Generalizability in Causal Inference

Carlos Cinelli
UCLA
@analisereal

Elias Bareinboim
Columbia University
@eliasbareinboim

Southern California Methods Conference
Riverside, September 2019

Outline

1. What is causal inference?
2. Observational causal inference (internal validity)
3. Transportability of causal effects
4. Recovering from selection bias
5. Data fusion

What is causal inference?

Causal assumptions → Causal conclusions

Inference: from______ to _______

Inference is always *from something to something*.

Inference: from______ to _______

Inference is always *from something to something*.

a. Statistics
   - from <u>sample</u> to <u>distribution</u>

Inference: from______ to _______

Inference is always *from something to something*.

a. Statistics
   - from <u>sample</u> to <u>distribution</u>

b. Observational Causal Inference
   - from <u>observational distribution</u> to <u>experimental distribution</u>

Inference: from______ to _______

Inference is always *from something to something*.

a. Statistics
- from <u>sample</u> to <u>distribution</u>

b. Observational Causal Inference
- from <u>observational distribution</u> to <u>experimental distribution</u>

c. Sampling Selection Bias
- from <u>study (obs/exp) distribution</u> to <u>general (obs/exp) distribution</u>

Inference: from______ to _______

Inference is always *from something to something*.

a. Statistics
- from sample to distribution

b. Observational Causal Inference
- from observational distribution to experimental distribution

c. Sampling Selection Bias
- from study (obs/exp) distribution to general (obs/exp) distribution

d. General Transportability
- from (obs/exp) distributions of populations A, B, C… to experimental distribution of a target population

Inference: from______ to ________

Inference is always *from something to something*.

a. Statistics
- from sample to distribution

b. Observational Causal Inference
- from observational distribution to experimental distribution

c. Sampling Selection Bias
- from study (obs/exp) distribution to general (obs/exp) distribution

d. General Transportability
- from (obs/exp) distributions of populations A, B, C… to experimental distribution of a target population

To make the leap from ________ to _______ we need ***a model.*** *The model allows us to go from assumptions to conclusions, and the assumptions of your model must be in the same level of the leap you want to make.*

Inference: from______ to _______

Inference is always *from something to something*.

a. Statistics
- from <u>sample</u> to <u>distribution</u>

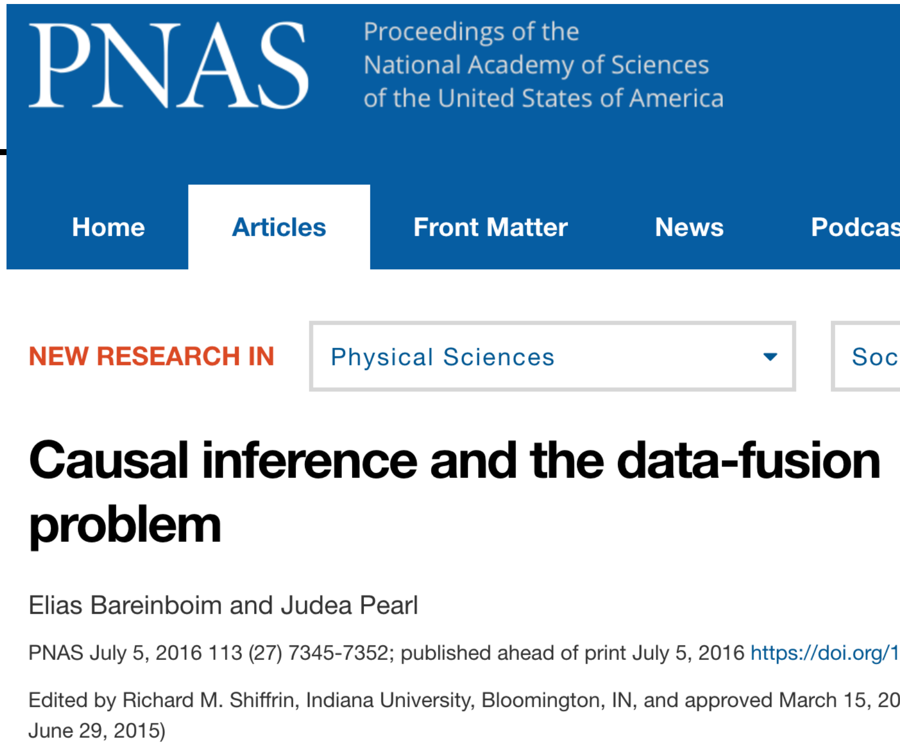
b. Observational Causal Inference
- from <u>observational distribution</u> to <u>experimental distribution</u>

c. Sampling Selection Bias
- from <u>study (obs/exp) distribution</u> to <u>general (obs/exp) distribution</u>

d. General Transportability
- from <u>(obs/exp) distributions of populations A, B, C…</u> to <u>experimental distribution of a target population</u>

- We will <u>model</u> problems of <u>selection bias</u> and <u>transportability</u>.
- Formal language to represent the problem (nonparametrically), reduce it to an exercise of symbolic calculus, and derive complete solutions.

Inference: from ______ to ______


PNAS Proceedings of the National Academy of Sciences of the United States of America

Home | Articles | Front Matter | News | Podcas

NEW RESEARCH IN Physical Sciences Soc

Causal inference and the data-fusion problem

Elias Bareinboim and Judea Pearl

PNAS July 5, 2016 113 (27) 7345-7352; published ahead of print July 5, 2016 https://doi.org/1

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved March 15, 20 June 29, 2015)

Inference is always *from something to something*.

a. Statistics
- from sample to distribution

b. Observational Causal Inference
- from observational distribution to experimental distribution

c. Sampling Selection Bias
- from study (obs/exp) distribution to general (obs/exp) distribution

d. General Transportability
- from (obs/exp) distributions of populations A, B, C… to experimental distribution of a target population

- We will model problems of selection bias and transportability.
- Formal language to represent the problem (nonparametrically), reduce it to an exercise of symbolic calculus, and derive complete solutions.

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y|do(x)],\ E^*[Y_x] = E^*[Y|do(x)]$ (on population Π^*)

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y|do(x)]$, $E^*[Y_x] = E^*[Y|do(x)]$ (on population Π^*)

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y|do(x)]$, $E^*[Y_x] = E^*[Y|do(x)]$ (on population Π^*)

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,*

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$$E[Y_x] = E[Y \mid do(x)],\ E^*[Y_x] = E^*[Y \mid do(x)] \text{ (on population } \Pi^*)$$

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,* $Y_{xz} = Y_x$, $P(Y \mid do(x), do(z)) = P(Y \mid do(x))$

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$$E[Y_x] = E[Y \mid do(x)],\ E^*[Y_x] = E^*[Y \mid do(x)] \text{ (on population } \Pi^*)$$

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,* $Y_{xz} = Y_x$, $P(Y \mid do(x), do(z)) = P(Y \mid do(x))$

Outputs:

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y \mid do(x)],\ E^*[Y_x] = E^*[Y \mid do(x)]$ (on population Π^*)

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,* $Y_{xz} = Y_x,\ \ P(Y \mid do(x), do(z)) = P(Y \mid do(x))$

Outputs:

1) Whether the data we have, plus what we already know is enough to answer what we want to know. And how.

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y|do(x)],\ E^*[Y_x] = E^*[Y|do(x)]$ (on population Π^*)

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,* $Y_{xz} = Y_x,\ \ P(Y|do(x), do(z)) = P(Y|do(x))$

Outputs:

1) Whether the data we have, plus what we already know is enough to answer what we want to know. And how.

2) Other logical ramifications of our assumptions (eg., test. implications)

Causal inference: causal models

Inputs:

1) What do we want to know? (Query)

*- A **property** of the **causal model** (ie, a causal parameter), eg: the expectation of Y if we experimentally set X to x, in a specific population.*

$E[Y_x] = E[Y|do(x)],\ E^*[Y_x] = E^*[Y|do(x)]$ (on population Π^*)

2) What data do we have? (Data)

- Observational? Experimental? Random sample? From which population?

3) What do we already know? (Causal Assumptions)

*- A **partial specification** of the **causal model,** eg: Z does not affect Y except through X,* $Y_{xz} = Y_x,\ \ P(Y|do(x), do(z)) = P(Y|do(x))$

Outputs:

We will go over each of these for TR and SB problems. But first a quick review.

1) Whether the data we have, plus what we already know is enough to answer what we want to know. And how.

2) Other logical ramifications of our assumptions (eg., test. implications)

Observational Causal Inference

Observational Distribution → Experimental Distribution

Observational causal inference

1) What do we want to know?

- *The distribution of Y if we experimentally set X to x*

$$P(Y_x) = P(Y|do(x))$$

Observational causal inference

1) What do we want to know?

- The distribution of Y if we experimentally set X to x

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

2) What data do we have?

- Observational data (joint distribution) from the population of interest

$P(Y, X, Z, \ldots)$

Observational causal inference

1) What do we want to know?

- The distribution of Y if we experimentally set X to x

$P(Y_x) = P(Y | do(x))$

2) What data do we have?

- Observational data (joint distribution) from the population of interest

$P(Y, X, Z, \ldots)$

3) What do we already know?

*- A **partial specification** of the **causal model:** exclusion restrictions, independence restrictions, parametric constraints.*

Observational causal inference

1) What do we want to know?

- The distribution of Y if we experimentally set X to x

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

2) What data do we have?

- Observational data (joint distribution) from the population of interest

$P(Y, X, Z, \ldots)$

3) What do we already know?

*- A **partial specification** of the **causal model:** exclusion restrictions, independence restrictions, parametric constraints.*

We need a language to formally represent what we want to know, the data we have and what we already know.

Observational causal inference

1) What do we want to know?

- The distribution of Y if we experimentally set X to x

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

2) What data do we have?

- Observational data (joint distribution) from the population of interest

$P(Y, X, Z, \ldots)$

3) What do we already know?

*- A **partial specification** of the **causal model:** exclusion restrictions, independence restrictions, parametric constraints.*

We need a language to formally represent what we want to know, the data we have and what we already know.

***Structural models:** combine the power of potential outcomes, structural equations, and graphs.*

The structural model

The ***structural model is our oracle.*** With a fully specified structural model we can answer ***any*** causal or counterfactual question.

Functional assignments

$$M: \quad Z = f_z(U_z)$$
$$X = f_x(Z, U_x)$$
$$Y = f_y(X, Z, U_y)$$

Distribution unobserved factors

$$P: \quad P(U_z, U_x, U_y)$$

Causal (and counterfactual) quantities are ***defined in terms of our model.***

$$M_x: \quad Z = f_z(U_z)$$

do(x) $X = x$

$$Y = f_y(X, Z, U_y)$$

$$E[Y_x] = E[Y \mid do(x)] = E[f_y(x, Z, U_y)]$$

The expectation of Y in the modified model where X is experimentally set to x.

The structural model

The ***structural model is our oracle.*** With a fully specified structural model we can answer ***any*** causal or counterfactual question.

Functional assignments **Distribution unobserved factors**

$$M: \quad Z = f_z(U_z)$$
$$X = f_x(Z, U_x)$$
$$Y = f_y(X, Z, U_y)$$

$$P: \quad P(U_z, U_x, U_y)$$

Causal (and counterfactual) quantities are ***defined in terms of our model.***

$$M_x: \quad Z = f_z(U_z)$$
do(x) $X = x$
$$Y = f_y(X, Z, U_y)$$

$$E[Y_x] = E[Y \mid do(x)] = E[f_y(x, Z, U_y)]$$

The expectation of Y in the modified model where X is experimentally set to x.

In most cases we don't have a fully specified model, but only a partial understanding of what is going on. How can we encode that knowledge?

Encoding what we know: causal diagrams

Causal diagrams provide a nonparametric, qualitative partial specification of a causal model. In its basic form, it encodes:

Encoding what we know: causal diagrams

Causal diagrams provide a nonparametric, qualitative partial specification of a causal model. In its basic form, it encodes:

1. *Absence of direct effects between variables (exclusion restrictions);*

Encoding what we know: causal diagrams

Causal diagrams provide a nonparametric, qualitative partial specification of a causal model. In its basic form, it encodes:

1. *Absence of direct effects between variables (exclusion restrictions);*
2. *Absence of unobserved common causes between variables (independence restrictions).*

Encoding what we know: causal diagrams

Causal diagrams provide a nonparametric, qualitative partial specification of a causal model. In its basic form, it encodes:

1. *Absence of direct effects between variables (exclusion restrictions);*
2. *Absence of unobserved common causes between variables (independence restrictions).*

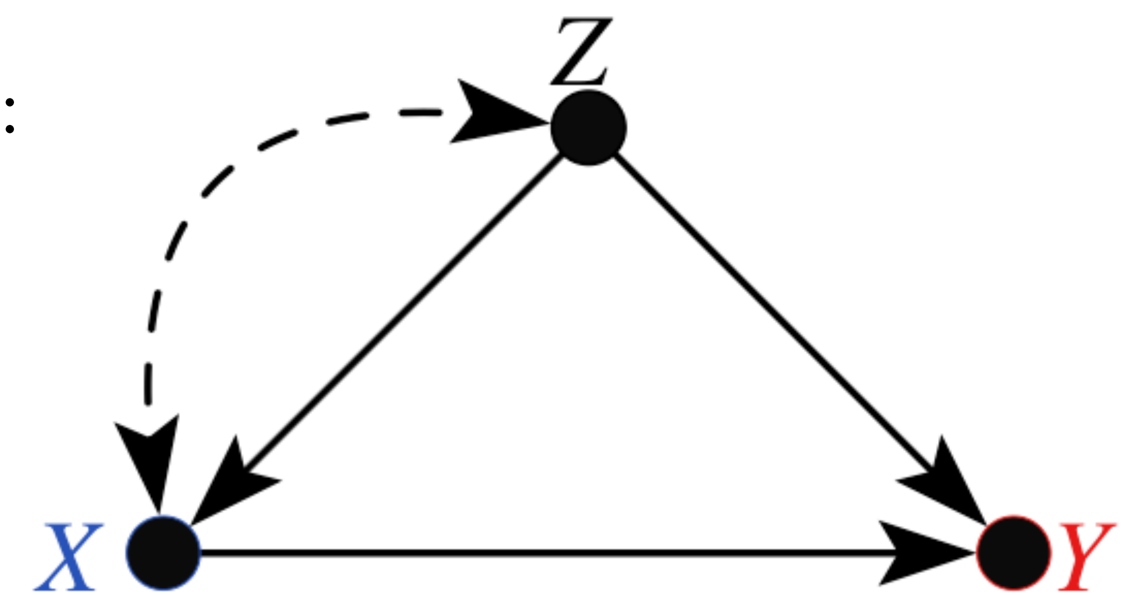
Functional assignments

$$M: \quad Z = f_z(U_z)$$
$$X = f_x(Z, U_x)$$
$$Y = f_y(X, Z, U_y)$$

Distribution unobserved factors

$$P: P(U_z, U_x, U_y) = P(U_z, U_x)P(U_y)$$

$G:$



Encoding what we know: causal diagrams

Causal diagrams provide a nonparametric, qualitative partial specification of a causal model. In its basic form, it encodes:

1. *Absence of direct effects between variables (exclusion restrictions);*
2. *Absence of unobserved common causes between variables (independence restrictions).*

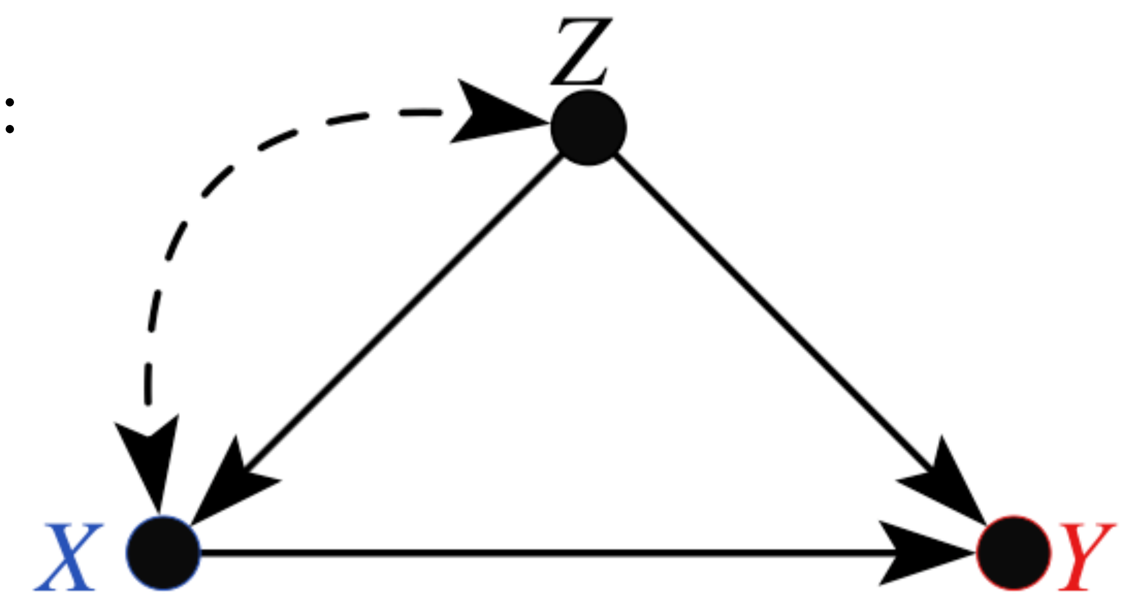
Functional assignments

$$M: \quad Z = f_z(U_z)$$
$$X = f_x(Z, U_x)$$
$$Y = f_y(X, Z, U_y)$$

Distribution unobserved factors

$$P: P(U_z, U_x, U_y) = P(U_z, U_x)P(U_y)$$

$G:$

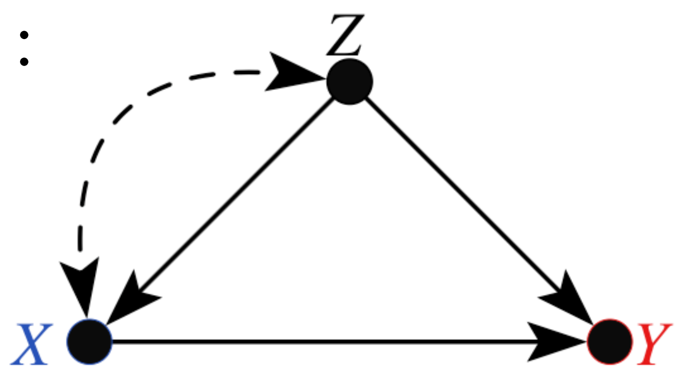


The question of whether our partial understanding + the data we have is sufficient for answering our query is known as the ***identification problem.***

The identification problem

We have data from P(Y, X, Z):

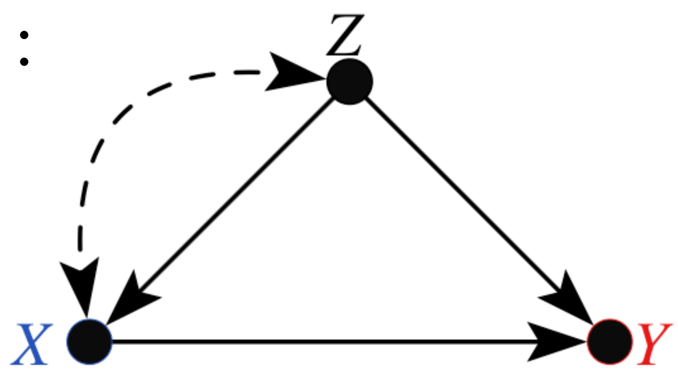
$G:$



The identification problem

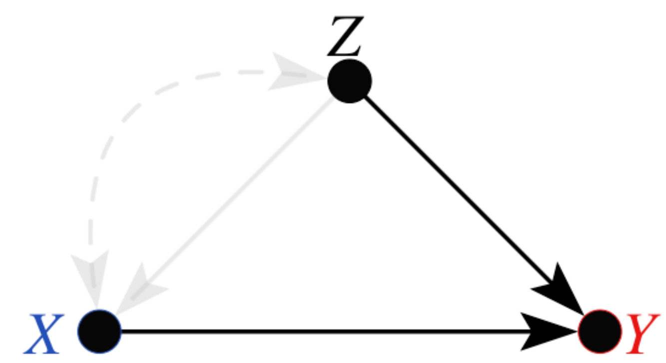
We have data from P(Y, X, Z):

$G:$



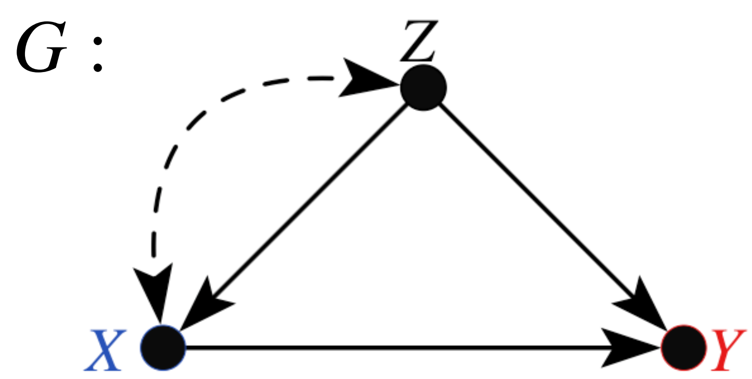
Want to make inference about P(Y|do(X)):

$G_{\overline{X}}:$

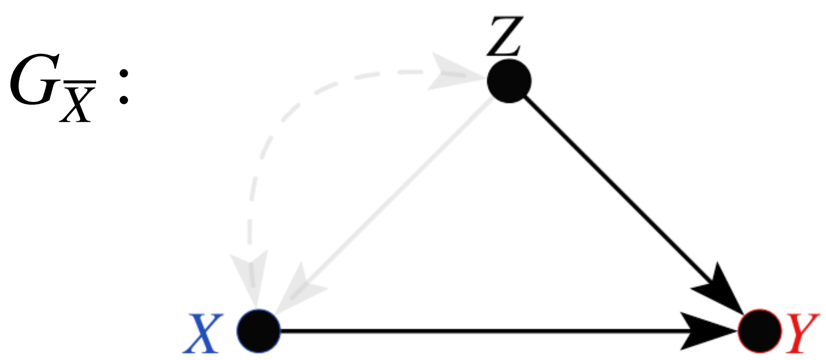


The identification problem

We have data from P(Y, X, Z):



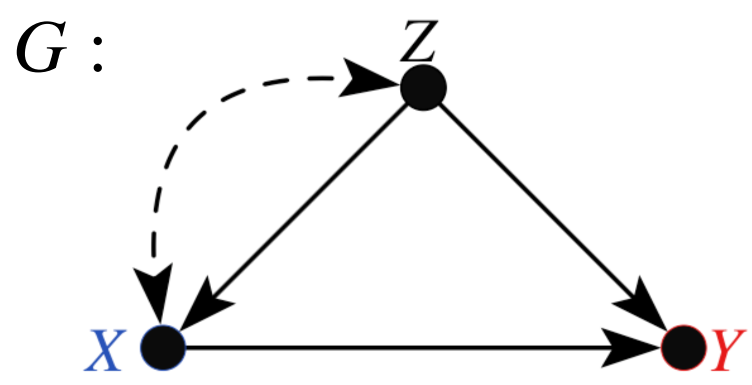
Want to make inference about P(Y|do(X)):



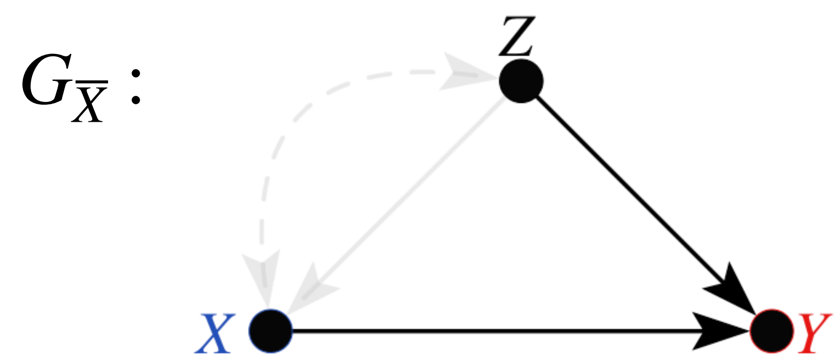
Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):

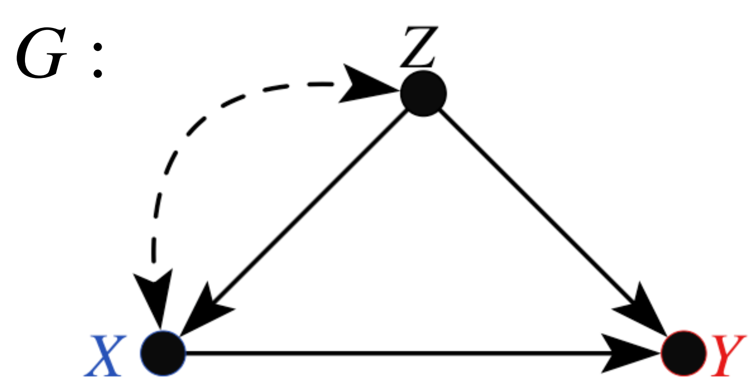


Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

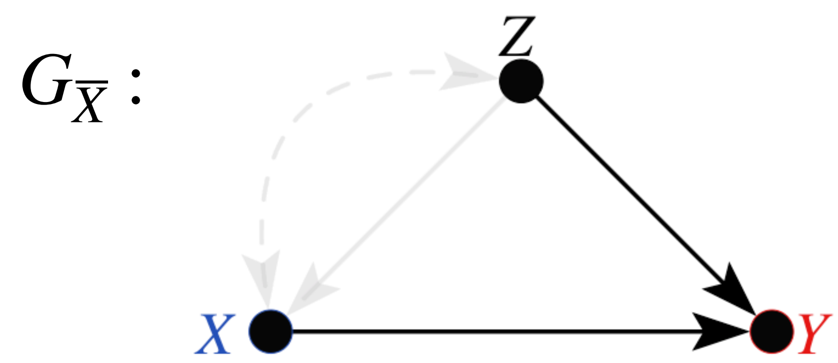
$$
\begin{aligned}
P(y \mid do(x)) &= \sum_z P(y \mid do(x), z) P(z \mid do(x)) \\
&= \sum_z P(y \mid x, z) P(z)
\end{aligned}
$$

The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):

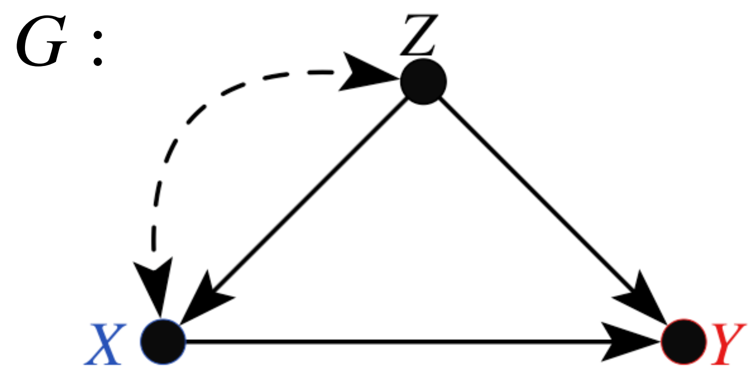


Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

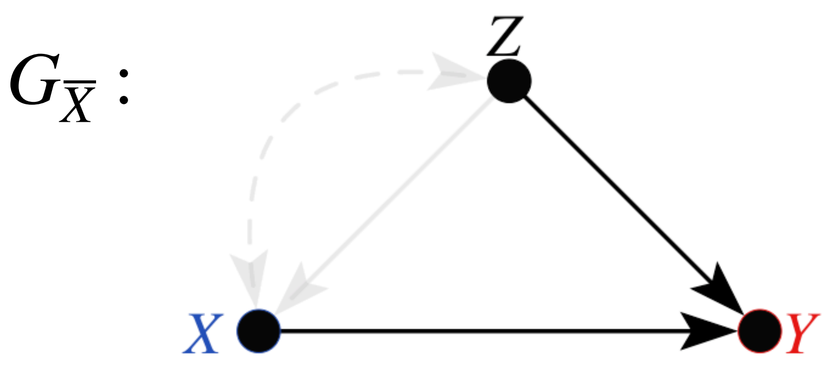
$$P(y \mid do(x)) = \sum_z P(y \mid do(x), z) P(z \mid do(x))$$
$$= \sum_z P(y \mid x, z) P(z)$$

The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):



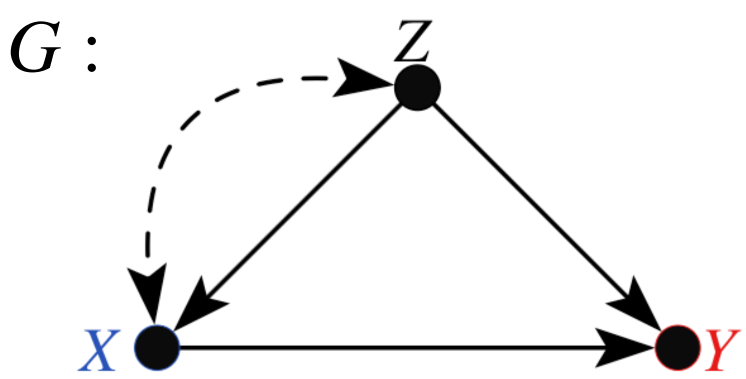
Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

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

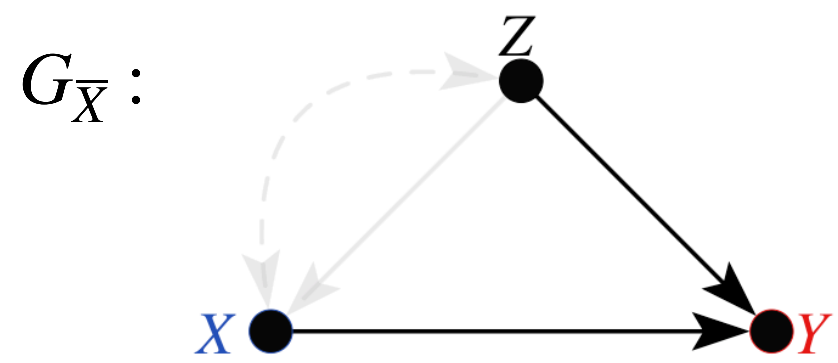
$$= \sum_z P(y \mid x, z) P(z)$$

The identification problem

We have data from P(Y, X, Z):



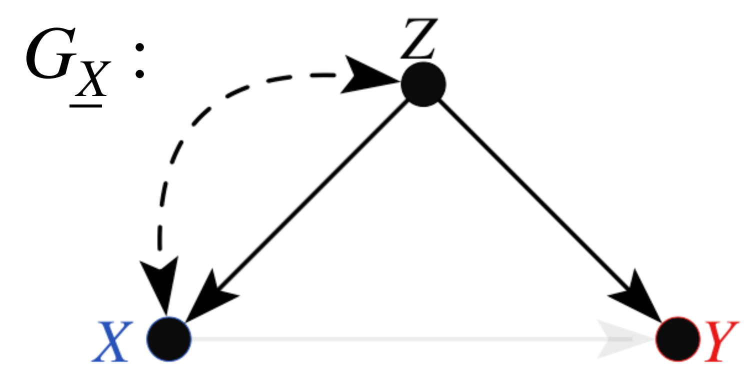
Want to make inference about P(Y|do(X)):



Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

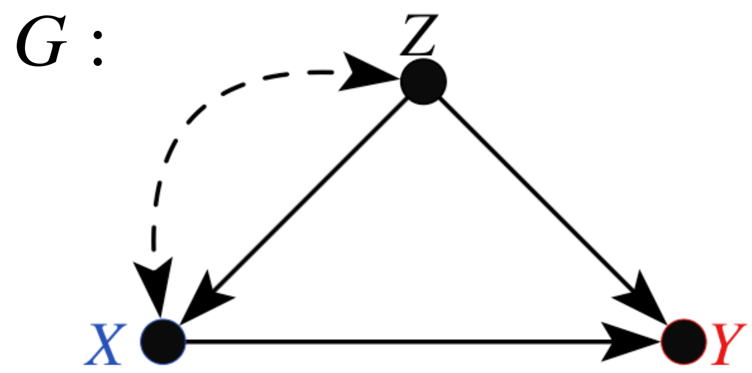
$$P(y \mid do(x)) = \sum_z P(y \mid do(x), z) P(z \mid do(x))$$

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

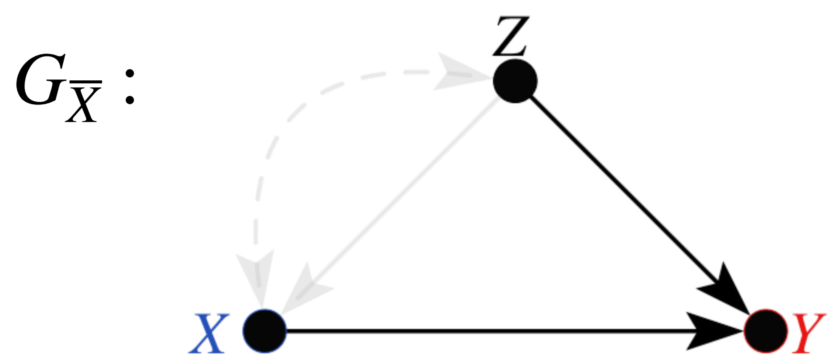


The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):



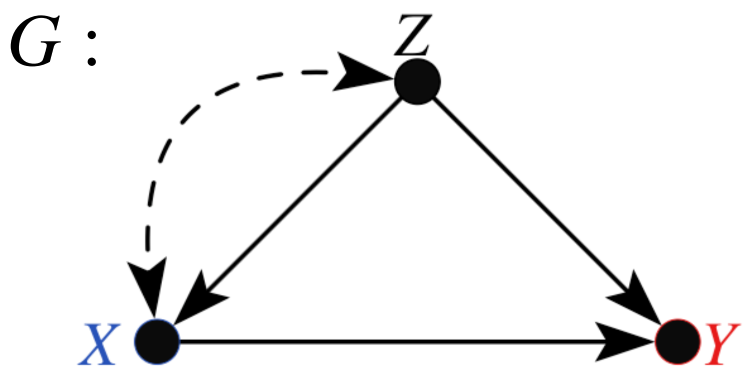
Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

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

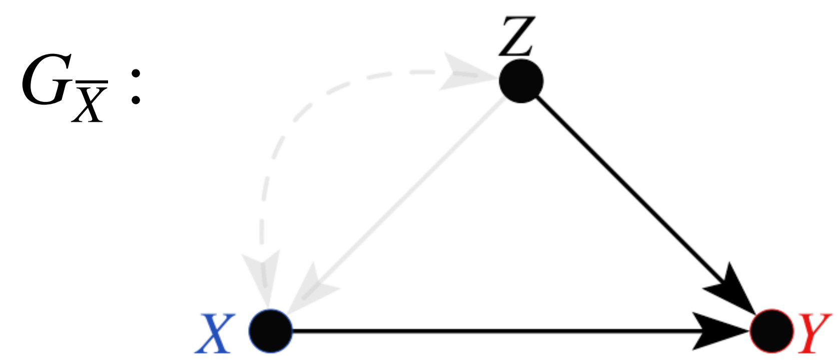
$$= \sum_z P(y \mid x, z) P(z)$$

The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):



Task: *to express* $P(Y \mid do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

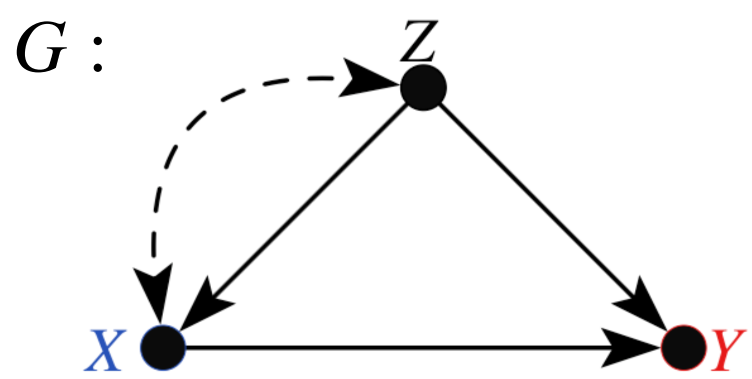
$$P(y \mid do(x)) = \sum_z P(y \mid do(x), z) P(z \mid do(x))$$

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

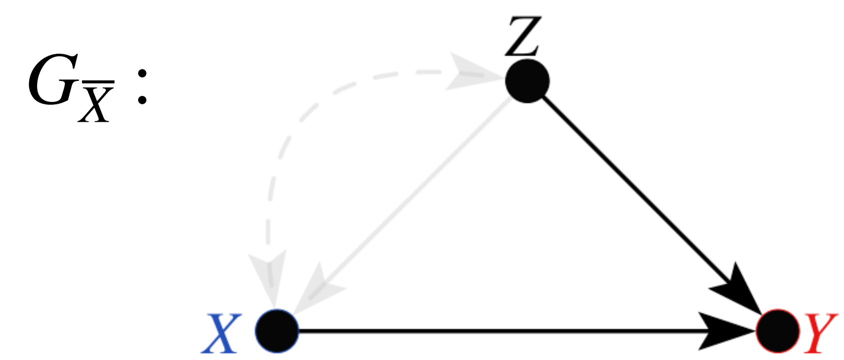
$$\left.\begin{array}{l} (Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \implies Y_x \perp\!\!\!\perp X \mid Z_x \\ (Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies Z_x = Z \end{array}\right. \implies (Y_x \perp\!\!\!\perp X \mid Z)$$

The identification problem

We have data from P(Y, X, Z):



Want to make inference about P(Y|do(X)):



Task: *to express* $P(Y|do(x)) = P(Y_x)$ *in terms of* $P(Y, X, Z)$. ***Symbolically,*** *this amounts to removing do() operators or counterfactual subscripts;* ***Graphically,*** *licensing assumptions checked via d-sep. in modified graphs.*

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

$$P(y_x) = \sum_z P(y_x|z)P(z) = \sum_z P(y_x|x, z)P(z) = \sum_z P(y|x, z)P(z)$$

$$\begin{array}{l} (Y \perp\!\!\!\perp X|Z)_{G_{\underline{X}}} \implies Y_x \perp\!\!\!\perp X|Z_x \\ (Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies Z_x = Z \end{array} \implies (Y_x \perp\!\!\!\perp X|Z)$$

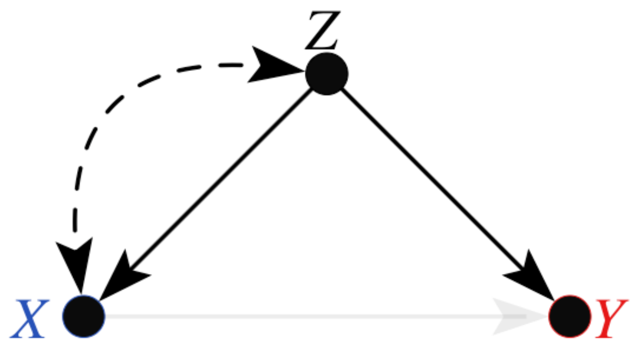
Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

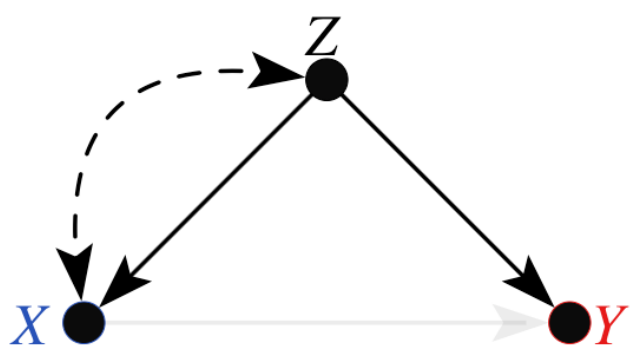
$G_{\underline{X}}$:



Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

$G_{\underline{X}}$:

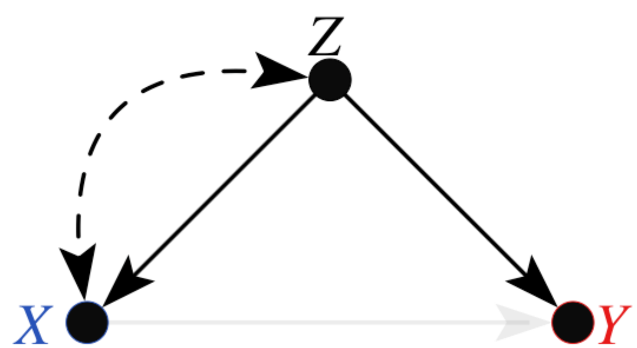


$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \implies P(y \mid do(x), z) = P(y \mid x, z)$$

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

$G_{\underline{X}}$:



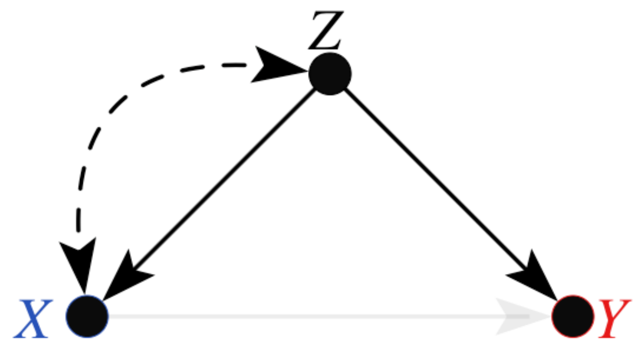
$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \implies P(y \mid do(x), z) = P(y \mid x, z)$$

If you block all confounding paths, seeing = doing.
(= checking indep. restriction)

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

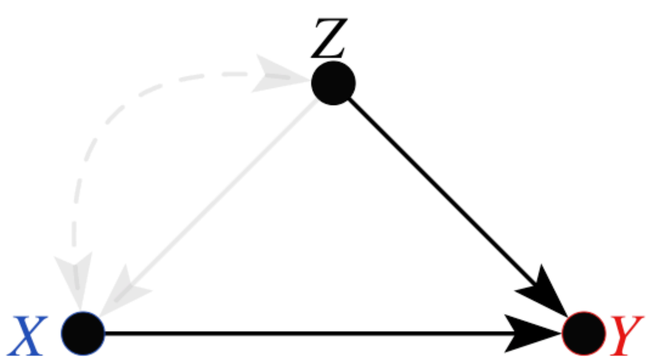
$G_{\underline{X}}$:



$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \implies P(y \mid do(x), z) = P(y \mid x, z)$$

If you block all confounding paths, seeing = doing.
(= checking indep. restriction)

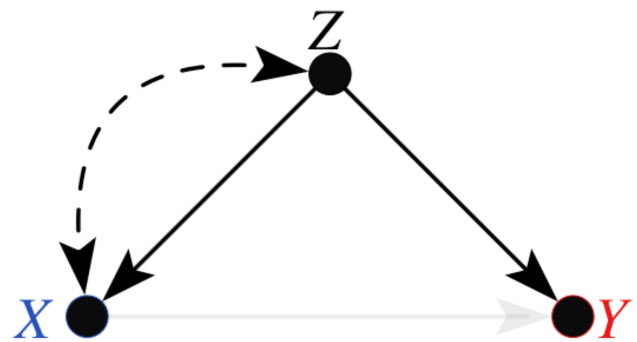
$G_{\overline{X}}$:



Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

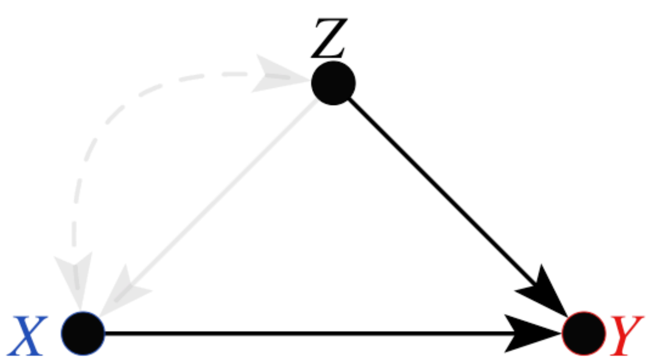
$G_{\underline{X}}$:



$$(Y \perp\!\!\!\perp X \mid Z)_{G_{\underline{X}}} \implies P(y \mid do(x), z) = P(y \mid x, z)$$

If you block all confounding paths, seeing = doing.
(= checking indep. restriction)

$G_{\overline{X}}$:

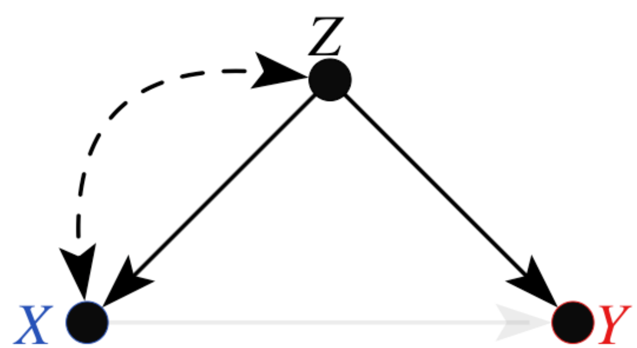


$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies P(z \mid do(x)) = P(z)$$

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

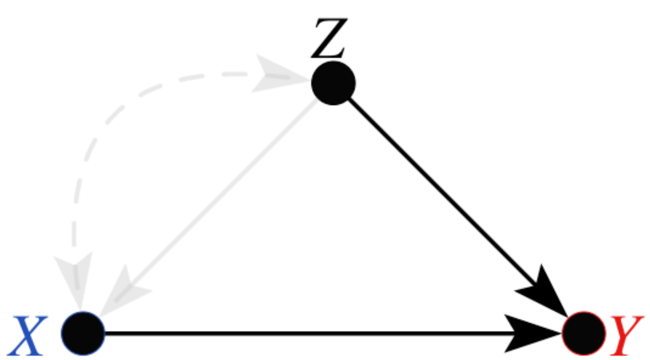
$G_{\underline{X}}$:



$$(Y \perp\!\!\!\perp X | Z)_{G_{\underline{X}}} \implies P(y | do(x), z) = P(y | x, z)$$

If you block all confounding paths, seeing = doing.
(= checking indep. restriction)

$G_{\overline{X}}$:



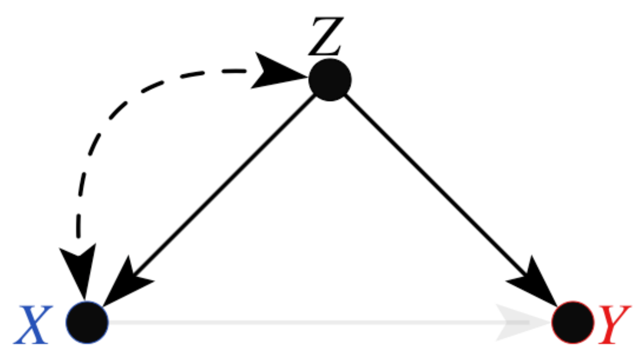
$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies P(z | do(x)) = P(z)$$

You can drop/include actions if there is no causal path from manipulated variable to target variable.
(= checking exclusion restriction)

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

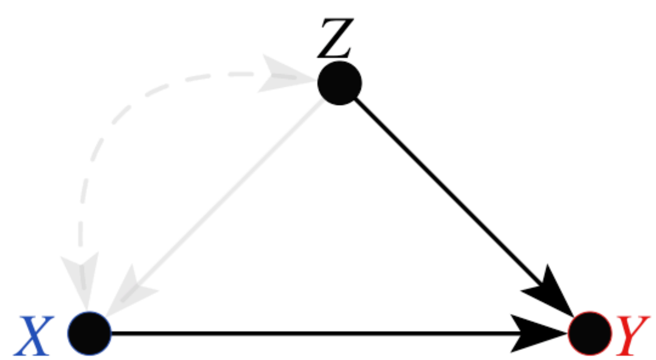
$G_{\underline{X}}$:



$$(Y \perp\!\!\!\perp X | Z)_{G_{\underline{X}}} \implies P(y | do(x), z) = P(y | x, z)$$

If you block all confounding paths, seeing = doing.
(= checking indep. restriction)

$G_{\overline{X}}$:



$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies P(z | do(x)) = P(z)$$

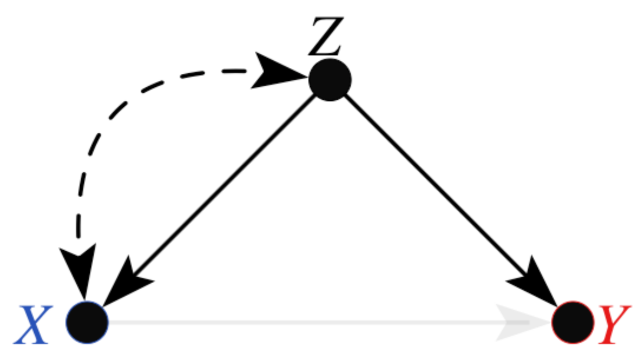
You can drop/include actions if there is no causal path from manipulated variable to target variable.
(= checking exclusion restriction)

This is the do-calculus (rule 1 can be derived from these two). Any identifiable causal effect can be derived via an application of those simple rules.

Complete solution: do-calculus

The previous derivation showcases the (simplified) manipulation rules you need to know for massaging causal expressions (+ basic probability theory).

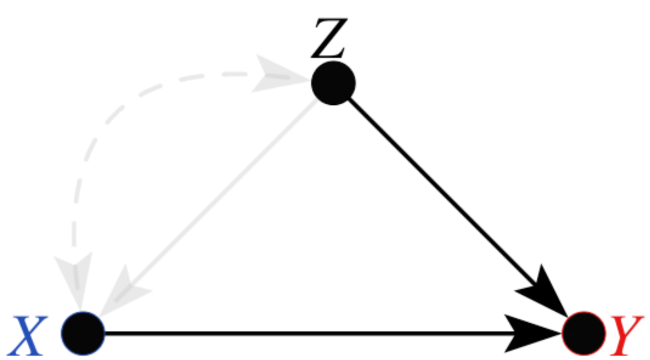
$G_{\underline{X}}$:



$$(Y \perp\!\!\!\perp X | Z)_{G_{\underline{X}}} \implies P(y | do(x), z) = P(y | x, z)$$

If you block all confounding paths, seeing = doing. (= checking indep. restriction)

$G_{\overline{X}}$:



$$(Z \perp\!\!\!\perp X)_{G_{\overline{X}}} \implies P(z | do(x)) = P(z)$$

You can drop/include actions if there is no causal path from manipulated variable to target variable. (= checking exclusion restriction)

This is the do-calculus (rule 1 can be derived from these two). Any identifiable causal effect can be derived via an application of those simple rules.

We also have ***complete algorithms:*** *completeness assures us that, if we can't find a solution, it is impossible to identify the effect* ***without extra assumptions.*** That is, no other method can do better.

Internal validity vs external validity

The previous discussion concerns obtaining a valid estimand for the causal effect in the specific population at hand– also known as “***internal validity”.***

Internal validity vs external validity

The previous discussion concerns obtaining a valid estimand for the causal effect in the specific population at hand– also known as "***internal validity".***

But science is about generalization: studies are usually done with the aim of being applicable to new settings. This is usually denoted by ***"external validity" or "generalizability".***

Internal validity vs external validity

The previous discussion concerns obtaining a valid estimand for the causal effect in the specific population at hand– also known as "***internal validity"*.**

But science is about generalization: studies are usually done with the aim of being applicable to new settings. This is usually denoted by ***"external validity" or "generalizability"*.**

Question: Is it possible to predict the effect of X on Y in a ***target population*** using data ***learned from experiments elsewhere, under different conditions?***

Internal validity vs external validity

The previous discussion concerns obtaining a valid estimand for the causal effect in the specific population at hand– also known as “***internal validity”.***

But science is about generalization: studies are usually done with the aim of being applicable to new settings. This is usually denoted by ***“external validity” or “generalizability”.***

Question: Is it possible to predict the effect of X on Y in a ***target population*** using data ***learned from experiments elsewhere, under different conditions?***

Answer: sometimes, yes.

Internal validity vs external validity

The previous discussion concerns obtaining a valid estimand for the causal effect in the specific population at hand– also known as "***internal validity".***

But science is about generalization: studies are usually done with the aim of being applicable to new settings. This is usually denoted by ***"external validity" or "generalizability".***

Question: Is it possible to predict the effect of X on Y in a ***target population*** using data ***learned from experiments elsewhere, under different conditions?***

Answer: sometimes, yes.

Our goal: extend our modeling tools to formally characterize ***when*** and ***how.***

Transportability

(exp/obs) dist pop A, B, …→ (exp/obs) dist target pop

Generalizability, External Validity… ?

- *"'External validity' asks the question of generalizability: To **what populations, settings**, treatment variables, and measurement variables **can this effect be generalized?"***
- Shadish, Cook and Campbell (2002)

Generalizability, External Validity… ?

- *"'External validity' asks the question of generalizability: To* ***what populations, settings****, treatment variables, and measurement variables* ***can this effect be generalized?"***
- Shadish, Cook and Campbell (2002)

- *"Extrapolation across studies requires '****some understanding of the reasons for the differences.****'"*
- Cox (1958)

Generalizability, External Validity… ?

- *"'External validity' asks the question of generalizability: To **what populations, settings**, treatment variables, and measurement variables **can this effect be generalized?"***

Shadish, Cook and Campbell (2002)

- *"Extrapolation across studies requires '**some understanding of the reasons for the differences.**'"*

Cox (1958)

*"An experiment is said to have "external validity" if the **distribution of outcomes realized by a treatment group** is the **same as the distribution of outcome that would be realized in an actual program**."*

Manski (2007)

Generalizability, External Validity… ?

- *"'External validity' asks the question of generalizability: To **what populations, settings**, treatment variables, and measurement variables **can this effect be generalized?"***

Shadish, Cook and Campbell (2002)

- *"Extrapolation across studies requires '**some understanding of the reasons for the differences.**'"*

Cox (1958)

*"An experiment is said to have "external validity" if the **distribution of outcomes realized by a treatment group** is the **same as the distribution of outcome that would be realized in an actual program**."*

Manski (2007)

How can we operationalize this?

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

1) ***What do we want to know? (Query)***

- *Causal effect on target population:* $E^*[Y \mid do(x)]$

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

1) ***What do we want to know? (Query)***

- *Causal effect on target population:* $E^*[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- *Obs./Exp. on source; obs. on target:* $P(V), P(V \mid do(z)), P^*(V)$

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

1) ***What do we want to know? (Query)***

- *Causal effect on target population:* $E^*[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- *Obs./Exp. on source; obs. on target:* $P(V), P(V \mid do(z)), P^*(V)$

3) ***What do we already know? (Causal Assumptions)***

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

1) ***What do we want to know? (Query)***

- *Causal effect on target population:* $E^*[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- *Obs./Exp. on source; obs. on target:* $P(V), P(V \mid do(z)), P^*(V)$

3) ***What do we already know? (Causal Assumptions)***

- *Need to* ***encode disparities/commonalities*** *between environments*

Formalizing transportability

Let us start with only two populations (more later):

$$\Pi = \langle P, M \rangle : \text{source population} \qquad \Pi^* = \langle P^*, M^* \rangle : \text{target population}$$

1) ***What do we want to know? (Query)***

   - Causal effect on target population: $E^*[Y \mid do(x)]$

2) ***What data do we have? (Data)***

   - Obs./Exp. on source; obs. on target: $P(V), P(V \mid do(z)), P^*(V)$

3) ***What do we already know? (Causal Assumptions)***

   *- Need to **encode disparities/commonalities** between environments*

   *- Our approach will be **nonparametric**, requiring only a **qualitative description of which mechanisms are suspected to be different***

Encoding disparities: selection nodes

We will extend our causal diagram with “selection nodes” (S) which indicates ***structural discrepancies*** between populations.

Encoding disparities: selection nodes

We will extend our causal diagram with "selection nodes" (S) which indicates ***structural discrepancies*** between populations.

Switching between the two populations is represented by conditioning on different values of S (or simply conditioning or not conditioning on S).

Encoding disparities: selection nodes

We will extend our causal diagram with "selection nodes" (S) which indicates ***structural discrepancies*** between populations.

Switching between the two populations is represented by conditioning on different values of S (or simply conditioning or not conditioning on S).

For instance, if P(y | do(x)) represents the experimental distribution of Y in the source domain Π and P*(y | do(x)) the experimental distribution of Y in the target domain Π^*, the selection node act as a "switcher", and accounts for *any discrepancy between the two populations. That is, by definition,*

Encoding disparities: selection nodes

We will extend our causal diagram with "selection nodes" (S) which indicates ***structural discrepancies*** between populations.

Switching between the two populations is represented by conditioning on different values of S (or simply conditioning or not conditioning on S).

For instance, if P(y | do(x)) represents the experimental distribution of Y in the source domain Π and P*(y | do(x)) the experimental distribution of Y in the target domain Π^*, the selection node act as a "switcher", and accounts for *any discrepancy between the two populations. That is, by definition,*

$$P^*(y \mid do(x)) = P(y \mid do(x), s)$$

Encoding disparities: selection nodes

We will extend our causal diagram with "selection nodes" (S) which indicates ***structural discrepancies*** between populations.

Switching between the two populations is represented by conditioning on different values of S (or simply conditioning or not conditioning on S).

For instance, if P(y | do(x)) represents the experimental distribution of Y in the source domain Π and P*(y | do(x)) the experimental distribution of Y in the target domain Π^*, the selection node act as a "switcher", and accounts for *any discrepancy between the two populations. That is, by definition,*

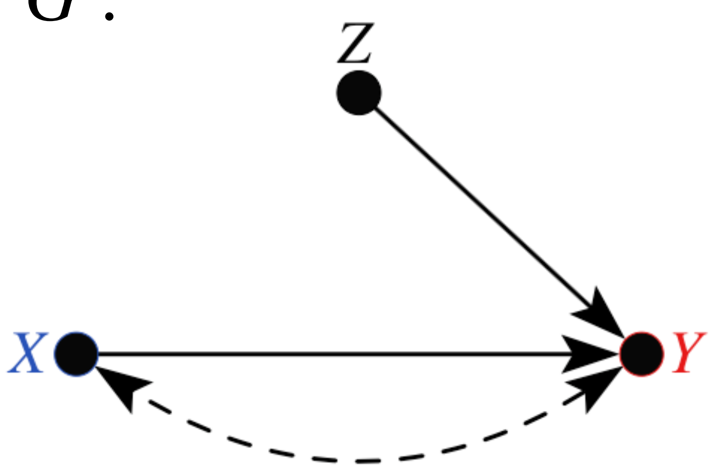
$$P^*(y \mid do(x)) = P(y \mid do(x), s)$$

Thus, ***symbolically***, our task is to ***remove conditioning on S on any do() expression*** (or counterfactual expression), since we do not have experimental data on the target domain.

Encoding disparities: selection diagrams

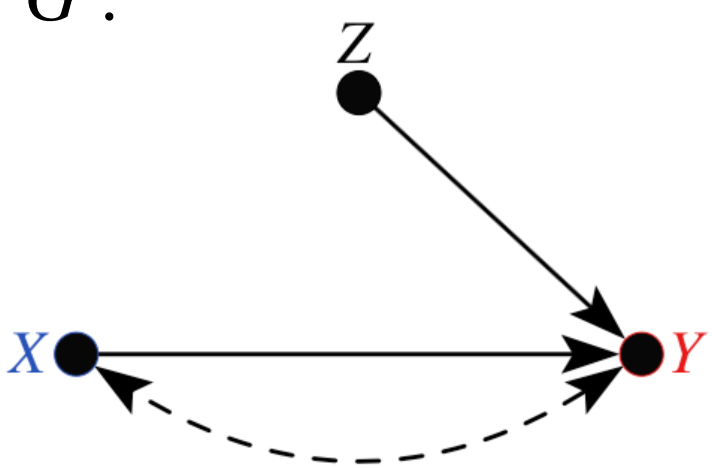
Encoding disparities: selection diagrams

$G:$

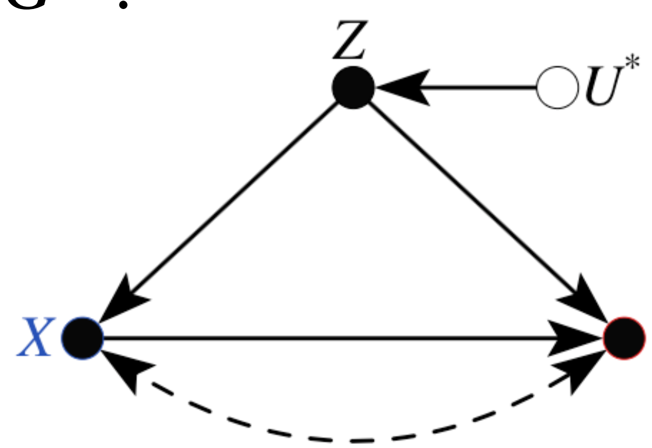


Encoding disparities: selection diagrams

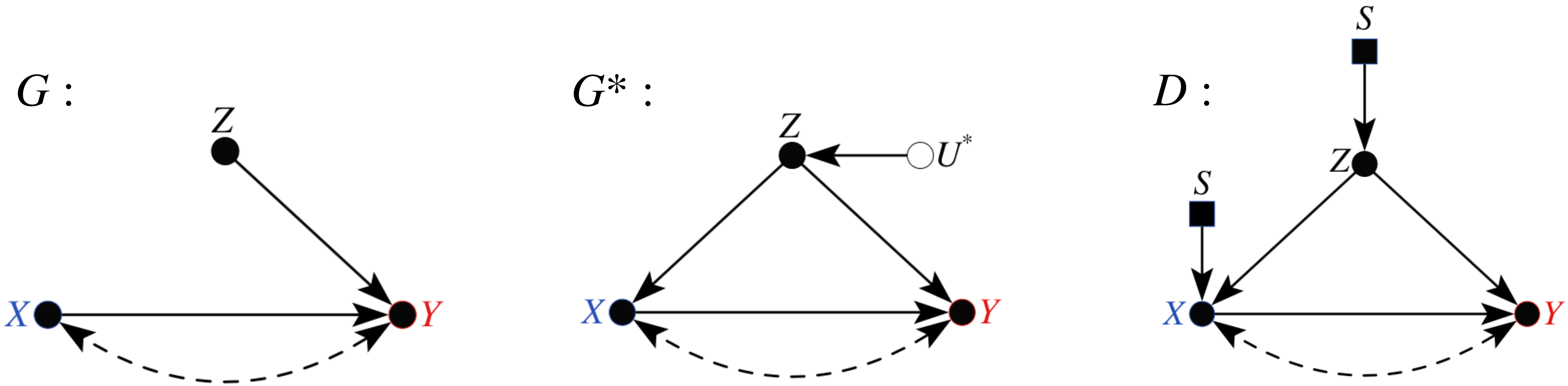
G :



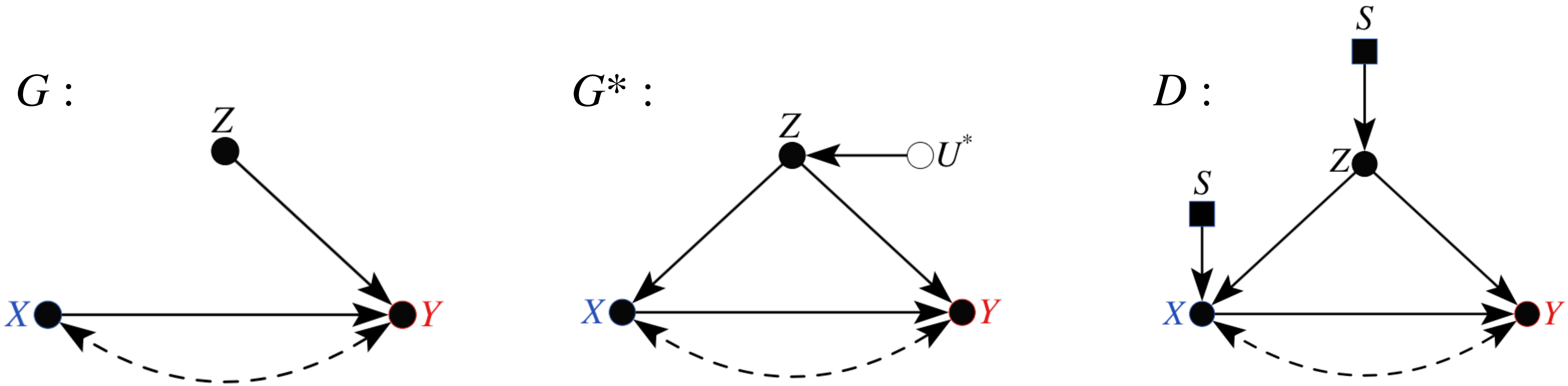
G^* :



Encoding disparities: selection diagrams

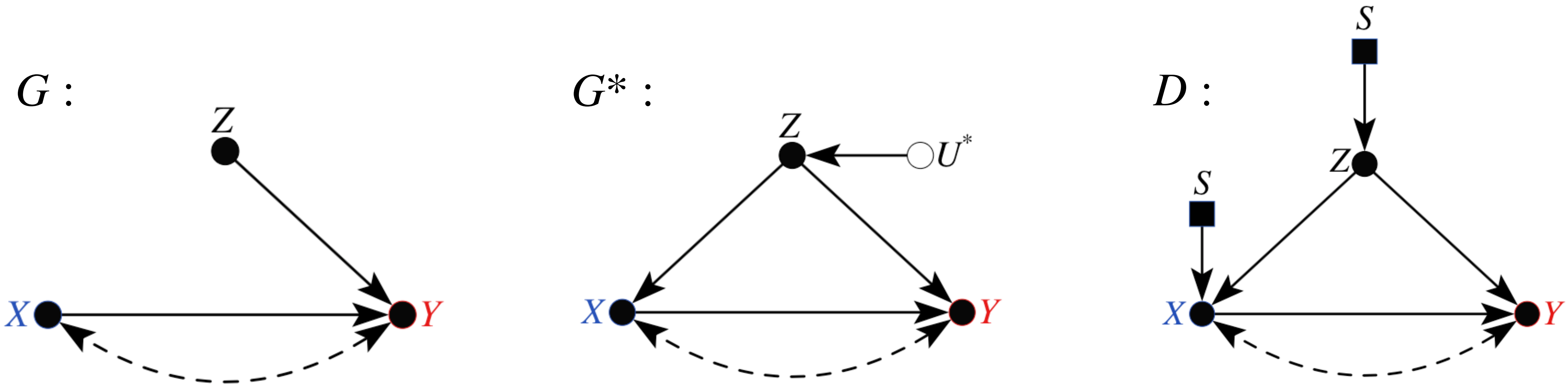


Encoding disparities: selection diagrams



The presence of an edge S→ Z means the ***local mechanism*** *that assigns values to Z* may be ***different,*** $f_z \neq f_z^*$ or $P(U_z) \neq P^*(U_z)$, between populations.

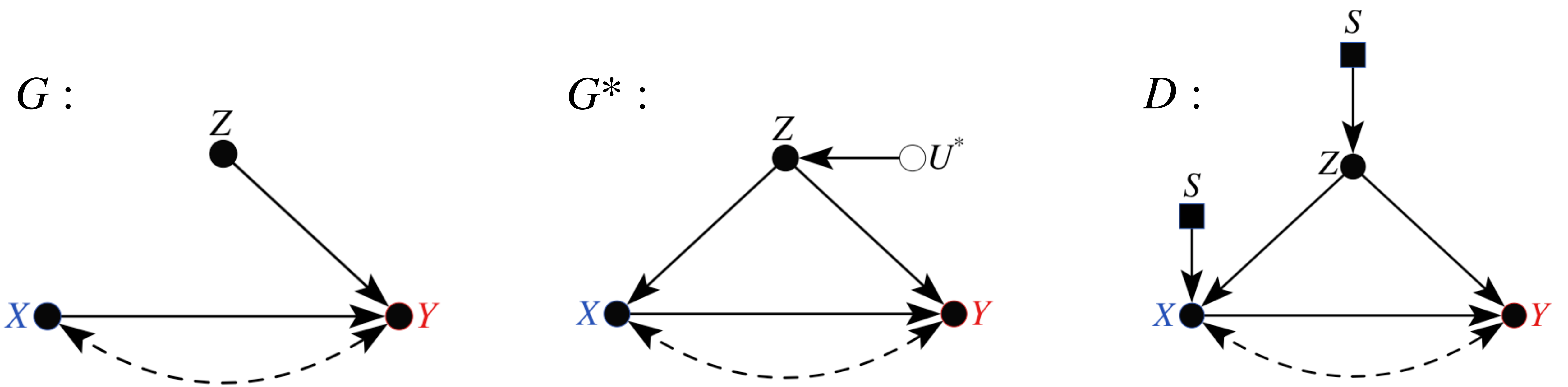
Encoding disparities: selection diagrams



The presence of an edge S→ Z means the ***local mechanism*** *that assigns values to Z* may be ***different,*** $f_z \neq f_z^*$ or $P(U_z) \neq P^*(U_z)$, between populations.

Conversely, ***absence*** of an edge S → Y represents the ***assumption*** that the ***local mechanism*** *that assigns values to Y* ***is the same*** in both populations.

Encoding disparities: selection diagrams

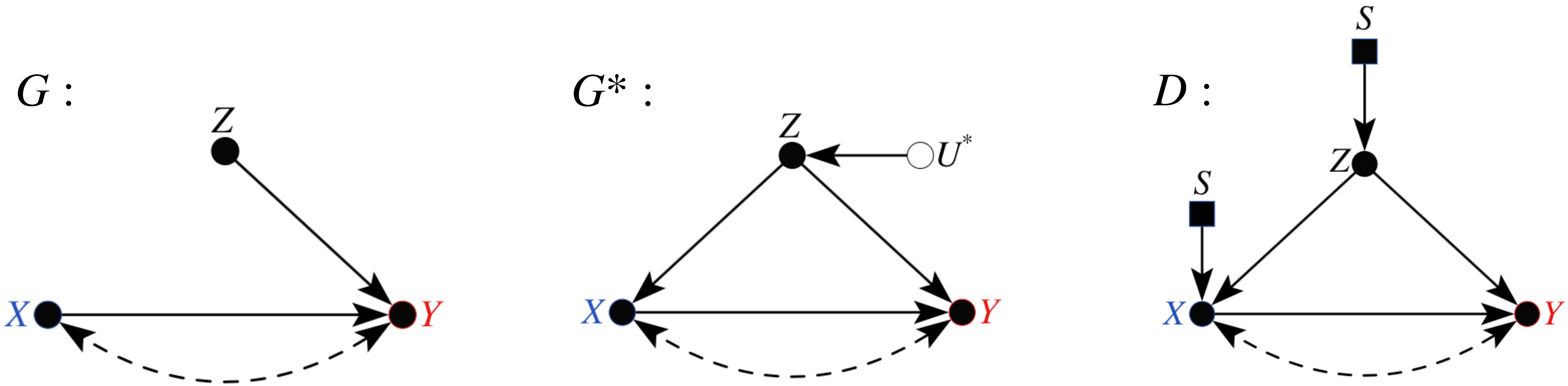


The presence of an edge S→ Z means the ***local mechanism*** *that assigns values to Z* may be ***different,*** $f_z \neq f_z^*$ or $P(U_z) \neq P^*(U_z)$, between populations.

Conversely, ***absence*** of an edge S → Y represents the ***assumption*** that the ***local mechanism*** *that assigns values to Y* ***is the same*** in both populations.

Thus, ***graphically***, we will check for ***separation of the source of discrepancy (S)*** from key variables in the terms that describe out target quantity .

Encoding disparities: selection diagrams



The presence of an edge S→ Z means the ***local mechanism*** *that assigns values to Z* may be ***different,*** $f_z \neq f_z^*$ or $P(U_z) \neq P^*(U_z)$, between populations.

Conversely, ***absence*** of an edge S → Y represents the ***assumption*** that the ***local mechanism*** *that assigns values to Y* ***is the same*** in both populations.

Thus, ***graphically***, we will check for ***separation of the source of discrepancy (S)*** from key variables in the terms that describe out target quantity .

For clarity, selection nodes (S) are represented by square nodes (▪).

Finding invariances: the basics

1) ***Trivial transportability***

- Effect estimable directly from obs. distribution in target (vanilla identification)

Finding invariances: the basics

1) ***Trivial transportability***

- Effect estimable directly from obs. distribution in target (vanilla identification)

2) ***Direct transportability***

- Transportable directly from source to target (Manksi called "external validity")

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

Finding invariances: the basics

1) ***Trivial transportability***

- Effect estimable directly from obs. distribution in target (vanilla identification)

2) ***Direct transportability***

- Transportable directly from source to target (Manksi called "external validity")

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- Reduced to checking d-separation in selection diagram

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$

Finding invariances: the basics

1) ***Trivial transportability***

- *Effect estimable directly from obs. distribution in target (vanilla identification)*

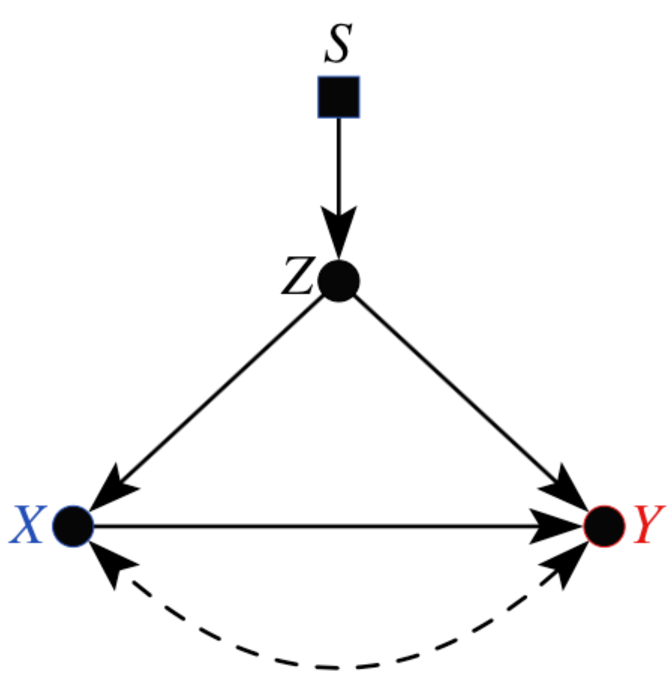
2) ***Direct transportability***

- *Transportable directly from source to target (Manksi called "external validity")*

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- *Reduced to checking d-separation in selection diagram*

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$



Finding invariances: the basics

1) ***Trivial transportability***

- *Effect estimable directly from obs. distribution in target (vanilla identification)*

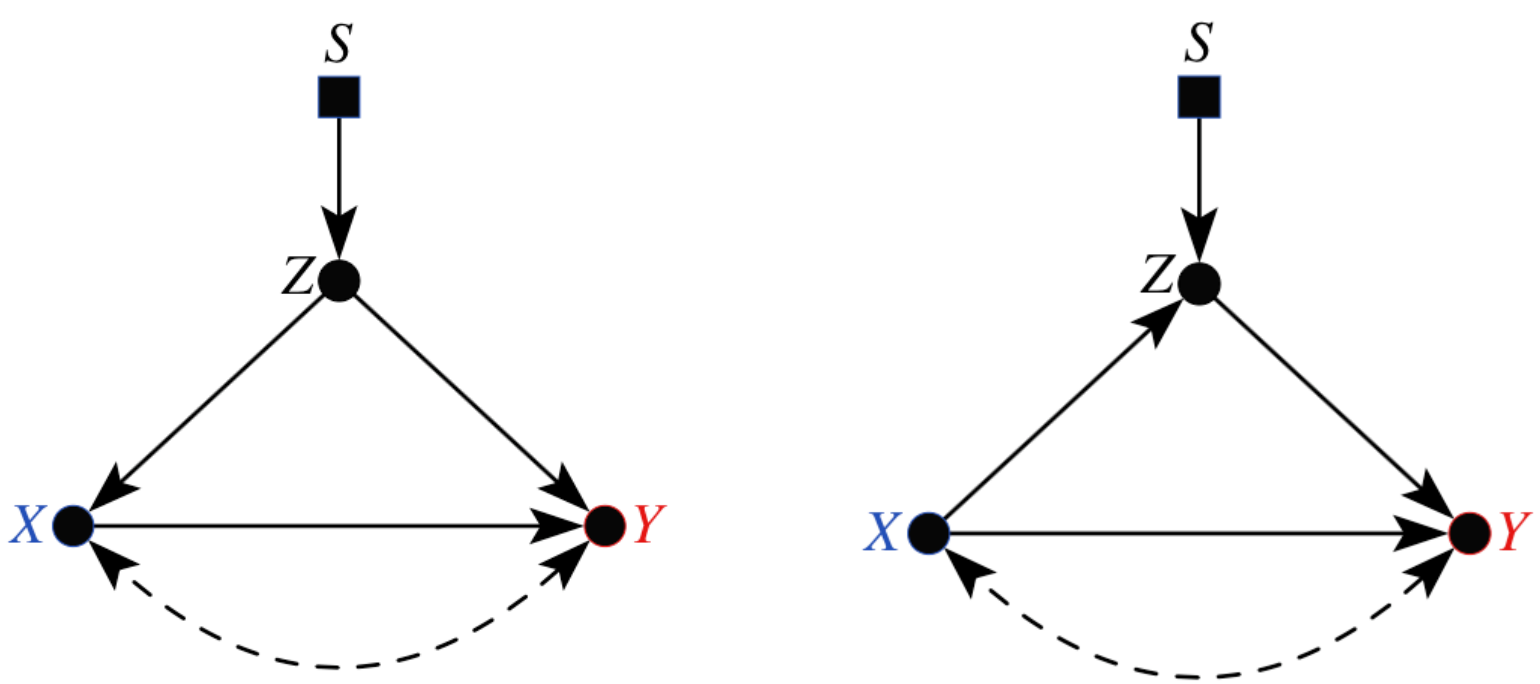
2) ***Direct transportability***

- *Transportable directly from source to target (Manksi called "external validity")*

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- *Reduced to checking d-separation in selection diagram*

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$



Finding invariances: the basics

1) ***Trivial transportability***

- *Effect estimable directly from obs. distribution in target (vanilla identification)*

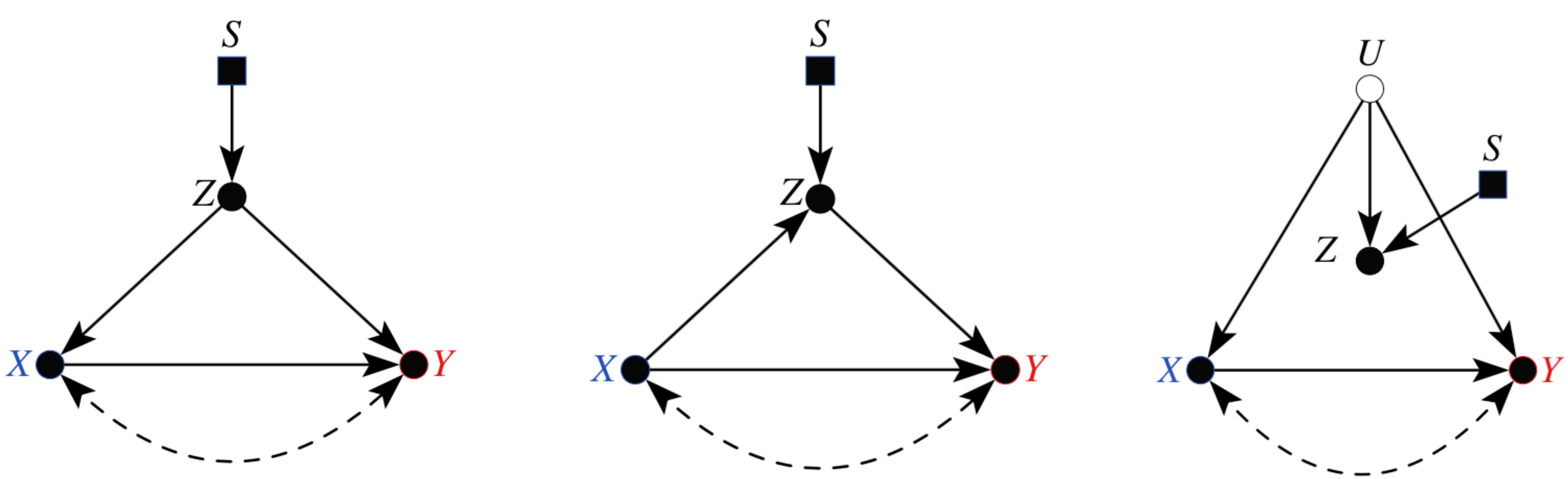
2) ***Direct transportability***

- *Transportable directly from source to target (Manksi called "external validity")*

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- *Reduced to checking d-separation in selection diagram*

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$



Finding invariances: the basics

1) ***Trivial transportability***

- Effect estimable directly from obs. distribution in target (vanilla identification)

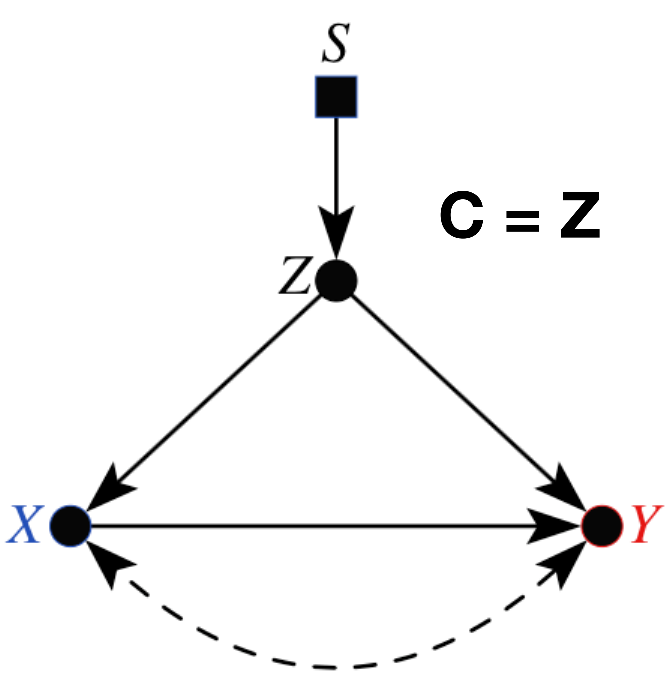
2) ***Direct transportability***

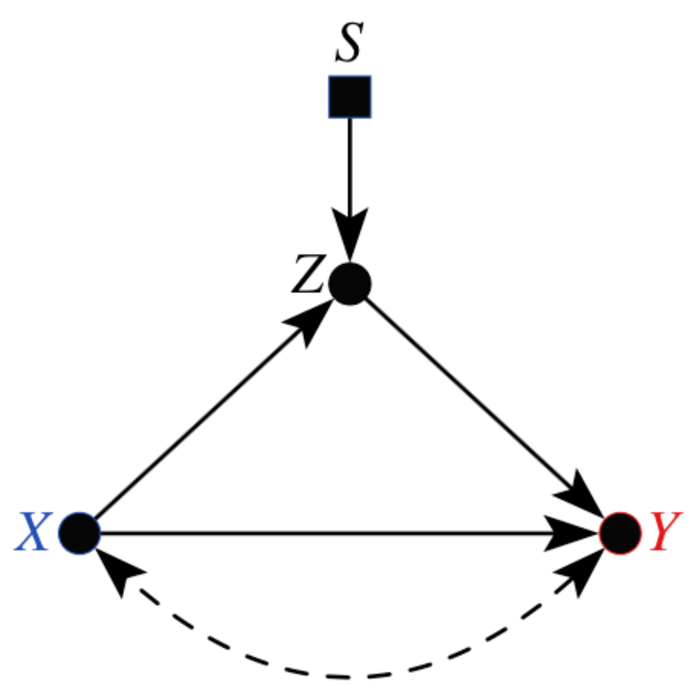
- Transportable directly from source to target (Manksi called "external validity")

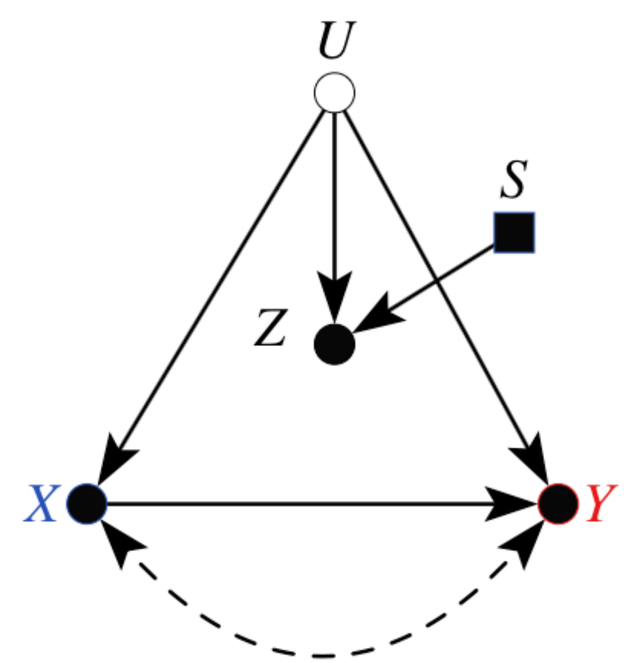
$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- Reduced to checking d-separation in selection diagram

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$







Finding invariances: the basics

1) ***Trivial transportability***

- *Effect estimable directly from obs. distribution in target (vanilla identification)*

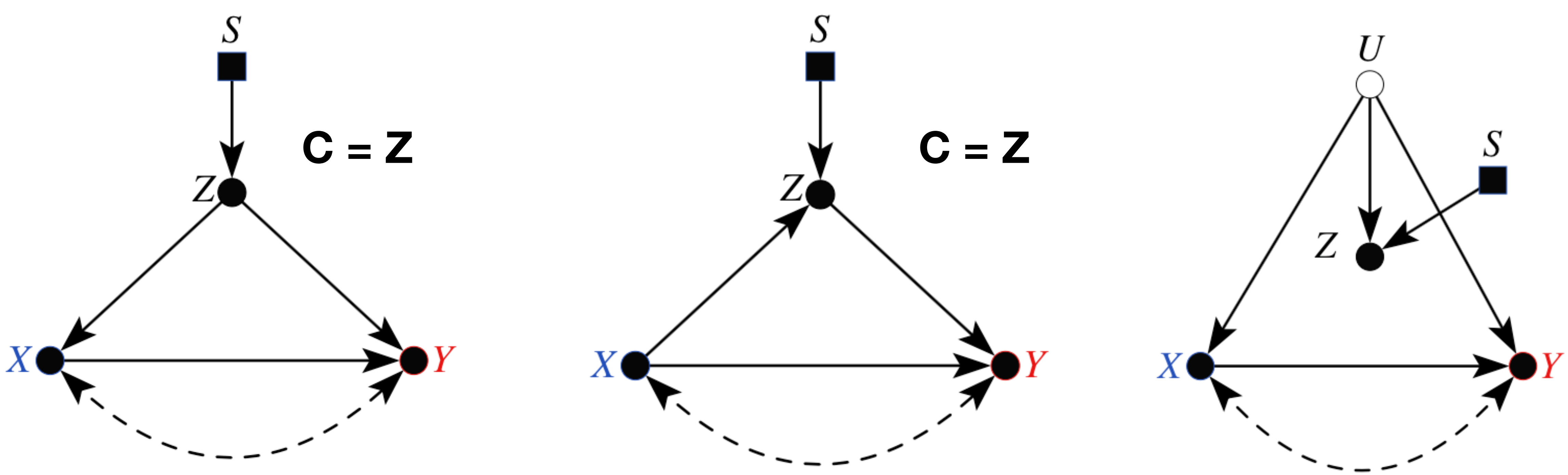
2) ***Direct transportability***

- *Transportable directly from source to target (Manksi called "external validity")*

$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- *Reduced to checking d-separation in selection diagram*

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$



Finding invariances: the basics

1) ***Trivial transportability***

- *Effect estimable directly from obs. distribution in target (vanilla identification)*

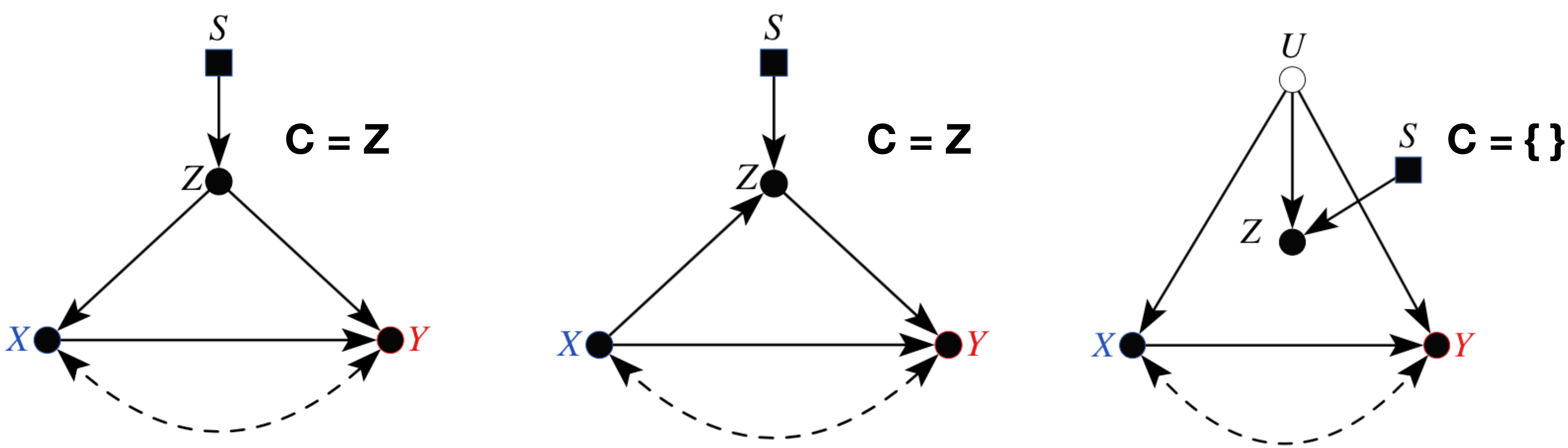
2) ***Direct transportability***

- *Transportable directly from source to target (Manksi called "external validity")*

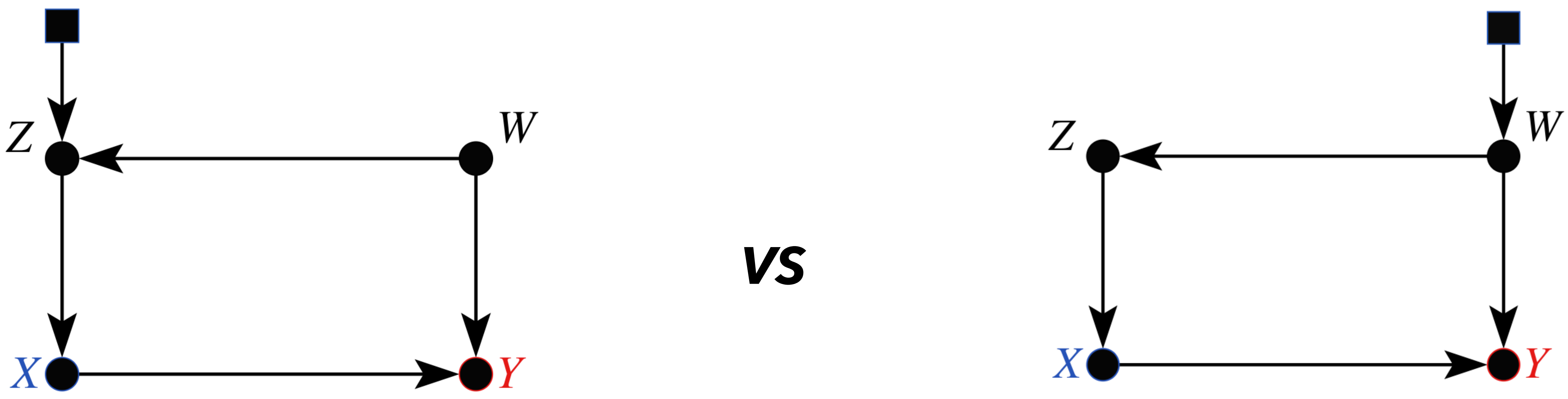
$$E^*[Y \mid do(x), z] = E[Y \mid do(x), z]$$

- *Reduced to checking d-separation in selection diagram*

$$(Y \perp\!\!\!\perp S \mid C, X)_{D_{\overline{X}}} \implies E^*[Y \mid do(x), c] = E[Y \mid do(x), c, s] = E[Y \mid do(x), c]$$

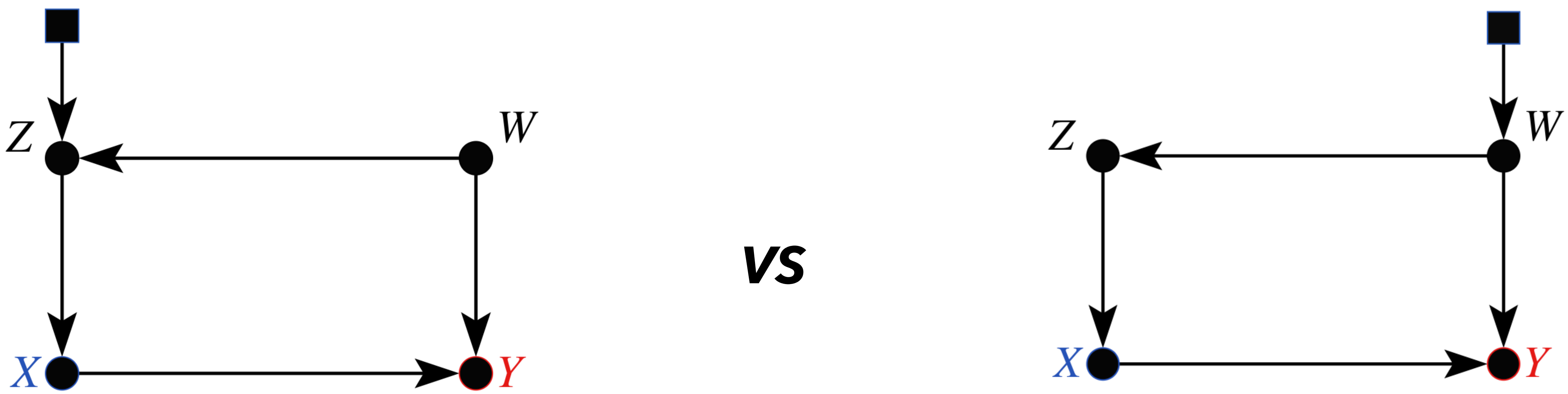


Two lessons



Both Z and W are valid adjustments for the identification of P(y|do(x)). But are they equally important for transporting the effect to Π^*?
(hint: use d-sep.)

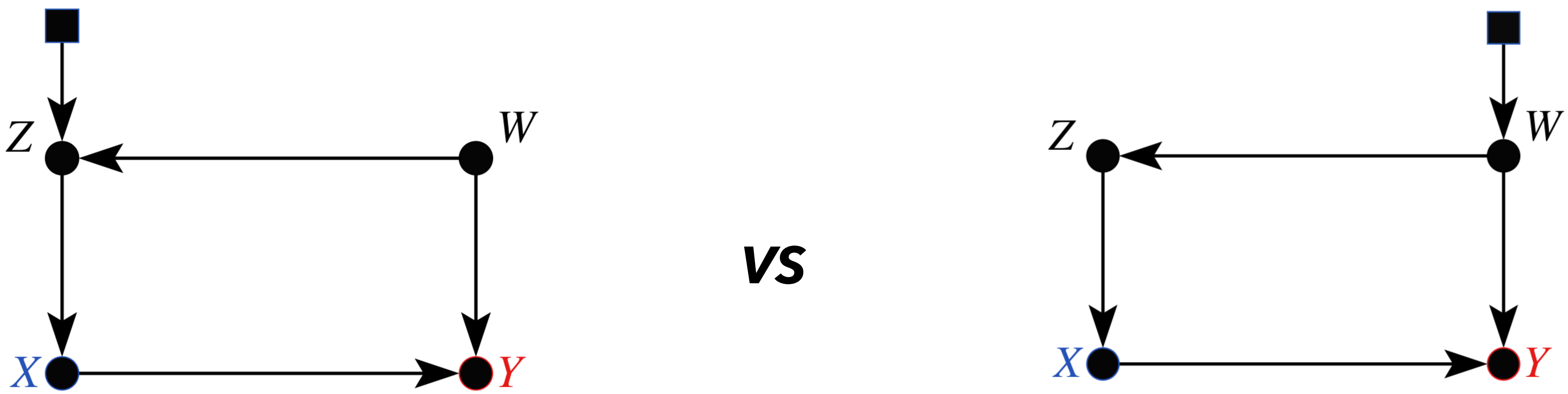
Two lessons



Both Z and W are valid adjustments for the identification of P(y|do(x)). But are they equally important for transporting the effect to Π^*?
(hint: use d-sep.)

Any selection node d-connected to Y only via X can be ignored.

Two lessons



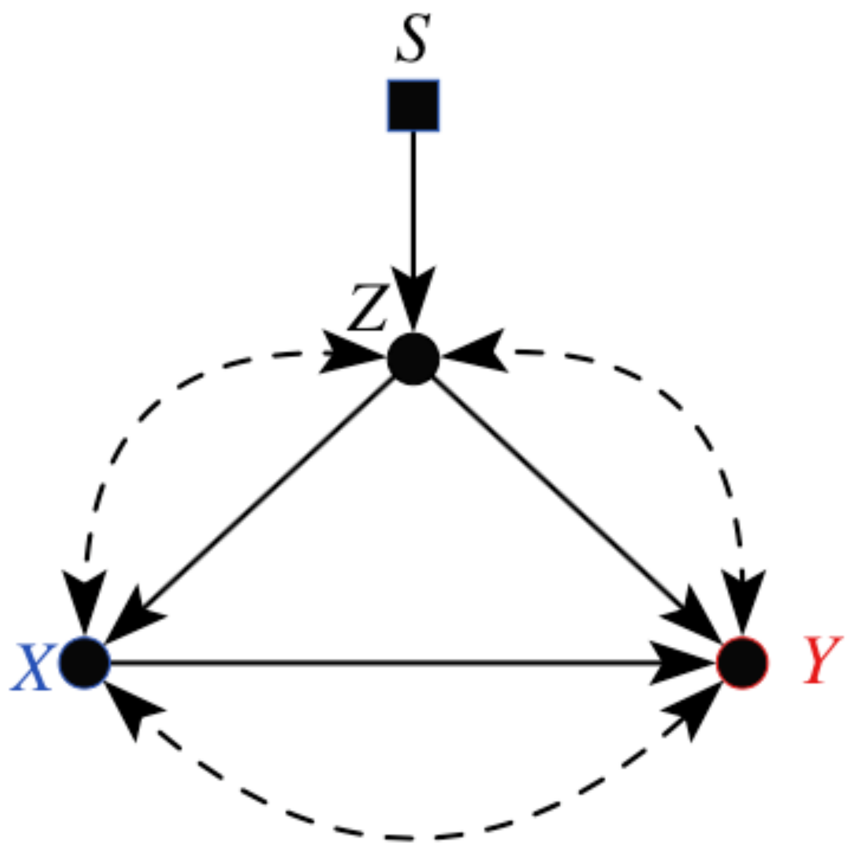
Both Z and W are valid adjustments for the identification of P(y|do(x)). But are they equally important for transporting the effect to Π^*?
(hint: use d-sep.)

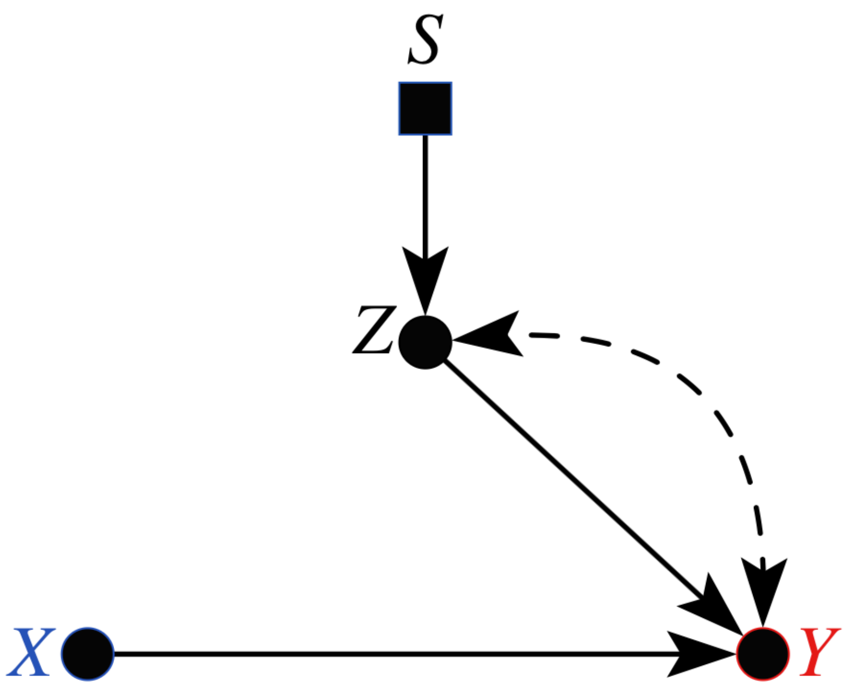
Any selection node d-connected to Y only via X can be ignored.

Lesson 1: *differences in propensity to receive treatment do not matter for transportability of causal effects. What matters are potential effect-modifiers.*

Two lessons

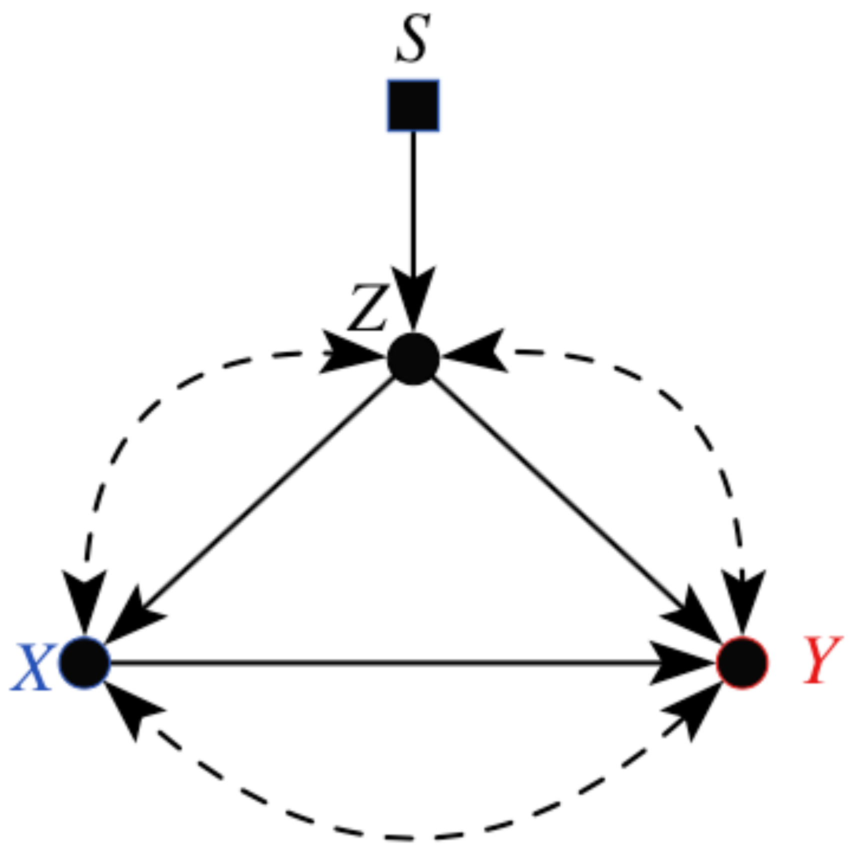
Is a randomized control trial really a gold standard?

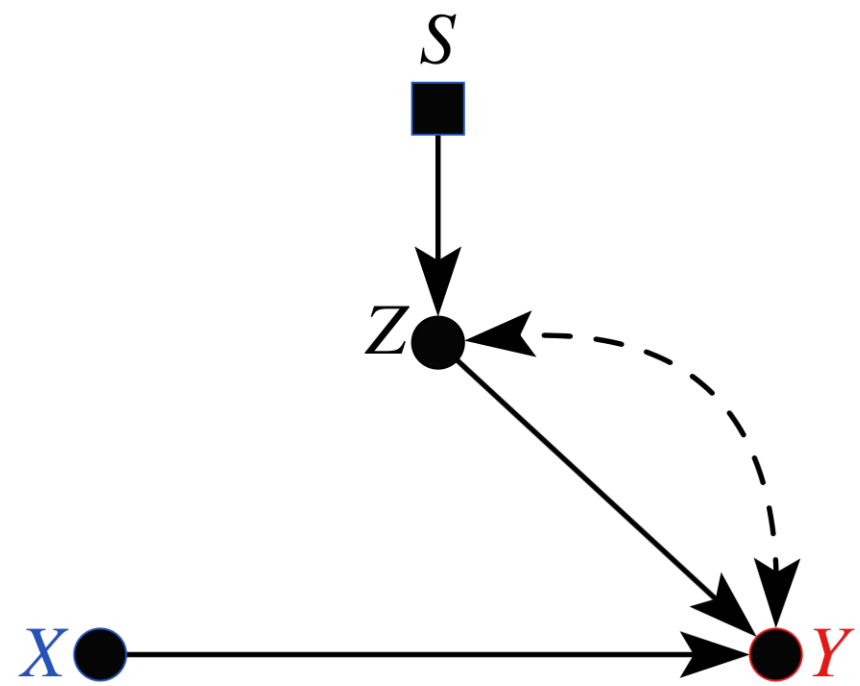




Two lessons

Is a randomized control trial really a gold standard?

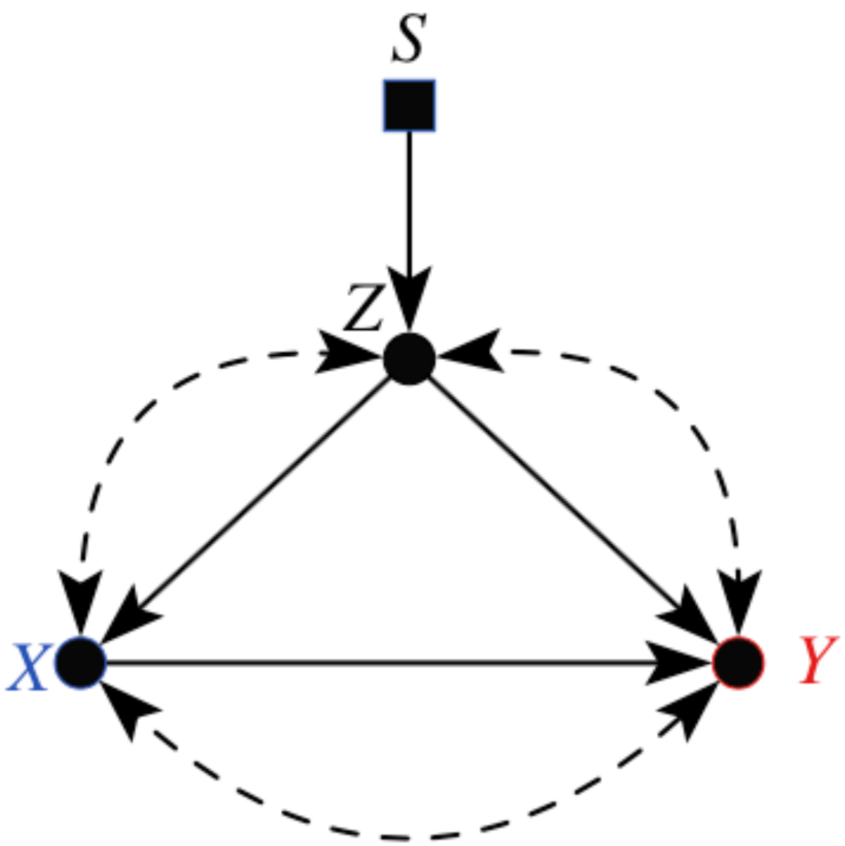


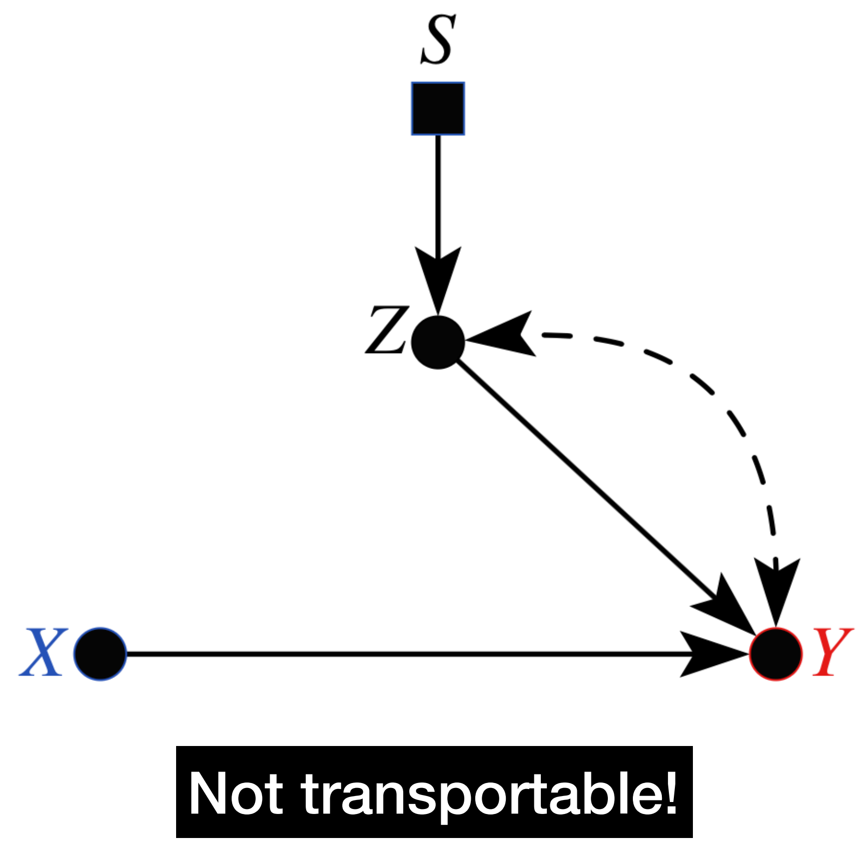


Not transportable!

Two lessons

Is a randomized control trial really a gold standard?





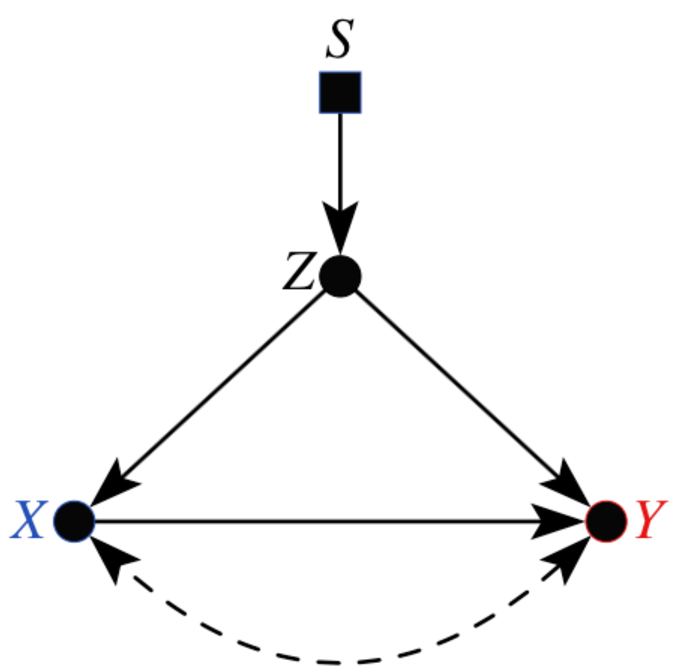
Lesson 2: unless one wants to confine experimental results to the strict conditions of the studied subpopulation, even with a perfect RCT one still needs to go through a transportability exercise (ie, causal modeling).

Finding invariances: beyond direct transportability

- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*

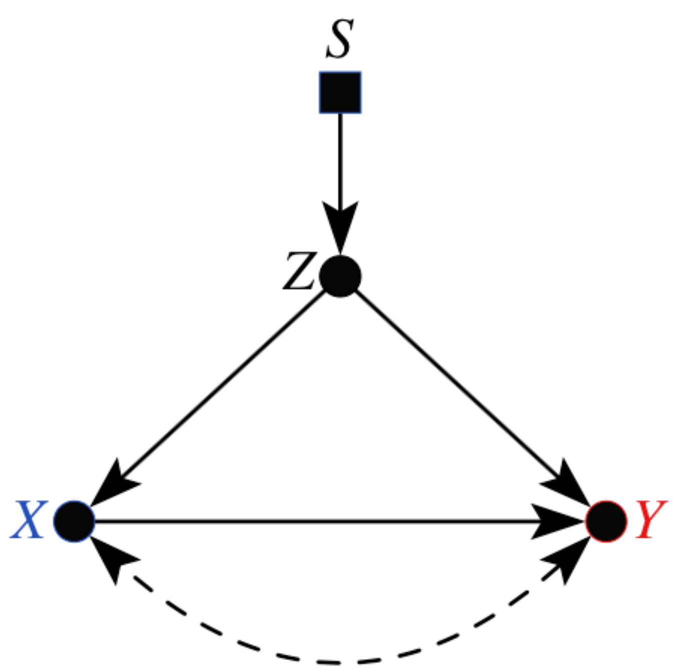
Finding invariances: beyond direct transportability

- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*



Finding invariances: beyond direct transportability

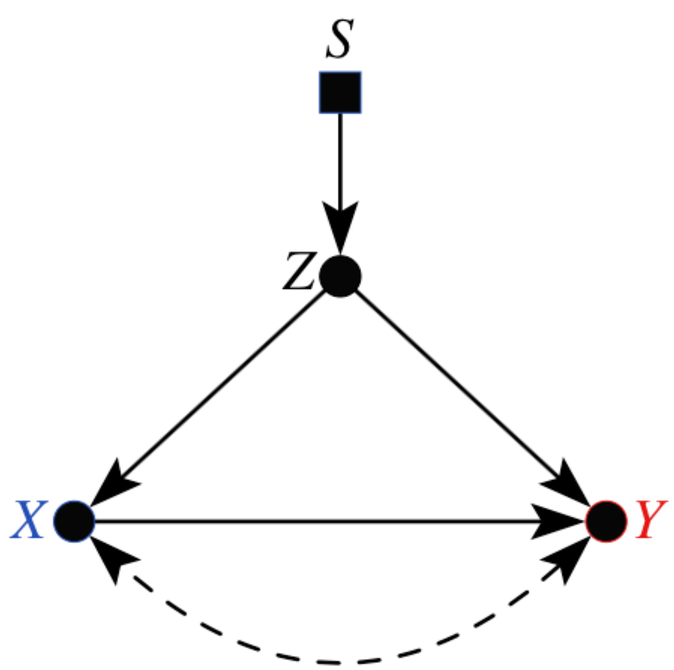
- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*



$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z-specific effect from source}} \quad \underbrace{P^*(z)}_{\text{weight from target dist.}}
\end{aligned}
$$

Finding invariances: beyond direct transportability

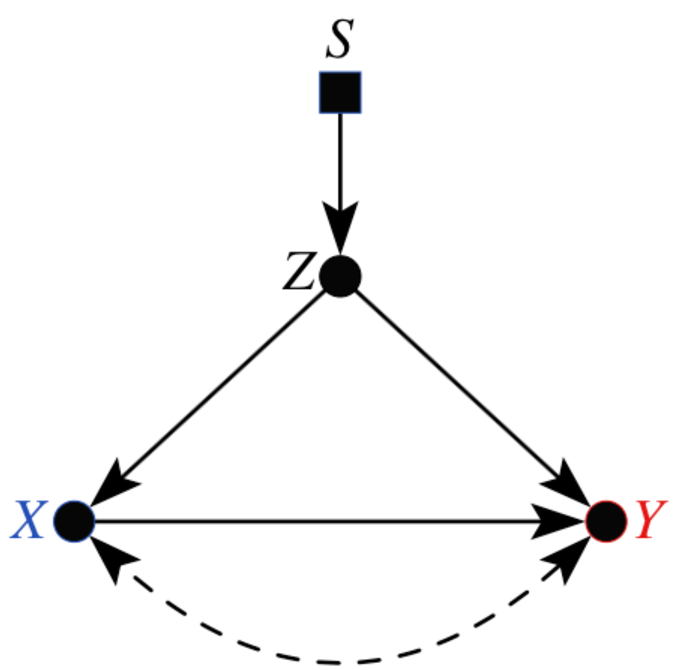
- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*



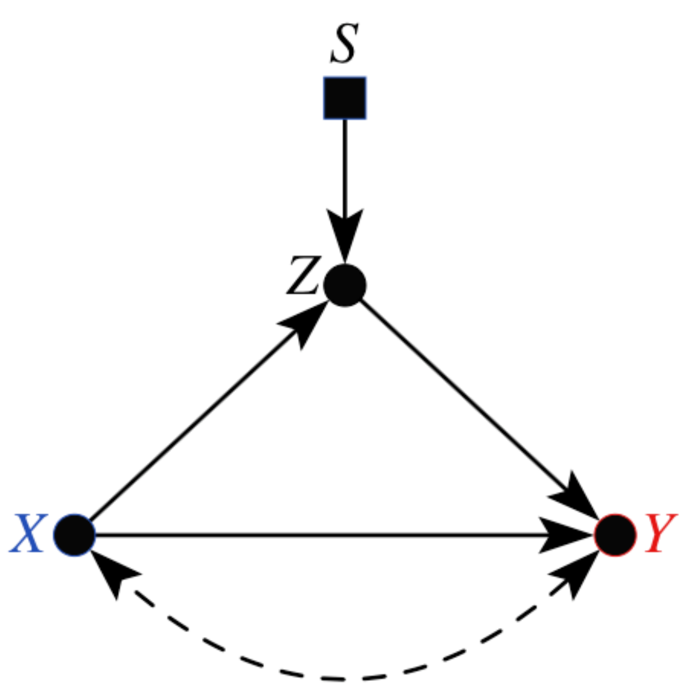
$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z-specific effect from source}} \; \underbrace{P^*(z)}_{\text{weight from target dist.}}
\end{aligned}
$$

Finding invariances: beyond direct transportability

- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*

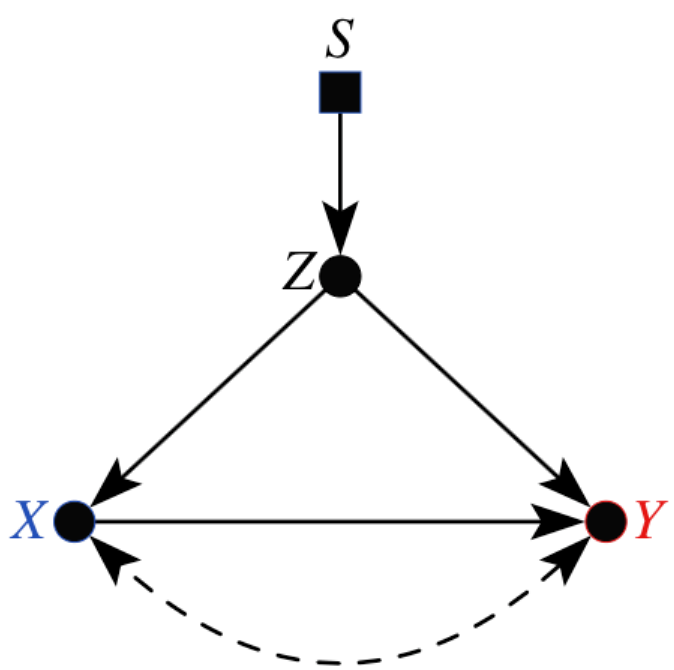


$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z-specific effect from source}} \underbrace{P^*(z)}_{\text{weight from target dist.}}
\end{aligned}
$$

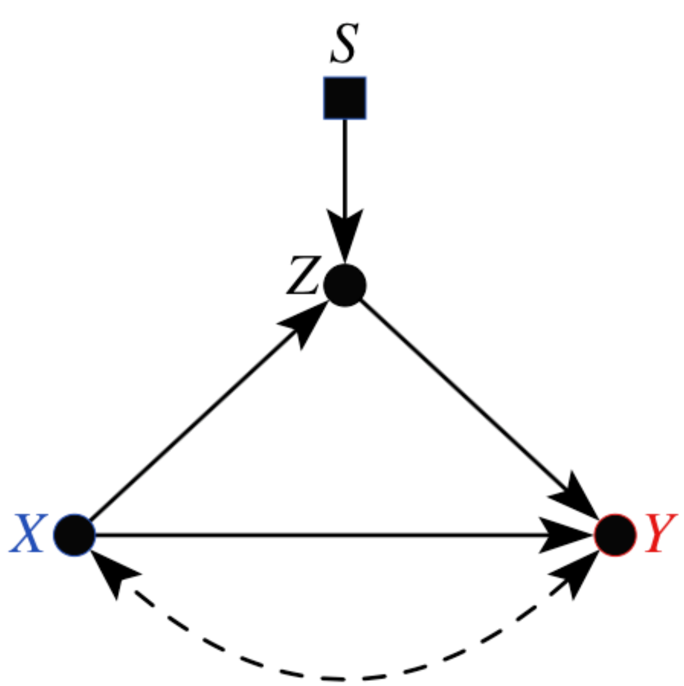


Finding invariances: beyond direct transportability

- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*



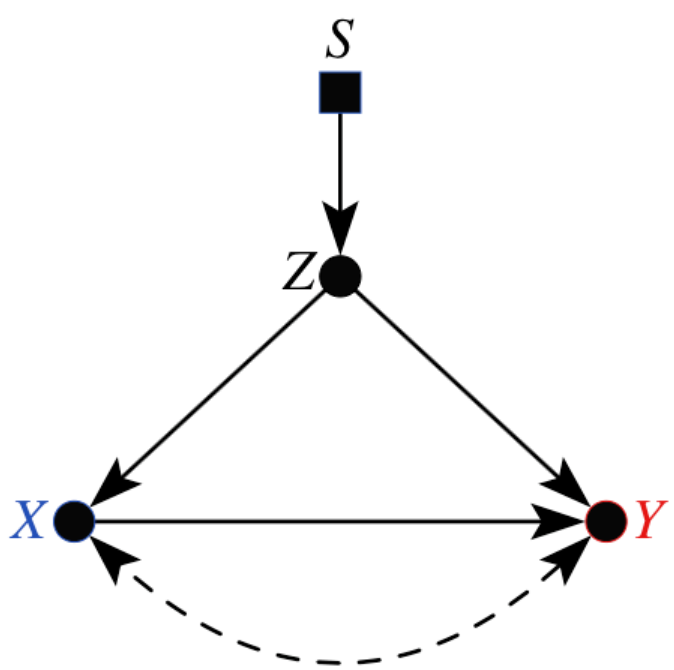
$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z-specific effect from source}} \; \underbrace{P^*(z)}_{\text{weight from target dist.}}
\end{aligned}
$$



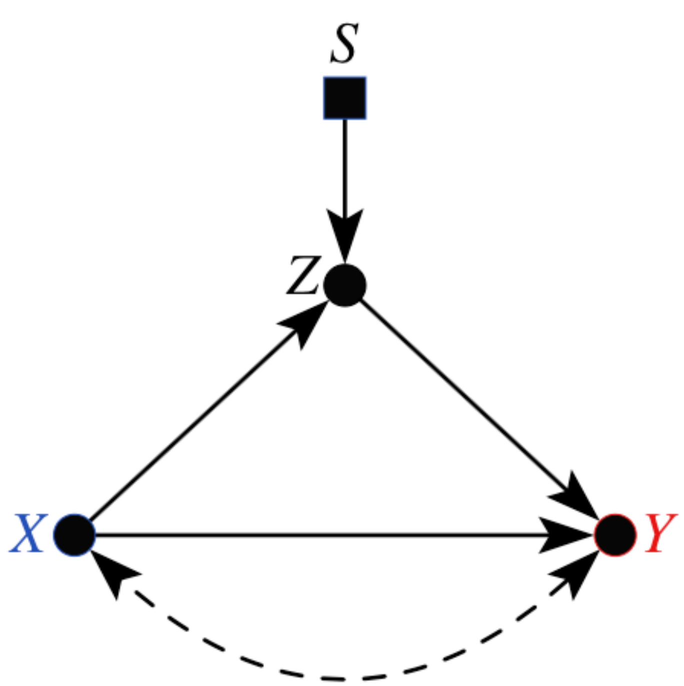
$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid x, s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z-specific effect from source}} \; \underbrace{P^*(z \mid x)}_{\text{weight from target dist.}}
\end{aligned}
$$

Finding invariances: beyond direct transportability

- *Many effects are not directly TR, but are TR after proper adjustment.*
- ***Strategy:*** *break relations that are not directly TR to find invariant pieces.*



$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z−specific effect from source}} \quad \underbrace{P^*(z)}_{\text{weight from target dist.}}
\end{aligned}
$$



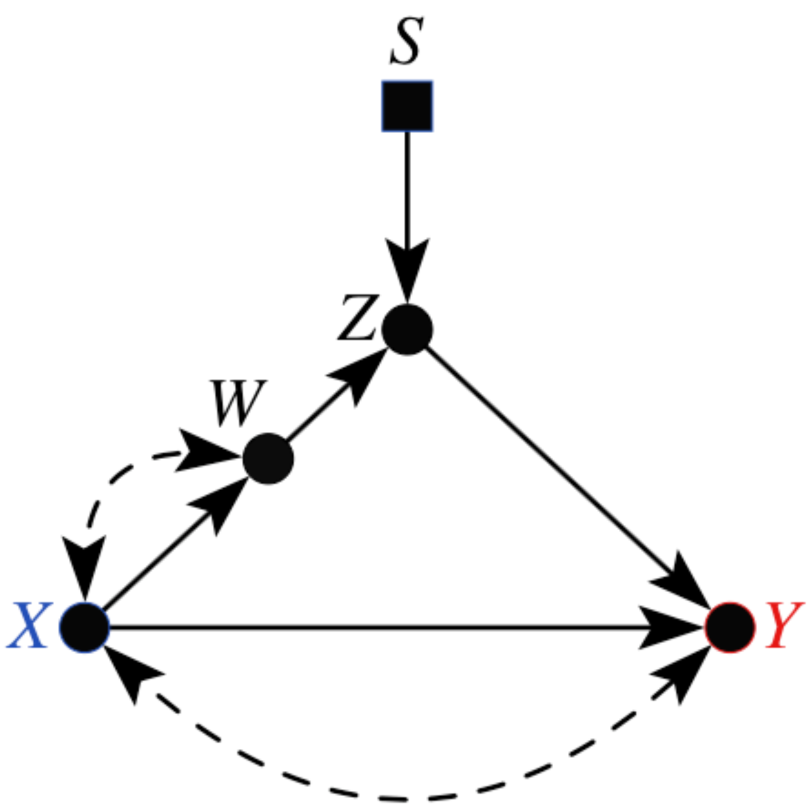
$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid x, s) \\
&= \sum_z \underbrace{E[Y \mid do(x), z]}_{\text{z−specific effect from source}} \quad \underbrace{P^*(z \mid x)}_{\text{weight from target dist.}}
\end{aligned}
$$

Finding invariances: beyond direct transportability

A more elaborate example:

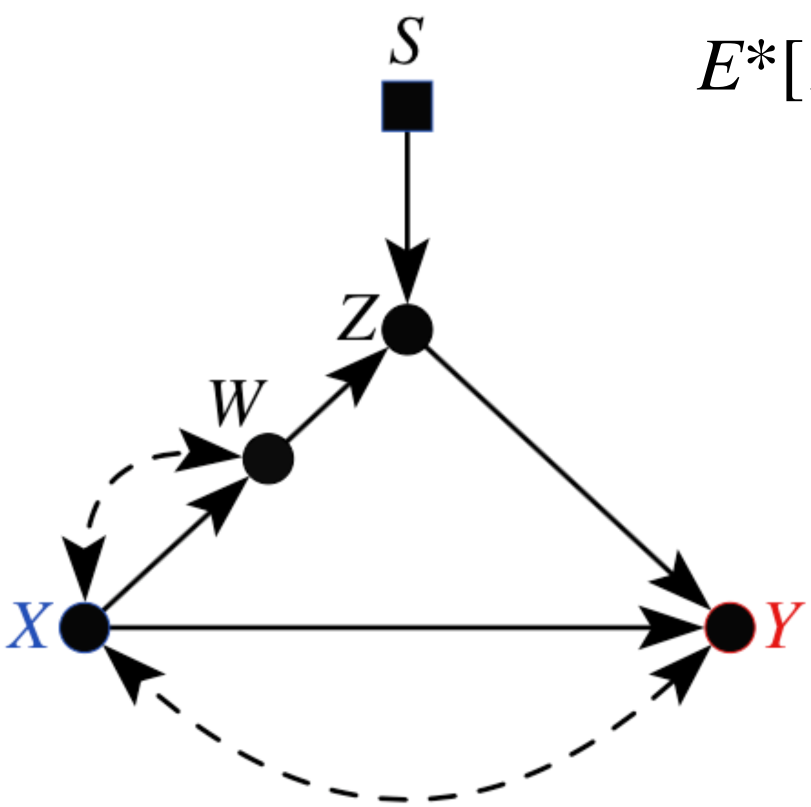
Finding invariances: beyond direct transportability

A more elaborate example:



Finding invariances: beyond direct transportability

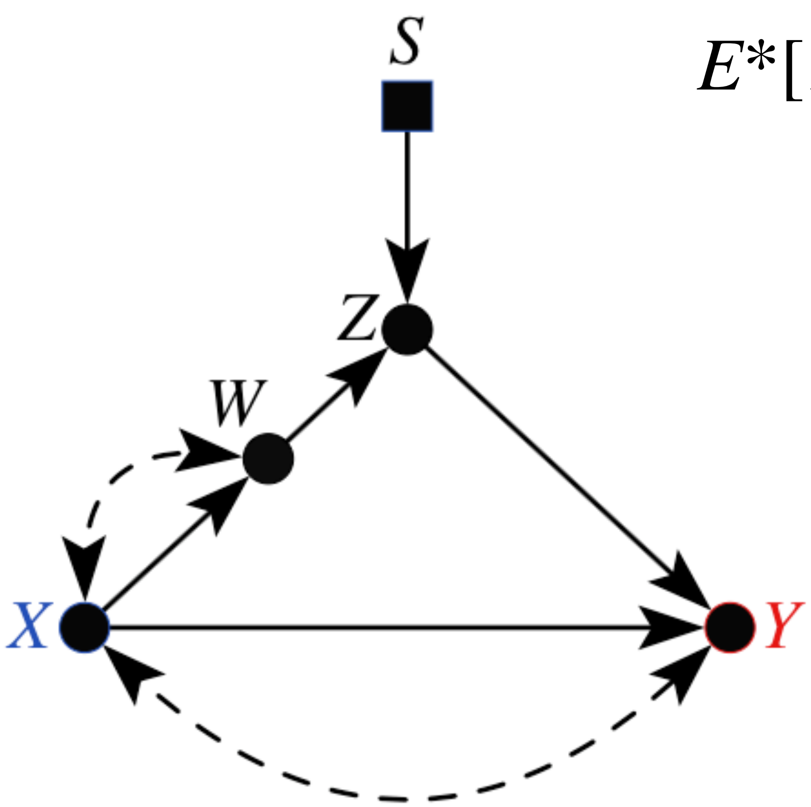
A more elaborate example:



$$
\begin{aligned}
E^*[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z] \sum_w P(z \mid do(x), w, s) P(w \mid do(x), s) \\
&= \sum_z E[Y \mid do(x), z] \sum_w P^*(z \mid w) P(w \mid do(x))
\end{aligned}
$$

Finding invariances: beyond direct transportability

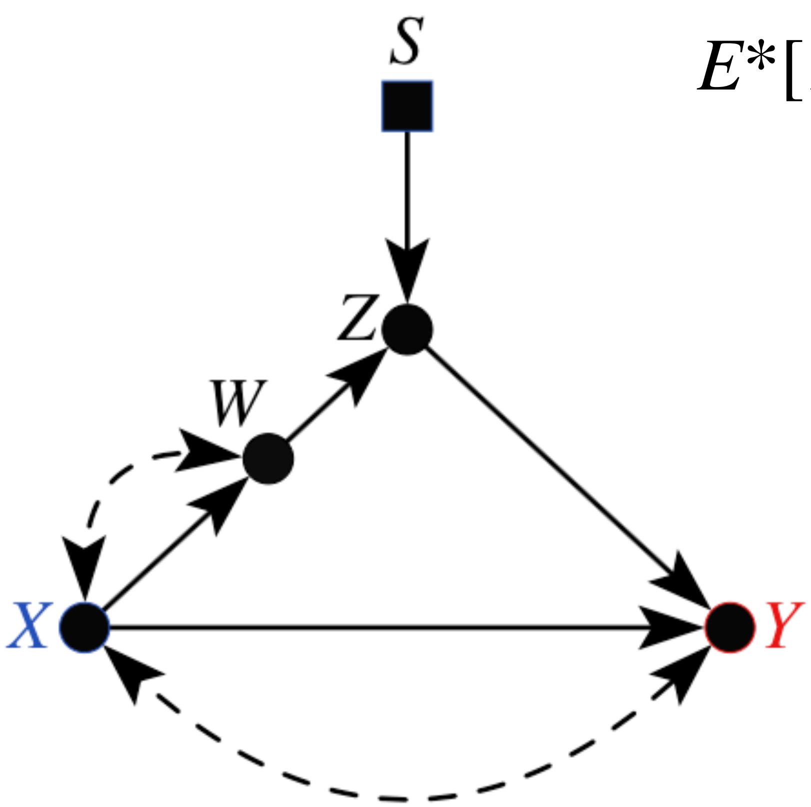
A more elaborate example:



$$E^*[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid do(x), z] \sum_w P(z \mid do(x), w, s) P(w \mid do(x), s)$$
$$= \sum_z E[Y \mid do(x), z] \sum_w P^*(z \mid w) P(w \mid do(x))$$

Finding invariances: beyond direct transportability

A more elaborate example:



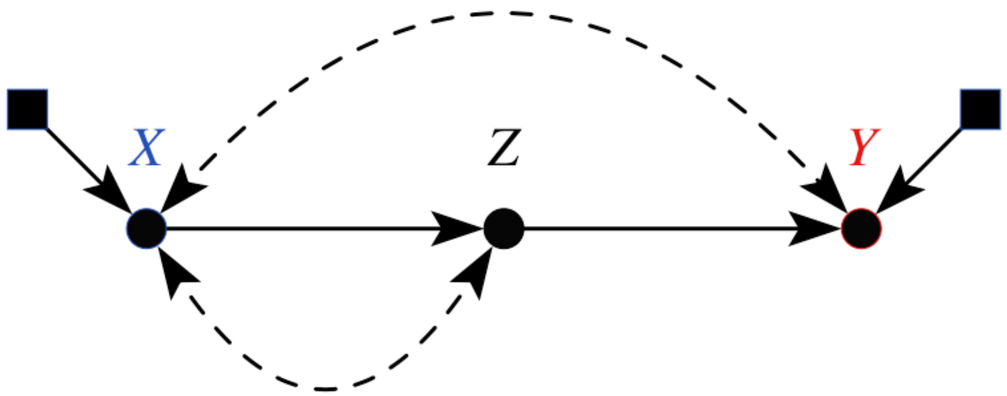
$$E^*[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid do(x), z] \sum_w P(z \mid do(x), w, s) P(w \mid do(x), s)$$
$$= \sum_z E[Y \mid do(x), z] \sum_w P^*(z \mid w) P(w \mid do(x))$$

Now let us extend to ***multiple populations****, each with* ***different experimental conditions:*** *for instance, in one domain only X was randomized while in another domain only Z was randomized… and so on.*

Multiple Populations

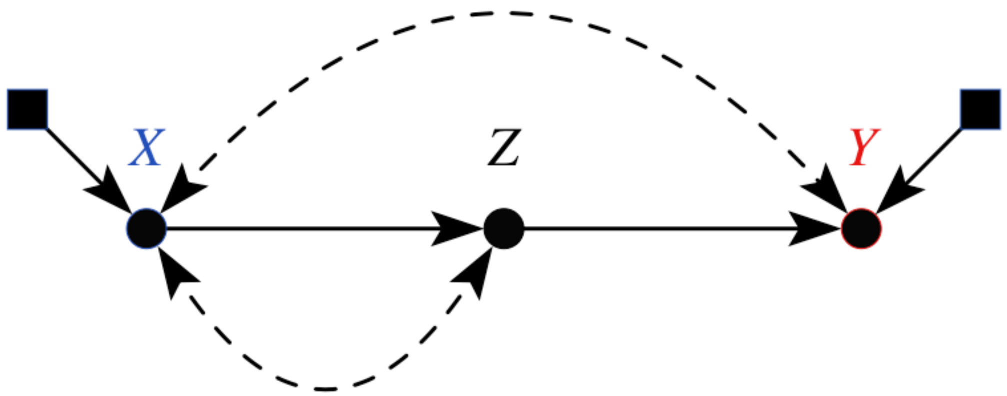
Multiple Populations

Π^A :



Multiple Populations

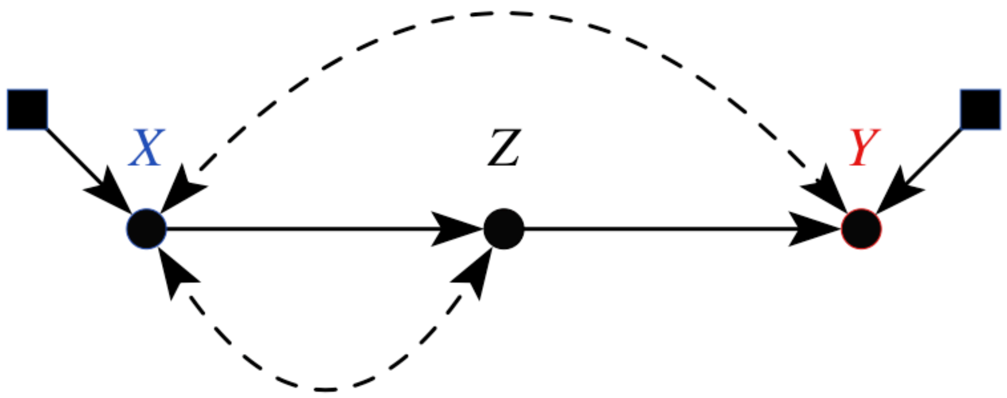
Π^A :



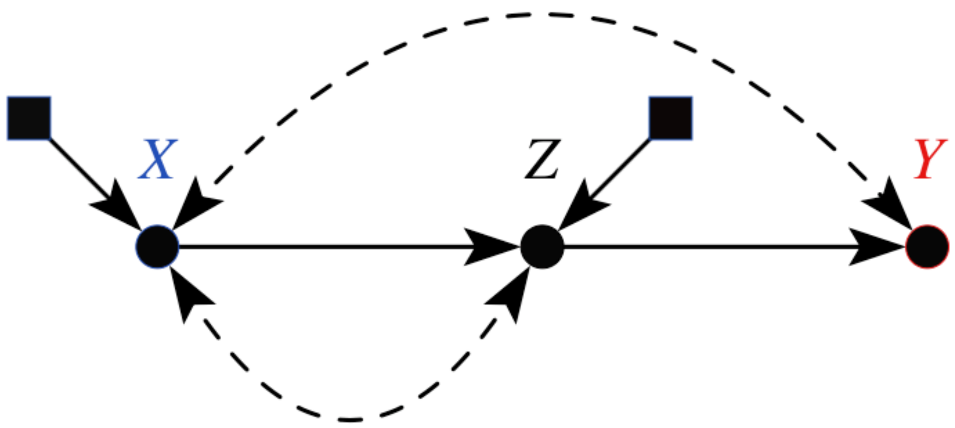
Not transportable from A.

Multiple Populations

Π^A :



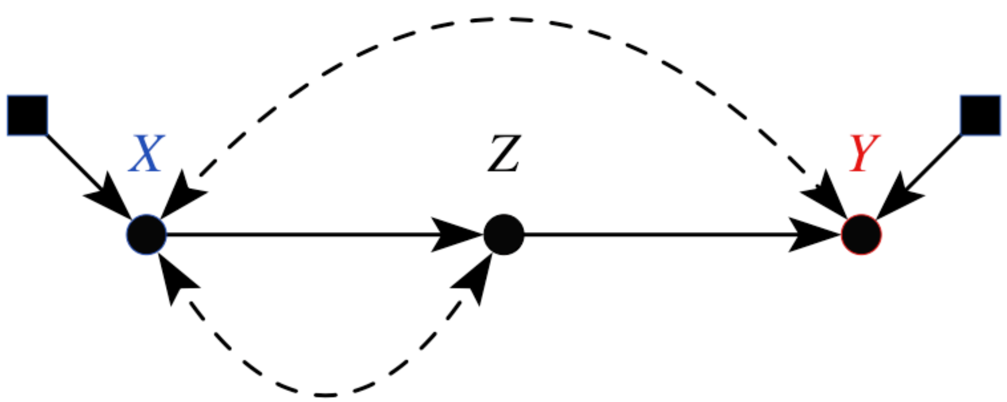
Π^B :



Not transportable from A.

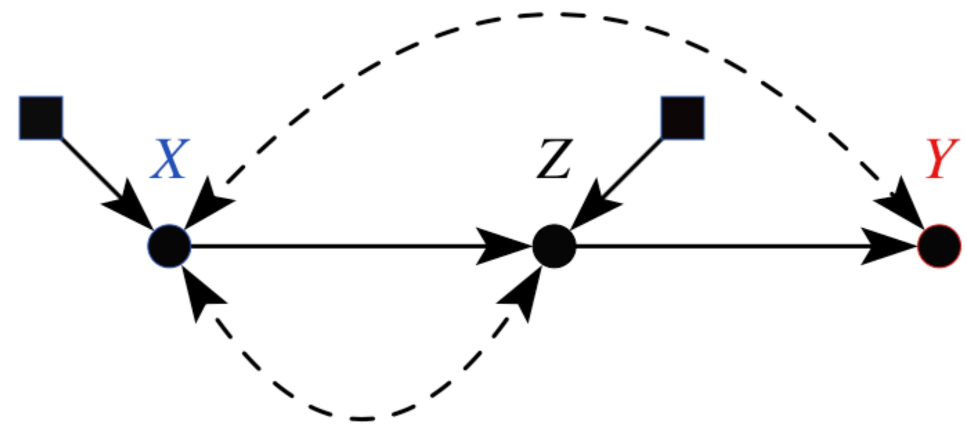
Multiple Populations

Π^A :



Not transportable from A.

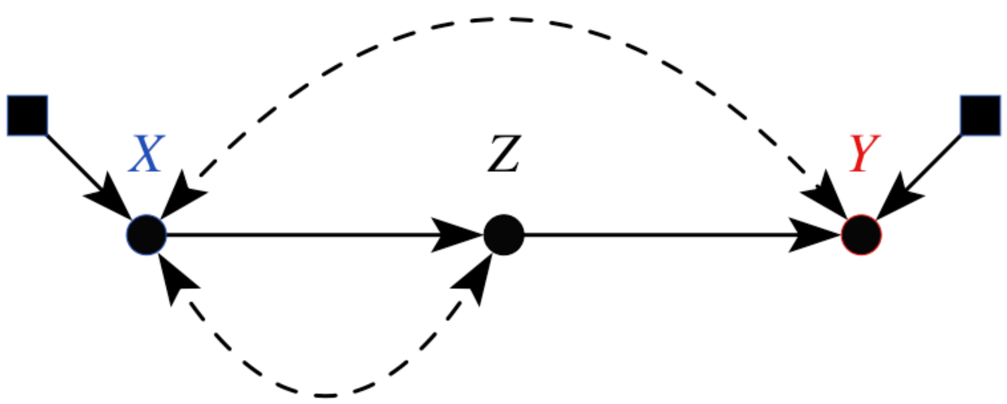
Π^B :



Not transportable from B.

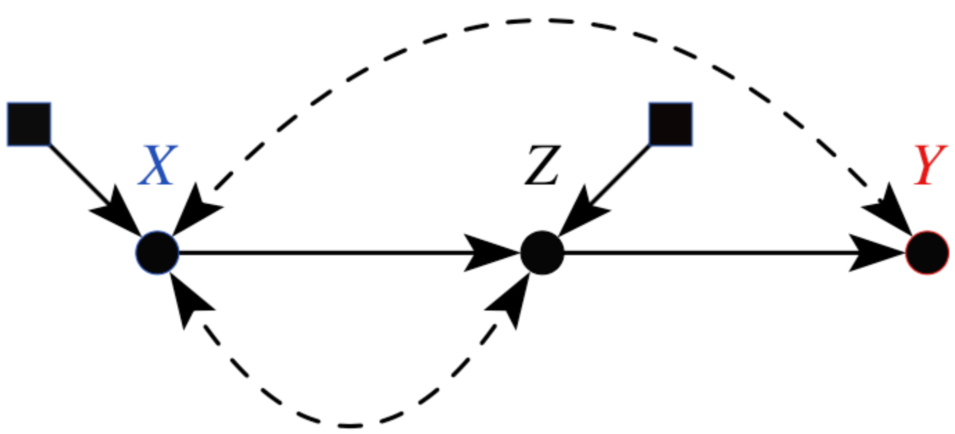
Multiple Populations

Π^A :



Not transportable from A.

Π^B :

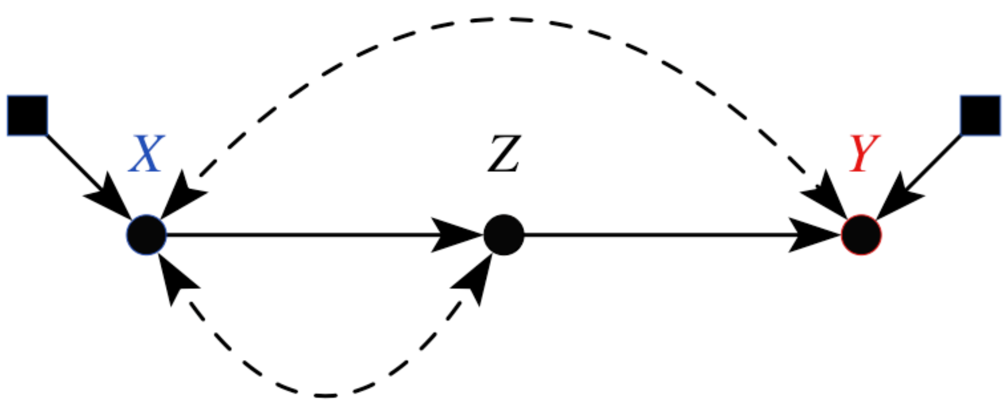


Not transportable from B.

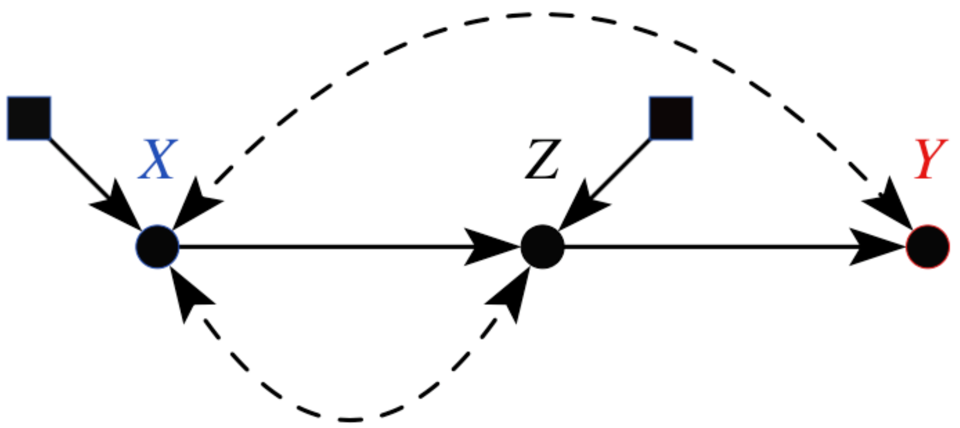
Not transportable at all?

Multiple Populations

Π^A :



Π^B :



Not transportable from A.

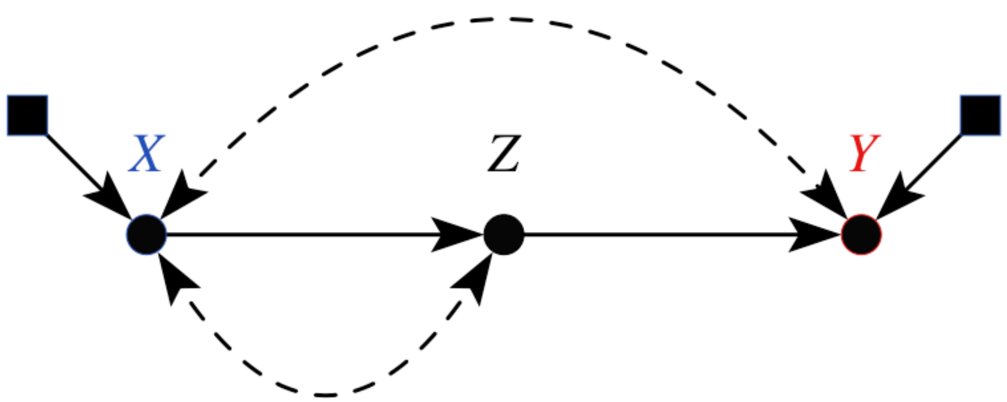
Not transportable from B.

Not transportable at all?

What if we combine the experimental results of A and B?

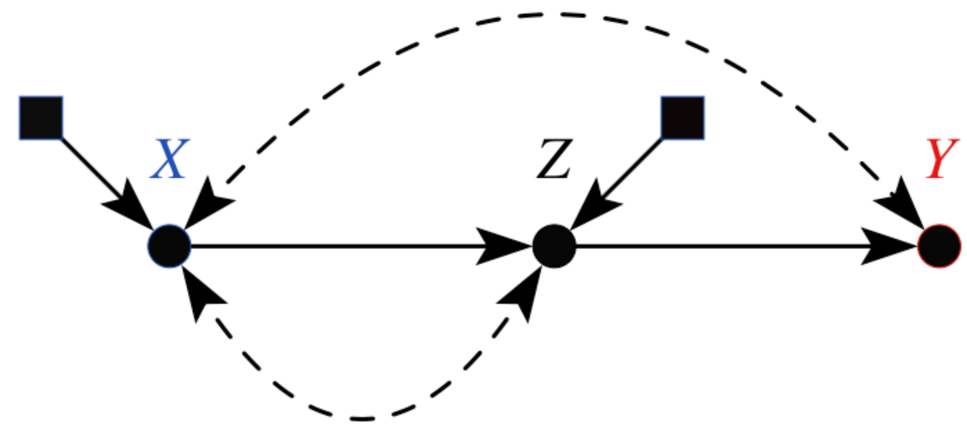
Multiple Populations

Π^A :



Not transportable from A.

Π^B :



Not transportable from B.

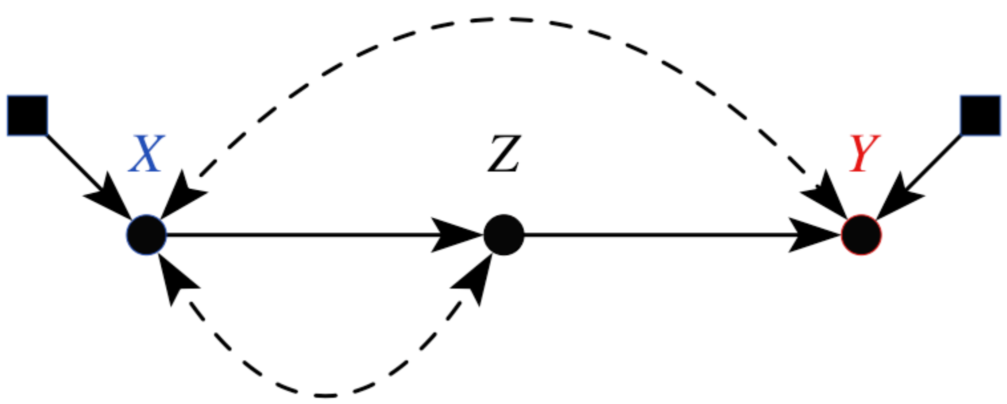
Not transportable at all?

What if we combine the experimental results of A and B?

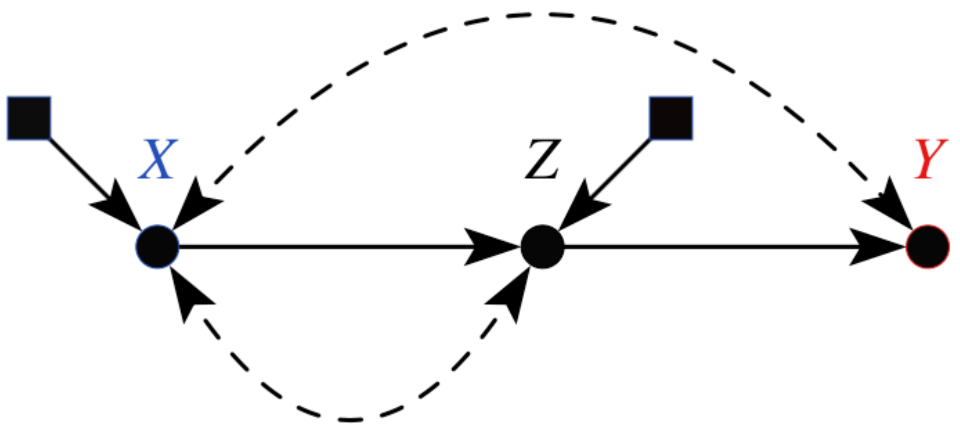
$$
\begin{aligned}
P^*(y \mid do(x)) &= \sum_z P^*(y \mid do(x), z) P^*(z \mid do(x)) \\
&= \sum_z P^*(y \mid do(x), do(z)) P^*(z \mid do(x)) \\
&= \sum_z P^*(y \mid do(z)) P^*(z \mid do(x)) \\
&= \sum_z \underbrace{P^B(y \mid do(z))}_{\text{RCT Z in B}} \underbrace{P^A(z \mid do(x))}_{\text{RCT X in A}}
\end{aligned}
$$

Multiple Populations

Π^A :



Π^B :



Not transportable from A.

Not transportable from B.

Not transportable at all?

What if we combine the experimental results of A and B?

$$P^*(y \mid do(x)) = \sum_z P^*(y \mid do(x), z) P^*(z \mid do(x))$$

$$= \sum_z P^*(y \mid do(x), do(z)) P^*(z \mid do(x))$$

$$= \sum_z P^*(y \mid do(z)) P^*(z \mid do(x))$$

$$= \sum_z \underbrace{P^B(y \mid do(z))}_{\text{RCT Z in B}} \underbrace{P^A(z \mid do(x))}_{\text{RCT X in A}}$$

General solution to transportability

You do not need to derive each case by hand.

General solution to transportability

You do not need to derive each case by hand.

We have a ***complete algorithms*** that can decide how to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions, so as to construct a valid estimate of the effect size for the target population.

General solution to transportability

You do not need to derive each case by hand.

We have a ***complete algorithms*** that can decide how to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions, so as to construct a valid estimate of the effect size for the target population.

What does completeness mean?

General solution to transportability

You do not need to derive each case by hand.

We have a ***complete algorithms*** that can decide how to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions, so as to construct a valid estimate of the effect size for the target population.

What does completeness mean?

- It means that if the algorithm can't find a solution, then it is ***impossible*** *to transport the causal effect of interest* ***without strengthening assumptions.***

General solution to transportability

You do not need to derive each case by hand.

We have a ***complete algorithms*** that can decide how to combine results of several experimental and observational studies, each conducted on a different population and under a different set of conditions, so as to construct a valid estimate of the effect size for the target population.

What does completeness mean?

- It means that if the algorithm can't find a solution, then it is ***impossible*** *to transport the causal effect of interest* ***without strengthening assumptions.***

FUSION DEMO 1

Selection Bias

Selected Subpopulation → General Population

Selection bias vs Confounding bias

Warning: some economists use "selection bias" to denote confounding bias

Selection bias vs Confounding bias

Warning: some economists use "selection bias" to denote confounding bias

e.g. Angrist and Pischke, MHE

Selection bias vs Confounding bias

Warning: some economists use "selection bias” to denote confounding bias

e.g. Angrist and Pischke, MHE

For them, "selection bias” is referring to preferential "selection to treatment”.

Selection bias vs Confounding bias

Warning: some economists use "selection bias" to denote confounding bias

e.g. Angrist and Pischke, MHE

For them, "selection bias" is referring to preferential "selection to treatment".

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

Selection bias vs Confounding bias

Warning: some economists use "selection bias" to denote confounding bias

e.g. Angrist and Pischke, MHE

For them, "selection bias" is referring to preferential "selection to treatment".

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

These should not be mixed, confounding bias and selection bias have different nature.

Selection bias vs Confounding bias

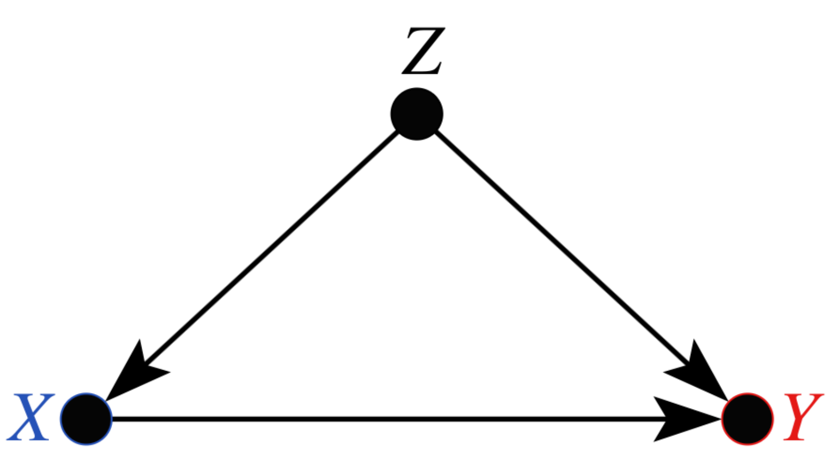
Warning: some economists use "selection bias" to denote confounding bias

e.g. Angrist and Pischke, MHE

For them, "selection bias" is referring to preferential "selection to treatment".

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

These should not be mixed, confounding bias and selection bias have different nature.



Selection bias vs Confounding bias

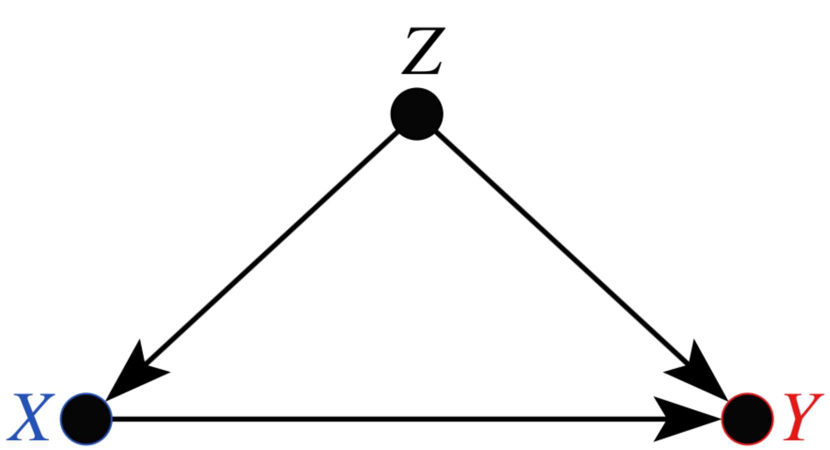
Warning: some economists use "selection bias" to denote confounding bias

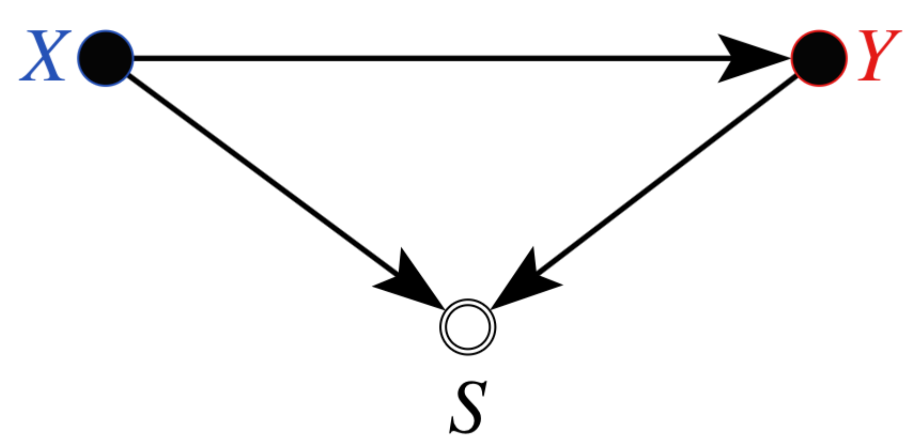
e.g. Angrist and Pischke, MHE

For them, "selection bias" is referring to preferential "selection to treatment".

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

These should not be mixed, confounding bias and selection bias have different nature.





Selection bias vs Confounding bias

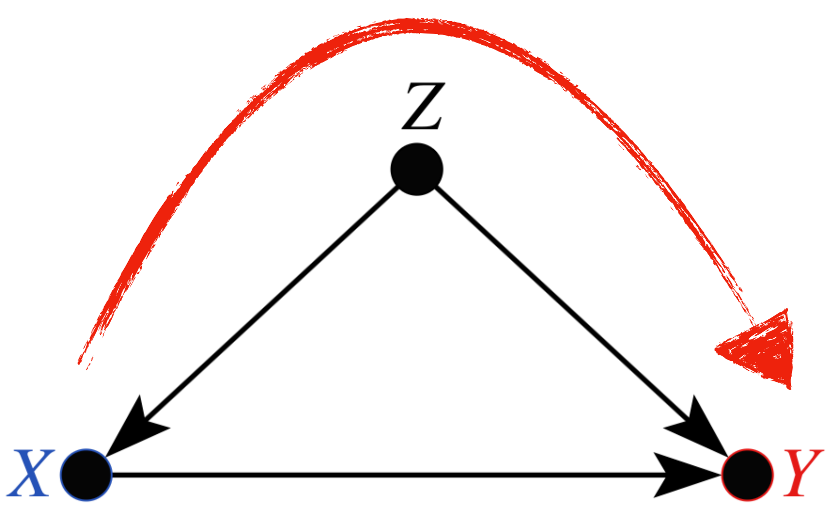
Warning: some economists use "selection bias” to denote confounding bias

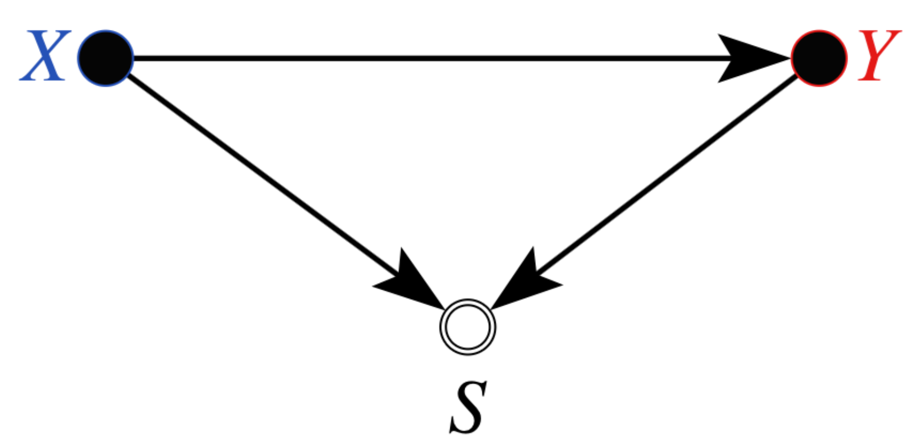
e.g. Angrist and Pischke, MHE

For them, "selection bias” is referring to preferential "selection to treatment”.

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

These should not be mixed, confounding bias and selection bias have different nature.





Selection bias vs Confounding bias

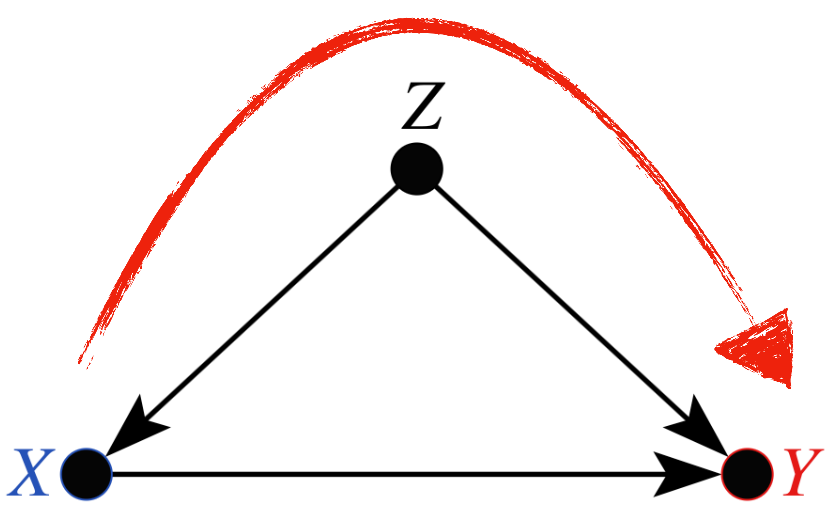
Warning: some economists use "selection bias" to denote confounding bias

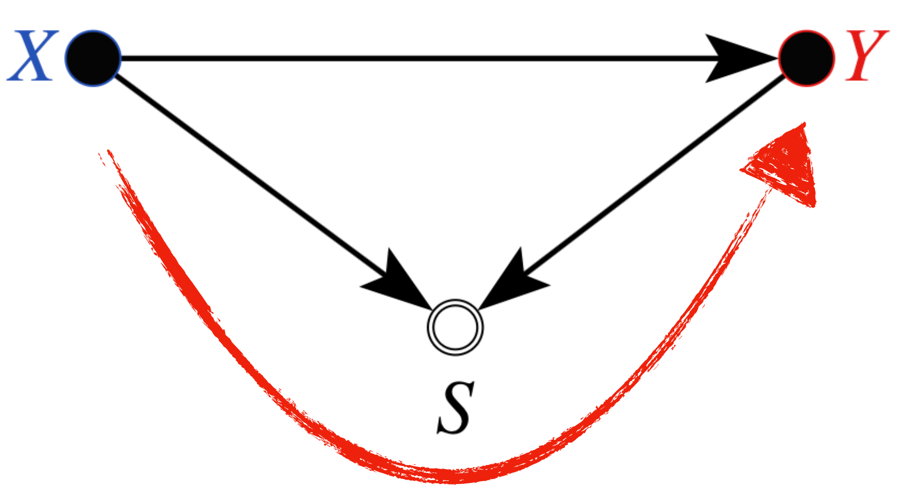
e.g. Angrist and Pischke, MHE

For them, "selection bias" is referring to preferential "selection to treatment".

More generally, we use selection bias to mean bias due to ***preferential selection of units into the study sample.***

These should not be mixed, confounding bias and selection bias have different nature.





Formalizing selection bias

1) ***What do we want to know? (Query)***

- *Conditional expectation on general population:* $P(y|x)$
- *Causal effect on general population:* $E[Y|do(x)]$

Formalizing selection bias

1) ***What do we want to know? (Query)***

- *Conditional expectation on general population:* $P(y \mid x)$
- *Causal effect on general population:* $E[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- *Observational/Experimental data in the study sample (S = 1). May or may not have census data for some variables Z in the general population.*

$$P(V \mid S = 1), P(V \mid do(z), S = 1), P(Z)$$

Formalizing selection bias

1) ***What do we want to know? (Query)***

- *Conditional expectation on general population:* $P(y \mid x)$
- *Causal effect on general population:* $E[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- *Observational/Experimental data in the study sample (S = 1). May or may not have census data for some variables Z in the general population.*

$$P(V \mid S = 1), P(V \mid do(z), S = 1), P(Z)$$

3) ***What do we already know? (Causal Assumptions)***

- Need to describe the selection process.

Formalizing selection bias

1) ***What do we want to know? (Query)***

- *Conditional expectation on general population:* $P(y|x)$
- *Causal effect on general population:* $E[Y|do(x)]$

2) ***What data do we have? (Data)***

- *Observational/Experimental data in the study sample (S = 1). May or may not have census data for some variables Z in the general population.*

$$P(V|S=1), P(V|do(z), S=1), P(Z)$$

3) ***What do we already know? (Causal Assumptions)***

- Need to describe the selection process.
- Some approaches in early literature invoked strong parametric assumptions (Heckman: linear, gaussian);

Formalizing selection bias

1) ***What do we want to know? (Query)***

- Conditional expectation on general population: $P(y \mid x)$

- Causal effect on general population: $E[Y \mid do(x)]$

2) ***What data do we have? (Data)***

- Observational/Experimental data in the study sample (S = 1). May or may not have census data for some variables Z in the general population.

$$P(V \mid S = 1), P(V \mid do(z), S = 1), P(Z)$$

3) ***What do we already know? (Causal Assumptions)***

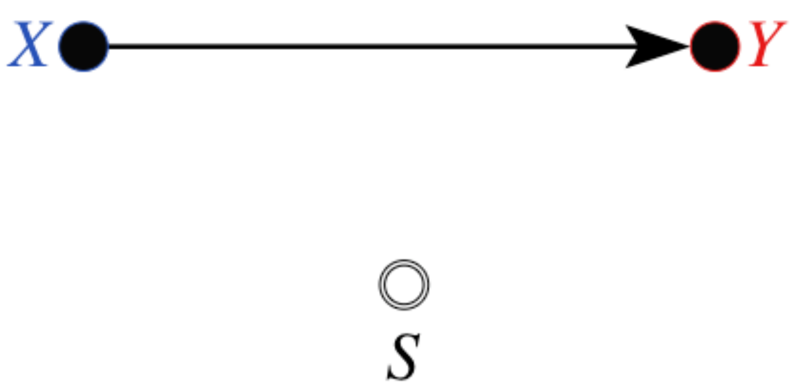
- Need to describe the selection process.
- Some approaches in early literature invoked strong parametric assumptions (Heckman: linear, gaussian);
- *Here:* ***nonparametric, qualitative description*** *of the determinants of inclusion of units in the study sample.*

Encoding the selection mechanism

Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.

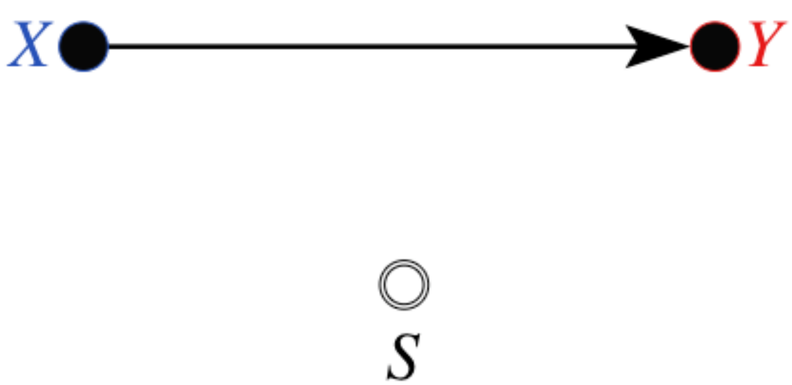
Encoding the selection mechanism

Again we extend our causal diagram with “selection nodes” (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



Encoding the selection mechanism

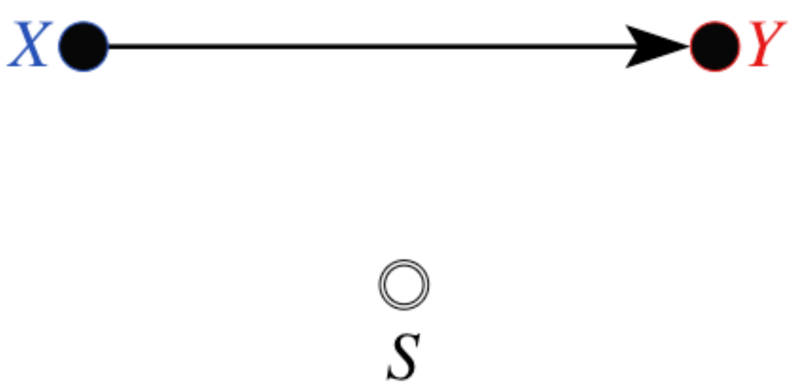
Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



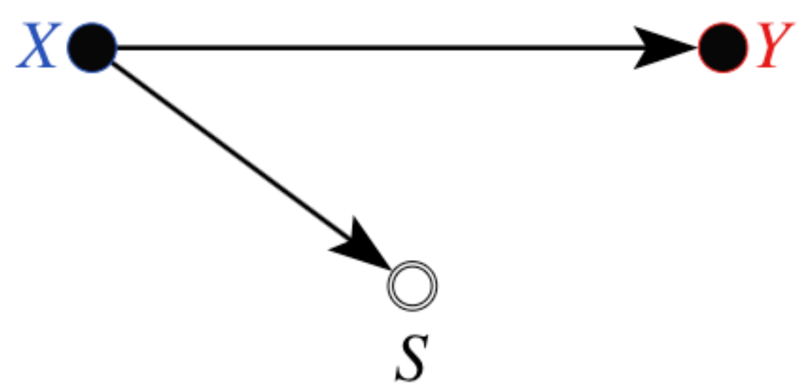
random sample

Encoding the selection mechanism

Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.

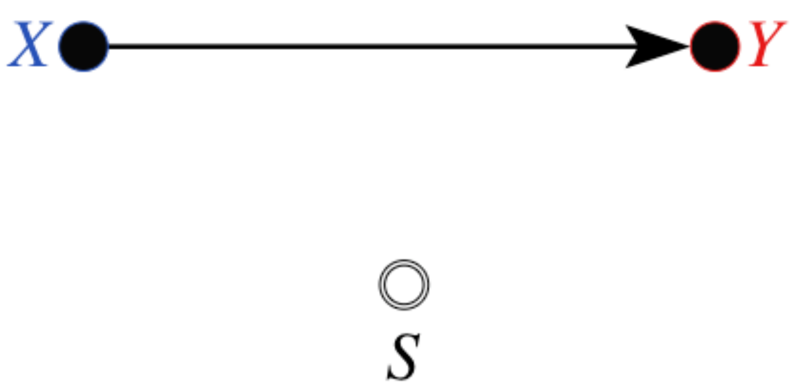


random sample

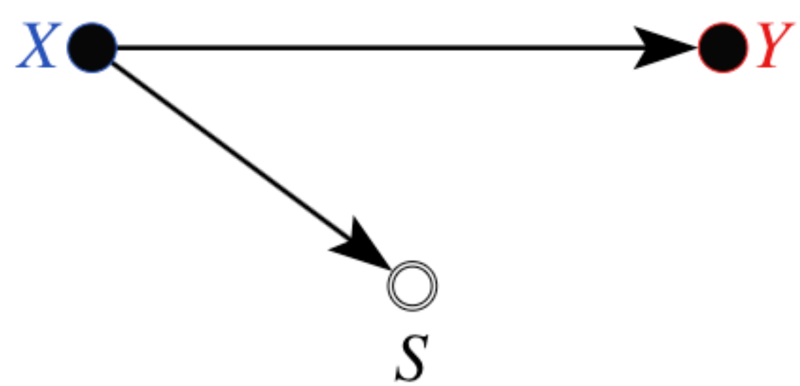


Encoding the selection mechanism

Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



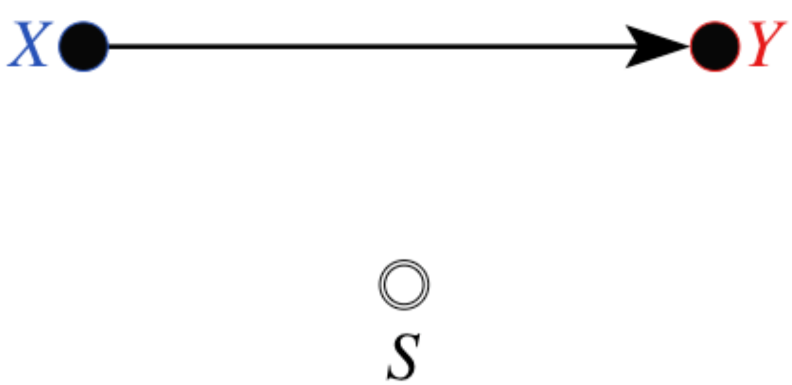
random sample



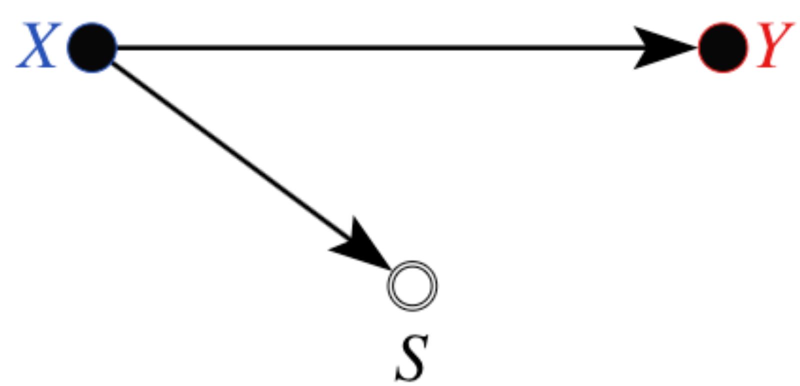
selection depends on X

Encoding the selection mechanism

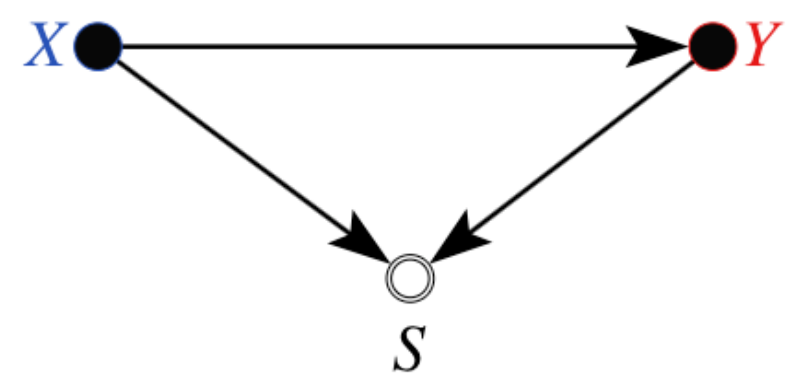
Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



random sample

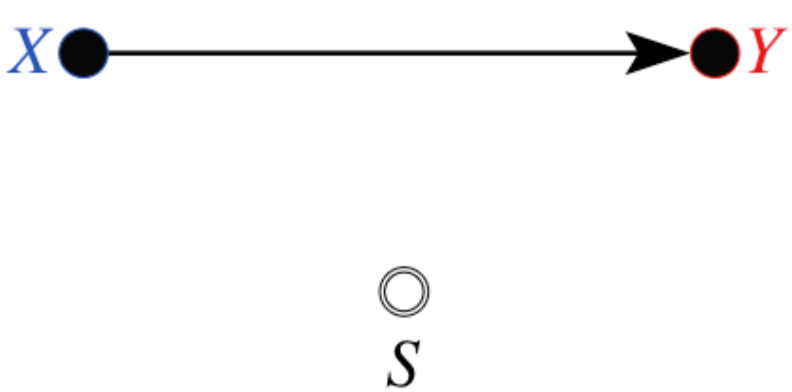


selection depends on X

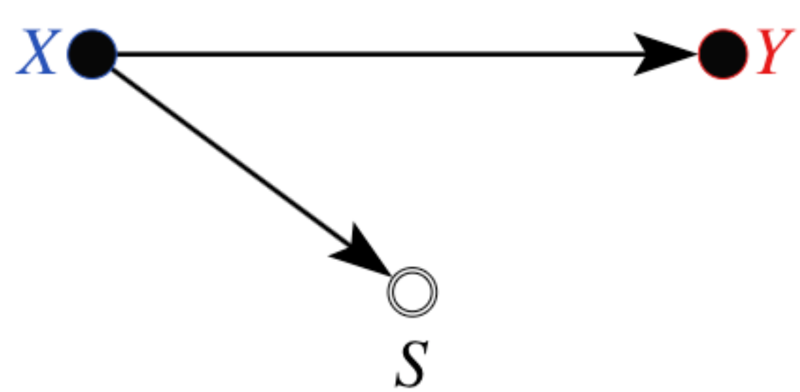


Encoding the selection mechanism

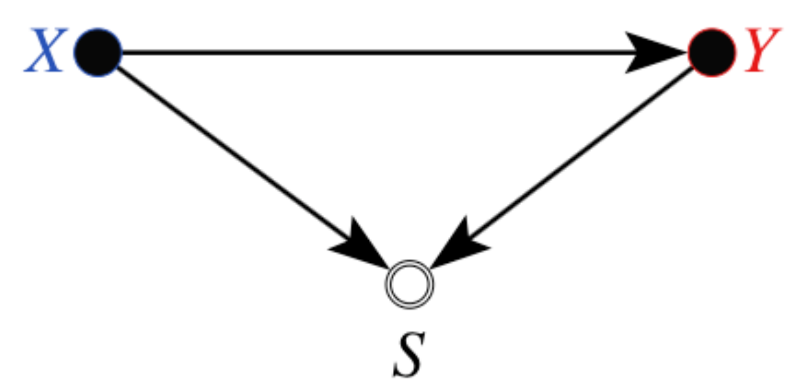
Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



random sample



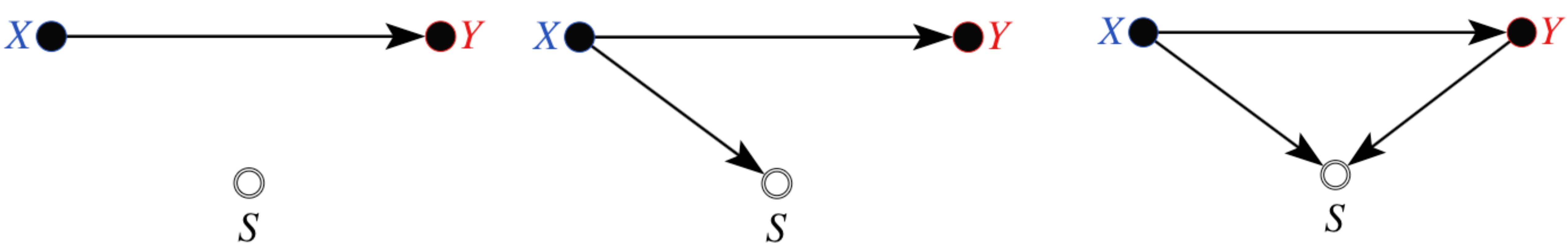
selection depends on X



selection depends on X, Y

Encoding the selection mechanism

Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.

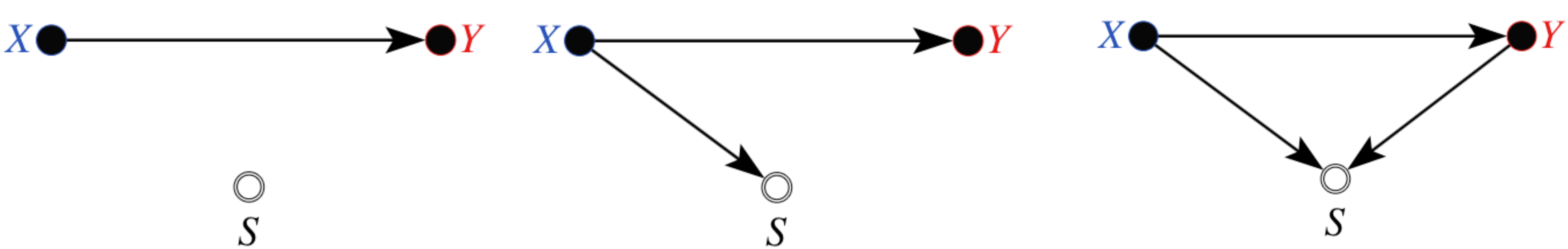


random sample | *selection depends on X* | *selection depends on X, Y*

Symbolically, our task is to express the query in terms of the available data, that is, the ***distribution under selection bias*** $P(V \mid S = 1)$ – or more concisely $P(V \mid s)$ – and the ***census data*** we have available (if any).

Encoding the selection mechanism

Again we extend our causal diagram with "selection nodes" (S) which now indicate ***selection to the study sample (S = 1), or not (S = 0).*** Our target of inference is a quantity on the population as a whole, ***not conditioning on S***.



random sample — *selection depends on X* — *selection depends on X, Y*

Symbolically, our task is to express the query in terms of the available data, that is, the ***distribution under selection bias*** $P(V \mid S = 1)$ – or more concisely $P(V \mid s)$ – and the ***census data*** we have available (if any).

Graphically, we will check for separation of the selection mechanism S from key variables of interest that compose our query.

Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

Recovering conditional distributions from selection

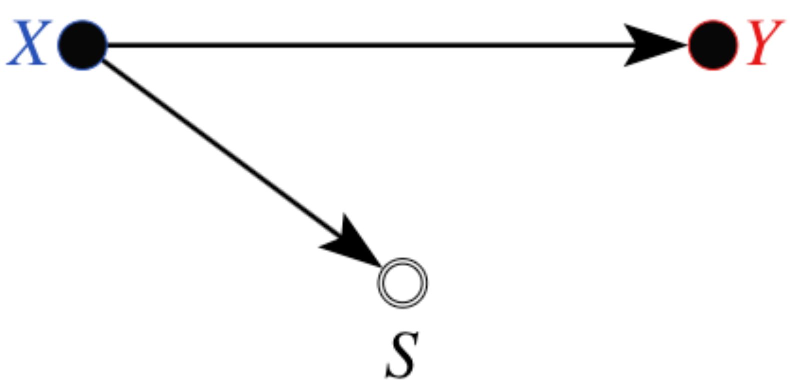
Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S | X)$

Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

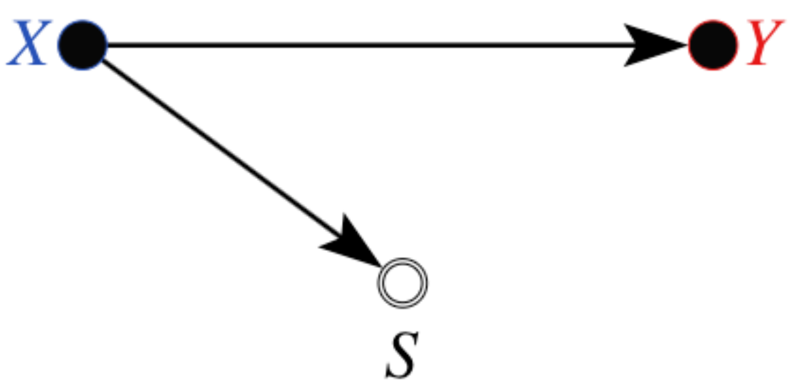
The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S | X)$



Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$

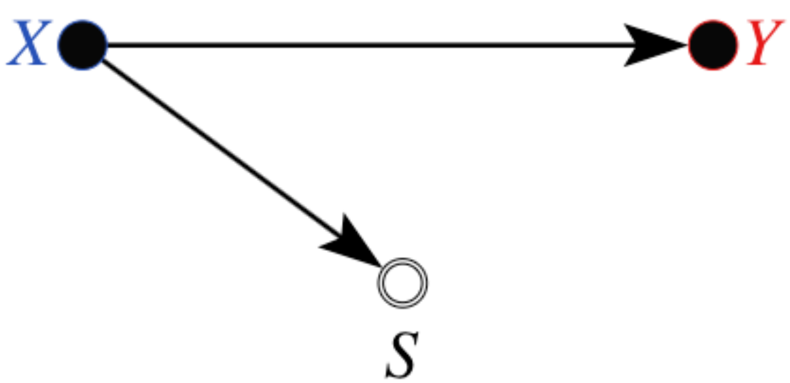


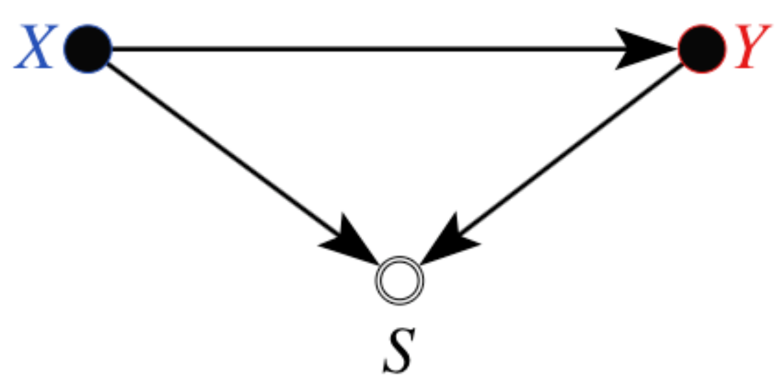
P(y | x) recoverable

Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



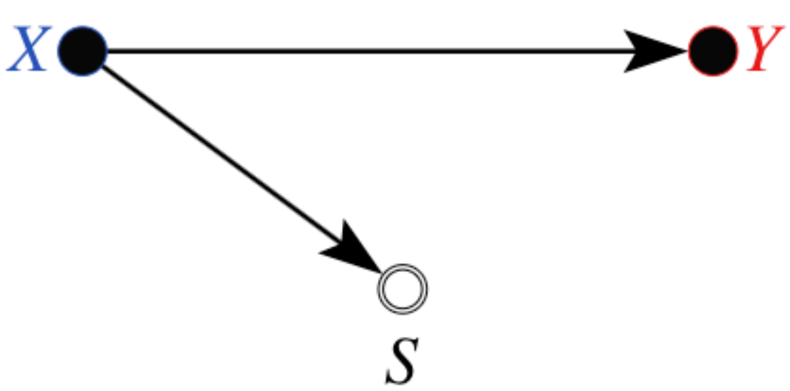


P(y | x) recoverable

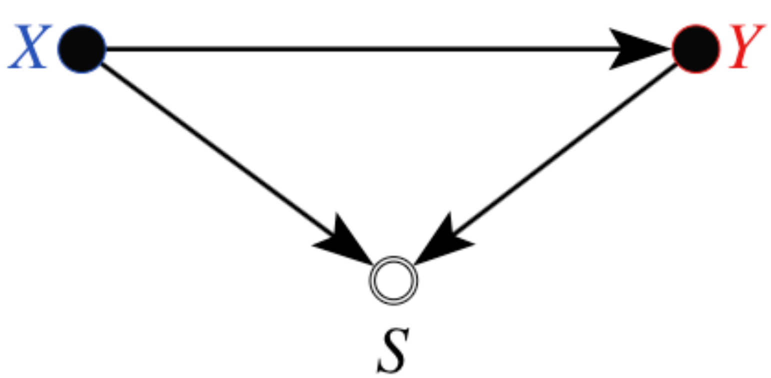
Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



P(y | x) recoverable

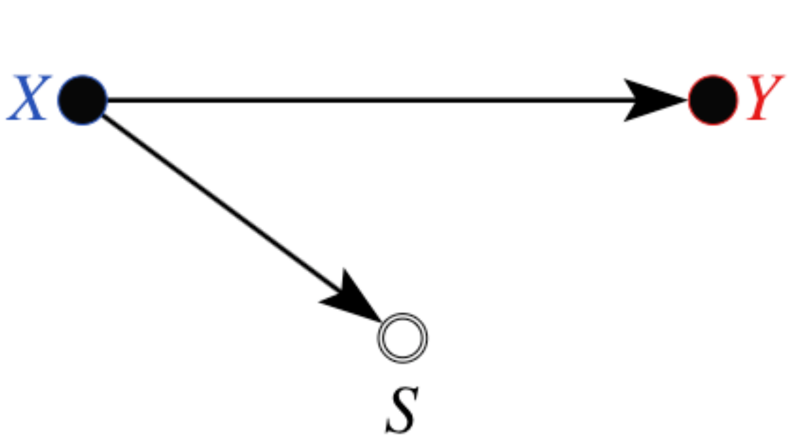


P(y | x) not recoverable

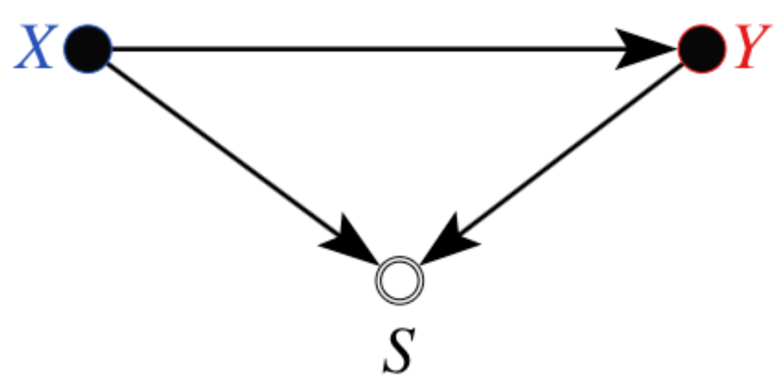
Recovering conditional distributions from selection

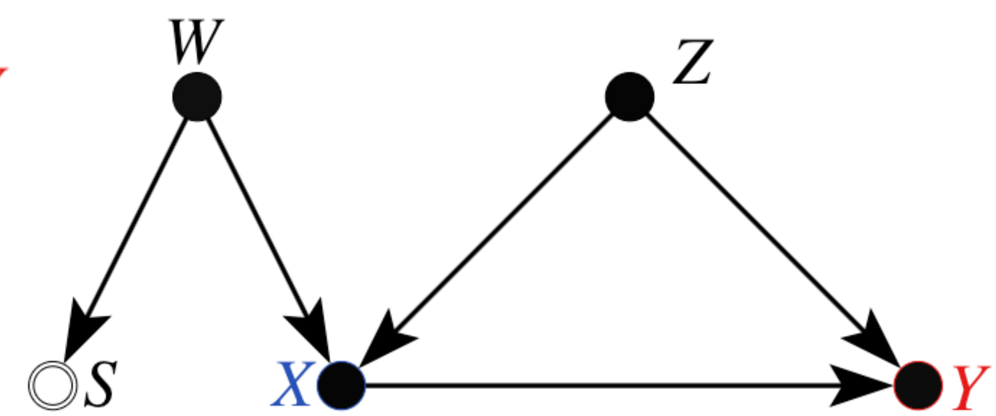
Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



P(y | x) recoverable



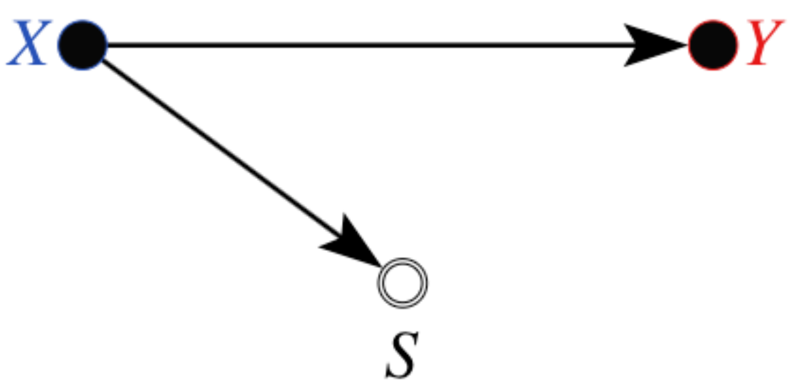


P(y | x) not recoverable

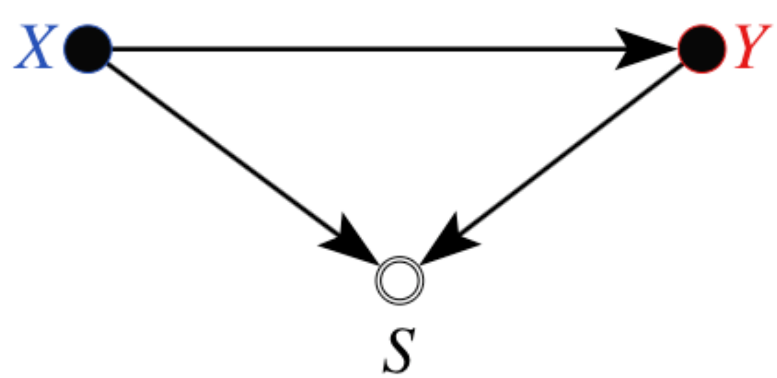
Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

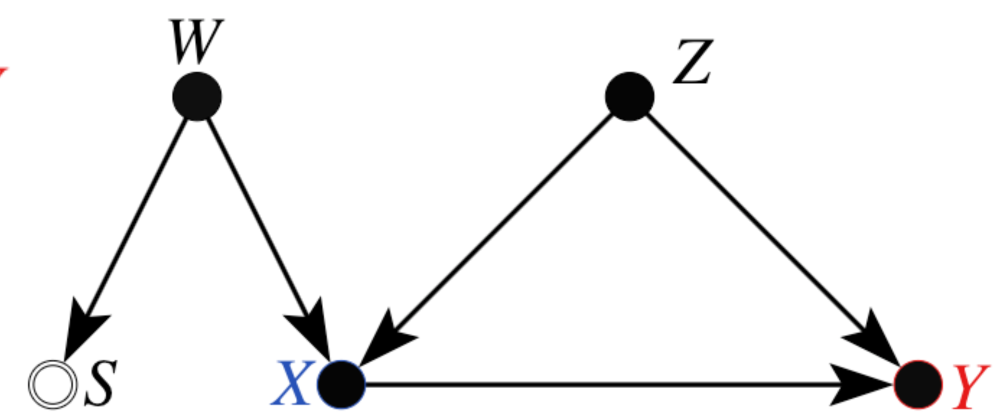
The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



P(y | x) recoverable



P(y | x) not recoverable

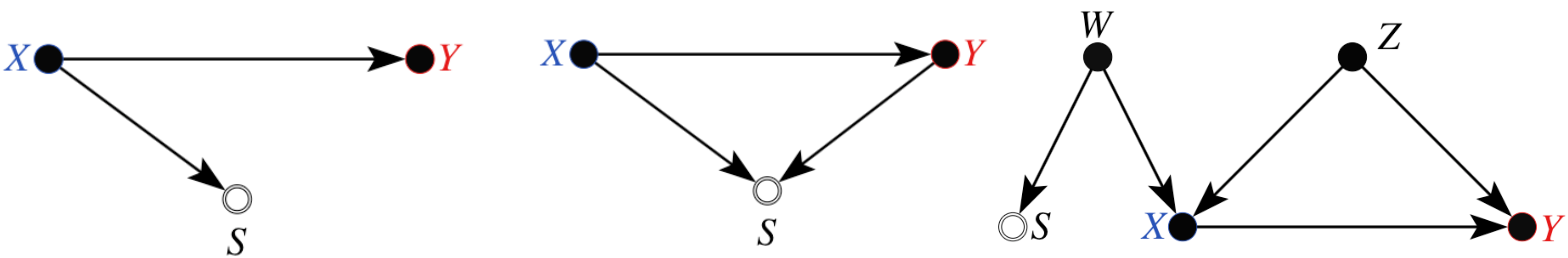


P(y | x) not recoverable

Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



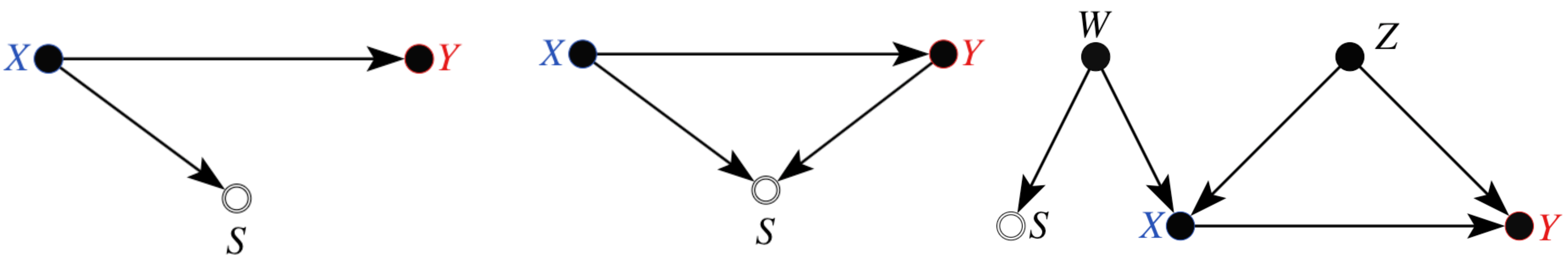
P(y | x) recoverable *P(y | x) not recoverable* *P(y | x) not recoverable*

Note this is different from recovering the causal effect P(y | do(x)).

Recovering conditional distributions from selection

Very simple necessary and sufficient condition for conditional distributions.

The conditional distribution P(y | x) is recoverable (without external data) if and only if: $(Y \perp\!\!\!\perp S \mid X)$



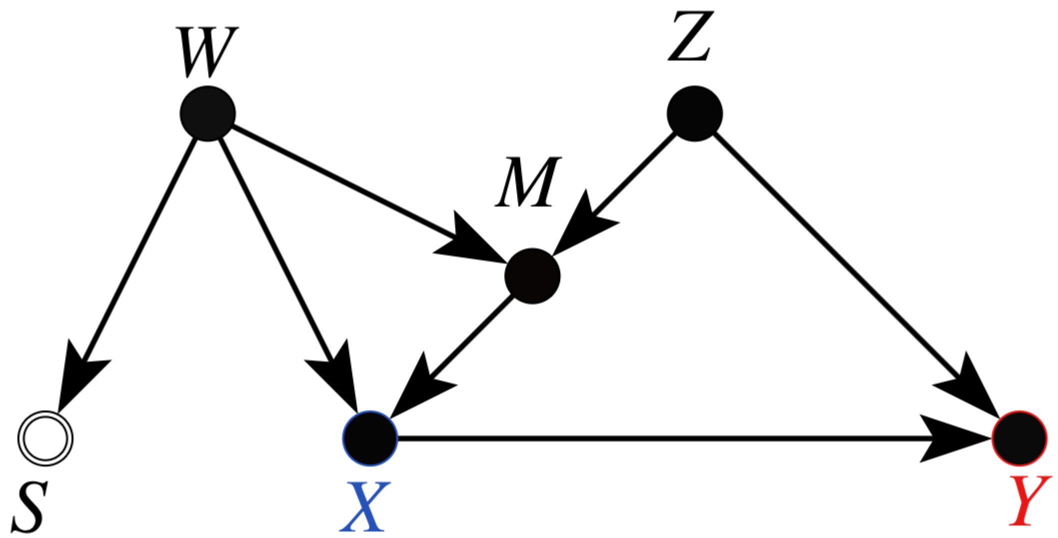
P(y | x) recoverable *P(y | x) not recoverable* *P(y | x) not recoverable*

Note this is different from recovering the causal effect P(y | do(x)).

For instance, in the third model, P(y|x) is not recoverable, while P(y|do(x)) is, as we show next.

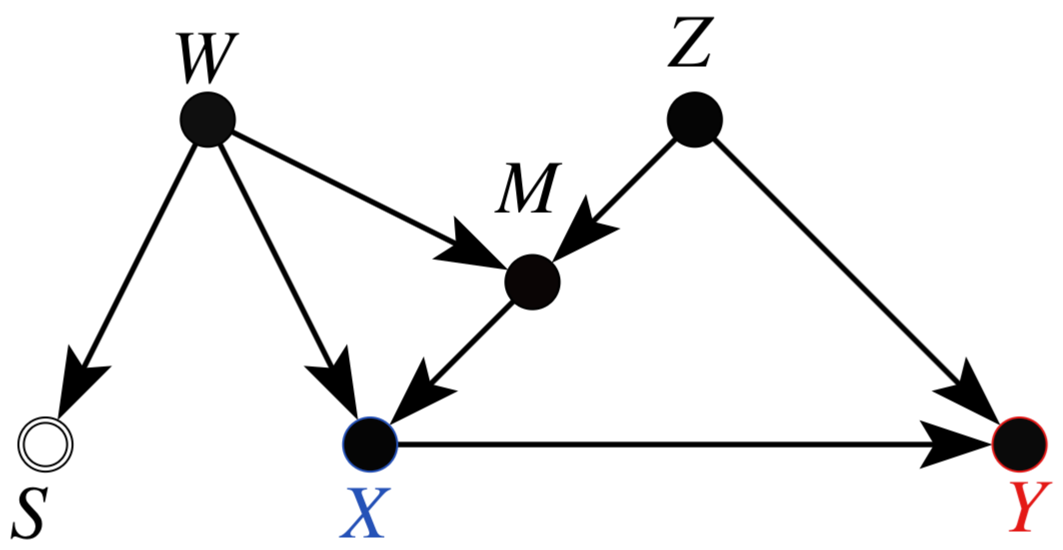
Recovering causal effects from selection and confounding

Do we need external data?



Recovering causal effects from selection and confounding

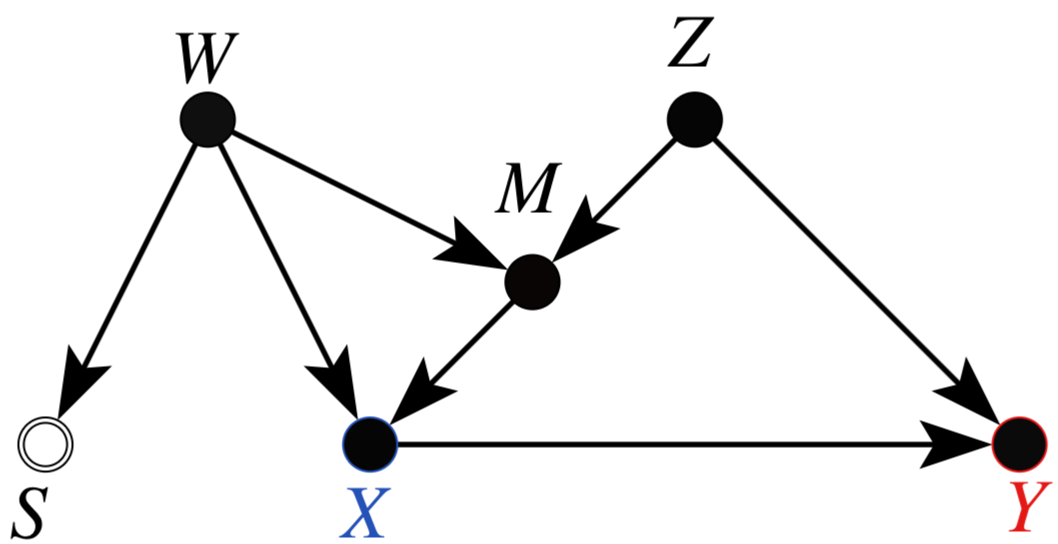
Do we need external data?



$$
\begin{aligned}
E[Y \mid do(x)] &= E[Y \mid do(x), s] \\
&= \sum_{z} E[Y \mid do(x), z, s] P(z \mid do(x), s) \\
&= \sum_{z} E[Y \mid x, z, s] P(z \mid s)
\end{aligned}
$$

Recovering causal effects from selection and confounding

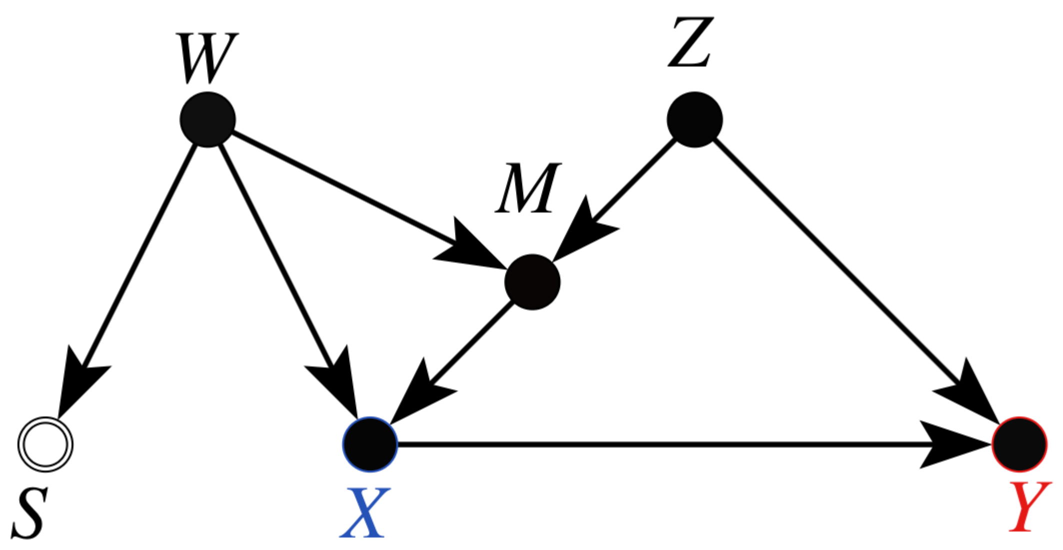
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Recovering causal effects from selection and confounding

Do we need external data?

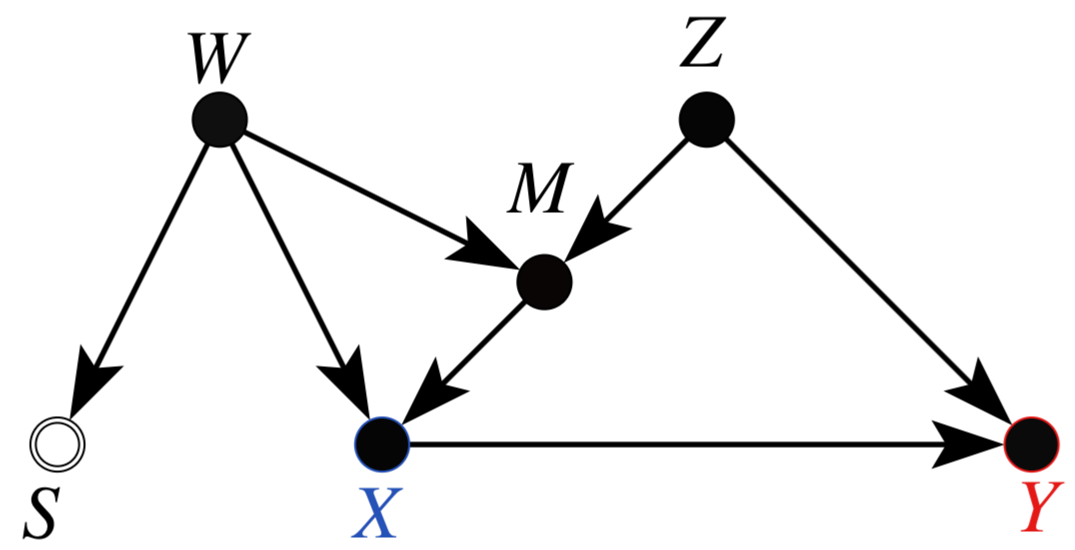


$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

Recovering causal effects from selection and confounding

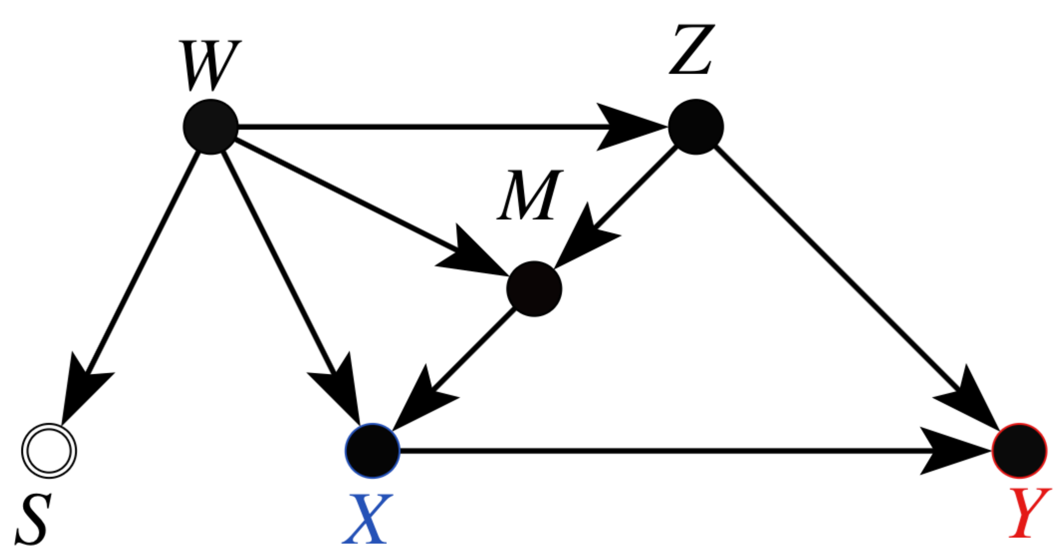
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

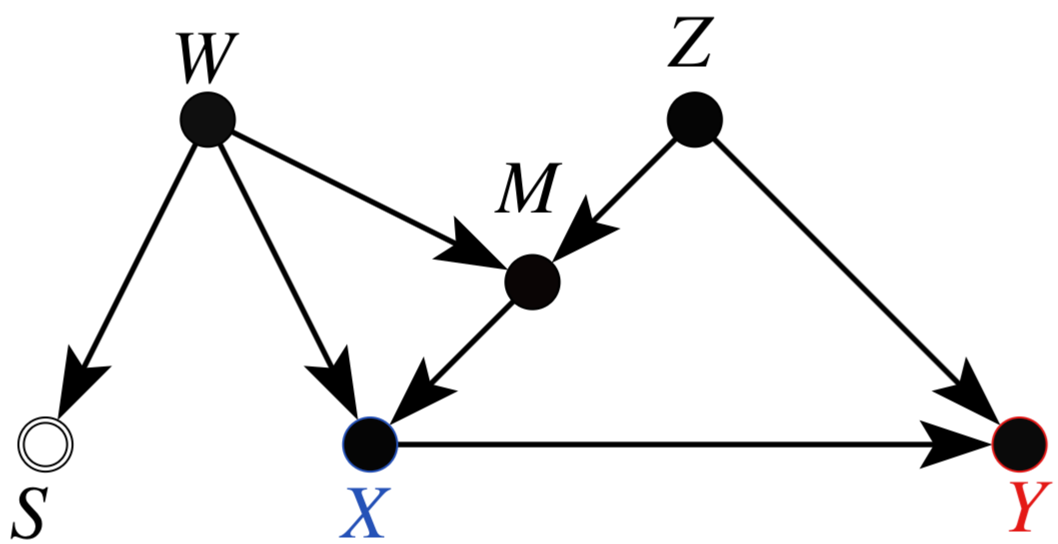
Don't need external data

Do we need external data?



Recovering causal effects from selection and confounding

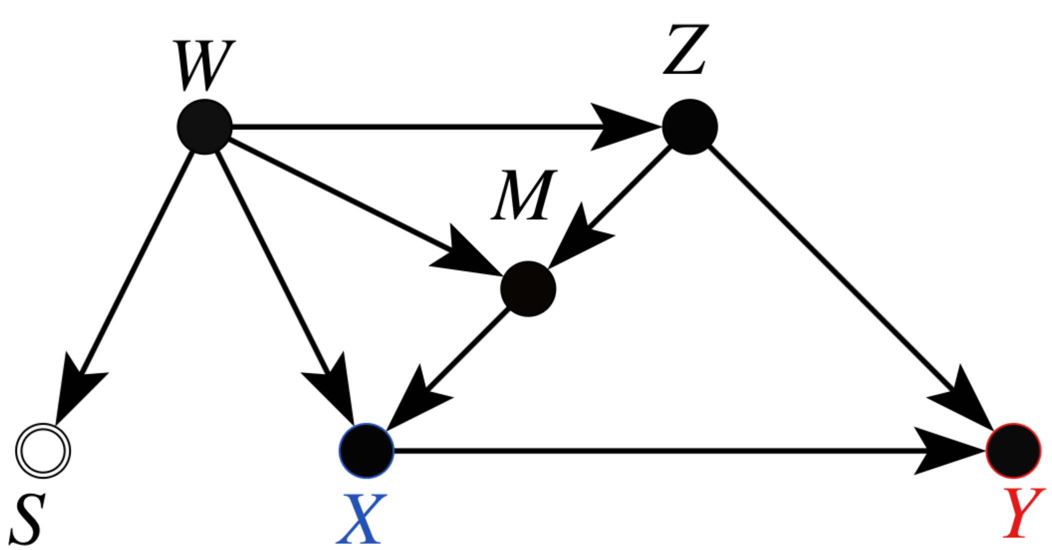
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

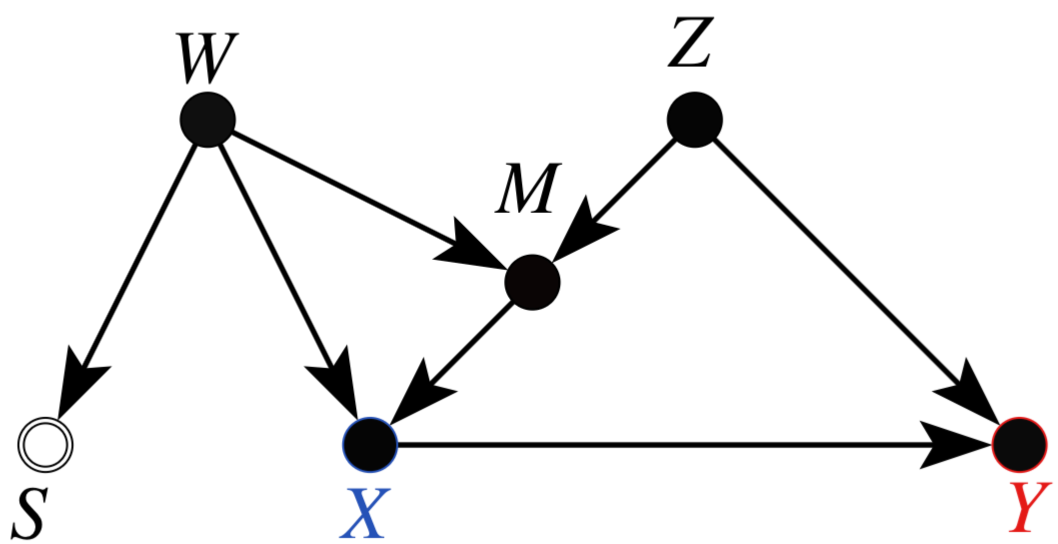
Do we need external data?



$$E[Y \mid do(x)] = \sum_z E[Y \mid do(x), z] P(z \mid do(x))$$
$$= \sum_z E[Y \mid do(x), z, s] P(z)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w) P(w)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w, s) P(w)$$

Recovering causal effects from selection and confounding

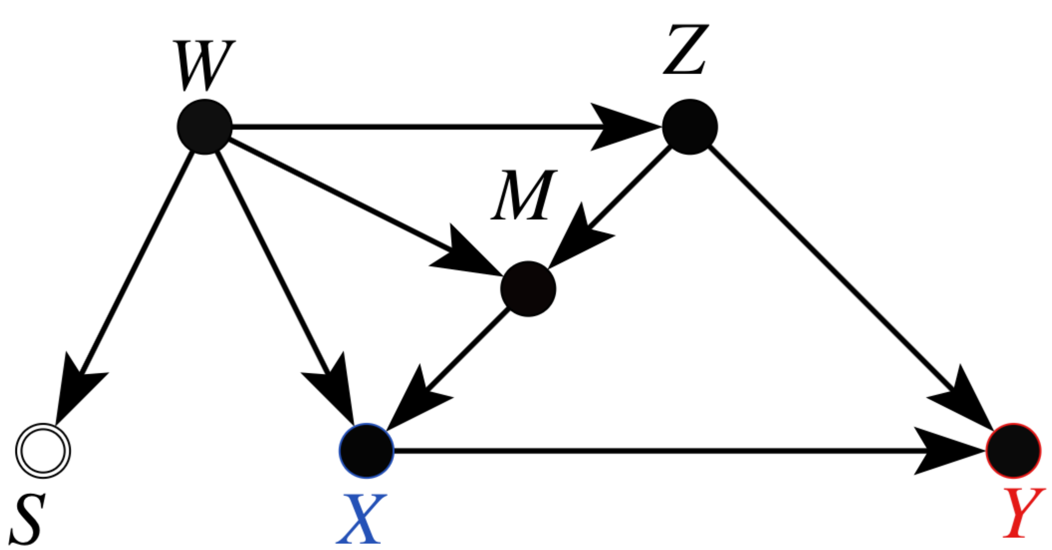
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

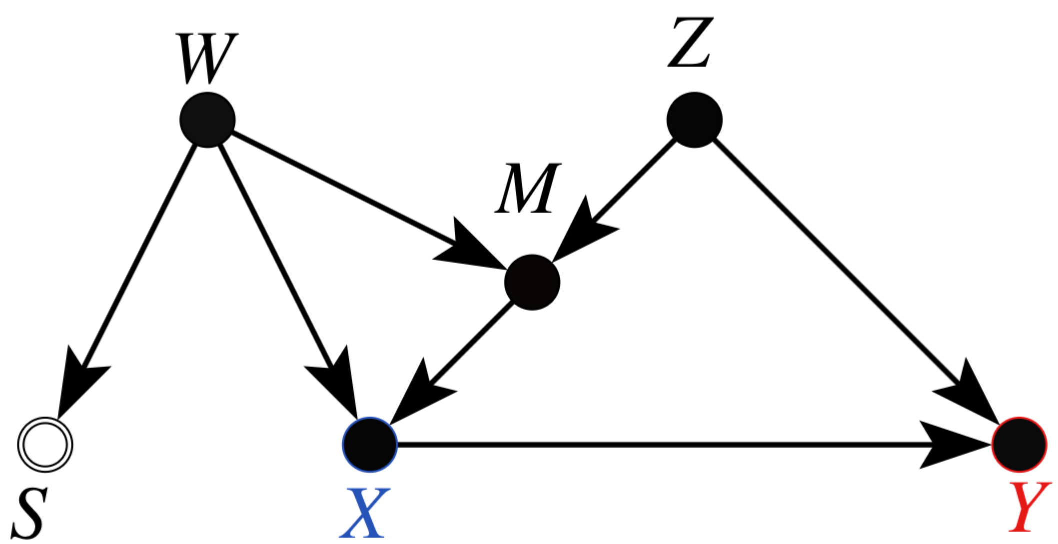
Do we need external data?



$$E[Y \mid do(x)] = \sum_z E[Y \mid do(x), z] P(z \mid do(x))$$
$$= \sum_z E[Y \mid do(x), z, s] P(z)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w) P(w)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w, s) P(w)$$

Recovering causal effects from selection and confounding

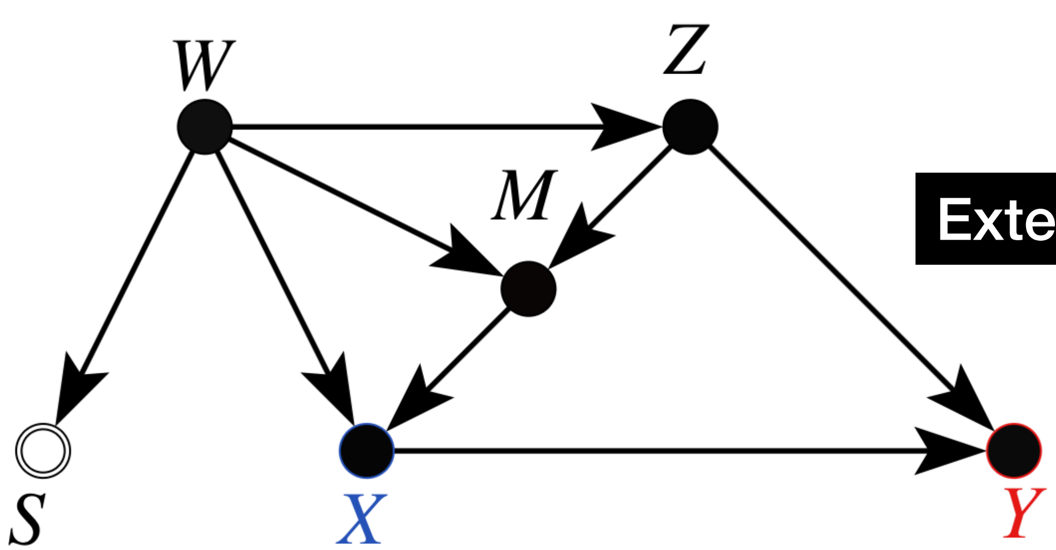
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

Do we need external data?

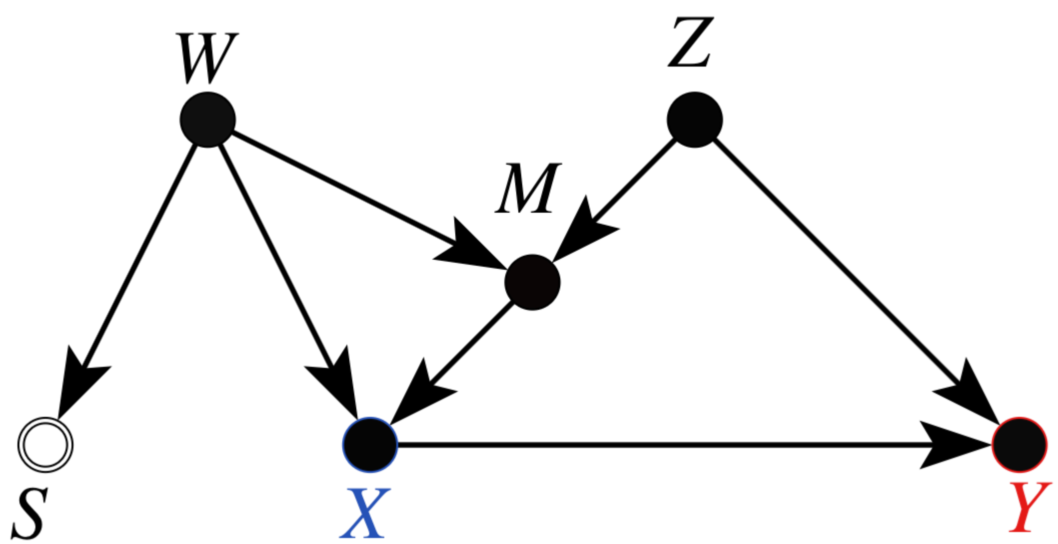


External data on Z

$$E[Y \mid do(x)] = \sum_z E[Y \mid do(x), z] P(z \mid do(x))$$
$$= \sum_z E[Y \mid do(x), z, s] P(z)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w) P(w)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w, s) P(w)$$

Recovering causal effects from selection and confounding

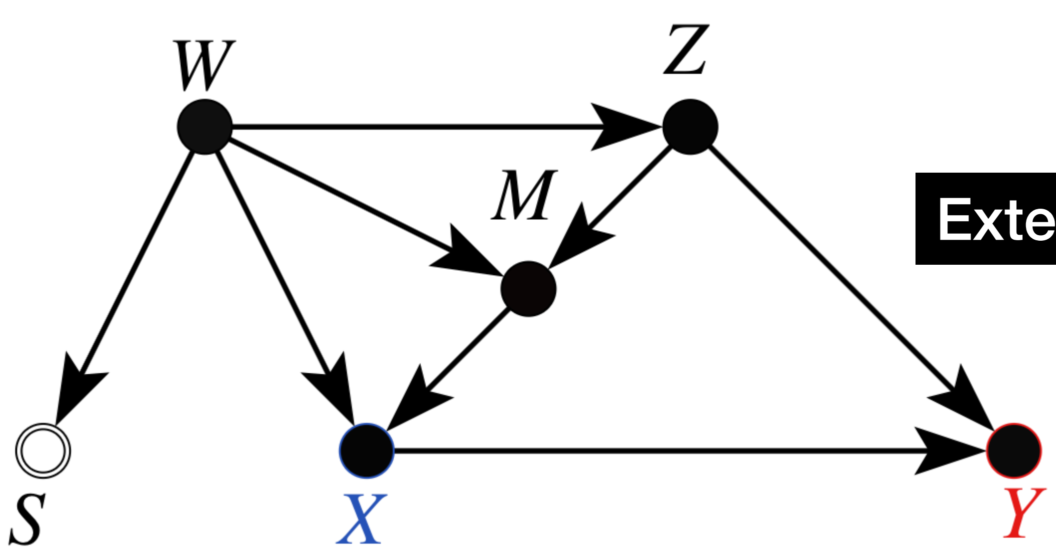
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

Do we need external data?

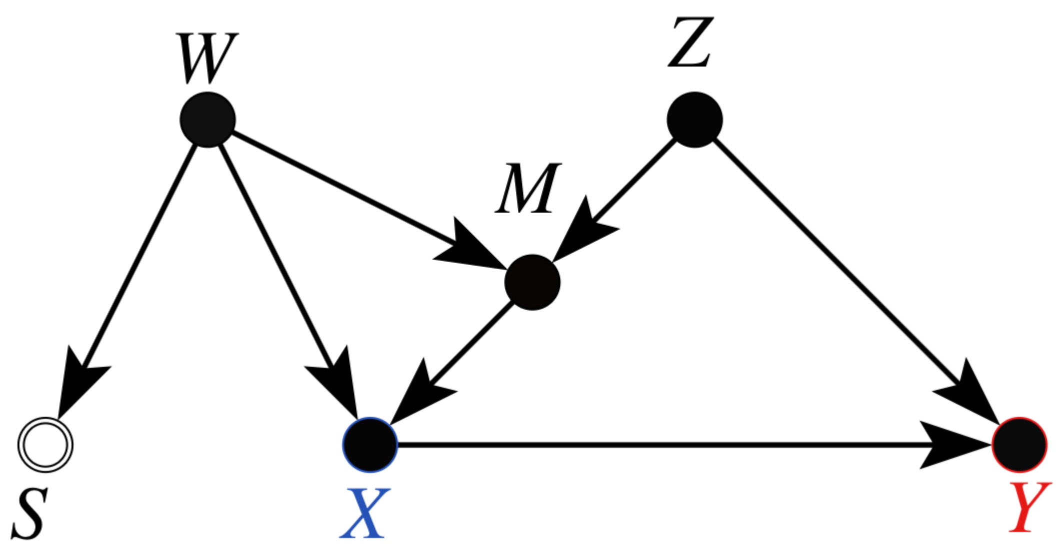


External data on Z

$$E[Y \mid do(x)] = \sum_z E[Y \mid do(x), z] P(z \mid do(x))$$
$$= \sum_z E[Y \mid do(x), z, s] P(z)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w) P(w)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w, s) P(w)$$

Recovering causal effects from selection and confounding

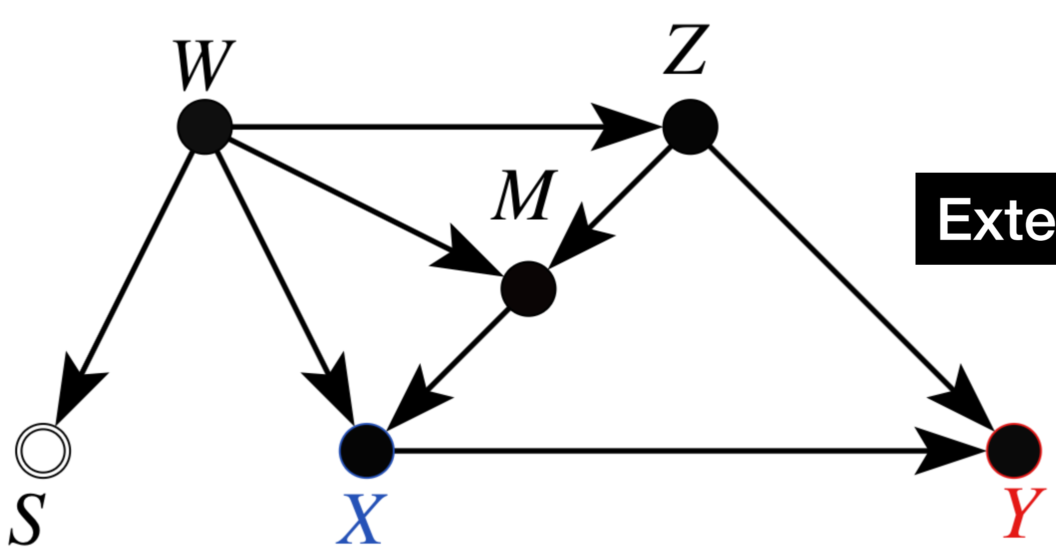
Do we need external data?



$$E[Y \mid do(x)] = E[Y \mid do(x), s]$$
$$= \sum_z E[Y \mid do(x), z, s] P(z \mid do(x), s)$$
$$= \sum_z E[Y \mid x, z, s] P(z \mid s)$$

Don't need external data

Do we need external data?



$$E[Y \mid do(x)] = \sum_z E[Y \mid do(x), z] P(z \mid do(x))$$
$$= \sum_z E[Y \mid do(x), z, s] P(z)$$

External data on Z

$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w) P(w)$$
$$= \sum_z E[Y \mid x, z, s] \sum_w P(z \mid w, s) P(w)$$

Or external data on W!

General solution to selection bias

Recovery without external data: *we have complete algorithms for recovering from selection and confounding biases, both for markovian and semi-markovian models.*

General solution to selection bias

Recovery without external data: *we have complete algorithms for recovering from selection and confounding biases, both for markovian and semi-markovian models.*

Again, why is completeness important? Completeness assures us, for instance, that Heckman's solution ***must*** rely on parametric assumptions.

General solution to selection bias

Recovery without external data: *we have complete algorithms for recovering from selection and confounding biases, both for markovian and semi-markovian models.*

Again, why is completeness important? Completeness assures us, for instance, that Heckman's solution ***must*** rely on parametric assumptions.

PS: proof of completeness is recent – Correa, Tian and Bareinboim (2019)

General solution to selection bias

Recovery without external data: *we have complete algorithms for recovering from selection and confounding biases, both for markovian and semi-markovian models.*

Again, why is completeness important? Completeness assures us, for instance, that Heckman's solution ***must*** rely on parametric assumptions.

PS: proof of completeness is recent – Correa, Tian and Bareinboim (2019)

Recovery using external data: *still an* **open question** *whether the current state-of-the-art algorithm is complete.*

General solution to selection bias

Recovery without external data: *we have complete algorithms for recovering from selection and confounding biases, both for markovian and semi-markovian models.*

Again, why is completeness important? Completeness assures us, for instance, that Heckman's solution ***must*** rely on parametric assumptions.

PS: proof of completeness is recent – Correa, Tian and Bareinboim (2019)

Recovery using external data: *still an* **open question** *whether the current state-of-the-art algorithm is complete.*

FUSION DEMO 2

Data Fusion

$(d_1, d_2, d_3, d_4) \rightarrow (d'_1, d'_2, d'_3, d'_4)$

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

		Dataset 1	**Dataset 2**	**Dataset 3**
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat.Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

		Dataset 1	Dataset 2	Dataset 3
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat.Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

-Observational Causal Inference: $(d_1, see(x), d_3, d_4) \rightarrow (d_1, do(x), d_3, d_4)$

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

		Dataset 1	Dataset 2	Dataset 3
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat.Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

-Observational Causal Inference: $(d_1, see(x), d_3, d_4) \rightarrow (d_1, do(x), d_3, d_4)$

- Sampling Selection Bias: $(d_1, d_2, select(age), d_4) \rightarrow (d_1, d_2, \{\}, d_4)$

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

		Dataset 1	Dataset 2	Dataset 3
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat.Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

-Observational Causal Inference: $(d_1, see(x), d_3, d_4) \rightarrow (d_1, do(x), d_3, d_4)$

- Sampling Selection Bias: $(d_1, d_2, select(age), d_4) \rightarrow (d_1, d_2, \{\}, d_4)$

- Transportability: $(LA, d_2, d_3, d_4) \rightarrow (NY, d_2, d_3, d_4)$

Putting it all together

We can describe each data collection as the tuple:

$(d_1, d_2, d_3, d_4) =$ (population, obs/exp., sampling selection, observed data)

		Dataset 1	Dataset 2	Dataset 3
d_1	Population	Los Angeles	New York	Texas
d_2	Obs. / Exp.	Experimental	Observational	Experimental
	Treat.Assign.	Randomized Z_1	-	Randomized Z_2
d_3	Sampling	Selection on Age	Selection on SES	-
d_4	Measured	X_1, Z_1, W, M, Y_1	X_1, X_2, Z_1, N, Y_2	X_2, Z_1, W, L, M, Y_1

-Observation ... $do(x), d_3, d_4)$

- Sampling Se ... $(d_1, d_2, \{\}, d_4)$

- Transportab ... $, d_3, d_4)$

In general: $(d_1, d_2, d_3, d_4) \rightarrow (d_1', d_2', d_3', d_4')$

Conclusions

Generalizing causal knowledge from heterogenous datasets require encoding assumptions about the data generating process.

Conclusions

Generalizing causal knowledge from heterogenous datasets require encoding assumptions about the data generating process.

The structural theory of causation combines graphical models, structural equations and potential outcomes to represent and tackle common problems of selection bias, transportability, and data fusion more generally.

Conclusions

Generalizing causal knowledge from heterogenous datasets require encoding assumptions about the data generating process.

The structural theory of causation combines graphical models, structural equations and potential outcomes to represent and tackle common problems of selection bias, transportability, and data fusion more generally.

This has led to necessary and sufficient conditions that fully characterize transportability and selection bias (non parameterically), as well as complete algorithms for finding those solutions (when they exist).

Conclusions

Generalizing causal knowledge from heterogenous datasets require encoding assumptions about the data generating process.

The structural theory of causation combines graphical models, structural equations and potential outcomes to represent and tackle common problems of selection bias, transportability, and data fusion more generally.

This has led to necessary and sufficient conditions that fully characterize transportability and selection bias (non parameterically), as well as complete algorithms for finding those solutions (when they exist).

Software under development: Causal Fusion.

Conclusions

Generalizing causal knowledge from heterogenous datasets require encoding assumptions about the data generating process.

The structural theory of causation combines graphical models, structural equations and potential outcomes to represent and tackle common problems of selection bias, transportability, and data fusion more generally.

This has led to necessary and sufficient conditions that fully characterize transportability and selection bias (non parameterically), as well as complete algorithms for finding those solutions (when they exist).

Software under development: Causal Fusion.

Thank you!

References

[1] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. Proceedings of the National Academy of Sciences, 113(27):7345–7352, 2016.

[2] E. Bareinboim and J. Pearl. Transportability from multiple environments with limited experiments: Completeness results. In Advances in neural information processing systems, 2014.

[3] E. Bareinboim and J. Tian. Recovering causal effects from selection bias. In Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015.

[4] E. Bareinboim, J. Tian, and J. Pearl. Recovering from selection bias in causal and statistical inference. In Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.

[5] J. Correa, J. Tian, and E. Bareinboim. Adjustment criteria for generalizing experimental findings. In International Conference on Machine Learning (ICML), 2019.

[6] J. D. Correa, J. Tian, and E. Bareinboim. Identification of causal effects in the presence of selection bias. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI), 2019.

[7] D. R. Cox. Planning of experiments. John Wiley and Sons, NY, 1958.

[8] C. F. Manski. Identification for prediction and decision. Harvard University Press, 2009.

[9] J. Pearl. Causality. Cambridge university press, 2009.

[10] J. Pearl. Generalizing experimental findings. Journal of Causal Inference, 3(2):259–266, 2015.

[11] J. Pearl, E. Bareinboim, et al. External validity: From do-calculus to transportability across populations. Statistical Science, 29(4):579–595, 2014.

[12] J. Pearl and D. Mackenzie. The Book of Why: The New Science of Cause and Effect. Hachette UK, 2018.

[13] W. R. Shadish, T. D. Cook, and D. T. Campbell. Experimental and quasi-experimental designs for generalized causal inference. Houghton-Mifflin, Boston, second edition, 2002.