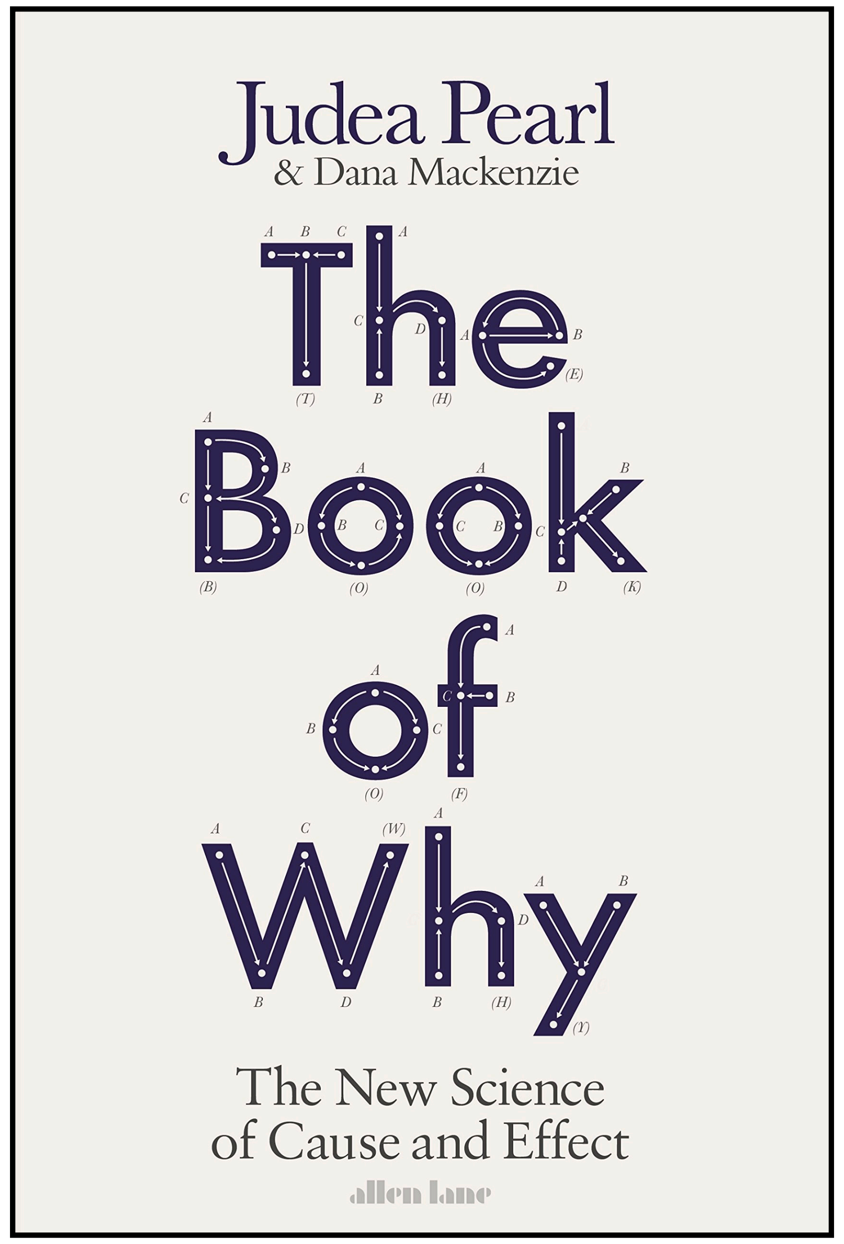
Deep Dive into SCMs and POMs

Chaochao Lu
(陆超超)

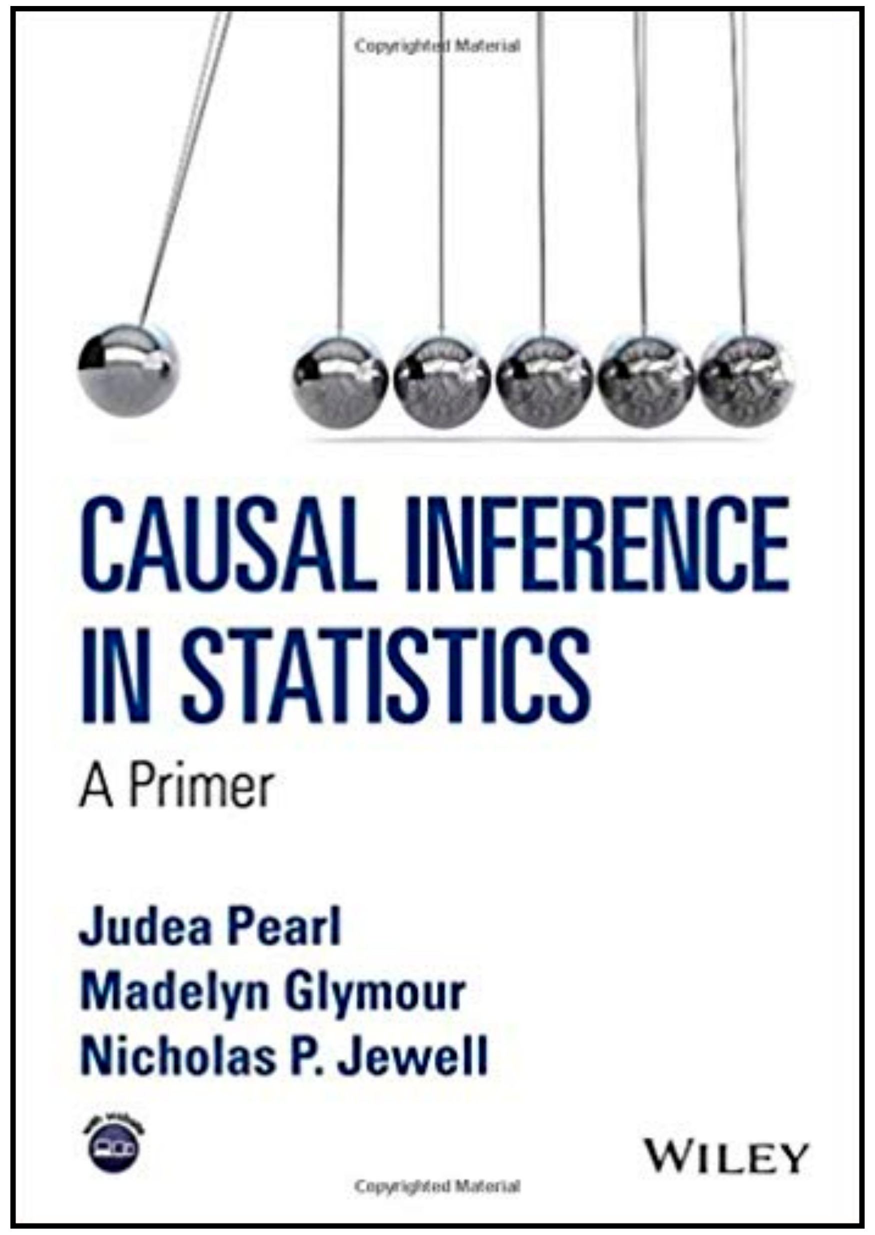
University of Cambridge
Max Planck Institute for Intelligent Systems

因果科学暑期学校 | 17 Aug 2021

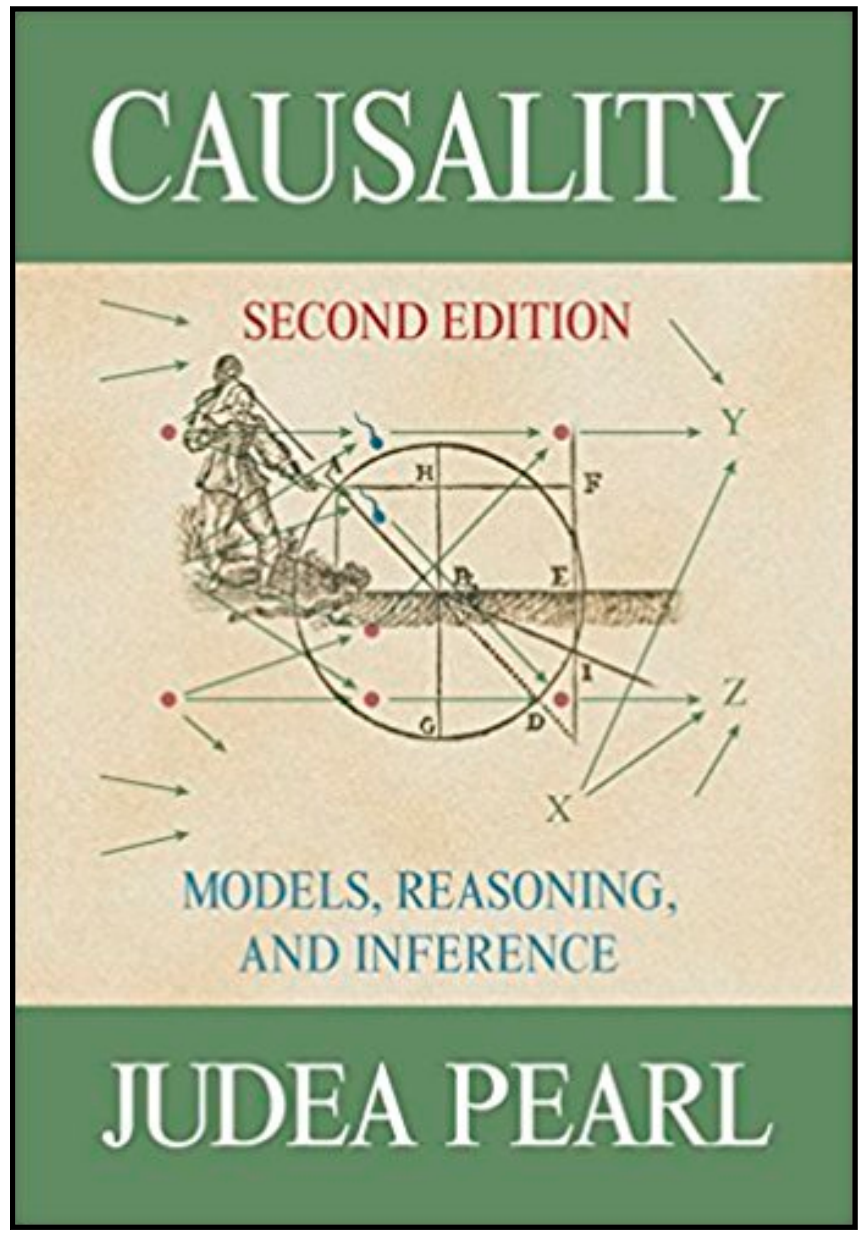
Disclaimer



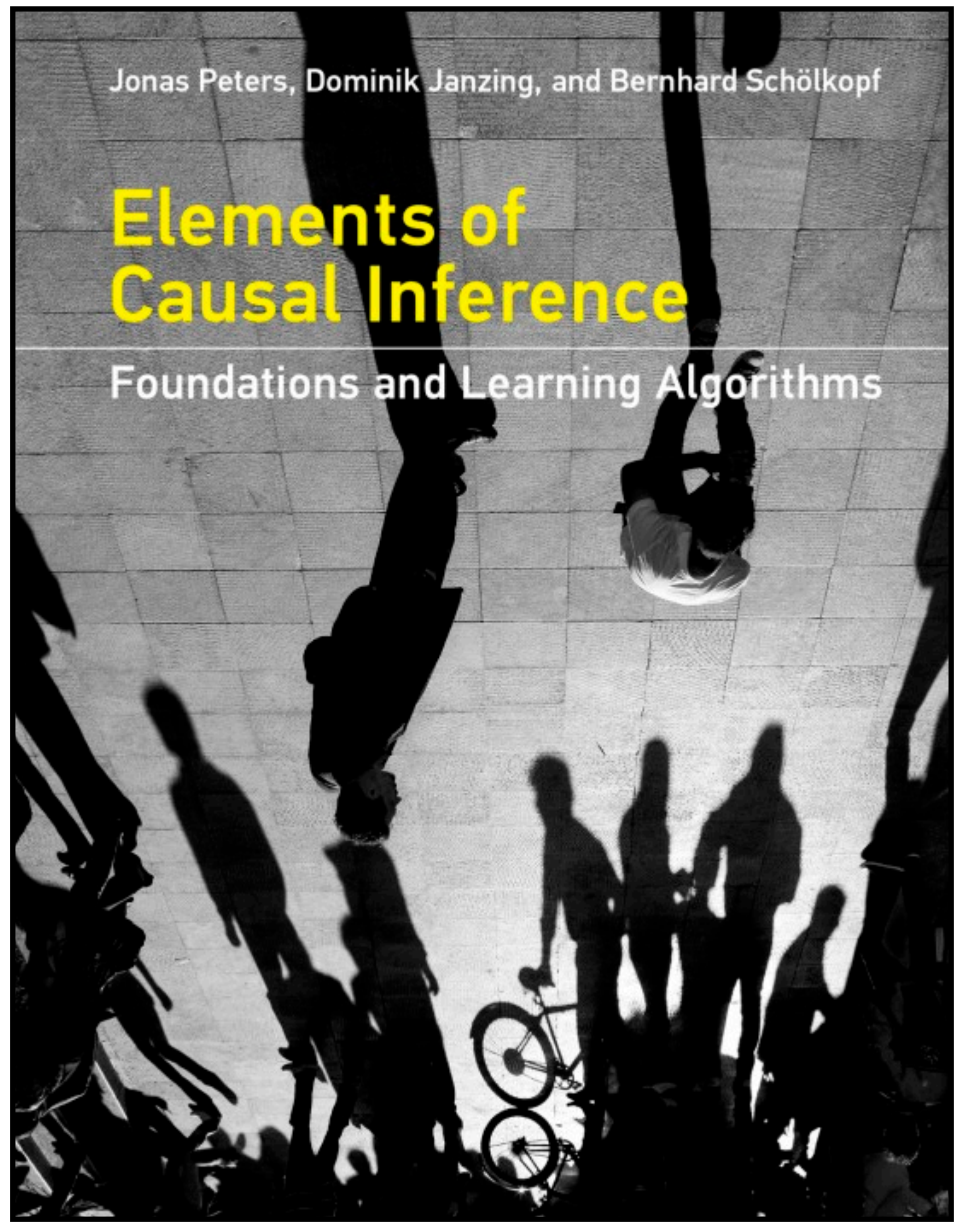
WHY



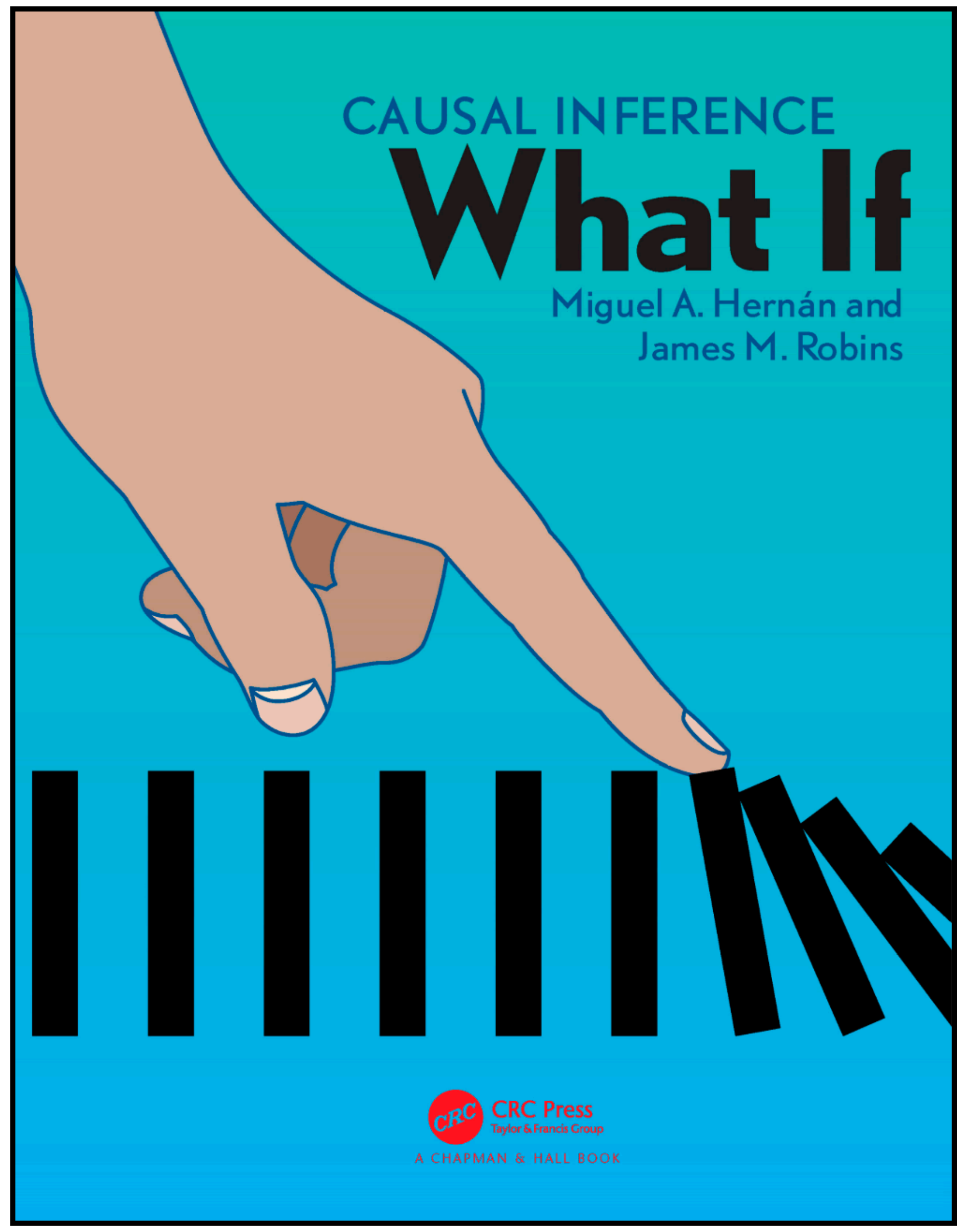
PRIMER



CAUSALITY



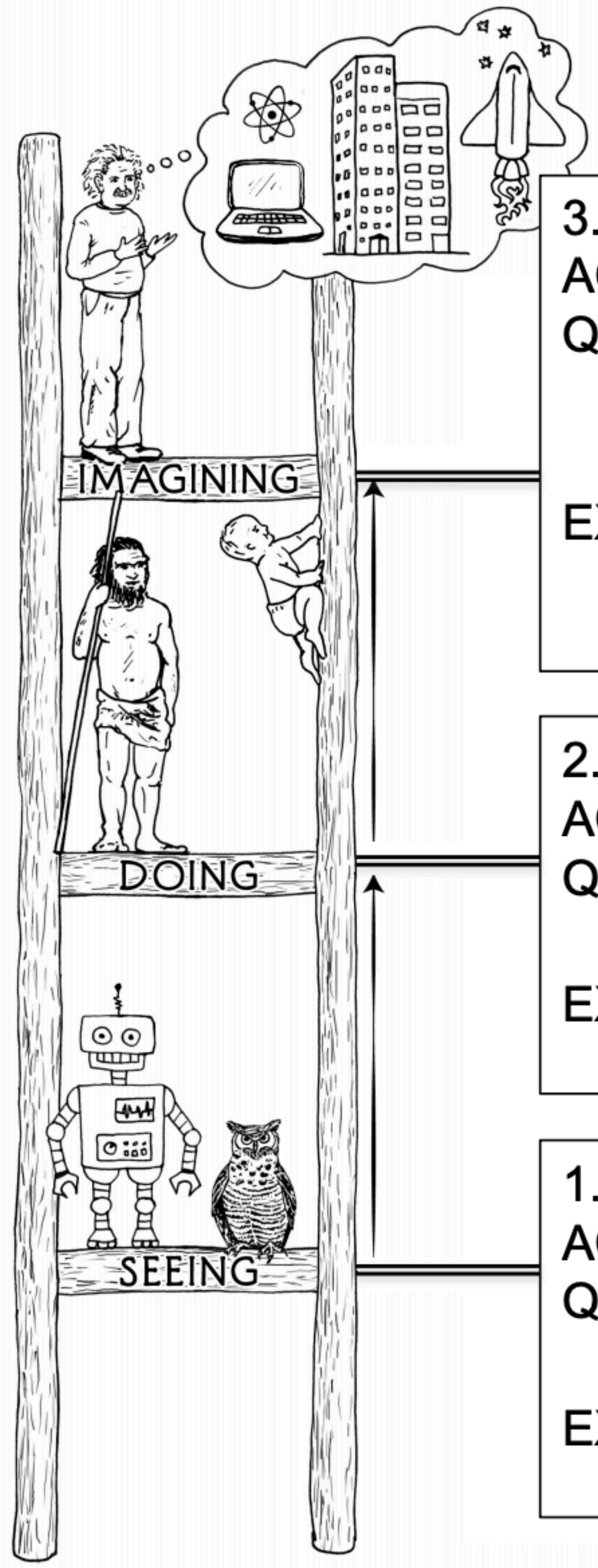
ELEMENTS



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?

$P(Y_{X'} \mid X)$

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?

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

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?

$P(Y \mid X)$

(WHY, CH1)

Structural Causal Models

Structural Causal Model

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

- $V = \{V_1, ..., V_n\}$ are endogenous variables
- $U = \{U_1, ..., U_m\}$ are background variables
- $F = \{f_1, ..., f_n\}$ 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

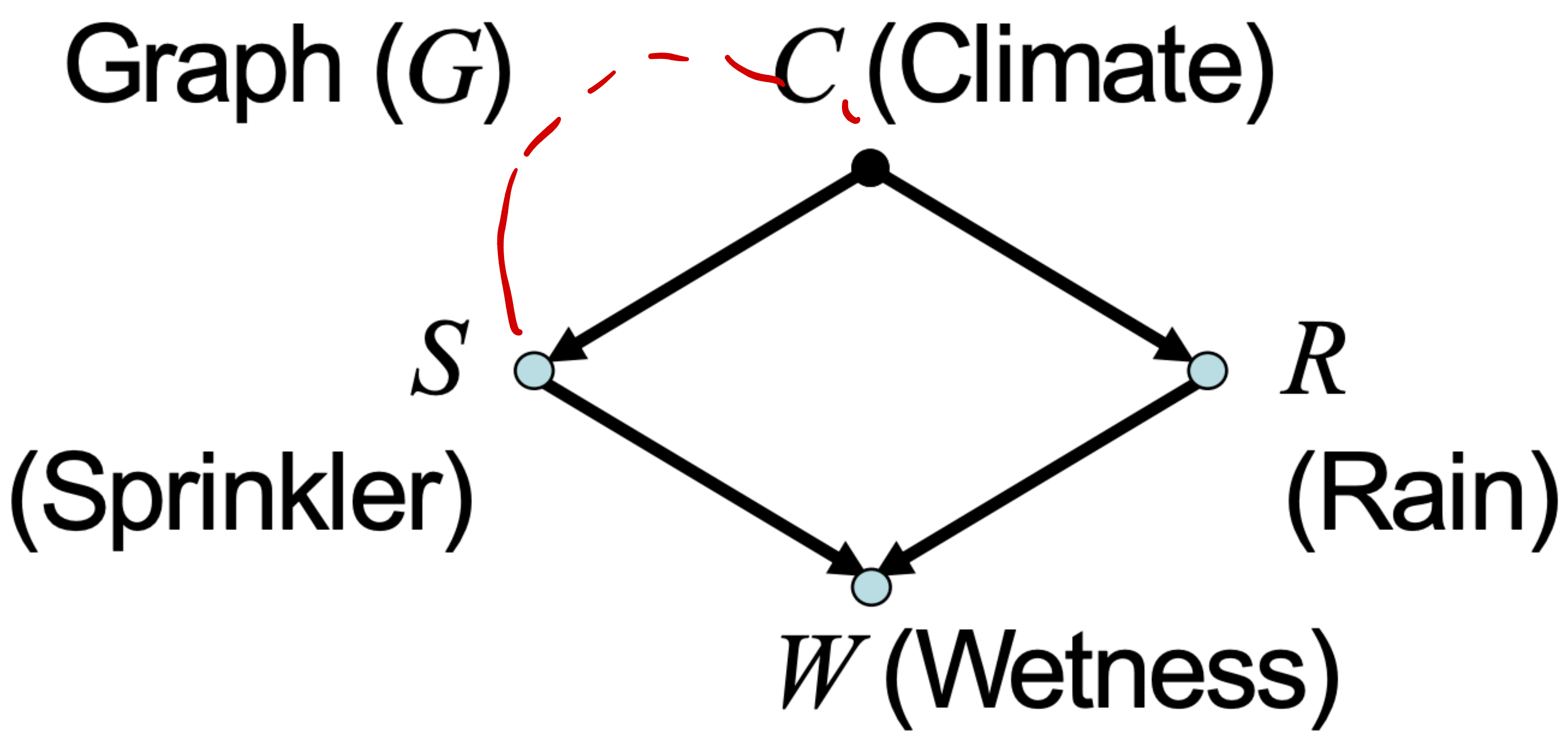
$P(u)$ and F induce a distribution $P(v)$ over observable variables

$V_1 := f(V_m, N)$

$U_1 \perp\!\!\!\perp U_2 \perp\!\!\!\perp \cdots \perp\!\!\!\perp U_m$ ✓

(CAUSALITY, CH7)

Graphical Representation



Model (M)

$$C = f_C(U_C)$$
$$S = f_S(C, U_S)$$
$$R = f_R(C, U_R)$$
$$W = f_W(S, R, U_W)$$

$U_c \not\perp\!\!\!\perp U_s$

Three Building Blocks

$$X \rightarrow Z \rightarrow Y$$

$$X \leftarrow Z \rightarrow Y$$

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

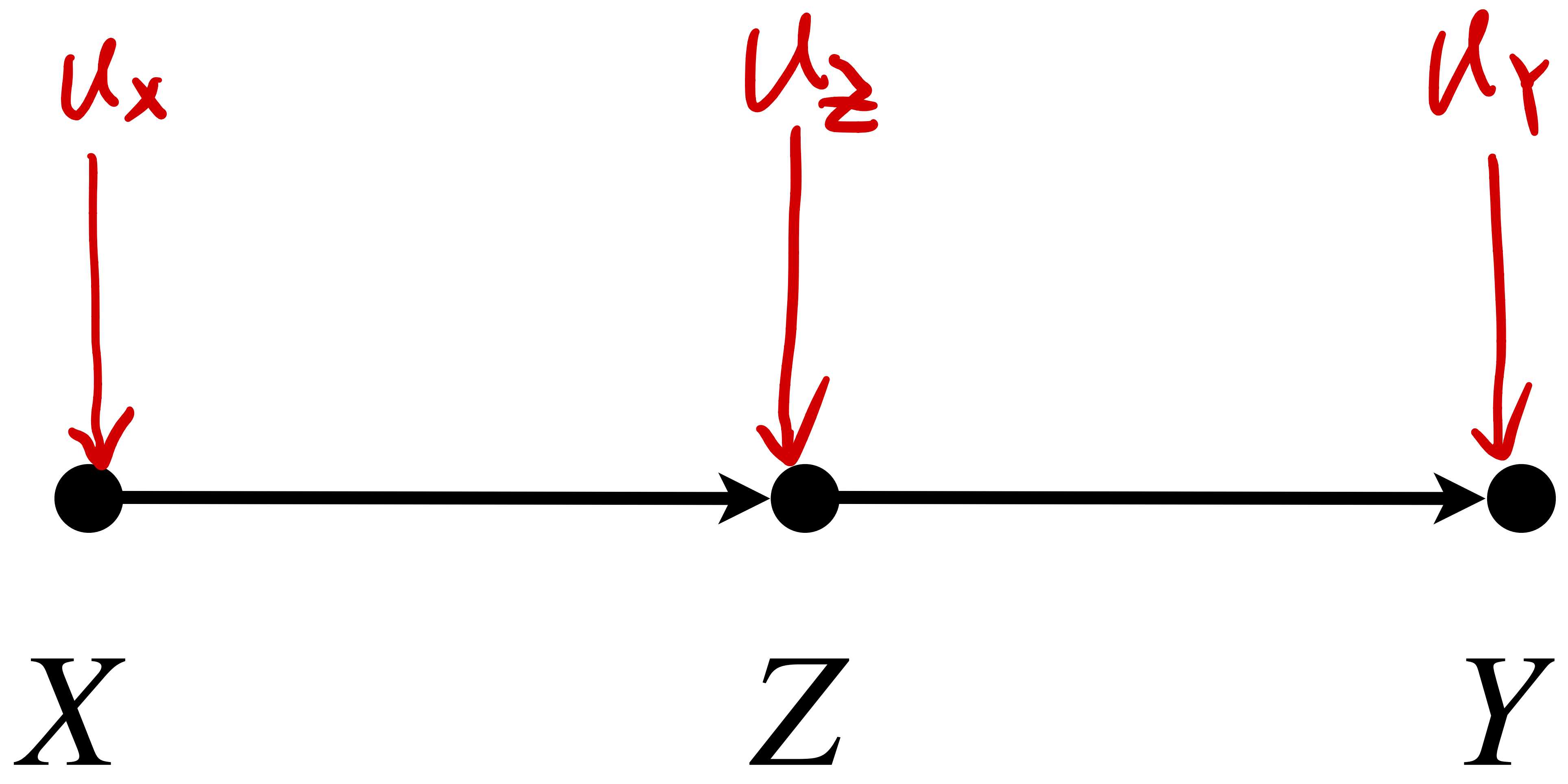
$$X \rightarrow Z \leftarrow Y$$

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

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

(PRIMER, CH2)

Chain



$X \perp\!\!\!\perp Y \mid \underline{Z}$

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

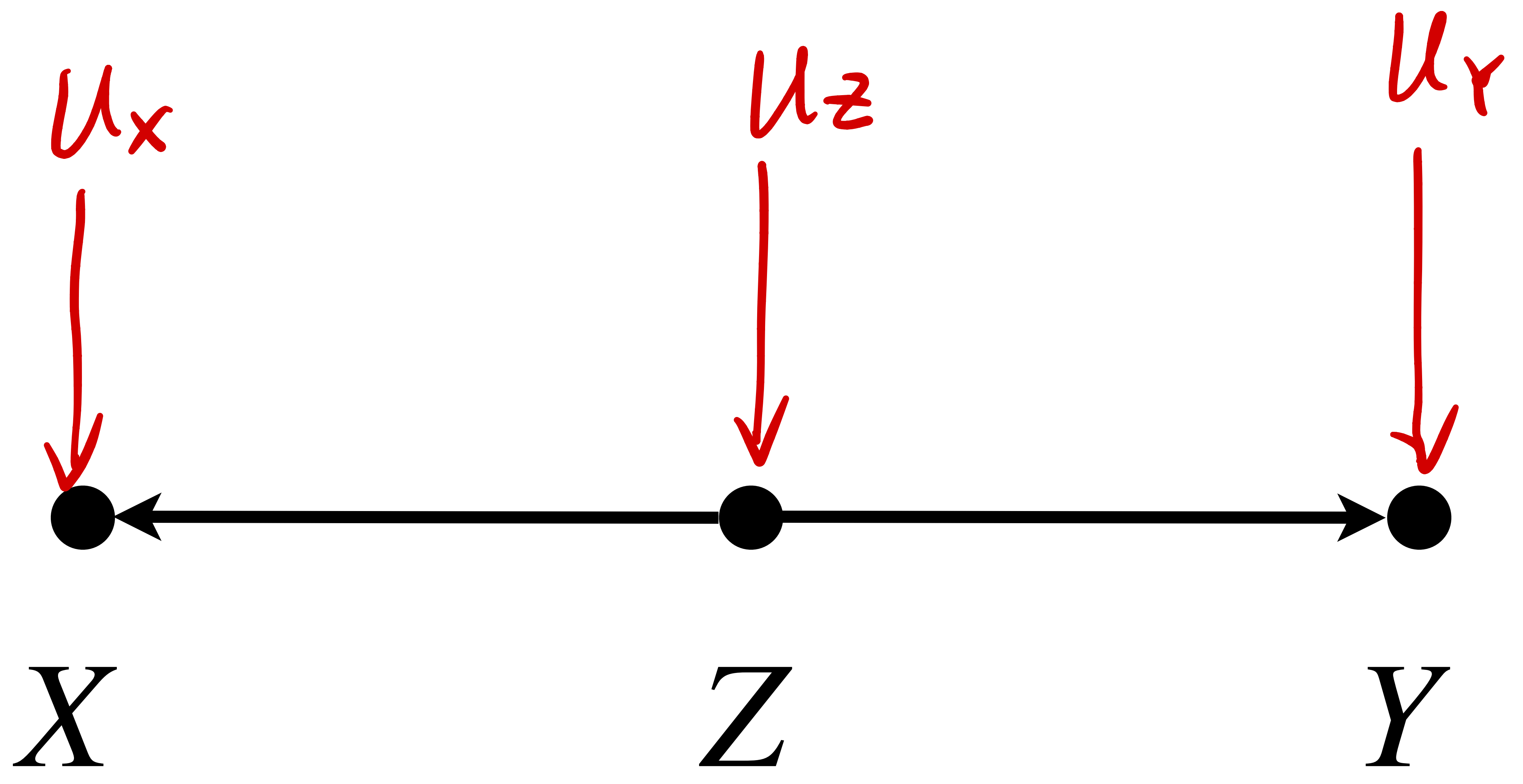
$X = f_X(U_X)$

$U_X \perp\!\!\!\perp U_Y$

$X \perp\!\!\!\perp Y \mid \underline{Z}$

(PRIMER, CH2)

Fork



$X \perp\!\!\!\perp Y \mid \underline{\underline{Z}}$

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

$X = f_X(\boxed{Z}, U_X)$

$U_X \perp\!\!\!\perp U_Y$

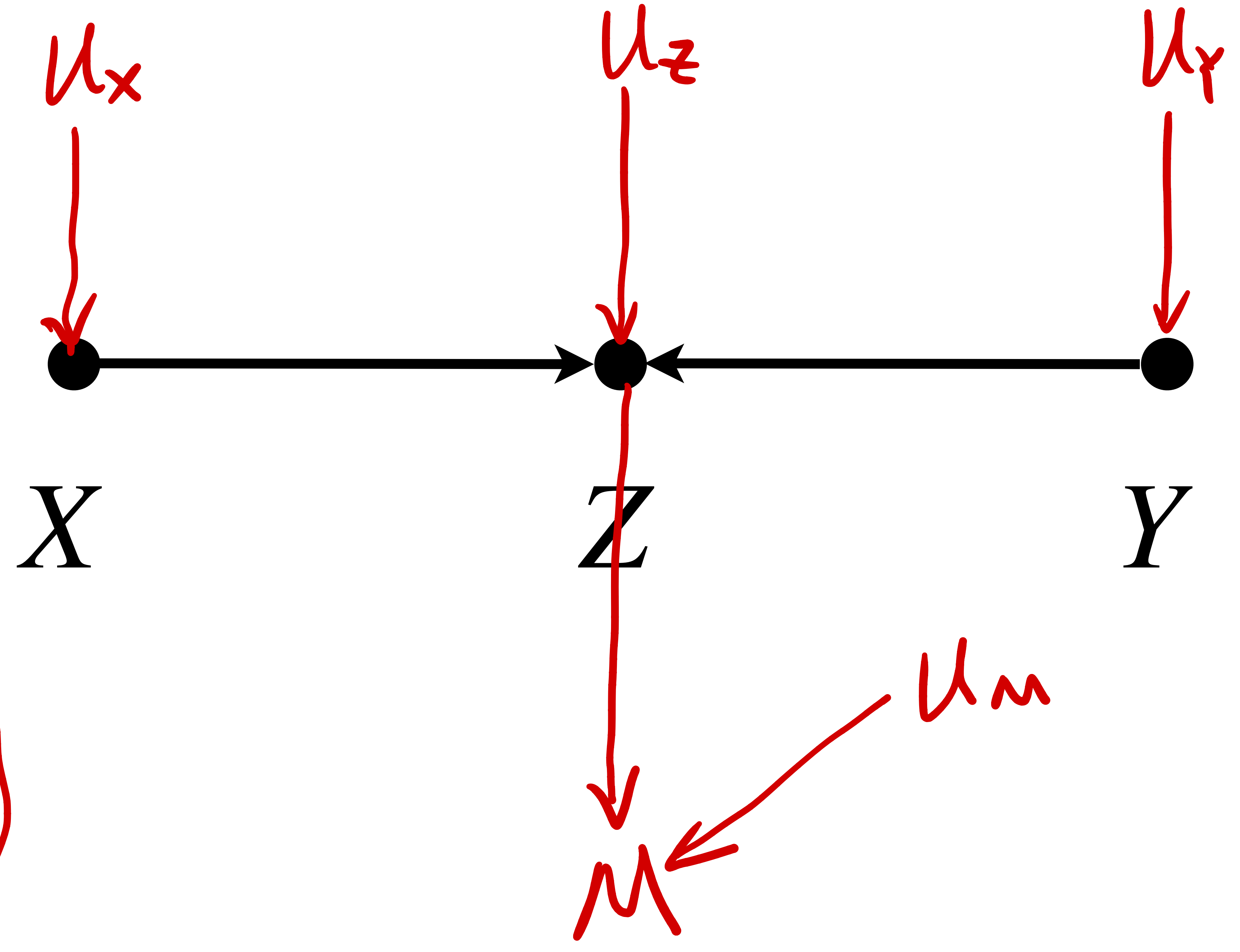
$X \perp\!\!\!\perp Y \mid Z$

(PRIMER, CH2)

Collider

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

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



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

$Y = f_Y(U_Y)$

$X = f_X(U_X)$

$\because U_X \perp\!\!\!\perp U_Y$

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

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

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

$M = f_M(Z, U_M) = f'_M(X, Y, U_Z, U_M)$

(PRIMER, CH2)

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.*

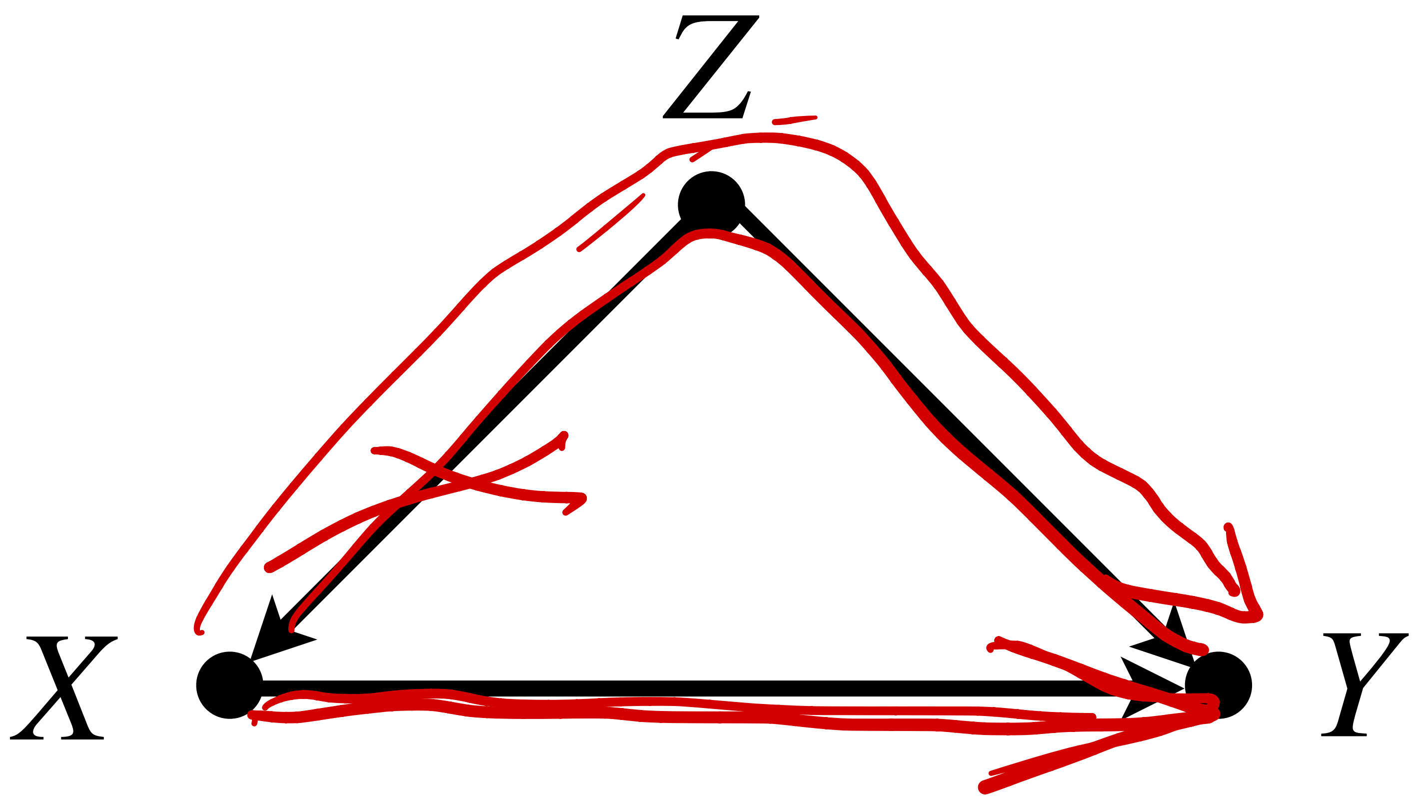
If Z blocks every path between two nodes X and Y, then X and Y are d-separated, conditional on Z, and thus are independent conditional on Z.

(PRIMER, CH2)

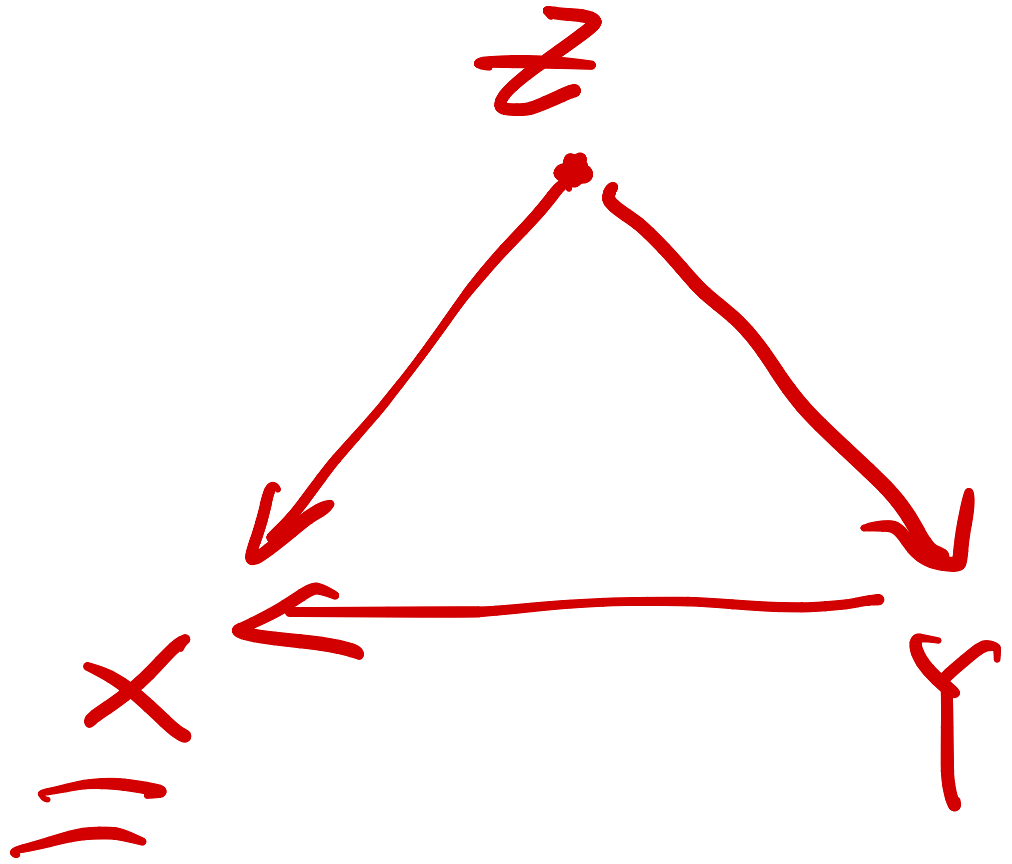
Back-Door Criterion

"Sure Thing"

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.*

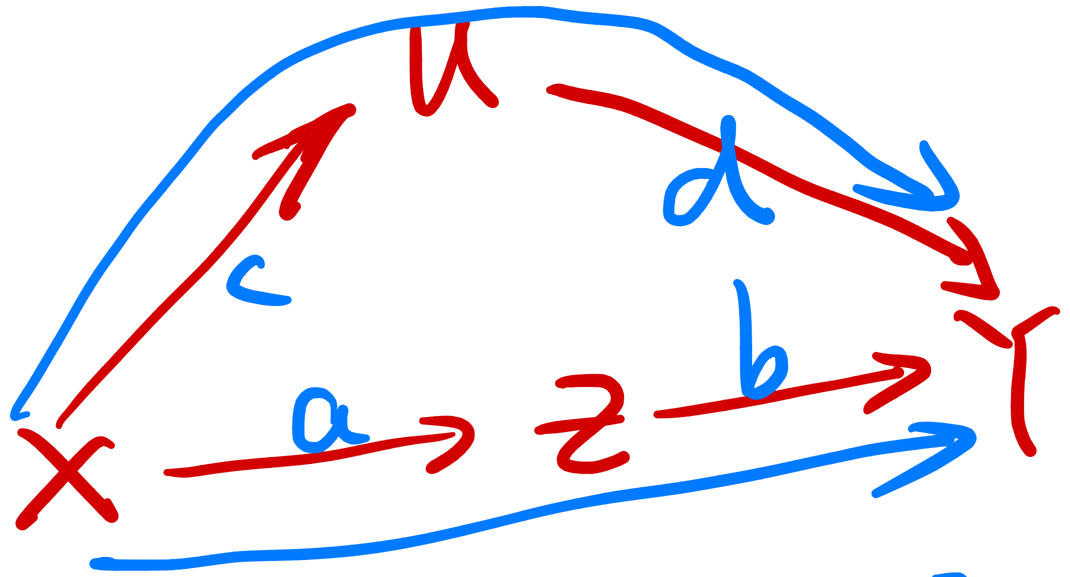


$$p(Y \mid do(x)) = \sum_z p(Y \mid x, \underline{\underline{z}})\, p(\underline{\underline{z}})$$

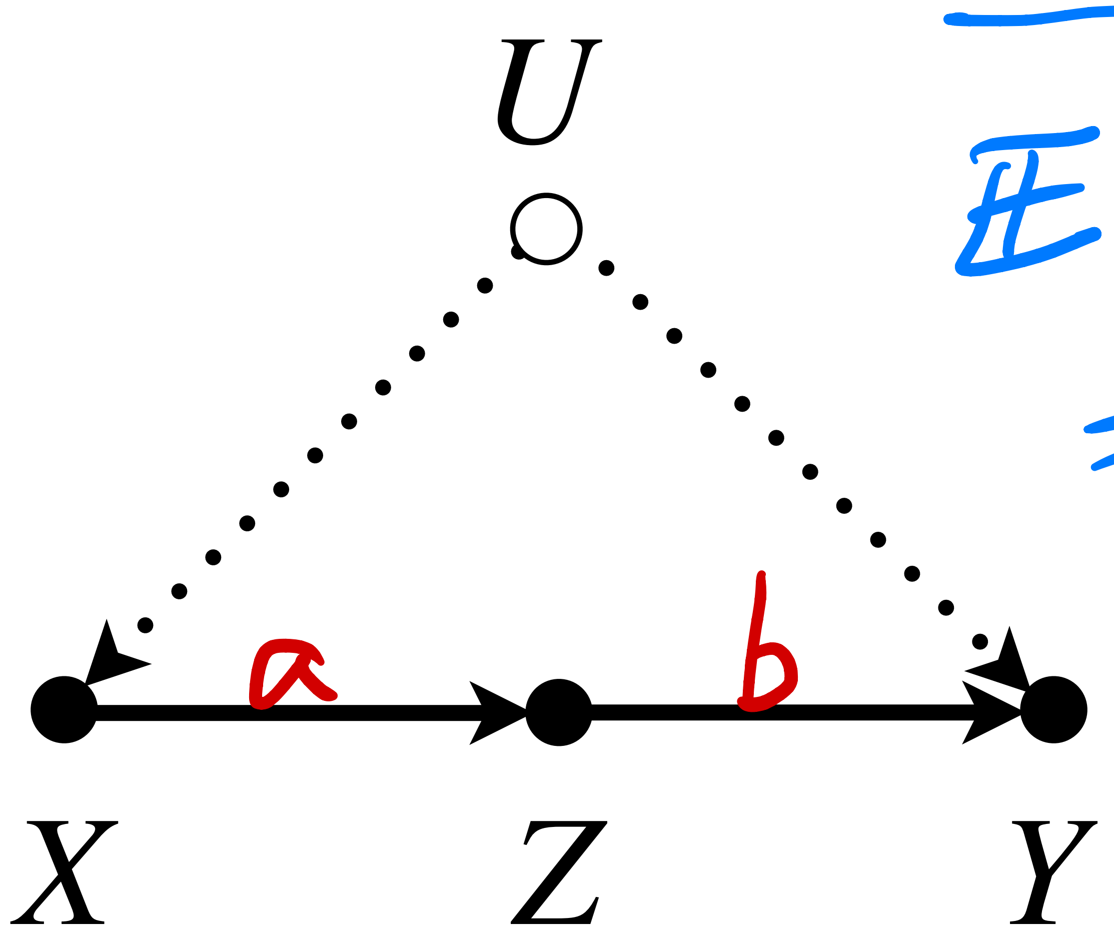


$$p(Y \mid do(x)) = p(Y)$$

Front-Door Criterion



$$E[Y \mid do(X)] = ab + cd$$



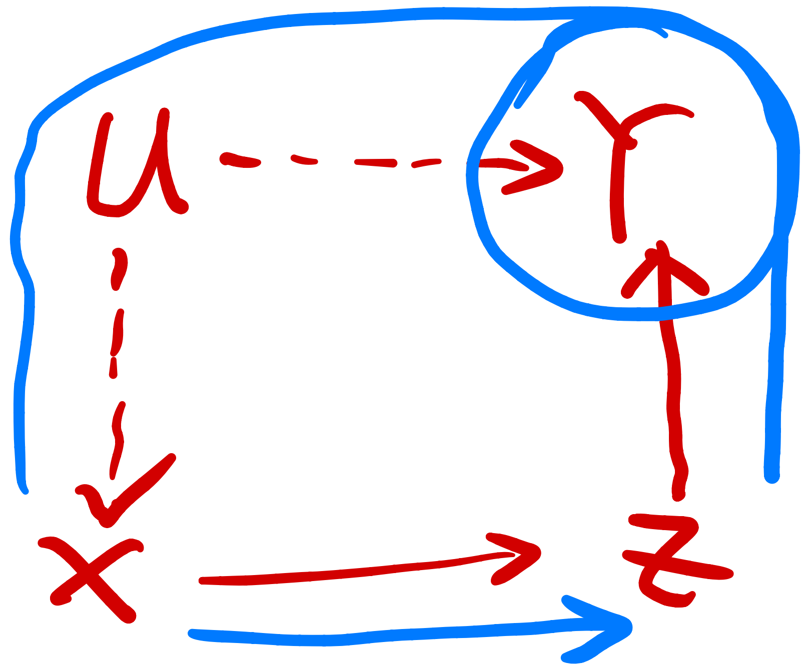
$P(Y \mid do(X))$

$E[Y \mid do(X)] = ab$

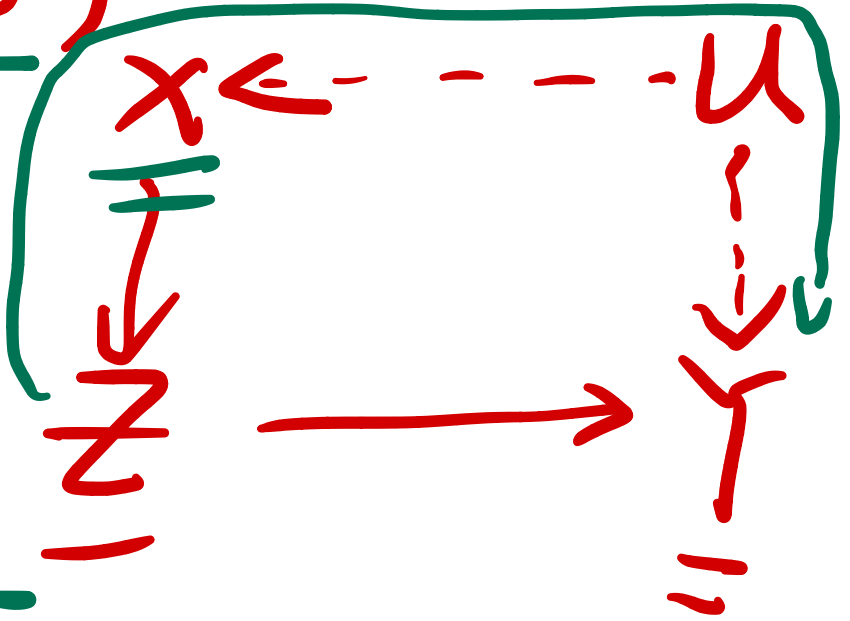
Step I: $P(Z \mid do(X))$

Step II: $P(Y \mid do(Z))$

Step I:
$$P(Z \mid do(X)) = \boxed{P(Z \mid X)} = \sum_{W} P(Z \mid X, W)\, P(W), \quad W = \emptyset$$



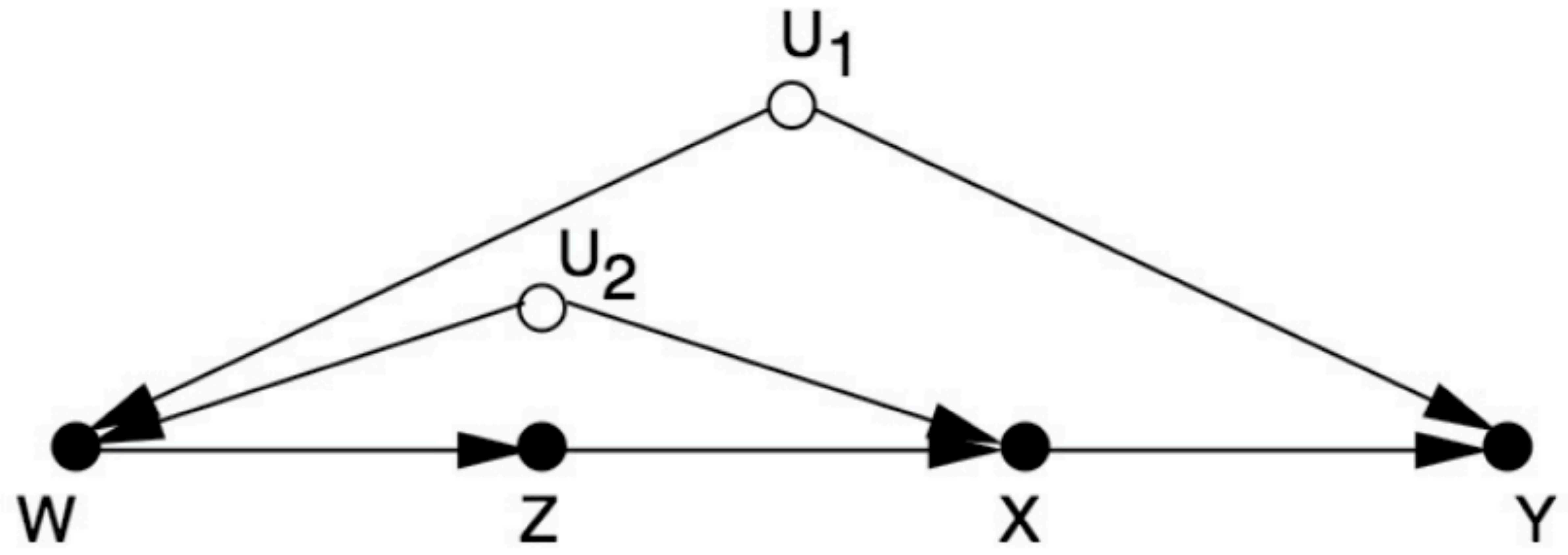
Step II:
$$P(Y \mid do(Z)) = \sum_{X} P(Y \mid Z, X)\, P(X)$$



$$P(Y \mid do(X)) = \underbrace{\sum_{z} P(z \mid x)}_{\text{Step I}} \underbrace{\sum_{x'} P(y \mid x', z)\, P(x')}_{\text{Step II}}$$

(PRIMER, CH3)

Front-Door or Back-Door?



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

(WHY, CH7)

do-Calculus

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

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

Rule 2 (*Action/observation exchange*):

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

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

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

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

(CAUSALITY, CH3)

Insertion/deletion of observations

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

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

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

Action/observation exchange

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

Rule 2 (*Action/observation exchange*): back-door

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

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

$$P(y \mid do(z)) = \sum_{w} P(y \mid z, w) P(w)$$

Insertion/deletion of actions

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

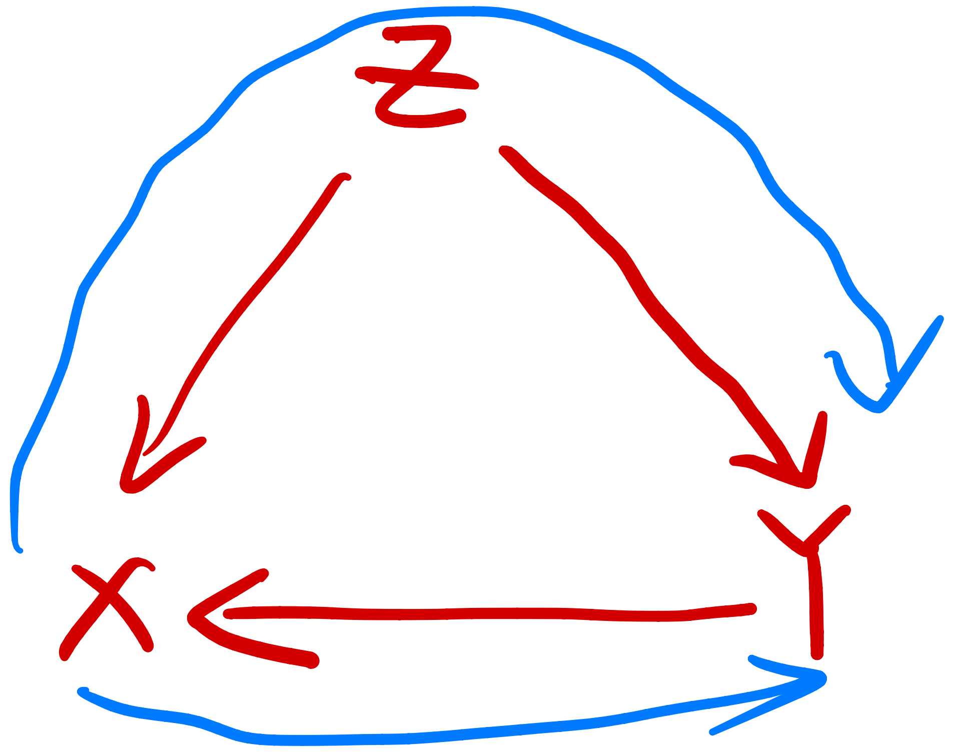
$$P(y \mid \hat{x}, \hat{z}, w) = P(y \mid \hat{x}, w) \text{ if } (Y \perp\!\!\!\perp Z \mid X, W)_{G_{\overline{X}, \overline{Z(W)}}},$$

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

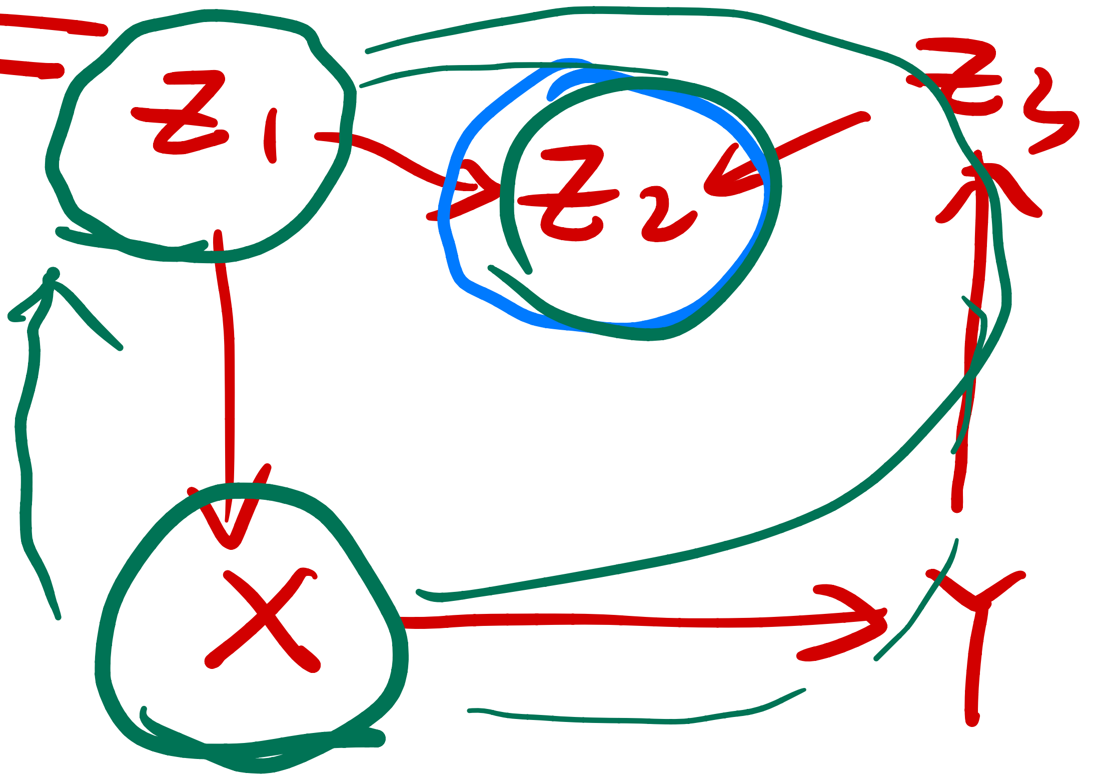
$$p(y \mid \hat{z}) = p(y)$$

$$z \to \cdots \to y$$

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



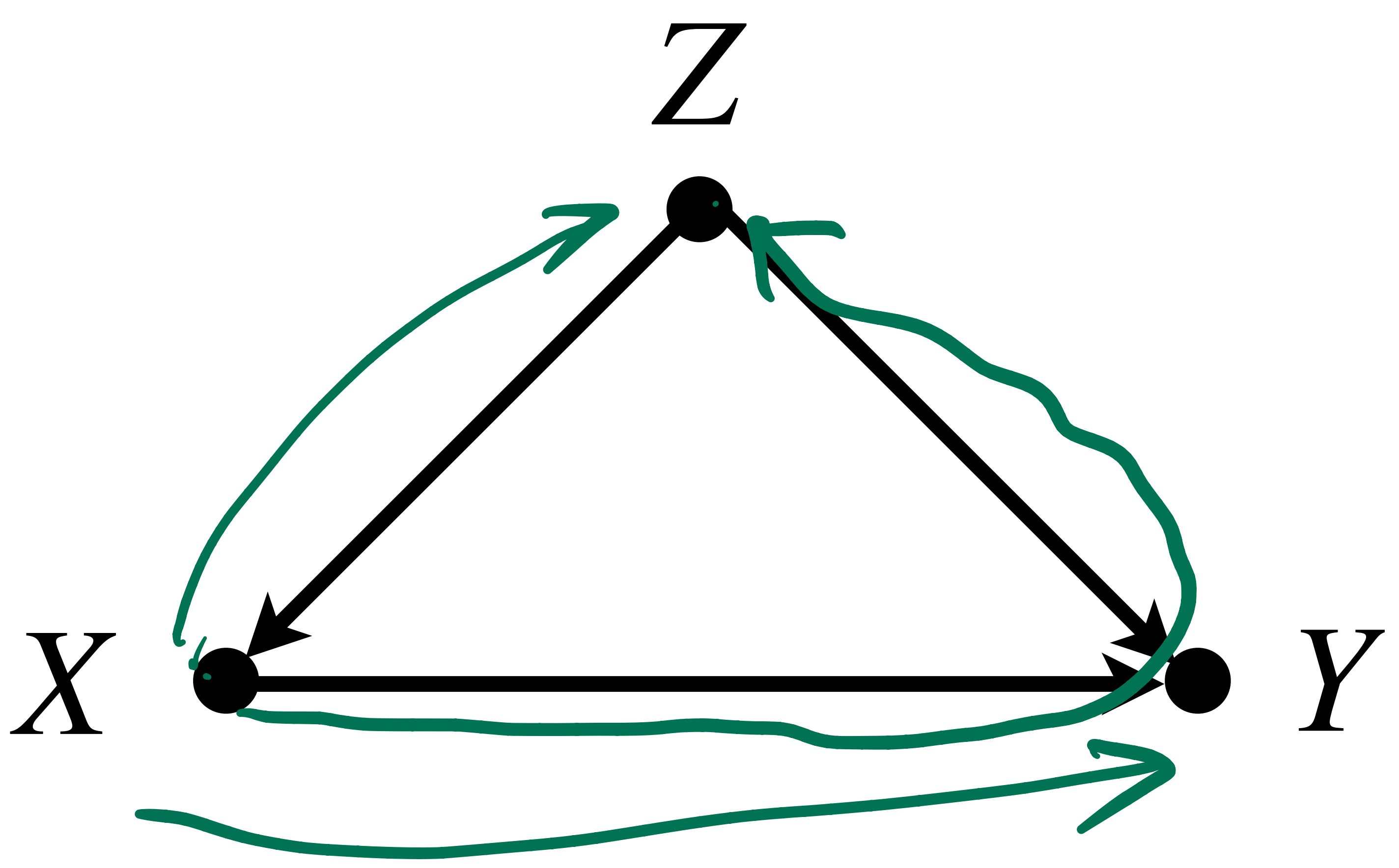
$$P(Y \mid do(x)) = P(Y)$$



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

$$P(z_1 \mid do(x), \boxed{z_2}) \neq P(z_1 \mid z_2)$$

Revisit Back-Door Criterion



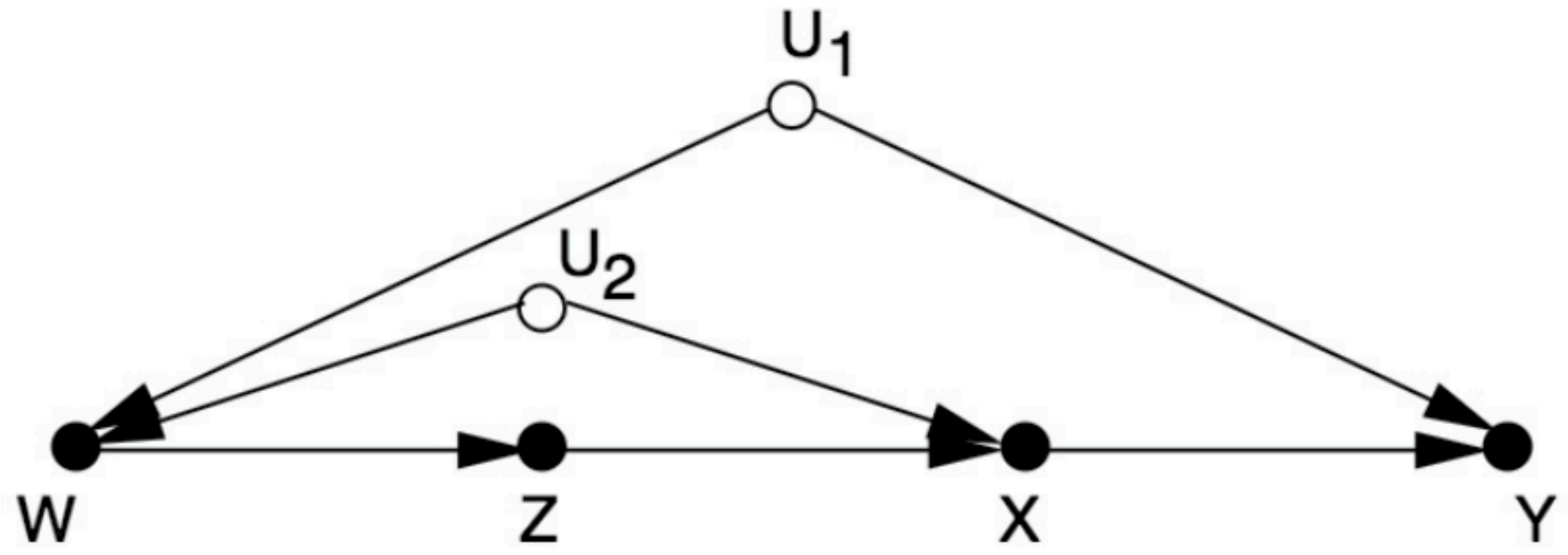
$$P(Y \mid do(x))$$

$$= \sum_z P(Y, z \mid do(x))$$

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

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

Let's do-Calculus!

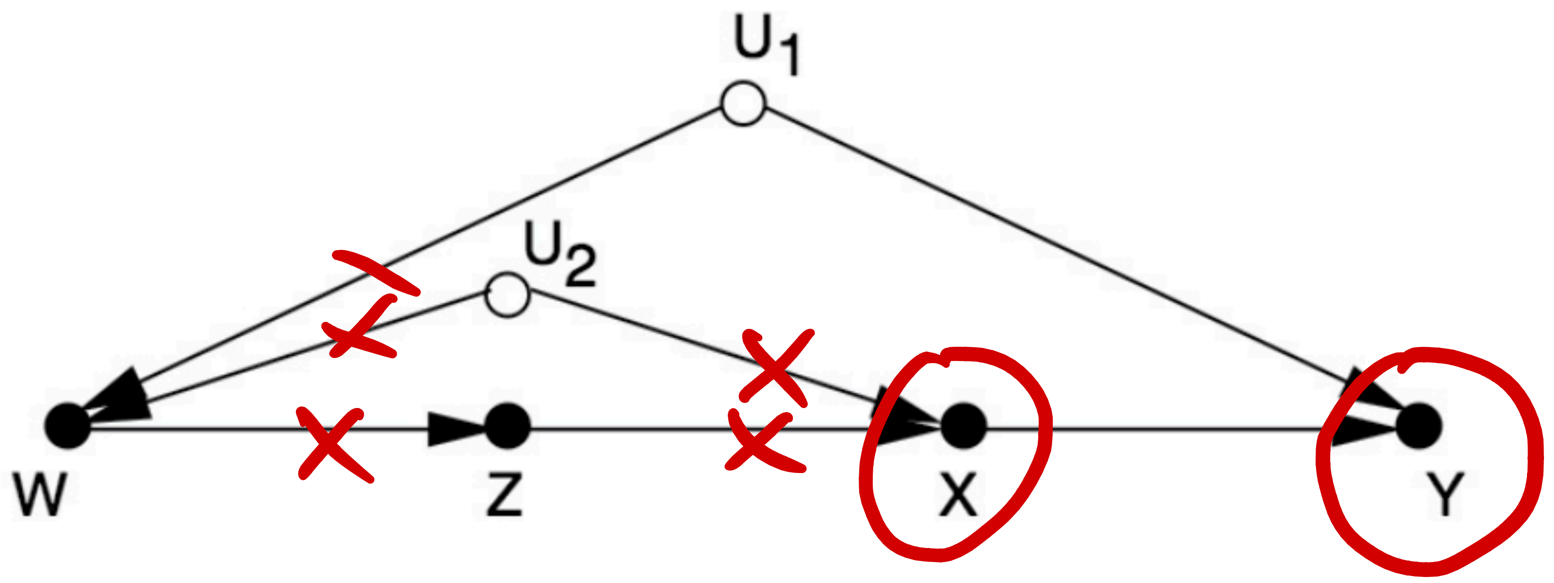


$$P(Y \mid do(X))\,!$$

(WHY, CH7)

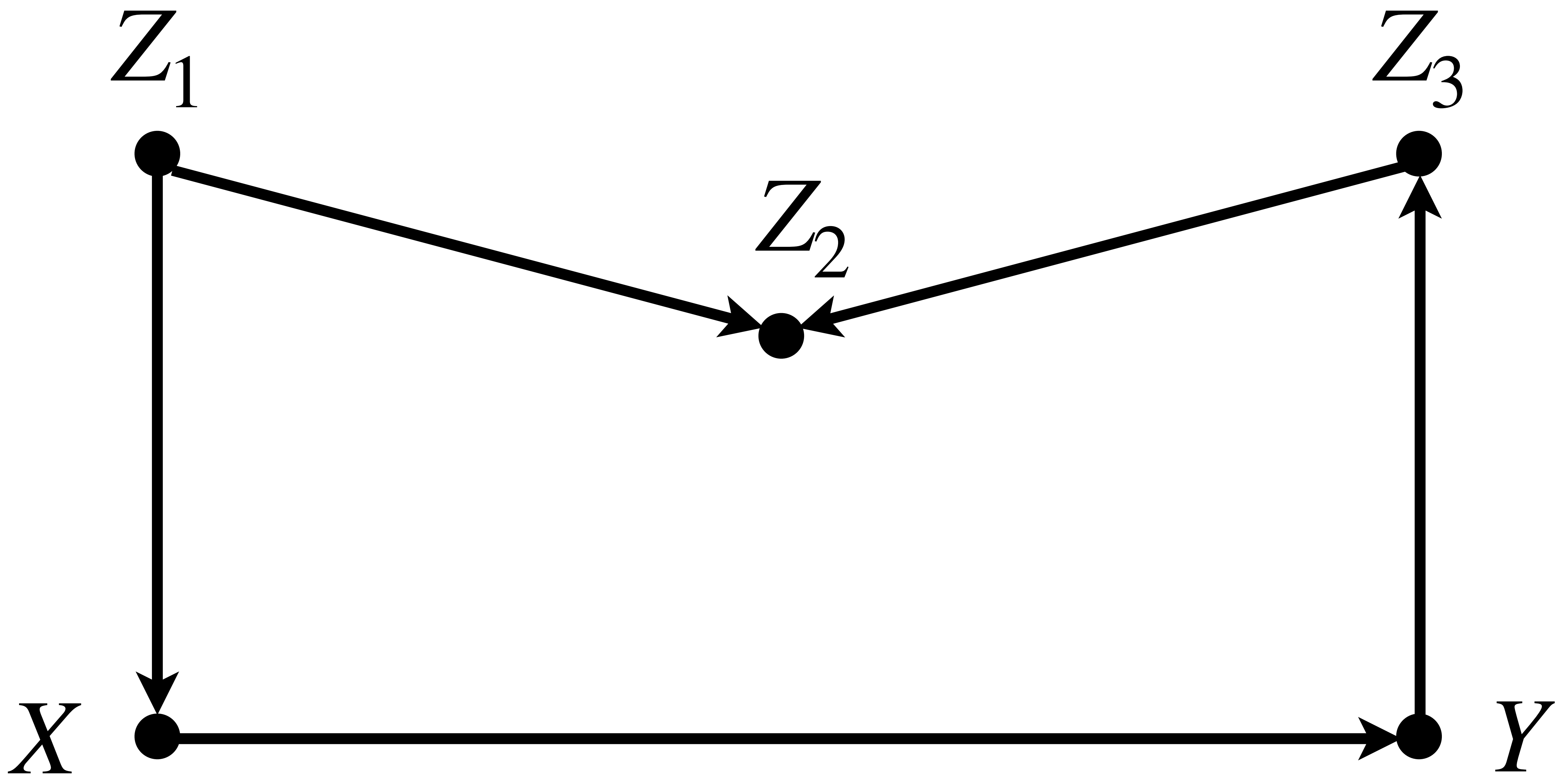
Let's *do*-Calculus!

$$p(Y \mid X) = \frac{p(Y, X)}{p(X)}$$



$$
\begin{aligned}
P(Y \mid do(X)) &\overset{Rule3}{=} P(Y \mid do(X), do(W), do(Z)) \overset{Rule2}{=} P(Y \mid X, do(W), do(Z)) \overset{Bayes}{=} \frac{P(Y, X \mid do(W), do(Z))}{P(X \mid do(W), do(Z))} \\
&\overset{Rule3}{=} \frac{P(Y, X \mid do(Z))}{P(X \mid do(W), do(Z))} \overset{Bayes}{=} \frac{\sum_W P(Y, X, W \mid do(Z))}{P(X \mid do(W), do(Z))} \overset{Bayes}{=} \frac{\sum_W P(Y, X \mid W, do(Z)) P(W \mid do(Z))}{P(X \mid do(W), do(Z))} \\
&\overset{Rule3}{=} \frac{\sum_W P(Y, X \mid W, do(Z)) P(W)}{P(X \mid do(W), do(Z))} \overset{Rule2}{=} \frac{\sum_W P(Y, X \mid W, Z) P(W)}{P(X \mid do(W), do(Z))} \overset{Rule3}{=} \frac{\sum_W P(Y, X \mid W, Z) P(W)}{P(X \mid do(Z))} \\
&\overset{Bayes}{=} \frac{\sum_W P(Y, X \mid W, Z) P(W)}{\sum_W P(X \mid W, do(Z)) P(W \mid do(Z))} \overset{Rule3}{=} \frac{\sum_W P(Y, X \mid W, Z) P(W)}{\sum_W P(X \mid W, do(Z)) P(W)} \\
&\overset{Rule2}{=} \frac{\sum_W P(Y, X \mid W, Z) P(W)}{\sum_W P(X \mid W, Z) P(W)}
\end{aligned}
$$

Let's *do*-Calculus?



$$P(Z_1 \mid do(X), Z_2)$$

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.

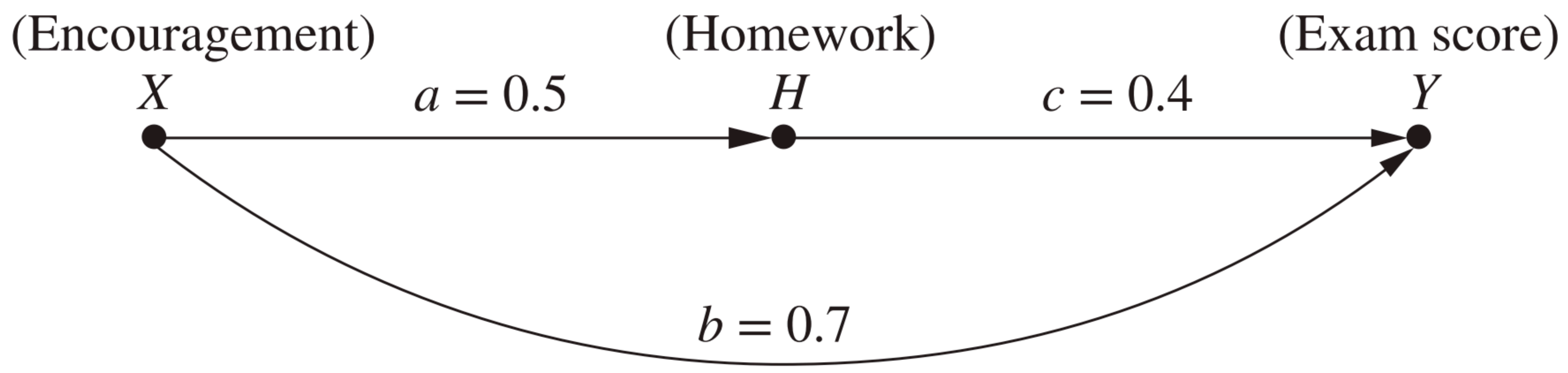
Step 1 **(*abduction*):** Update the probability $P(u)$ to obtain $P(u \mid e)$.

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



$$
\begin{aligned}
X &= U_X \\
H &= a \cdot X + U_H \\
Y &= b \cdot X + c \cdot H + U_Y
\end{aligned}
$$

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?

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

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

(PRIMER, CH4)

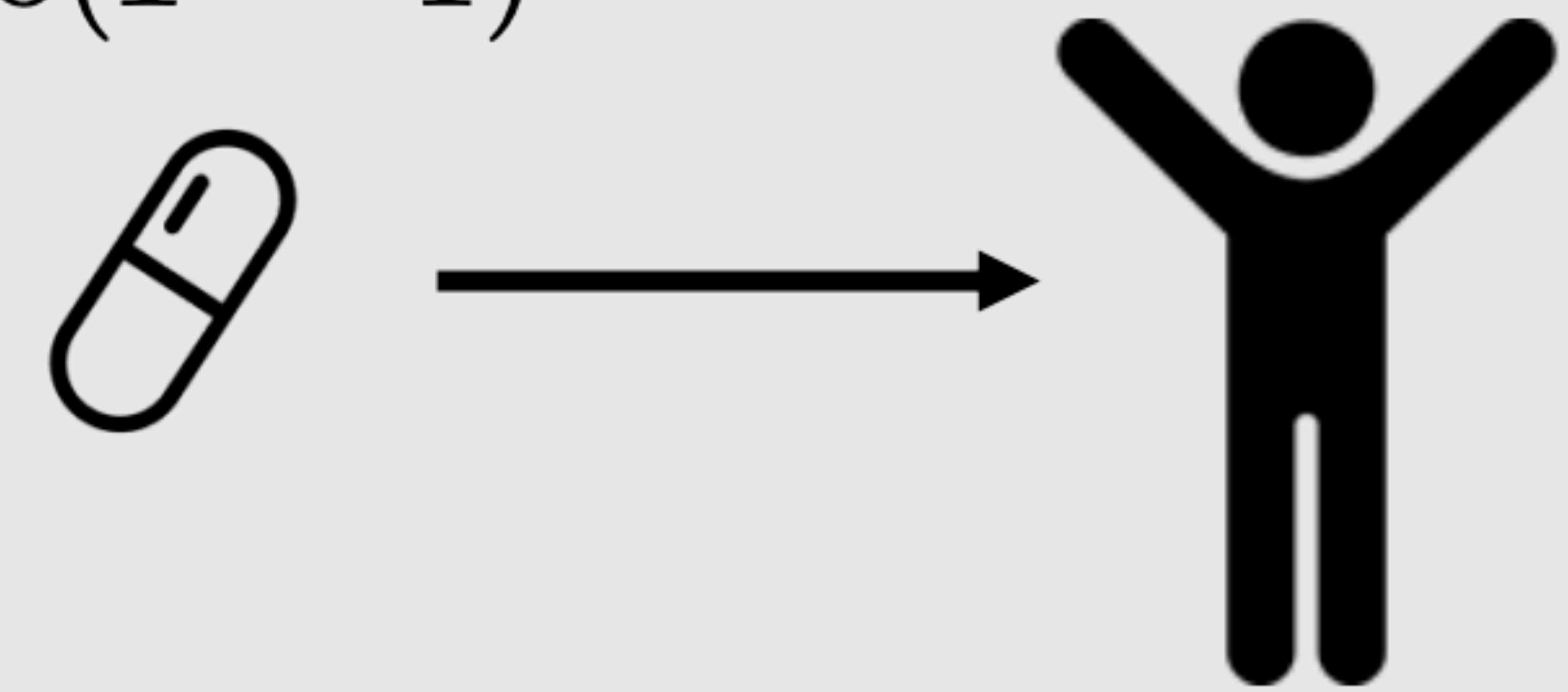
Potential Outcome Models

Potential Outcomes: Notations

value variable

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

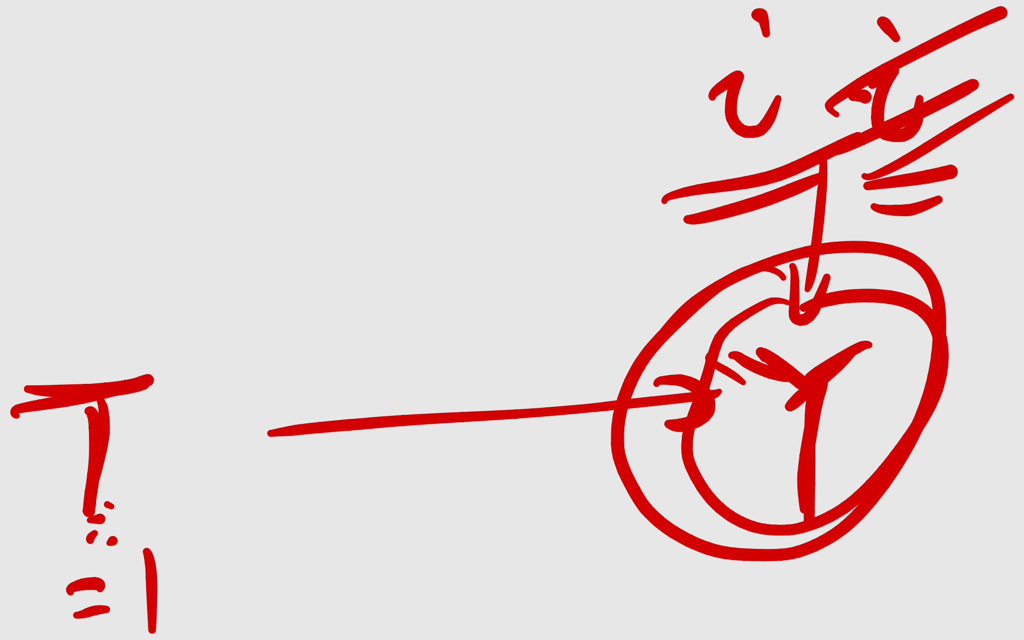
$do(T = 1)$



Y(1).

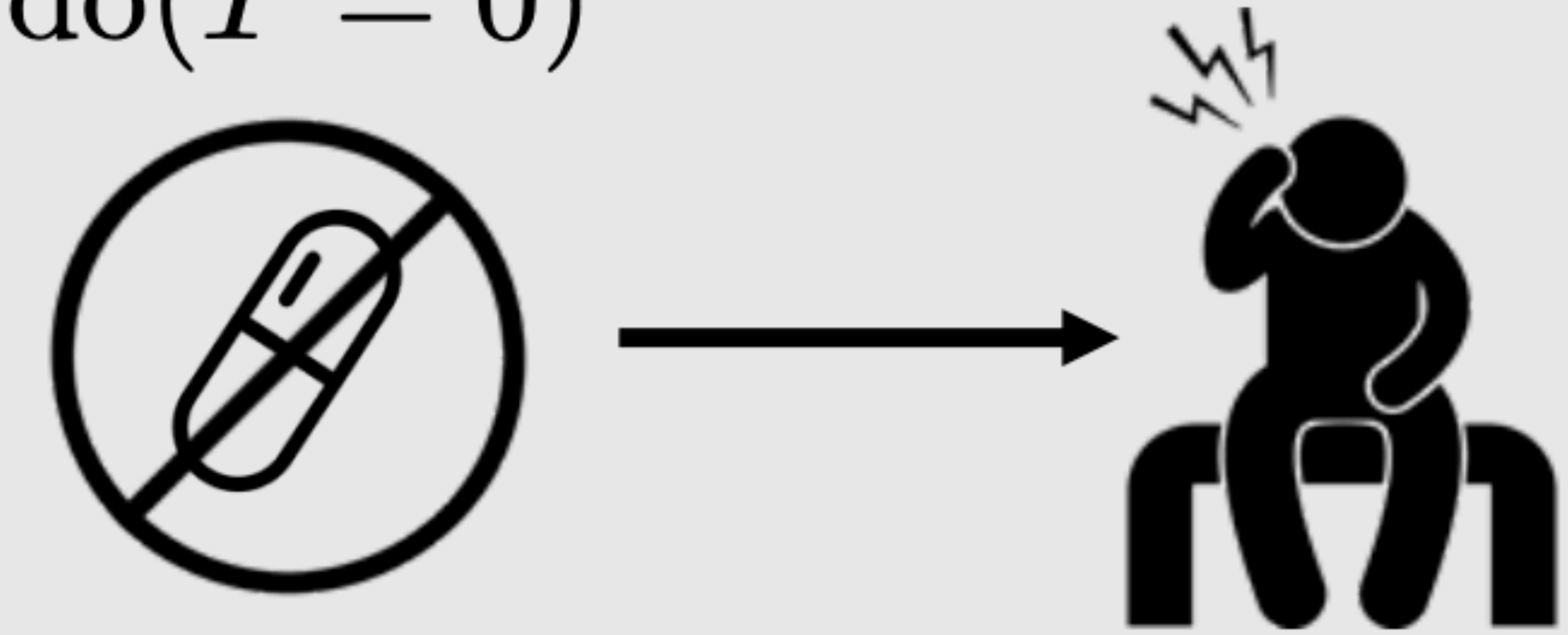
variable

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



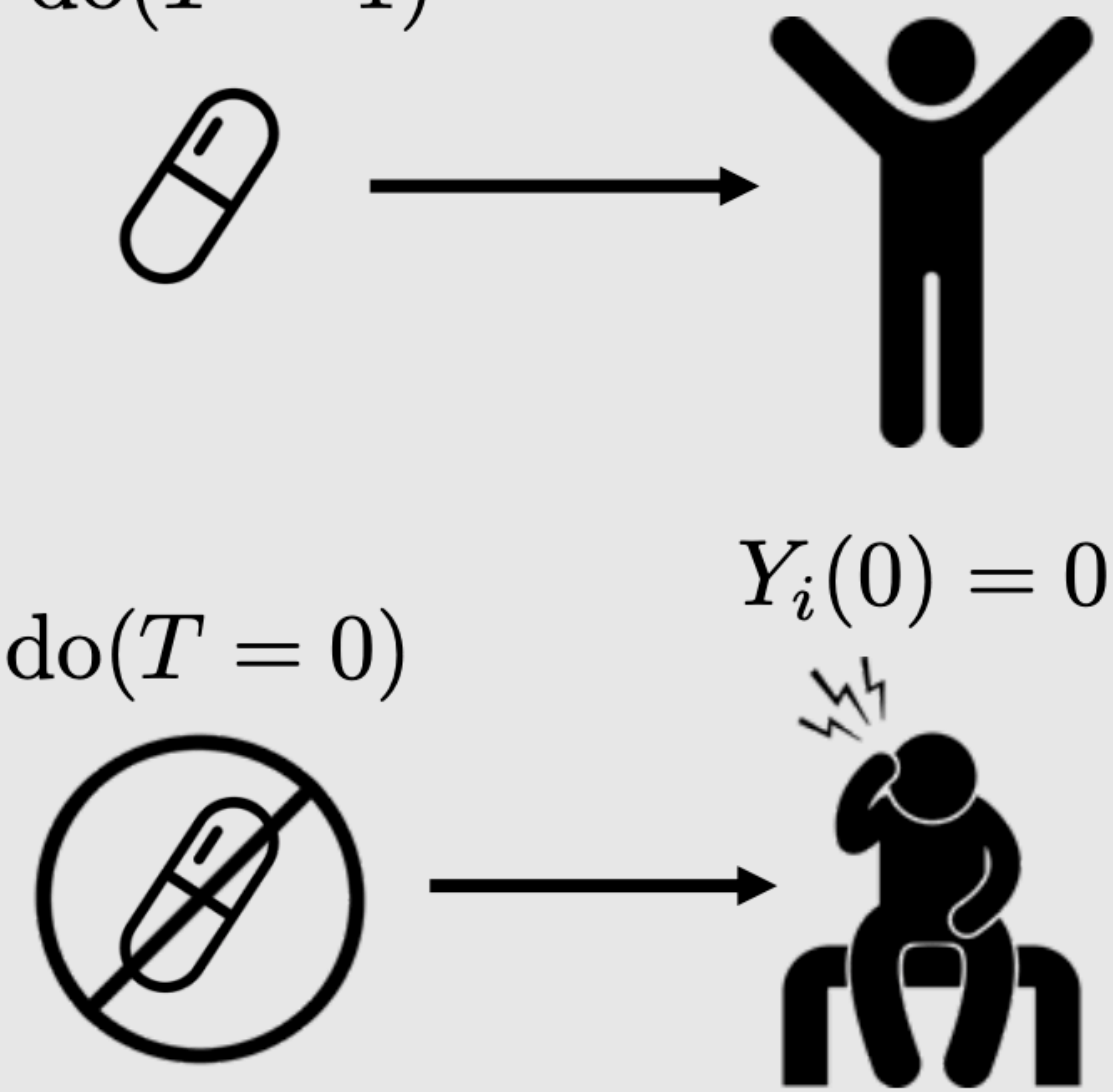
$$Y_i|_{do(T=0)} \triangleq Y_i(0)$$

$do(T = 0)$



(Brady Neal, CH2)

Potential Outcomes: Notations



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

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

$Y(1) - Y(0)$

(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

(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	?	?

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

(Brady Neal, CH2)

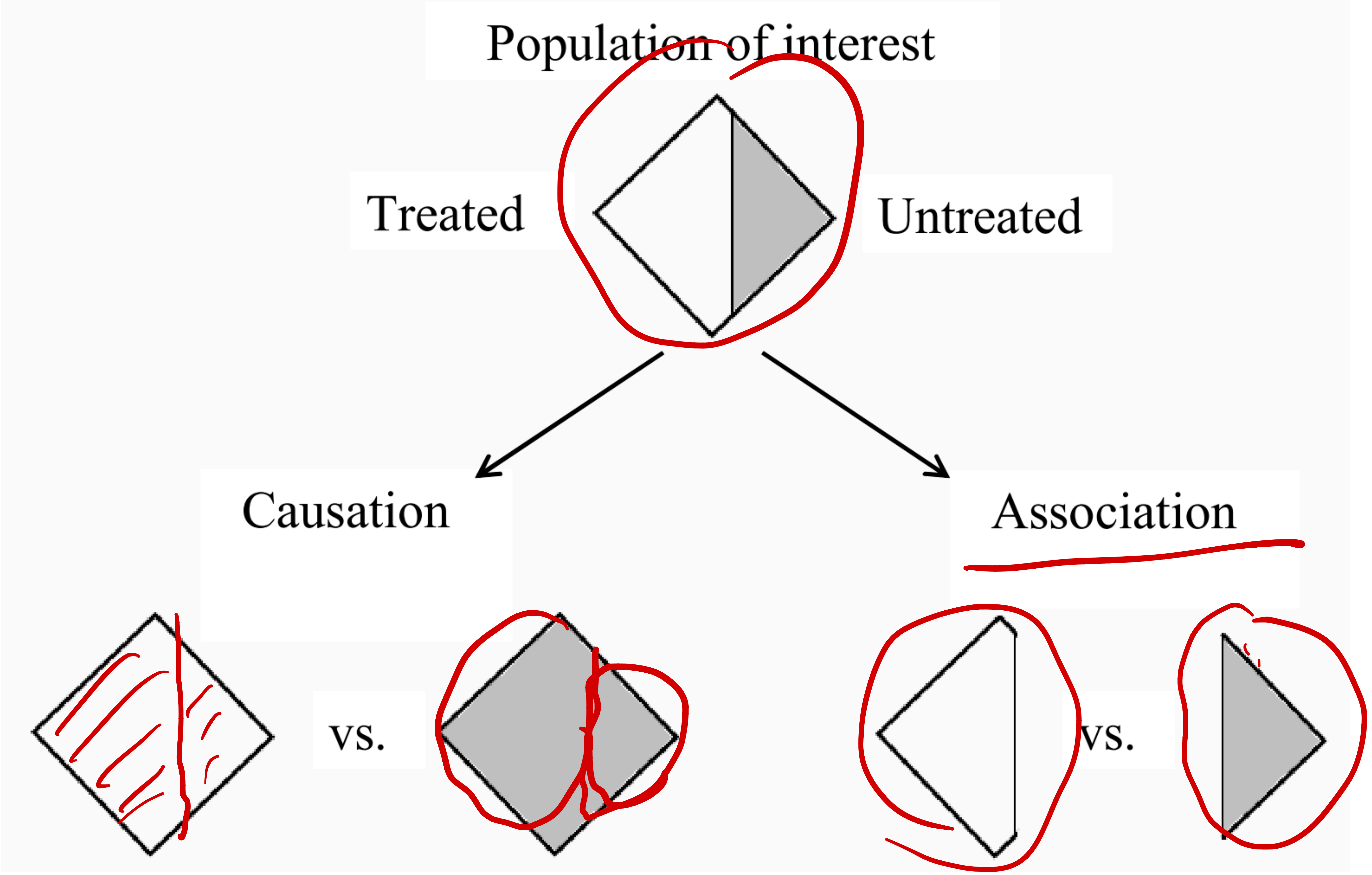
Average Treatment Effect (ATE)

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)

Causation versus Association



$\mathbb{E}[Y_{T=1}]$ $\boxed{\mathbb{E}[Y_{T=0}]}$ $\mathbb{E}[Y \mid T=1]$ $\mathbb{E}[Y \mid T=0]$

(WHATIF, CH1)

Consistency Rule

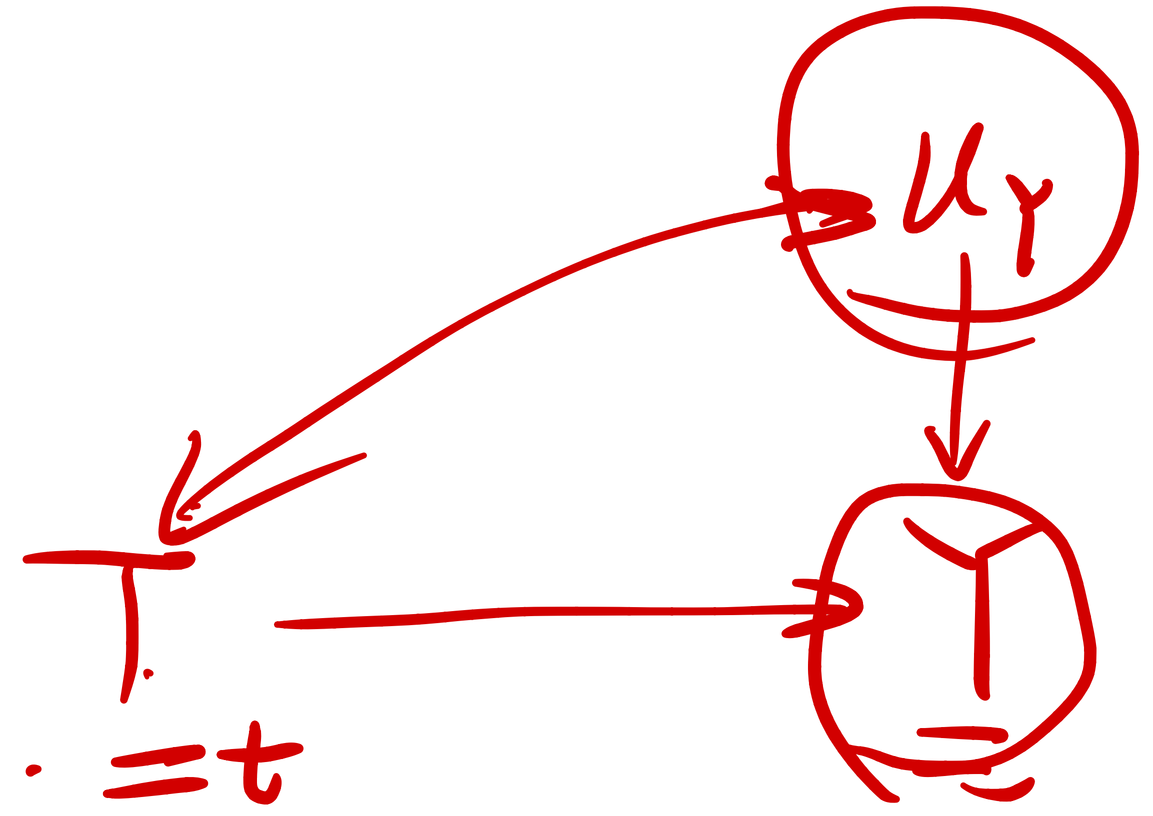
i	T	Y	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

$Y=y$
$T=t$

$T=t$ $Y=y$

$$P(Y(t) \mid T = t) = P(Y \mid T = t)$$

Ignorability/Exchangeability



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

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

$\Downarrow$

U_Y

$U_Y \perp\!\!\!\perp T$

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

(Brady Neal, CH2)

Conditional Ignorability/Exchangeability

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

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

(Brady Neal, CH2)

Standardization = back-door

$$p(Y \mid do(A)) = \sum_L p(Y \mid A, L)\, p(L)$$

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

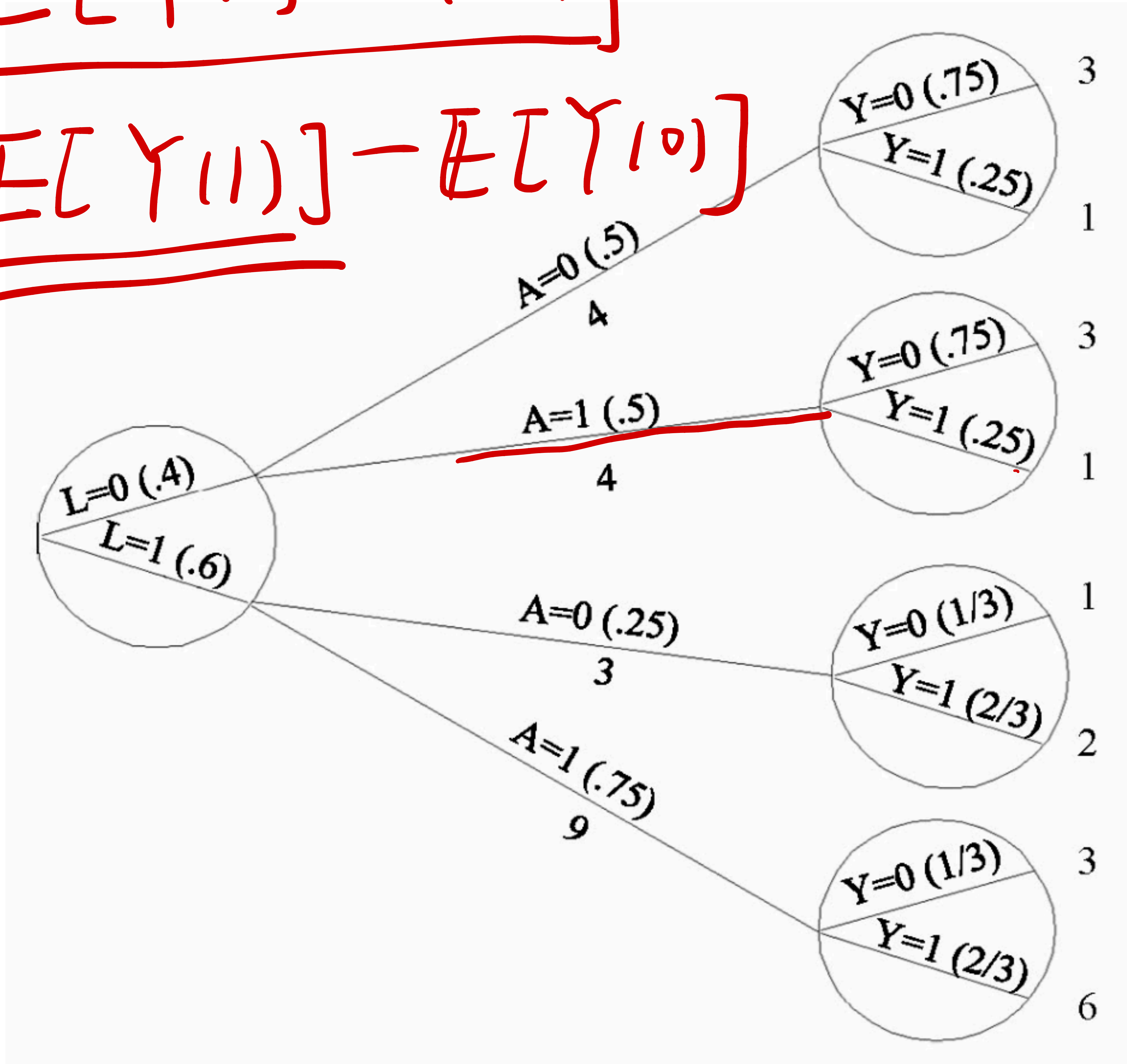
(WHATIF, CH2)

Inverse Probability Weighting

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

$\underline{E[Y(1) - Y(0)]}$

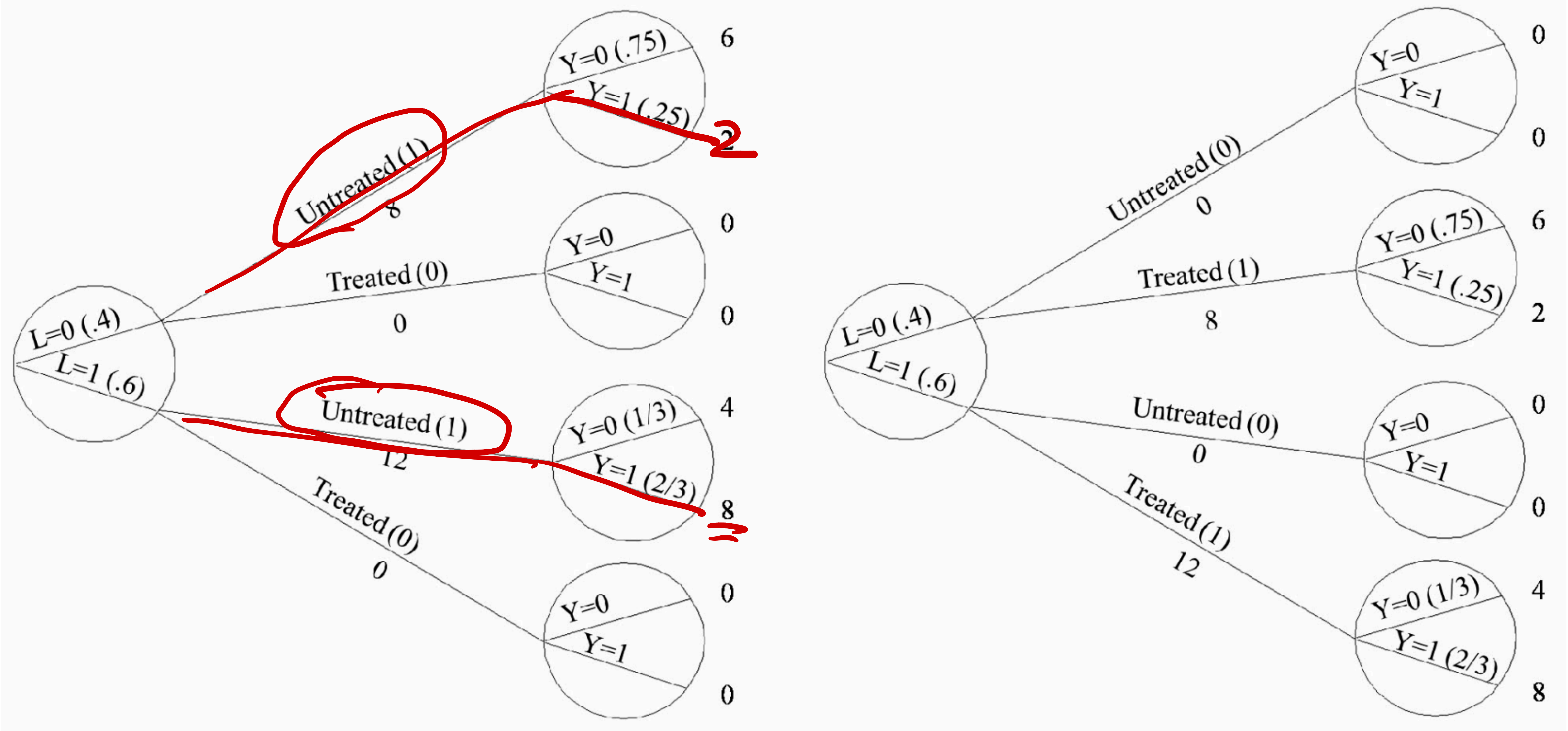
$= \underline{E[Y(1)]} - E[Y(0)]$

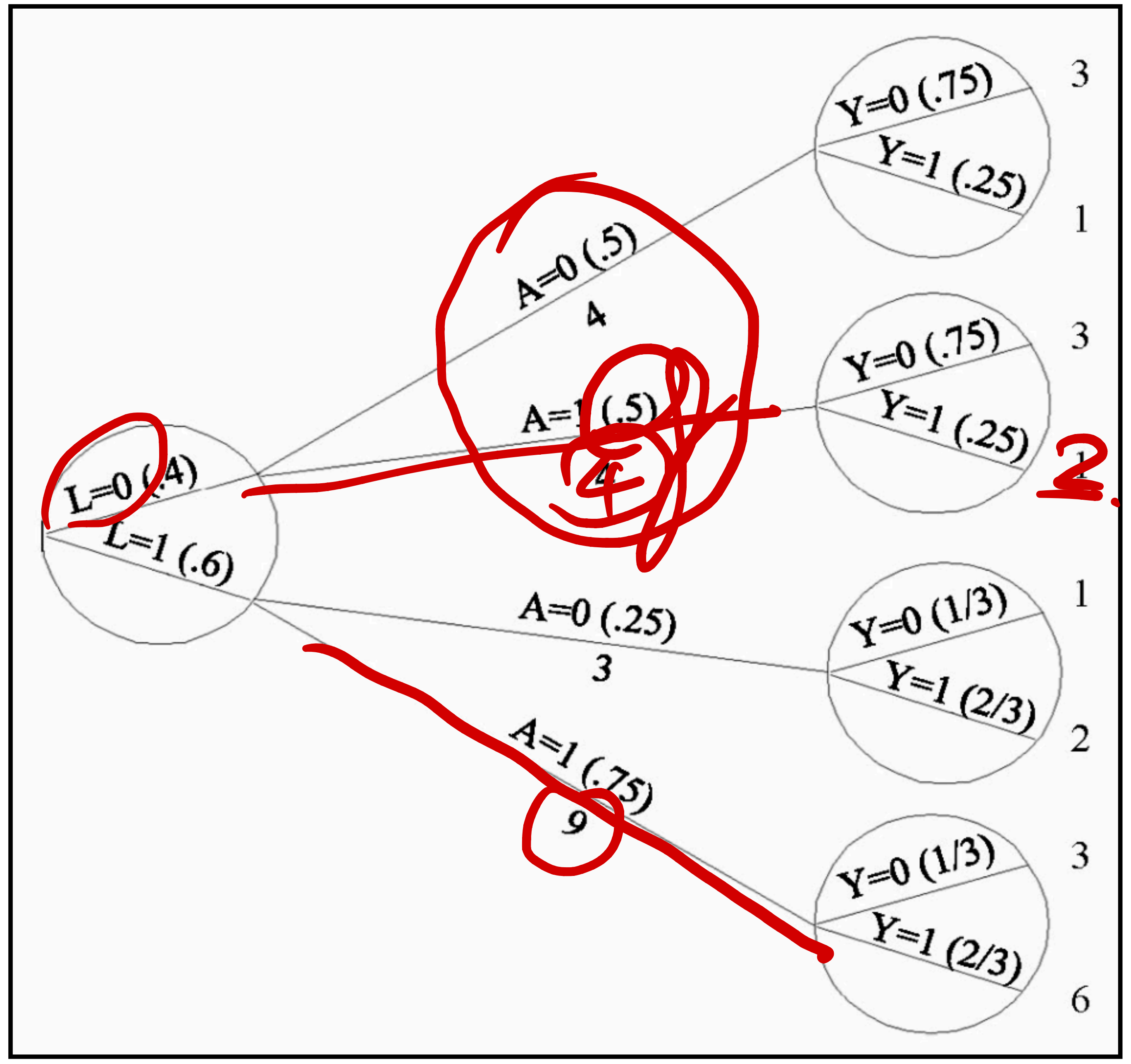


(WHATIF, CH2)

Inverse Probability Weighting

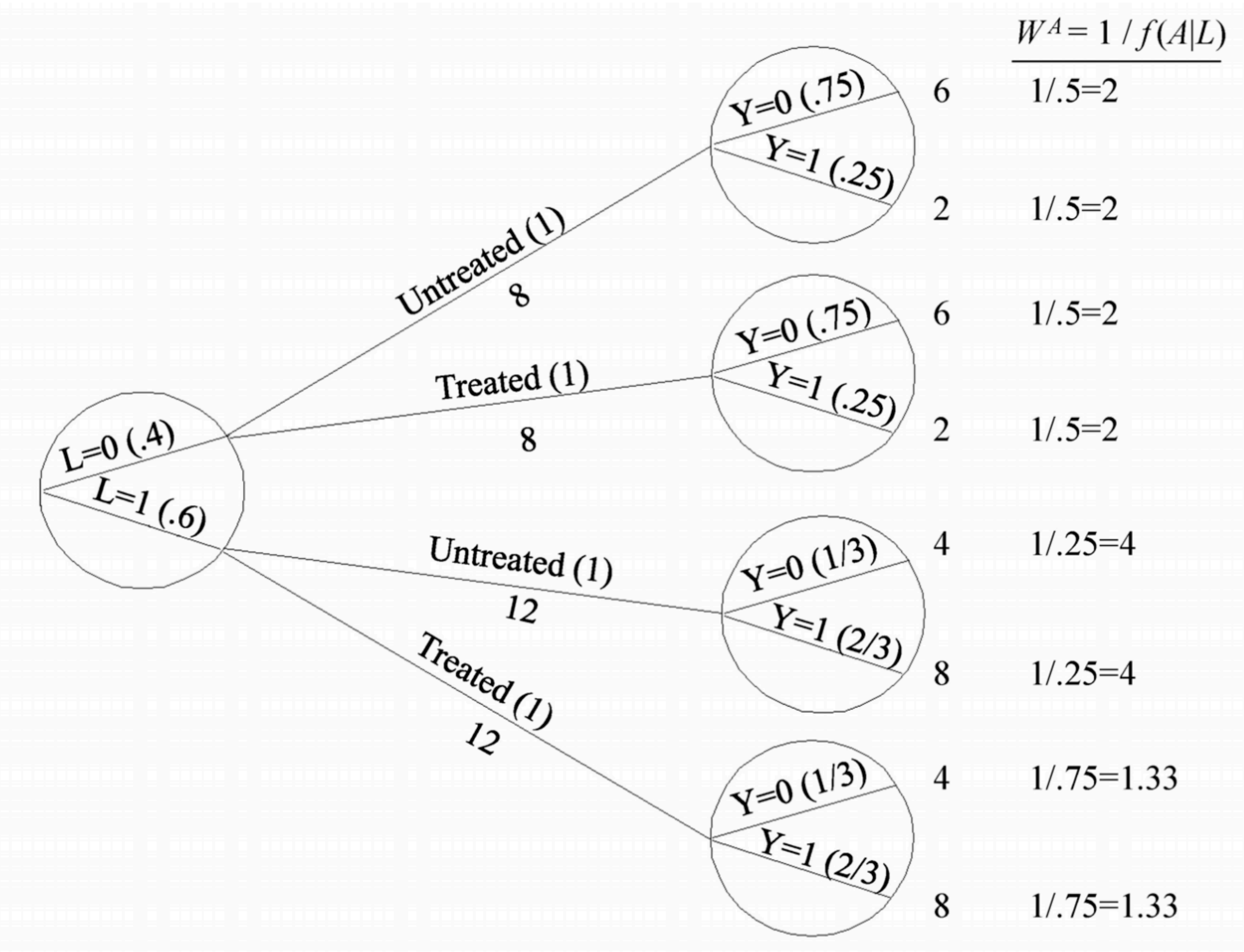
$E[Y_{A=1}=1]$

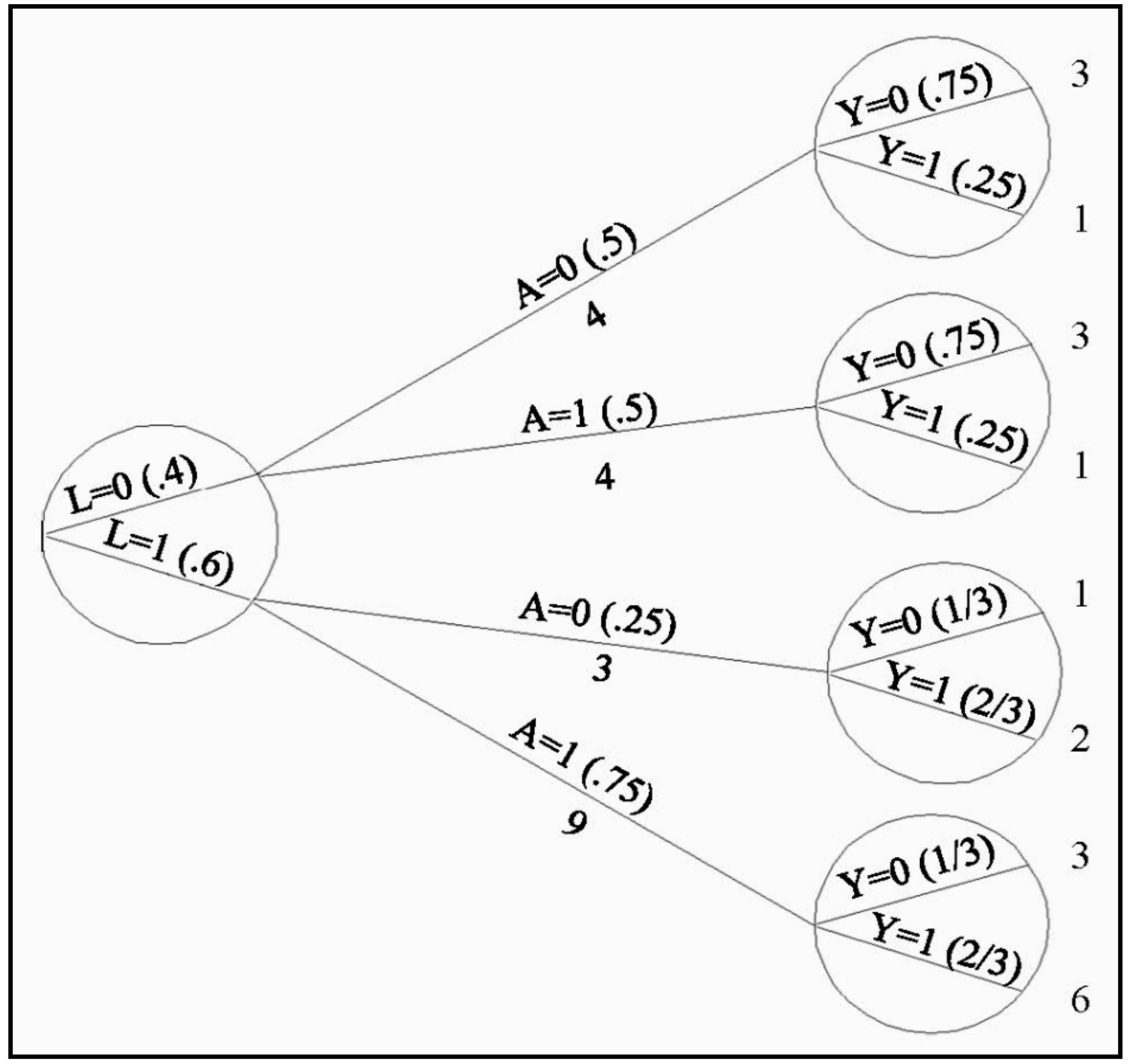




(WHATIF, CH2)

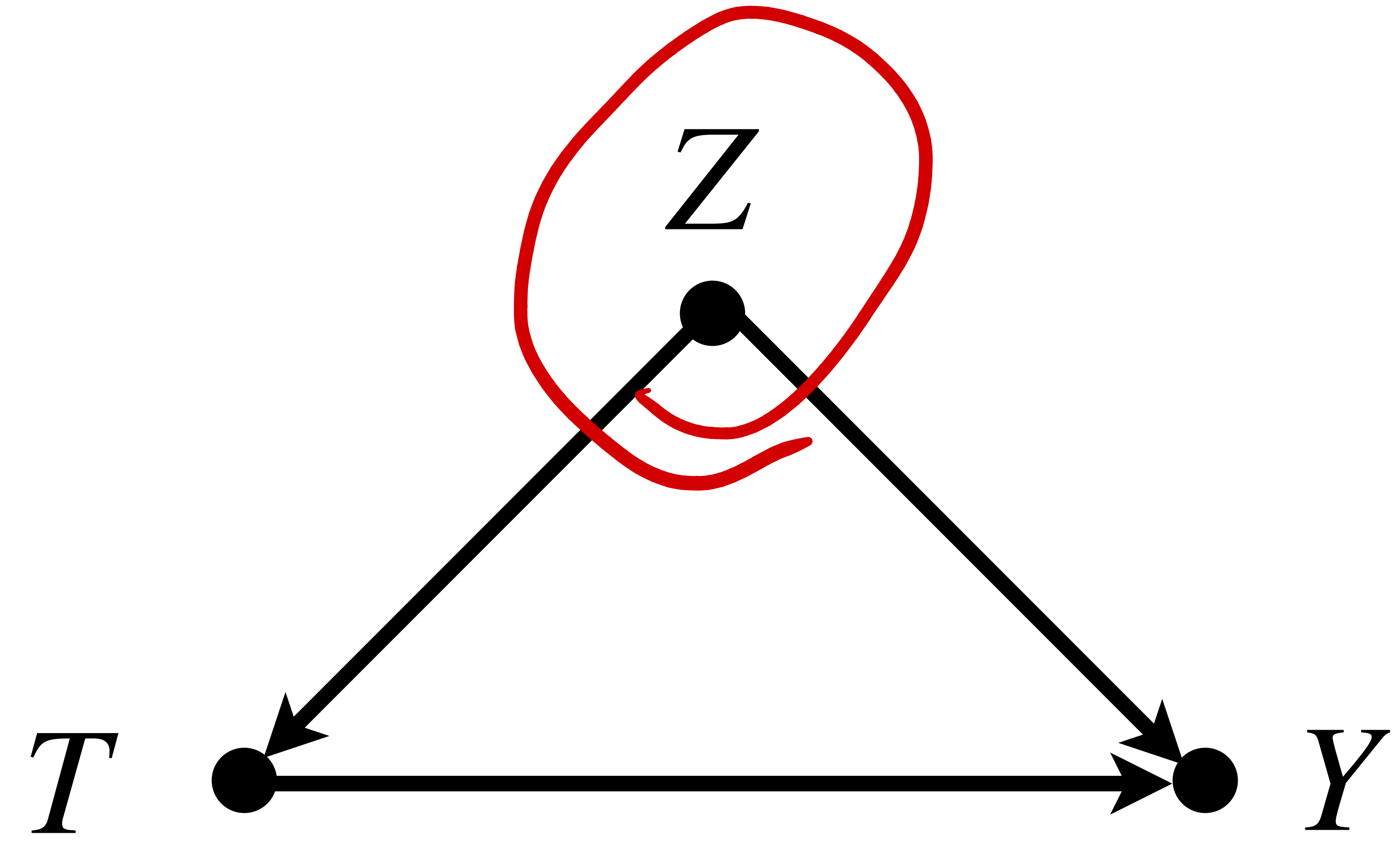
Inverse Probability Weighting





(WHATIF, CH2)

Propensity Score Theorem



$$Y(t) \perp\!\!\!\perp T \mid Z \implies Y(t) \perp\!\!\!\perp T \mid e(Z) \quad = 1$$

$$e(Z) \triangleq P(T = 1 \mid Z)$$

$$T \perp\!\!\!\perp Z \mid e(Z)$$

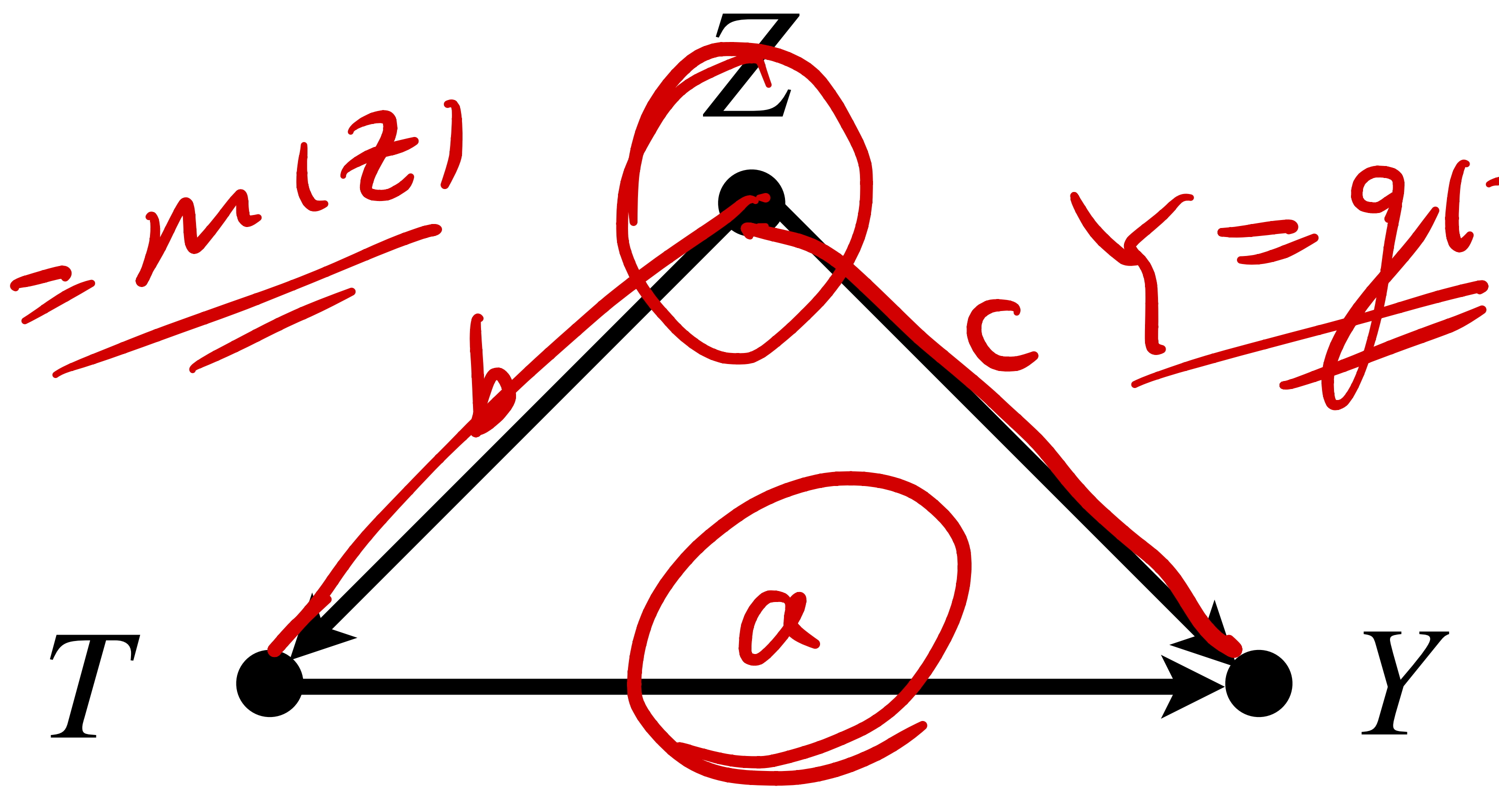
(Brady Neal, CH6)

Double Machine Learning

G_α

$T = m(z)$ $\quad$ $Y = g(z)$

$Y = a \cdot T + \beta \cdot Z$



Stage 1:

- Fit a model to predict Y from Z to get the predicted $\hat{Y}$
- Fit a model to predict T from Z to get the predicted $\hat{T}$

Stage 2:

- Partial out Z by fitting a model to predict $Y - \hat{Y}$ from $T - \hat{T}$

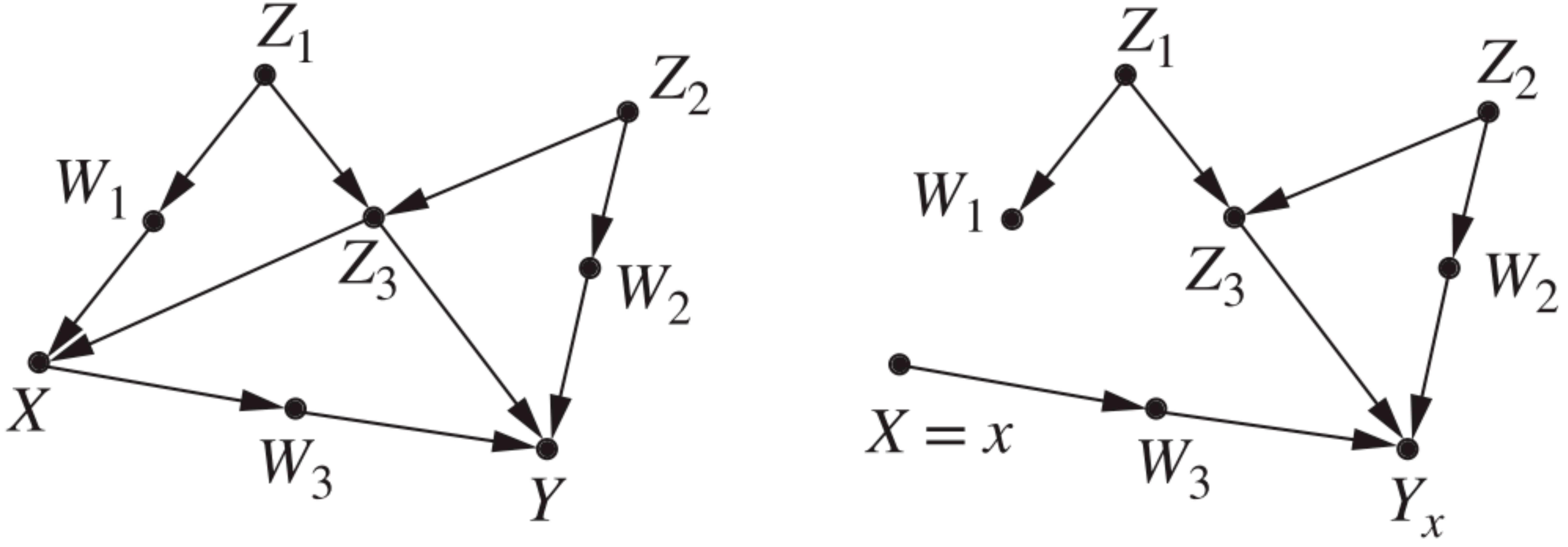
(Chernozhukov et al. 2018)

The Great Power of Graphs

“Logic void of representation is metaphysics.”

–Judea Pearl

Visualizing Counterfactuals

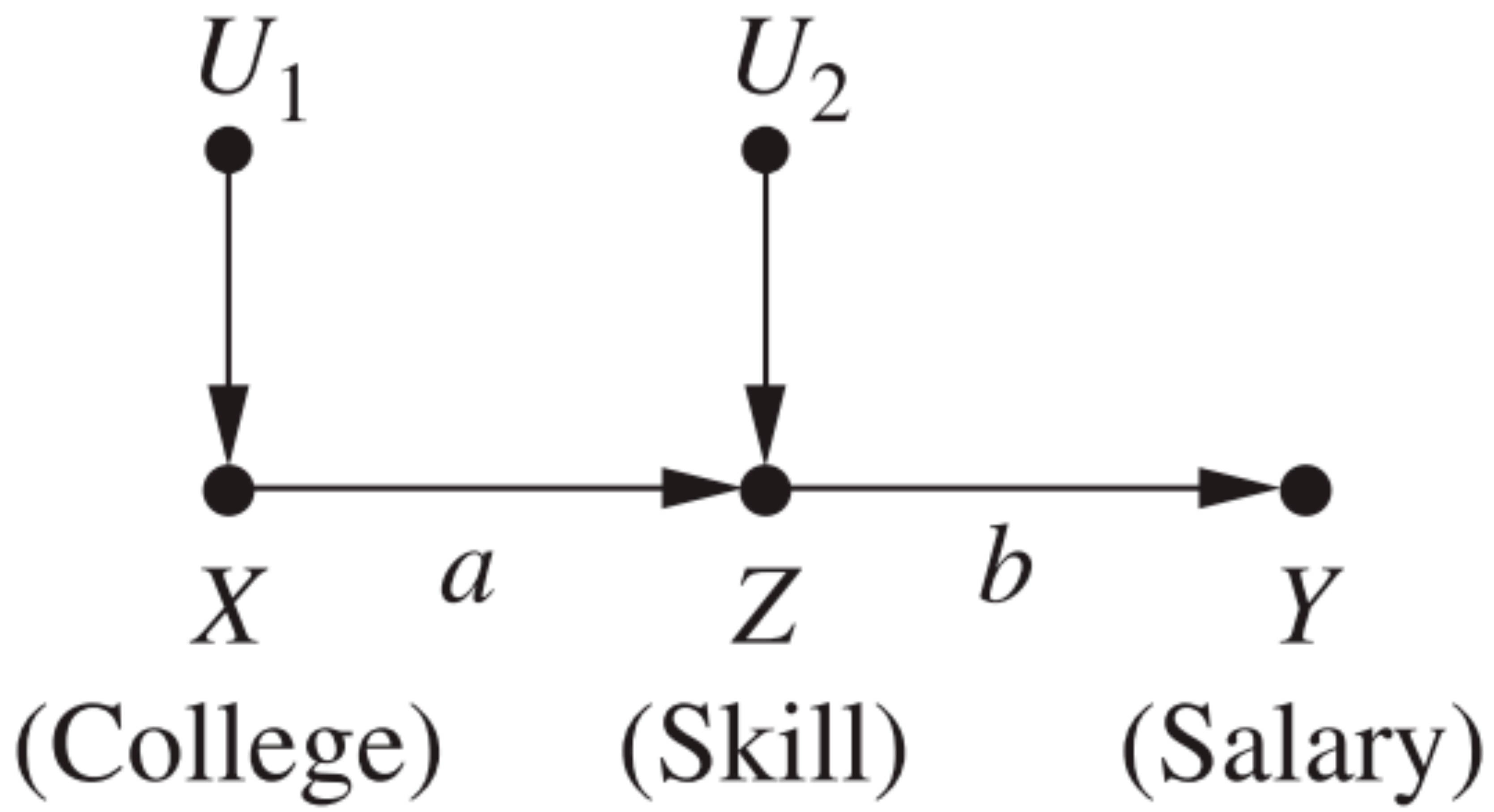


(a)

(b)

(PRIMER, CH4)

Example

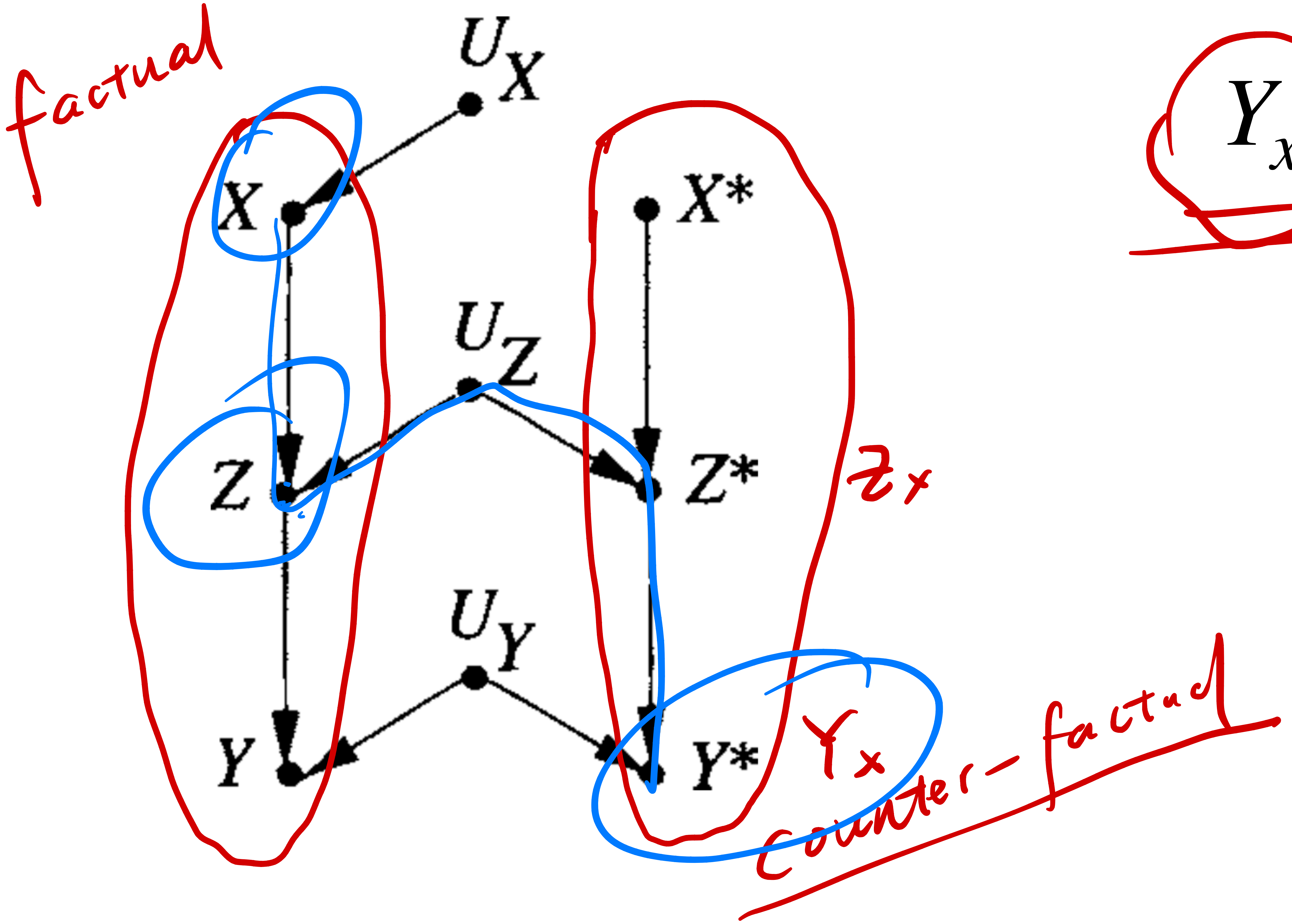


$$\mathbb{E}[Y_{X=1} \,|\, Z = 1]$$

(PRIMER, CH4)

The Twin Network Method

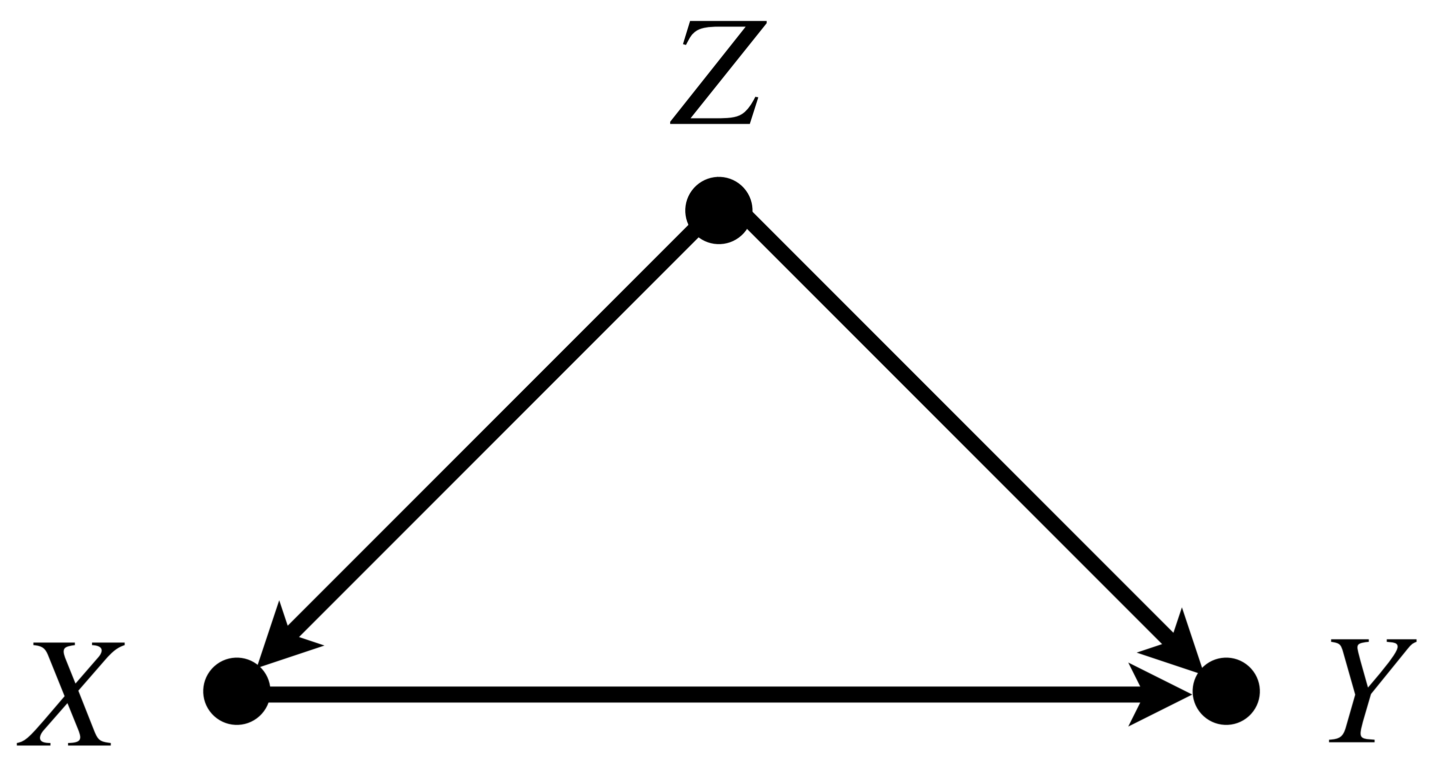
$$X \longrightarrow Z \longrightarrow Y$$



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

(CAUSALITY, CH7)

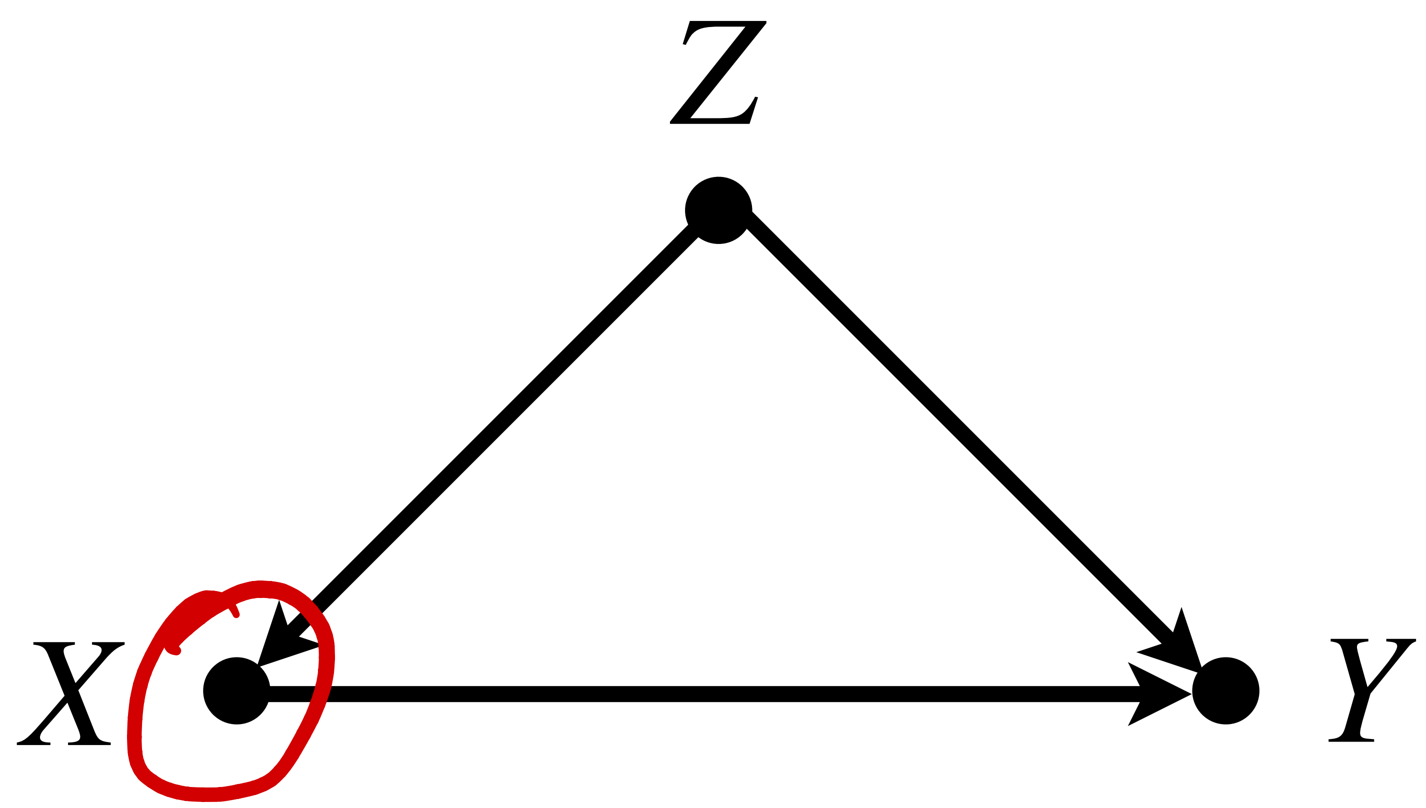
Example

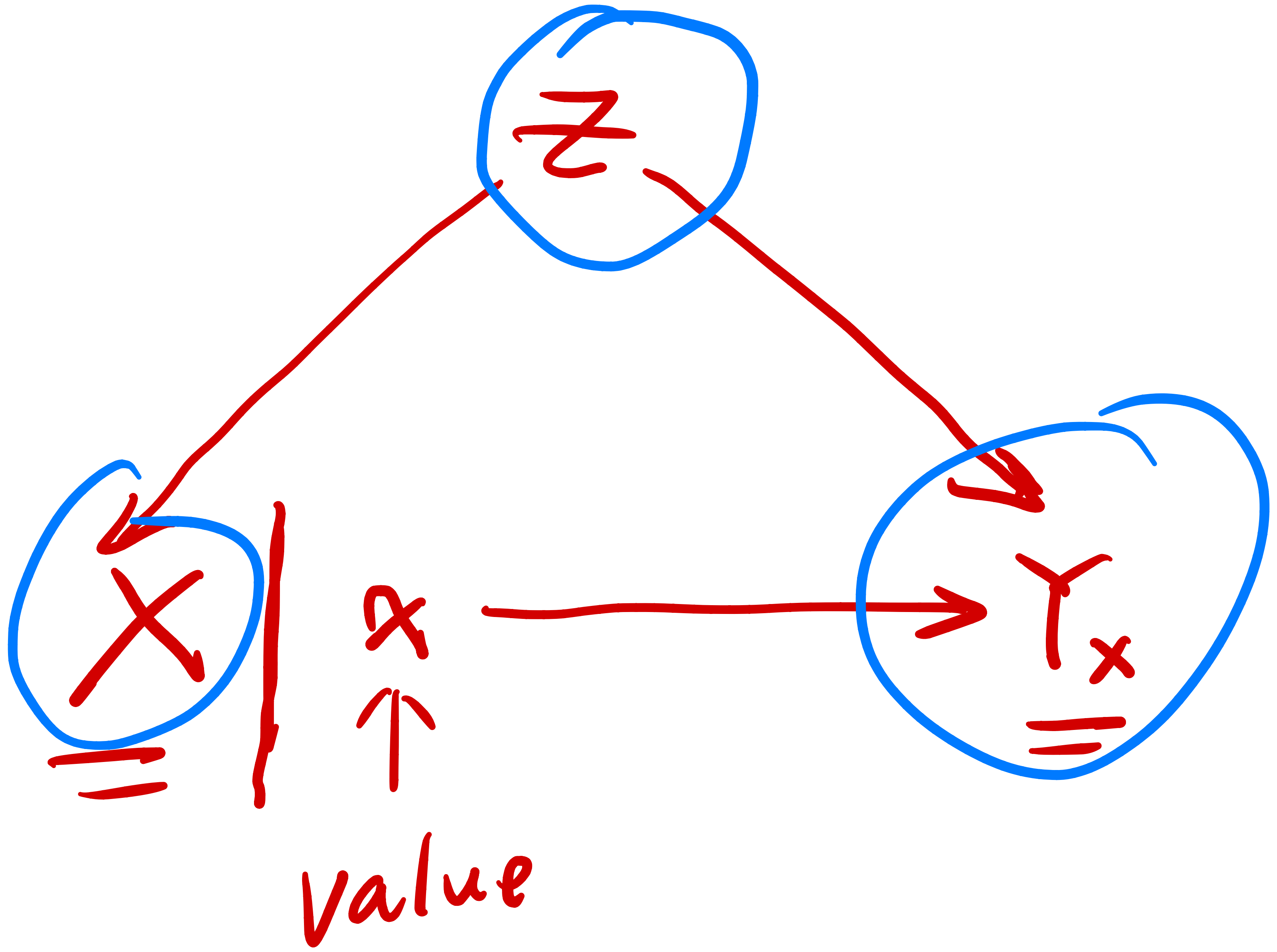


$$(Y(0), Y(1)) \perp\!\!\!\perp X$$

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

Single-World Intervention Graph (SWIG)





$Y(x)$

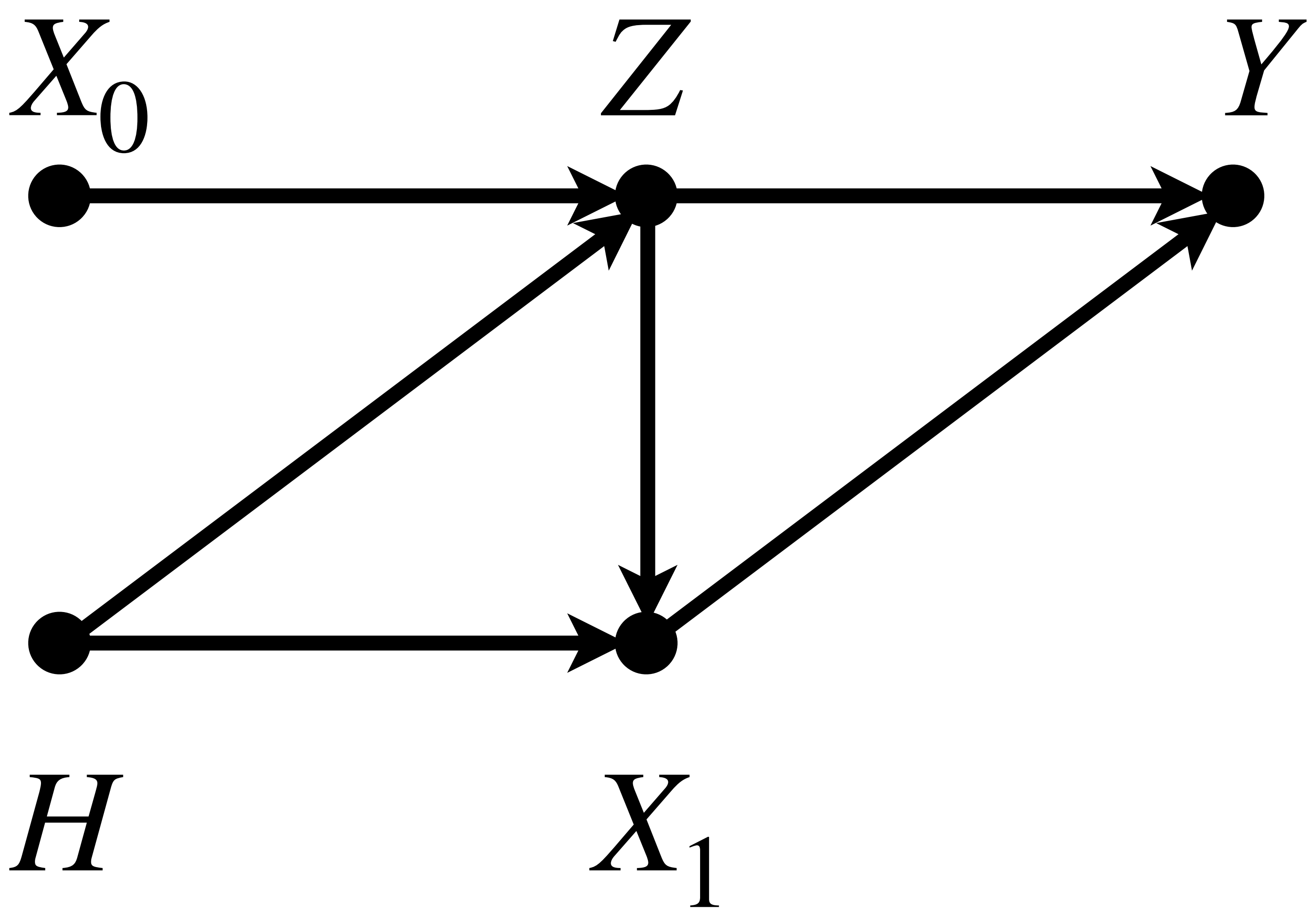
$$(Y(0), Y(1)) \perp\!\!\!\perp X$$

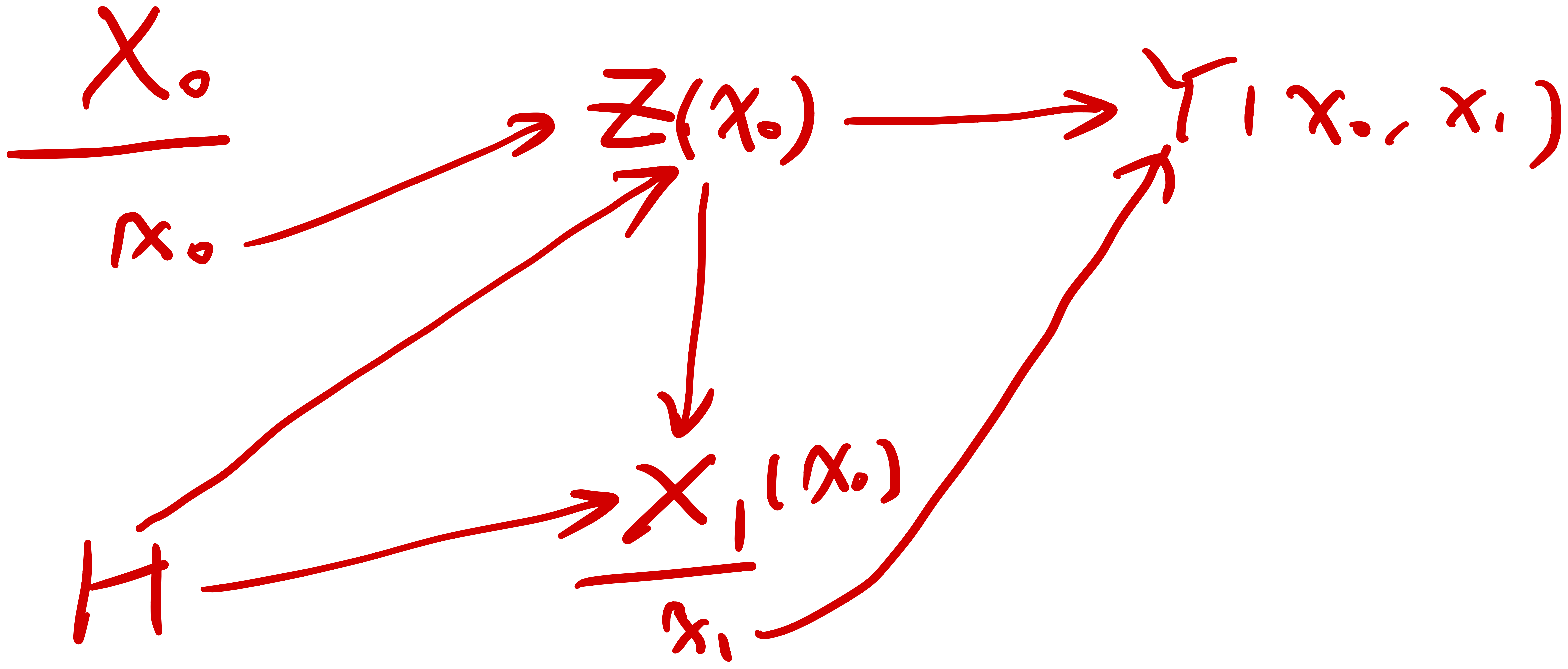
$$(Y(0), Y(1)) \perp\!\!\!\perp X \mid Z$$

$Y(x) \perp\!\!\!\perp X \mid Z$

(WHATIF, CH7)

Example Using SWIG

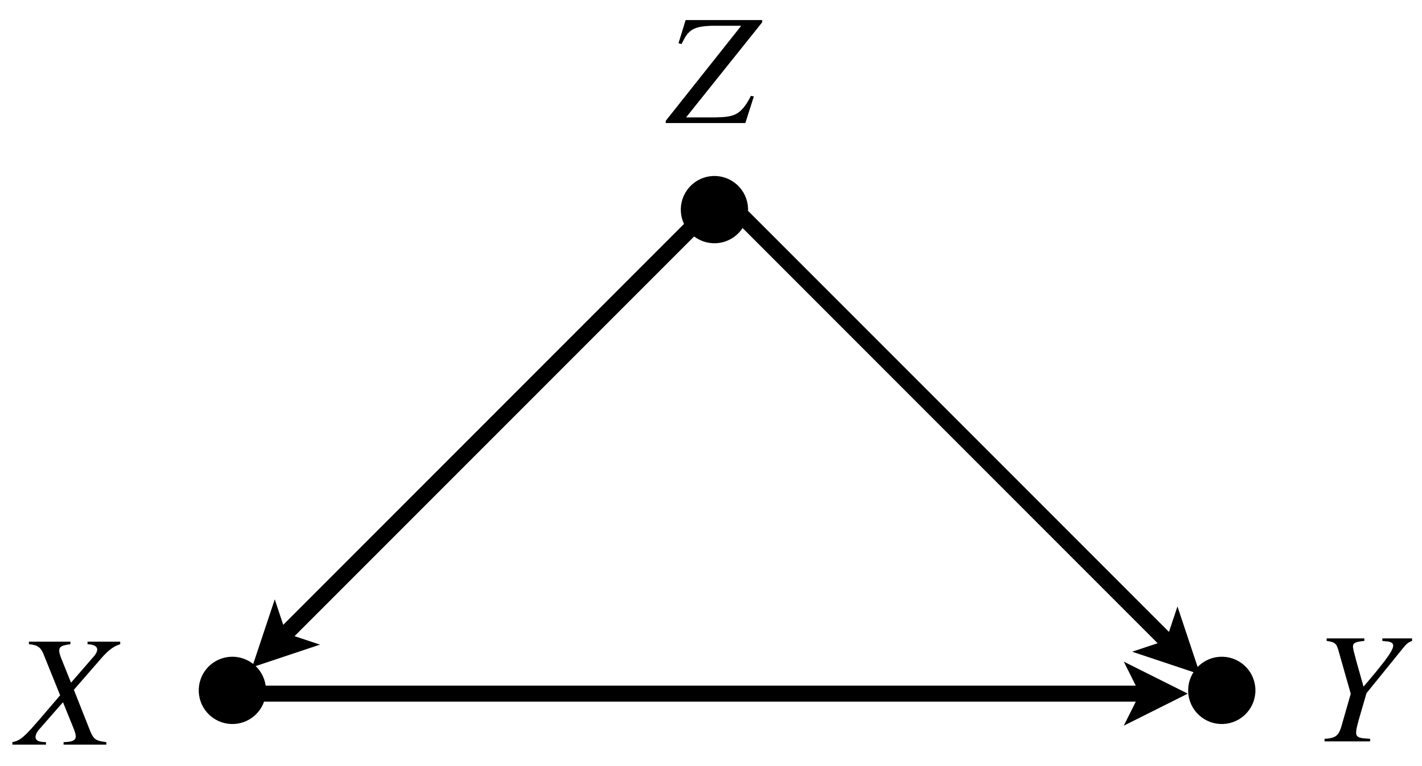




$$Y(x_0, x_1) \perp\!\!\!\perp X_1(x_0) \mid Z(x_0), X_0 = x_0$$

$$Y(x_0, x_1) \perp\!\!\!\perp X_1 \mid Z, X_0 = x_0$$

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**.

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**: $X = x \Longrightarrow Y_x = Y$

(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

$Y = f(pa_y, u)$

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

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

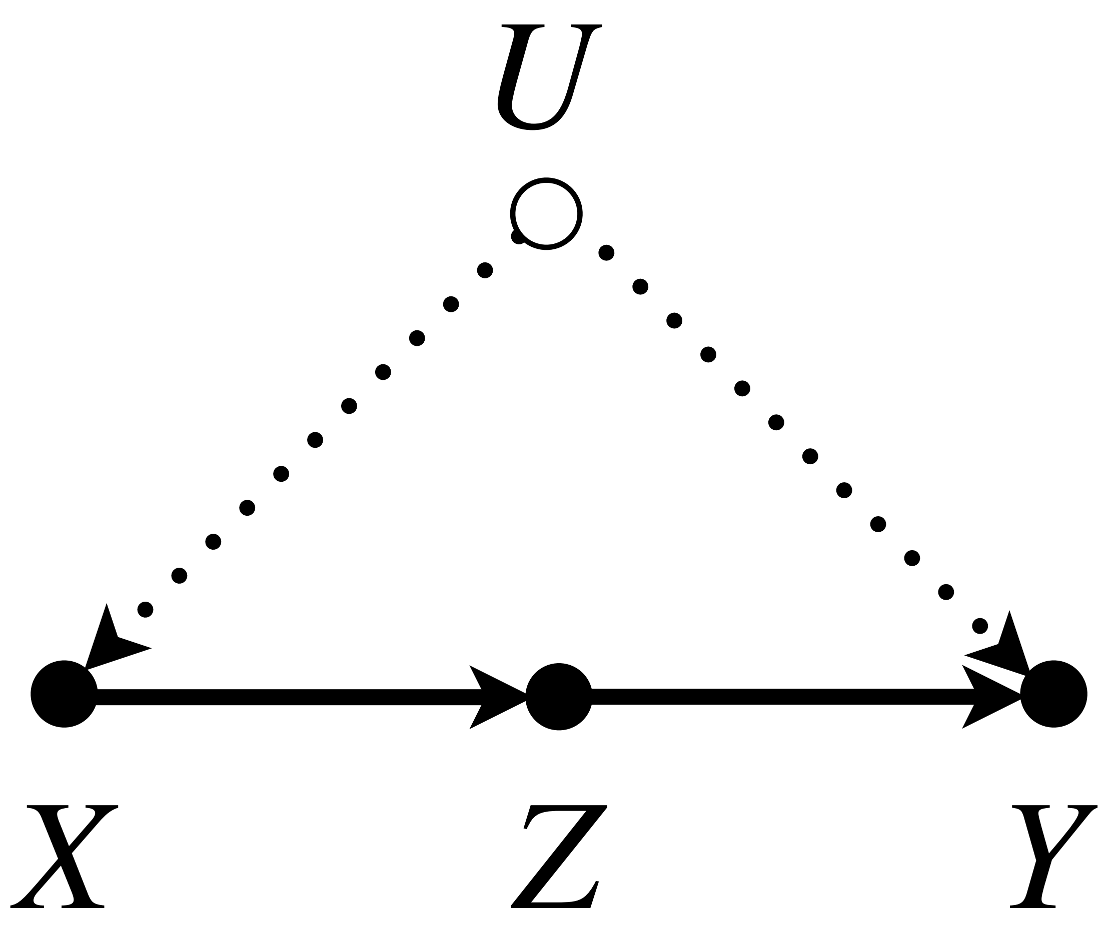
$$Y_{pa_Y} \perp\!\!\!\perp \{Z_{1_{pa_{Z_1}}}, \ldots, Z_{k_{pa_{Z_k}}}\}$$

$$\Downarrow \qquad \Downarrow \qquad \Downarrow$$

$$U_Y \perp\!\!\!\perp U_{Z_1} \perp\!\!\!\perp U_{Z_k}$$

(CAUSALITY, CH7)

Example



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

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

Axiomatic Characterization

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

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

$x = \emptyset$

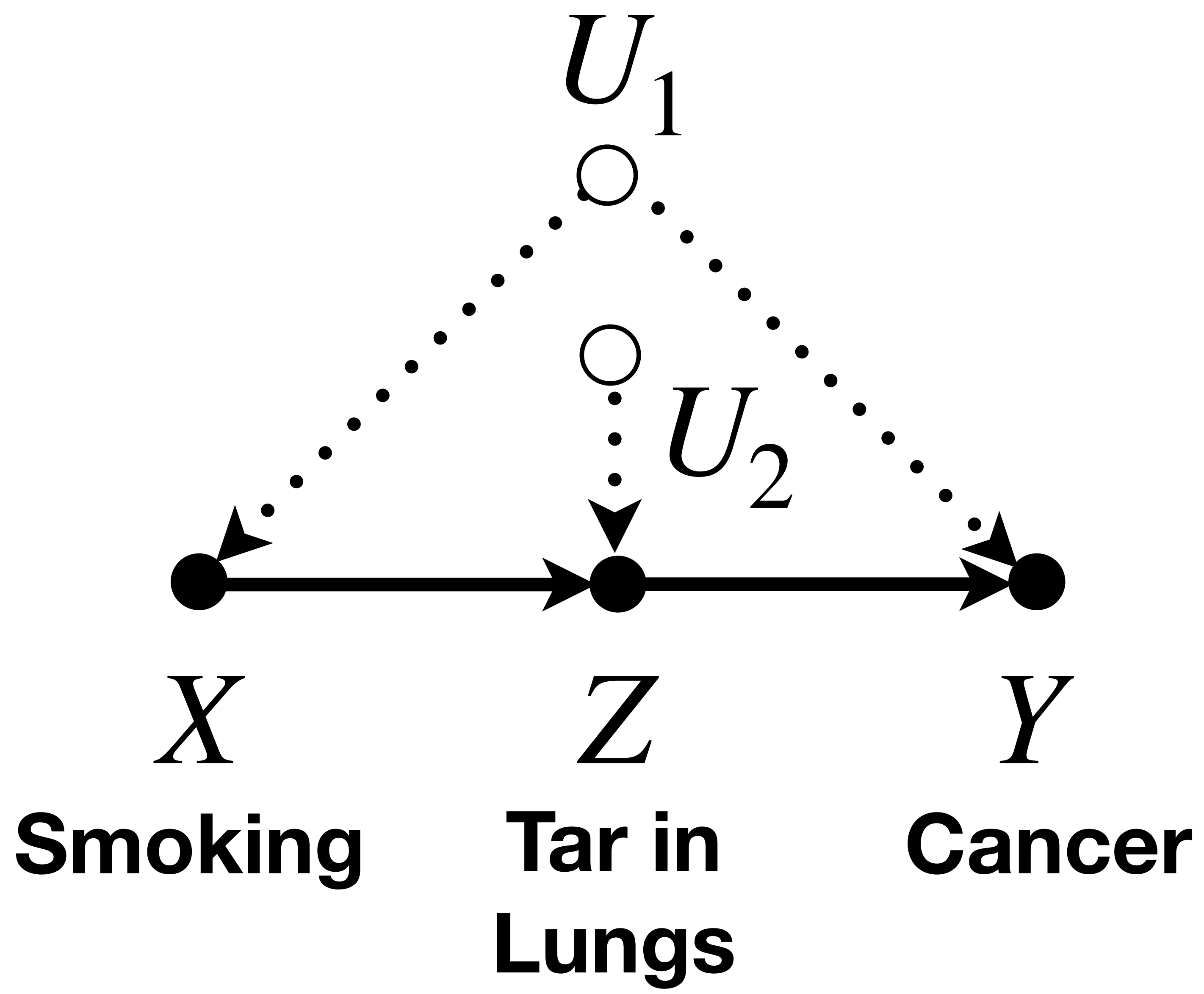
$$\frac{Y(x)}{Y_i(x)}$$

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

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

(CAUSALITY, CH7)

Example from Counterfactual Logic



(CAUSALITY, CH7)

Example from Counterfactual Logic

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

Composition:

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

Effectiveness:

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

Task 1

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

(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:

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

Effectiveness:

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

Task 2

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

(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:

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

Effectiveness:

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

Task 3

Compute $P(Y_x = y)$
(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**.

(CAUSALITY, CH7)

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
Structural Causal Model
Composition
Effectiveness
Reversibility

Thank you.

Q & A

https://causallu.com