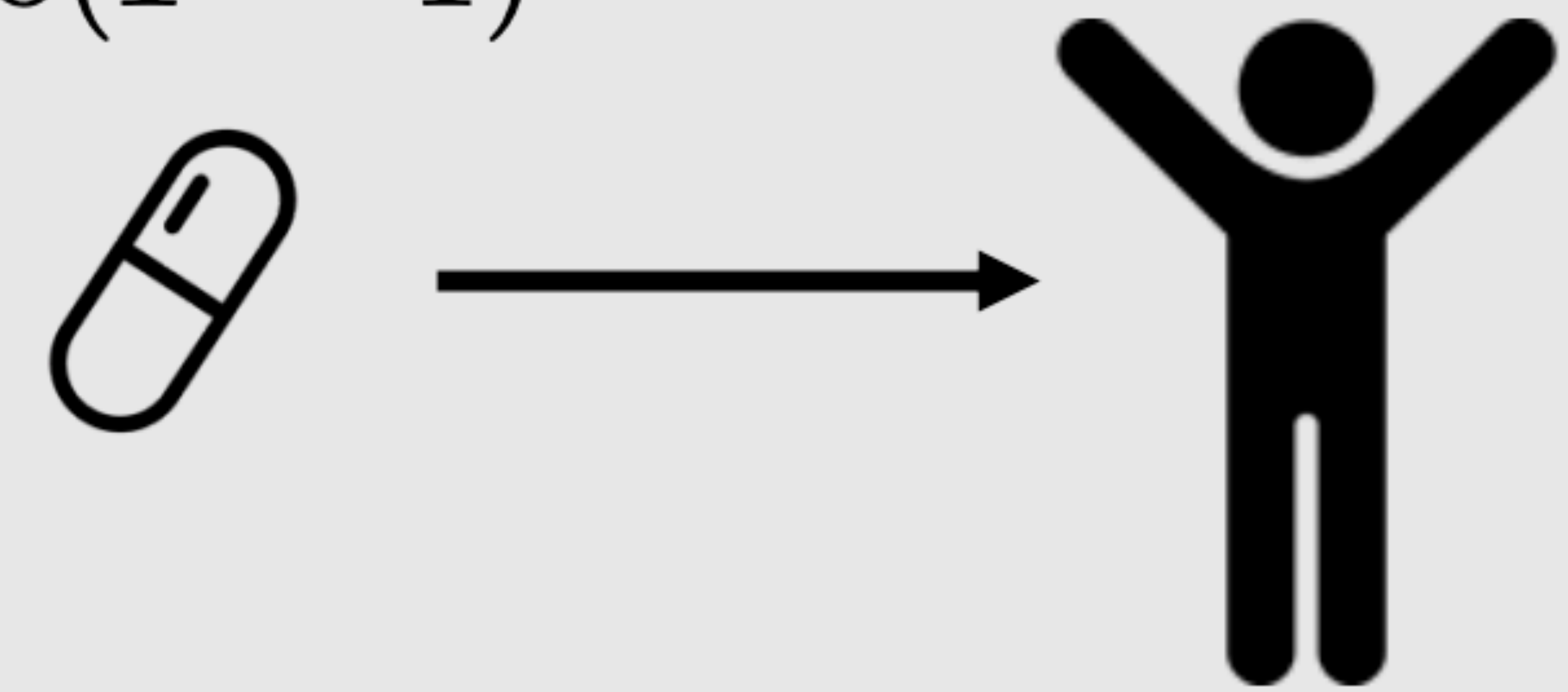
(PRIMER, CH4)

Potential Outcome Models

Potential Outcomes: Notation

$\text{do}(T = 1)$

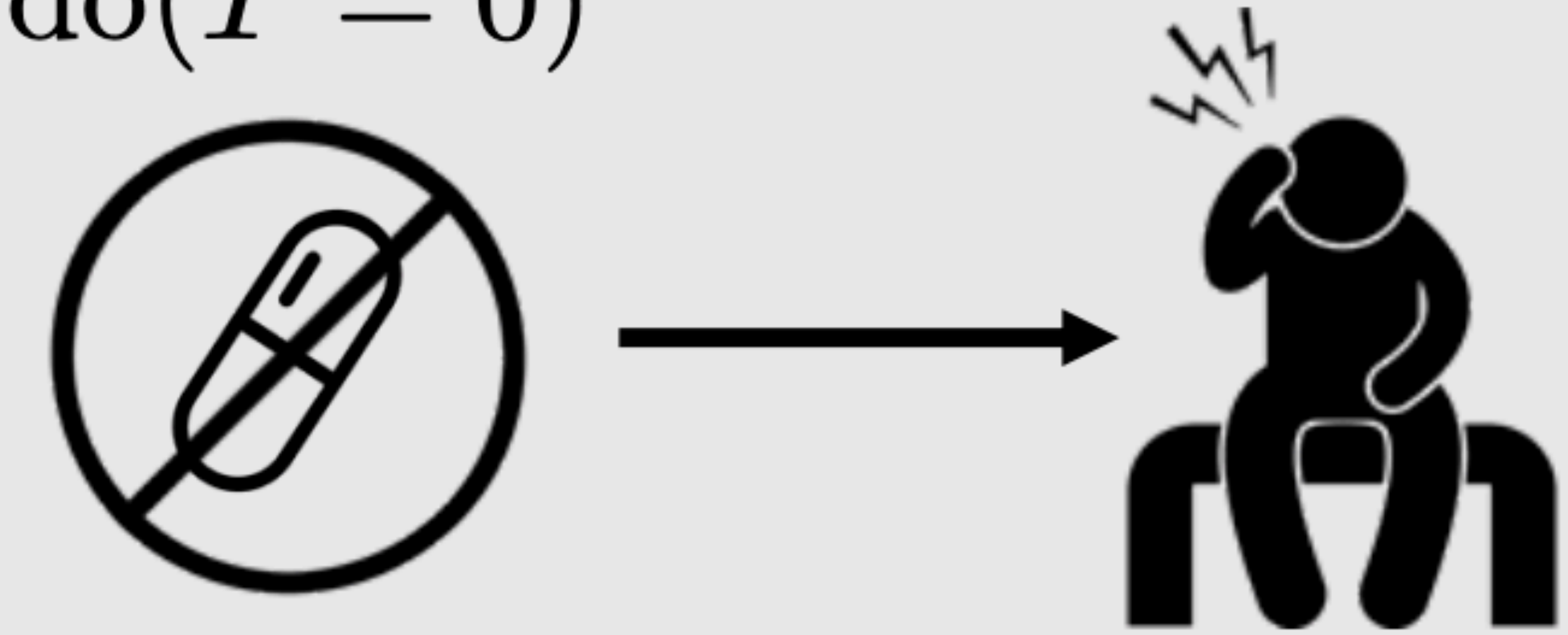
$Y_i|_{\text{do}(T=1)} \triangleq Y_i(1)$



$Y_{T=1}(u)$ u

$\text{do}(T = 0)$

$Y_i|_{\text{do}(T=0)} \triangleq Y_i(0)$

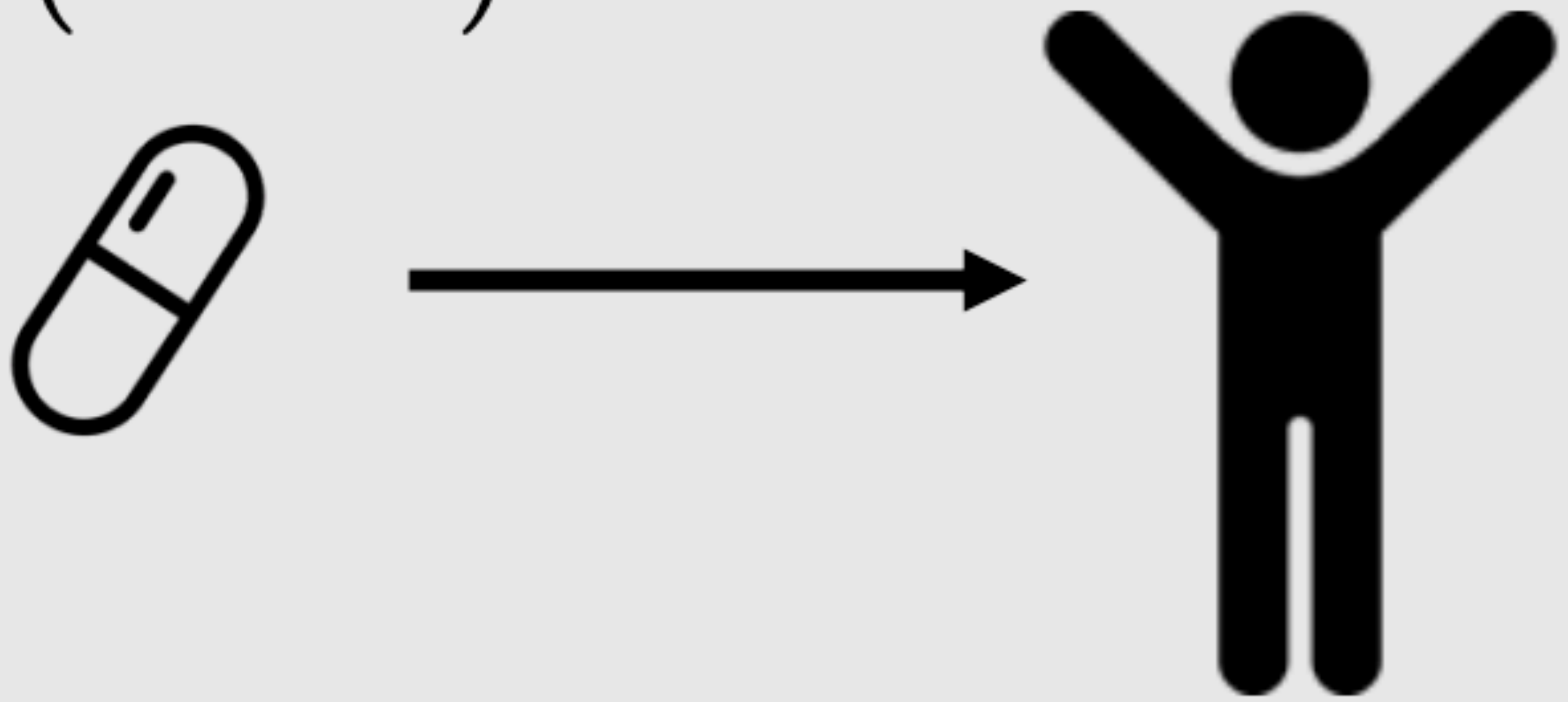


T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

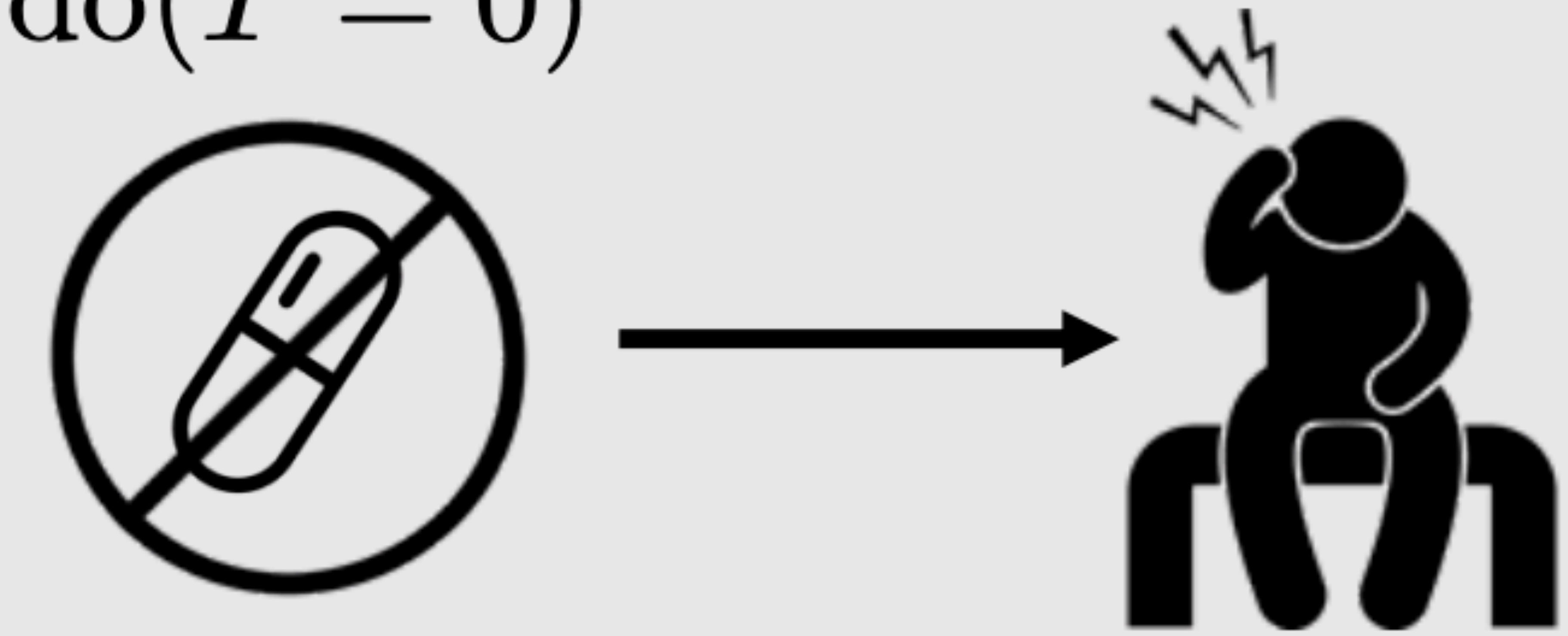
(Brady Neal, CH2)

Potential Outcomes: Notation

$$\mathrm{do}(T = 1) \qquad Y_i(1) = 1$$



$$\mathrm{do}(T = 0) \qquad Y_i(0) = 0$$



T	: observed treatment
Y	: observed outcome
i	: used in subscript to denote a specific unit/individual
$Y_i(1)$	: potential outcome under treatment
$Y_i(0)$	: potential outcome under no treatment

Causal effect

$$Y_i(1) - Y_i(0) = 1$$

(Brady Neal, CH2)

Fundamental Problem

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

T : observed treatment
Y : observed outcome
i : used in subscript to denote a specific unit/individual
$Y_i(1)$: potential outcome under treatment
$Y_i(0)$: potential outcome under no treatment

$$E[Y(1) - Y(0)]$$
$$= E[Y(1)] - E[Y(0)]$$
$$\downarrow \text{Assumptions}$$
$$\overset{?}{=} E[Y \mid T=1] - E[Y \mid T=0]$$

(Brady Neal, CH2)

Average Treatment Effect (ATE)

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

(Brady Neal, CH2)

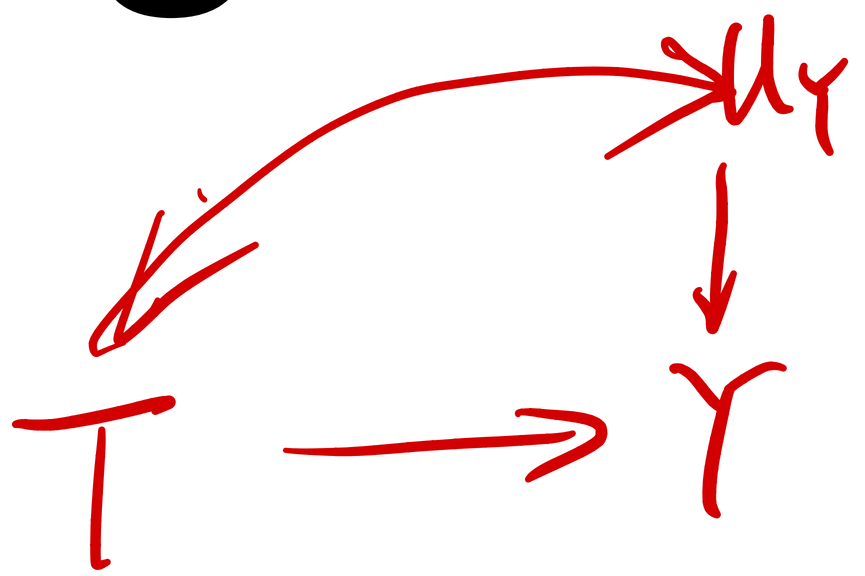
Average Treatment Effect (ATE)

i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \overset{?}{=} \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]$$

(Brady Neal, CH2)

Ignorability/Exchangeability



$$(Y(0), Y(1)) \perp\!\!\!\perp T$$

$f(U_Y)$

$Y(t) \perp\!\!\!\perp T$

$$\begin{aligned}\mathbb{E}[Y(1) - Y(0)] &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &= \mathbb{E}[Y(1) \mid T = 1] - \mathbb{E}[Y(0) \mid T = 0] \\ &= \mathbb{E}[Y \mid T = 1] - \mathbb{E}[Y \mid T = 0]\end{aligned}$$

Consistency: $X = x \Rightarrow Y_x(u) = y$

(Brady Neal, CH2)

Conditional Ignorability/Exchangeability

$$(Y(0), Y(1)) \perp\!\!\!\perp T \mid X$$

back-door

$$\begin{aligned}
\mathbb{E}[Y(1) - Y(0)] &= \mathbb{E}_X \mathbb{E}[Y(1) - Y(0) \mid X] \\
&= \mathbb{E}_X[\mathbb{E}[Y(1) \mid X] - \mathbb{E}[Y(0) \mid X]] \\
&= \mathbb{E}_X[\mathbb{E}[Y(1) \mid T = 1, X] - \mathbb{E}[Y(0) \mid T = \overset{0}{\cancel{1}}, X]] \\
&= \mathbb{E}_X[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]]
\end{aligned}$$

(Brady Neal, CH2)

The Great Power of Graphs

“Logic void of representation is metaphysics.”

–Judea Pearl

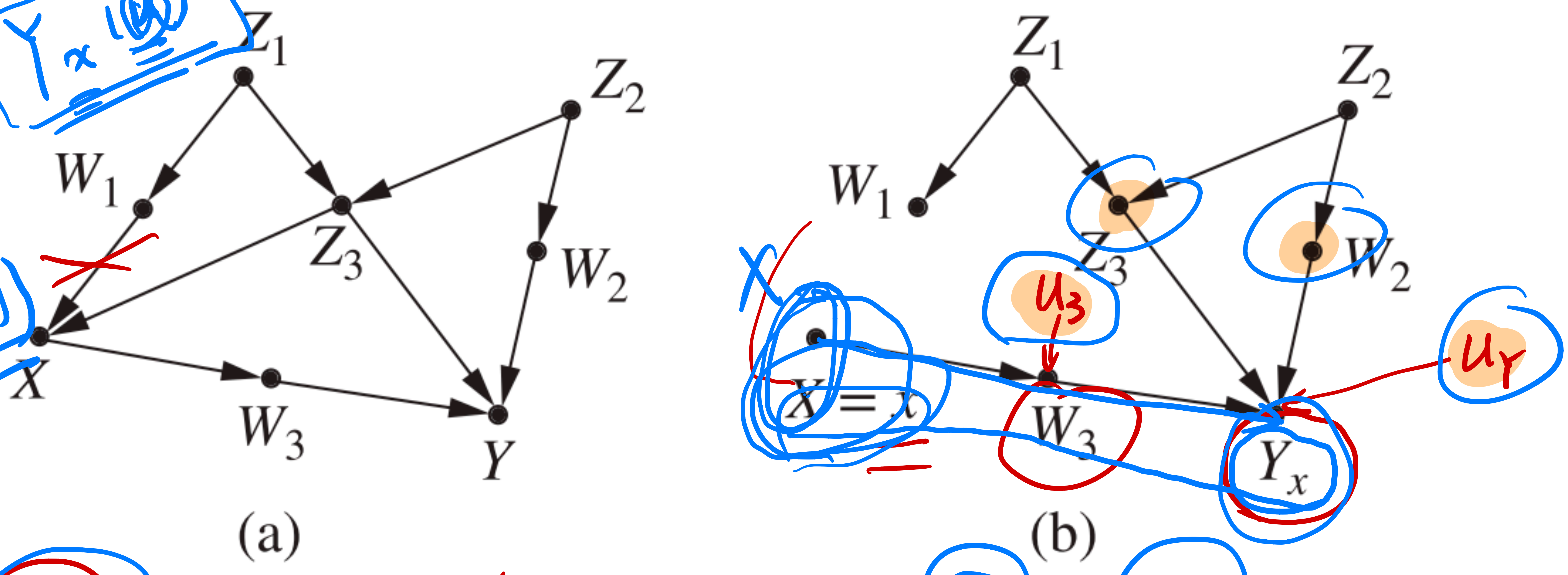
Visualizing Counterfactuals

$Y_i(x) \triangleq Y_u(x)$ value

$\triangleq \boxed{Y_x(u)}$

$Y_x(u)$

$P(Y \mid do(x))$



Y_x

$X \perp\!\!\!\perp Y_x$

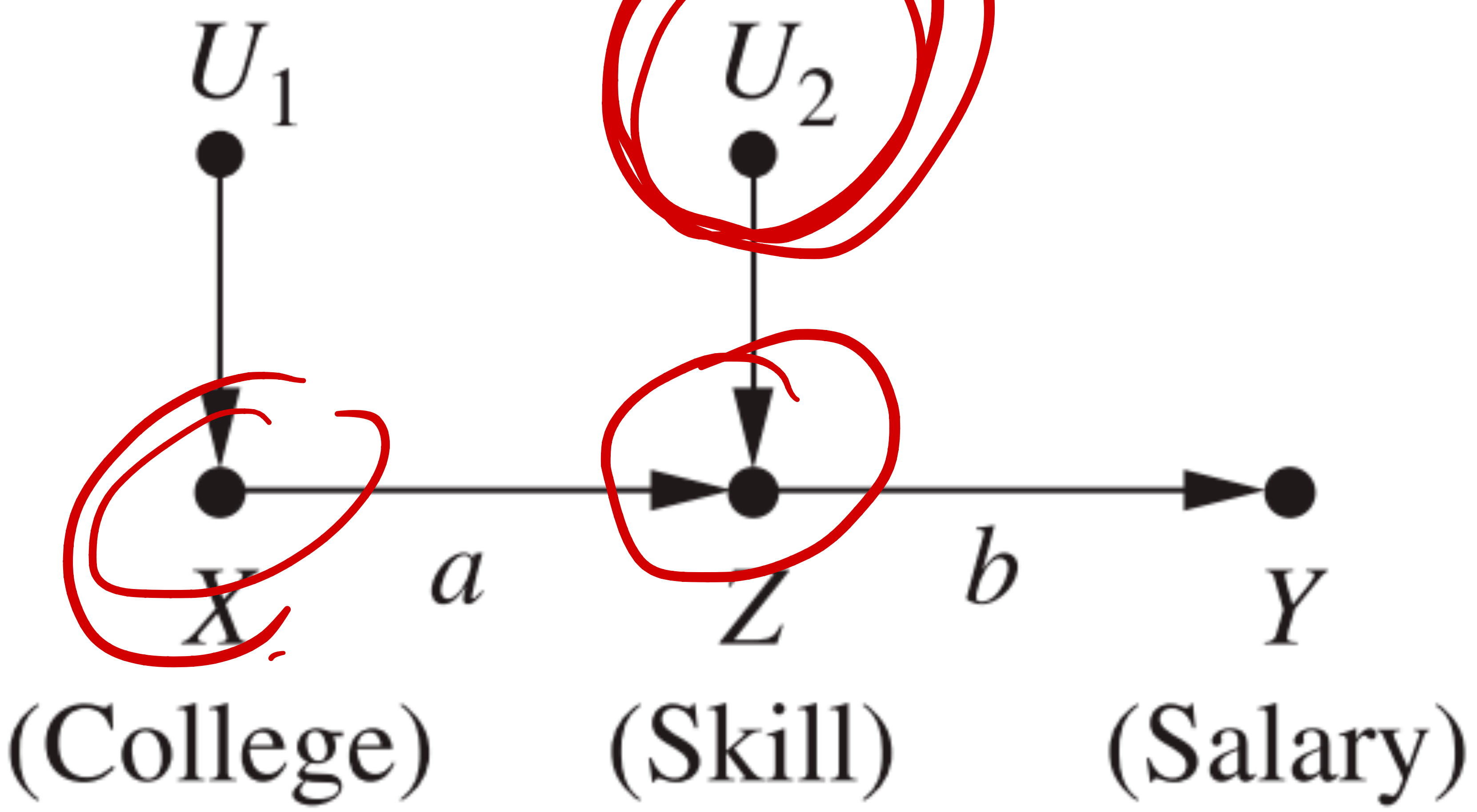
$X \perp\!\!\!\perp Y_x$

$W_3 = f_{W_3}(X = x, U_3)$

$Y_x = f(\cancel{W_3}\ U_3, Z_3, W_2, U_Y)$

(PRIMER, CH4)

Example



$$X \perp\!\!\!\perp Y_{(x)} \mid Z$$

$$Y = f(X, U_2)$$

α

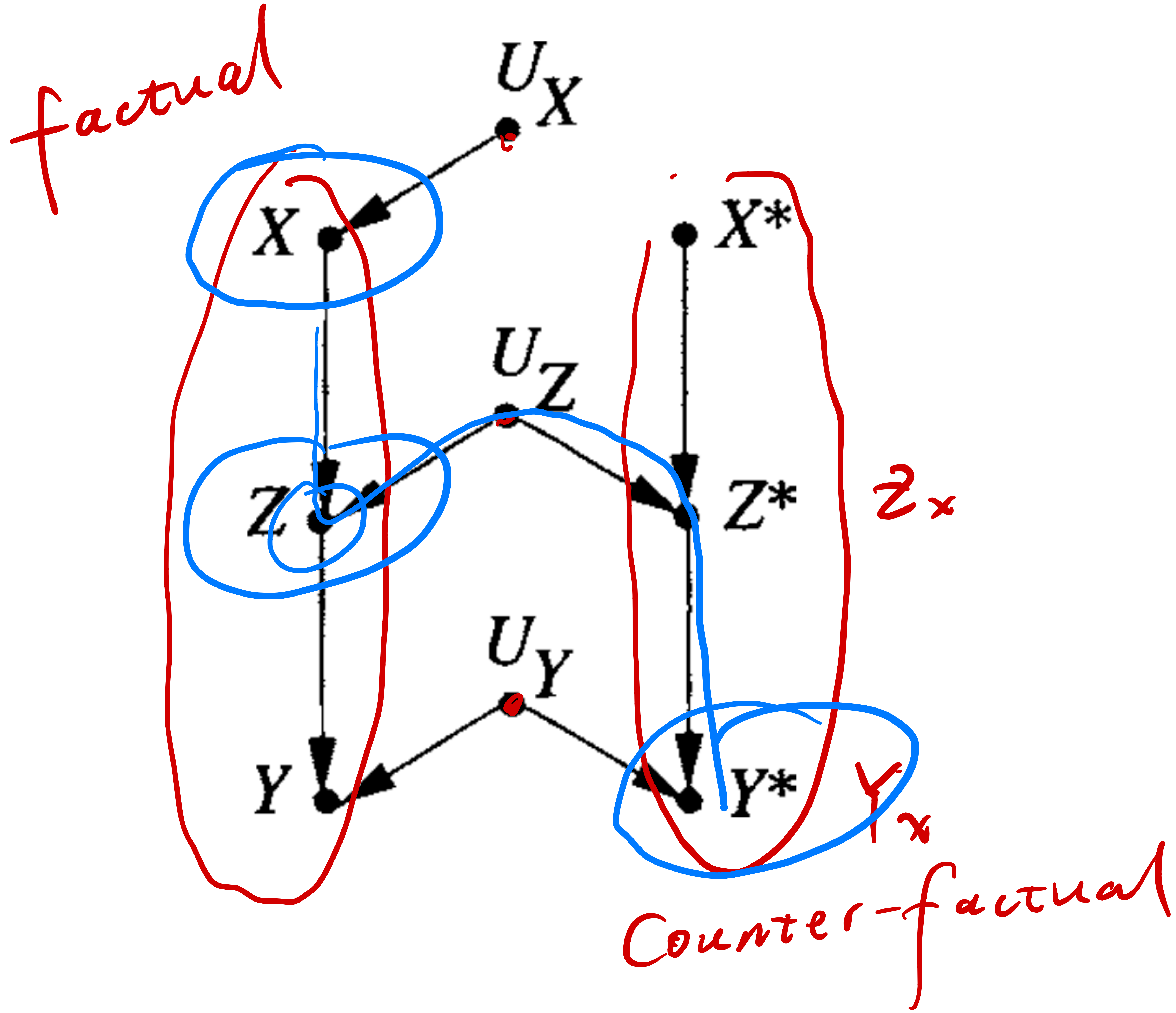
$$X \perp\!\!\!\perp U_2 \mid Z$$

$$\mathbb{E}[Y_{X=1} | Z = 1]$$

(PRIMER, CH4)

The Twin Network Method

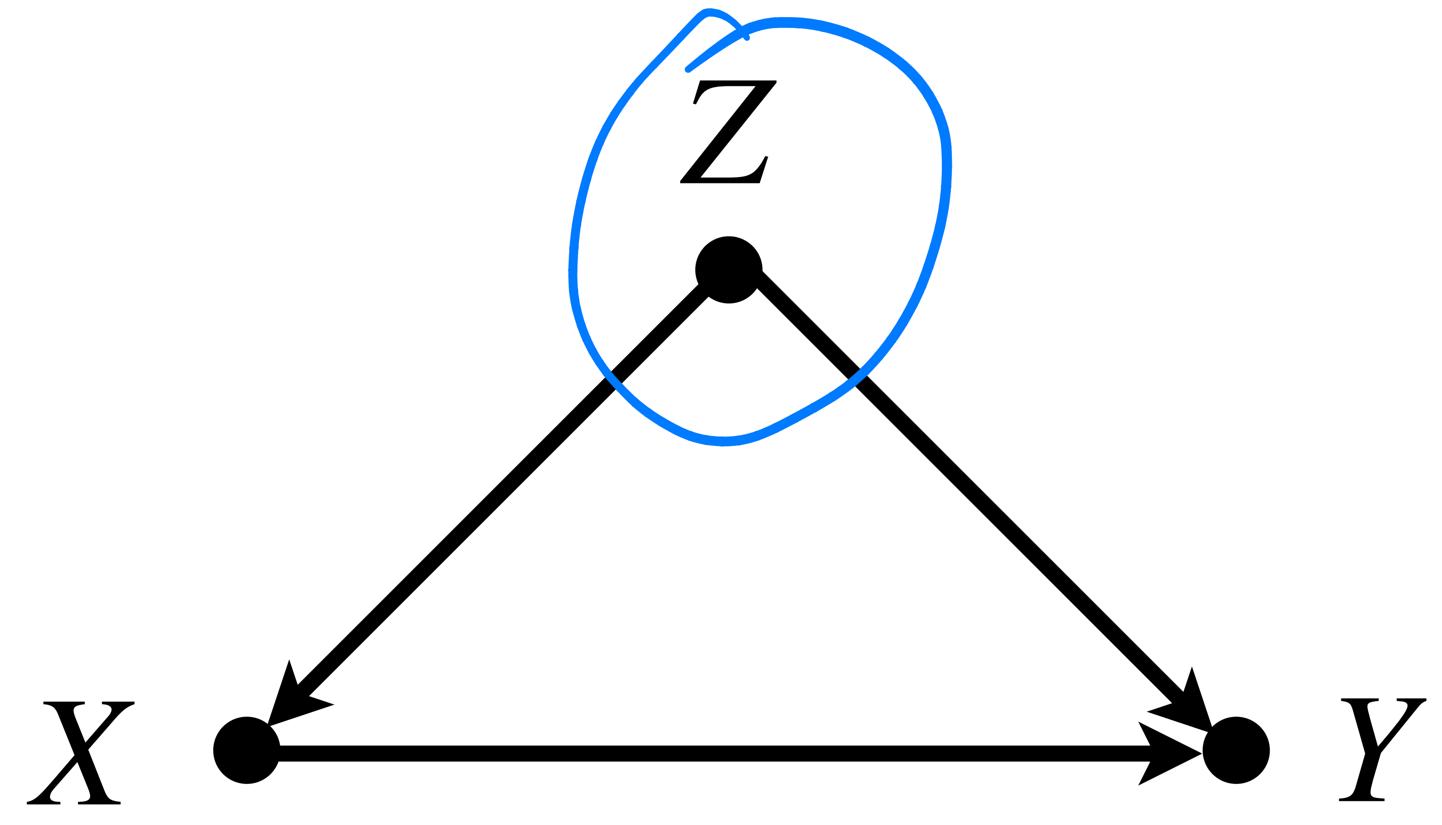
$x \longrightarrow z \longrightarrow Y$

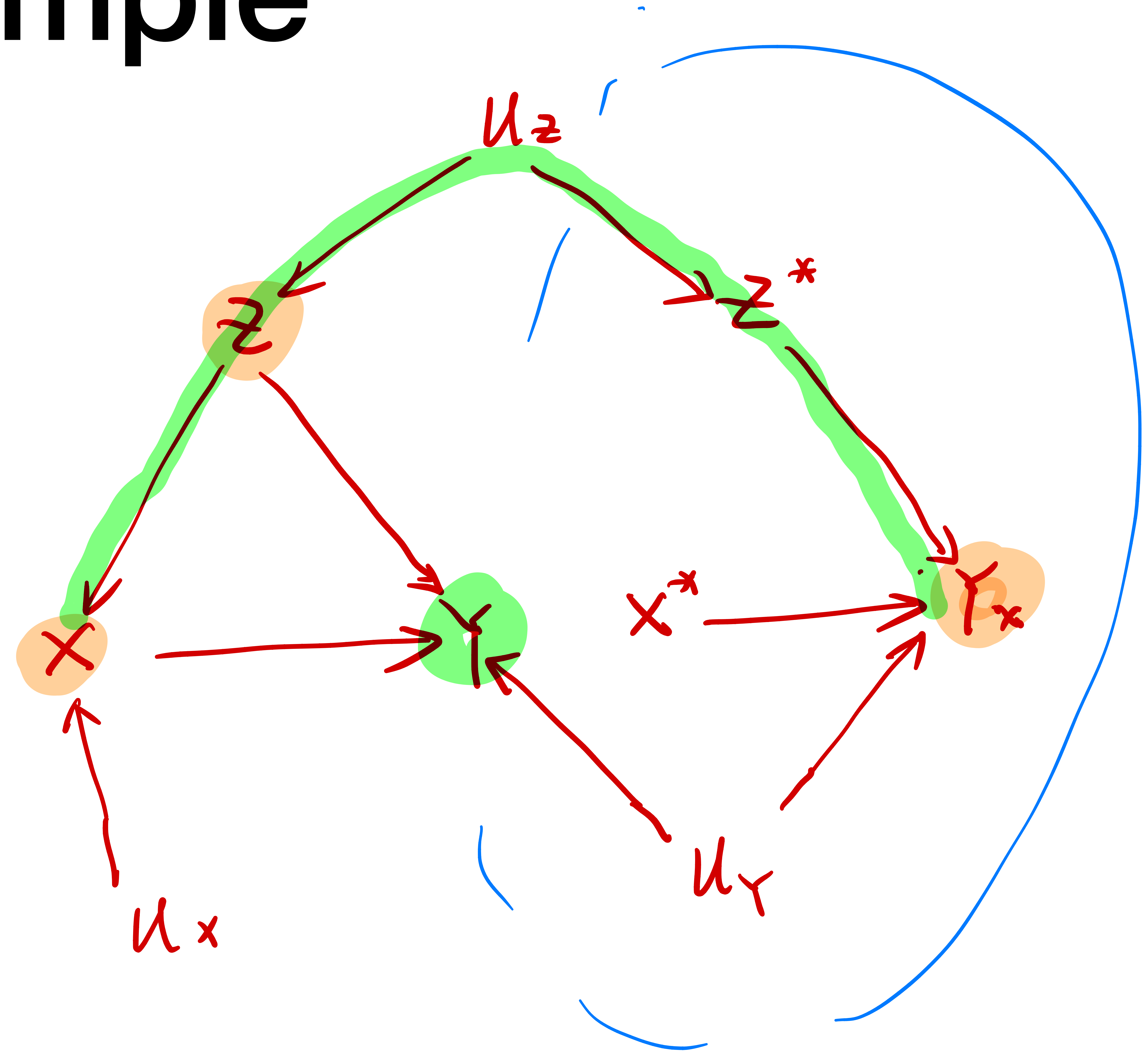


$$Y_x \perp\!\!\!\perp X \mid Z \quad ?$$

(CAUSALITY, CH7)

Example

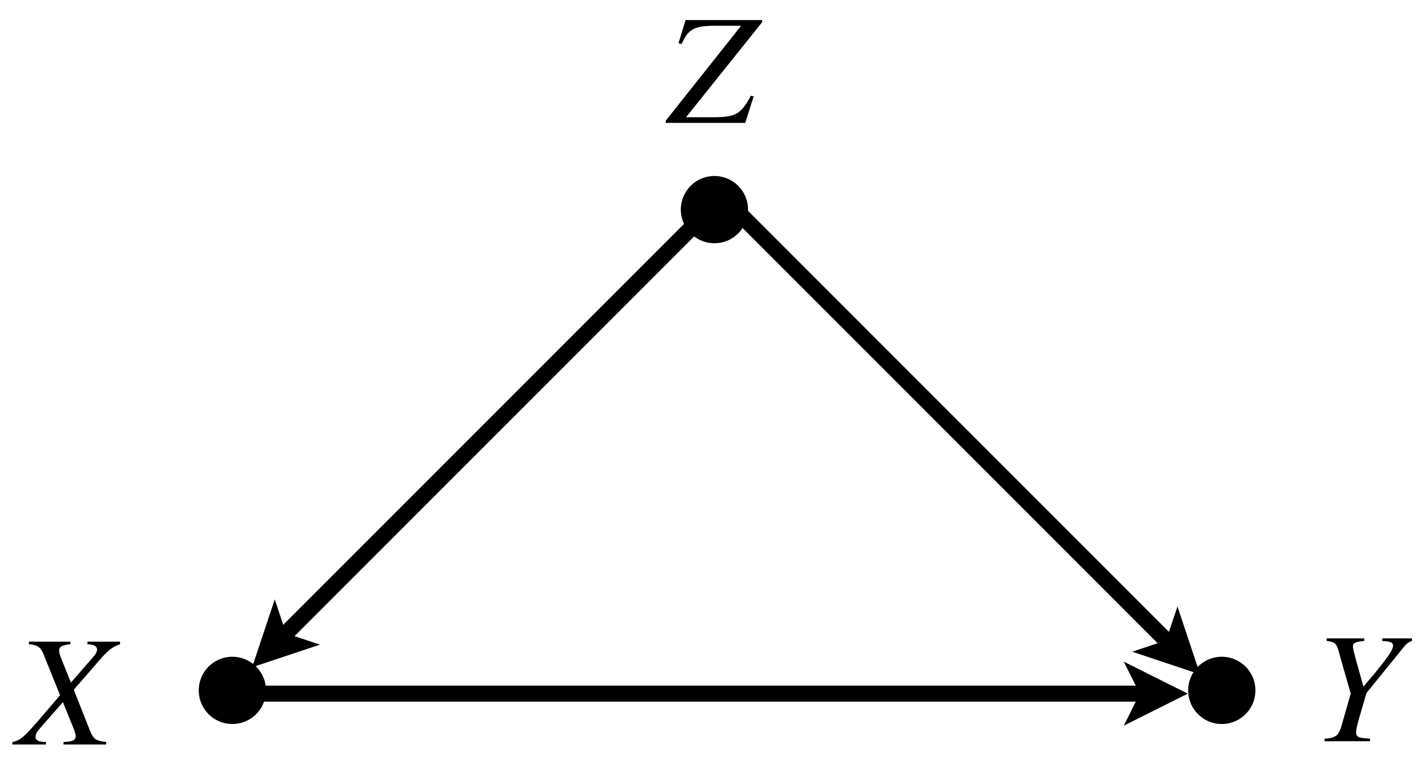




$(Y(0), Y(1)) \perp\!\!\!\perp X$

$(Y(0), Y(1)) \perp\!\!\!\perp X \mid Z$

Counterfactual Interpretation of Backdoor



Theorem 4.3.1 (Counterfactual Interpretation of Backdoor) *If a set Z of variables satisfies the backdoor condition relative to (X, Y), then, for all x, the counterfactual Y_x is conditionally independent of X given Z*

$$P(Y_x|X, Z) = P(Y_x|Z) \tag{4.15}$$

(PRIMER, CH4)

Connections

How does POM work?

- **“Mud does not cause rain.”**

- The probability of the counterfactual event “**rain if it were not muddy**” is the same as the probability of “**rain if it were muddy**”.

- **Causal judgements** are expressed as constraints on probability functions **involving counterfactual variables**.

(CAUSALITY, CH3)

How does POM work?

- The potential-outcome analysis proceeds by imaging **observed distribution** $P(x_1, \ldots, x_n)$ as marginal distribution of **an augmented probability function** P^* defined over **both observed and counterfactual variables**.

- For example, $P(y \mid do(x))$ is phrased as $P^*(Y_x = y)$.

- The potential-outcome approach views **the variable** Y **under** $do(X)$ to be a different **counterfactual variable** Y_x.

- The **counterfactual variable** Y_x can be connected to **observed variable** X **and** Y via **consistency constraints:** $X = x \implies Y_x = Y$

(CAUSALITY, CH3)

From Graphs to Potential Outcomes

- **Exclusion restrictions**: For every variable Y having parents PA_Y and for every set of variables S disjoint of PA_Y, we have

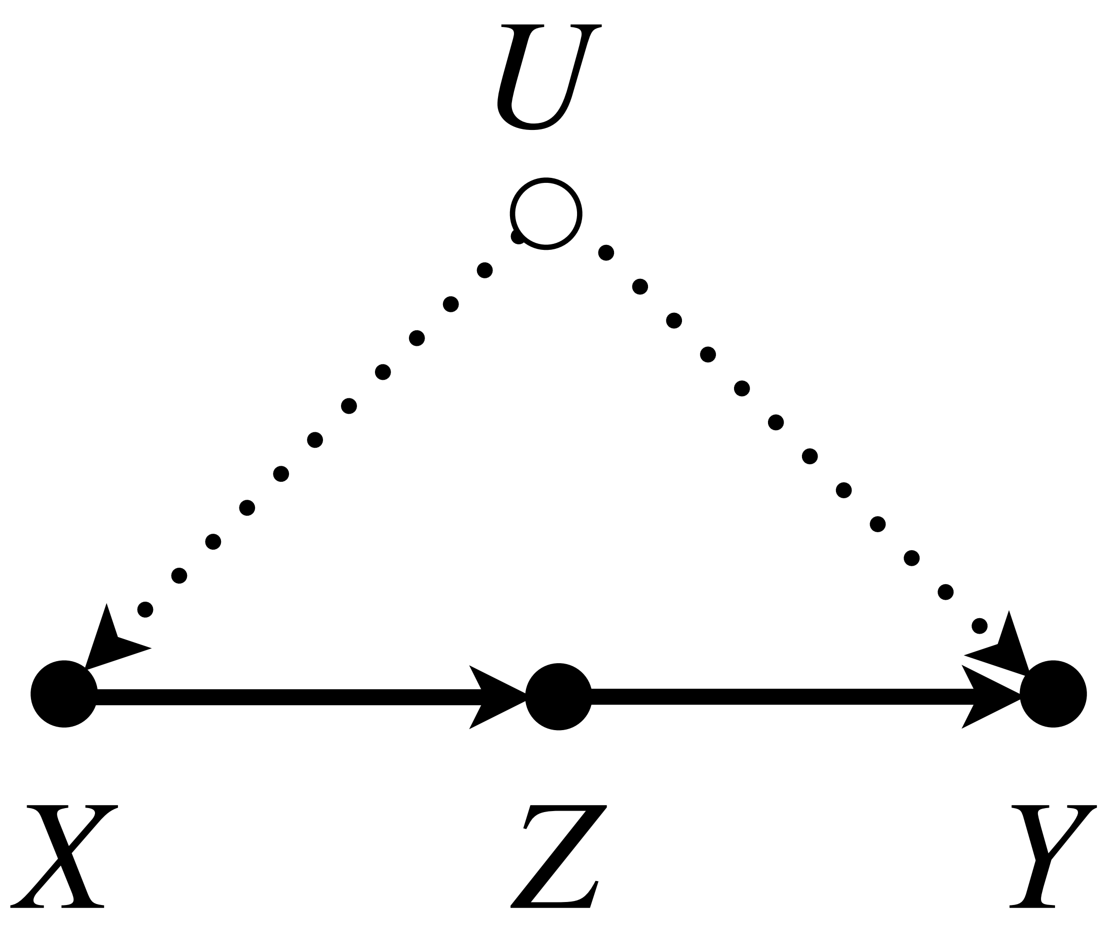
$$Y_{pa_Y}(u) = Y_{pa_Y, s}(u)$$

- **Independence restrictions**: If $Z_1, \ldots, Z_k$ is any set of nodes not connected to Y via dashed arcs, we have

$$\underset{\Downarrow \atop \text{"}U_Y\text{"}}{Y_{pa_Y}} \perp\!\!\!\perp \{\underset{\Downarrow \atop U_{Z_1}}{Z_{1_{pa_{Z_1}}}}, \ldots, \underset{\Downarrow \atop U_{Z_k}}{Z_{k_{pa_{Z_k}}}}\}$$

(CAUSALITY, CH7)

Example



$$Y_{pa_Y}(u) = Y_{pa_Y, s}(u)$$

$$Y_{pa_Y} \perp\!\!\!\perp \{Z_{1_{pa_{Z_1}}}, \ldots, Z_{k_{pa_{Z_k}}}\}$$

$PA_X = \{\emptyset\}$, $PA_Z = \{X\}$, $PA_Y = \{Z\}$

$Z_X = Z_{X,Y}$

$X_Y = X_{Y,X} = X_Z = X$

$Y_Z = Y_{Z,X}$

$Z_X \perp\!\!\!\perp Y_Z, X$

(CAUSALITY, CH7)

Axiomatic Characterization

- **Composition**: For any three sets of endogenous variables X, Y, and W in a causal model, we have

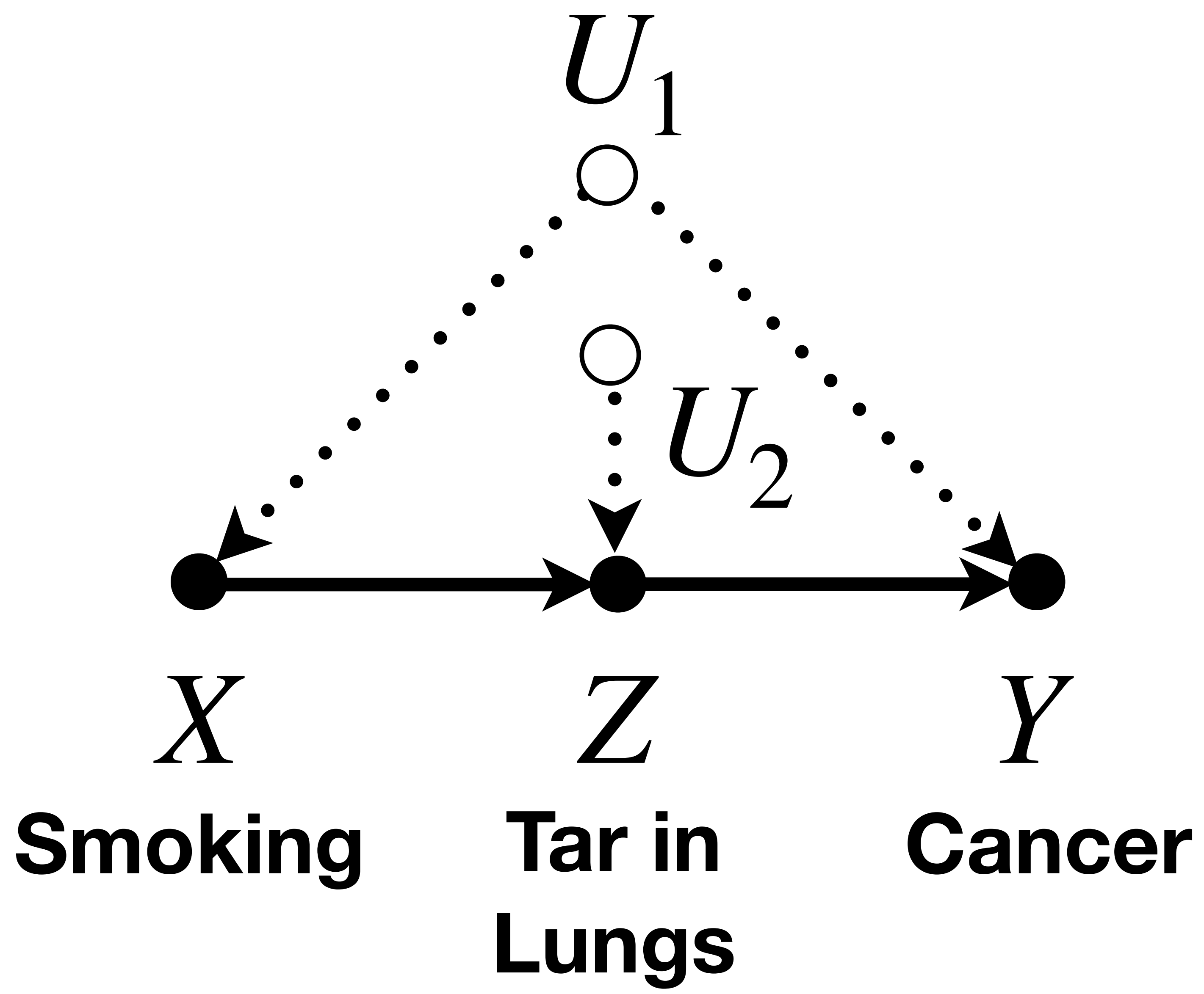
$$W_x(u) = w \implies Y_{xw}(u) = Y_x(u)\,.$$

- **Effectiveness**: For all sets of variables, we have

$$X_{xw}(u) = x.$$

(CAUSALITY, CH7)

Example from Counterfactual Logic



(CAUSALITY, CH7)

Example from Counterfactual Logic

$$Z_x(u) = Z_{yx}(u),$$
$$X_y(u) = X_{zy}(u) = X_z(u) = X(u),$$
$$Y_z(u) = Y_{zx}(u),$$
$$Z_x \perp\!\!\!\perp \{Y_z, X\}.$$

Composition:

$$W_x(u) = w \implies Y_{xw}(u) = Y_x(u).$$

Effectiveness:

$$X_{xw}(u) = x.$$

Task 1

Compute $P(Z_x = z)$
(i.e., the causal effect of smoking on tar).

$X \to Z$

$P(Z_x = z) = P(Z_x = z \mid x)$

$Z_x(u)$

$= P(Z = z \mid x) = P(z \mid x)$

$\boxed{X = x, \quad Y = y}$

$Y_x = y$

(CAUSALITY, CH7)

Example from Counterfactual Logic

$$Z_x(u) = Z_{yx}(u),$$

$$X_y(u) = X_{zy}(u) = X_z(u) = X(u),$$

$$Y_z(u) = Y_{zx}(u),$$

$$Z_x \perp\!\!\!\perp \{Y_z, X\}.$$

Composition:

$$W_x(u) = w \implies Y_{xw}(u) = Y_x(u).$$

Effectiveness:

$$X_{xw}(u) = x.$$

Task 2

Compute $P(Y_z = y)$
(i.e., the causal effect of tar on cancer).

$Z \longrightarrow Y$

$$P(Y_z = y) = \sum_x P(Y_z = y \mid x)\, P(x)$$

$$P(Y_z = y \mid x) = P(Y_z = y \mid x, Z_x = z)$$

$$= P(Y_z = y \mid x, z)$$

$$= P(y \mid x, z)$$

(CAUSALITY, CH7)

Example from Counterfactual Logic

$$Z_x(u) = Z_{yx}(u),$$
$$X_y(u) = X_{zy}(u) = X_z(u) = X(u),$$
$$Y_z(u) = Y_{zx}(u),$$
$$Z_x \perp\!\!\!\perp \{Y_z, X\}\,.$$

Composition:

$$W_x(u) = w \implies Y_{xw}(u) = Y_x(u)\,.$$

Effectiveness:

$$X_{xw}(u) = x\,.$$

Task 3

Compute $P(Y_x = y)$ $X \longrightarrow Y$
(i.e., the causal effect of smoking on cancer).

(CAUSALITY, CH7)

POM versus SCM

- $Y_x(u)$ stands for the outcome of experimental unit u under a **hypothetical experimental** condition $X = x$.

- In **POM**, $Y_x(u)$ is **NOT** derived from a **causal model** or from any formal representation of **scientific knowledge**, but is taken as a **primitive**.

- $Y_x(u)$ is connected to **the reality** only via **the consistency rule**.

- Consequently, **POM** does NOT provide a mathematical model, **without the guarantee on completeness**.

POM versus SCM

- **The formal equivalence** between **POM** and **SCM** covers issues of **semantics** and **expressiveness** but does **NOT** imply **equivalence** in conceptualisation or **practical usefulness**.

- **SCMs** and their associated **graphs** are particularly useful as means of expressing **assumptions** about **cause-effect relationships**.

- **The major weakness** of **POM** lies in the requirement that assumptions be articulated as **conditional independence relationships** involving **counterfactual variables**.

- The most compelling reason for molding causal assumption in the language of **graphs** is that **such assumptions are needed before the data are gathered**.

(CAUSALITY, CH7)


Independence Restrictions
Exclusion Restrictions
Potential Outcome Model
Composition
Structural Causal Model
Effectiveness
Reversibility