因果推理读书会

Causal AI

图模型和因果推理基础 - Part II

报告人：付鑫玉，匹兹堡大学

研究方向：人机协作

Xinyu.fu@pitt.edu

Thanks Sanna Tyrv¨ainen for providing part of the slides

集智斑图 Swarma Pattern

Agenda

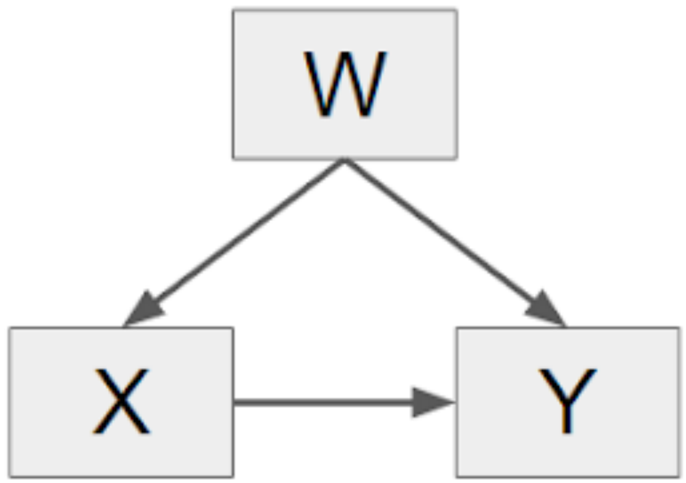
- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - Backdoor Adjustment with Examples
    - Frontdoor Adjustment with Examples
    - Instrumental Variable Analysis
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

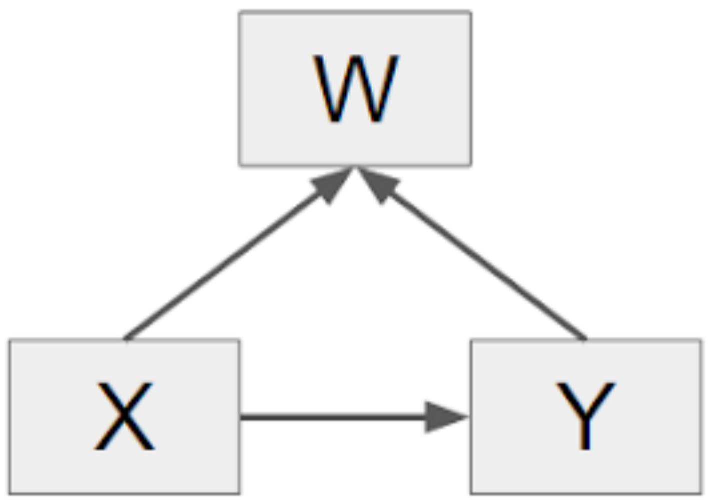
Agenda

- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - Backdoor Adjustment with Examples
    - Frontdoor Adjustment with Examples
    - Instrumental Variable Analysis
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

Identification

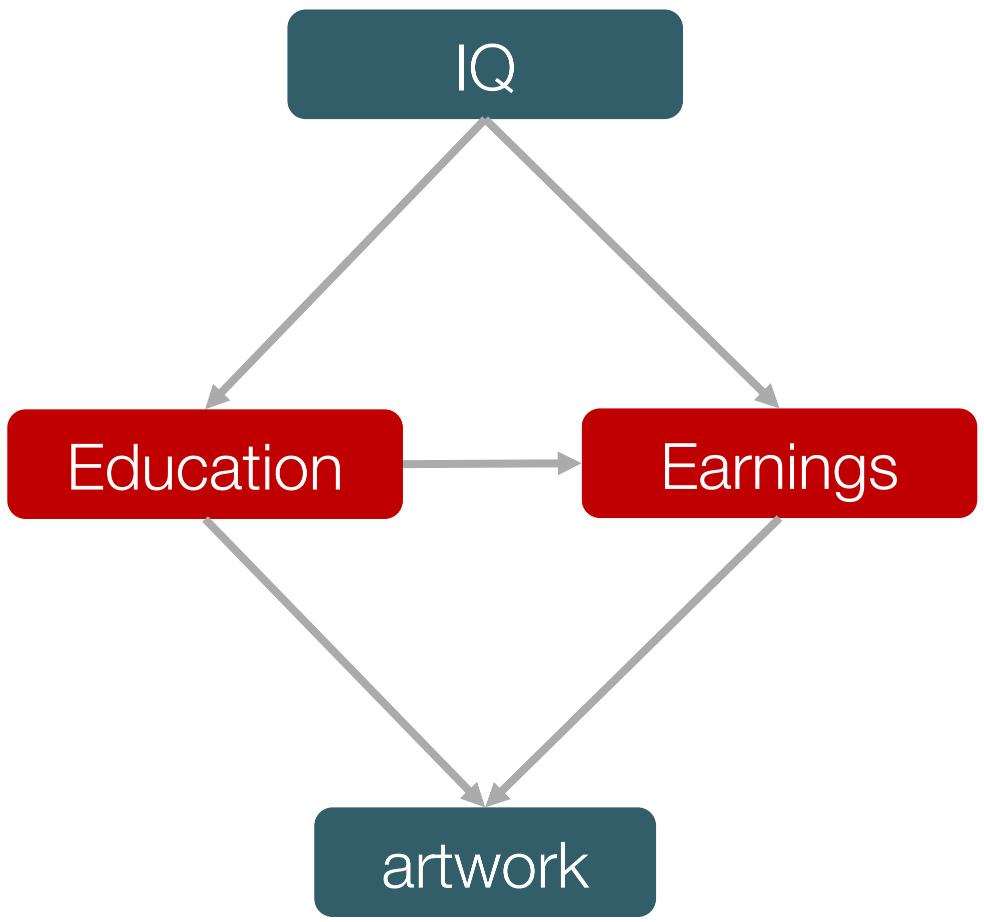
One of the most prominent uses of DAGs in causal inference is to help decide whether and how the available data identifies a desired causal target of inference under the assumed causal model.





Motivating example

- Question: the causal effect of education attainment on earnings
- Dataset: education, earnings, IQ, spent on artwork

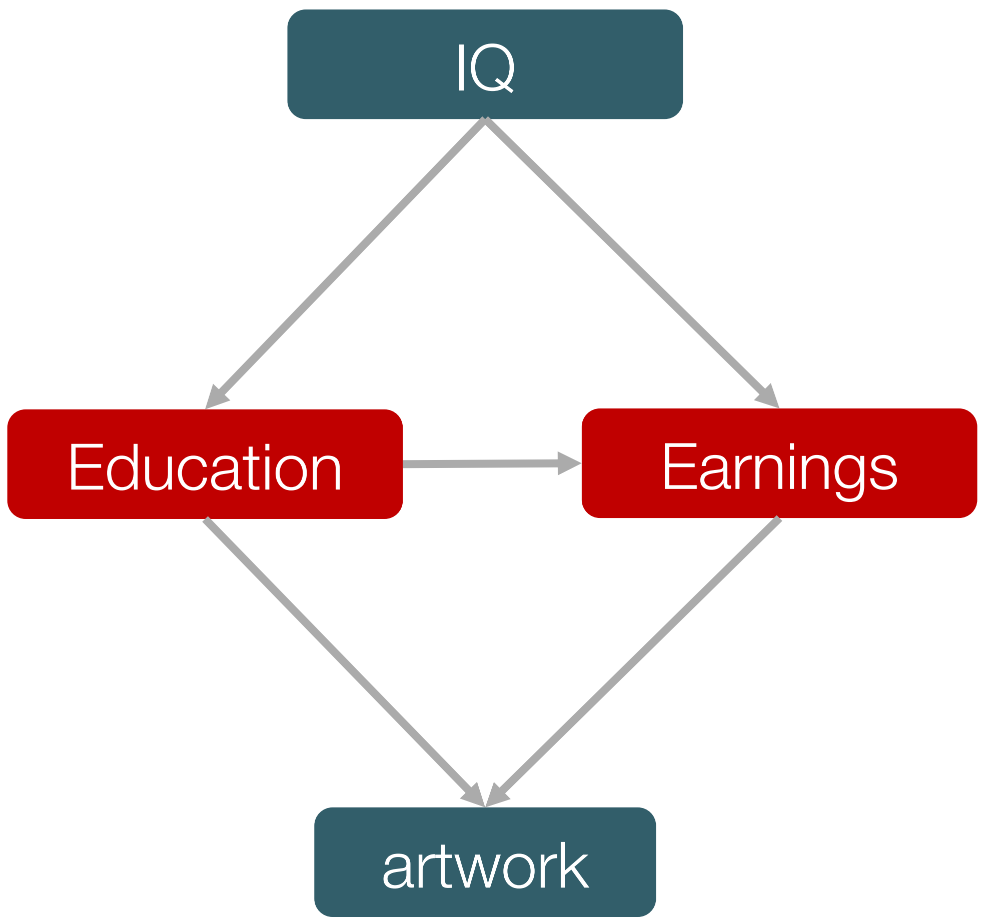


From which can we get an unbiased estimation?

````
```{r}
summary(lm(earnings ~ edu))
summary(lm(earnings ~ edu + IQ))
summary(lm(earnings ~ edu + IQ + art))
```
````

Motivating example

- Question: the causal effect of education attainment on earnings
- Dataset: education, earnings, IQ, spent on artwork




```
```{r}
N <- 100000

#generate data
IQ <- rnorm(N)
edu <- .5 * IQ + rnorm(N)
earnings <- .3 * IQ + .4 * edu + rnorm(N)
art <- 1.2 * edu + .6 * earnings + rnorm(N)
```
```

From which can we get an unbiased estimation?

```
```{r}
summary(lm(earnings ~ edu))
summary(lm(earnings ~ edu + IQ))
summary(lm(earnings ~ edu + IQ + art))
```
```

```
Call:
lm(formula = earnings ~ edu)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.2825 -0.6950 -0.0023  0.6929  4.4687 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.344e-05  3.274e-03   -0.01    0.992    
edu          5.181e-01  2.925e-03  177.12   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.035 on 99998 degrees of freedom
Multiple R-squared:  0.2388,	Adjusted R-squared:  0.2388 
F-statistic: 3.137e+04 on 1 and 99998 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = earnings ~ edu + IQ)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.2078 -0.6729 -0.0015  0.6727  3.9517 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.001230   0.003162  -0.389    0.697    
edu          0.398195   0.003158 126.088   <2e-16 ***
IQ           0.299418   0.003525  84.952   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9999 on 99997 degrees of freedom
Multiple R-squared:   0.29,	Adjusted R-squared:   0.29 
F-statistic: 2.043e+04 on 2 and 99997 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = earnings ~ edu + IQ + art)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6666 -0.5782  0.0003  0.5773  3.7976 

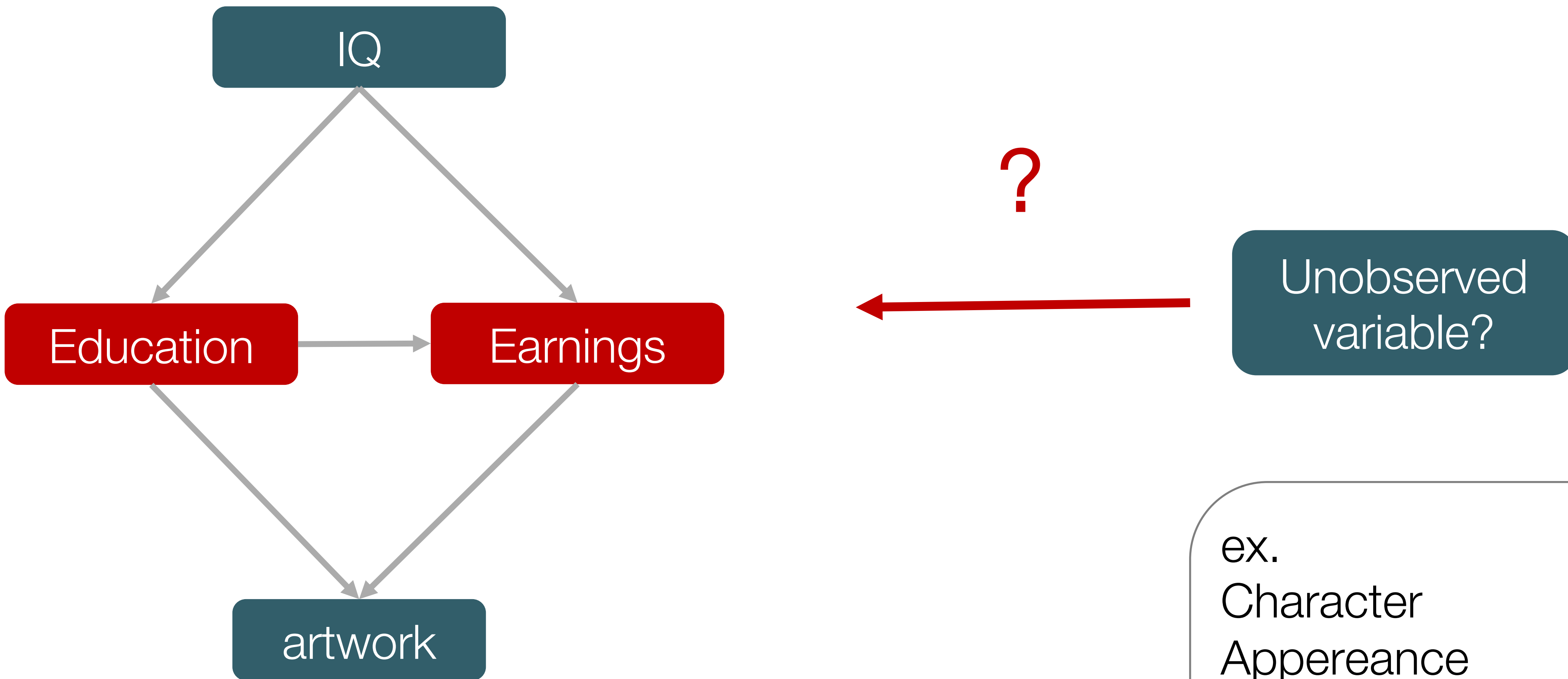
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.001869   0.002708   -0.69     0.49    
edu         -0.237545   0.004293  -55.33   <2e-16 ***
IQ           0.218788   0.003048   71.79   <2e-16 ***
art          0.443131   0.002324  190.68   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8563 on 99996 degrees of freedom
Multiple R-squared:  0.4793,	Adjusted R-squared:  0.4793 
F-statistic: 3.069e+04 on 3 and 99996 DF,  p-value: < 2.2e-16
```

````
```{r}
N <- 100000

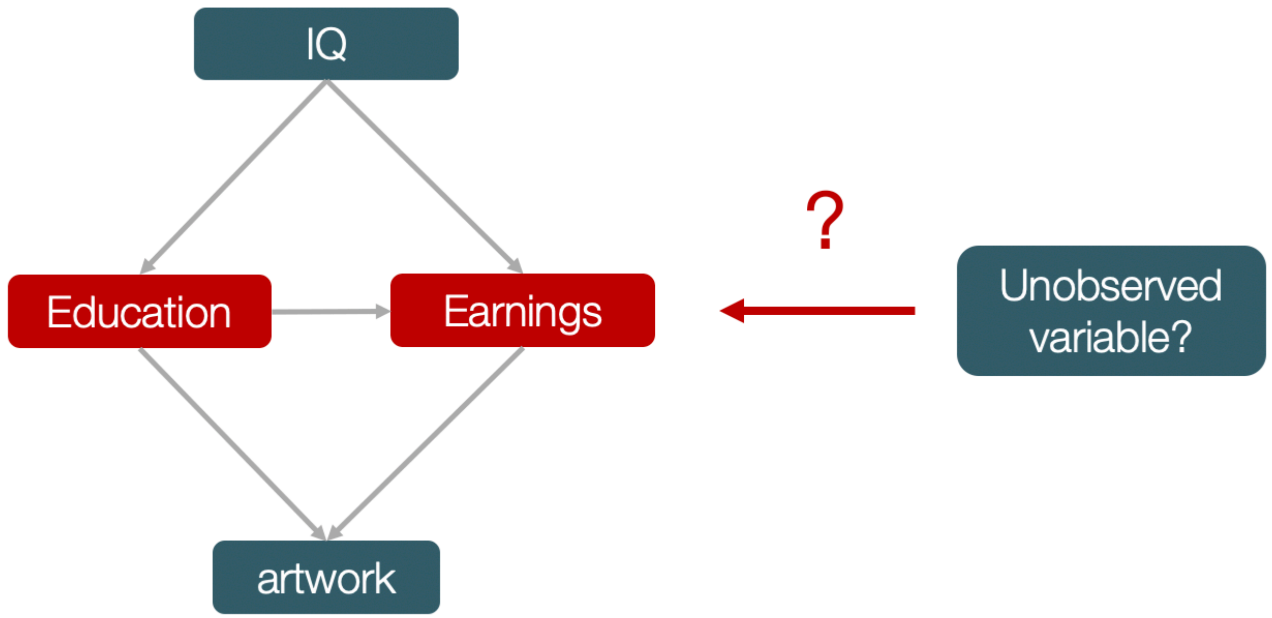
#generate data
IQ <- rnorm(N)
edu <- .5 * IQ + rnorm(N)
earnings <- .3 * IQ + .4 * edu + rnorm(N)
art <- 1.2 * edu + .6 * earnings + rnorm(N)
```
````

What if we have a hidden/unobservable variable?



ex.
Character
Appereance
Family status
Parents’ job and education
Partners’ job and education
……

The challenge: Confounders and Confounding



1. Unmeasured variables(Hidden variables)
2. Measured variables → confounders or not?

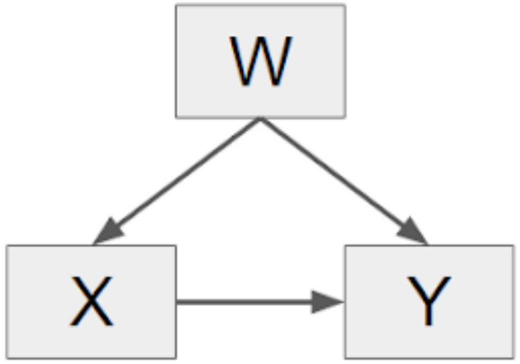
Variables	Confounder	Not a confounder
Measured	Adjustment strategies **(This talk)**	Exclude from estimation
Unmeasured	Big problem ignorability assumption (no unmeansured confounding)	Nevermind ☺

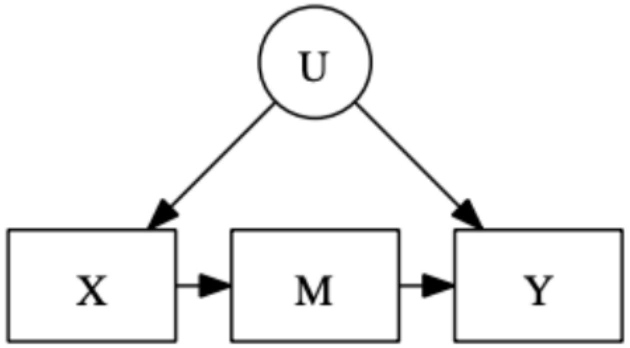
Agenda

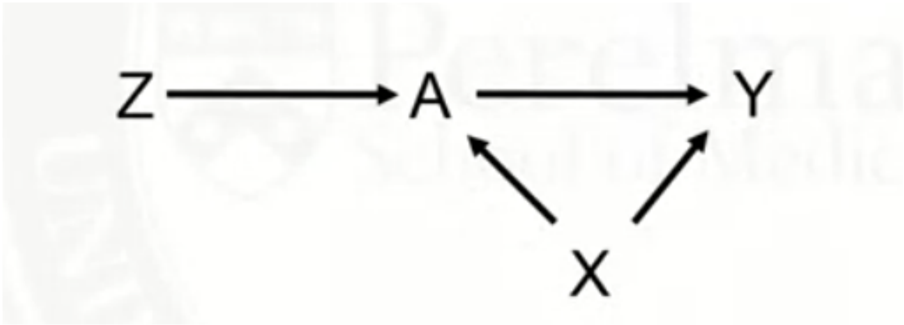
- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - Backdoor Adjustment with Examples
    - Frontdoor Adjustment with Examples
    - Instrumental Variable Analysis
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

Identification Strategies/adjustments

- The back-door criterion: identification by conditioning
- The front-door criterion: identification by mechanisms
- Instrumental variables





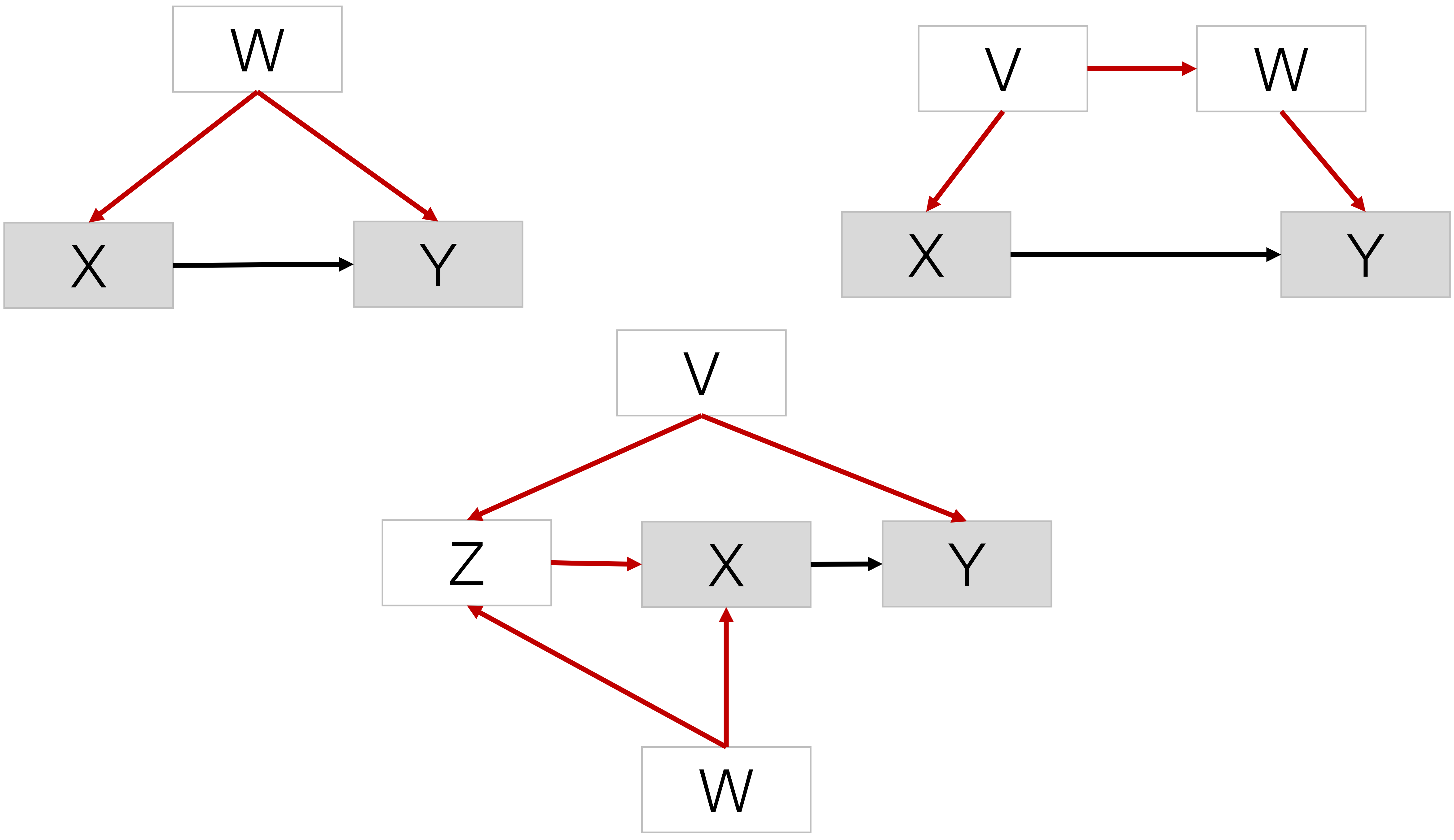


Agenda

- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - **Backdoor Adjustment with Examples**
    - Frontdoor Adjustment with Examples
    - Instrumental Variable Analysis
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

Back-door path

A back-door path is a undirected path between X and Y with an arrow into X.

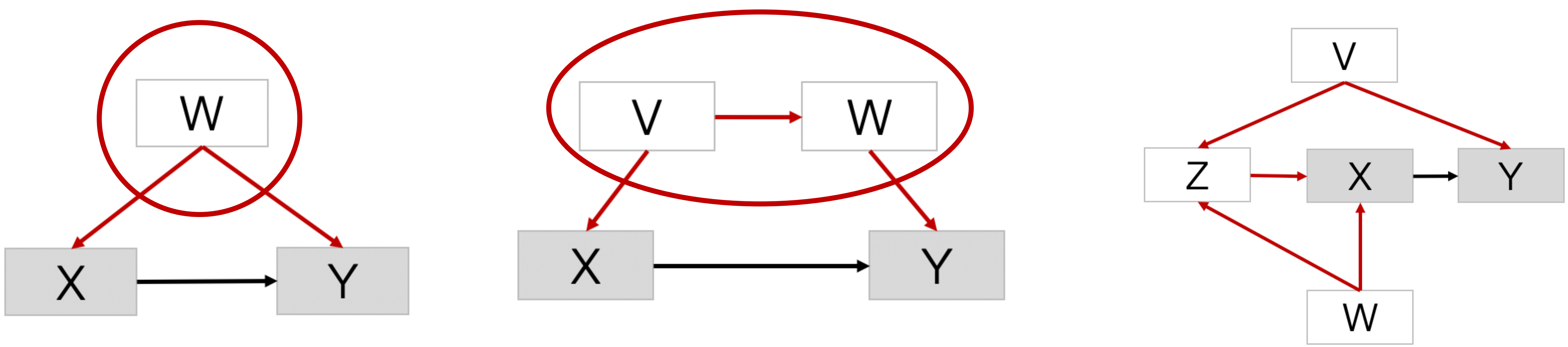


Backdoor path criterion

A set of variable X is sufficient to control for confouding if:

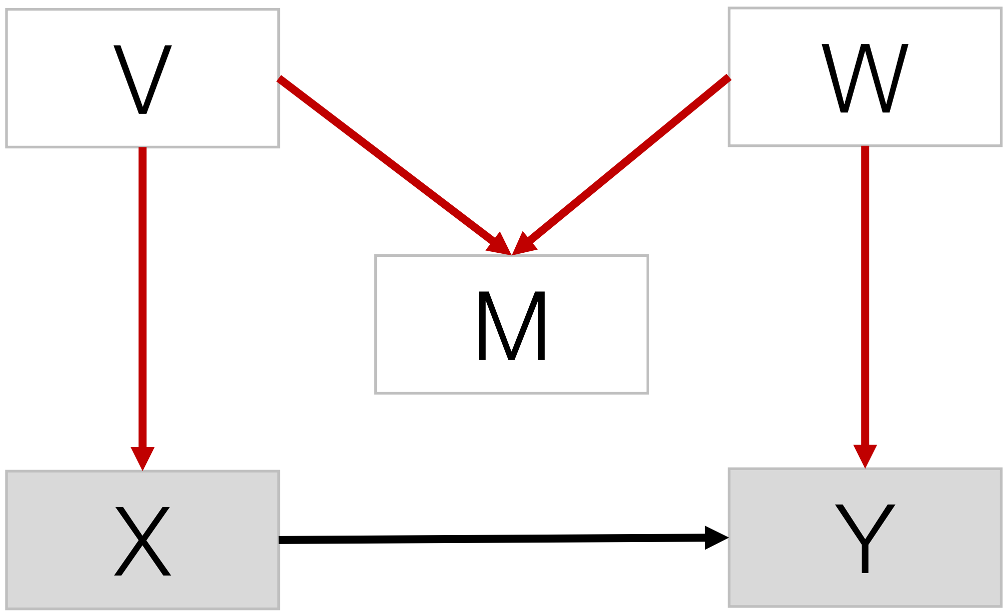
- It blocks all backdoor path from treatment to the outcome

- it does not include any descendants of treatment



$$\Pr\left(Y|do(X=x)\right)=\sum_{s}\Pr\left(Y|X=x,S=s\right)\Pr\left(S=s\right)$$

Backdoor path criterion - Quiz



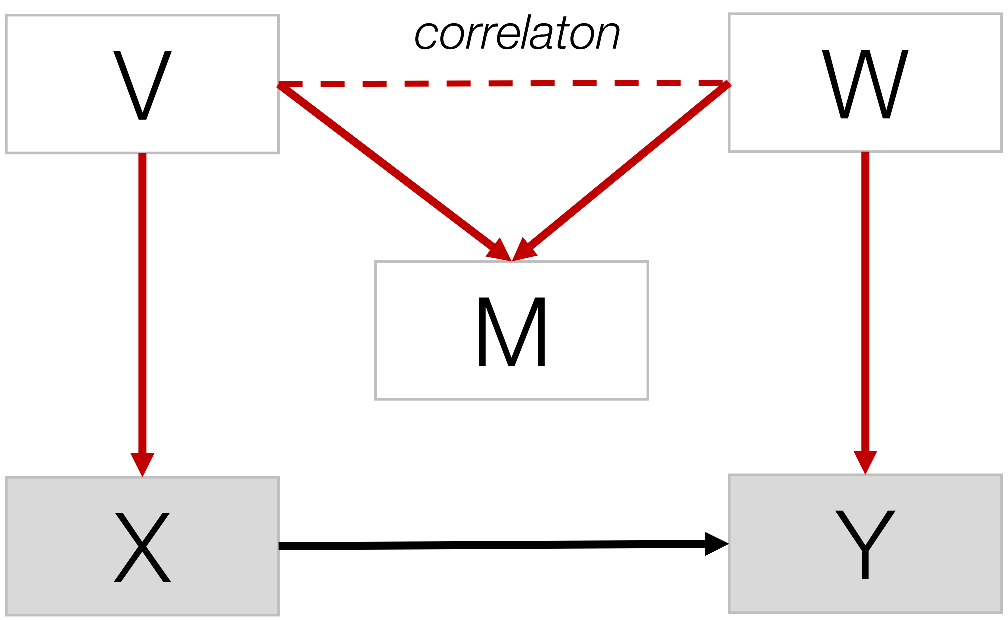
A set of variable X is sufficient to control for confouding if:

- It blocks all backdoor path from treatment to the outcome
- it does not include any descendants of treatment

Sets of variables that are NOT sufficient to control for confouding:

A. {}
B. {V}
C. {W}
D. {M}
E. {M,W}
F. {M,V}
G. {M,V,W}

Backdoor path criterion - Quiz



A set of variable X is sufficient to control for confouding if:

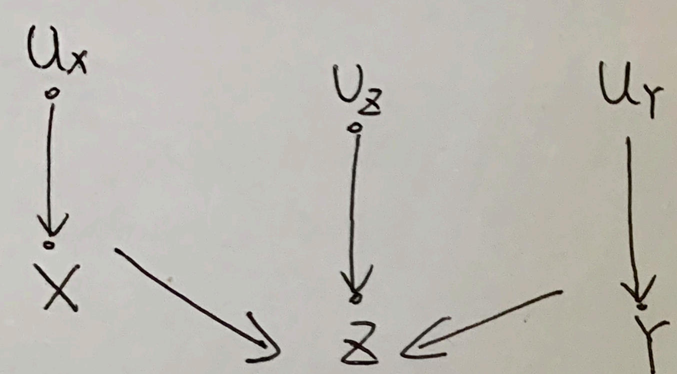
- It blocks all backdoor path from treatment to the outcome

- it does not include any descendants of treatment

- V and M are likely dependent
- W and M are likely dependent
- V and W are independent
- V and W are dependent conditional on M

Sets of variables that are NOT sufficient to control for confouding:

A. {}
B. {V}
C. {W}
D. {M}
E. {M,W}
F. {M,V}
G. {M,V,W}



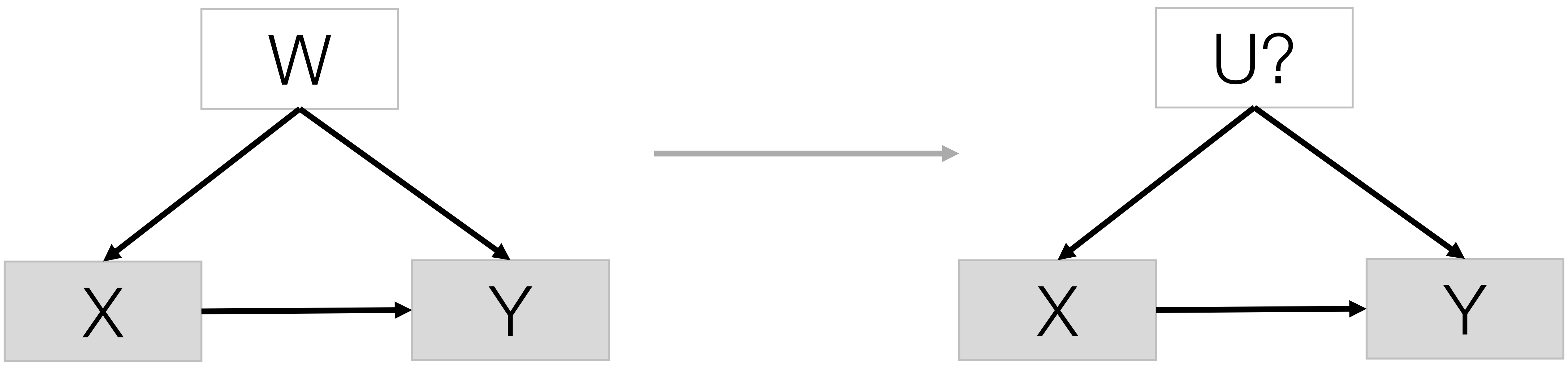
$$f_X = X = U_X$$

$$f_Y: \; Y = U_Y$$

$$f_Z: \; Z = X + Y + U_Z =: c \qquad \Rightarrow X = c - U_Z - Y$$

$$P(X = x \mid Y = y, Z = c) \neq P(X = x \mid Z = c)$$

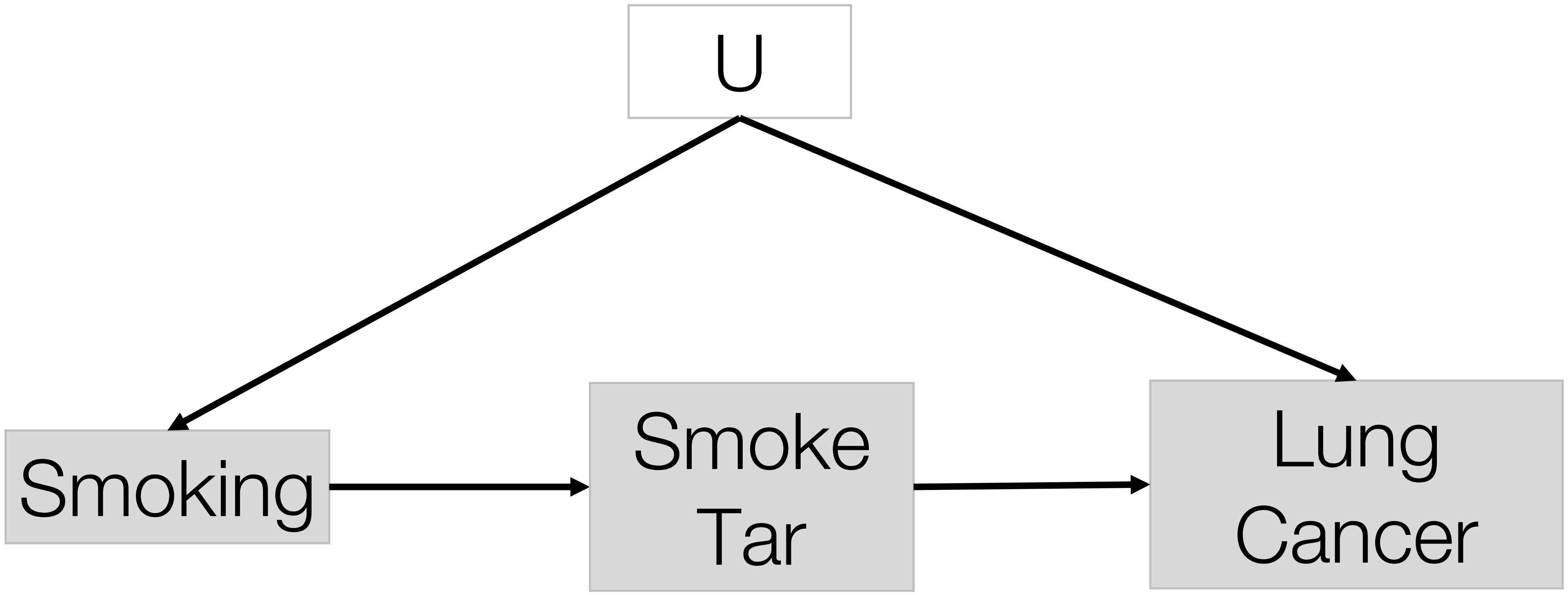
What if W is not observable or cannot be measured?



Agenda

- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
    - **Perspective 1 - Adjustments**
        - Backdoor Adjustment with Examples
        - **Frontdoor Adjustment with Examples**
        - Instrumental Variable Analysis
    - **Perspective 2 - Do-calculus**
        - Basic Notations
        - 3 Rules of Do-calculus with Examples to Explain WHY They are True
        - Examples to Show How It Works
    - **Connecting Do-Calculus with Adjustment perspective**

How can we identify causal effect of smoking on lung cancer if an unobservable *U* exists?



If we know that smoking causes lung cancer ONLY through increasing smoke tar.

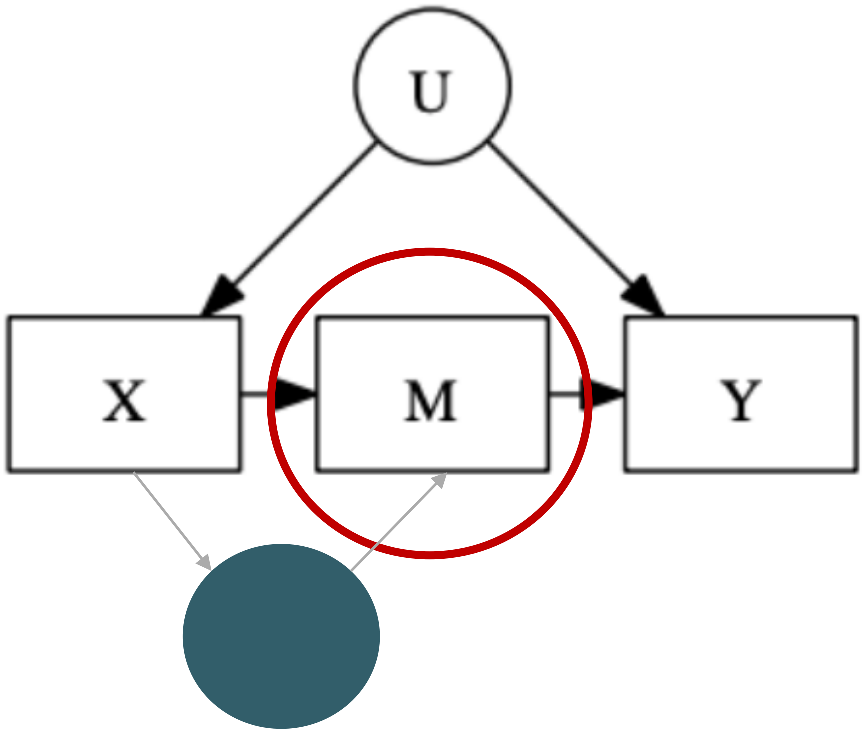
Then we just need to show:

- No confounders between smoking and smoke tar
- No direct influence from smoking on lung cancer

Front-door criterion

A set of variables M satisfies the front-door criterion when:

- M blocks all directed paths from X to Y
- There are no unblocked back-door paths from X to M
- X blocks all back-door pahts from M to Y



$$\Pr(Y|do(X=x)) = \sum_{m} \Pr(M=m|X=x) \sum_{x'} \Pr(Y|X=x', M=m) \Pr(X=x') \tag{12}$$

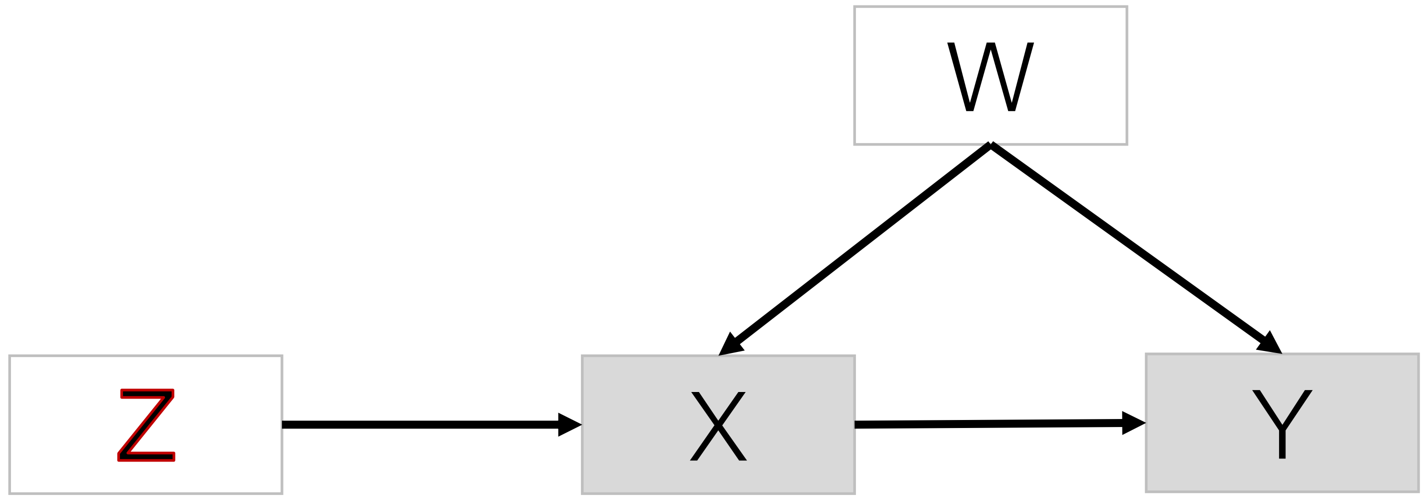
Agenda

- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - Backdoor Adjustment with Examples
    - Frontdoor Adjustment with Examples
    - **Instrumental Variable Analysis**
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

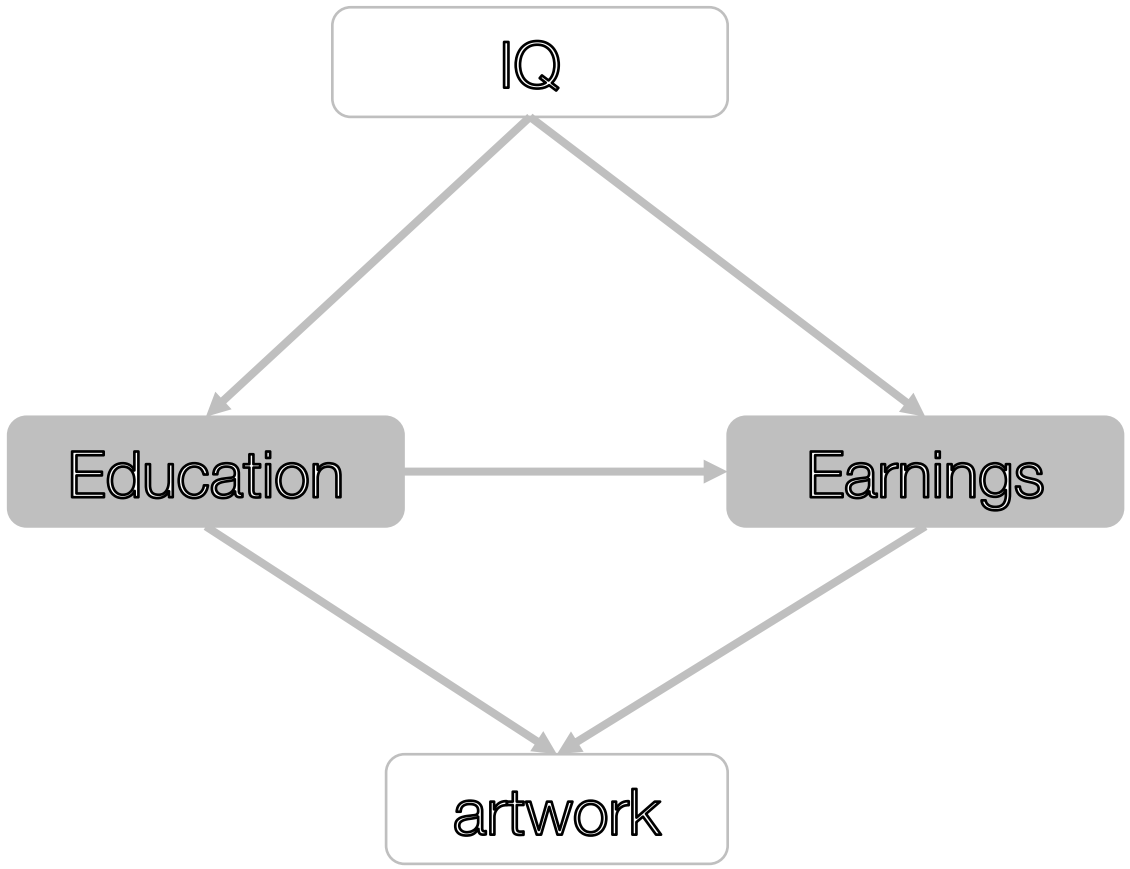
Instrumental variables

Z is an IV:

- It affects treatment X
- But it does not (directly) affect the outcome Y

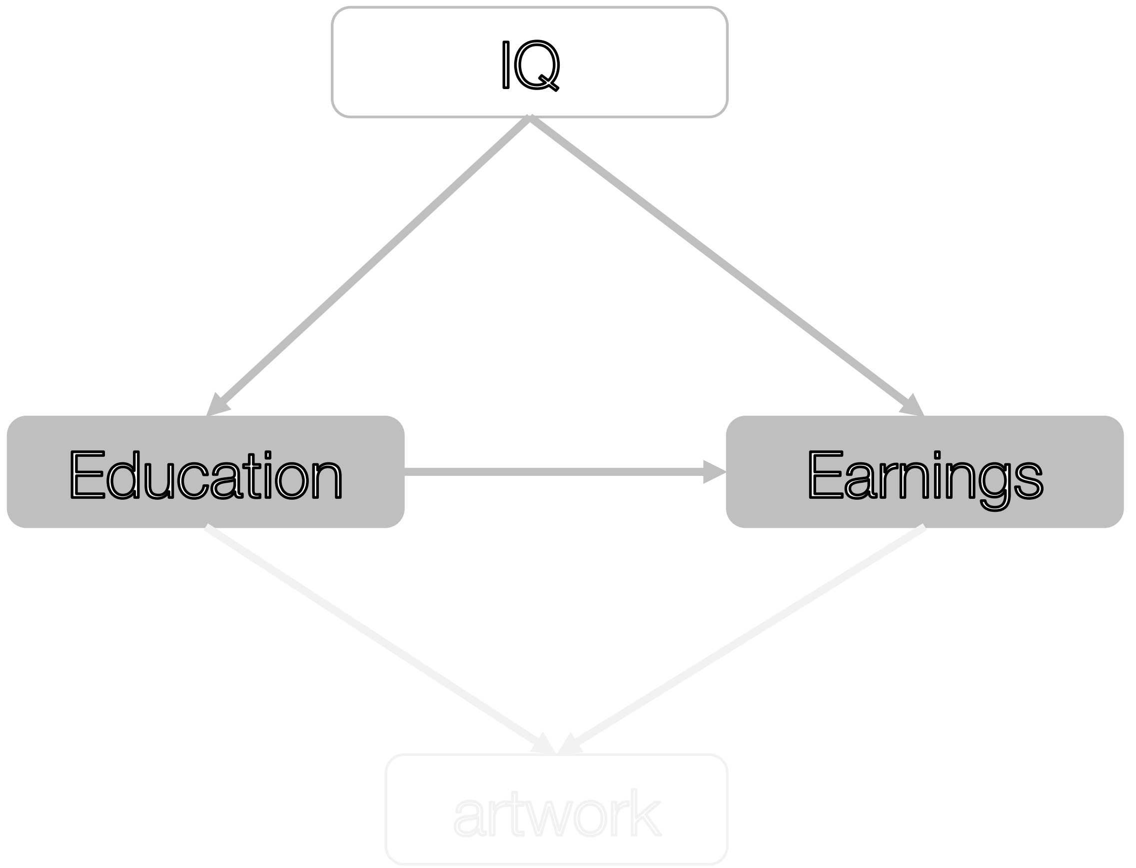


Flow of thinking



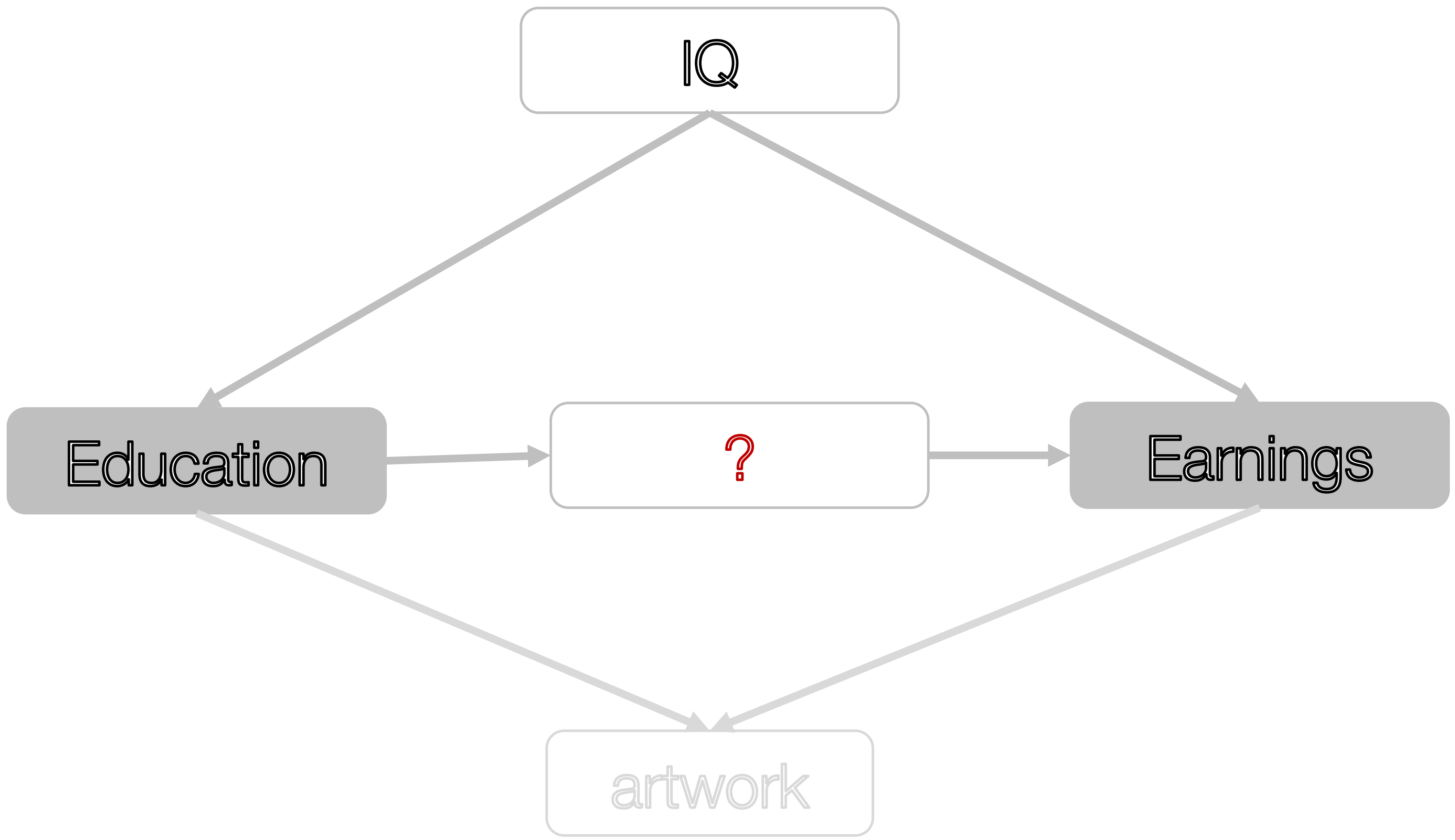
1. IQ is a confounder, but artwork is not

Flow of thinking



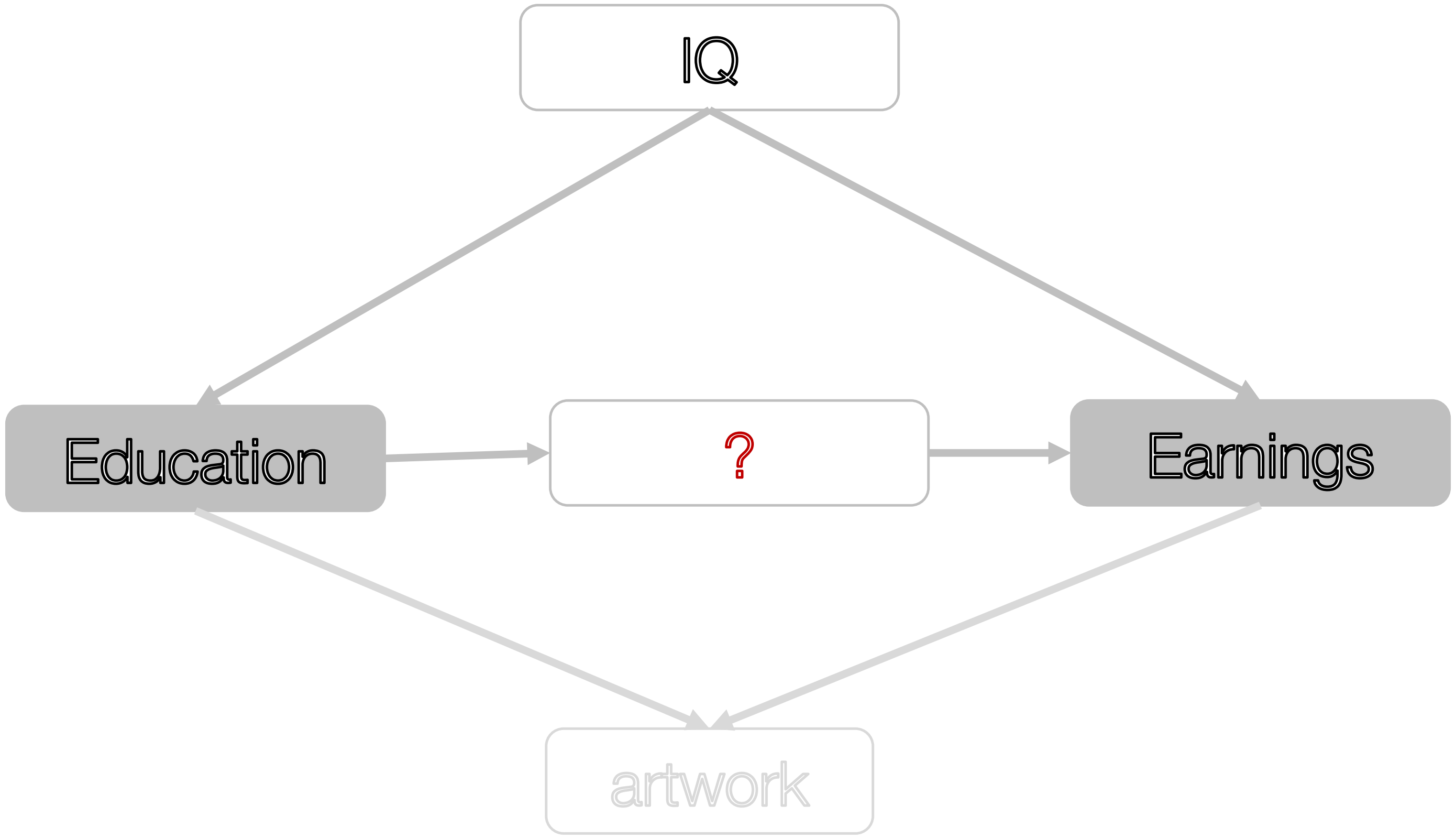
1. IQ is a confounder, but artwork is not
2. **If IQ is observable → apply backdoor adjustment**

Flow of thinking



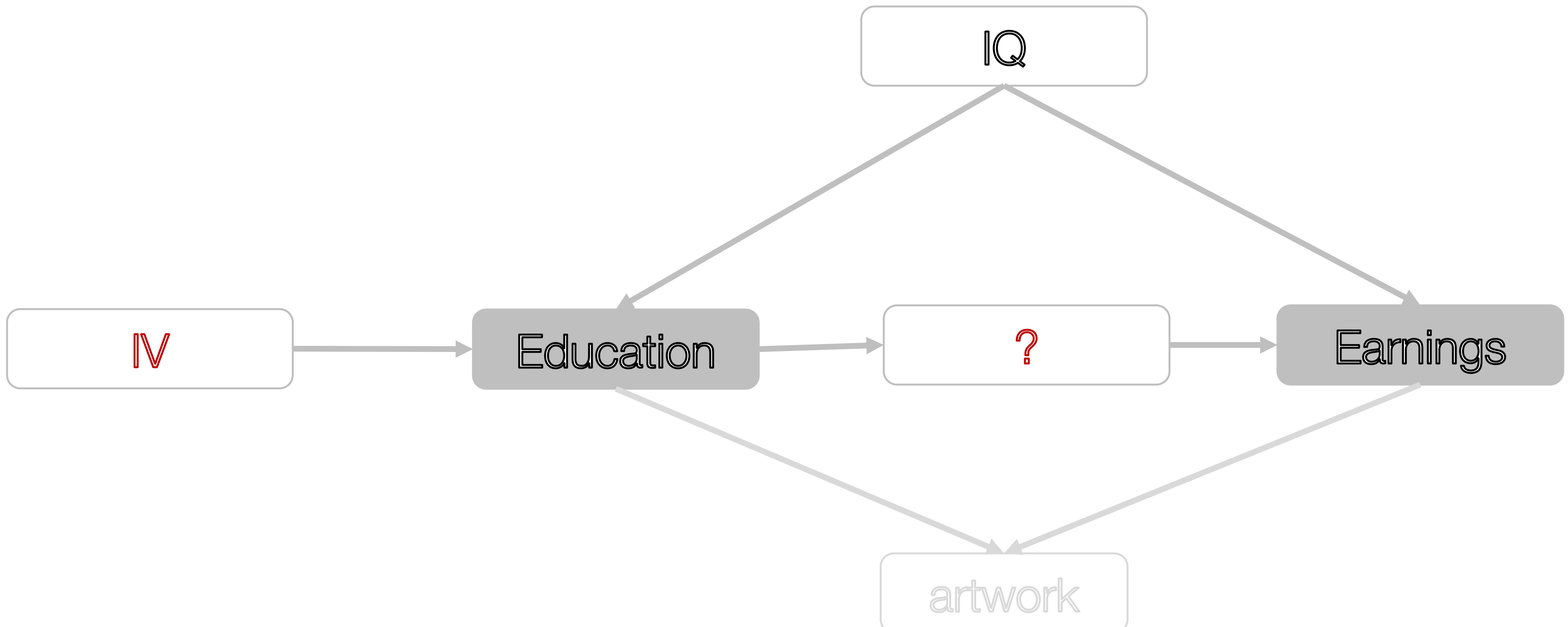
1. IQ is a confounder, but artwork is not
2. If IQ is observable → apply backdoor adjustment
3. If IQ is not observable:
    1. **Try applying front door adjustment: Does education influence earnings through a single mechanism, e.x. amount of knowledge? (Maybe not, social network also matters)**

Flow of thinking



1. IQ is a confounder, but artwork is not
2. If IQ is observable → apply backdoor adjustment
3. If IQ is not observable:
    1. **Try applying front door adjustment: Does education influence earnings through a single mechanism, e.x. amount of knowledge? (Maybe not, social network also matters)**

Flow of thinking



1. IQ is a confounder, but artwork is not
2. If IQ is observable → apply backdoor adjustment
3. If IQ is not observable:
   1. Try applying front door adjustment: Does education influence earnings through a single mechanism, e.x. amount of knowledge? (Maybe not, social network also matters)
   2. **Try using instrument variables:**

   e.g. Angrist & Krueger (1991): birth quarter; compulsory education laws

Why using adjustments?

1. Not all variables are necessary to be observed – we only need to observe variables satisfying backdoor or frontdoor criterion
2. Helps us detect confounders and design observational studies

Agenda

- **Identification problem & the role of DAGs (a motivating example)**
- **Identification strategies**
  - **Perspective 1 - Adjustments**
    - Backdoor Adjustment with Examples
    - Frontdoor Adjustment with Examples
    - Instrumental Variable Analysis
  - **Perspective 2 - Do-calculus**
    - Basic Notations
    - 3 Rules of Do-calculus with Examples to Explain WHY They are True
    - Examples to Show How It Works
  - **Connecting Do-Calculus with Adjustment perspective**

Notations

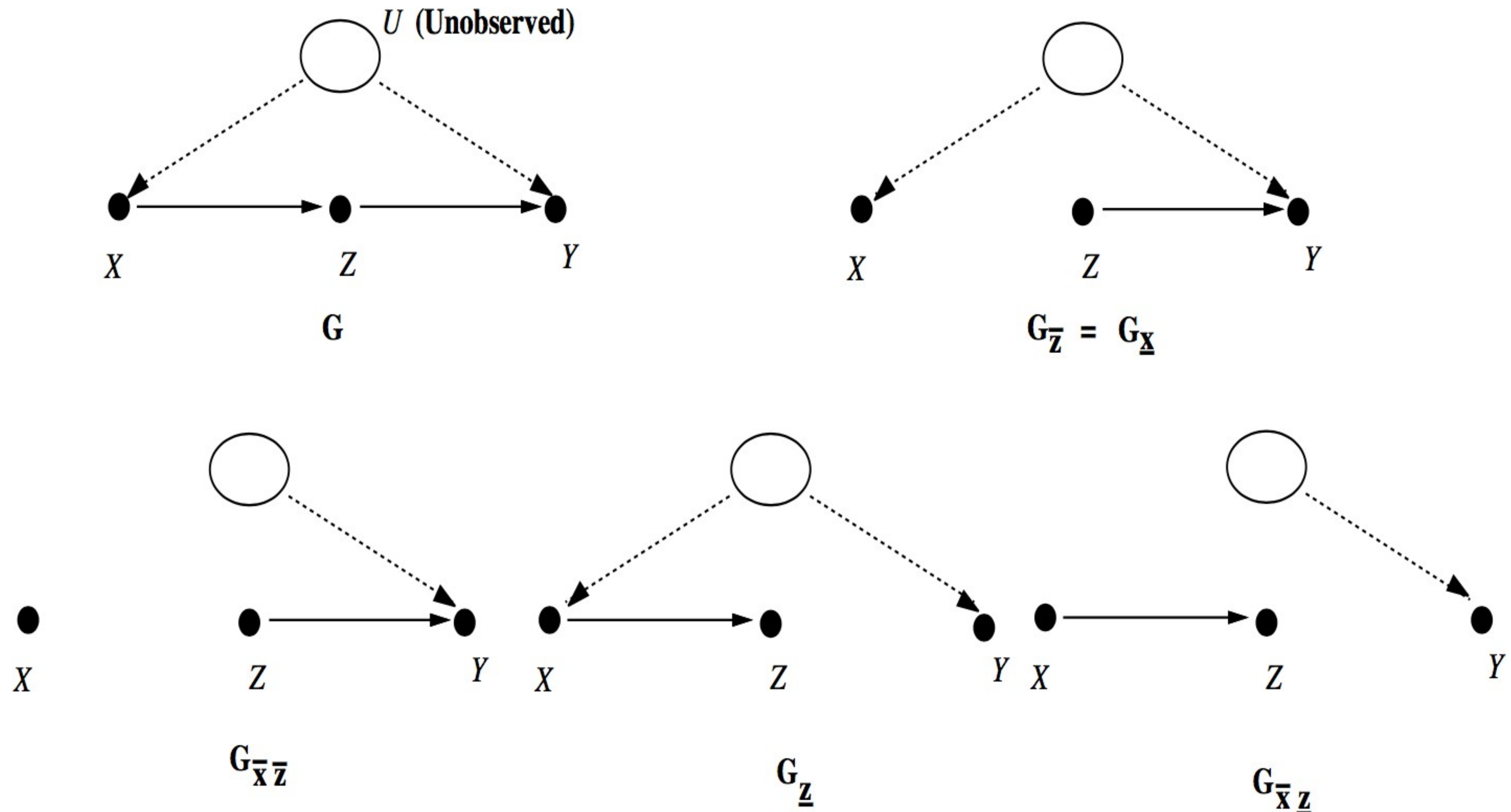
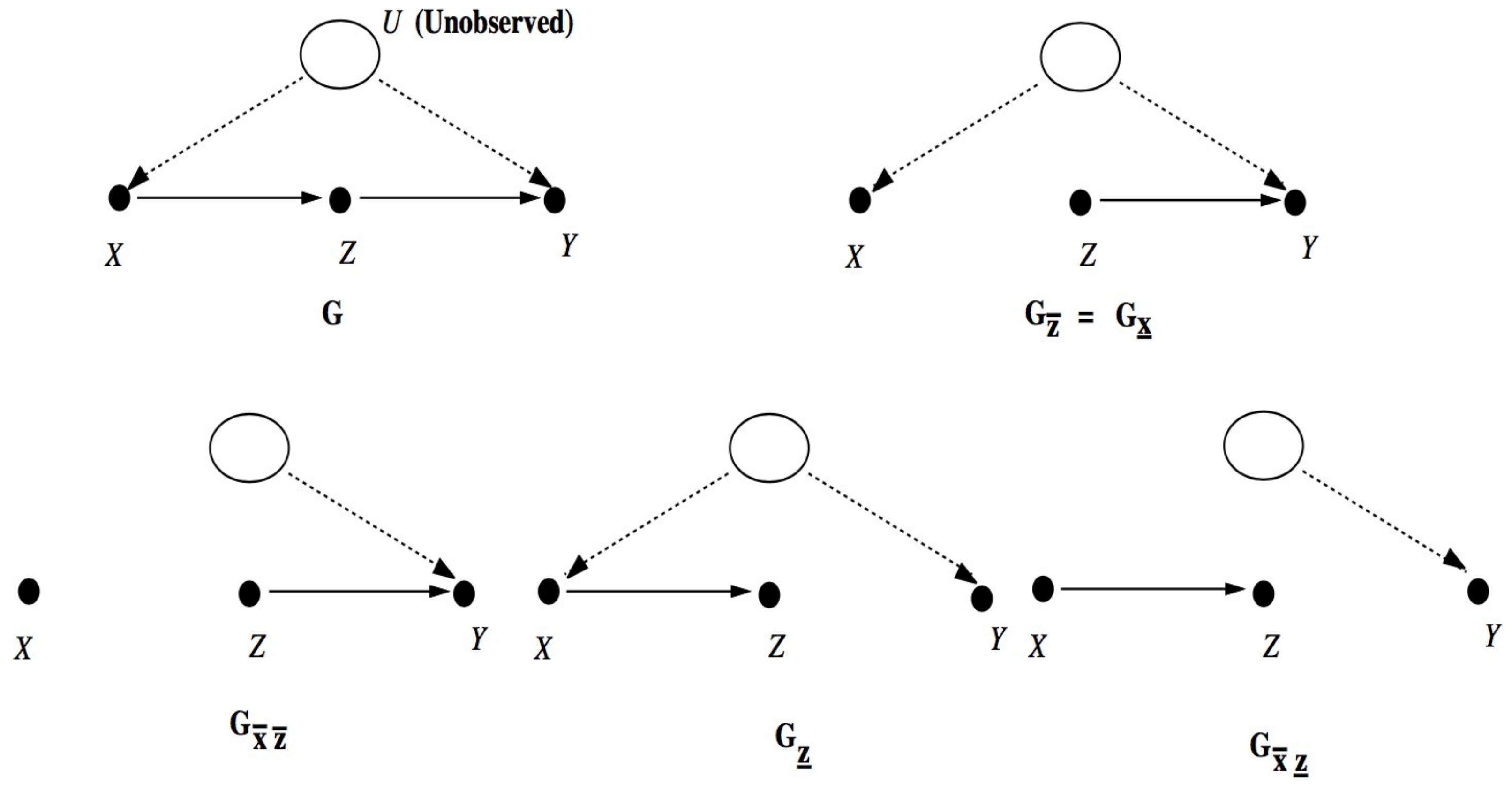


Image: Judea Pearl

Notations

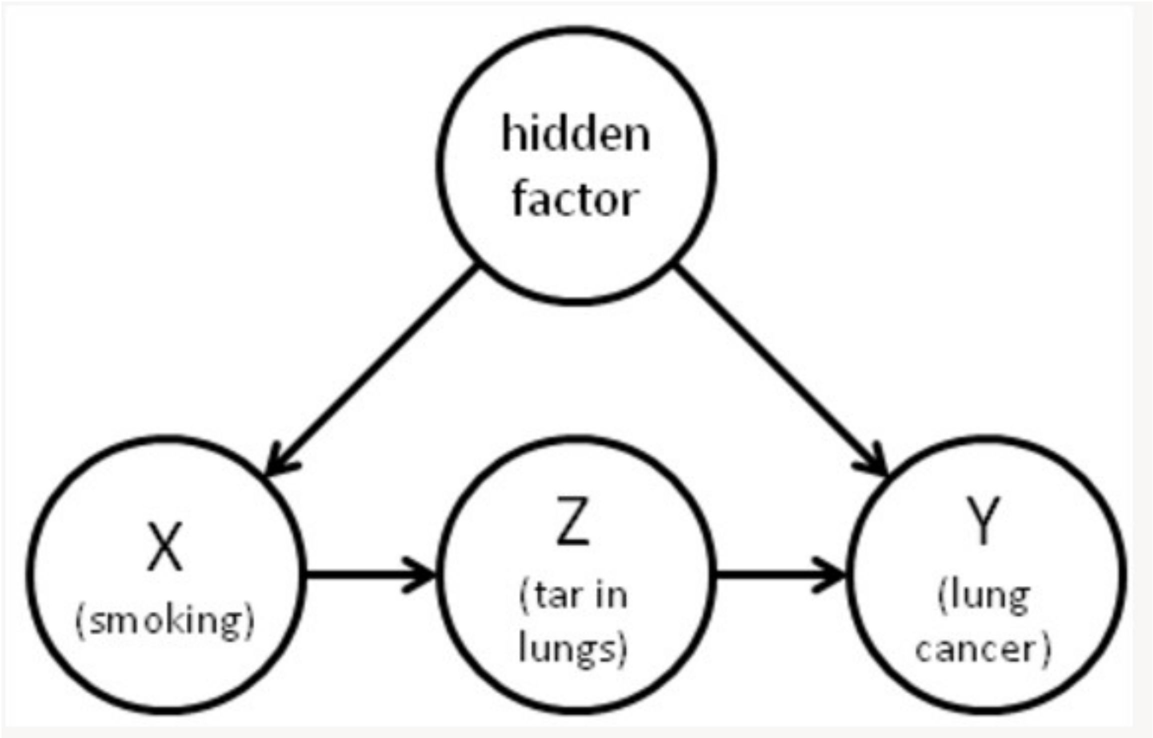


Notation: a graph G, W, X, Y, Z are disjoint subsets of the variables. $G_{\overline{X}}$ denotes the perturbed graph in which all edges pointing to X have been deleted, and $G_{\underline{X}}$ denotes the perturbed graph in which all edges pointing from X have been deleted. $Z(W)$ denote the set of nodes in Z which are not ancestors of W

Image: Judea Pearl

Example: Smoking and lung cancer

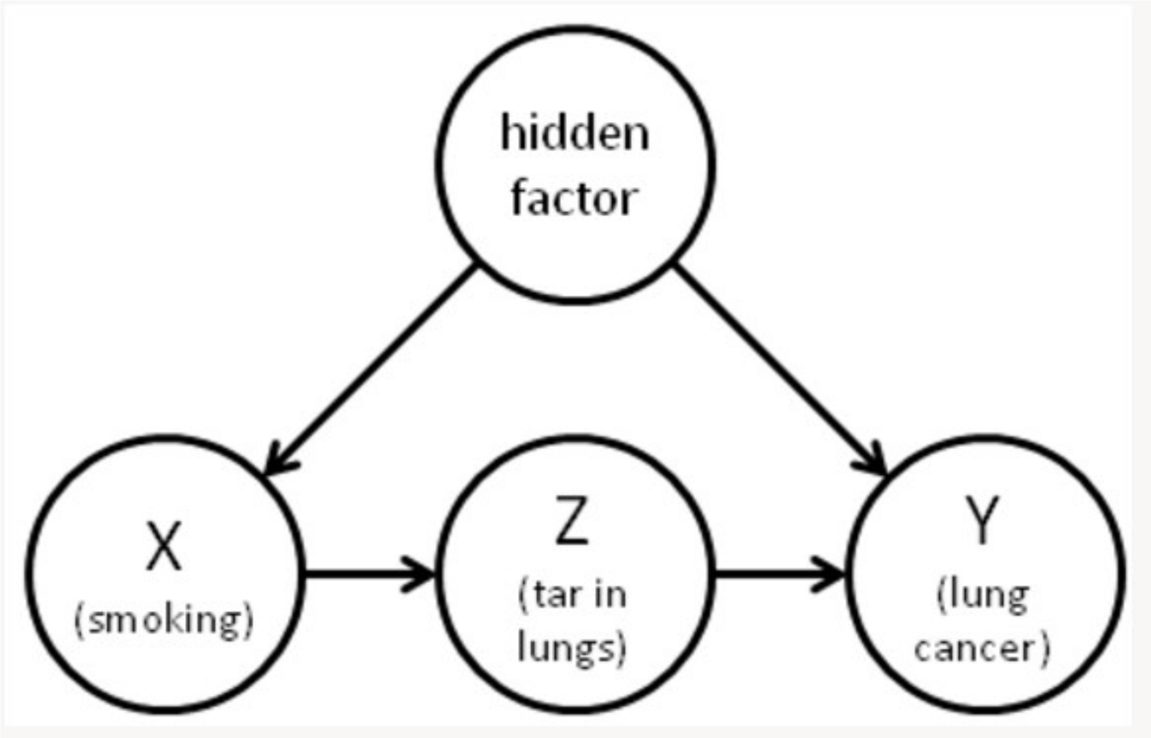
$$p(y|do(x)) = \sum_z p(y|z, do(x))p(z|do(x))$$



Note: We have no information about the hidden variable that could cause both smoking and cancer

Example: Smoking and lung cancer

$$p(y|do(x)) = \sum_z p(y|z, do(x))p(z|do(x))$$

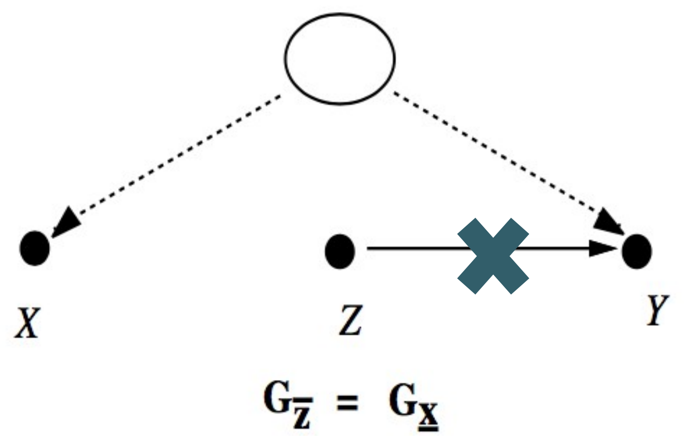


Note: We have no information about the hidden variable that could cause both smoking and cancer

Pearl’s 3 rules

- Ignoring observations/Insertion/deletion of observations

$$p(y|do(x), z, w) = p(y|do(x), w) \text{ if } (Y \perp Z|X, W)_{G_{\overline{X}}}$$



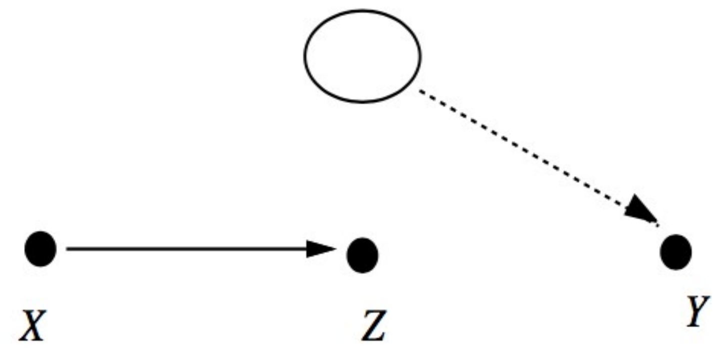
Pearl's 3 rules

- Ignoring observations/Insertion/deletion of observations

$$p(y|do(x), z, w) = p(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

- Action/Observation exchange (the back-door criterion)

$$p(y|do(x), do(z), w) = p(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X},\underline{Z}}}$$



$G_{\overline{X}\,\underline{Z}}$

Pearl's 3 rules

- Ignoring observations/Insertion/deletion of observations

$$p(y|do(x), z, w) = p(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$$

- Action/Observation exchange (the back-door criterion)

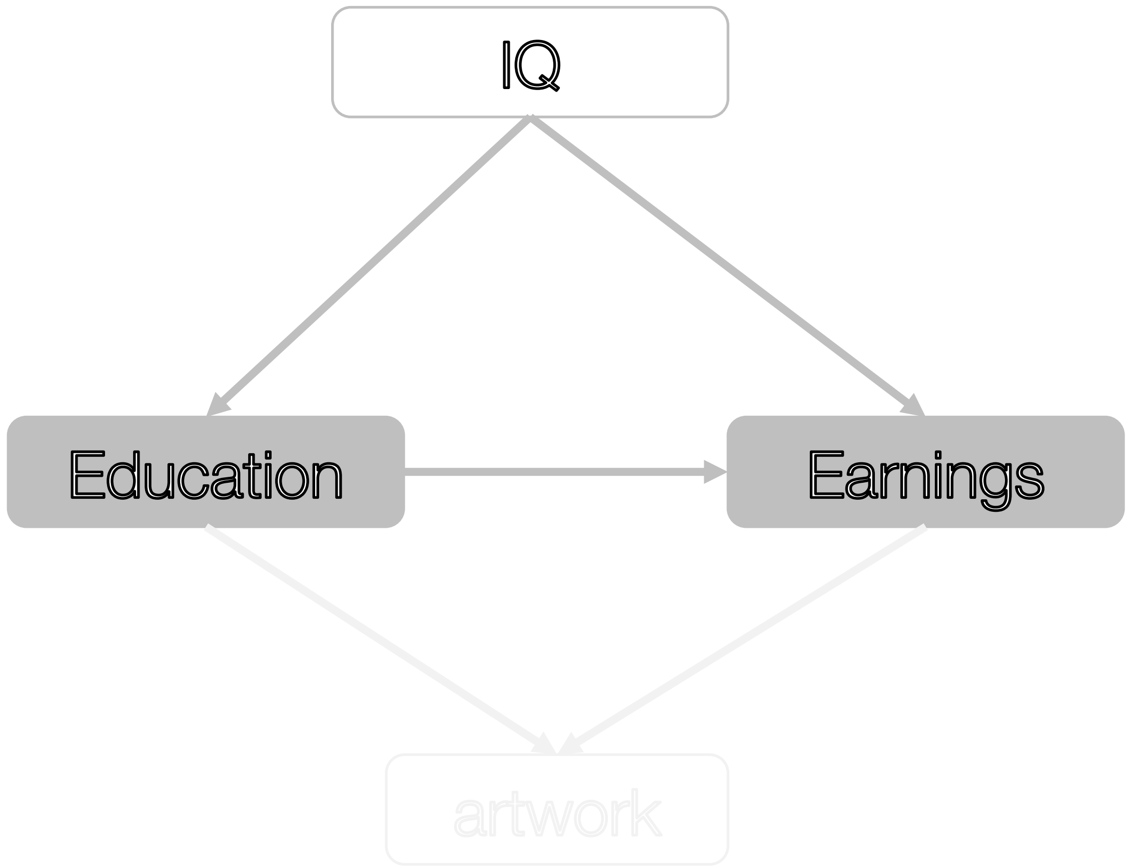
$$p(y|do(x), do(z), w) = p(y|do(x), z, w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}, \underline{Z}}}$$

- Ignoring actions/interventions

$$p(y|do(x), do(z), w) = p(y|do(x), w) \text{ if } (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}, \overline{Z(W)}}}$$

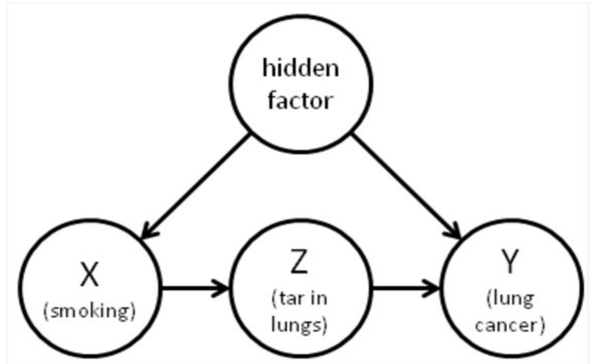
Notation: a graph G, W, X, Y, Z are disjoint subsets of the variables. $G_{\overline{X}}$ denotes the perturbed graph in which all edges pointing to X have been deleted, and $G_{\underline{X}}$ denotes the perturbed graph in which all edges pointing from X have been deleted. $Z(W)$ denote the set of nodes in Z which are not ancestors of W

Flow of thinking



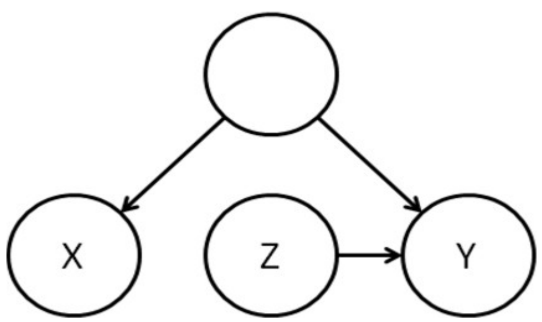
1. IQ is a confounder, but artwork is not
2. **If IQ is observable → apply backdoor adjustment**

Example



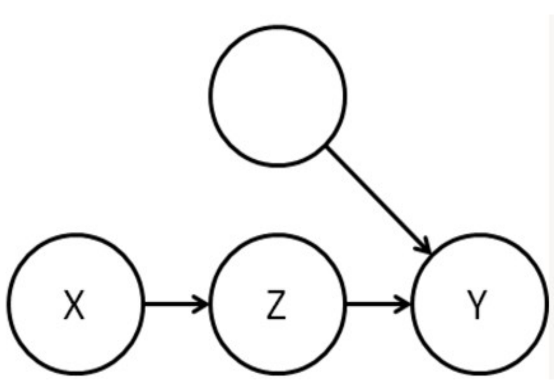
We can't try to apply rule 1 because there is no observations to ignore, we would just have $p(y|do(x)) = p(y|do(x))$.

Try apply rule 2: We would have $p(y|do(x)) = p(y|x)$, that is, the intervention doesn't matter. It's condition is $(Y \perp X)_{G_{\underline{X}}}$:



Y and X are not d-separated, because they have a common ancestor.
$= \Rightarrow$ Rule 2 can't be applied

Try apply rule 3: We would have $p(y|do(x)) = p(y)$, that is, an intervention to force someone to smoke has no impact on whether they get cancer. It's condition is $(Y \perp X)_{G_X}$:



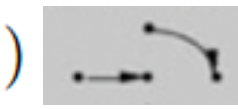
Y and X are not d-separated, because we have unblocked path between them.
$= \Rightarrow$ Rule 3 can't be applied

Example

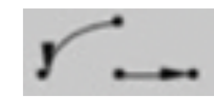
$$p(y|do(x)) = \sum_z p(y|z, do(x))p(z|do(x))$$

$$p(y|do(x), do(z), w) = p(y|do(x), z, w)$$

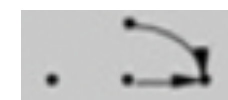
$$= \sum_z p(y|z, do(x))p(z|x) \qquad \text{(rule 2: } (Z \perp\!\!\!\perp X)_{G_{\underline{X}}}\text{)}$$



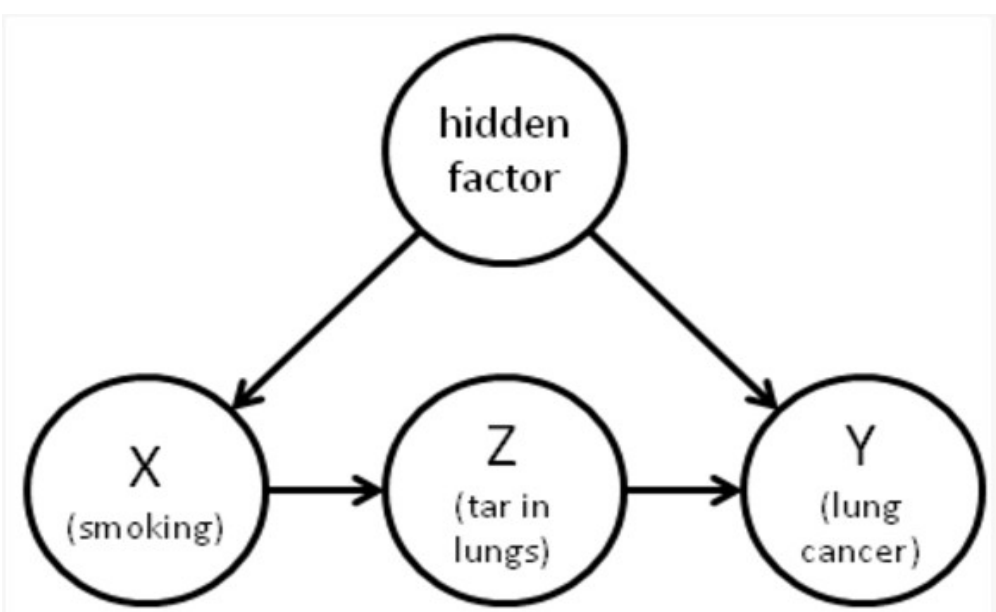
$$= \sum_z p(y|do(z), do(x))p(z|x) \qquad \text{(rule 2: } (Y \perp\!\!\!\perp Z|X)_{G_{\overline{X},\underline{Z}}}\text{)}$$



$$= \sum_z p(y|do(z))p(z|x) \qquad \text{(rule 3: } (Y \perp\!\!\!\perp X|Z)_{G_{\overline{Z},\overline{X}}}\text{)}$$



$$p(y|do(x), do(z), w) = p(y|do(x), w)$$



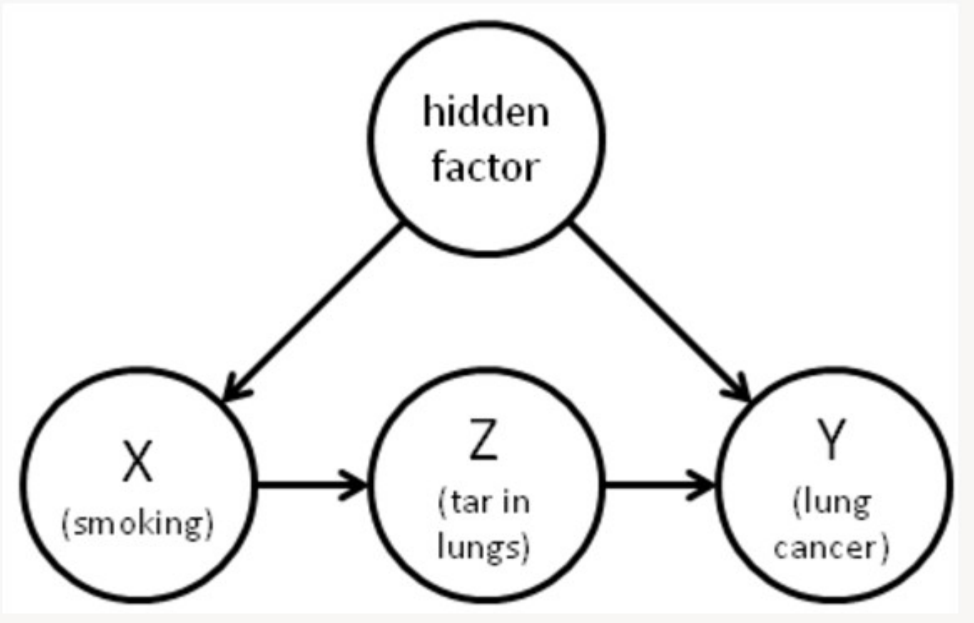
Example

We can use the same approach to the first term on the right hand side:

$$
\begin{aligned}
p(y|do(z)) &= \sum_x p(y|x, do(z))p(x|do(z)) \\
&= \sum_x p(y|x, z)p(x) \qquad\qquad \text{(rule 2 + rule 3)}
\end{aligned}
$$

Finally we can combine these results:

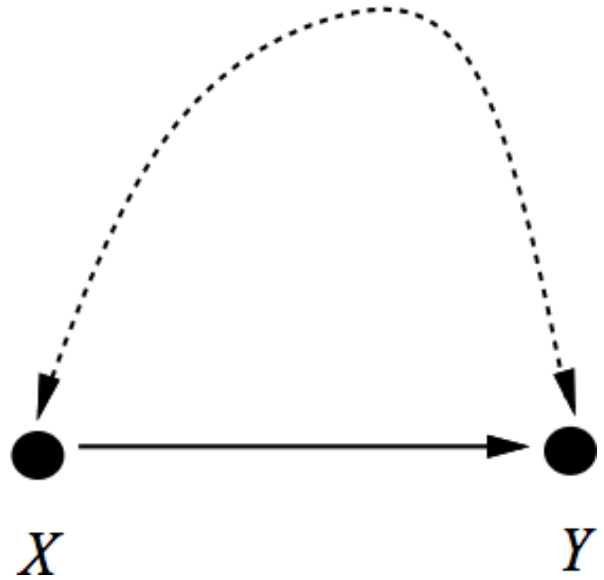
$$
p(y|do(x)) = \sum_{z,x'} p(y|x', z)p(z|x)p(x')
$$



We can now compare $p(y)$ and $p(y|x)$. The needed probabilities can be observed directly from experimental data: What part of smokers have lung cancer, how many of them have tar in their lungs etc.

Example: Summary

- The analysis would have not worked if the graph had missed the tar variable, Z, because there is no general way to compute $p(y|do(x))$ from any observed distributions whenever the causal model includes subgraph shown the figure below
- Causal Calculus can be used to analyze causality in more complicated (and more unethical) situations than RCT
- Causal Calculus can also be used to test whether unobserved variables are missed by removing all do terms from the relation
- Not all models are acyclic. See for example Modeling Discrete Interventional Data Using Directed Cyclic Graphical Models (UAI 2009) by Mark Schmidt and Kevin Murphy



Check-list questions

- Identification problem – What are the chanlleges?
- Why using adjustments?
    - Not all variables are necessary to be observed – we only need to observe variables satisfying backdoor or frontdoor criterion
    - Helps us detect confounders and design observational studies
- Flow of thinking – How to identify a causal effect?

References

- Chp.15 & Chp.16 from ***Handbook of Graphical Models***, by Marloes Maathuis, Mathias Drton, Steffen Lauritzen and Martin Wainwright
- A Probabilistic Calculus of Actions (UAI1994) by Judea Pearl
- Tutorial by Michael Nielsen: http://www.michaelnielsen.org/ddi/if-correlation-doesnt-imply-causation-then-what-does/
- Tutorial in Formalised Thinking: https://formalisedthinking.wordpress.com/2010/08/20/pearls-formalisation-of-causality-sequence-index/
- Judea Pearl, Causality: Models, reasoning, and inference, Cambridge University Press, 2000
- Introduction to Judea Pearl's Do-Calculus, Robert Tucci, 2013

Q & A