



3

THE EFFECTS OF INTERVENTIONS

3.1 INTERVENTIONS



When we collect data on factors associated with wildfires, we are actually searching for something we can intervene upon in order to decrease wildfire frequency.

When we perform a study on a new cancer drug, we are trying to identify how a patient's illness responds when we intervene upon it by medicating the patient.



When we research the correlation between violent television and acts of aggression in children, we are trying to determine whether intervening to reduce children's access to violent television will reduce their aggressiveness.

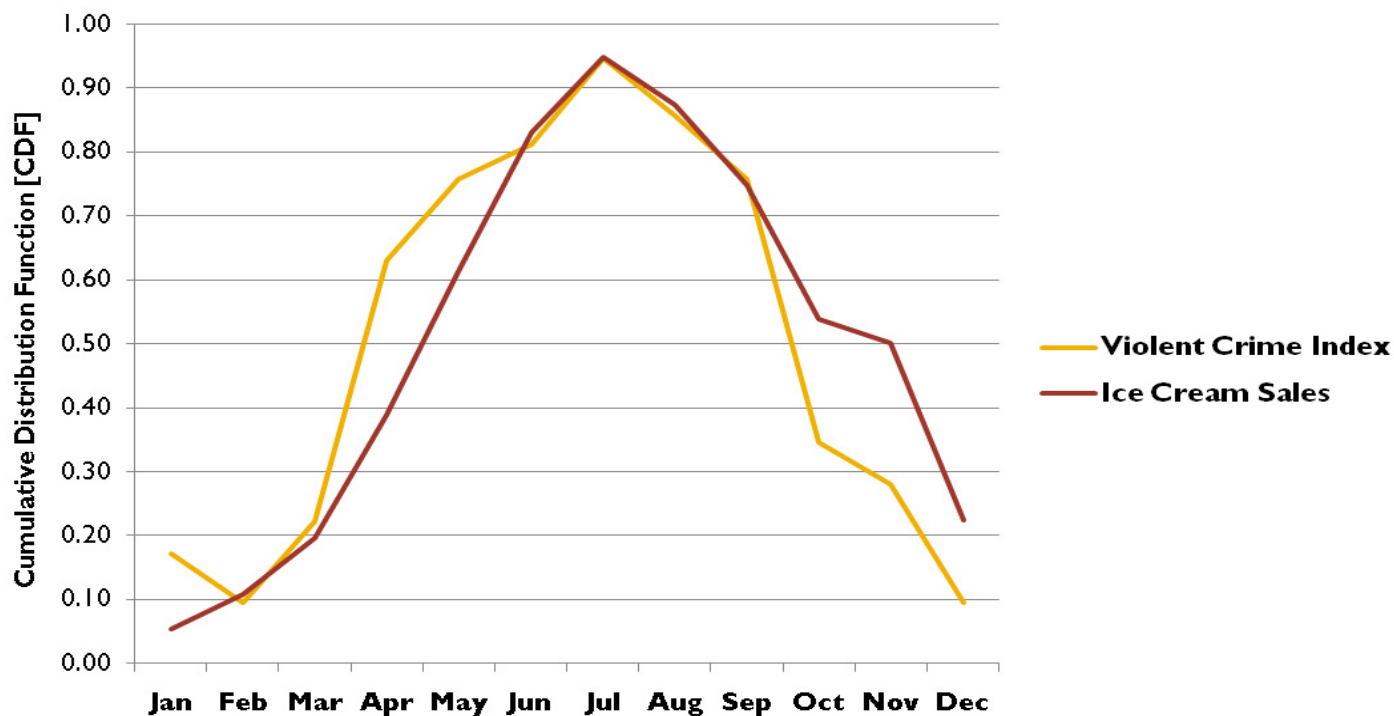
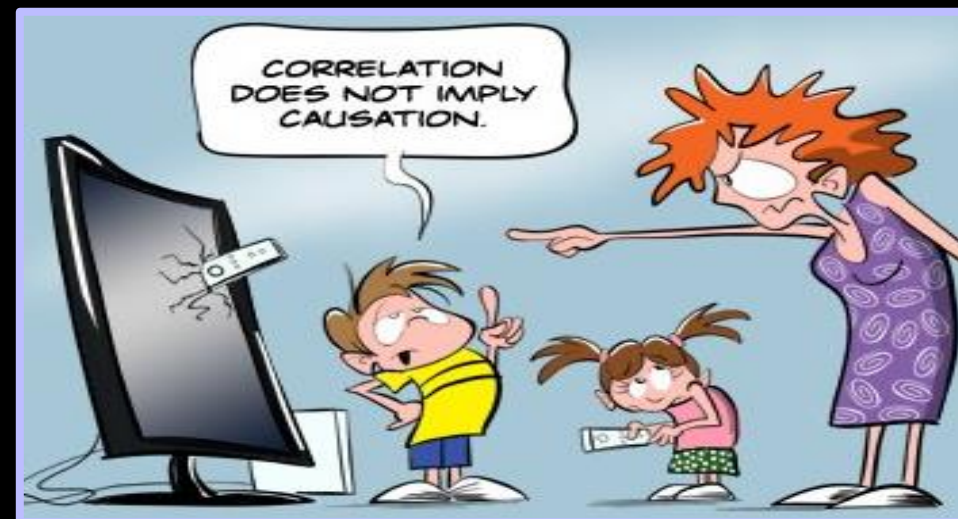
The ultimate aim of many statistical studies is to predict the effects of interventions.

3.1 INTERVENTIONS

As you have undoubtedly heard many times in statistics classes,

“correlation is not causation.”

A mere association between two variables does not necessarily or even usually mean that one of those variables causes the other.

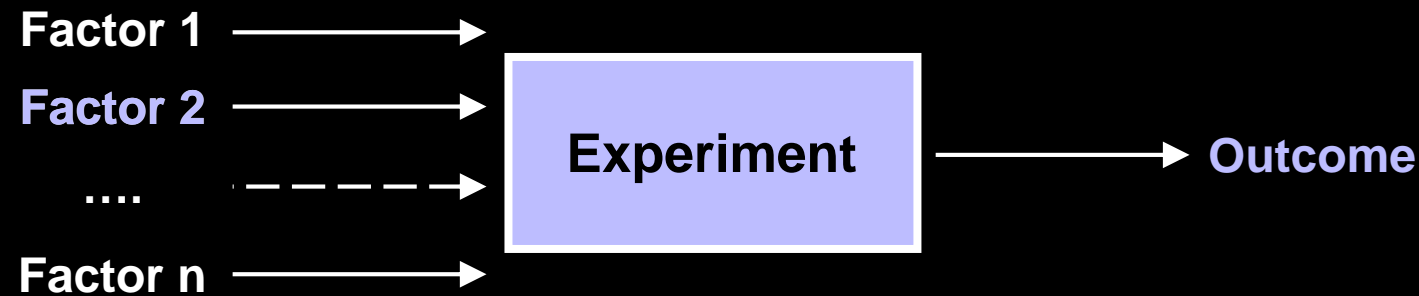


The famous example of this property is that an increase in ice cream sales is correlated with an increase in violent crime—not because ice cream causes crime, but because both ice cream sales and violent crime are more common in hot weather.

3.1 INTERVENTIONS

For this reason, the **randomized controlled experiment** is considered the golden standard of statistics.

In a properly randomized controlled experiment, all factors that influence the outcome variable are either static, or vary at random, except for one—so any change in the outcome variable must be due to that one input variable (factor).



**static, or vary at random,
except for one**

Factor 1 static, all other factors
vary at random, except **Factor 2**

**any change of the outcome variable must
be due to that one input variable (factor)**

If the value of **Outcome** changes,
then it is due to **Factor 2**

3.1 INTERVENTIONS

Unfortunately, many questions do not lend themselves to randomized controlled experiments.



We cannot control the weather, so we can't randomize the variables that affect wildfires.

Even randomized drug trials can run into problems when **participants drop out, fail to take their medication, or misreport their usage.**



We could conceivably randomize the participants in a study about violent television, but it would be **difficult to effectively control how much television each child watches**, and nearly impossible to know whether we were controlling them effectively or not.



3.1 INTERVENTIONS

In cases where randomized controlled experiments are not practical, researchers instead perform **observational studies**, in which they merely record data, rather than controlling it.

The problem of such studies is that it is difficult to untangle the causal from the merely correlative.

Our **common sense** tells us that intervening on ice cream sales is unlikely to have any effect on crime, but the facts are not always so clear.



Reducing
sales

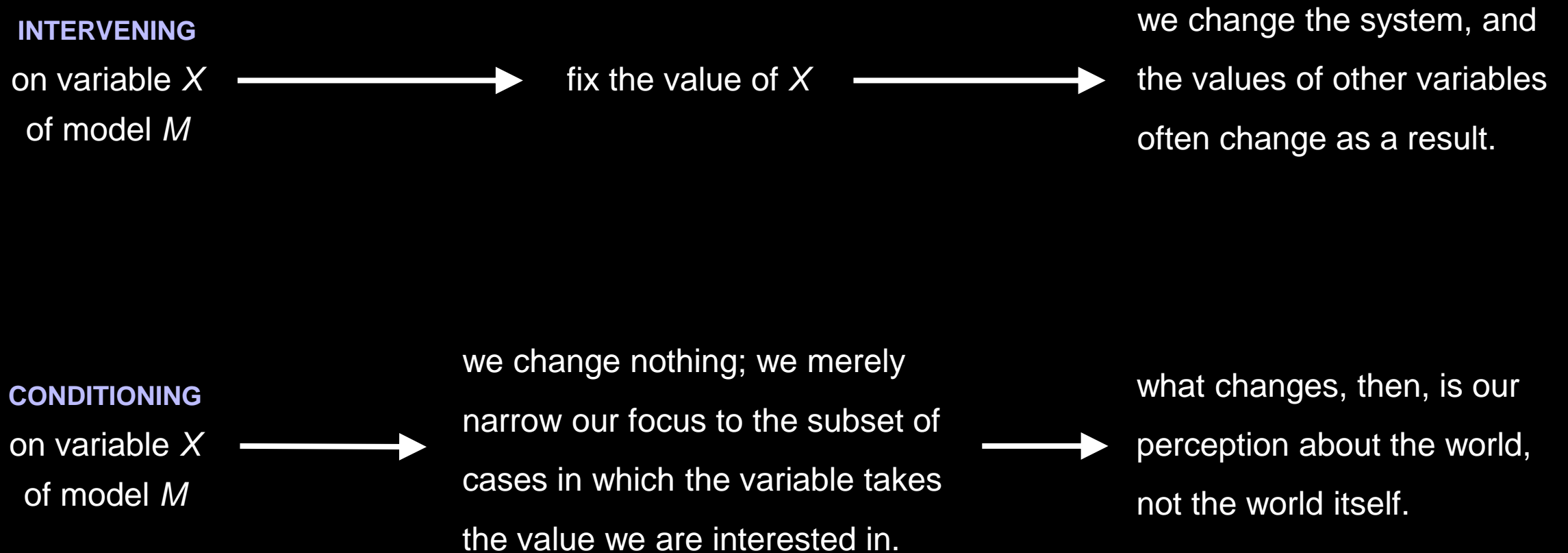


Reduces
crimes



3.1 INTERVENTIONS

The difference between **intervening** on a variable and **conditioning** on that variable should, hopefully, be obvious.

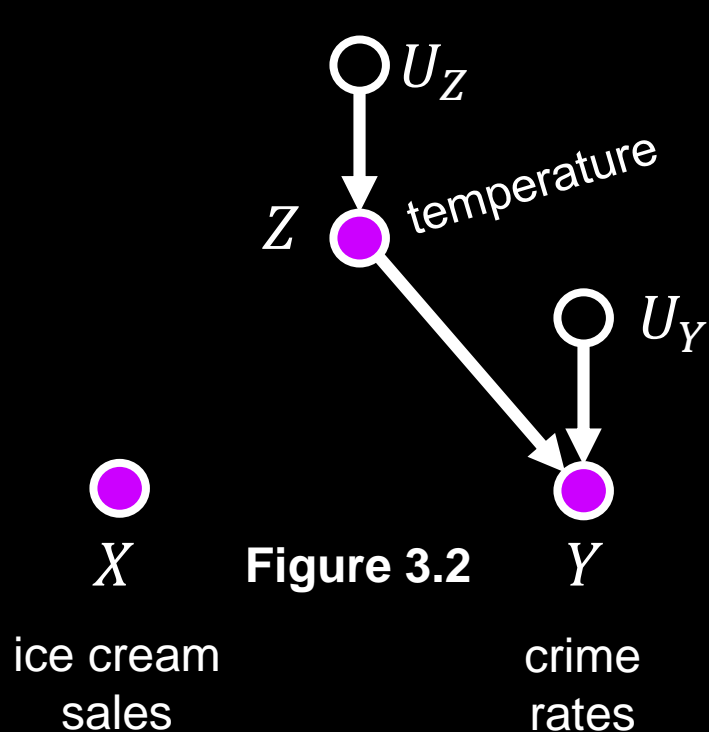
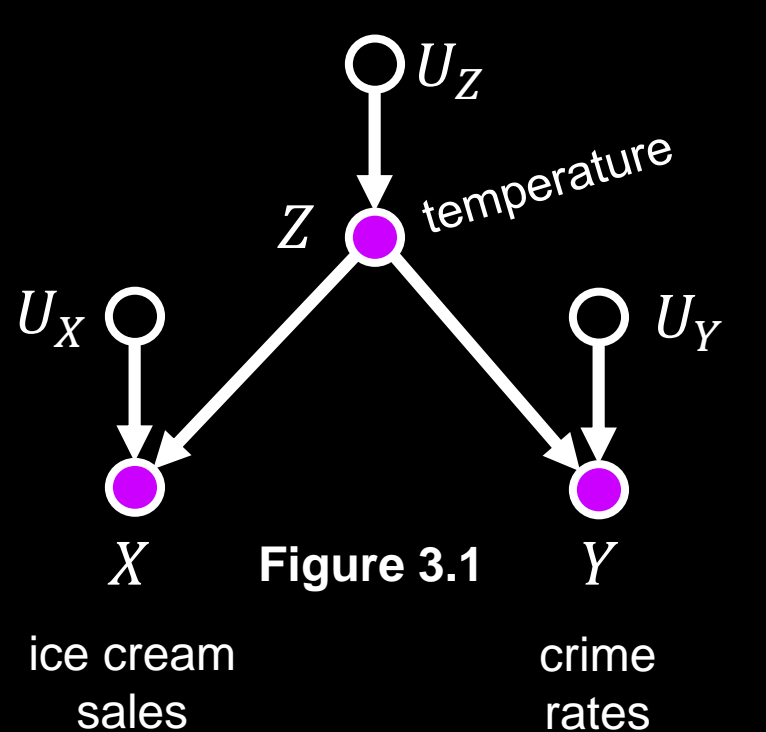


3.1 INTERVENTIONS

Consider, for instance, **Figure 3.1** that shows a graphical model of our ice cream sales example, with

- X as ice cream sales,
- Y as crime rates, and
- Z as temperature.

When we intervene to fix the value of a variable, we curtail the natural tendency of that variable to vary in response to other variables in nature.



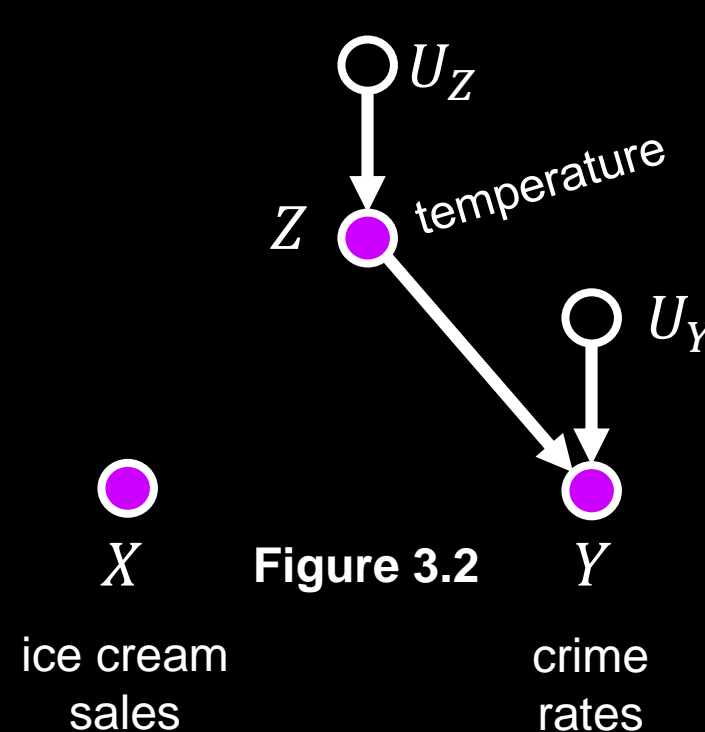
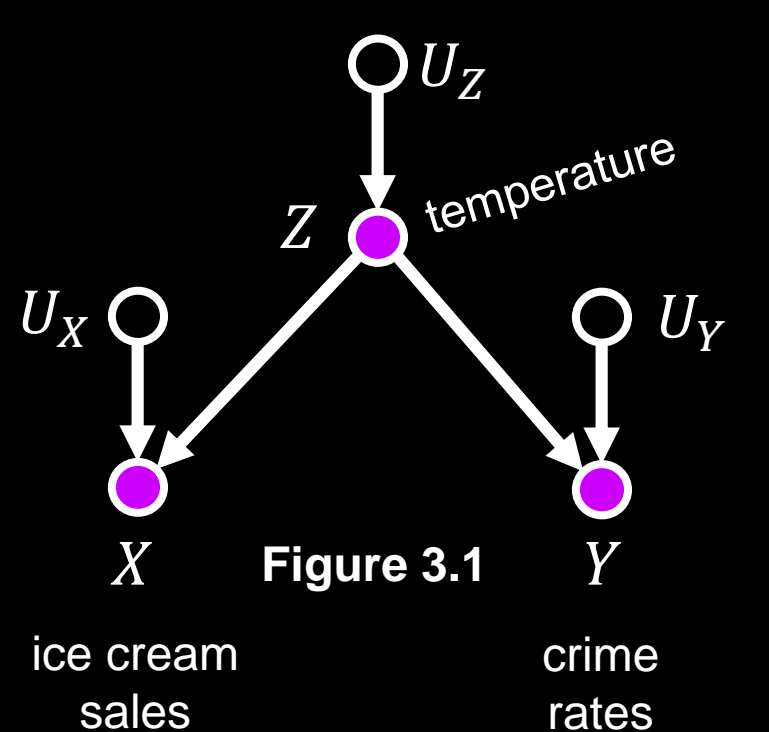
This amounts to performing a kind of **surgery on the graphical model**, removing all edges directed into that variable.

If we were to intervene to make ice cream sales (X) low (say, by shutting down all ice cream shops), we would have the graphical model shown in **Figure 3.2**.

3.1 INTERVENTIONS

When we examine correlations in this new graph (**Figure 3.2**), we find that crime rates (Y) are, totally independent of (i.e., uncorrelated with) ice cream sales (X) since the latter is no longer associated with temperature (Z).

In other words, even if we vary the level at which we hold X (ice cream sales) constant, that variation will not be transmitted to variable Y (crime rates).



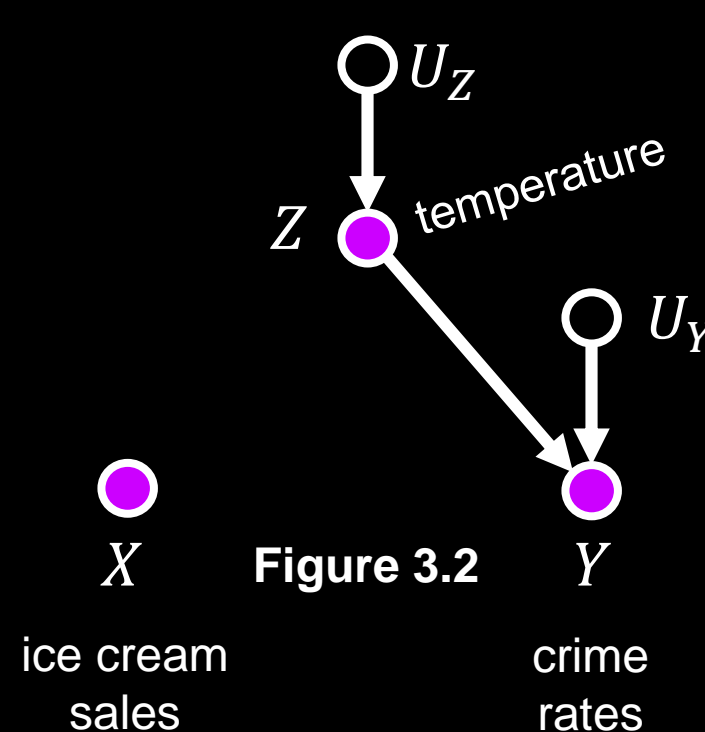
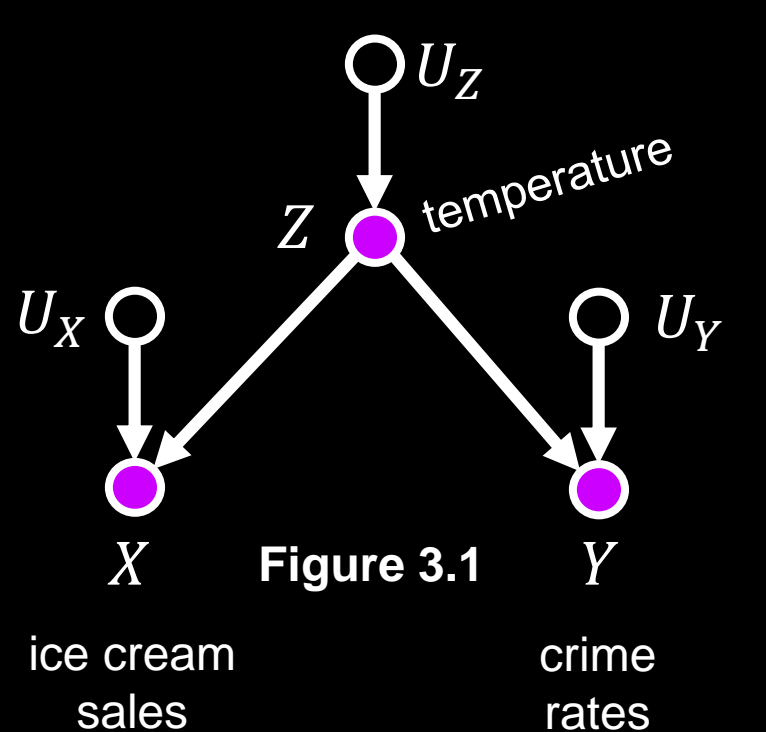
We see that intervening on a variable results in a totally different pattern of dependencies than conditioning on a variable.

Moreover, the **conditioning pattern** can be obtained directly from the data set, using the procedures described in **Part 1**, while the **intervening pattern** varies depending on the structure of the causal graph.

3.1 INTERVENTIONS

When we examine correlations in this new graph (**Figure 3.2**), we find that crime rates (Y) are, totally independent of (i.e., uncorrelated with) ice cream sales (X) since the latter is no longer associated with temperature (Z).

It is the graph that instructs us which arrow should be removed for any given intervention.



We see that intervening on a variable results in a totally different pattern of dependencies than conditioning on a variable.

Moreover, the **conditioning pattern** can be obtained directly from the data set, using the procedures described in **Part 1**, while the **intervening pattern** varies depending on the structure of the causal graph.

3.1 INTERVENTIONS

In notation, we distinguish between cases where

- a variable X takes a value x naturally

$$X = x$$

$$P(Y = y|X = x)$$

probability that $Y = y$ conditional on finding $X = x$

population distribution of Y among individuals whose X value is x .

- we fix $X = x$

$$do(X = x)$$

$P(Y = y|do(X = x))$ probability that $Y = y$ when we intervene to make $X = x$

population distribution of Y if everyone in the population had their X value fixed at x .

$$P(Y = y|do(X = x), Z = z)$$

conditional probability of $Y = y$, given $Z = z$, in the distribution created by the intervention $do(X = x)$.

3.1 INTERVENTIONS

Using **do-expressions** and **graph surgery**, we can begin to **untangle the causal relationships from the correlative**.

We now learn methods that can, astoundingly, tease out causal information from purely observational data, assuming of course that the graph constitutes a valid representation of reality.

It is worth noting here that we are making a tacit assumption that

- **The intervention** has no “**side effects**,” that is, that assigning the value x for the variable X for an individual does not alter subsequent variables in a direct way.

For example,

- being “assigned” a drug might have a different effect on recovery than
- being forced to take the drug against one’s religious objections.

When side effects are present, they need to be specified explicitly in the model.

3.2 THE ADJUSTMENT FORMULA

The ice cream example represents an extreme case in which the correlation between X and Y was totally spurious from a causal perspective, because there was no causal path from X to Y .

Most real-life situations are not so clear-cut. To explore a more realistic situation, let us examine **Figure 3.3**, in which Y responds to both Z and X .

Such a model could represent, for example, the first story we encountered for **Simpson's paradox**, where X stands for drug usage, Y stands for recovery, and Z stands for gender.

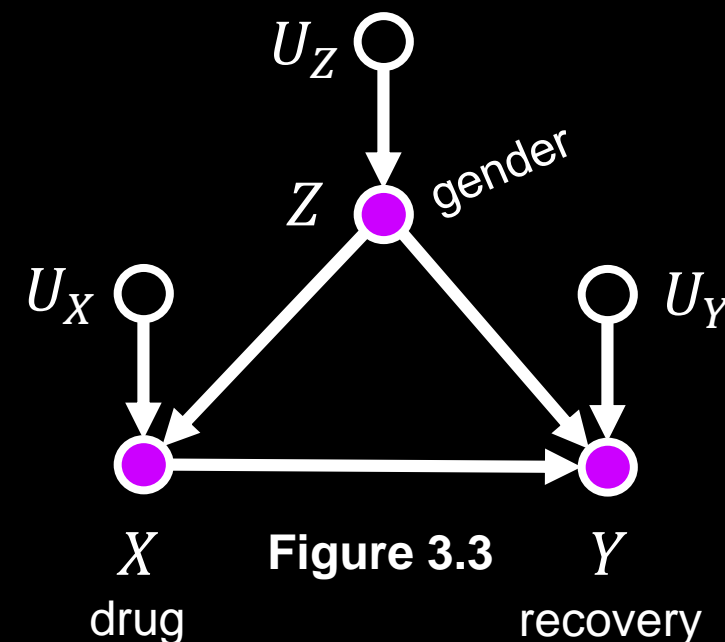
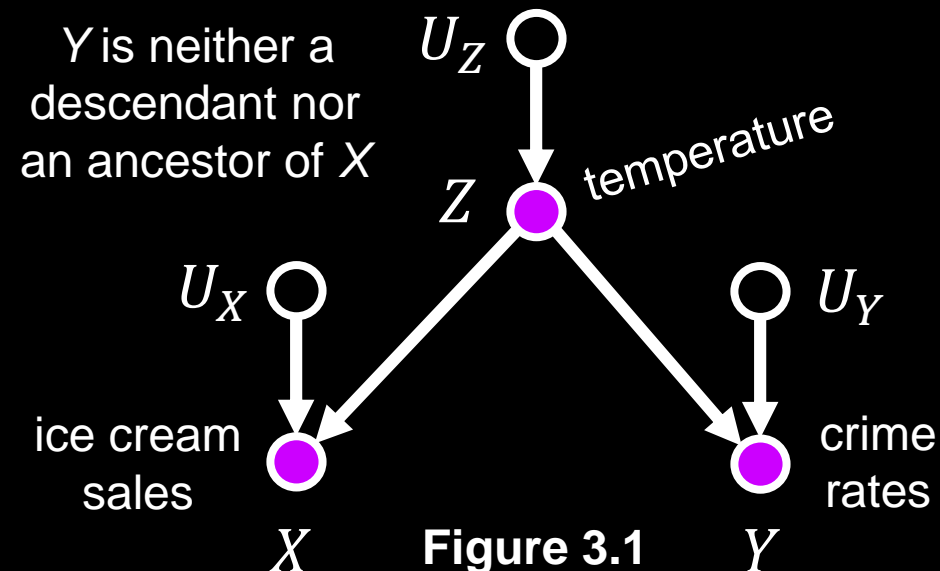


Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

3.2 THE ADJUSTMENT FORMULA

The ice cream example represents an extreme case in which the correlation between X and Y was totally spurious from a causal perspective, because there was no causal path from X to Y .

Most real-life situations are not so clear-cut. To explore a more realistic situation, let us examine **Figure 3.3**, in which Y responds to both Z and X .

Such a model could represent, for example, the first story we encountered for **Simpson's paradox**, where X stands for drug usage, Y stands for recovery, and Z stands for gender.

To find out how effective the drug is in the population, we imagine a hypothetical intervention by which we administer the drug uniformly to the entire population and compare the recovery rate to what would obtain under the complementary intervention, where we prevent everyone from using the drug.

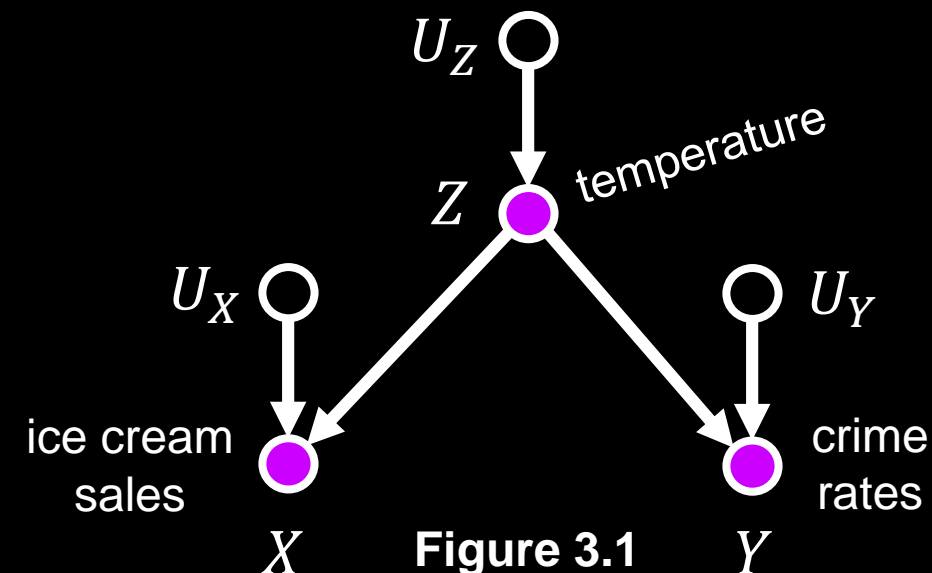


Figure 3.1

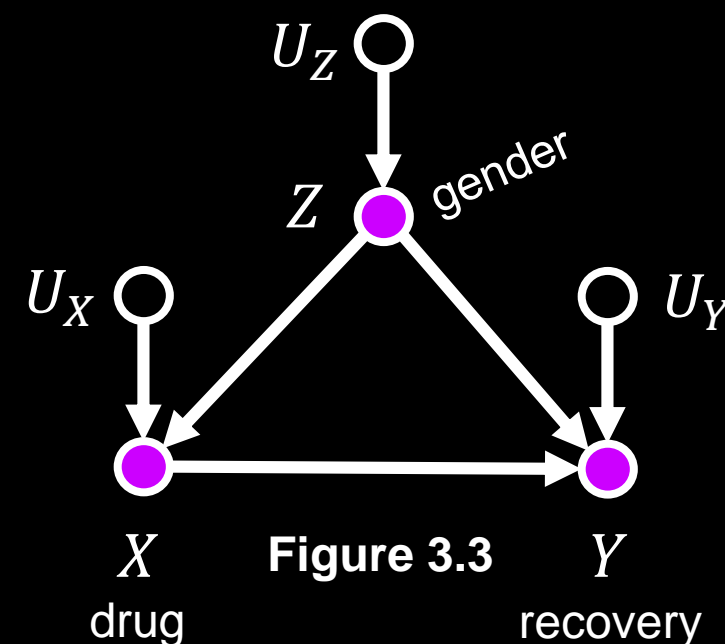


Figure 3.3

3.2 THE ADJUSTMENT FORMULA

first intervention, i.e.,
administer the drug
uniformly to the entire
population

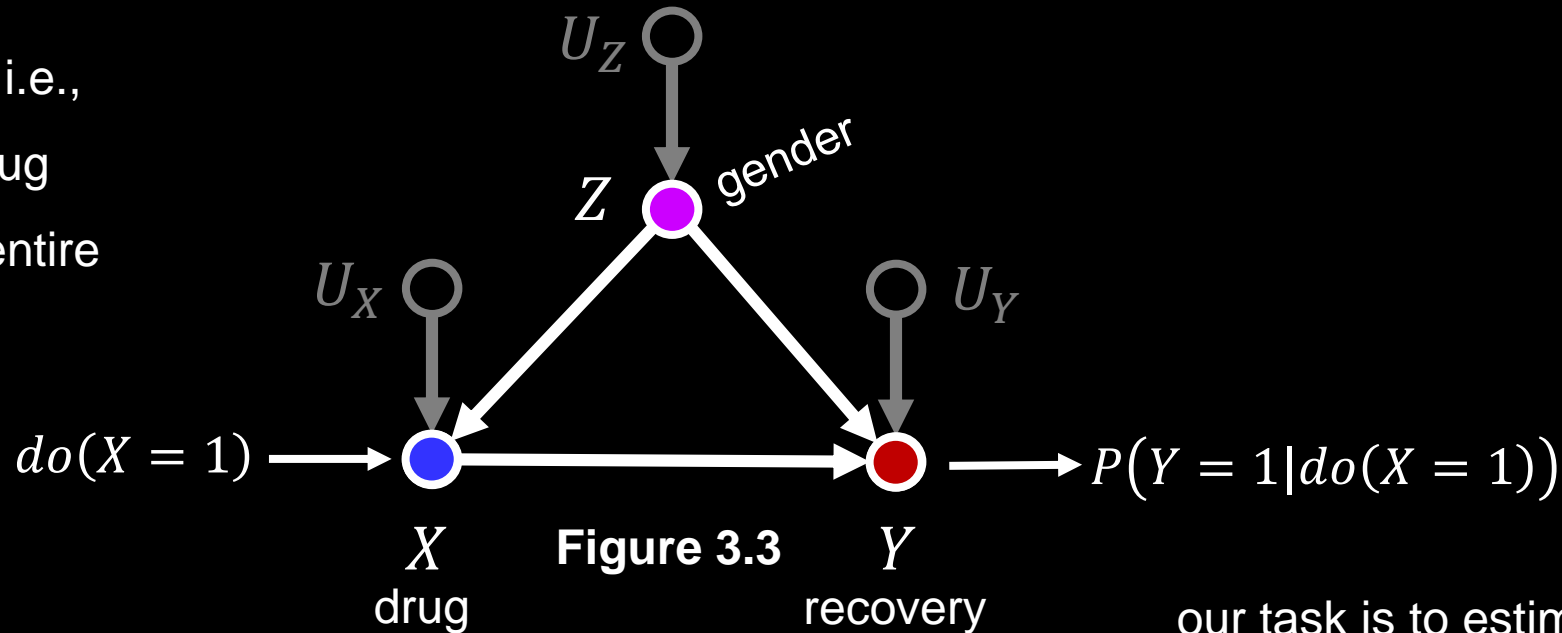


Figure 3.3

our task is to estimate the difference

$$P(Y = 1 | do(X = 1)) - P(Y = 1 | do(X = 0))$$

causal effect difference or
Average Causal Effect (ACE)

second intervention,
i.e., prevent everyone
from using the drug

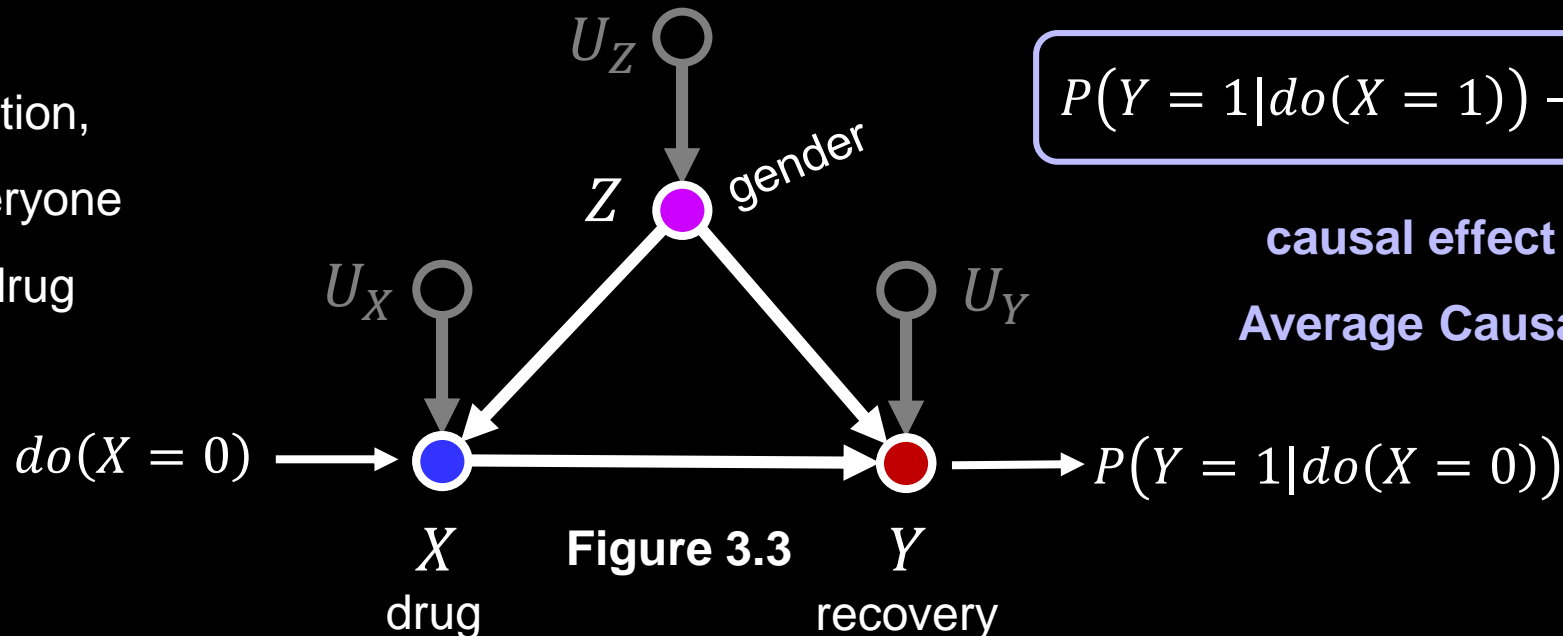


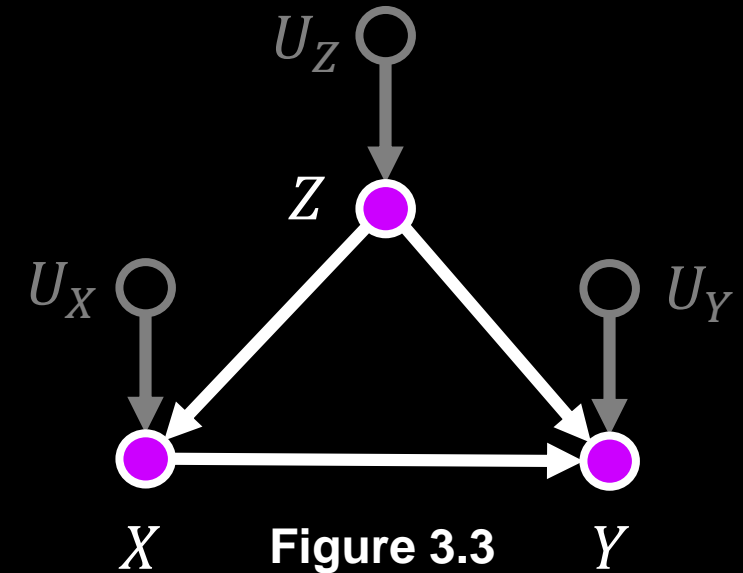
Figure 3.3

3.2 THE ADJUSTMENT FORMULA

In general, however, if X and Y can each take on more than one value, we would wish to predict the general causal effect

$$P(Y = y | do(X = x))$$

where x and y are any two values that X and Y can take on.



For example, x may be the dosage of the drug and y the patient's blood pressure.

We know from first principles that causal effects cannot be estimated from the data set itself without a causal story.

That was the lesson of **Simpson's paradox**:

The data itself was not sufficient even for determining whether the effect of the drug was positive or negative.

But with the aid of the graph in **Figure 3.3**, we can compute the magnitude of the causal effect from the data.

3.2 THE ADJUSTMENT FORMULA

To do so, we simulate the intervention in the form of a **graph surgery** on the **original model** (Figure 3.3) just as we did in the ice cream example.

The causal effect

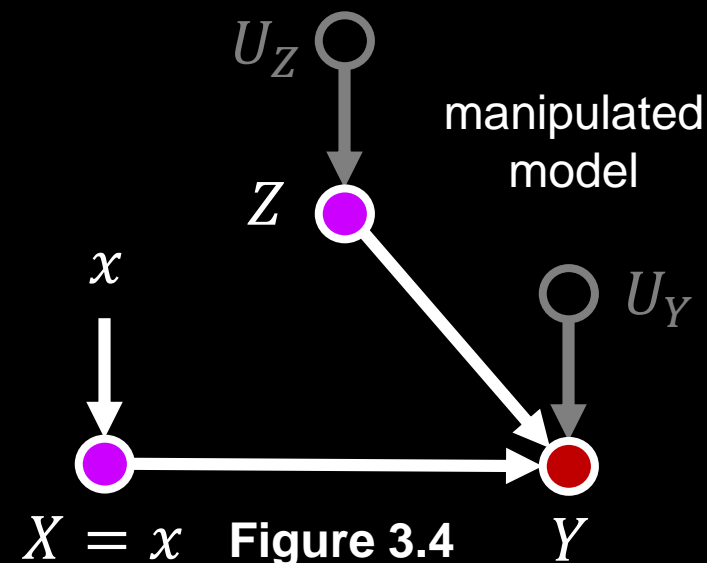
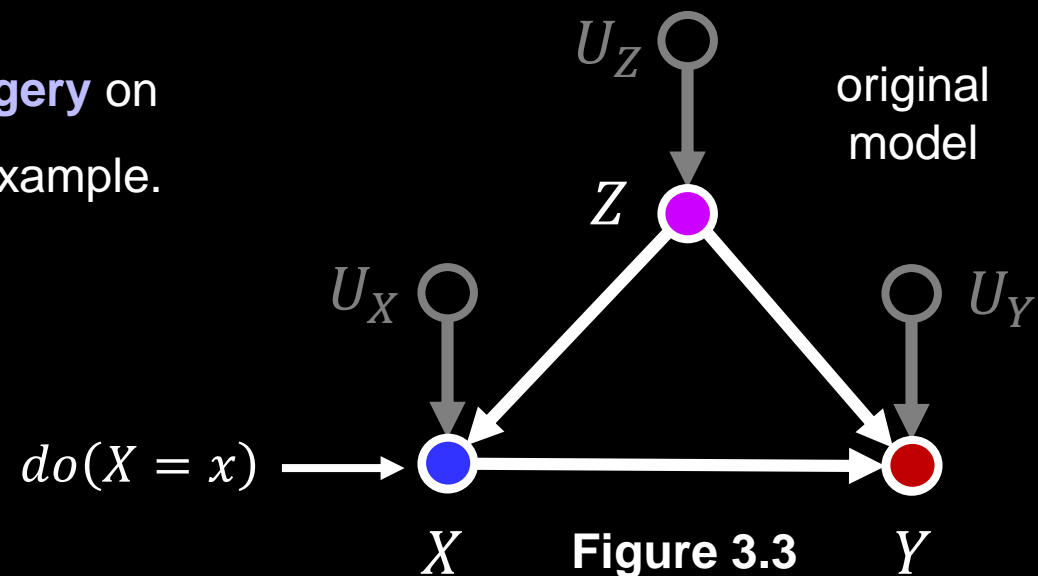
$$P(Y = y | do(X = x))$$

is equal to the conditional probability

$$P_m(Y = y | X = x)$$

that prevails in the **manipulated model** of Figure 3.4.

This, of course, also resolves the question of whether the correct answer lies in the aggregated or the Z -specific table—when we determine the answer through an intervention, there's only one table to contend with.

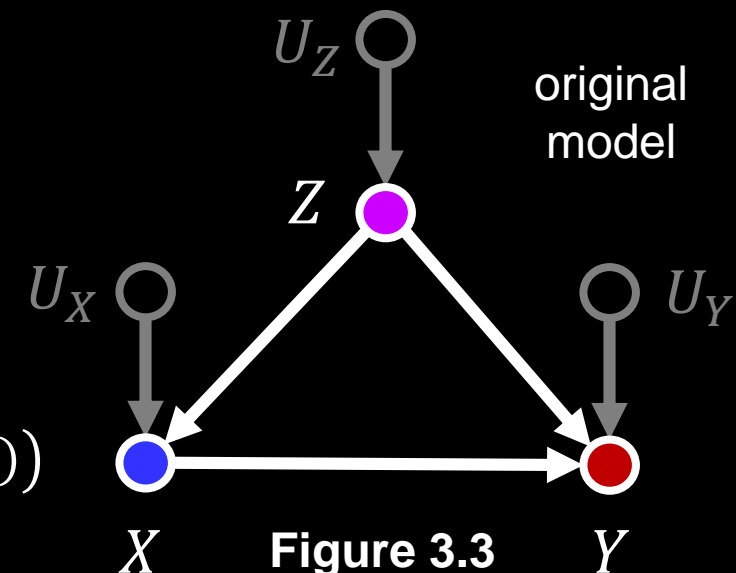


3.2 THE ADJUSTMENT FORMULA

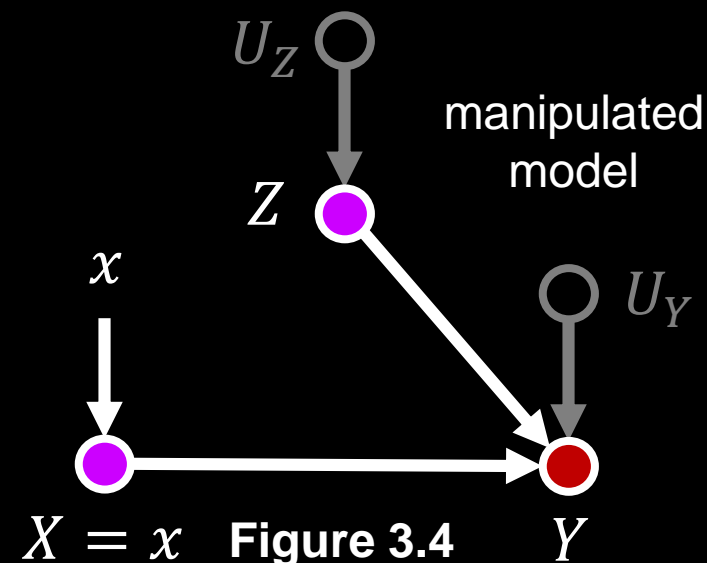
The key to computing the causal effect lies in the observation that P_m , the **manipulated probability (manipulated model)**, shares two essential properties with P (the **original probability (original model)** function that prevails in the **preintervention model** of Figure 3.3).

- the marginal probability $P(Z = z)$ is invariant under the intervention, because the process determining Z is not affected by removing the arrow from Z to X . (proportions of males and females remain the same, before and after the intervention)
- the conditional probability $P(Y = y|X = x, Z = z)$ is invariant, because the process by which Y responds to X and Z , $Y = f(x, z, u_Y)$, remains the same, regardless of whether X changes spontaneously or by deliberate manipulation.

$$P(Y = y|do(X = x))$$



$$P_m(Y = y|X = x)$$

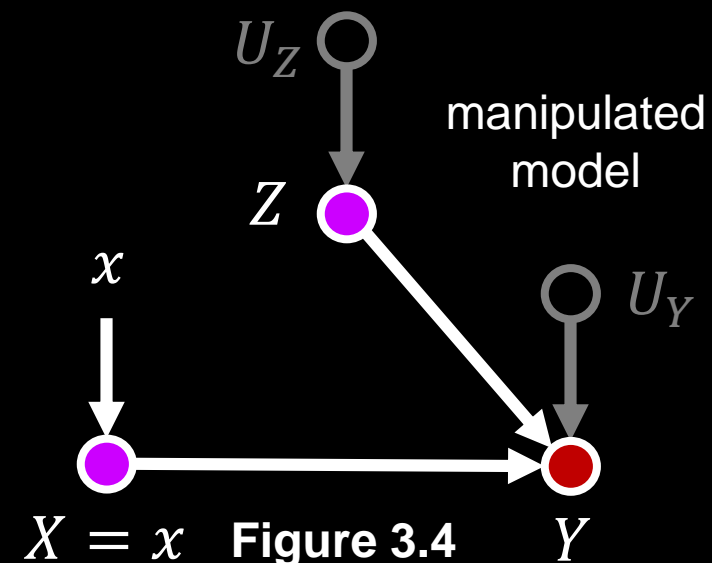
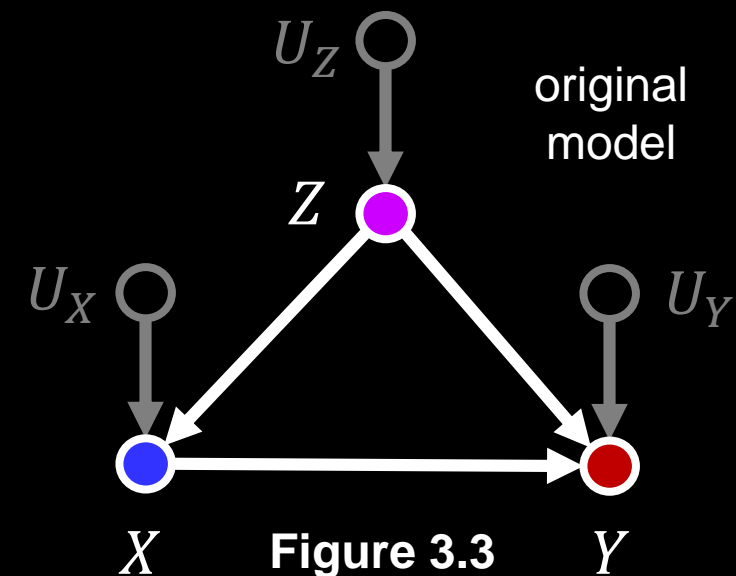


3.2 THE ADJUSTMENT FORMULA

We can therefore write two equations of invariance:

$$P_m(Z = z) = P(Z = z)$$

$$P_m(Y = y|X = x, Z = z) = P(Y = y|X = x, Z = z)$$



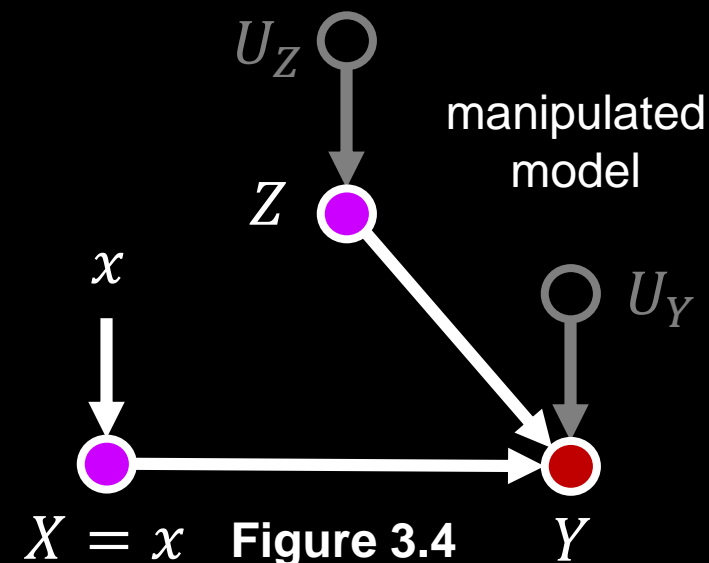
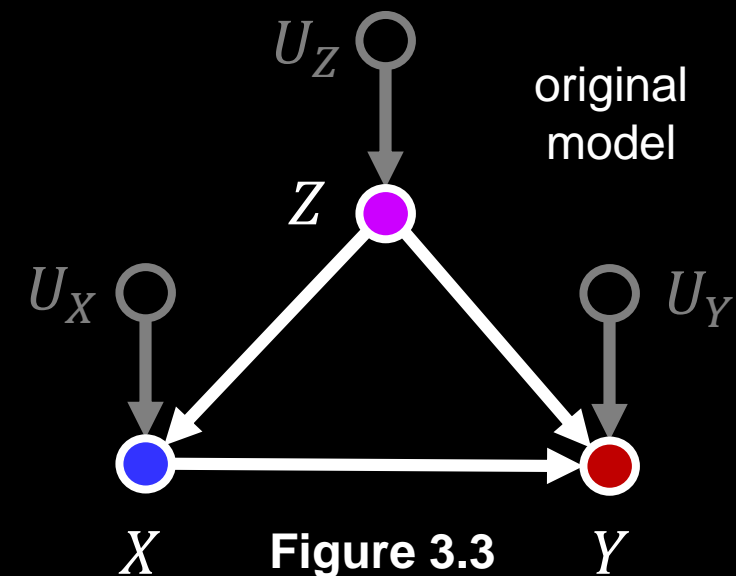
3.2 THE ADJUSTMENT FORMULA

We can therefore write two equations of invariance:

$$P_m(Z = z) = P(Z = z)$$

$$P_m(Y = y|X = x, Z = z) = P(Y = y|X = x, Z = z)$$

We can also use the fact that Z and X are d-separated (collider) in the manipulated model (**Figure 3.4**) and are, therefore, independent under the intervention distribution.



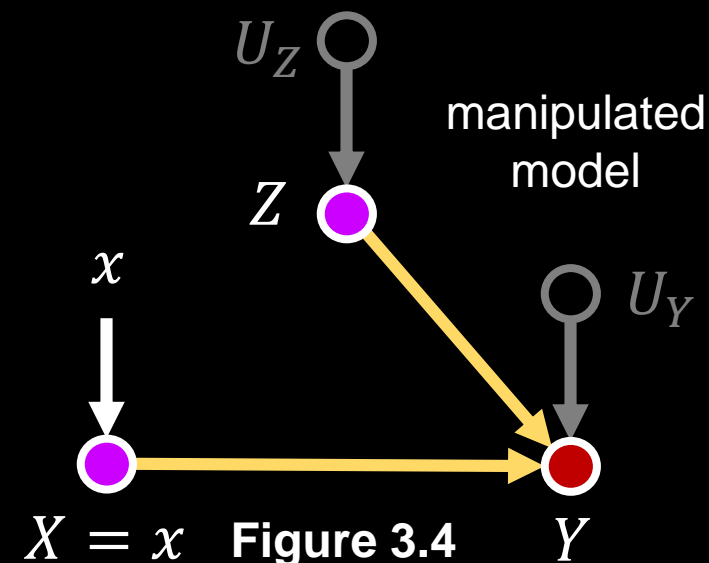
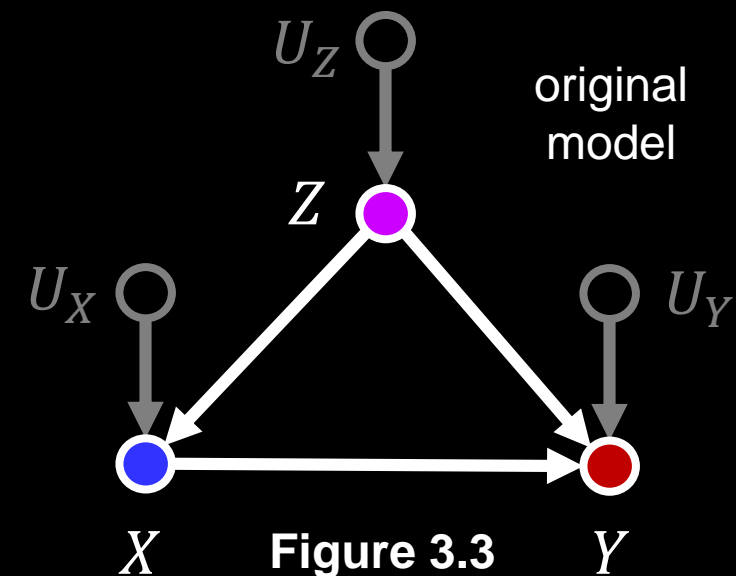
3.2 THE ADJUSTMENT FORMULA

We can therefore write two equations of invariance:

$$P_m(Z = z) = P(Z = z)$$

$$P_m(Y = y|X = x, Z = z) = P(Y = y|X = x, Z = z)$$

We can also use the fact that Z and X are d-separated (**collider**) in the manipulated model (**Figure 3.4**) and are, therefore, independent under the intervention distribution.



3.2 THE ADJUSTMENT FORMULA

We can therefore write two equations of invariance:

$$P_m(Z = z) = P(Z = z)$$

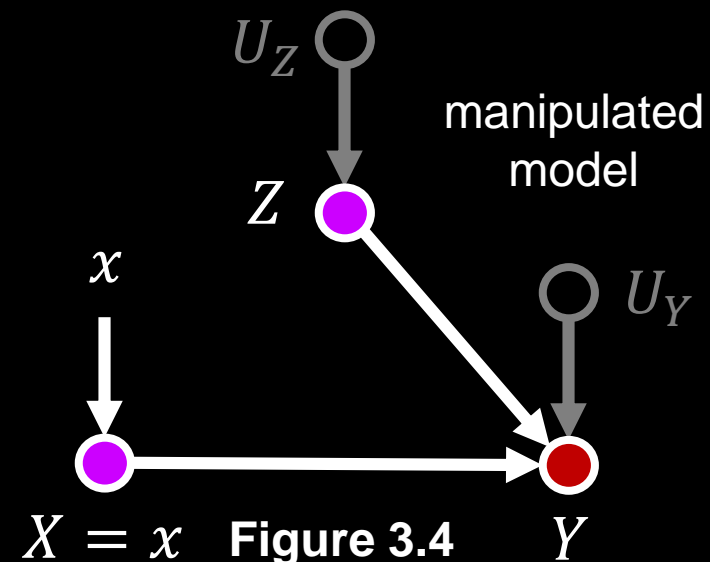
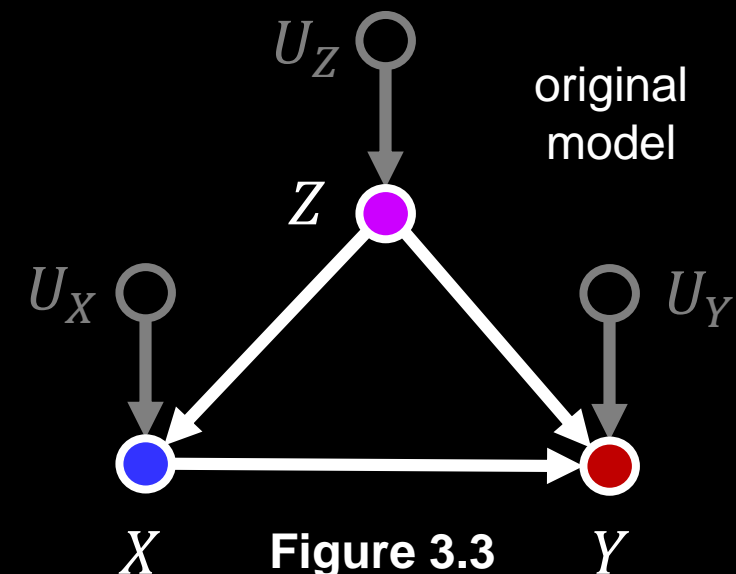
$$P_m(Y = y|X = x, Z = z) = P(Y = y|X = x, Z = z)$$

We can also use the fact that Z and X are d-separated (collider) in the manipulated model (**Figure 3.4**) and are, therefore, independent under the intervention distribution.

This tells us that $P_m(Z = z|X = x) = P_m(Z = z) = P(Z = z)$

Putting these considerations together, we have

$$\begin{aligned} P(Y = y|do(X = x)) &= P_m(Y = y|X = x) && \text{(by definition)} \\ &= \sum_z P_m(Y = y|X = x, Z = z) P_m(Z = z|X = x) \\ &= \sum_z P_m(Y = y|X = x, Z = z) P_m(Z = z) \end{aligned}$$



3.2 THE ADJUSTMENT FORMULA

To recap, we have

$$P(Y = y | do(X = x)) = \sum_z P_m(Y = y | X = x, Z = z) P_m(Z = z)$$

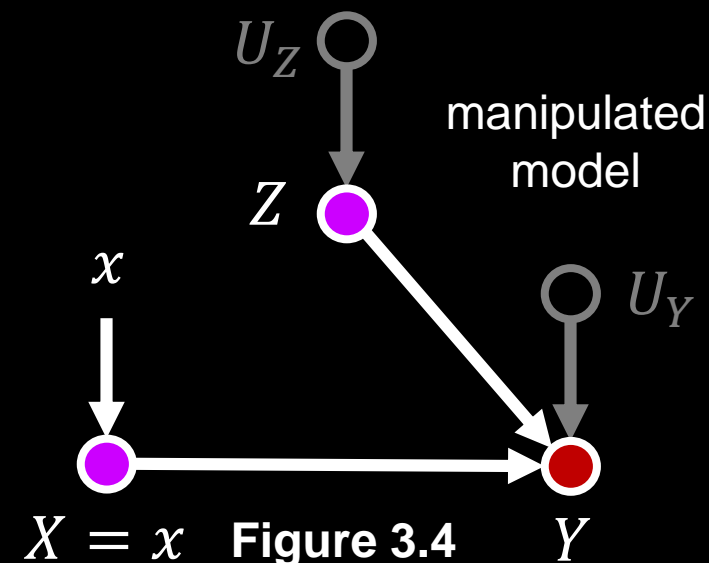
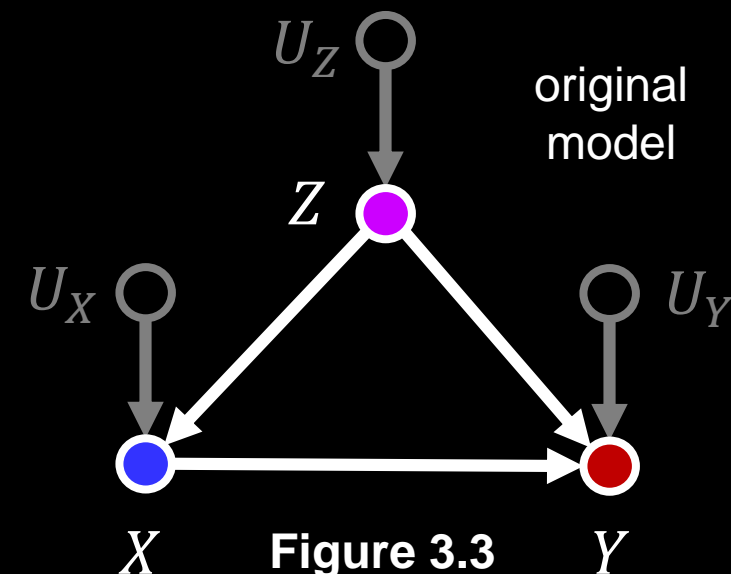
Finally, using the invariance relations,

$$P_m(Z = z) = P(Z = z)$$

$$P_m(Y = y | X = x, Z = z) = P(Y = y | X = x, Z = z)$$

we obtain a formula for the causal effect, in terms of preintervention probabilities (original model):

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



3.2 THE ADJUSTMENT FORMULA

The **Adjustment Formula** computes the association between X and Y for each value z of Z , then averages over those values.

This procedure is referred to as “adjusting for Z ” or “controlling for Z .”

It can be estimated directly from the data, since it consists only of conditional probabilities, each of which can be computed by the filtering procedure described in **Part 1**.

No adjustment is needed in a randomized controlled experiment since, in such a setting, the data are generated by a model which already possesses the structure of **Figure 3.4**, hence, $P_m = P$ regardless of any factors Z that affect Y .

The Adjustment Formula

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

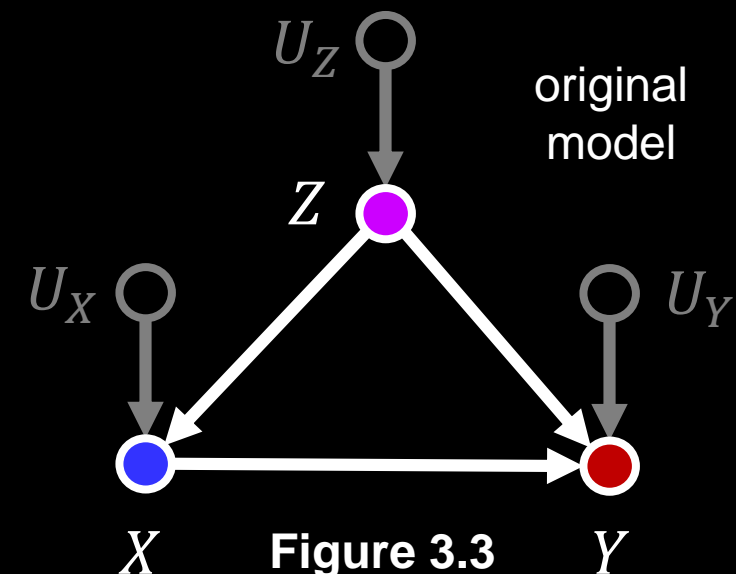


Figure 3.3

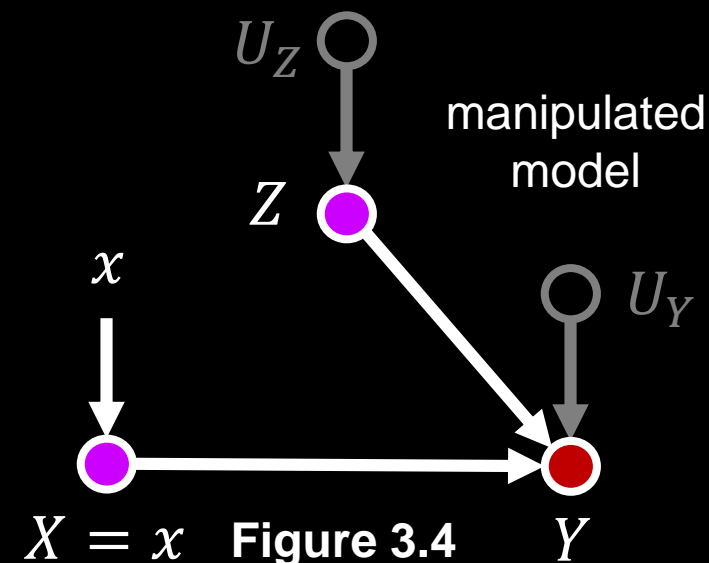


Figure 3.4

3.2 THE ADJUSTMENT FORMULA

This derivation of the adjustment formula constitutes therefore a formal proof that randomization gives us the quantity we seek to estimate, namely

$$P(Y = y | do(X = x))$$

In practice, investigators use adjustments in randomized experiments as well, for the purpose of minimizing sampling variations (Cox 1958).

The Adjustment Formula

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

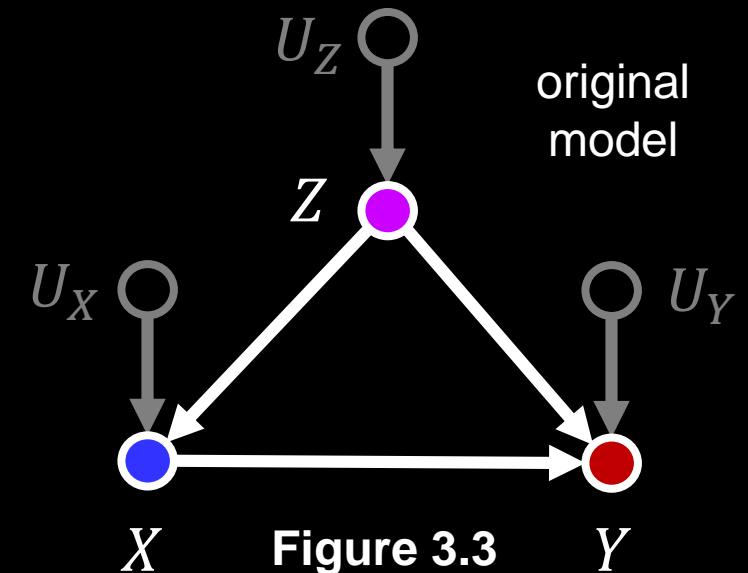


Figure 3.3

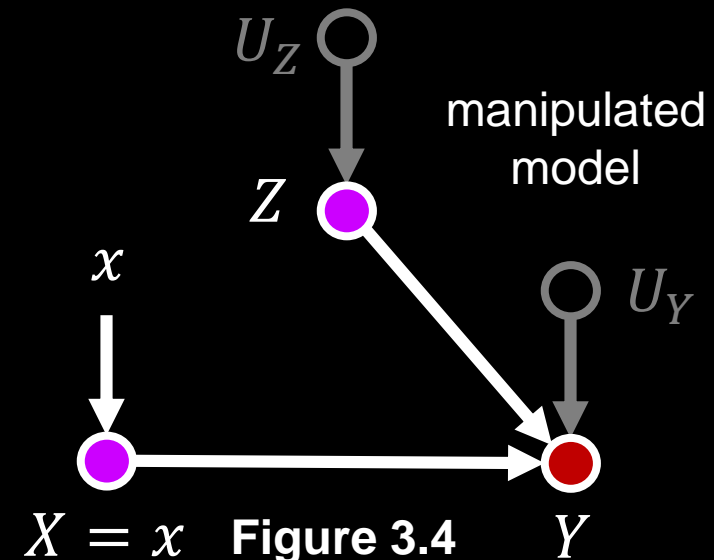


Figure 3.4

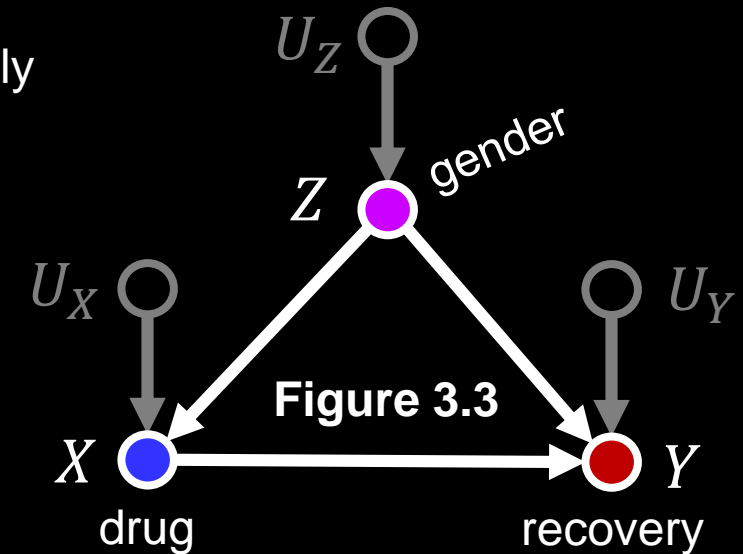
3.2 THE ADJUSTMENT FORMULA

To demonstrate the working of the adjustment formula, let us apply it numerically to **Simpson's story**, with

- $X = 1$ standing for the patient taking the drug,
- $Z = 1$ standing for the patient being male, and
- $Y = 1$ standing for the patient recovering.

we have

$$P(Y = 1 | do(X = 1)) = P(Y = 1 | X = 1, Z = 1) P(Z = 1) + P(Y = 1 | X = 1, Z = 0) P(Z = 0)$$



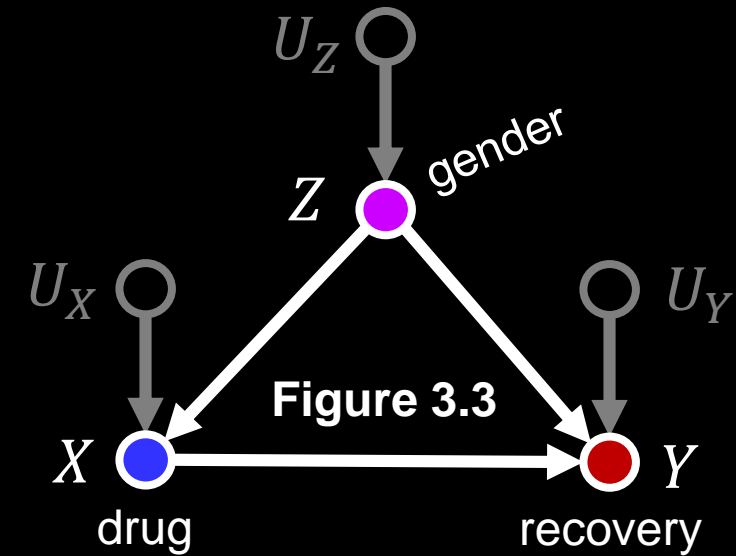
The Adjustment Formula

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

3.2 THE ADJUSTMENT FORMULA

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%



$$P(Y = 1|do(X = 1)) = P(Y = 1|X = 1, Z = 1) P(Z = 1) + P(Y = 1|X = 1, Z = 0) P(Z = 0)$$

$$P(Y = 1|do(X = 1)) = 0.93 \times \frac{(87 + 270)}{700} + 0.73 \times \frac{(263 + 80)}{700} = 0.832$$

$$P(Y = 1|do(X = 0)) = 0.87 \times \frac{(87 + 270)}{700} + 0.69 \times \frac{(263 + 80)}{700} = 0.7818$$

Thus, comparing the effect of **drug-taking** ($X = 1$) to the effect of **nontaking** ($X = 0$), we obtain

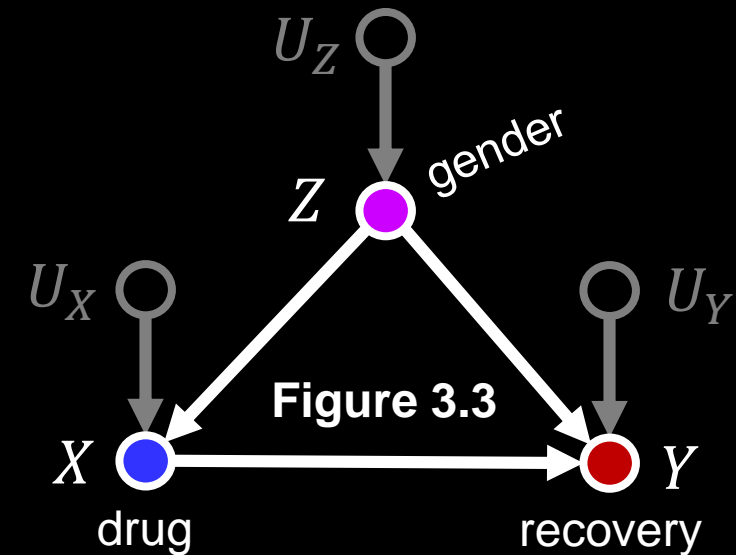
$$ACE = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0)) = 0.832 - 0.7818 = 0.0502$$

giving a **clear positive advantage to drug-taking**.

3.2 THE ADJUSTMENT FORMULA

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%



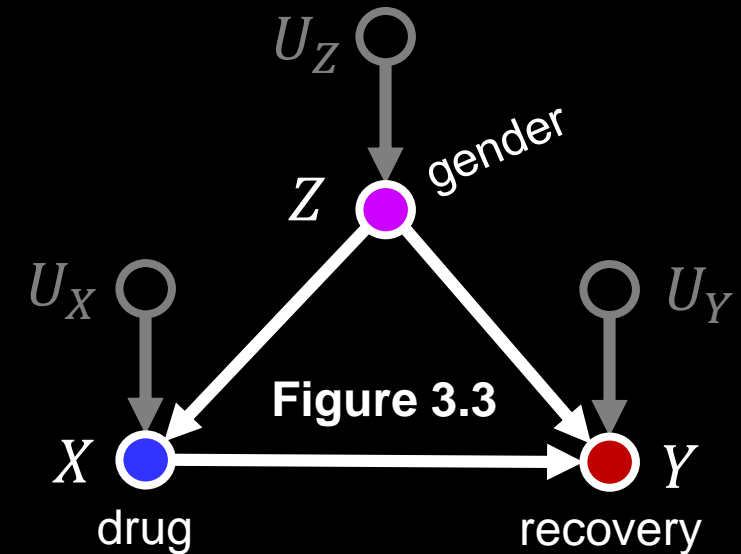
A more informal interpretation of ACE here is that it is simply the difference in the fraction of the population that would recover if everyone took the drug compared to when no one takes the drug.

$$ACE = P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0)) = 0.832 - 0.7818 = 0.0502$$

3.2 THE ADJUSTMENT FORMULA

We see that the adjustment formula instructs us to

- condition on gender,
- find the benefit of the drug separately for males and females,
- average the result using the percentage of males and females in the population.



$$P(Y = 1 | do(X = 1)) = \underbrace{P(Y = 1 | X = 1, Z = 1)}_{\substack{\text{probability to recover} \\ \text{for male drug-takers}}} \underbrace{P(Z = 1)}_{\substack{\text{average using} \\ \text{percentage males}}} + \underbrace{P(Y = 1 | X = 1, Z = 0)}_{\substack{\text{probability to recover} \\ \text{for female drug-takers}}} \underbrace{P(Z = 0)}_{\substack{\text{average using} \\ \text{percentage females}}}$$

↑
↑

condition on gender (male)
condition on gender (female)

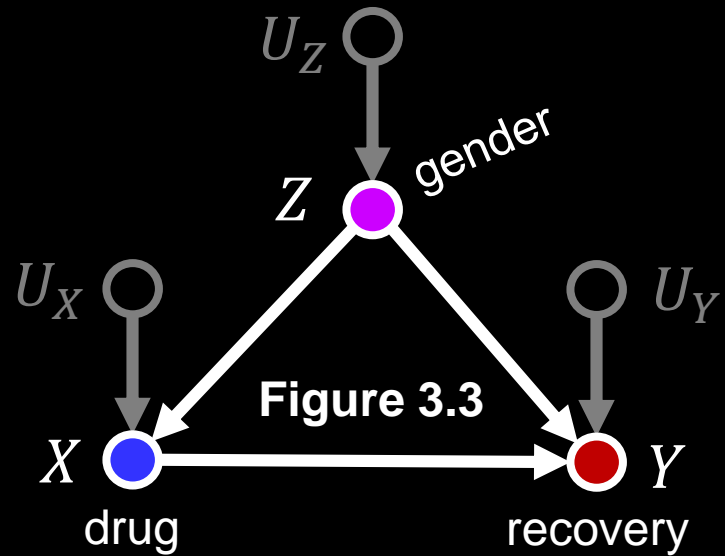
3.2 THE ADJUSTMENT FORMULA

It also thus instructs us to ignore the aggregated population data

$$P(Y = 1|X = 1)$$

$$P(Y = 1|X = 0)$$

from which we might (falsely) conclude that the drug has a negative effect overall.



$$ACE = P(Y = 1|X = 1) - P(Y = 1|X = 0) = 0.78 - 0.83 = -0.05$$

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug			No Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Men	87	81	93%	270	234	87%
Women	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

3.2 THE ADJUSTMENT FORMULA

These simple examples might give us the impression that whenever we face the dilemma of whether to condition on a third variable Z , the adjustment formula prefers the Z -specific analysis over the nonspecific analysis.

But we know this is not so, recalling the blood pressure example of Simpson's paradox given in **Table 1.2**.

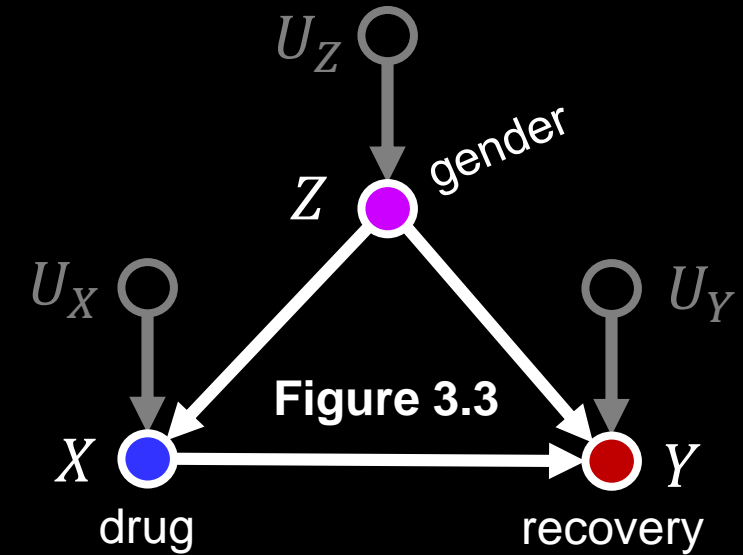


Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No Drug			Drug		
	patients	recovered	% recovered	patients	recovered	% recovered
Low BP	87	81	93%	270	234	87%
High BP	263	192	73%	80	55	69%
Combined data	350	273	78%	350	289	83%

There we argued that the more sensible method would be not to condition on blood pressure, but to examine the unconditional population table directly.

How would the adjustment formula cope with situations like that?

3.2 THE ADJUSTMENT FORMULA

The graph in **Figure 3.5** represents the causal story in the blood pressure example,

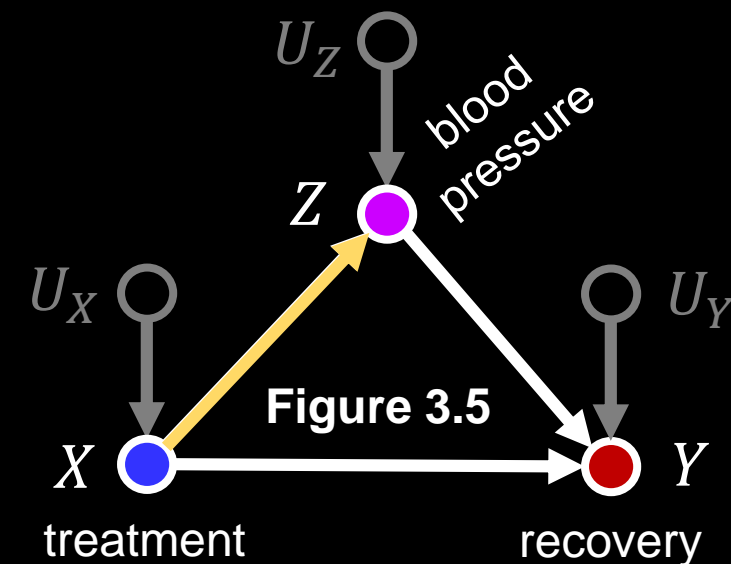
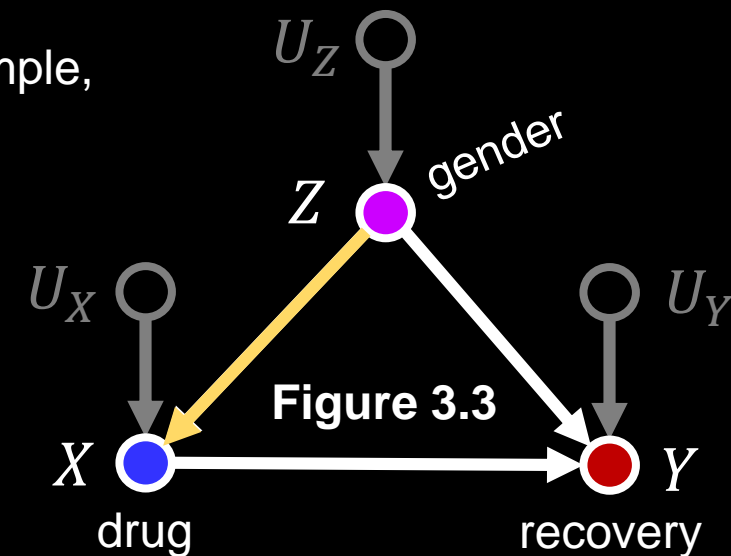
- X treatment,
- Z blood pressure,
- Y recovery.

It is the same as **Figure 3.3**, but with the arrow between X and Z reversed, reflecting the fact that the treatment has an effect on blood pressure and not the other way around.

Let us try now to evaluate the causal effect

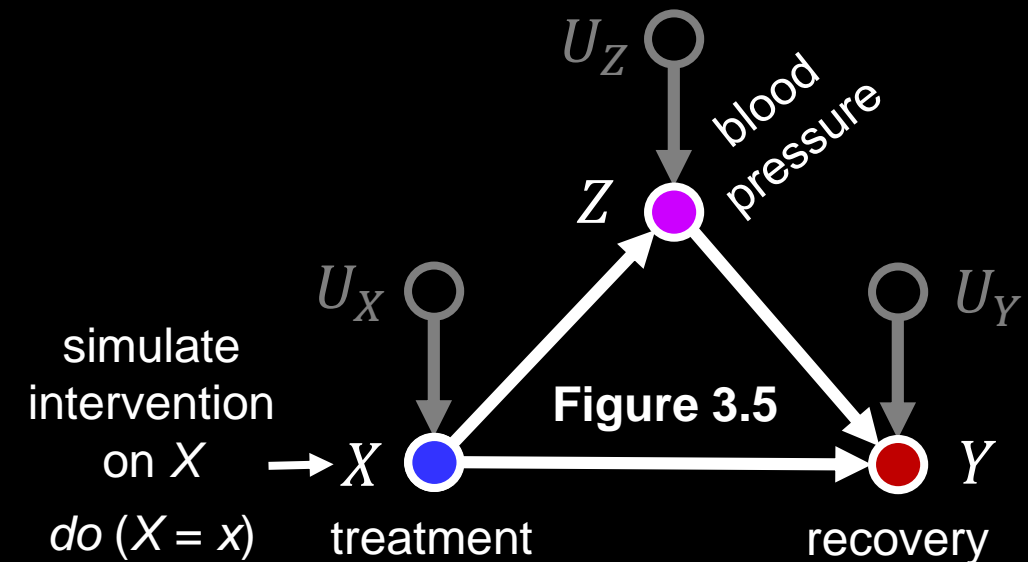
$$P(Y = 1 | do(X = 1))$$

associated with this model as we did with the gender example.



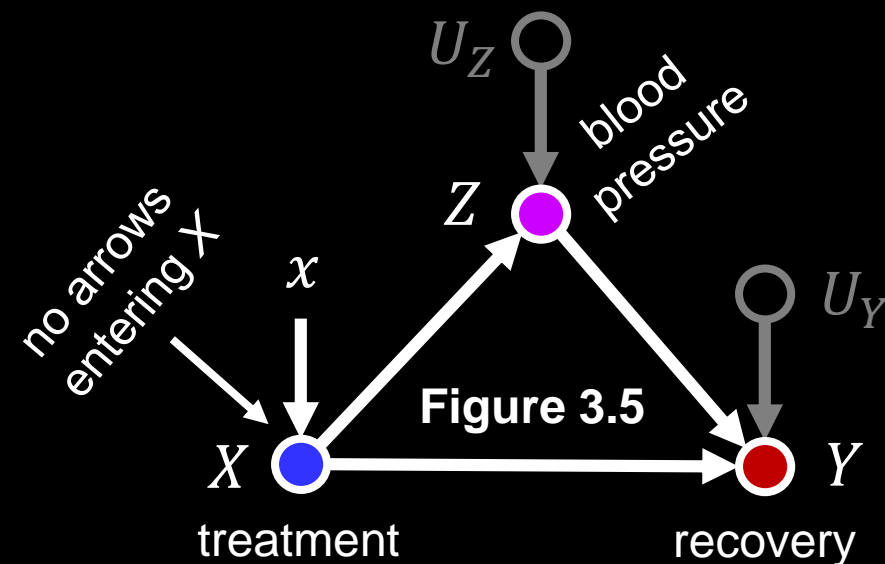
3.2 THE ADJUSTMENT FORMULA

- We simulate an intervention and then examine the adjustment formula that emanates from the simulated intervention.
- In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X .



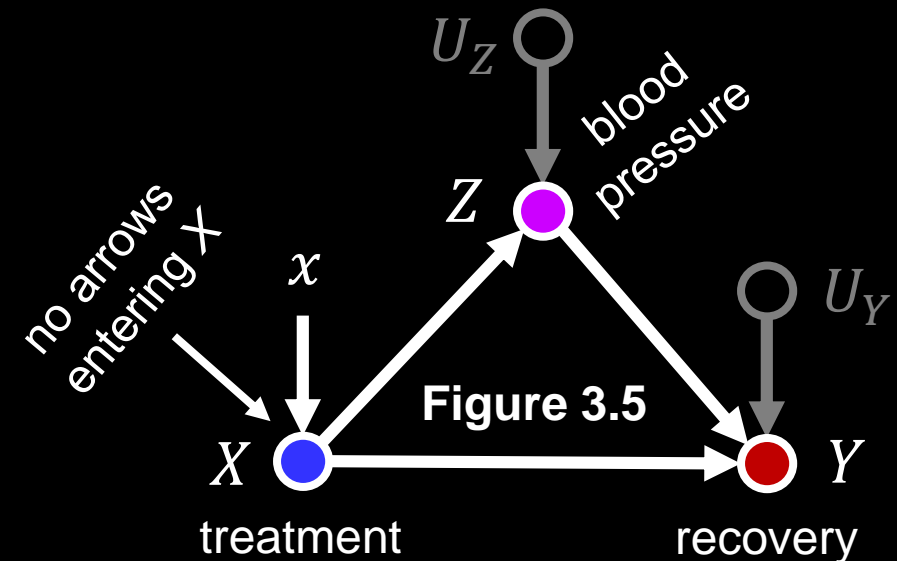
3.2 THE ADJUSTMENT FORMULA

- We simulate an intervention and then examine the adjustment formula that emanates from the simulated intervention.
- In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X .
- In our case, however, the graph of **Figure 3.5** shows no arrow entering X , since X has no parents.



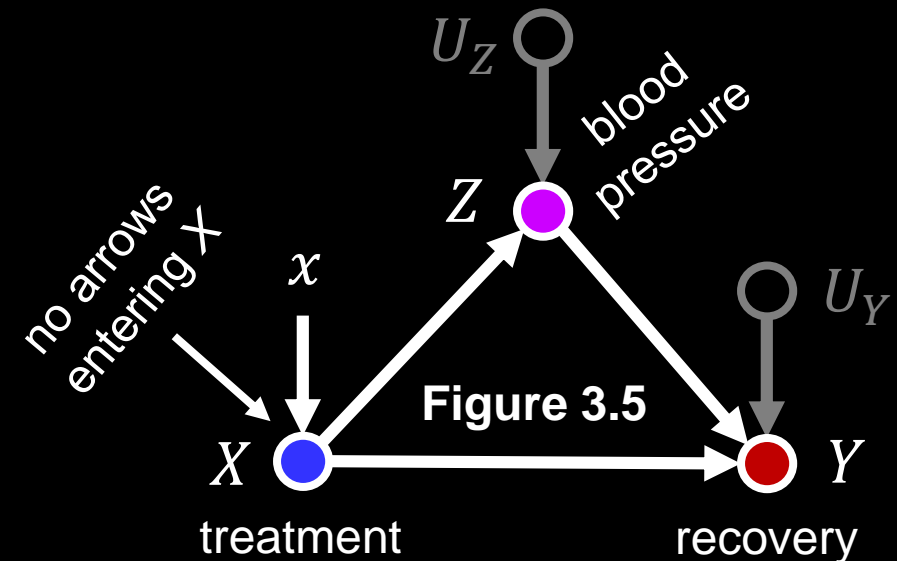
3.2 THE ADJUSTMENT FORMULA

- We simulate an intervention and then examine the adjustment formula that emanates from the simulated intervention.
- In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X .
- In our case, however, the graph of **Figure 3.5** shows no arrow entering X , since X has no parents.
- This means that no surgery is required; the conditions under which data were obtained were such that treatment was assigned “as if randomized.”



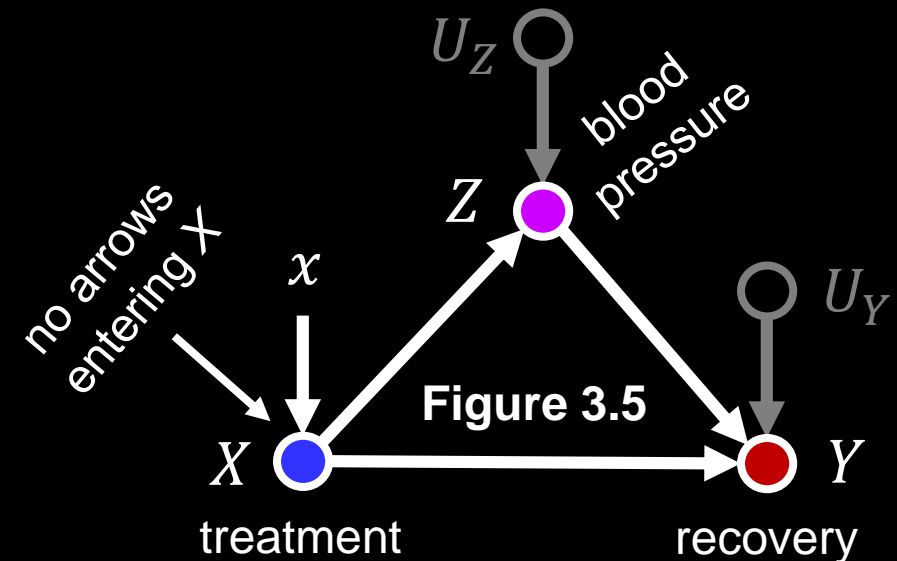
3.2 THE ADJUSTMENT FORMULA

- We simulate an intervention and then examine the adjustment formula that emanates from the simulated intervention.
- In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X .
- In our case, however, the graph of **Figure 3.5** shows no arrow entering X , since X has no parents.
- This means that no surgery is required; the conditions under which data were obtained were such that treatment was assigned “as if randomized.”
- If there was a factor that would make subjects prefer or reject treatment, such a factor should show up in the model;



3.2 THE ADJUSTMENT FORMULA

- We simulate an intervention and then examine the adjustment formula that emanates from the simulated intervention.
- In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X .
- In our case, however, the graph of **Figure 3.5** shows no arrow entering X , since X has no parents.
- This means that no surgery is required; the conditions under which data were obtained were such that treatment was assigned “as if randomized.”
- If there was a factor that would make subjects prefer or reject treatment, such a factor should show up in the model;
- the absence of such a factor gives us the license to treat X as a randomized treatment.



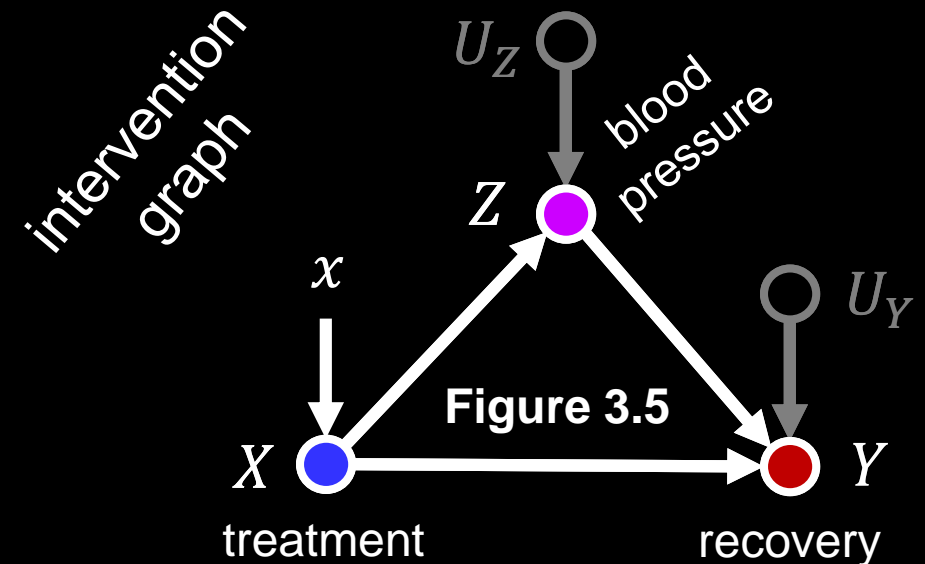
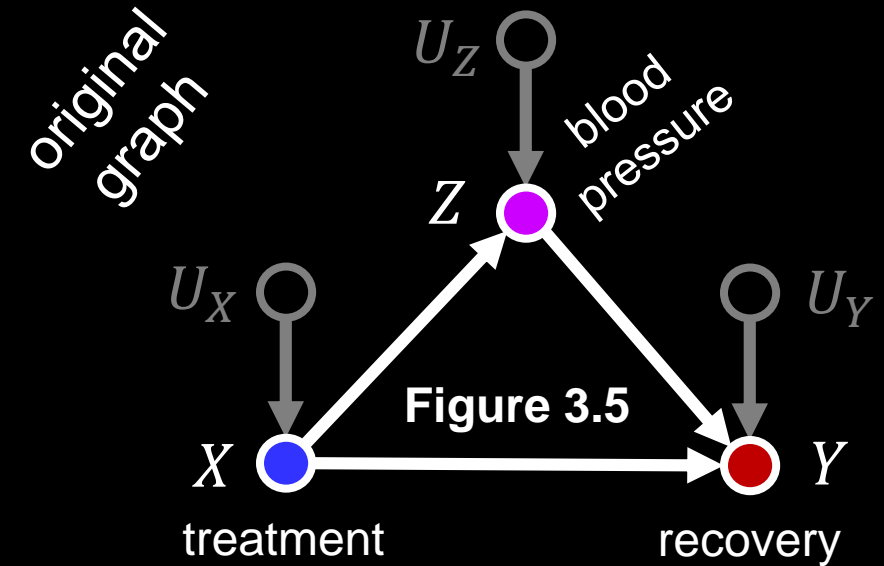
3.2 THE ADJUSTMENT FORMULA

Under such conditions, the **intervention graph** is equal to the **original graph**—no arrow need be removed—and the **adjustment formula** reduces to

$$P(Y = y|do(X = x)) = P(Y = y|X = x)$$

which can be obtained from our adjustment formula by letting the empty set be the element adjusted for.

Obviously, if we were to adjust for blood pressure Z , we would obtain an incorrect assessment—one corresponding to a model in which blood pressure Z causes people to seek treatment X .



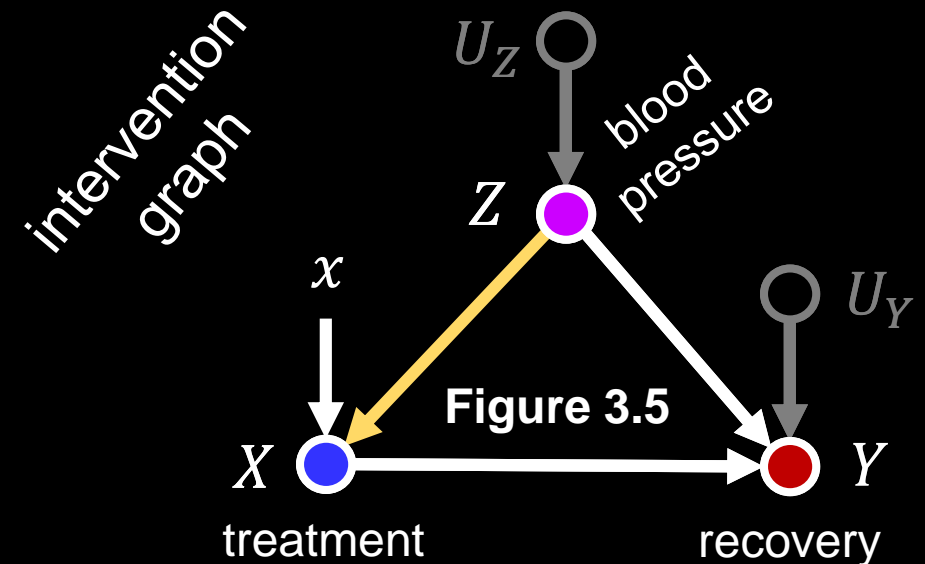
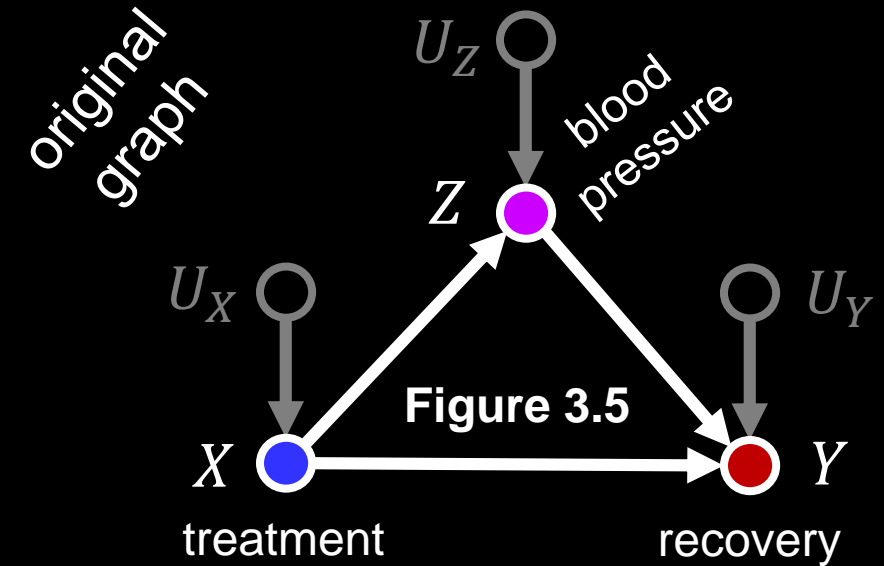
3.2 THE ADJUSTMENT FORMULA

Under such conditions, the **intervention graph** is equal to the **original graph**—no arrow need be removed—and the **adjustment formula** reduces to

$$P(Y = y|do(X = x)) = P(Y = y|X = x)$$

which can be obtained from our adjustment formula by letting the empty set be the element adjusted for.

Obviously, if we were to adjust for blood pressure Z , we would obtain an incorrect assessment—one corresponding to a model in which **blood pressure Z causes people to seek treatment X** .



3.2.1 THE ADJUSTMENT FORMULA: TO ADJUST OR NOT TO ADJUST?

We are now in a position to understand what variable, or set of variables, Z can legitimately be included in the adjustment formula.

The intervention procedure, which led to the adjustment formula, dictates that Z should coincide with the parents $pa(X)$ of X , because it is the influence of these parents that we neutralize when we fix X by external manipulation $do(X)$.

We can therefore write a general adjustment formula and summarize it in a rule:

Rule 1 (The Causal Effect Rule)

Given a graph G in which a set of variables $pa(X)$ are designated as the parents of X , the causal effect of X on Y is given by

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

where z ranges over all the combinations of values that the variables in $pa(X)$ can take.

3.2.1 THE ADJUSTMENT FORMULA: TO ADJUST OR NOT TO ADJUST?

If we multiply and divide the summand

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

by

$$P(X = x|pa(X) = z)$$

we get a more convenient form:

$$\begin{aligned} P(Y = y|do(X = x)) &= \sum_z P(Y = y|X = x, pa(X) = z) P(pa(X) = z) \frac{P(X = x|pa(X) = z)}{P(X = x|pa(X) = z)} \\ &= \sum_z \frac{P(Y = y|X = x, pa(X) = z) P(X = x|pa(X) = z) P(pa(X) = z)}{P(X = x|pa(X) = z)} \\ &= \sum_z \frac{P(Y = y, X = x, pa(X) = z)}{P(X = x|pa(X) = z)} \end{aligned}$$

3.2.1 THE ADJUSTMENT FORMULA: TO ADJUST OR NOT TO ADJUST?

The formula

$$P(Y = y | do(X = x)) = \sum_z \frac{P(Y = y, X = x, pa(X) = z)}{\boxed{P(X = x | pa(X) = z)}} \longleftarrow \begin{array}{l} \text{Propensity} \\ \text{Score} \end{array}$$

explicitly displays the role played by the parents of X in predicting the results Y of interventions ($do(X)$).

We can appreciate now what role the causal graph plays in resolving Simpson's paradox, and, more generally, what aspects of the graph allow us to predict causal effects from purely statistical data.

We need the graph in order to determine the identity of X 's parents—the set of factors that, under nonexperimental conditions, would be sufficient for determining the value of X , or the probability of that value.

3.2.1 THE ADJUSTMENT FORMULA: TO ADJUST OR NOT TO ADJUST?

This result alone is astounding; using graphs and their underlying assumptions, we were able to identify causal relationships in purely observational data.

But, from this discussion, readers may be tempted to conclude that the role of graphs is fairly limited;

- once we identify the parents of X , the rest of the graph can be discarded, and the causal effect can be evaluated mechanically from the adjustment formula.

In the next slides we show that things may not be so simple.

In most practical cases, the set of X 's parents will contain unobserved variables that would prevent us from calculating the conditional probabilities in the adjustment formula.

Luckily, as we will see in future slides, we can adjust for other variables in the model to substitute for the unmeasured elements of $pa(X)$.

3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

In deriving the adjustment formula, we assumed

- an intervention on a single variable X ,
- whose parents were disconnected,

so as to simulate the absence of their influence after intervention.

However, social and medical policies occasionally involve multiple interventions, such as those that dictate the value of several variables simultaneously, or those that control a variable over time.

To represent multiple interventions, it is convenient to resort to the product decomposition that a graphical model imposes on joint distributions (**Rule of Product Decomposition**)

$$P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | pa(X_i))$$

3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

According to the **Rule of Product Decomposition**, the **preintervention distribution** in the model of **Figure 3.3** is given by the product

$$P(x, y, z) = P(z) P(x|z) P(y|x, z)$$

whereas the **postintervention distribution**, governed by the model of **Figure 3.4** is given by the product

$$P(z, y|do(x)) = P_m(z) P_m(y|x, z) = P(z) P(y|x, z)$$

with the factor $P(x|z)$ purged from the product, since X becomes parentless as it is fixed at $X = x$.

This coincides with the adjustment formula, because to evaluate $P(y|do(x))$ we need to marginalize (or sum) over z , which gives

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

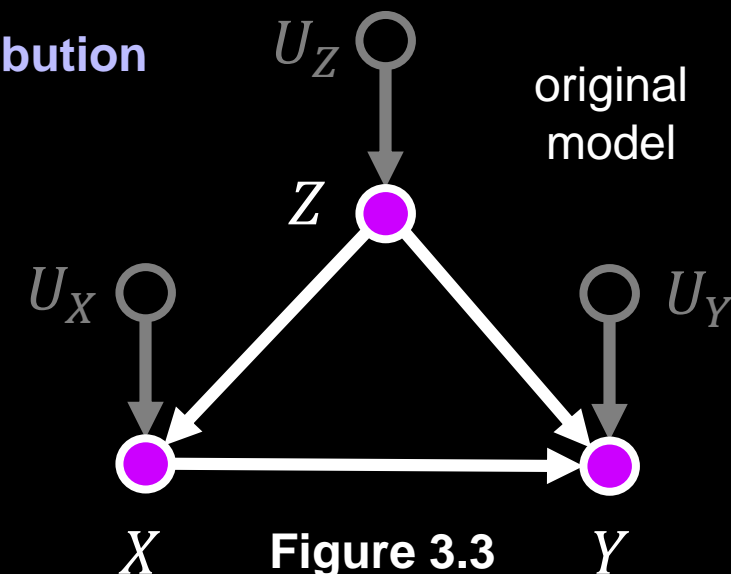


Figure 3.3

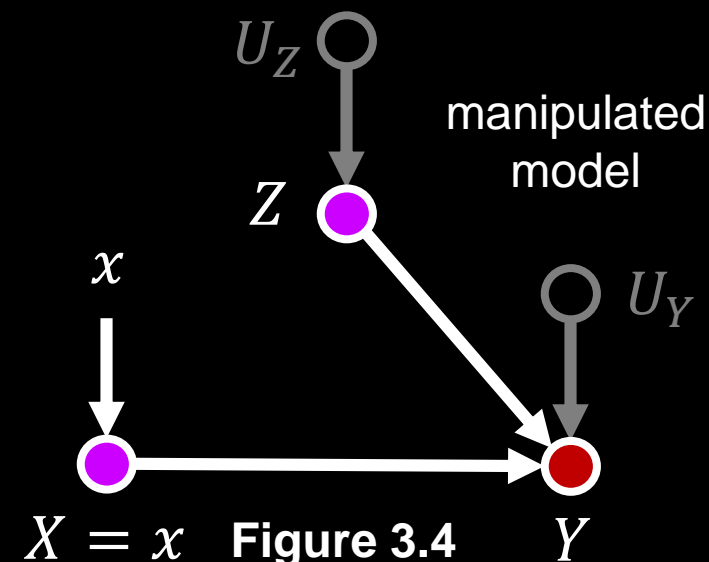


Figure 3.4

3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

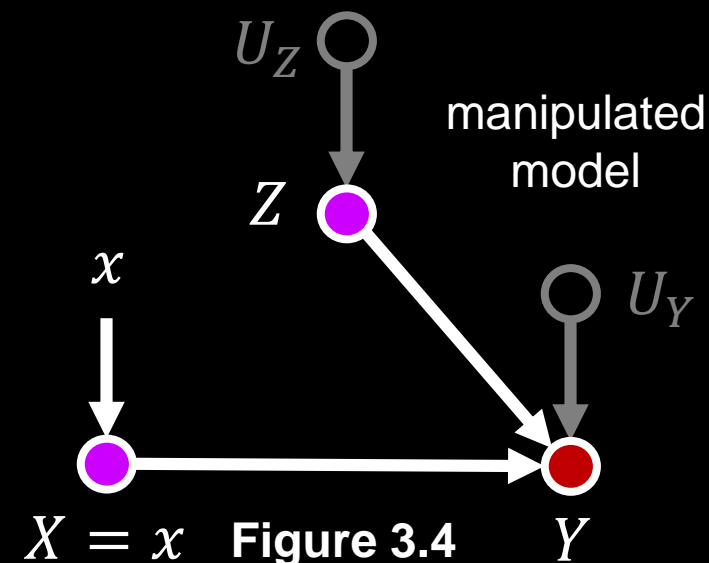
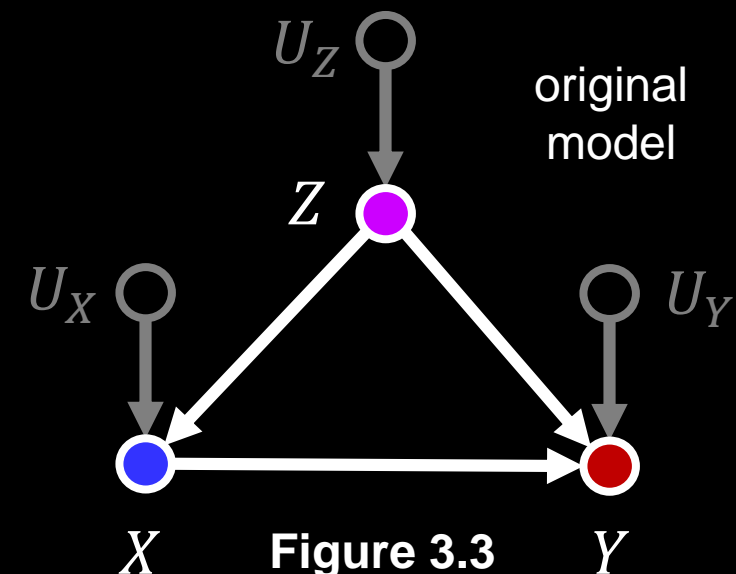
This consideration also allows us to generalize the adjustment formula to multiple interventions, that is, interventions that fix the values of a set of variables X to constants.

We simply write down the product decomposition of the preintervention distribution, and strike out all factors that correspond to variables in the intervention set X .

Formally, we write

$$P(x_1, x_2, \dots, x_n | do(x)) = \prod_{i=1}^n P(x_i | pa(x_i)), \quad \forall i : X_i \notin X$$

This came to be known as the **truncated product formula** or **g-formula**.



3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

To illustrate, assume that we intervene on the model of **Figure 2.9** and set X to x and Z_3 to z_3 .

$$P(z_1, z_2, w, y | do(X = x, Z_3 = z_3)) = P(z_1) P(z_2) P(x | \cancel{z_1}, z_3) P(z_3 | \cancel{z_1}, z_2) P(w | x) P(y | w, z_3, z_2)$$

The **postintervention distribution** of the other variables in the model is obtained by deleting the factors

$$P(x | z_1, z_3)$$

$$P(z_3 | z_1, z_2)$$

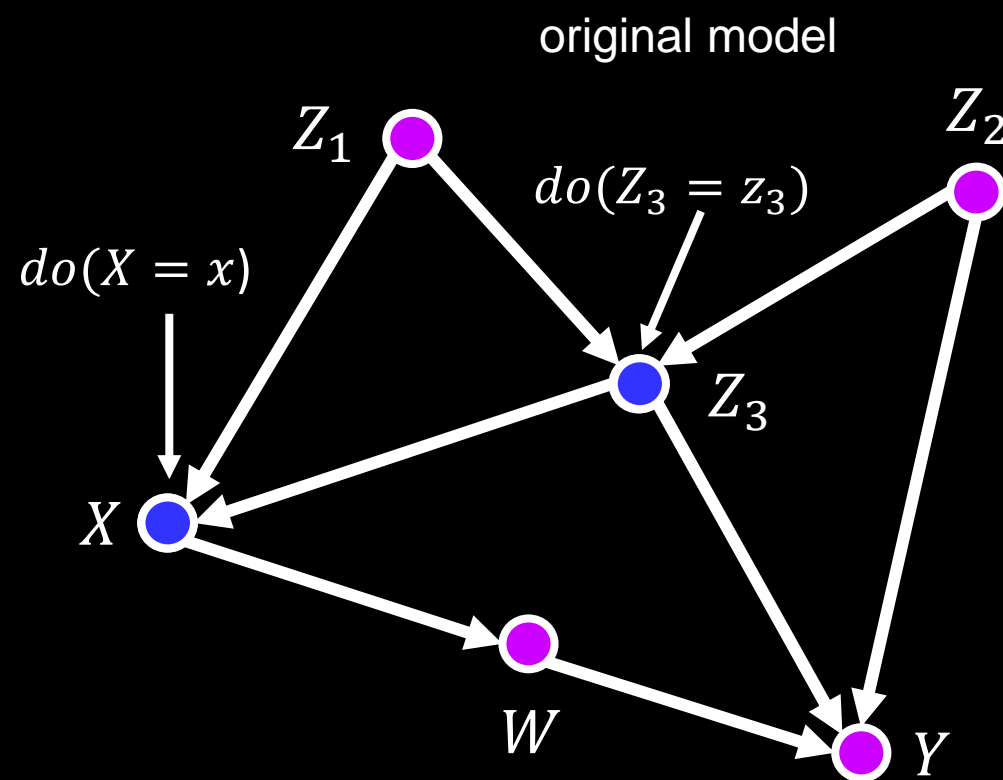


Figure 2.9

3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

To illustrate, assume that we intervene on the model of **Figure 2.9** and set X to x and Z_3 to z_3 .

$$P(z_1, z_2, w, y | do(X = x, Z_3 = z_3)) = P(z_1) P(z_2) P(x | \cancel{z_1}, z_3) P(z_3 | \cancel{z_1}, z_2) P(w | x) P(y | w, z_3, z_2)$$

The **postintervention distribution** of the other variables in the model is obtained by deleting the factors

$$P(x | z_1, z_3)$$

$$P(z_3 | z_1, z_2)$$

$$P(x | do(x)) = 1$$

$$P(z_3 | do(z_3)) = 1$$

Therefore, the **postintervention distribution** of the other variables is

$$P(z_1, z_2, w, y | do(X = x, Z_3 = z_3)) = P(z_1) P(z_2) P(w | x) P(y | w, z_3, z_2)$$

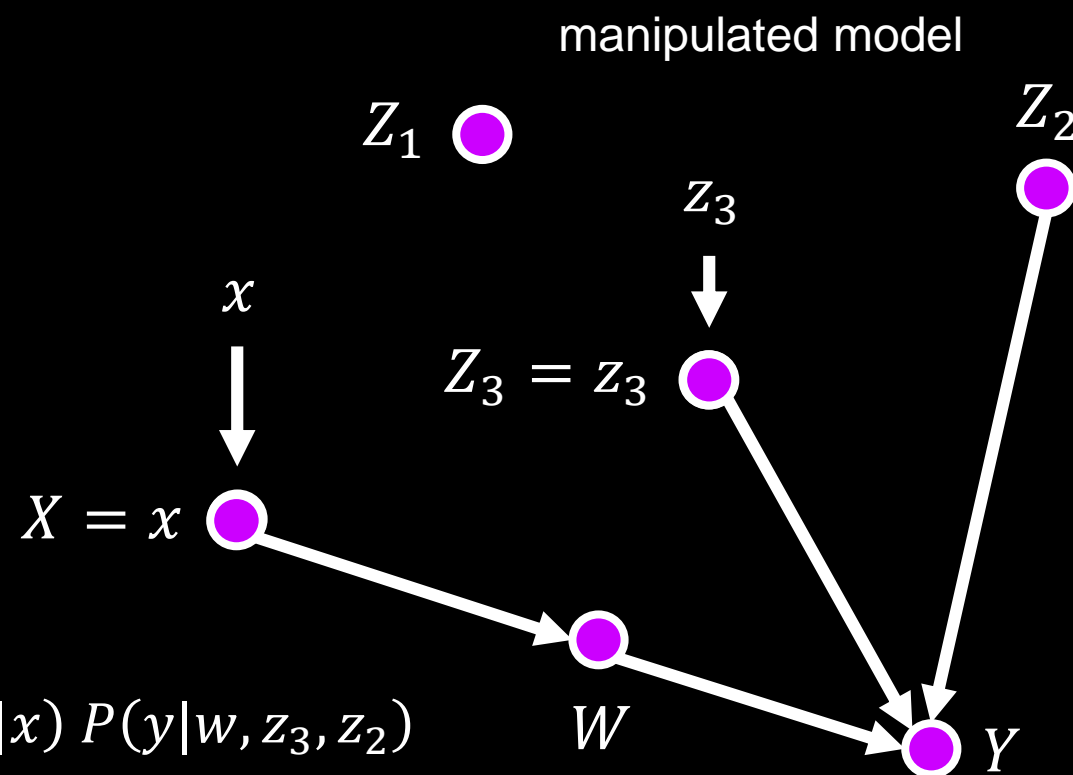


Figure 2.9

3.2.2 THE ADJUSTMENT FORMULA: MULTIPLE INTERVENTIONS AND THE TRUNCATED PRODUCT RULE

It is interesting to note that combining

$$P(x, y, z) = P(z) P(x|z) P(y|x, z) \quad \text{preintervention distribution}$$

and

$$P(z, y|do(x)) = P_m(z) P_m(y|x, z) = P(z) P(y|x, z) \quad \text{postintervention distribution}$$

we get a simple relation between the pre-and postintervention distributions:

$$P(z, y|do(x)) = \frac{P(x, y, z)}{P(x|z)}$$

It tells us that the conditional probability $P(x|z)$ is all we need to know in order to predict the effect of an intervention $do(x)$ from nonexperimental data governed by the distribution $P(x, y, z)$.

3.3 THE BACKDOOR CRITERION

In the previous section, we came to the conclusion that we should adjust for a variable's parents, when trying to determine its effect on another variable.

But often, we know, or believe, that the variables have unmeasured parents that, though represented in the graph, may be inaccessible for measurement.

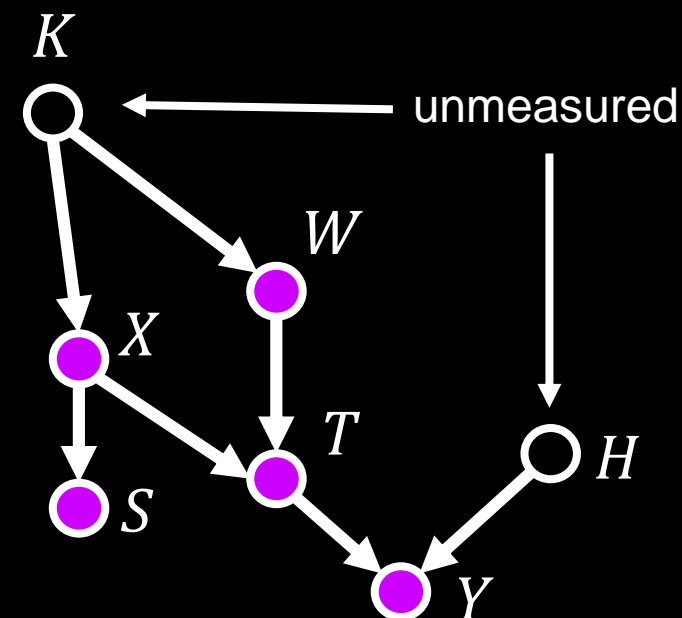
In those cases, we need to find an alternative set of variables to adjust for.

This dilemma unlocks a deeper statistical question:

Under what conditions does a causal story permit us to compute the causal effect of one variable on another, from data obtained by passive observations, with no interventions?

Since we have decided to represent causal stories with graphs, the question becomes a graph-theoretical problem:

Under what conditions, is the structure of the causal graph sufficient for computing a causal effect from a given data set?



3.3 THE BACKDOOR CRITERION

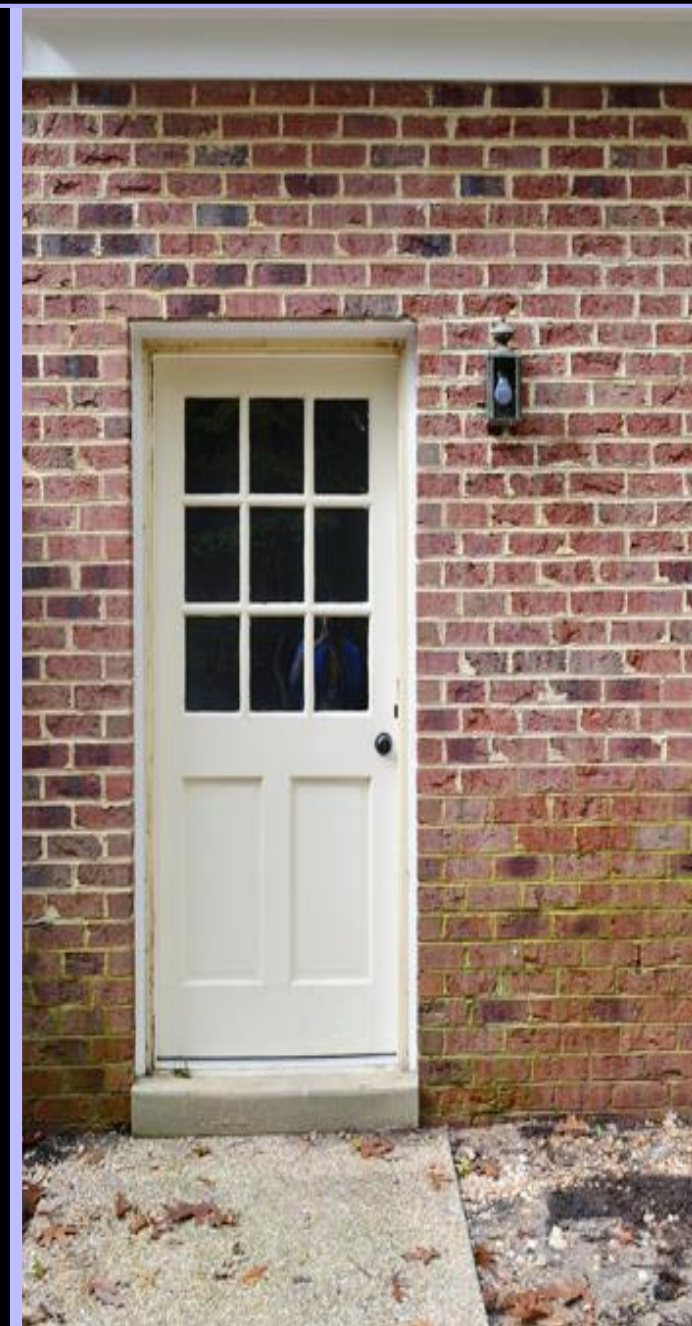
The answer to that question is long enough—and important enough—that we will spend the rest of the lecture addressing it.

But one of the most important tools we use to determine whether we can compute a causal effect is a simple test called the **backdoor criterion**.

Using it, we can determine, for any two variables X and Y in a causal model represented by a DAG, which set of variables Z in that model should be conditioned on when searching for the causal relationship between X and Y .

Definition 3.3.1 (The Backdoor Criterion)

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



3.3 THE BACKDOOR CRITERION

If a set of variables Z satisfies the backdoor criterion for X and Y , then the causal effect of X on Y is given by the formula

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

just as when we adjust for $pa(X)$.

(Note that $pa(X)$ always satisfies the backdoor criterion.)

Definition 3.3.1 (The Backdoor Criterion)

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



3.3 THE BACKDOOR CRITERION

If a set of variables Z satisfies the backdoor criterion for X and Y , then the causal effect of X on Y is given by the formula

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

just as when we adjust for $pa(X)$.

(Note that $pa(X)$ always satisfies the backdoor criterion.)

The logic behind the backdoor criterion is fairly straightforward.

In general, we would like to condition on a set of nodes Z such that we

1. block all spurious paths between X and Y .
2. leave all directed paths from X to Y unperturbed.
3. create no new spurious paths.



3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

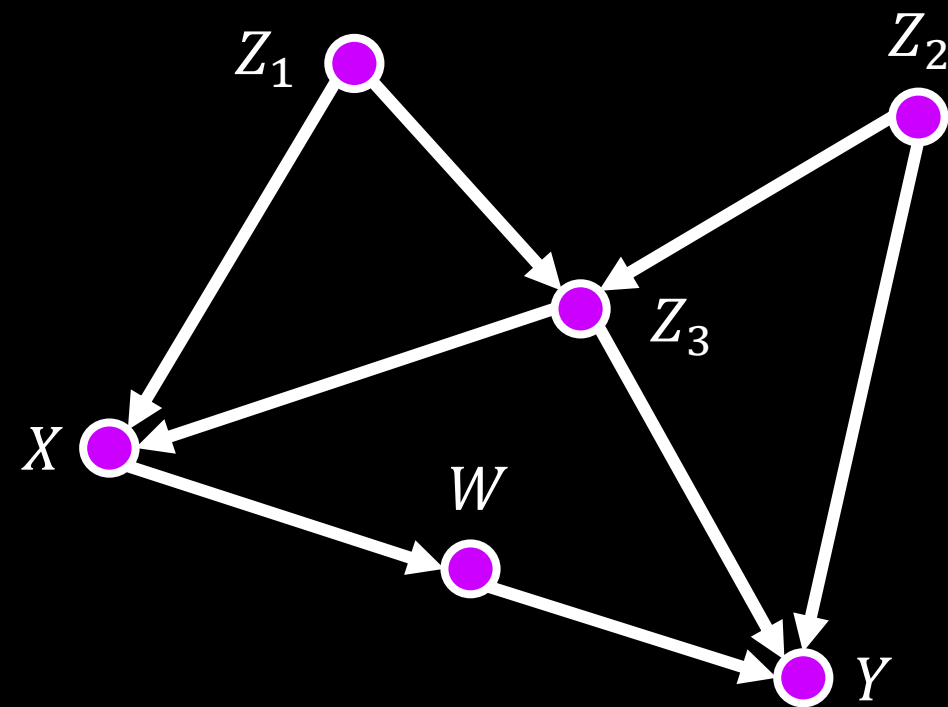


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

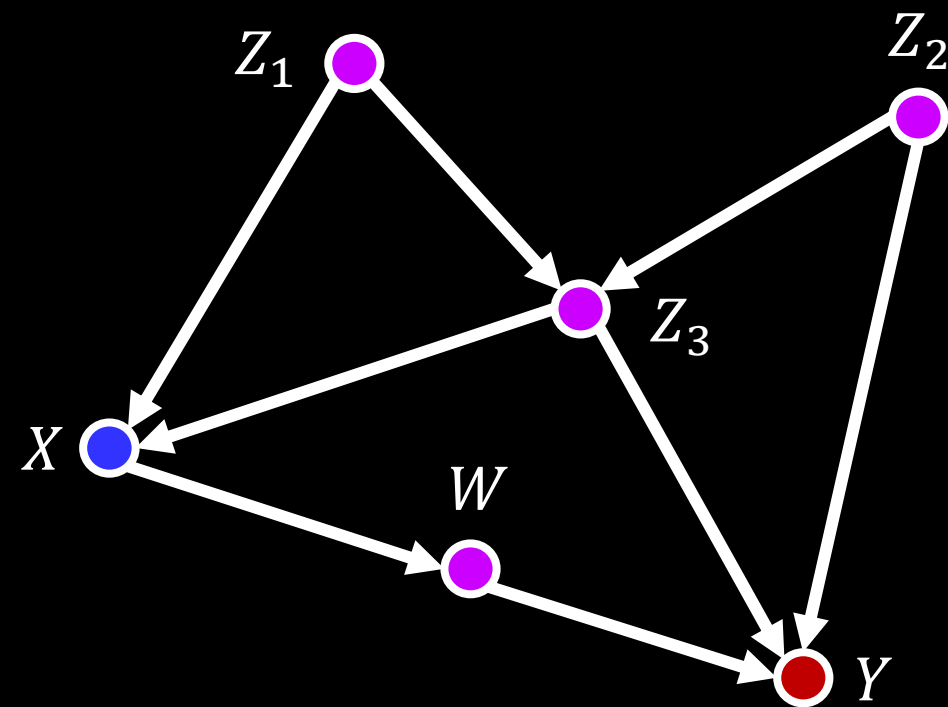


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

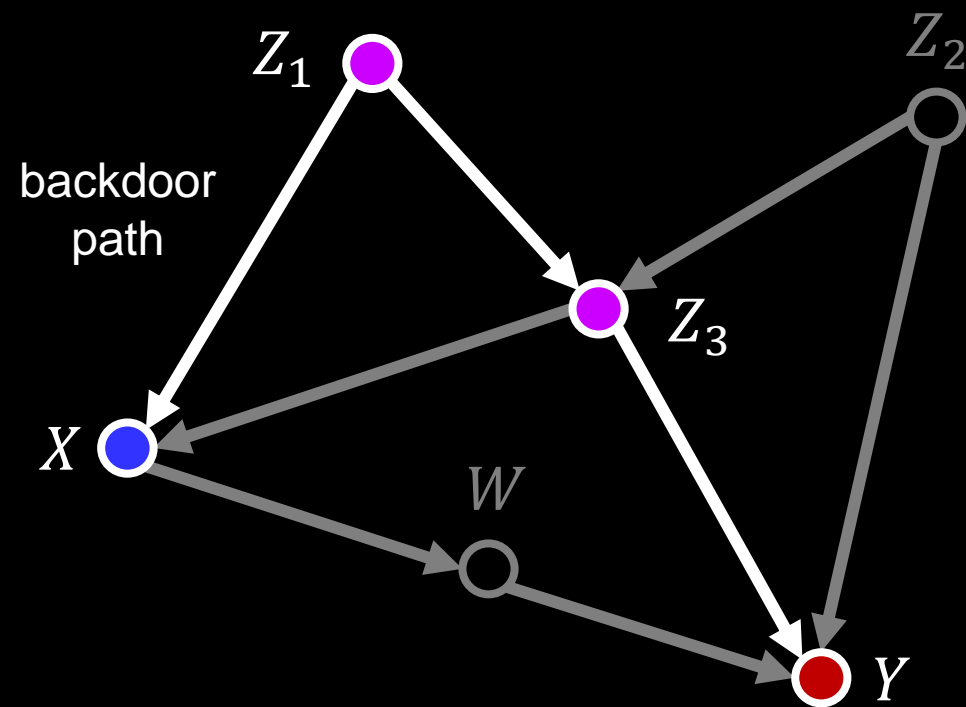


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

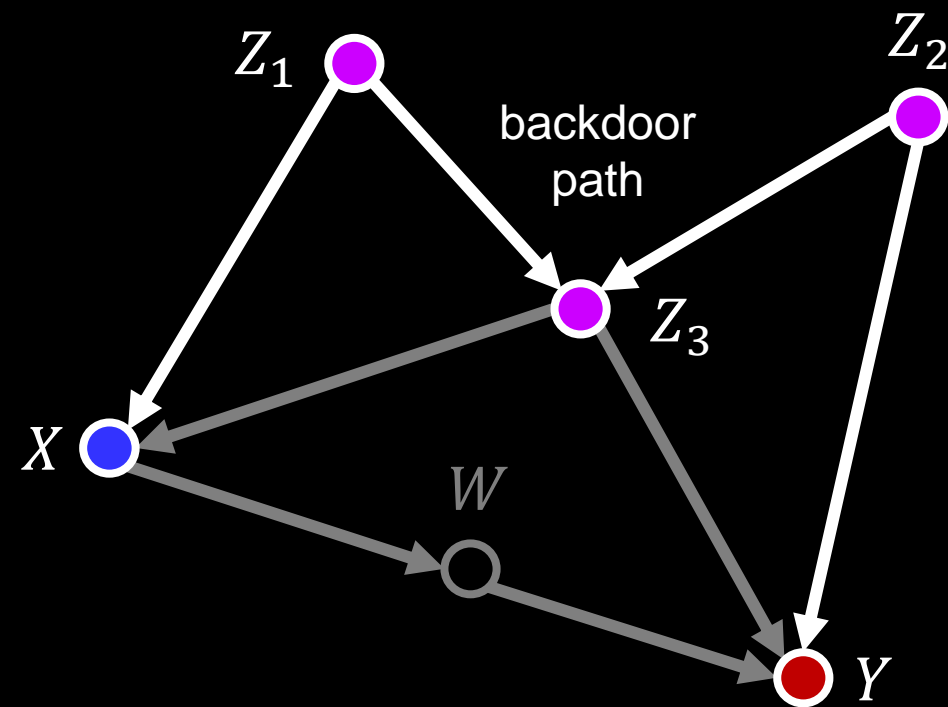


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

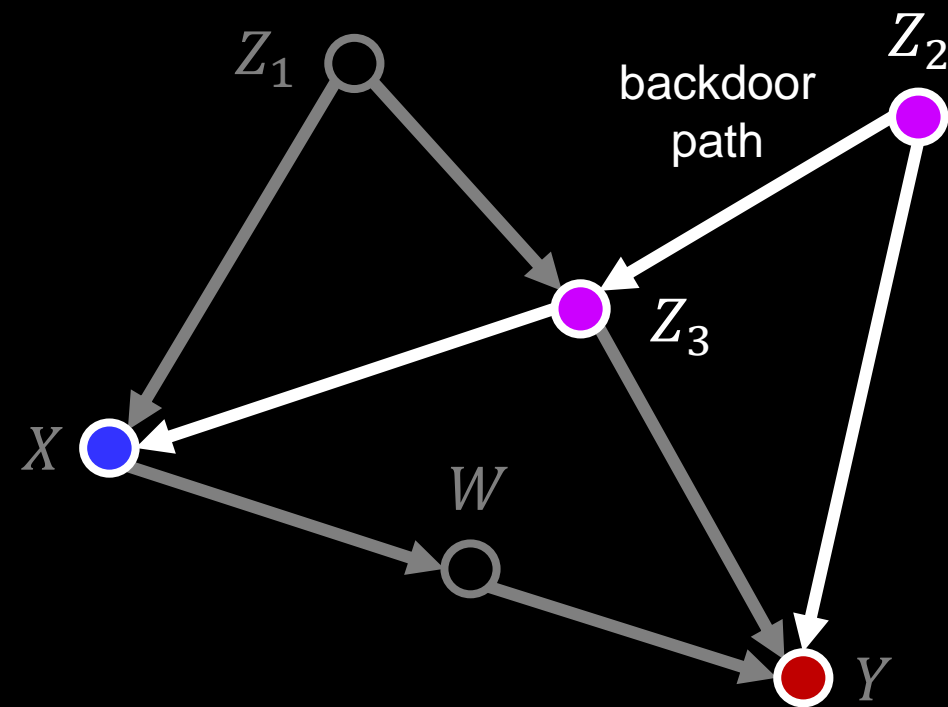


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

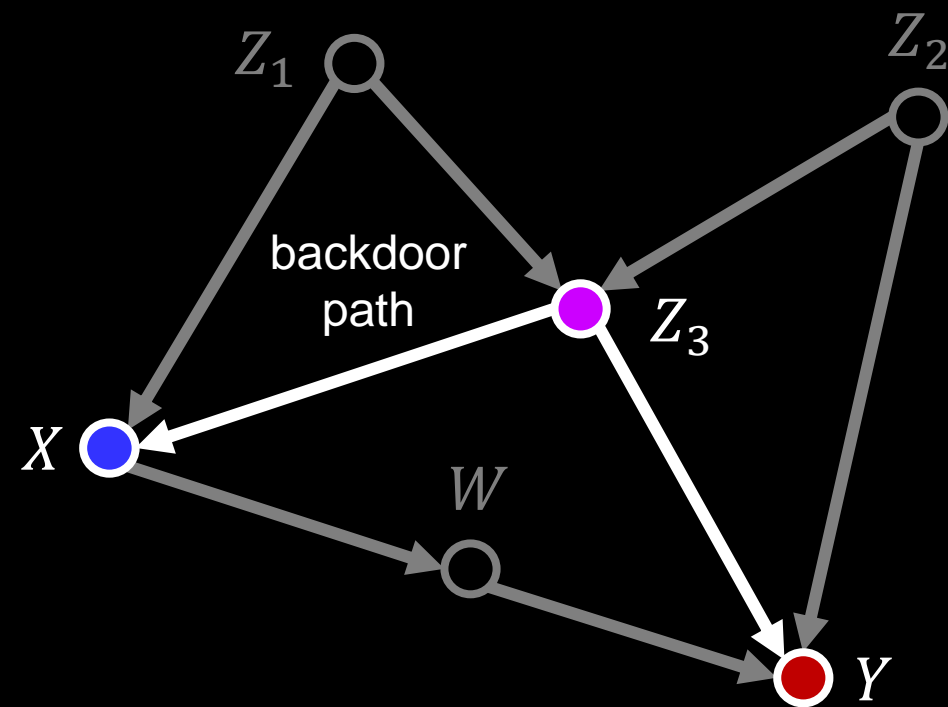


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

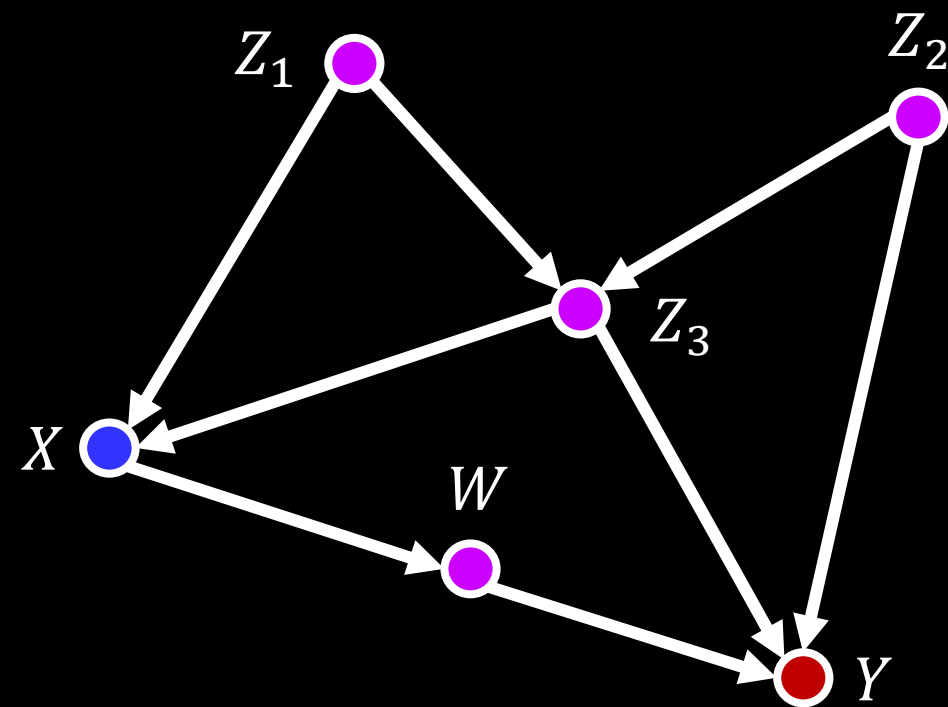


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

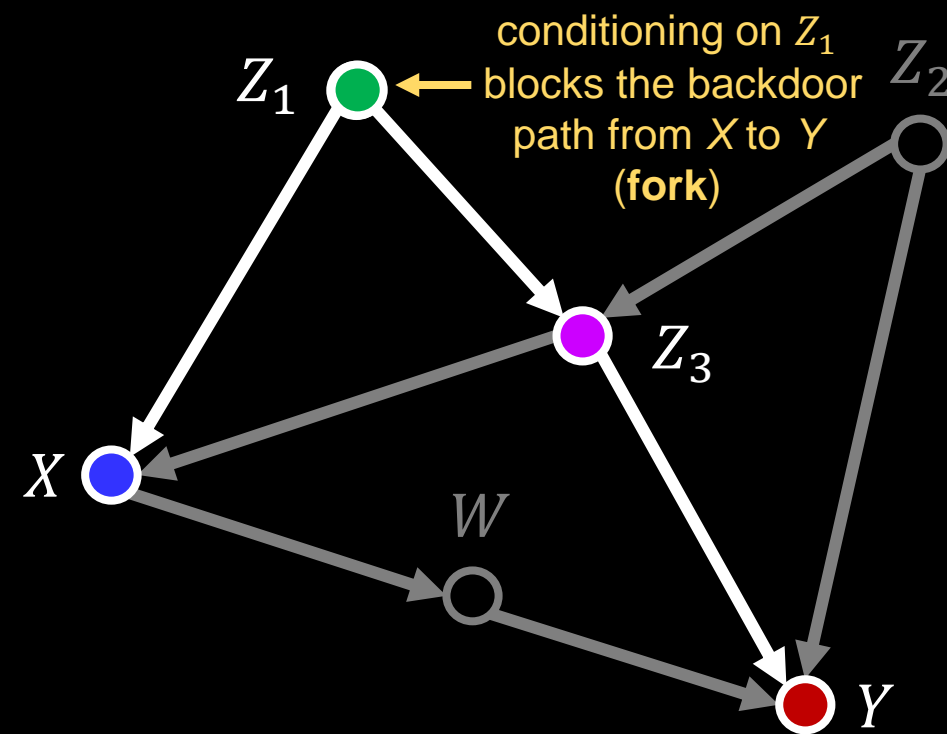


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

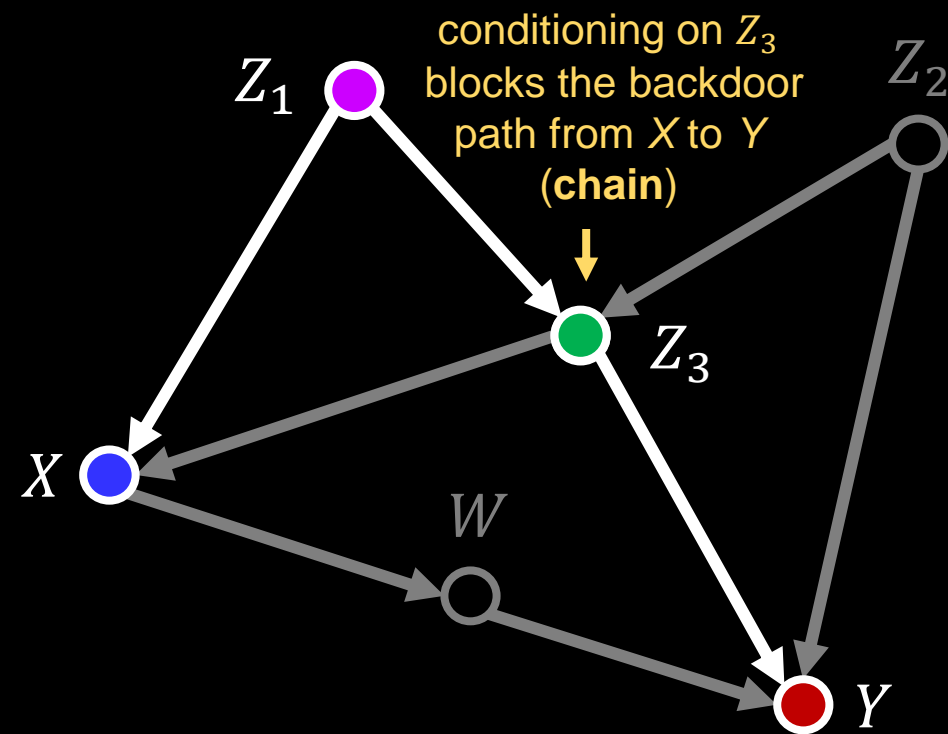


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

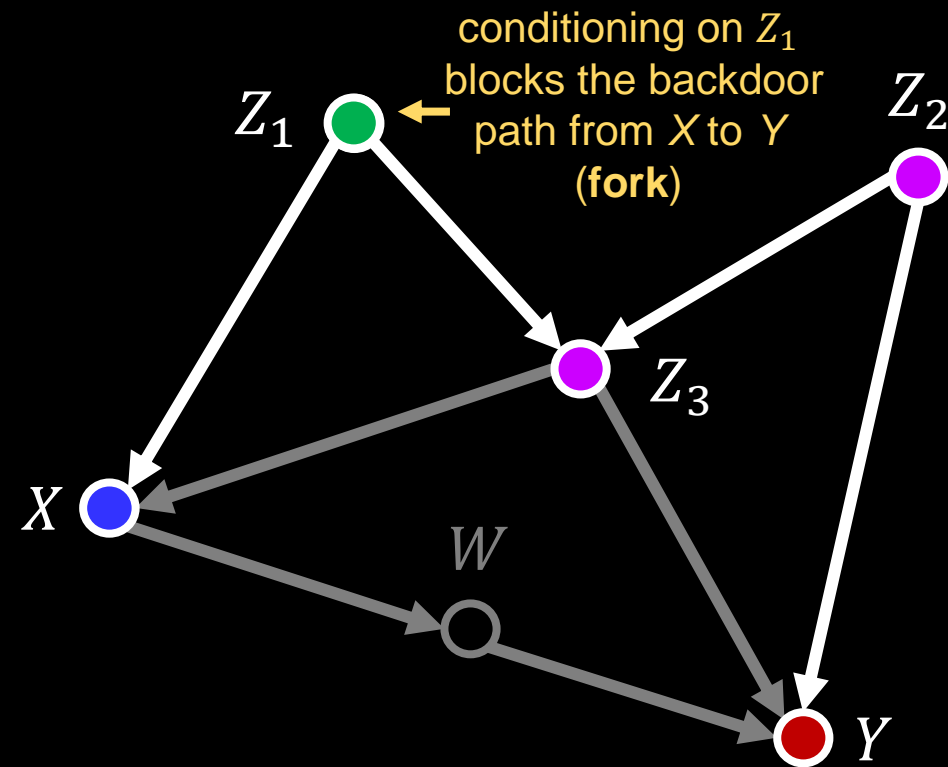


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

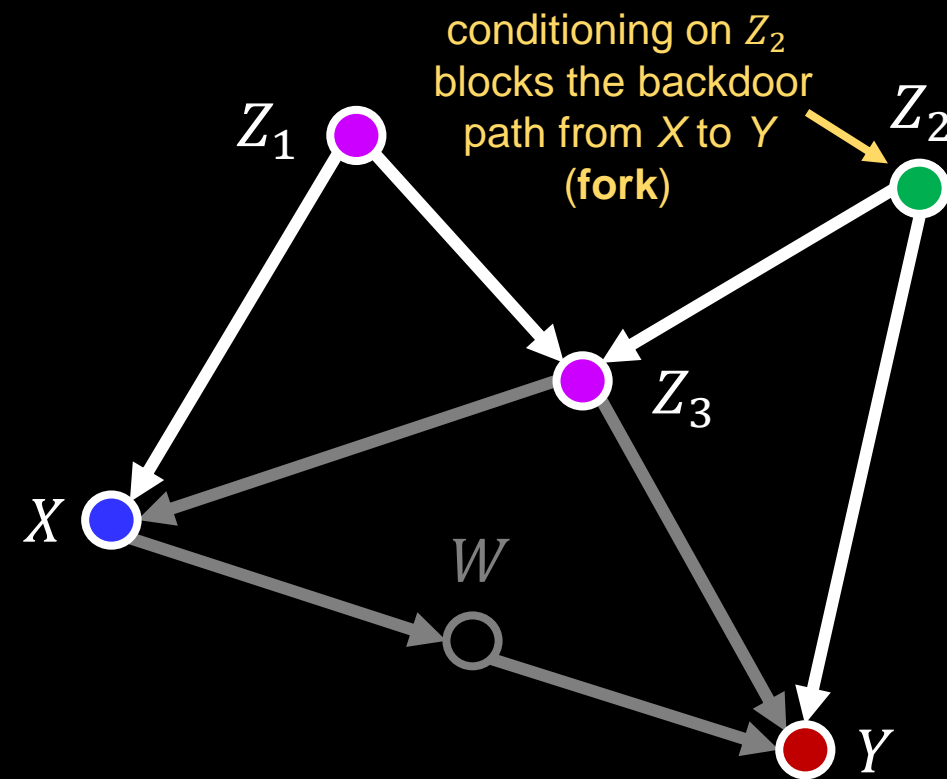


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

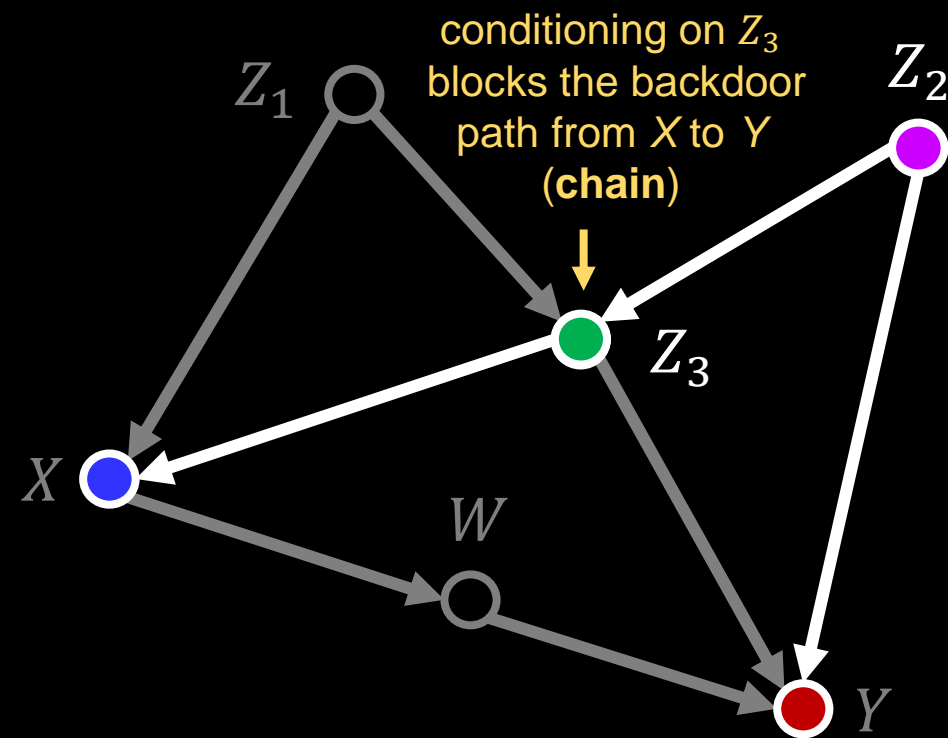


Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .

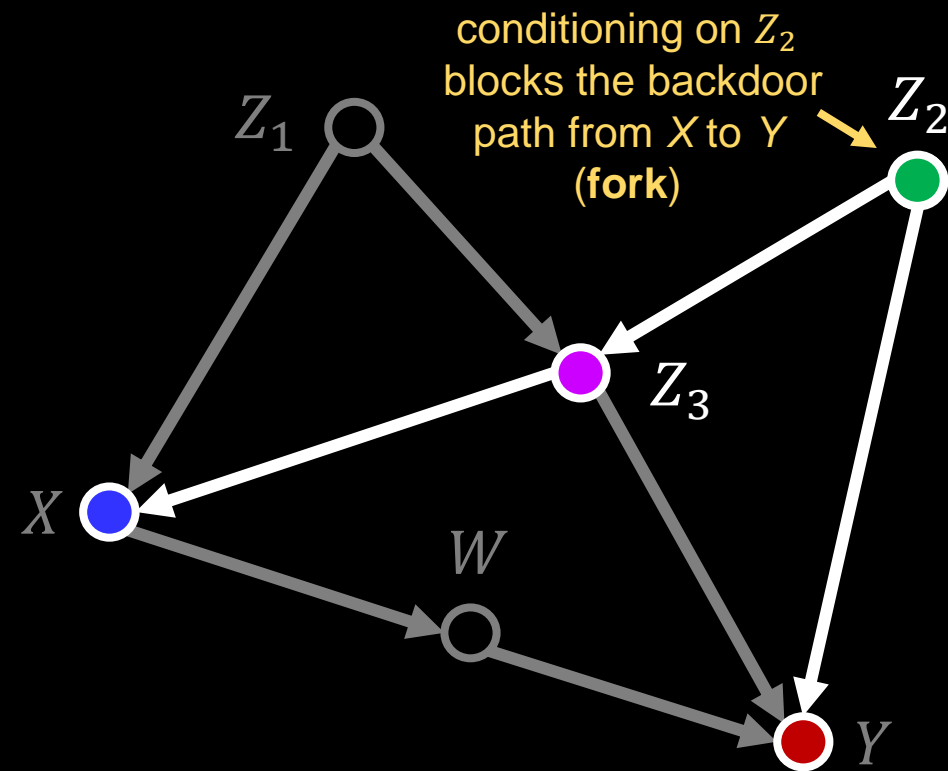


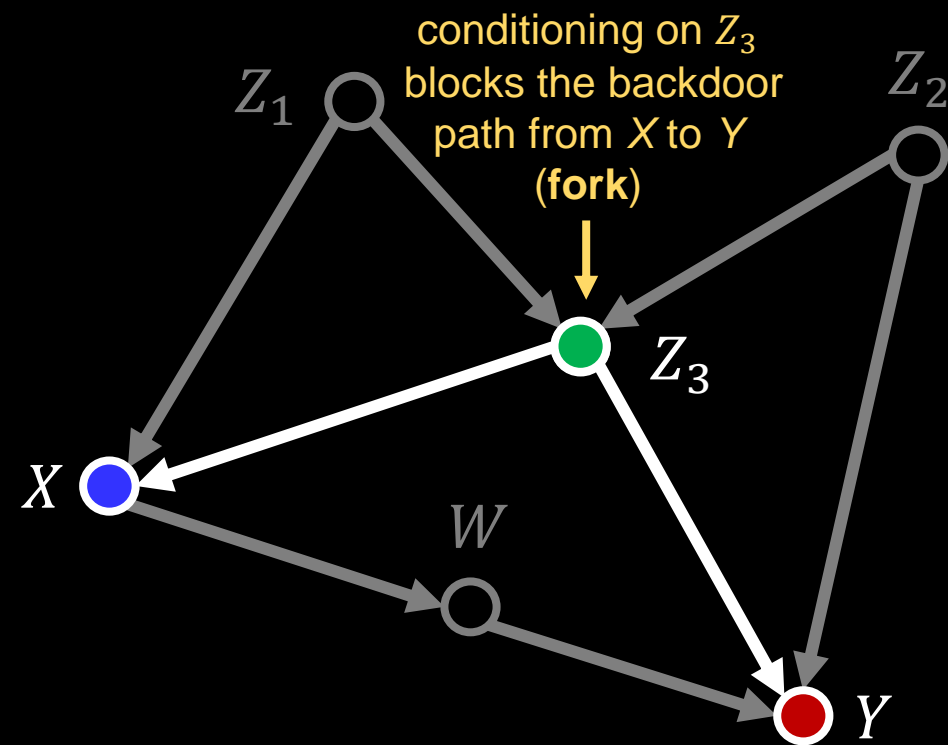
Figure 2.9

3.3 THE BACKDOOR CRITERION

1. Block all spurious paths between X and Y .

When trying to find the causal effect of X on Y , we want the nodes we condition on to block any “backdoor” path in which one end has an arrow into X , because such paths may make X and Y dependent, but are obviously not transmitting causal influences from X , and if we do not block them, they will confound the effect that X has on Y .

We condition on backdoor paths so as to fulfill our first requirement, i.e., block all spurious paths between X and Y .



Mind this case!!!

Figure 2.9

3.3 THE BACKDOOR CRITERION

2. Leave all directed paths from X to Y unperturbed.

However, we don't want to condition on any nodes that are descendants of X .

Descendants of X would be affected by an intervention on X and might themselves affect Y ; conditioning on them would block those pathways.

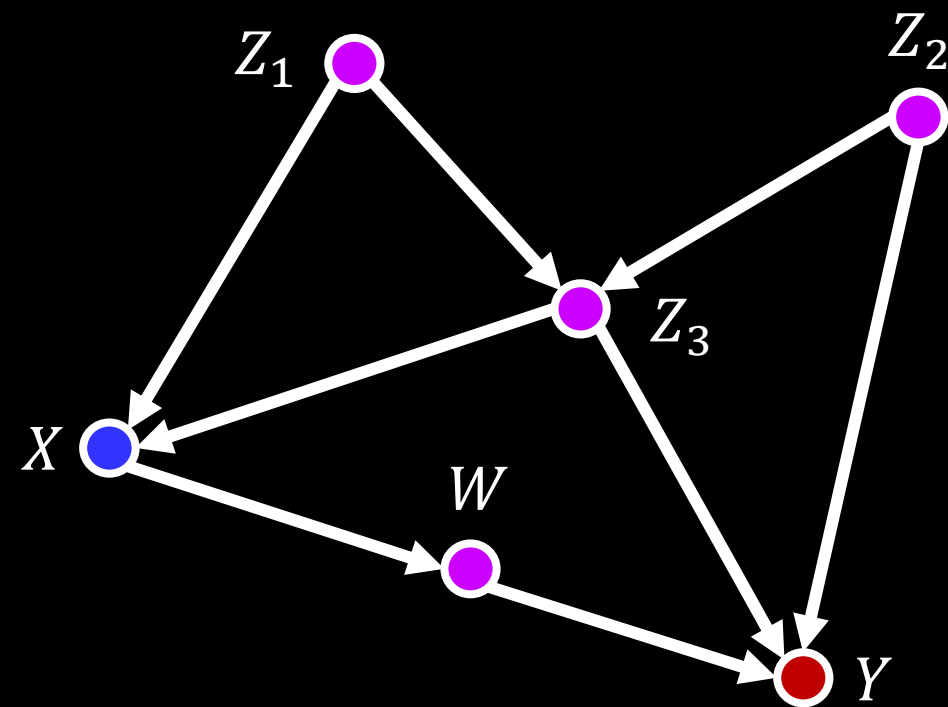


Figure 2.9

3.3 THE BACKDOOR CRITERION

2. Leave all directed paths from X to Y unperturbed.

However, we don't want to condition on any nodes that are descendants of X .

Descendants of X would be affected by an intervention on X and might themselves affect Y ; conditioning on them would block those pathways.

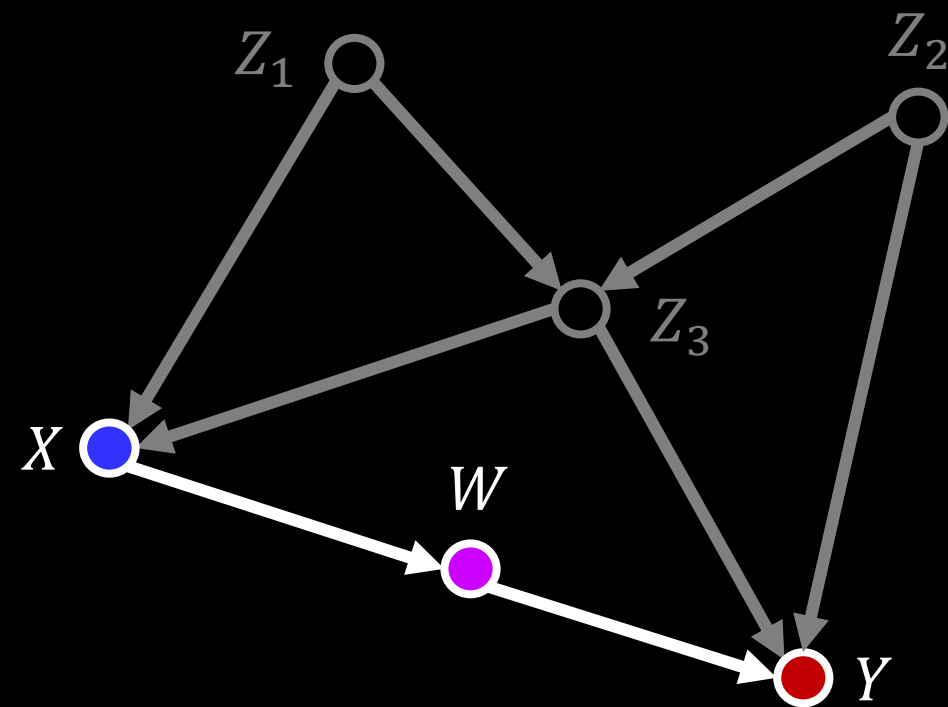


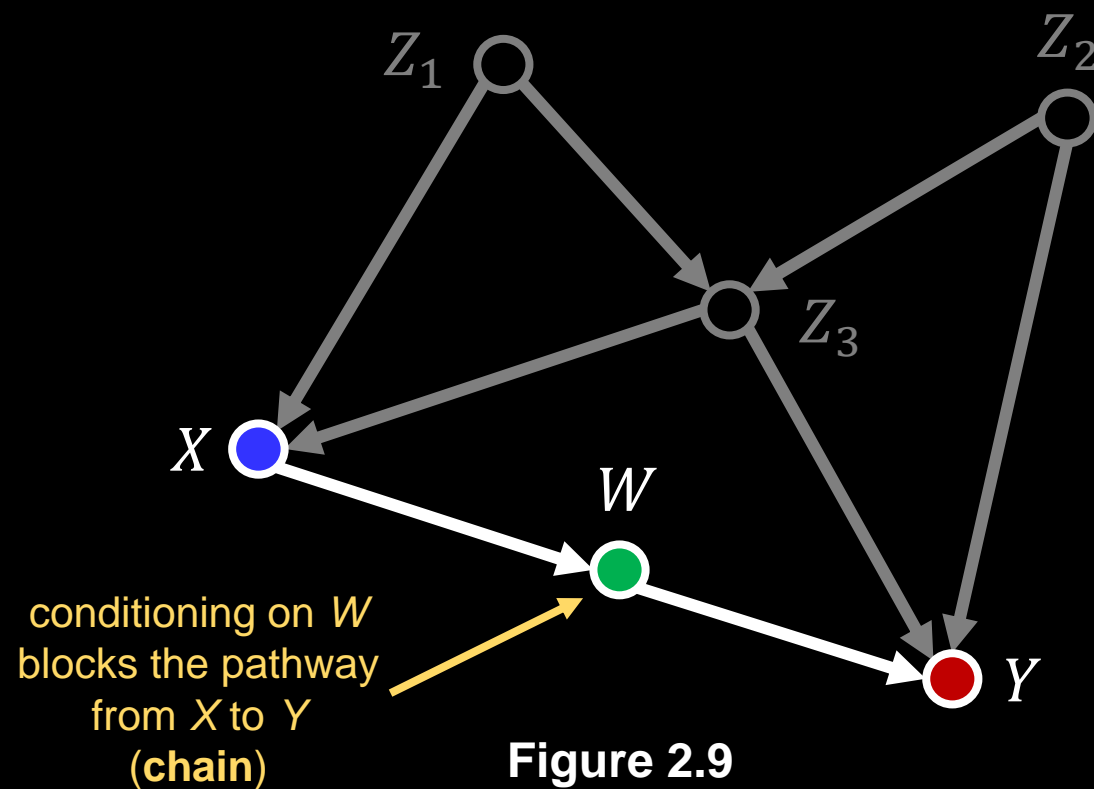
Figure 2.9

3.3 THE BACKDOOR CRITERION

2. Leave all directed paths from X to Y unperturbed.

However, we don't want to condition on any nodes that are descendants of X .

Descendants of X would be affected by an intervention on X and might themselves affect Y ; conditioning on them would block those pathways.



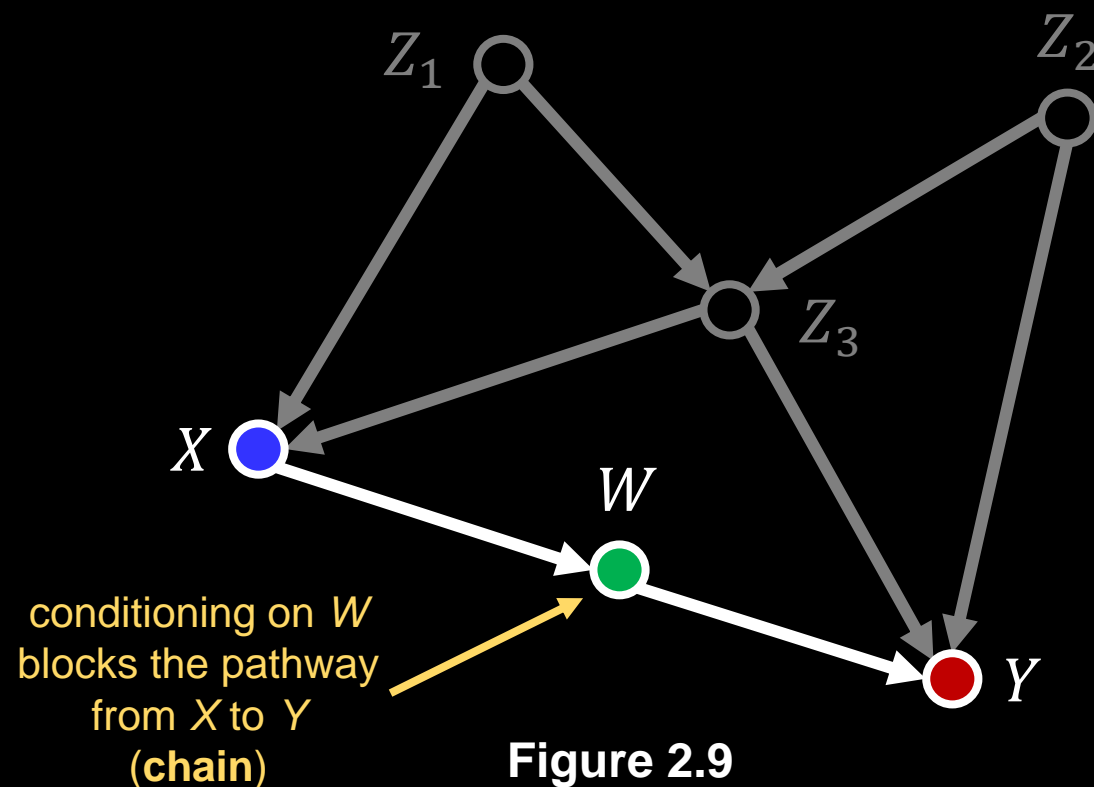
3.3 THE BACKDOOR CRITERION

2. Leave all directed paths from X to Y unperturbed.

However, we don't want to condition on any nodes that are descendants of X .

Descendants of X would be affected by an intervention on X and might themselves affect Y ; conditioning on them would block those pathways.

Therefore, we don't condition on descendants of X so as to fulfill our second requirement, i.e., leave all directed paths from X to Y unperturbed.



3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

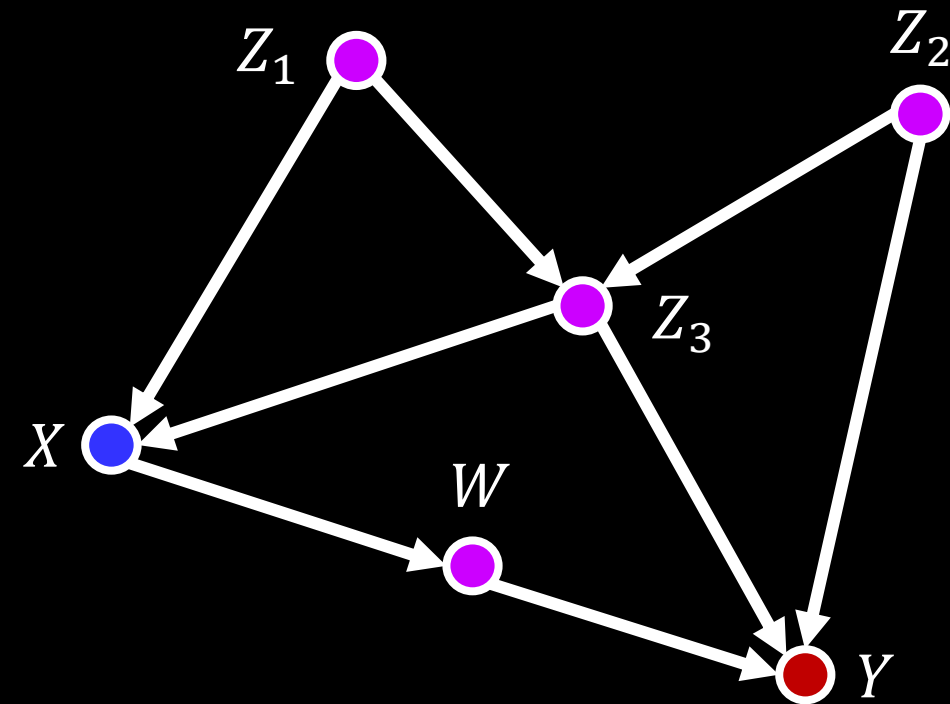


Figure 2.9

3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

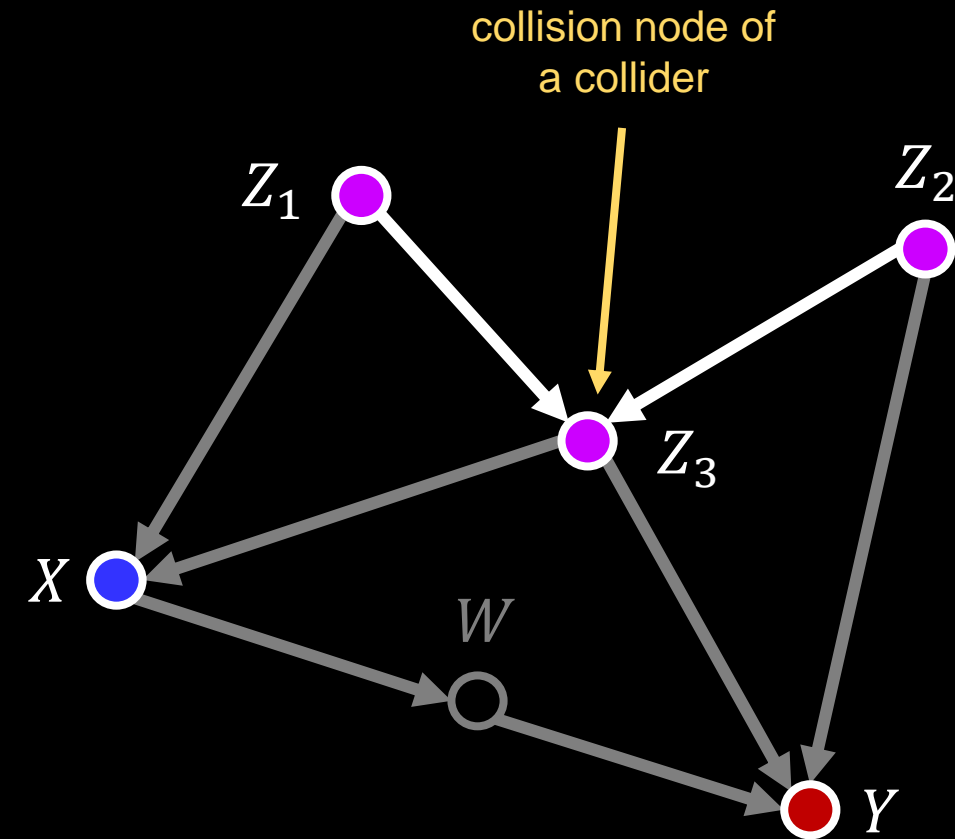


Figure 2.9

3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

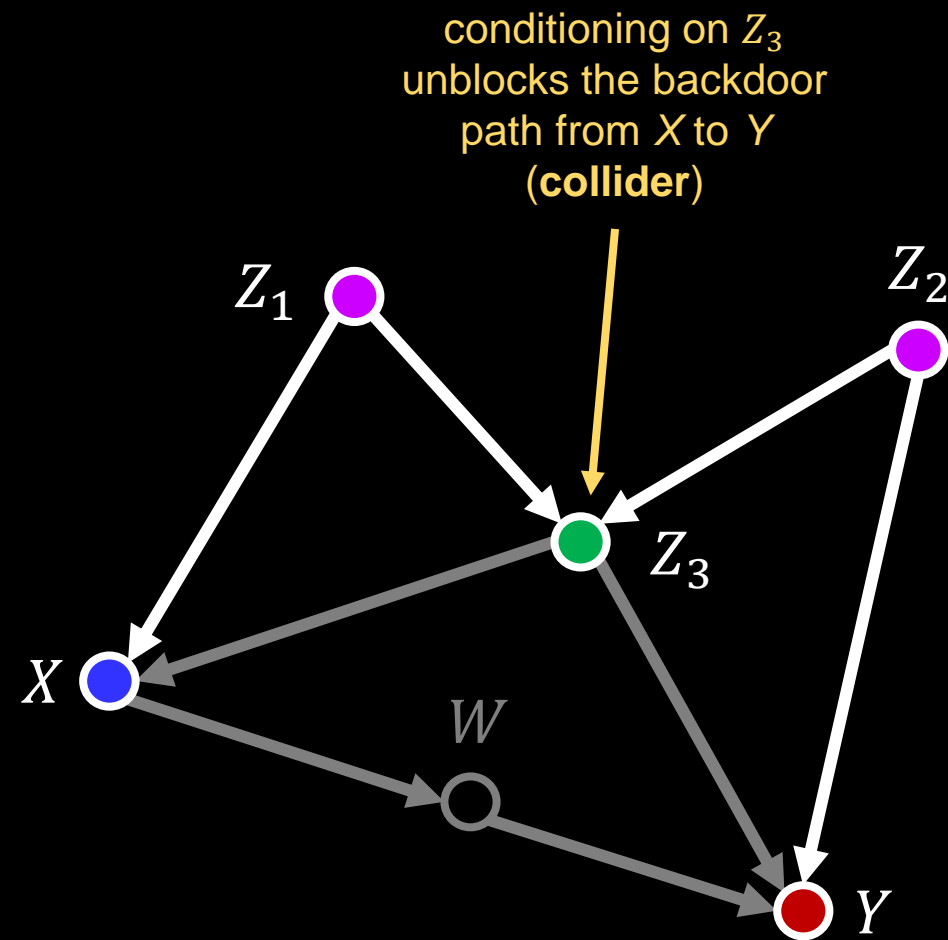


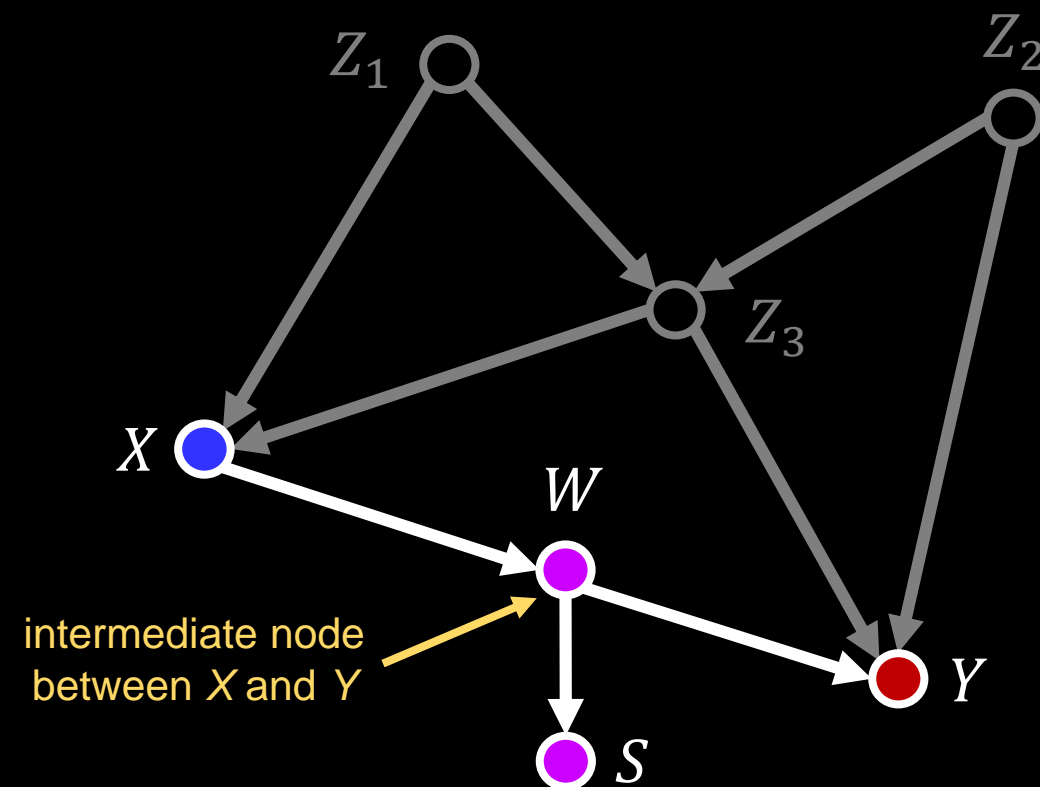
Figure 2.9

3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., node S)

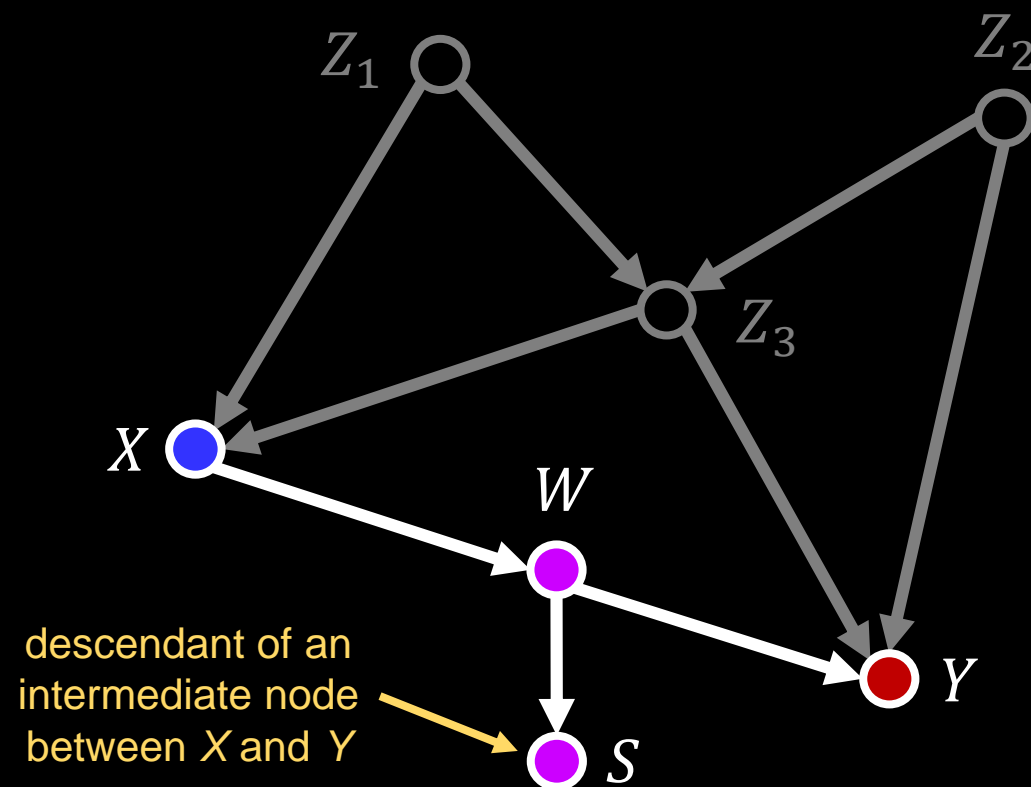


3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., node S)

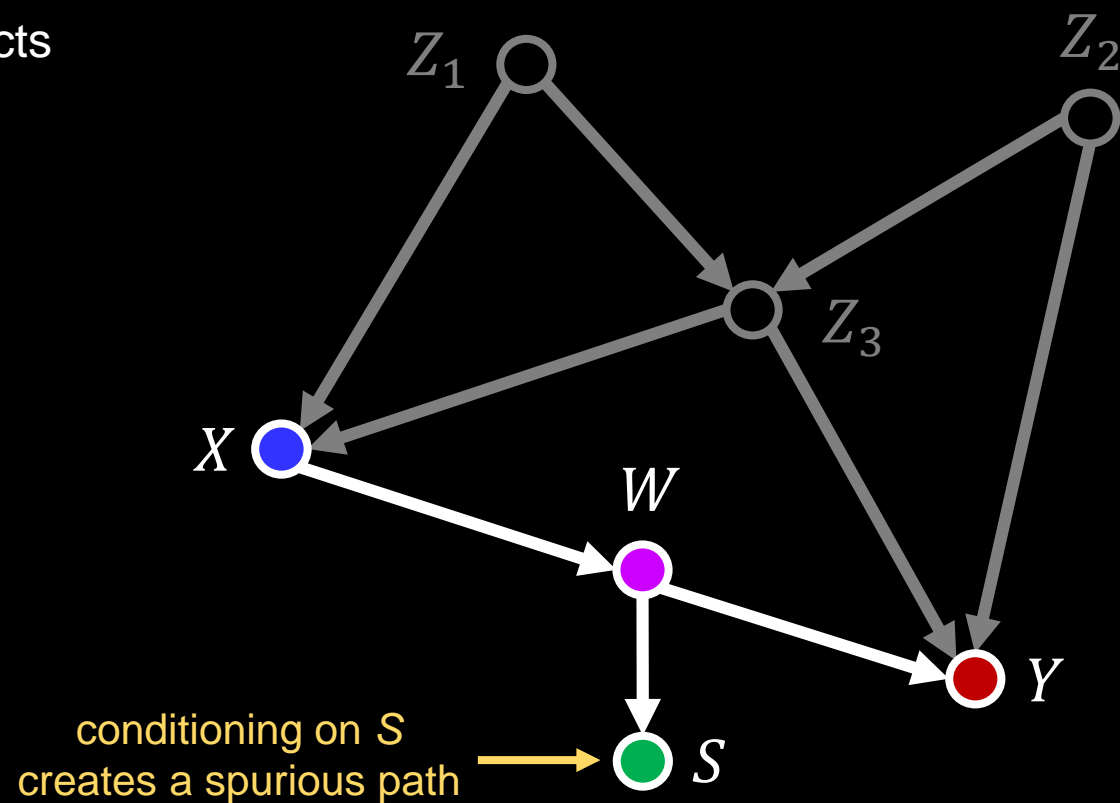


3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., node S)

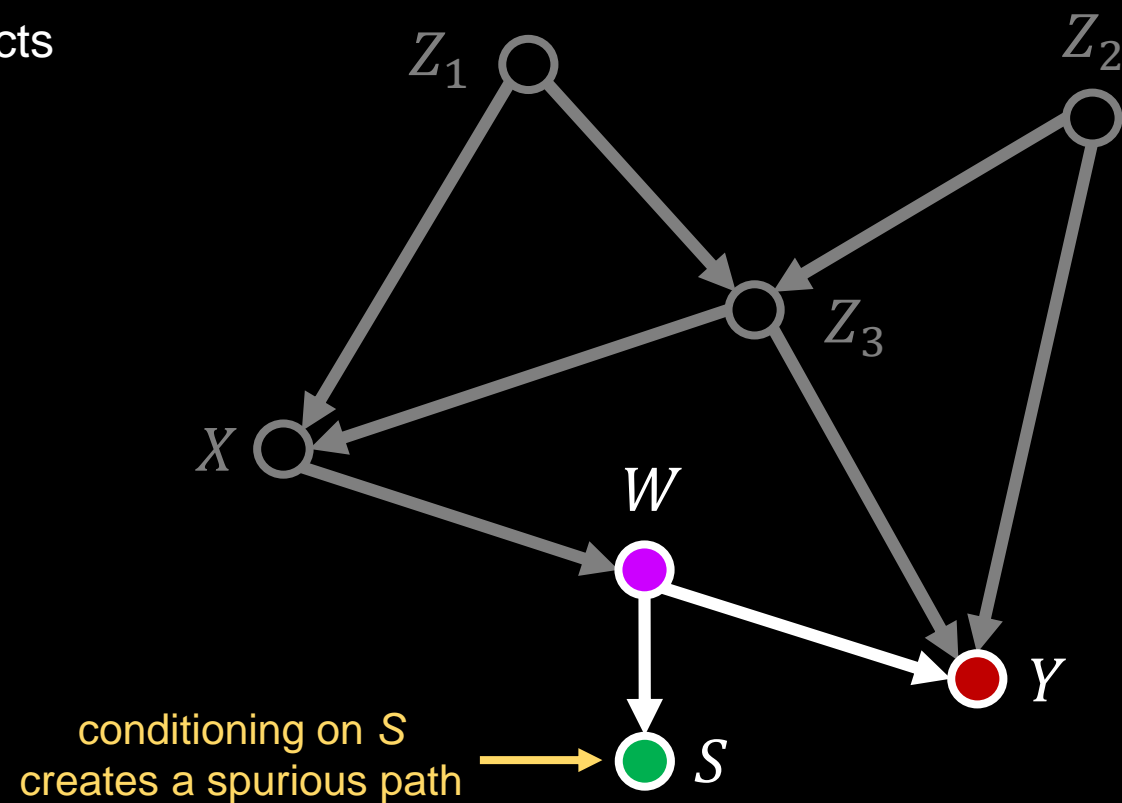


3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., node S)



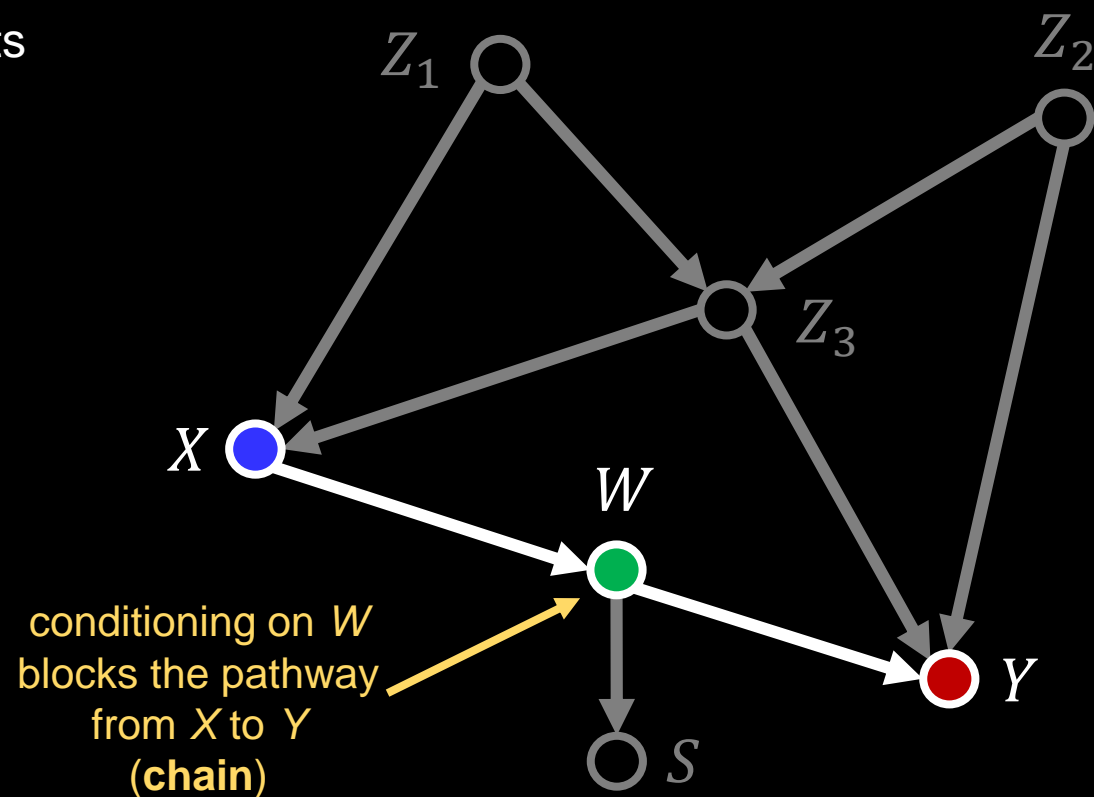
3.3 THE BACKDOOR CRITERION

3. Create no new spurious paths.

Finally, to comply with the third requirement, we should refrain from conditioning on any collider that would unblock a new path between X and Y .

The requirement of excluding descendants of X also protects us from conditioning on children of intermediate nodes between X and Y (e.g., node S)

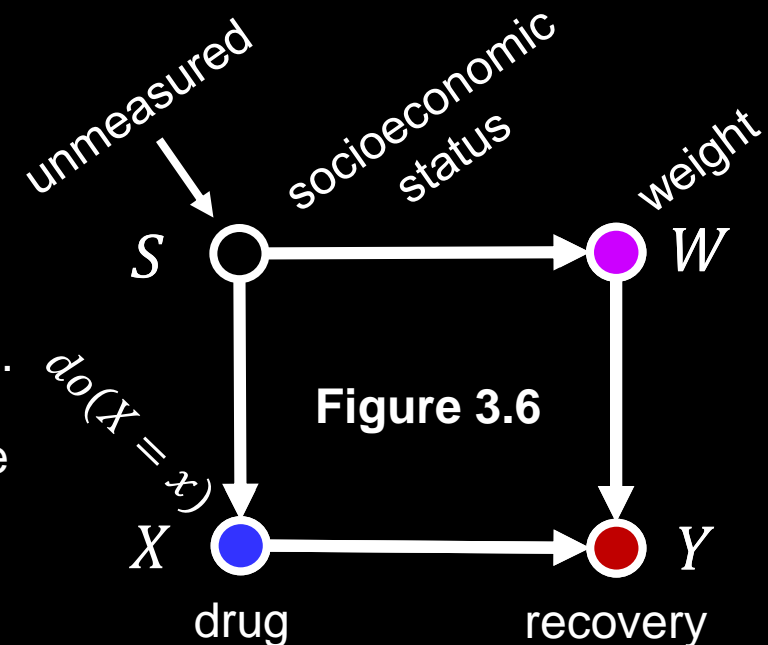
Such conditioning would distort the passage of causal association between X and Y , similar to the way conditioning on their parents would (node W).



3.3 THE BACKDOOR CRITERION

To see what this means in practice, let's look at a concrete example, shown in **Figure 3.6**, where:

- we are trying to gauge the effect of a **drug** (X) on **recovery** (Y).
- we have also measured **weight** (W), which has an effect on **recovery** (Y).
- we know that **socioeconomic status** (S) affects both **weight** (W) and the choice to receive **drug** (X) —but the study we are consulting did not record **socioeconomic status** (S).



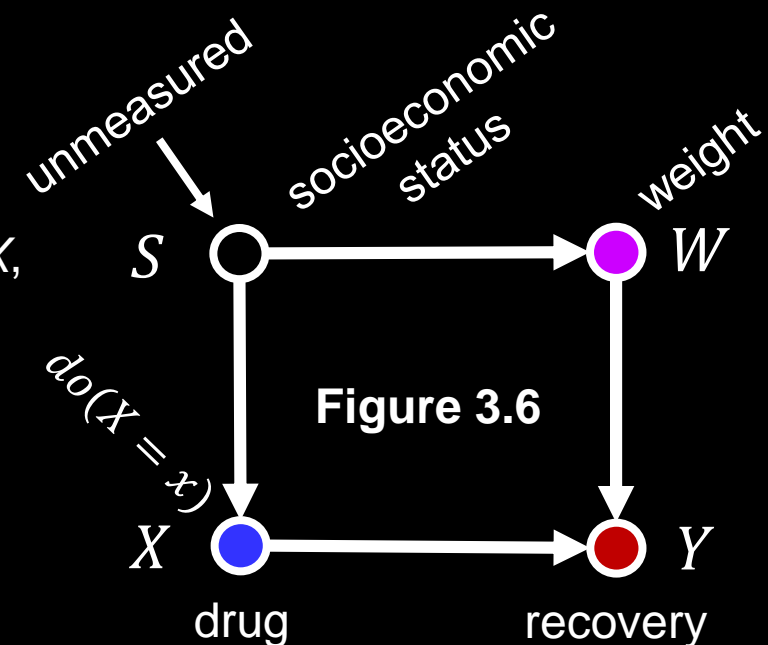
3.3 THE BACKDOOR CRITERION

Instead, we search for an observed variable that fits the **backdoor criterion** from X to Y .

A brief examination of the graph shows that W , which is not a descendant of X , also blocks the backdoor path

$$X \leftarrow S \rightarrow W \rightarrow Y.$$

Therefore, W meets the backdoor criterion.



Definition 3.3.1 (The Backdoor Criterion)

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

3.3 THE BACKDOOR CRITERION

Instead, we search for an observed variable that fits the **backdoor criterion** from X to Y .

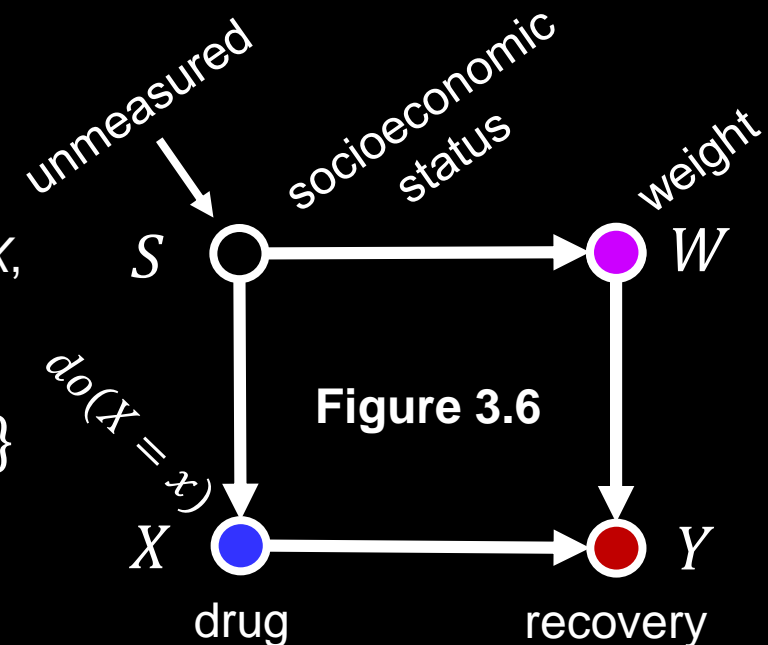
A brief examination of the graph shows that W , which is not a descendant of X , also blocks the backdoor path

$$X \leftarrow S \rightarrow W \rightarrow Y.$$

$$Z = \{W\}$$

Therefore, W meets the backdoor criterion.

- $Z = \{W\}$ is not a descendant of X
- $Z = \{W\}$ blocks every path between X and Y that contains an arrow into X



Definition 3.3.1 (The Backdoor Criterion)

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

3.3 THE BACKDOOR CRITERION

Instead, we search for an observed variable that fits the **backdoor criterion** from X to Y .

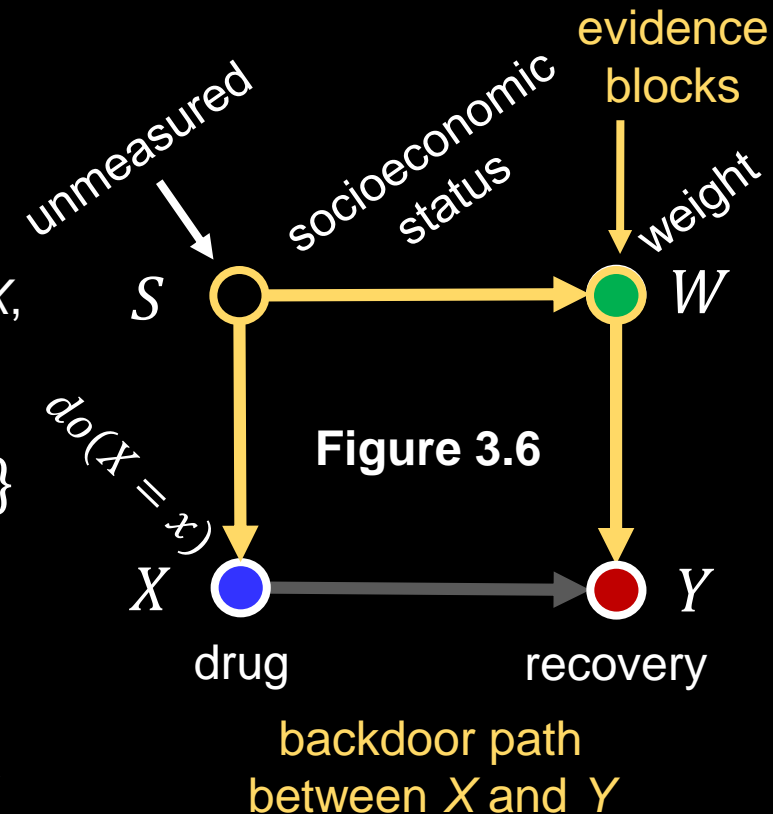
A brief examination of the graph shows that W , which is not a descendant of X , also blocks the backdoor path

$$X \leftarrow S \rightarrow W \rightarrow Y.$$

$$Z = \{W\}$$

Therefore, W meets the backdoor criterion.

- $Z = \{W\}$ is not a descendant of X
- $Z = \{W\}$ blocks every path between X and Y that contains an arrow into X



Definition 3.3.1 (The Backdoor Criterion)

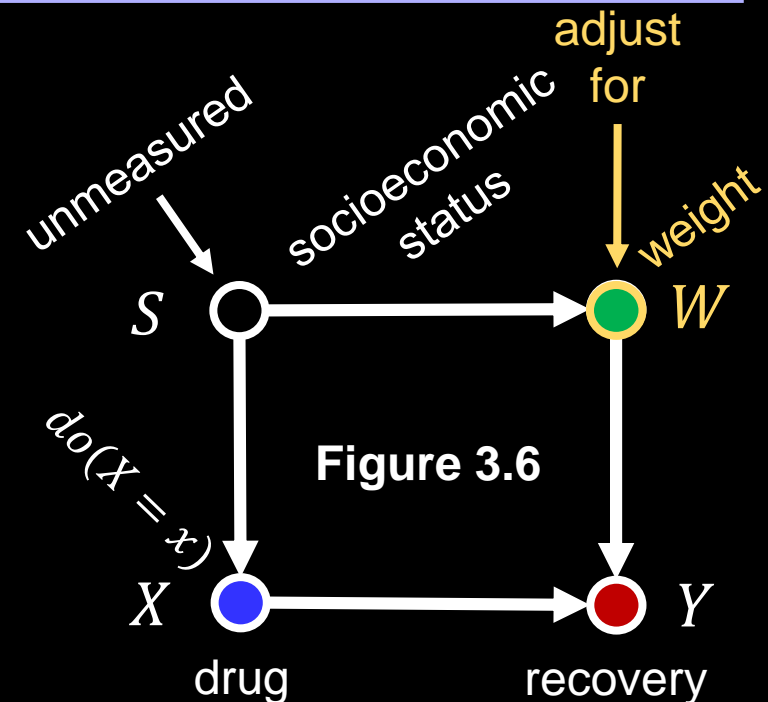
Given an ordered pair of variables (X, Y) in a directed acyclic graph G , a set of variables Z satisfies the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X , and Z blocks every path between X and Y that contains an arrow into X .

3.3 THE BACKDOOR CRITERION

So long as the causal story conforms to the graph in **Figure 3.6**, **adjusting for W** will give us the causal effect of X on Y .

Using the **adjustment formula**, we find

$$P(Y = y | do(X = x)) = \sum_w P(Y = y | X = x, W = w) P(W = w)$$



This sum can be estimated from our observational data, so long as W is observed.

With the help of the backdoor criterion, you can easily and algorithmically come to a conclusion about a pressing policy concern, even in complicated graphs.

3.3 THE BACKDOOR CRITERION

Consider the model in **Figure 2.8**, and assume again that we wish to evaluate the effect of X on Y .

What variables should we condition on to obtain the correct effect?

The question boils down to finding a set of variables Z that satisfy the backdoor criterion, but since there are no backdoor paths from X to Y , the answer is trivial:

The empty set satisfies the criterion, hence no adjustment is needed.

- no node in Z is a descendant of X

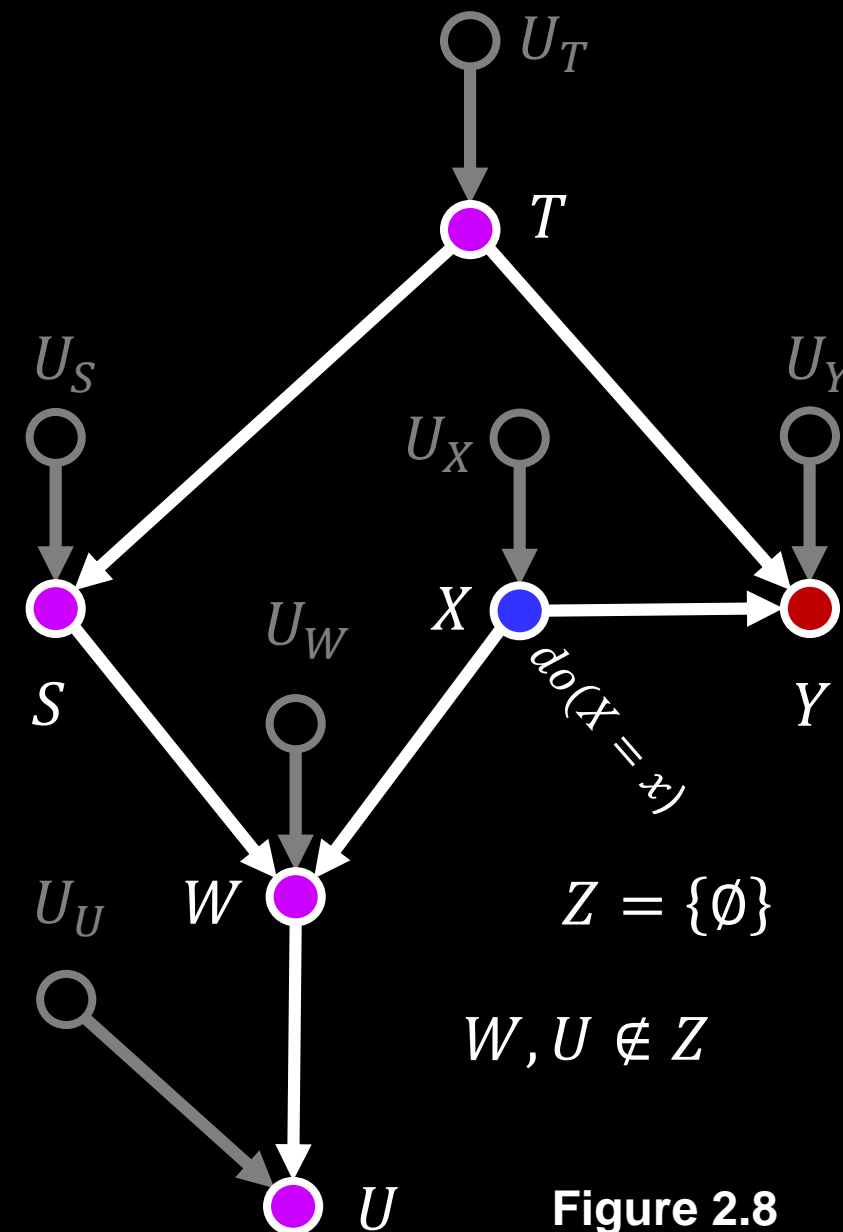


Figure 2.8

3.3 THE BACKDOOR CRITERION

Consider the model in **Figure 2.8**, and assume again that we wish to evaluate the effect of X on Y .

What variables should we condition on to obtain the correct effect?

The question boils down to finding a set of variables Z that satisfy the backdoor criterion, but since there are no backdoor paths from X to Y , the answer is trivial:

The empty set satisfies the criterion, hence no adjustment is needed.

- no node in Z is a descendant of X
- Z blocks every path between X and Y that contains an arrow into X

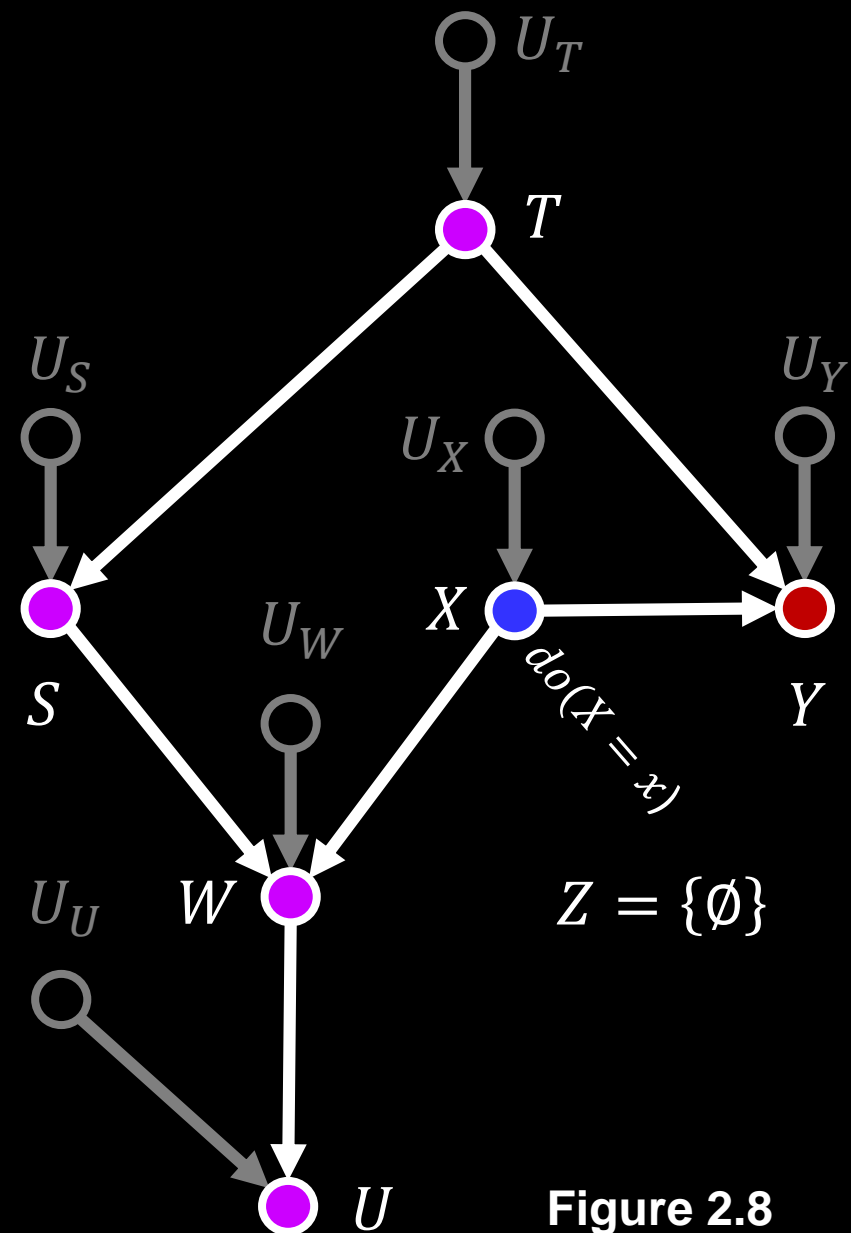


Figure 2.8

3.3 THE BACKDOOR CRITERION

Consider the model in **Figure 2.8**, and assume again that we wish to evaluate the effect of X on Y .

What variables should we condition on to obtain the correct effect?

The question boils down to finding a set of variables Z that satisfy the backdoor criterion, but since there are no backdoor paths from X to Y , the answer is trivial:

The empty set satisfies the criterion, hence no adjustment is needed.

- no node in Z is a descendant of X
- Z blocks every path between X and Y that contains an arrow into X

The answer is

$$P(Y = y | do(X = x)) = P(Y = y | X = x)$$

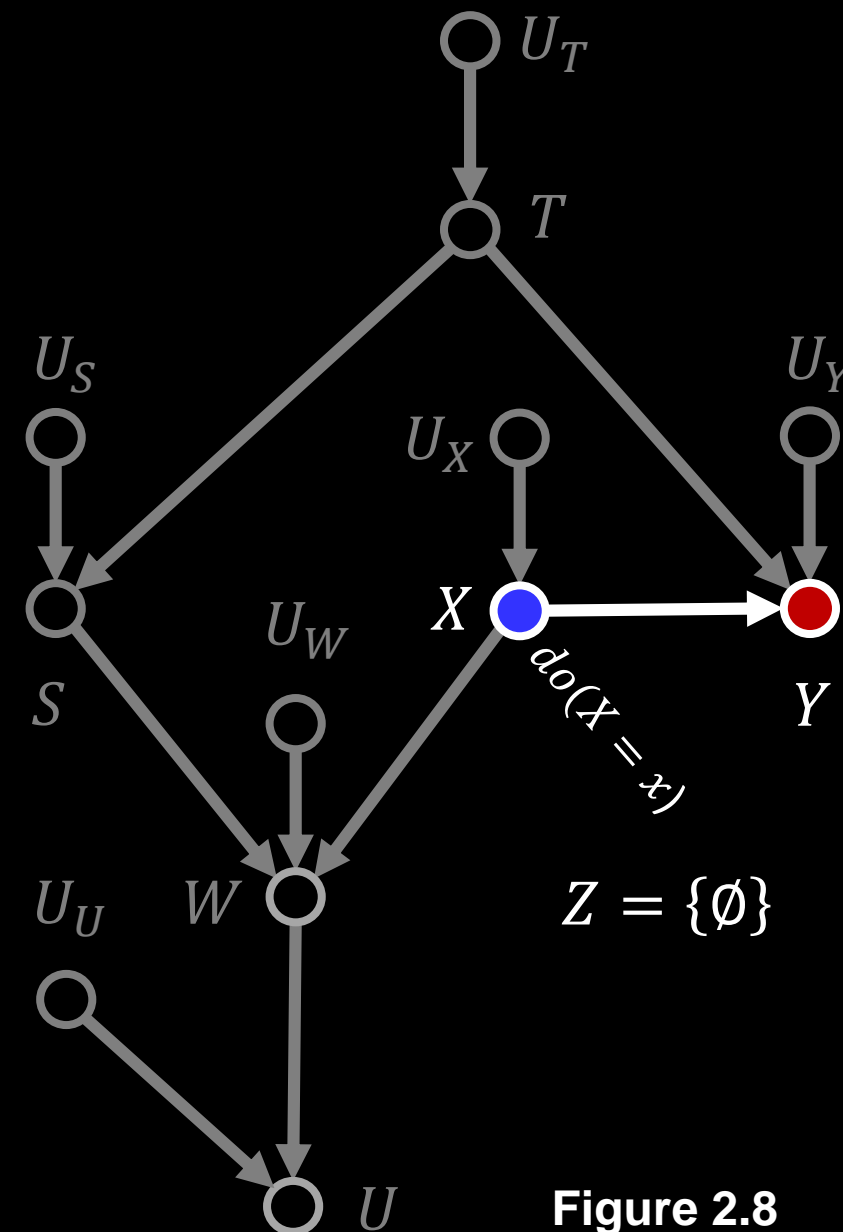


Figure 2.8

3.3 THE BACKDOOR CRITERION

Suppose, however, that we were to adjust for W .

Would we get the correct result for the effect of X on Y ?

Since W is a collider, conditioning on W would open the path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

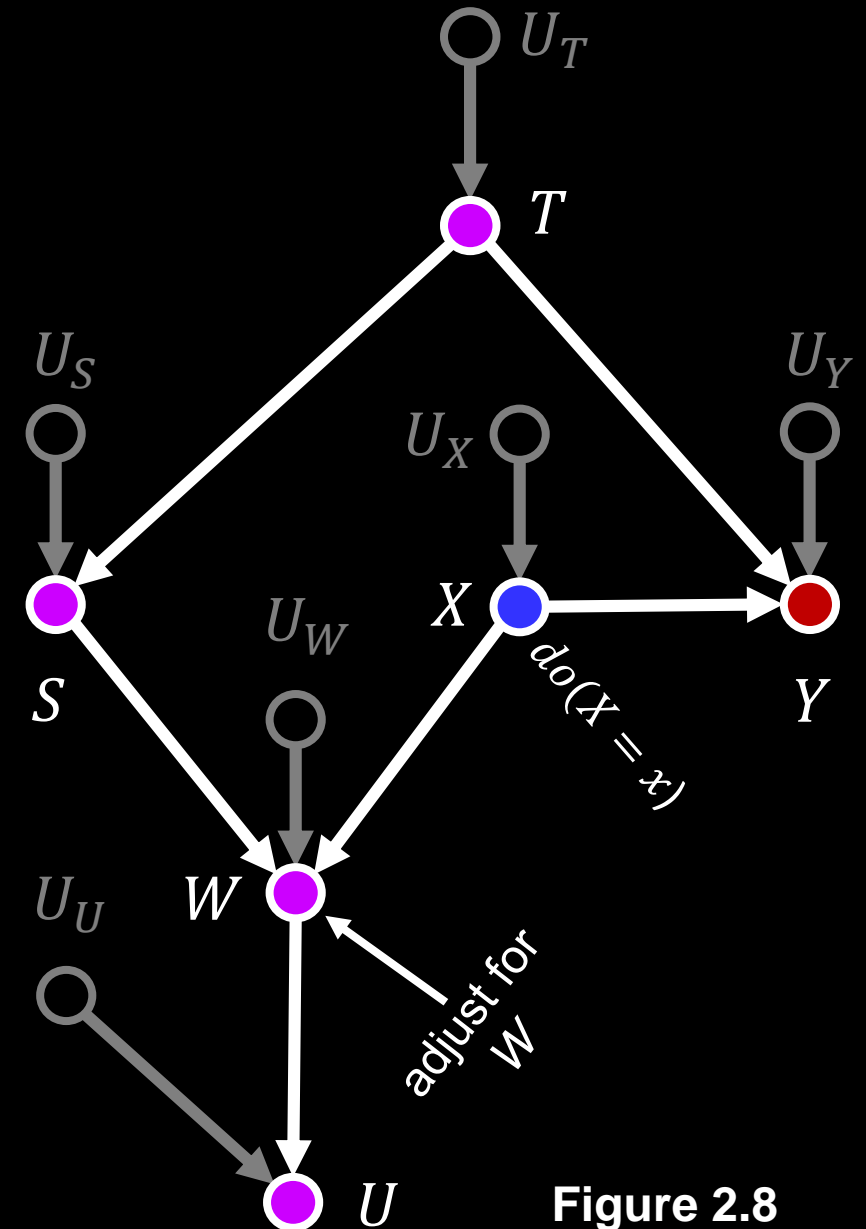


Figure 2.8

3.3 THE BACKDOOR CRITERION

Suppose, however, that we were to adjust for W .

Would we get the correct result for the effect of X on Y ?

Since W is a collider, conditioning on W would open the path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

This path is spurious since it lies outside the causal pathway from X to Y .

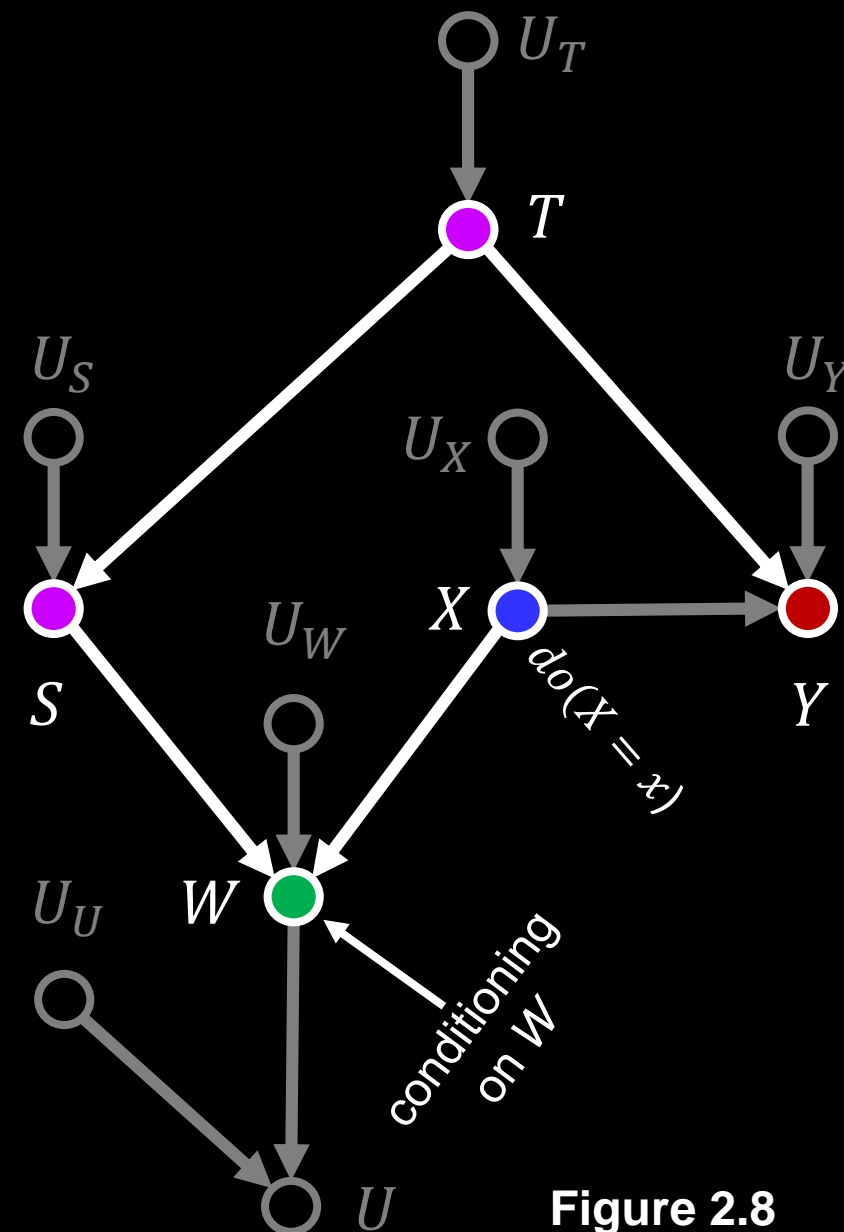


Figure 2.8

3.3 THE BACKDOOR CRITERION

Suppose, however, that we were to adjust for W .

Would we get the correct result for the effect of X on Y ?

Since W is a collider, conditioning on W would open the path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

This path is spurious since it lies outside the causal pathway from X to Y .

Opening this path will create bias and yield an erroneous answer.

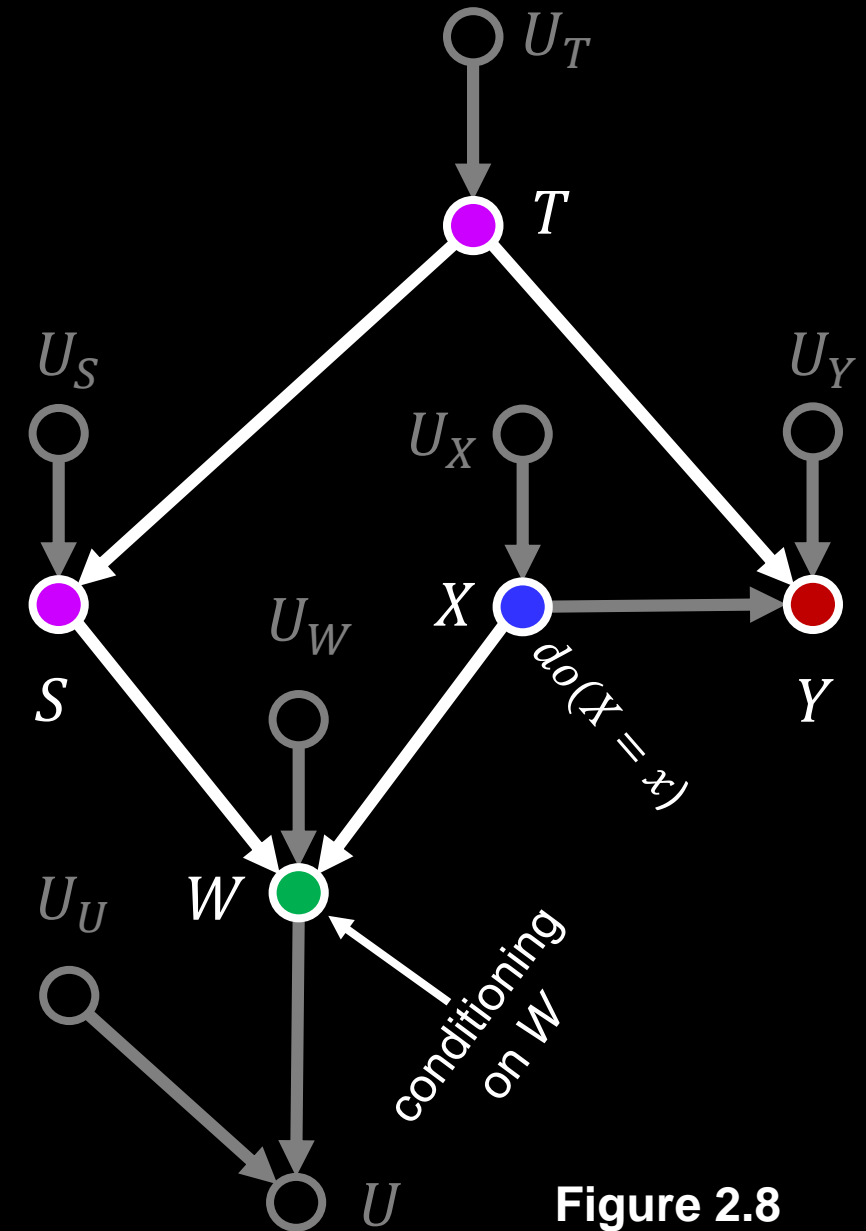


Figure 2.8

3.3 THE BACKDOOR CRITERION

Suppose, however, that we were to adjust for W .

Would we get the correct result for the effect of X on Y ?

Since W is a collider, conditioning on W would open the path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

This path is spurious since it lies outside the causal pathway from X to Y .

Opening this path will create bias and yield an erroneous answer.

This means that computing the association between X and Y for each value of W separately will not yield the correct effect of X on Y , and it might even give the wrong effect for each value of W .

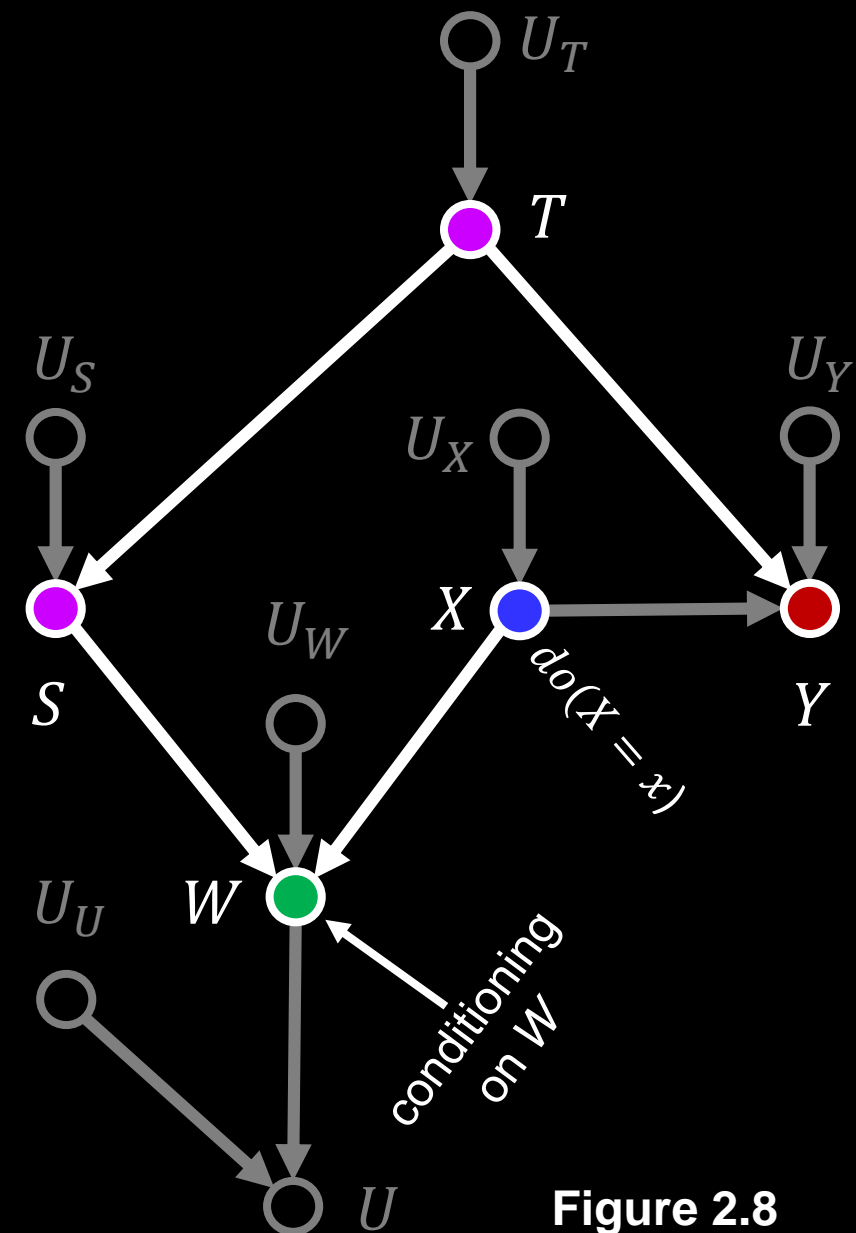


Figure 2.8

3.3 THE BACKDOOR CRITERION

How then do we compute the causal effect of X on Y for a specific value w of W ?

In **Figure 2.8** W may represent, for example, the level of posttreatment pain of a patient, and we might be interested in assessing the effect of X on Y for only those patients who did not suffer any pain.

Specifying the value of W amounts to conditioning on $W = w$, and this, as we have realized, opens a spurious path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

from X to Y by virtue of the fact that W is a collider.

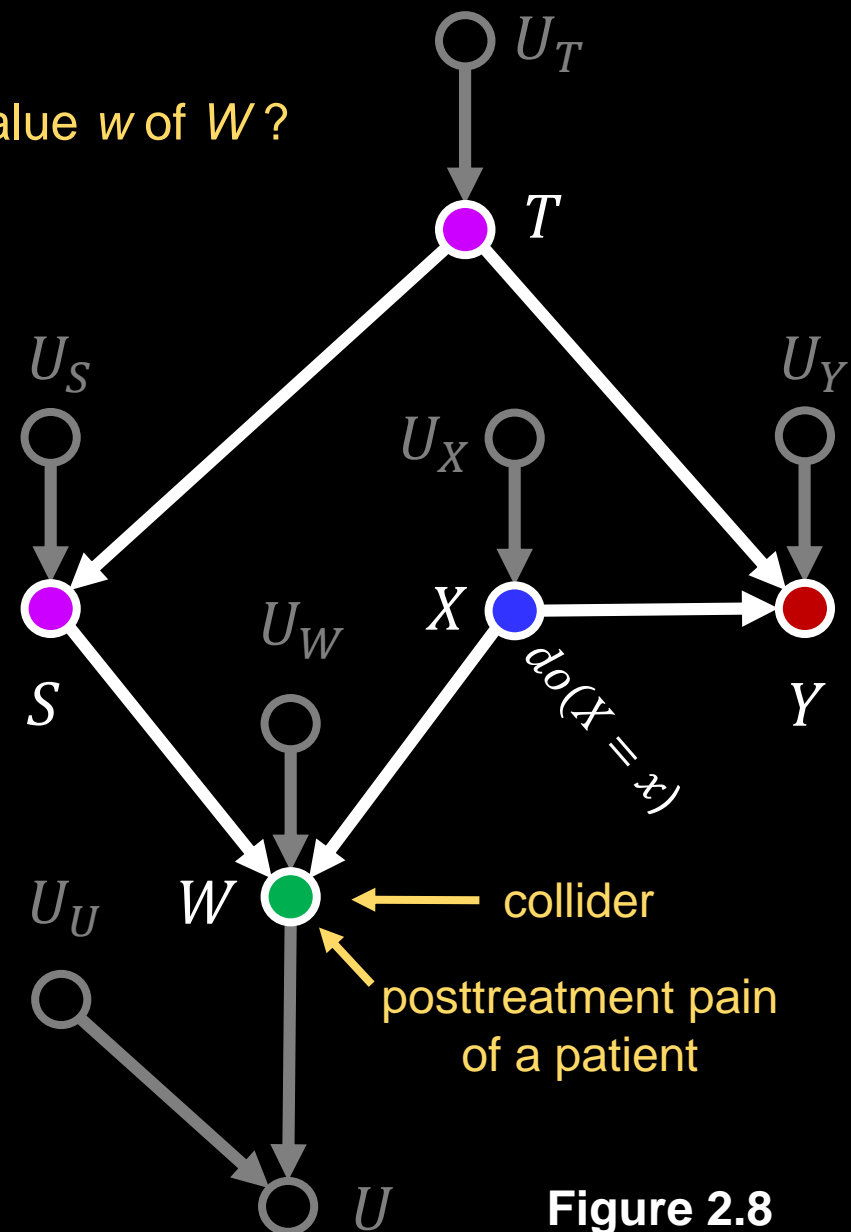


Figure 2.8

3.3 THE BACKDOOR CRITERION

How then do we compute the causal effect of X on Y for a specific value w of W ?

In **Figure 2.8** W may represent, for example, the level of posttreatment pain of a patient, and we might be interested in assessing the effect of X on Y for only those patients who did not suffer any pain.

Specifying the value of W amounts to conditioning on $W = w$, and this, as we have realized, opens a spurious path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

from X to Y by virtue of the fact that W is a collider.

The answer is that we still have the option of blocking that path using other variables.

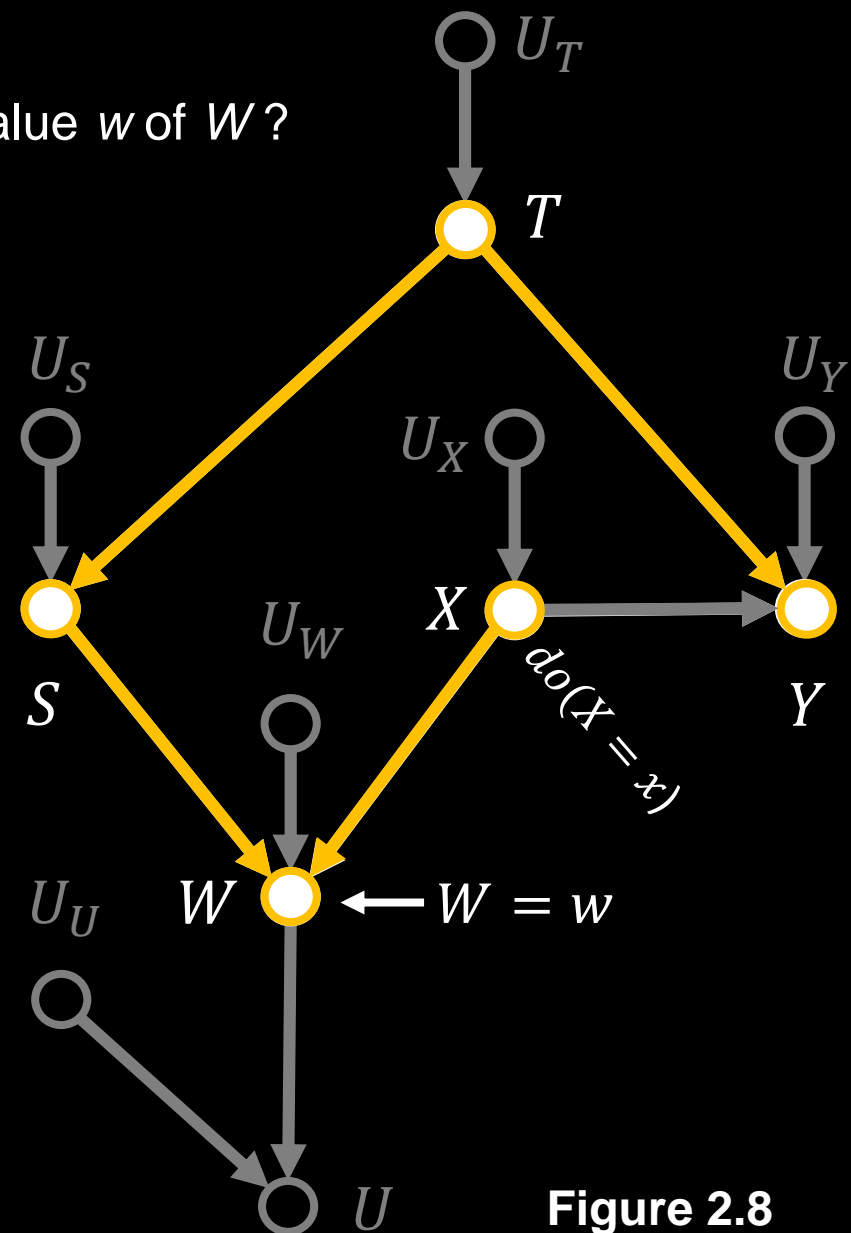


Figure 2.8

3.3 THE BACKDOOR CRITERION

How then do we compute the causal effect of X on Y for a specific value w of W ?

For example, if we condition on T , we would block the spurious path

$$X \rightarrow W \leftarrow S \leftarrow T \rightarrow Y$$

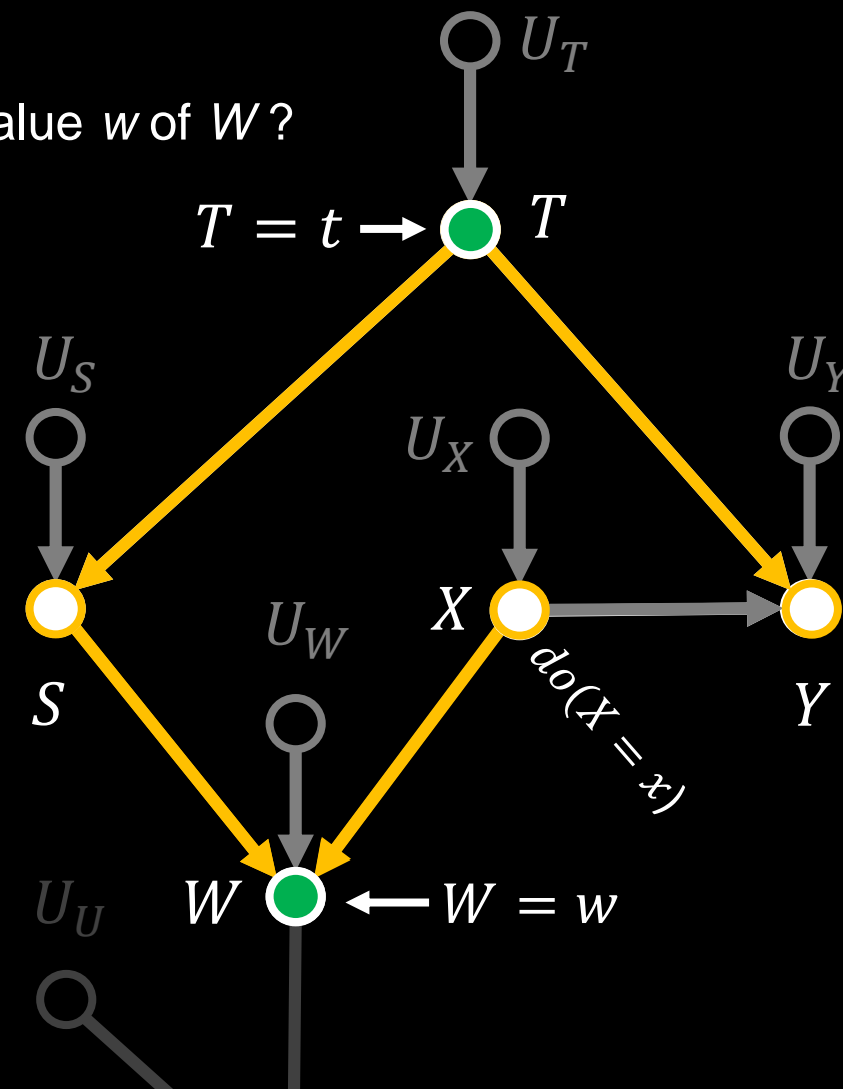
even if W is part of the conditioning set.

Thus to compute the **w-specific causal effect**, written

$$P(y|do(x), w)$$

we adjust for T , and obtain

$$P(Y = y|do(X = x), W = w) = \sum_t P(Y = y|X = x, W = w, T = t) P(T = t|X = x, W = w)$$



3.3 THE BACKDOOR CRITERION

$$P(Y = y|do(X = x), W = w) = \sum_t P(Y = y|X = x, W = w, T = t) P(T = t|X = x, W = w)$$

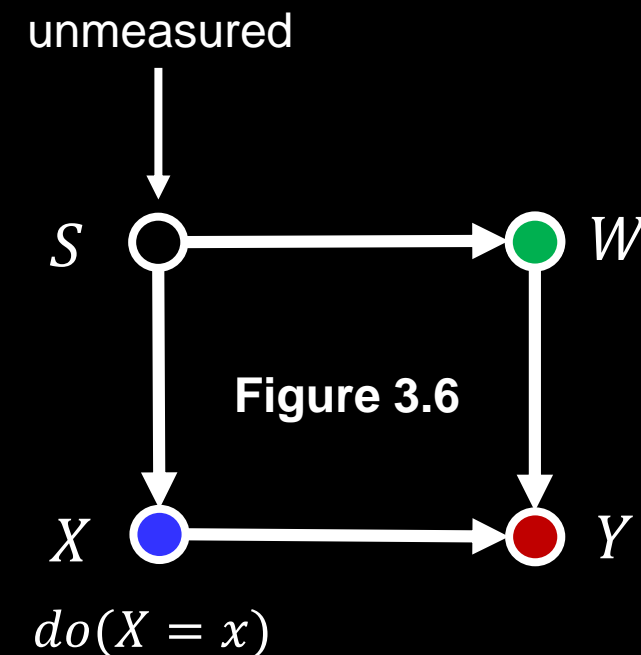
Computing such W -specific causal effects is an essential step in examining effect modification or moderation, that is, the degree to which the causal effect of X on Y is modified by different values of W .

Consider, again, the model in **Figure 3.6**, and suppose we wish to test whether the causal effect for units at level $W = w$ is the same as for units at level $W = w'$. (W may represent any pretreatment variable, such as age, sex, or ethnicity).

This question calls for comparing two causal effects,

$$P(Y = y|do(X = x), W = w)$$

$$P(Y = y|do(X = x), W = w')$$



3.3 THE BACKDOOR CRITERION

$$P(Y = y|do(X = x), W = w) = \sum_t P(Y = y|X = x, W = w, T = t) P(T = t|X = x, W = w)$$

In the specific example of **Figure 3.6**, the answer is simple, because W satisfies the backdoor criterion.

So, all we need to compare are the conditional probabilities

$$P(Y = y|X = x, W = w)$$

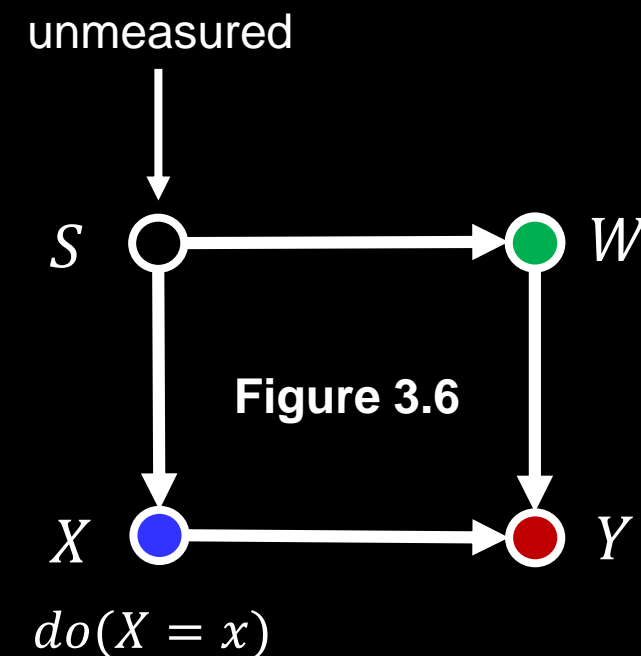
$$P(Y = y|X = x, W = w')$$

no summation is required.

In the more general case, where W alone does not satisfy the backdoor criterion, yet a larger set,

$$T \cup W$$

does, we need to adjust for members of T , which yields the formula on top.



3.3 THE BACKDOOR CRITERION

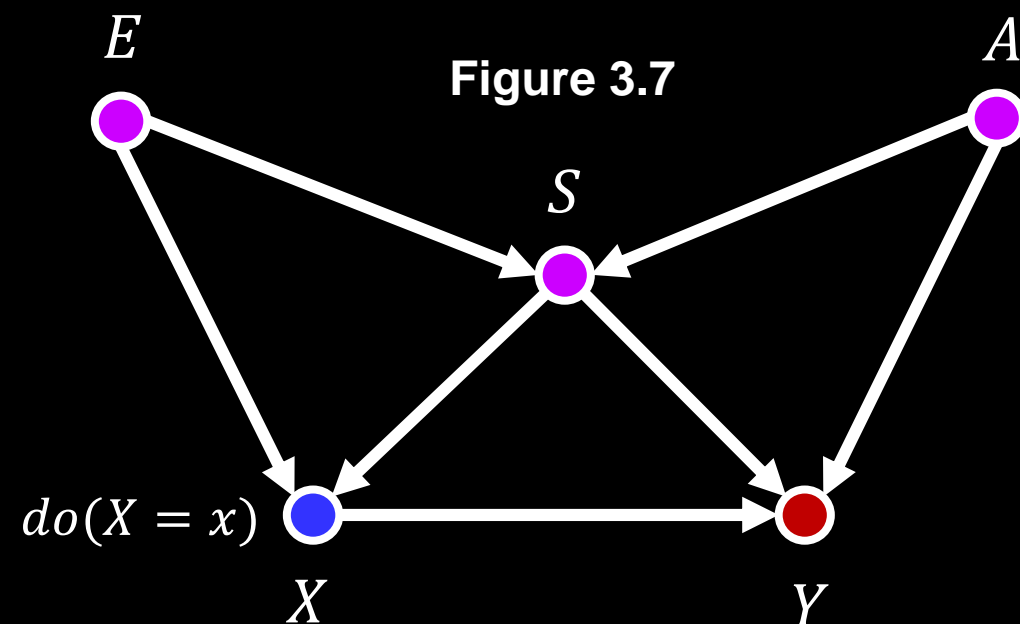
From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

Conditioning on S will unblock this path and will violate the backdoor criterion.



3.3 THE BACKDOOR CRITERION

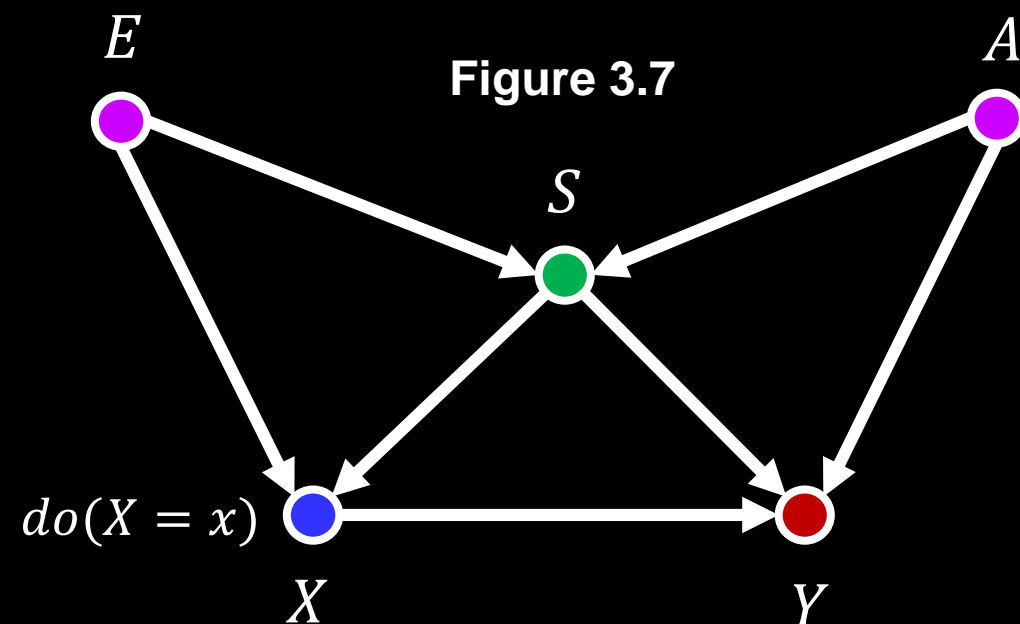
From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

Conditioning on S will unblock this path and will violate the backdoor criterion.



3.3 THE BACKDOOR CRITERION

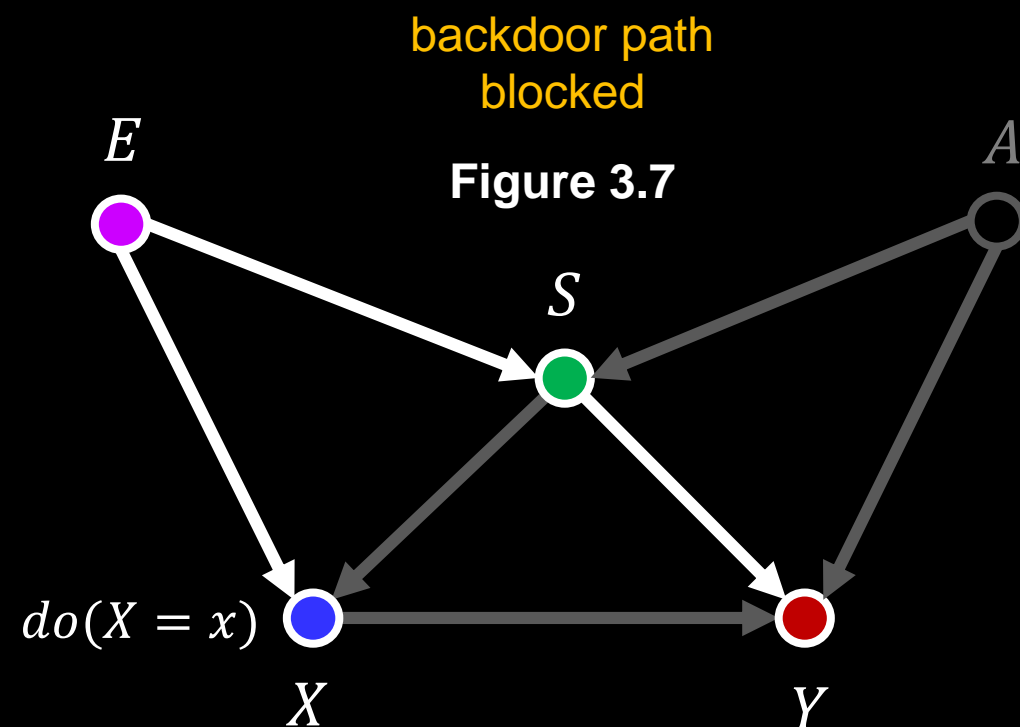
From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

Conditioning on S will unblock this path and will violate the backdoor criterion.



3.3 THE BACKDOOR CRITERION

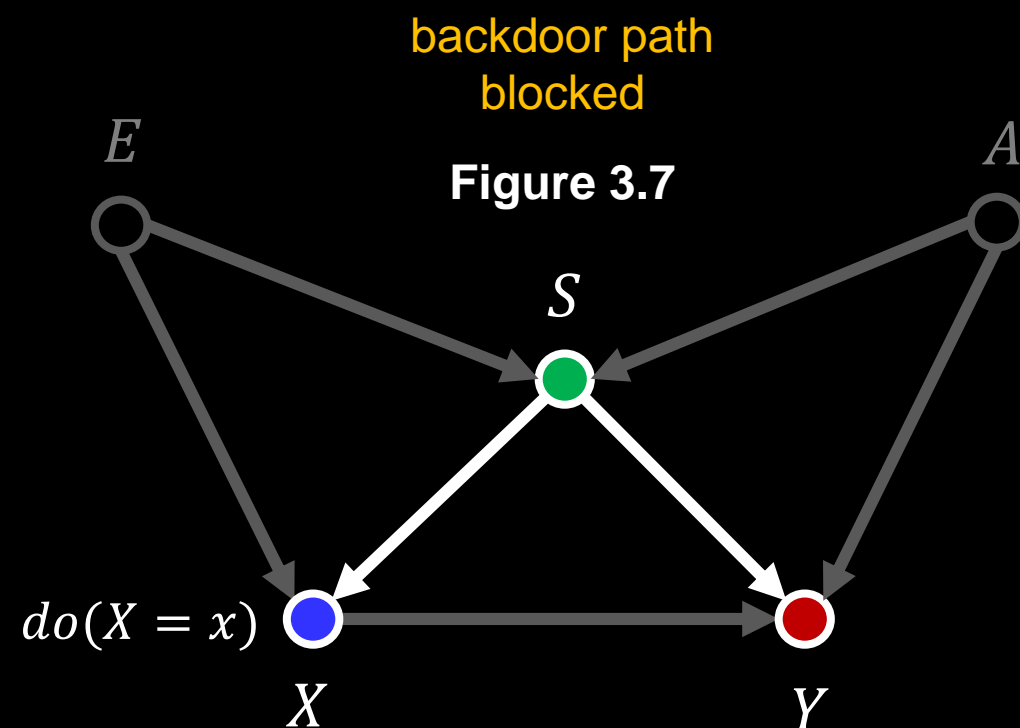
From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

Conditioning on S will unblock this path and will violate the backdoor criterion.



3.3 THE BACKDOOR CRITERION

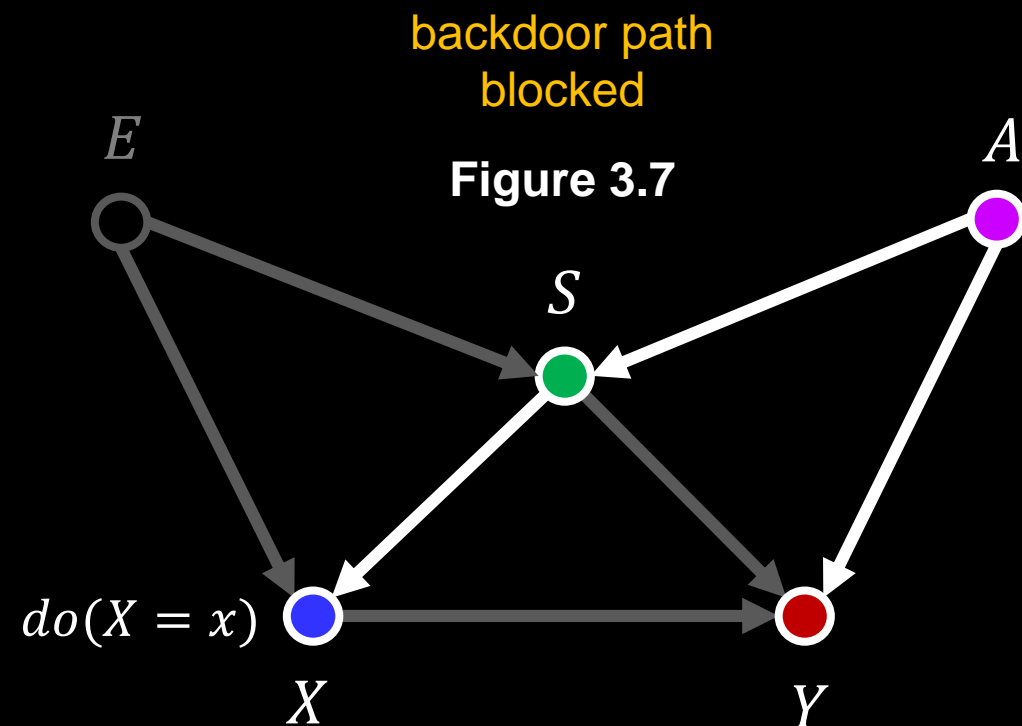
From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

Conditioning on S will unblock this path and will violate the backdoor criterion.



3.3 THE BACKDOOR CRITERION

From the examples seen thus far, readers may get the impression that one should refrain from adjusting for colliders.

Such adjustment is sometimes unavoidable, as seen in **Figure 3.7**.

Here, there are four backdoor paths from X to Y , all traversing variable S , which is a collider on the path

$$X \leftarrow E \rightarrow S \leftarrow A \rightarrow Y.$$

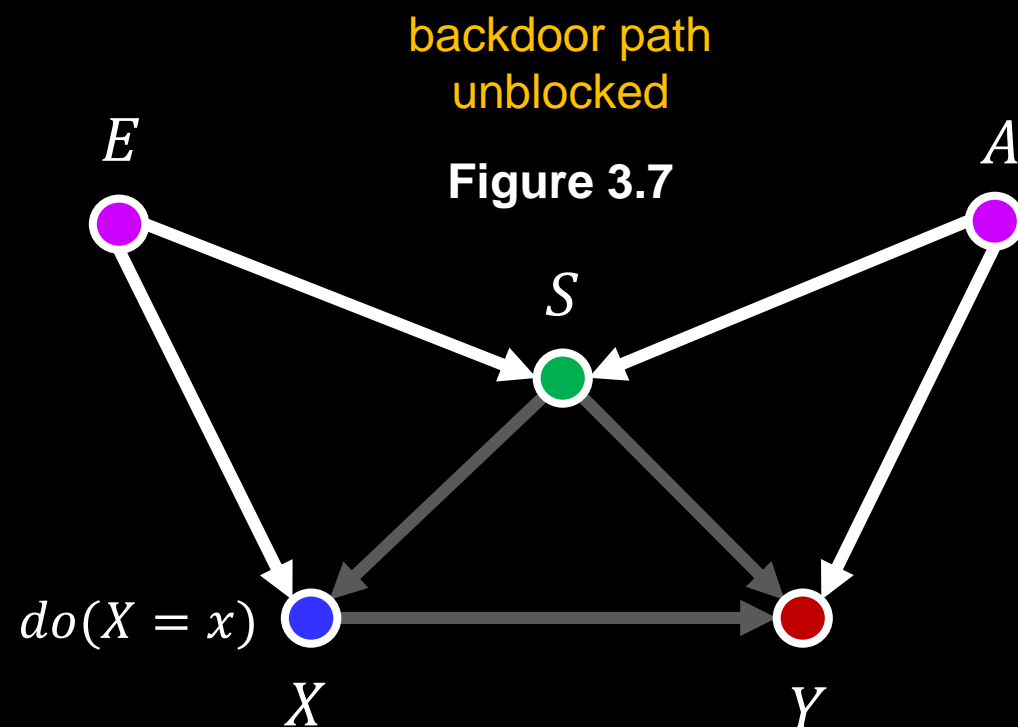
Conditioning on S will unblock this path and will violate the backdoor criterion.

To block all backdoor paths, we need to condition on one of the following sets:

$$\{E, S\}, \{A, S\}, \text{ or } \{E, S, A\}.$$

Each of these contains S .

We see, therefore, that S , a collider, must be adjusted for in any set that yields an unbiased estimate of the effect of X on Y .



3.3 THE BACKDOOR CRITERION

The backdoor criterion has some further possible benefits.

Consider the fact that

$$P(Y = y | do(X = x))$$

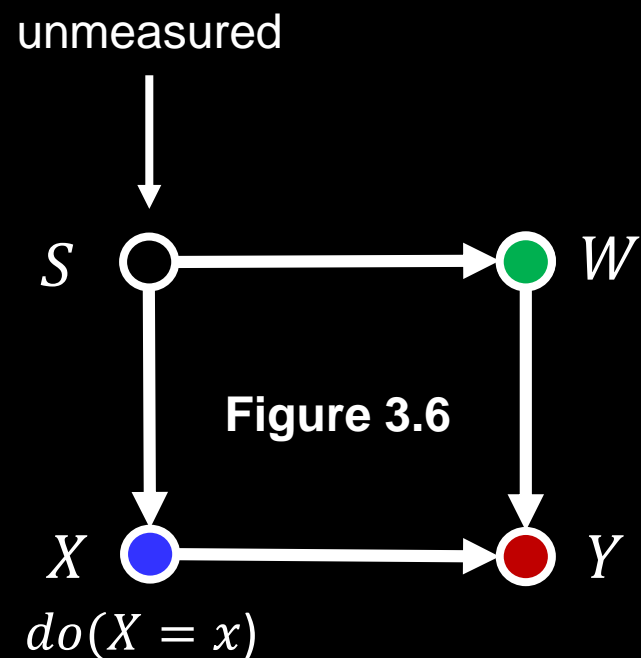
is an empirical fact of nature, not a byproduct of our analysis.

That means that any suitable variable or set of variables that we adjust on—whether it be $pa(X)$ or any other set that conforms to the backdoor criterion—must return the same result for

$$P(Y = y | do(X = x))$$

In the case we looked at in **Figure 3.6**, this means that

$$P(Y = y | X = x) = \sum_w P(Y = y | X = x, W = w) P(W = w) = \sum_s P(Y = y | X = x, S = s) P(S = s)$$



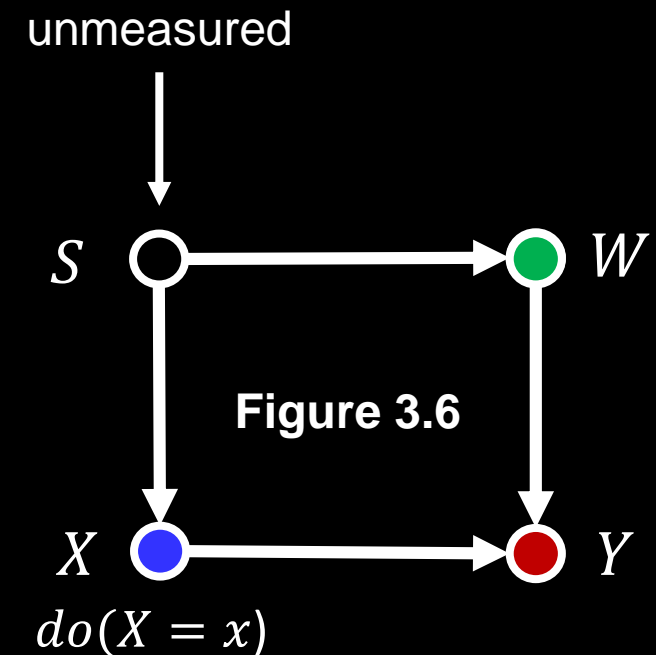
3.3 THE BACKDOOR CRITERION

The equality at the bottom of the slide is useful in two ways:

- in the cases where we have multiple observed sets of variables suitable for adjustment (e.g., in **Figure 3.6**, if both W and S had been observed), it provides us with a choice of which variables to adjust for.

This could be useful for any number of practical reasons—perhaps one set of variables is more expensive to measure than the other, or more prone to human error, or simply has more variables and is therefore more difficult to calculate.

- the equality constitutes a testable constraint on the data when all the adjustment variables are observed, much like the rules of d-separation. If we are attempting to fit a model that leads to such an equality on a data set that violates it, we can discard that model.



$$P(Y = y|X = x) = \sum_w P(Y = y|X = x, W = w) P(W = w) = \sum_s P(Y = y|X = x, S = s) P(S = s)$$

3.4 THE FRONT-DOOR CRITERION



The backdoor criterion provides us with a simple method of identifying sets of covariates that should be adjusted for when we seek to estimate causal effects from nonexperimental data.

It does not, however, exhaust all ways of estimating such effects.

The **do-operator** can be applied to graphical patterns that do not satisfy the backdoor criterion to identify effects that on first sight seem to be beyond one's reach.

One such pattern, called **front-door**, is now discussed.



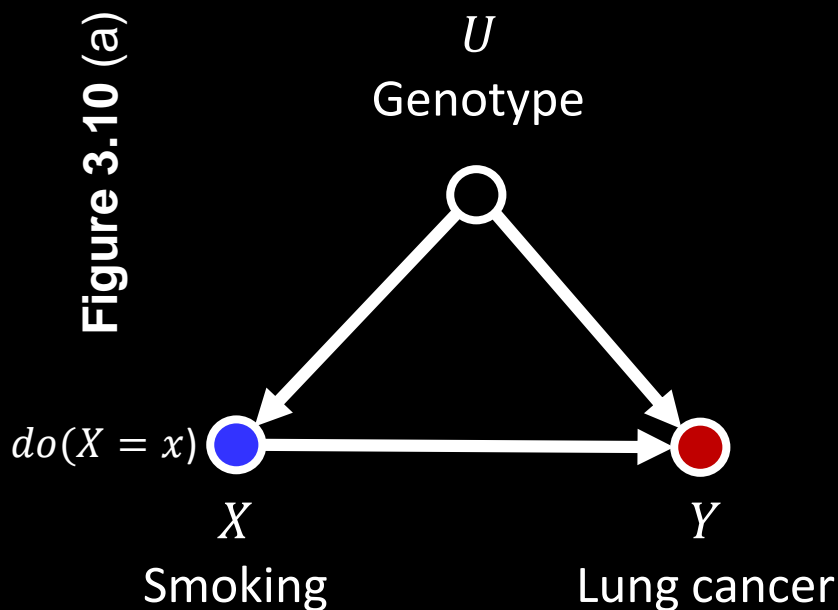
3.4 THE FRONT-DOOR CRITERION

Consider the century-old debate on the relation between smoking and lung cancer.

In the years preceding 1970, the tobacco industry has managed to prevent antismoking legislation by promoting the theory that the observed correlation between smoking and lung cancer could be explained by some sort of **carcinogenic genotype** that also induces an inborn craving for nicotine.



Figure 3.10 (a)



A graph depicting this example is shown in **Figure 3.10 (a)**.

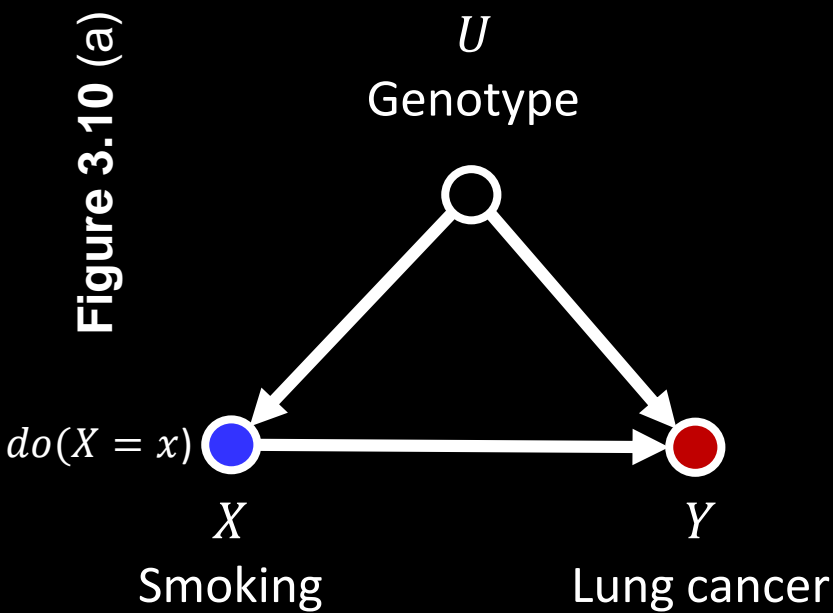
This graph does not satisfy the backdoor condition because the variable U is unobserved and hence cannot be used to block the backdoor path $(X \leftarrow U \rightarrow Y)$ from X to Y .

3.4 THE FRONT-DOOR CRITERION

The causal effect of smoking on lung cancer is not identifiable in this model; one can never ascertain which portion of the observed correlation between X and Y is spurious, attributable to their common effect, U , and what portion is genuinely causative.

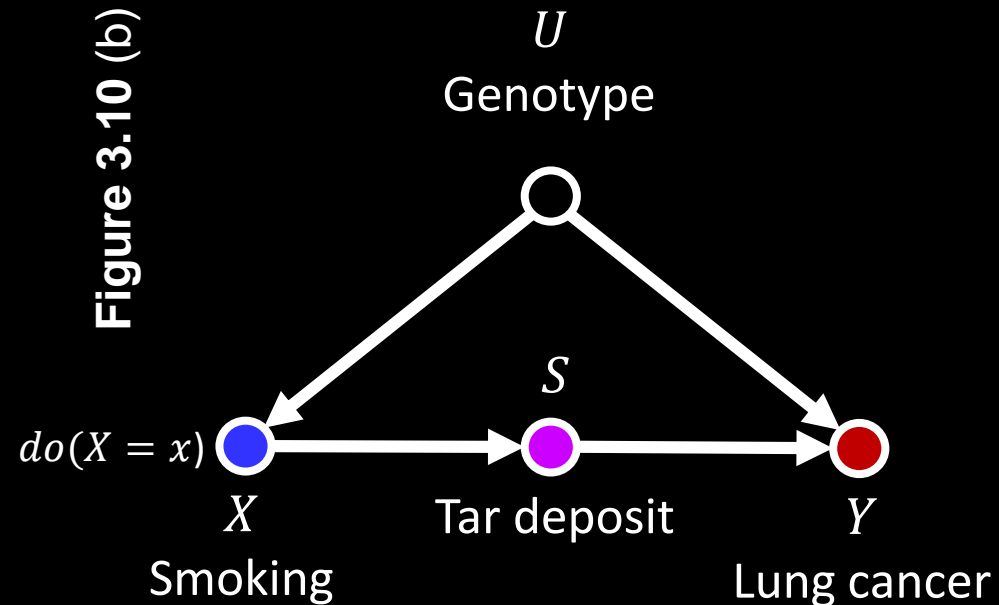
(We note, however, that even in these circumstances, much compelling work has been done to quantify how strong the (unobserved) associates between both U and X , and U and Y , must be in order to entirely explain the observed association between X and Y .)

Figure 3.10 (a)



However, we can go much further by considering the model in **Figure 3.10 (b)**, where an additional measurement is available: the amount of tar deposits in patients' lungs.

Figure 3.10 (b)



3.4 THE FRONT-DOOR CRITERION

The model in **Figure 3.10** (b) does not satisfy the backdoor criterion, because there is still no variable capable of blocking the spurious path

$$X \leftarrow U \rightarrow Y.$$

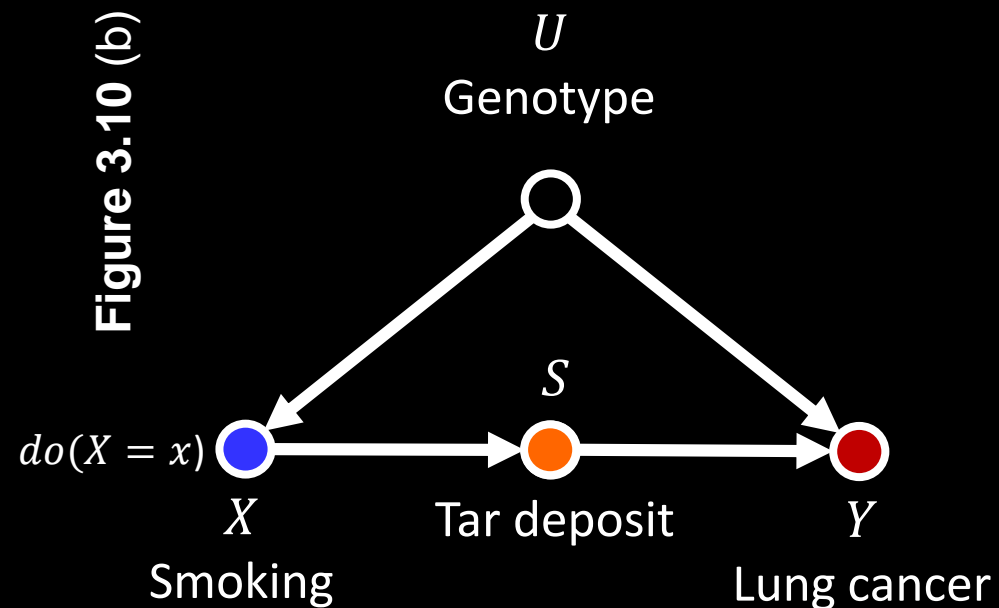
We see, however, that the causal effect of X on Y

$$P(Y = y | do(X = x))$$

is nevertheless identifiable in this model, through two consecutive applications of the backdoor criterion.

How can the intermediate variable S help us to assess the effect of X on Y ?

The answer is not at all trivial: as the following quantitative example shows, it may lead to heated debate.



3.4 THE FRONT-DOOR CRITERION

Assume that a careful study was undertaken, in which the following factors were measured simultaneously on a randomly selected sample of 800,000 subjects considered to be at very high risk of cancer (because of environmental exposures such as smoking, asbestos, radon, and the like).

1. Whether the subject smoked
2. Amount of tar in the subject's lungs
3. Whether lung cancer has been detected in the patient.

The data from this study are presented in **Table 3.1**, where, for simplicity, all three variables are assumed to be binary. All numbers are given in thousands.

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
	400		400		800	
	380	20	20	380	400	400
No cancer	323 85%	1 5%	18 90%	38 10%	341 85.25%	39 9.75%
Cancer	57 15%	19 95%	2 10%	342 90%	59 14.75%	361 90.25%

3.4 THE FRONT-DOOR CRITERION

Two opposing interpretations can be offered for these data.



The tobacco industry argues that the table proves the beneficial effect of smoking.

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	400	400	400	400	800	800
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
No cancer	380	20	20	380	400	400
	323	1	18	38	341	39
Cancer	85%	5%	90%	10%	85.25%	9.75%
	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

3.4 THE FRONT-DOOR CRITERION

Two opposing interpretations can be offered for these data.



The tobacco industry argues that the table proves the beneficial effect of smoking.

They point to the fact that only **14.75%** of the smokers have developed lung cancer, compared to **90.25%** of the nonsmokers.

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	400	400	400	400	800	800
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
No cancer	380	20	20	380	400	400
	323	1	18	38	341	39
Cancer	85%	5%	90%	10%	85.25%	9.75%
	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

3.4 THE FRONT-DOOR CRITERION

Two opposing interpretations can be offered for these data.



Moreover, within each of two subgroups, Tar and No Tar, smokers show a much lower percentage of cancer than nonsmokers. (These numbers are obviously contrary to empirical observations but well illustrate our point that observations are not to be trusted.)

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	400	400	400	400	800	800
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
No cancer	380	20	20	380	400	400
	323	1	18	38	341	39
	85%	5%	90%	10%	85.25%	9.75%
Cancer	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

3.4 THE FRONT-DOOR CRITERION

Two opposing interpretations can be offered for these data.



Moreover, within each of two subgroups, Tar and No Tar, smokers show a much lower percentage of cancer than nonsmokers. (These numbers are obviously contrary to empirical observations but well illustrate our point that observations are not to be trusted.)

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	400	400	400	400	800	800
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
No cancer	380	20	20	380	400	400
	323	1	18	38	341	39
Cancer	85%	5%	90%	10%	85.25%	9.75%
	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

3.4 THE FRONT-DOOR CRITERION



However, the antismoking lobbyists argue that the table tells an entirely different story—that smoking would actually increase, not decrease, one’s risk of lung cancer.

Their argument goes as follows:

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers 400		Non Smokers 400		All Subjects 800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
	380	20	20	380	400	400
No cancer	323 85%	18 90%	1 5%	38 10%	324 81.00%	56 14.00%
Cancer	57 15%	2 10%	19 95%	342 90%	76 19.00%	344 86.00%

3.4 THE FRONT-DOOR CRITERION



However, the antismoking lobbyists argue that the table tells an entirely different story—that smoking would actually increase, not decrease, one’s risk of lung cancer.

Their argument goes as follows:

- If you choose to smoke, then your chances of building up tar deposits are **95%** (380/400), compared to **5%** (20/400) if you choose not to smoke.

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers 400		Non Smokers 400		All Subjects 800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
	380	20	20	380	400	400
No cancer	323 85%	18 90%	1 5%	38 10%	324 81.00%	56 14.00%
Cancer	57 15%	2 10%	19 95%	342 90%	76 19.00%	344 86.00%

3.4 THE FRONT-DOOR CRITERION



However, the antismoking lobbyists argue that the table tells an entirely different story—that smoking would actually increase, not decrease, one’s risk of lung cancer.

Their argument goes as follows:

- If you choose to smoke, then your chances of building up tar deposits are 95% (380/400), compared to 5% (20/400) if you choose not to smoke.
- To evaluate the effect of tar deposits, we look separately at two groups, smokers and nonsmokers, as done in **Table 3.2**.

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers 400		Non Smokers 400		All Subjects 800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
	380	20	20	380	400	400
No cancer	323 85%	18 90%	1 5%	38 10%	324 81.00%	56 14.00%
Cancer	57 15%	2 10%	19 95%	342 90%	76 19.00%	344 86.00%

3.4 THE FRONT-DOOR CRITERION

It appears that tar deposits have a harmful effect in both groups;

- in smokers it increases cancer rates from 10% to 15%, and

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers		Non Smokers		All Subjects	
	400		400		800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
No cancer	380	20	20	380	400	400
	85%	90%	5%	10%	81.00%	14.00%
Cancer	57	2	19	342	76	344
	15%	10%	95%	90%	19.00%	86.00%

3.4 THE FRONT-DOOR CRITERION

It appears that tar deposits have a harmful effect in both groups;

- in smokers it increases cancer rates from 10% to 15%, and
- in nonsmokers it increases cancer rates from 90% to 95%.

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers		Non Smokers		All Subjects	
	400		400		800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
No cancer	380	20	20	380	400	400
	85%	90%	5%	10%	81.00%	14.00%
Cancer	57	2	19	342	76	344
	15%	10%	95%	90%	19.00%	86.00%

3.4 THE FRONT-DOOR CRITERION

It appears that tar deposits have a harmful effect in both groups;

- in smokers it increases cancer rates from 10% to 15%, and
- in nonsmokers it increases cancer rates from 90% to 95%.

Thus, regardless of whether I have a natural craving for nicotine, I should avoid the harmful effect of tar deposits, and no-smoking offers very effective means of avoiding them.

Table 3.2 Reorganization of the data set of Table 3.1 showing the percentage of cancer cases in each smoking-tar category (number in thousands)

	Smokers		Non Smokers		All Subjects	
	400		400		800	
	Tar	No Tar	Tar	No Tar	Tar	No Tar
	380	20	20	380	400	400
No cancer	323	18	1	38	324	56
	85%	90%	5%	10%	81.00%	14.00%
Cancer	57	2	19	342	76	344
	15%	10%	95%	90%	19.00%	86.00%

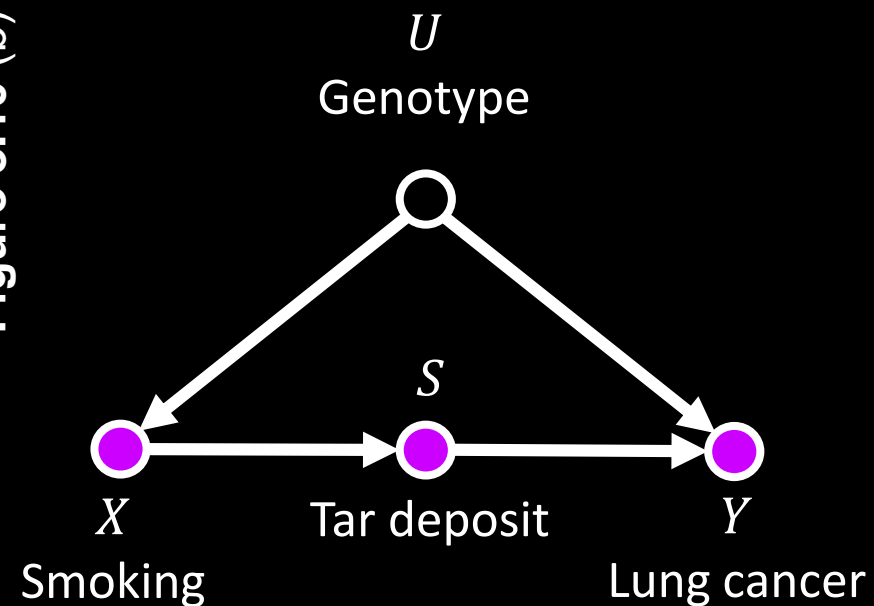
3.4 THE FRONT-DOOR CRITERION

The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Figure 3.10 (b)



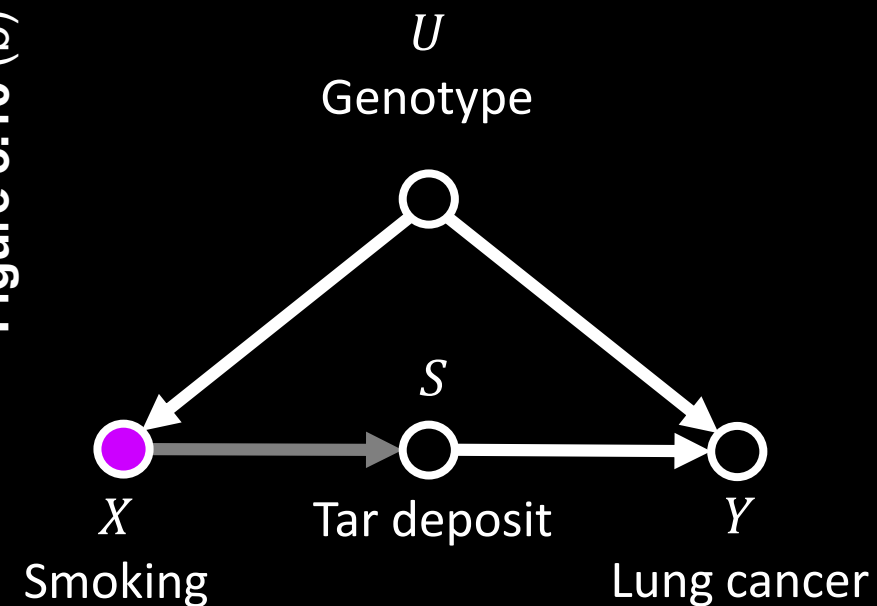
3.4 THE FRONT-DOOR CRITERION

The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Figure 3.10 (b)



3.4 THE FRONT-DOOR CRITERION

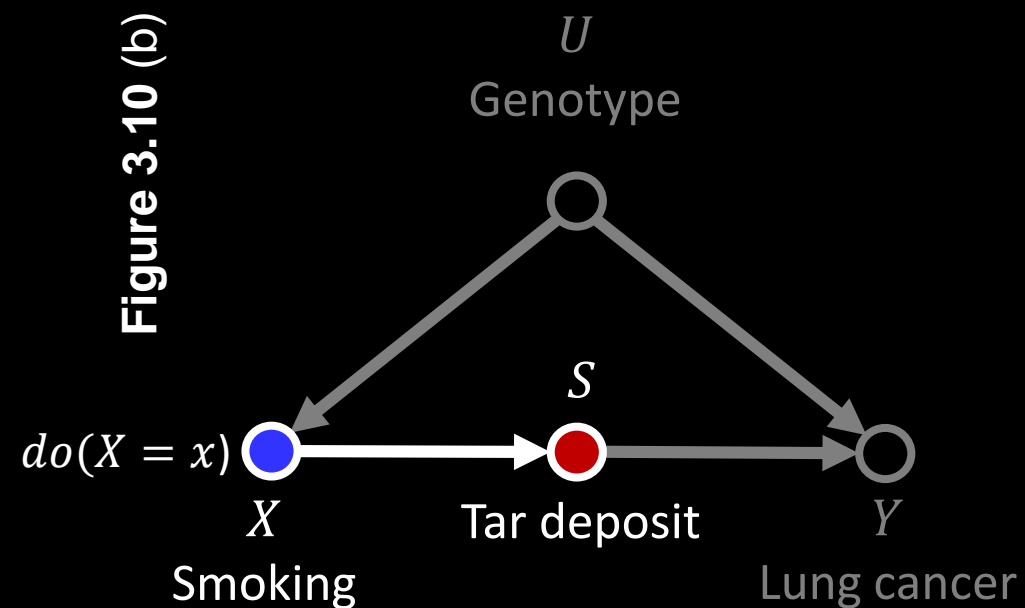
The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Thus, we can immediately write

$$P(S = s | do(X = x)) =$$



3.4 THE FRONT-DOOR CRITERION

The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

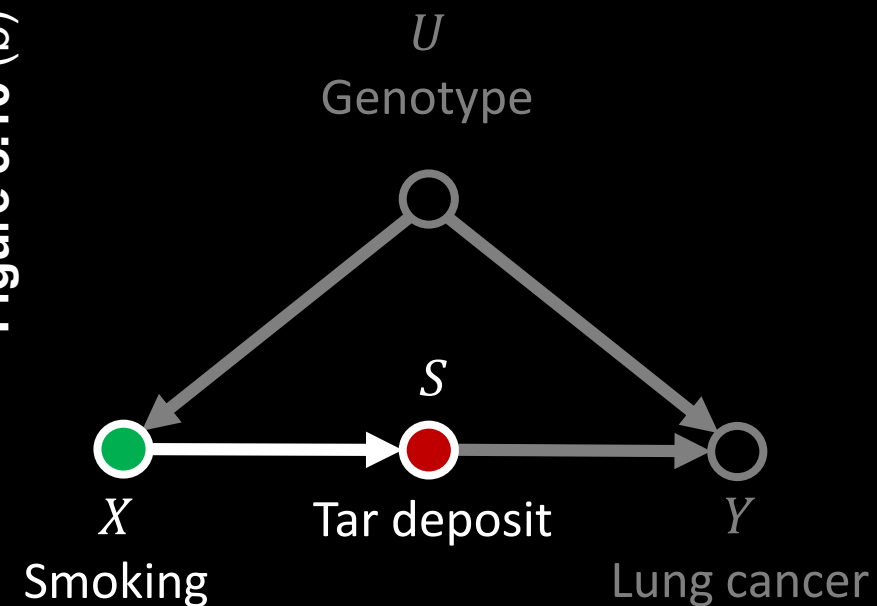
First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Thus, we can immediately write

$$P(S = s | do(X = x)) = P(S = s | X = x)$$

Figure 3.10 (b)



3.4 THE FRONT-DOOR CRITERION

The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Thus, we can immediately write

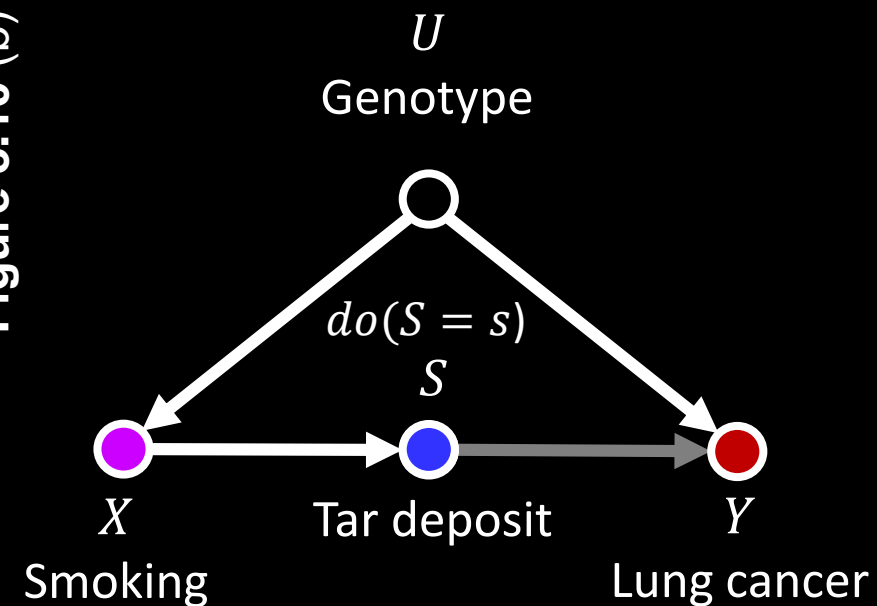
$$P(S = s | do(X = x)) = P(S = s | X = x)$$

Next we note that the effect of S on Y is also identifiable, since the backdoor path from S to Y , namely

$$S \leftarrow X \leftarrow U \rightarrow Y,$$

can be blocked by conditioning on X .

Figure 3.10 (b)



3.4 THE FRONT-DOOR CRITERION

The graph of **Figure 3.10** (b) enables us to decide between these two groups of statisticians.

First, we note that the effect of X on S is identifiable, since there is no unblocked backdoor path from X to S .

($Y \notin Z$)

Thus, we can immediately write

$$P(S = s | do(X = x)) = P(S = s | X = x)$$

Next we note that the effect of S on Y is also identifiable, since the backdoor path from S to Y , namely

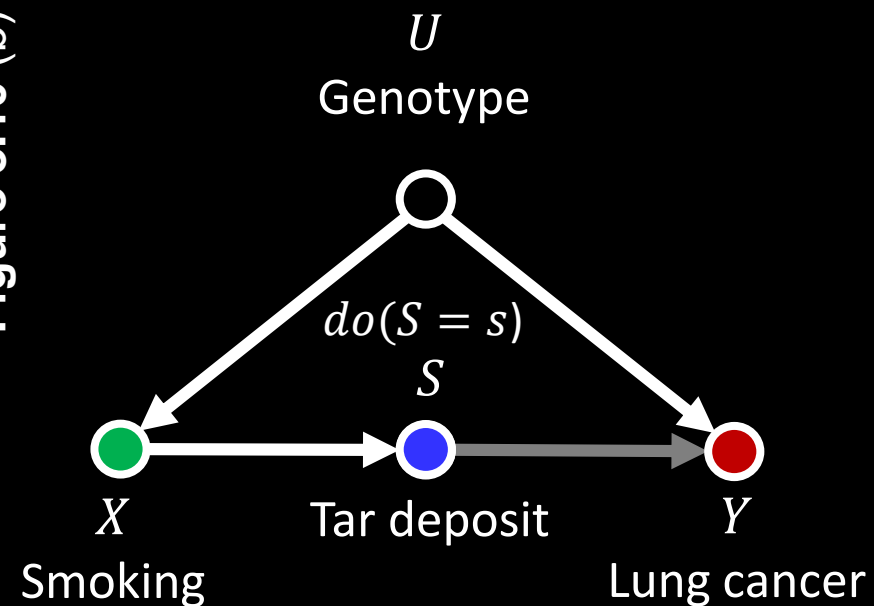
$$S \leftarrow X \leftarrow U \rightarrow Y,$$

can be blocked by conditioning on X .

Thus, we can write

$$P(Y = y | do(S = s)) = \sum_{x'} P(Y = y | S = s, X = x') P(X = x')$$

Figure 3.10 (b)



3.4 THE FRONT-DOOR CRITERION

Both $P(S = s|do(X = x)) = P(S = s|X = x)$

and $P(Y = y|do(S = s)) = \sum_{x'} P(Y = y|S = s, X = x') P(X = x')$

are obtained through the adjustment formula, the first by conditioning on the null set, and the second by adjusting for X .

We are now going to chain together the two partial effects to obtain the overall effect of X on Y .

3.4 THE FRONT-DOOR CRITERION

Both $P(S = s|do(X = x)) = P(S = s|X = x)$

and $P(Y = y|do(S = s)) = \sum_{x'} P(Y = y|S = s, X = x') P(X = x')$

are obtained through the adjustment formula, the first by conditioning on the null set, and the second by adjusting for X .

We are now going to chain together the two partial effects to obtain the overall effect of X on Y .

The reasoning goes as follows:

- If nature chooses to assign S the value s , then the probability of Y would be $P(Y = y|do(S = s))$.
- But the probability that nature would choose to do that (to set $S = s$), given that we choose to set X at x , is $P(S = s|do(X = x))$.

Therefore, summing over all states s of S , we have

$$P(Y = y|do(X = x)) = \sum_s P(Y = y|do(S = s)) P(S = s|do(X = x))$$

3.4 THE FRONT-DOOR CRITERION

$$P(Y = y | do(S = s)) = \sum_{x'} P(Y = y | S = s, X = x') P(X = x')$$

$$P(S = s | do(X = x)) = P(S = s | X = x)$$

By evaluating the terms
on the right-hand side of

$$P(Y = y | do(X = x)) = \sum_s P(Y = y | do(S = s)) P(S = s | do(X = x))$$

$$P(Y = y | do(X = x)) = \sum_s \boxed{}$$

$$P(Y = y|do(S = s)) = \sum_{x'} P(Y = y|S = s, X = x') P(X = x')$$

$$P(S = s|do(X = x)) = P(S = s|X = x)$$

By evaluating the terms
on the right-hand side of

$$P(Y = y|do(X = x)) = \sum_s P(Y = y|do(S = s)) P(S = s|do(X = x))$$

we obtain the following **do-free expression**

$$P(Y = y|do(X = x)) = \sum_s \sum_{x'} P(Y = y|S = s, X = x') P(X = x') P(S = s|X = x)$$

Front-Door Formula

3.4 THE FRONT-DOOR CRITERION

Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar 400		No Tar 400		All Subjects 800	
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
No cancer	380	20	20	380	400	400
	323	1	18	38	341	39
	85%	5%	90%	10%	85.25%	9.75%
Cancer	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

Applying this formula to the data in **Table 3.1**, we see that the tobacco industry was wrong;

- tar deposits have a harmful effect in that they make lung cancer more likely and smoking, by increasing tar deposits, increases the chances of causing this harm.

The data in **Table 3.1** are obviously unrealistic and were deliberately crafted so as to surprise readers with counterintuitive conclusions that may emerge from naive analysis of observational data.

In reality, we would expect observational studies to show positive correlation between smoking and lung cancer.

3.4 THE FRONT-DOOR CRITERION

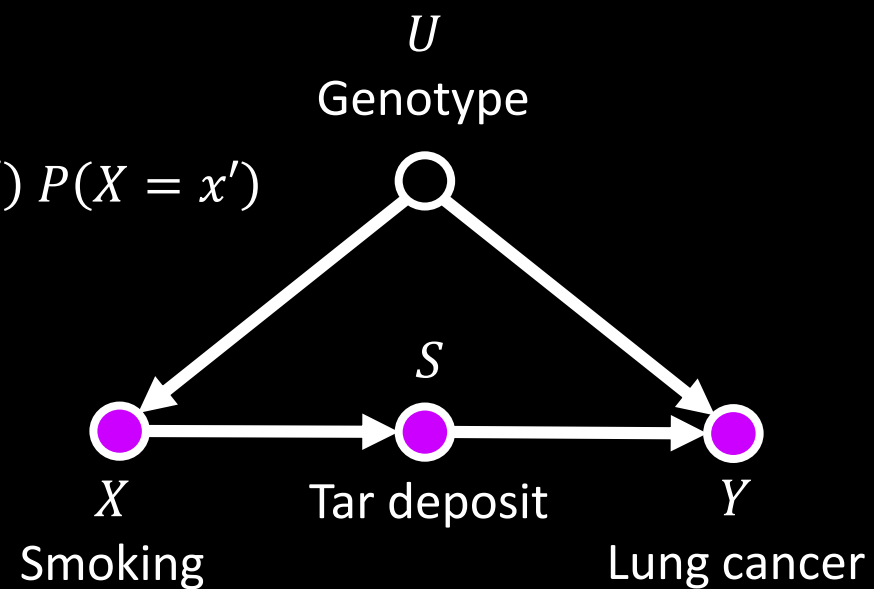
Table 3.1 A hypothetical data set of randomly selected samples showing the percentage of cancer cases for smokers and nonsmokers in each tar category (numbers in thousands)

	Tar		No Tar		All Subjects	
	400		400		800	
	Smokers	Non Smokers	Smokers	Non Smokers	Smokers	Non Smokers
	380	20	20	380	400	400
No cancer	323	1	18	38	341	39
	85%	5%	90%	10%	85.25%	9.75%
Cancer	57	19	2	342	59	361
	15%	95%	10%	90%	14.75%	90.25%

The estimand

$$P(Y = y|do(X = x)) = \sum_s P(S = s|X = x) \sum_{x'} P(Y = y|S = s, X = x') P(X = x')$$

could then be used for confirming and quantifying the harmful effect of smoking on cancer.



3.4 THE FRONT-DOOR CRITERION

The preceding analysis can be generalized to structures, where multiple paths lead from X to Y .

Definition 3.4.1 (Front-Door)

A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if

1. Z intercepts all directed paths from X to Y .
2. There is no unblocked backdoor path from X to Z .
3. All backdoor paths from Z to Y are blocked by X .

Theorem 3.4.1 (Front-Door Adjustment)

If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

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



3.4 THE FRONT-DOOR CRITERION

The preceding analysis can be generalized to structures, where multiple paths lead from X to Y .

Definition 3.4.1 (Front-Door)

A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if

1. Z intercepts all directed paths from X to Y .
2. There is no unblocked backdoor path from X to Z .
3. All backdoor paths from Z to Y are blocked by X .

The combination of the adjustment formula, the backdoor criterion, and the front-door criterion covers numerous scenarios.

It proves the enormous, even revelatory, power that causal graphs have in not merely representing, but actually discovering causal information.



3.6 INVERSE PROBABILITY WEIGHING



By now, the astute reader may have noticed a problem with our intervention procedures.

The **backdoor and front-door criteria** tell us whether it is possible to predict the results of hypothetical interventions from data obtained in an observational study.



Moreover, they tell us that we can make this prediction without simulating the intervention and without even thinking about it.

All we need to do is identify a set Z of covariates satisfying one of the criteria, plug this set into the adjustment formula, and we're done: the resulting expression is guaranteed to provide a valid prediction of how the intervention will affect the outcome.

3.6 INVERSE PROBABILITY WEIGHING

This is lovely in theory, but in practice, adjusting for Z may prove problematic.

It entails looking at each value or combination of values of Z separately, estimating the conditional probability of Y given X in that stratum and then averaging the results.

As the number of strata increases, adjusting for Z will encounter both computational and estimational difficulties.

Since the set Z can be comprised of dozens of variables, each spanning dozens of discrete values, the summation required by the adjustment formula may be formidable, and the number of data samples falling within each $Z = z$ cell may be too small to provide reliable estimates of the conditional probabilities involved.



Curse of Dimensionality

SMALL SAMPLE SIZE

Assuming that the function $P(X = x|Z = z)$ is available to us, we can use it to

- generate artificial samples that act as though they were drawn from the postintervention probability P_m , rather than $P(x, y, z)$.
- Once we obtain such fictitious samples, we can evaluate $P(Y = y|do(x))$ by simply counting the frequency of the event $Y = y$, for each stratum $X = x$ in the sample.

In this way, we skip the labor associated with summing over all strata $Z = z$; we essentially let nature do the summation for us.

The idea of estimating probabilities using fictitious samples is not new to us; it was used all along, though implicitly, whenever we estimated conditional probabilities from finite samples.

3.6 INVERSE PROBABILITY WEIGHING

In **Part 1**, we characterized conditioning as a process of filtering—that is, ignoring all cases for which the condition $X = x$ does not hold, and normalizing the surviving cases, so that their total probabilities would add up to one.

The net result of this operation is that the probability of each case such that **Age < 45**

is boosted by a factor

$$\frac{1}{P(X = x)}$$



Age of U.S. voters in the 2012 presidential election.

Age < 45

Filtering Table 1.3 by Age < 45

TABLE 1.3 Age breakdown of voters in 2012 election (all numbers in thousands)

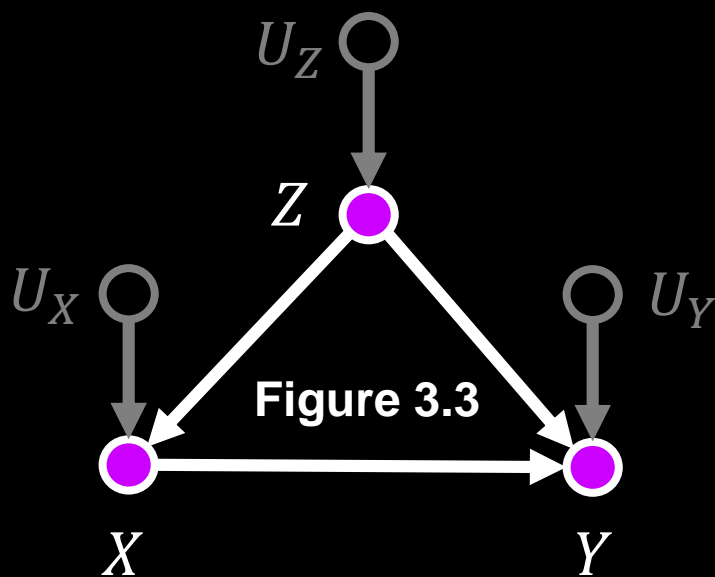
Age Group	# of voters
18-29	20,539
30-44	30,756
	132,949

3.6 INVERSE PROBABILITY WEIGHING

In **Part 1**, we characterized conditioning as a process of filtering—that is, ignoring all cases for which the condition $X = x$ does not hold, and normalizing the surviving cases, so that their total probabilities would add up to one.

The net result of this operation is that the probability of each case such that **Age < 45** is boosted by a factor

$$\frac{1}{P(X = x)}$$



This can be seen directly from **Bayes' rule**, which tells us that

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

In other words, to find the probability of each row in the surviving table, we multiply the unconditional probability,

$$P(Y = y, Z = z, X = x)$$

by the constant

$$\frac{1}{P(X = x)}$$

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = x)$ operation and ask how the probability of each case changes as a result of this operation. The answer is given to us by the **adjustment formula**, which reads

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

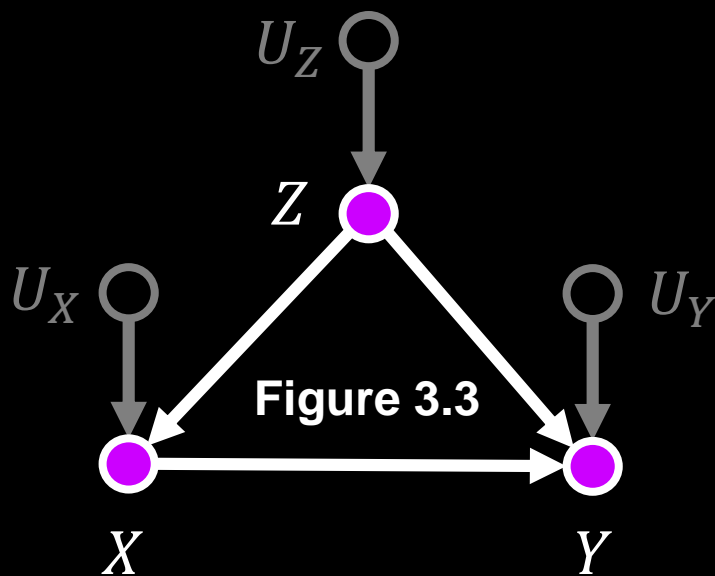
Multiplying and dividing the expression inside the sum by the propensity score $1/P(X = x|Z = z)$, we get

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

Upon realizing the numerator is none other but the pretreatment distribution of (X, Y, Z) , we can write

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

and the answer becomes clear: each case $(Y = y, X = x, Z = z)$ in the population should boost its probability by a factor equals to $1/P(X = x|Z = z)$. (Hence the name “**inverse probability weighing.**”)



3.6 INVERSE PROBABILITY WEIGHING

This provides us with a simple procedure of estimating $P(Y = y|do(X = x))$ when we have finite samples.

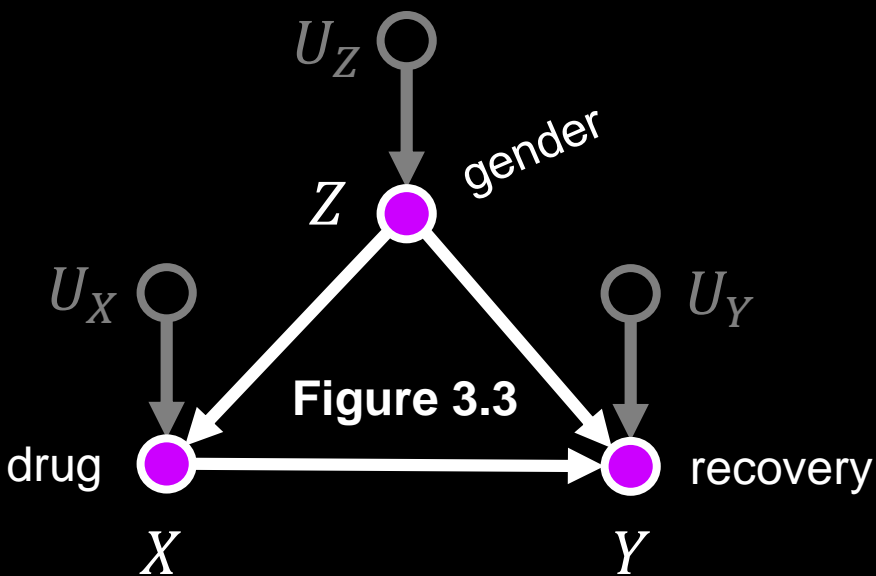
If we weigh each available sample by a factor $= 1/P(X = x|Z = z)$, we can then treat the reweighted samples as if they were generated from P_m , not P , and proceed to estimate $P(Y = y|do(x))$ accordingly.

This is best demonstrated in an example.

Table 3.3 returns to our **Simpson's paradox** example of the drug that seems to help men and women but to hurt the general population.

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036



3.6 INVERSE PROBABILITY WEIGHING

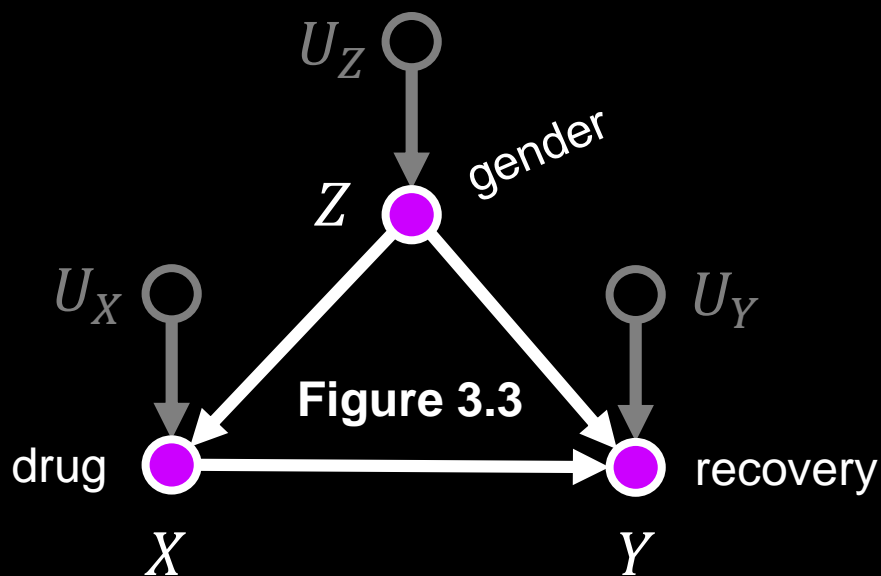
We'll use the same data we used before but presented this time as a weighted table.

In this case,

- X represents whether or not the patient took the drug,
- Y represents whether the patient recovered, and
- Z represents the patient's gender.

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036



3.6 INVERSE PROBABILITY WEIGHING

If we condition on “ $X = \text{yes}$,” we get the data set shown in **Table 3.4**, which was formed in two steps.

- all rows with $X = \text{no}$ were excluded.
- the weights given to the remaining rows were “renormalized,” that is, multiplied by a constant so as to make them sum to one.

Table 3.4 Conditional probability distribution $P(Y, Z | X)$ for drug users ($X = \text{yes}$) in the population of Table 3.3

X	Y	Z	% population
yes	yes	male	0.232
yes	yes	female	0.548
yes	no	male	0.018
yes	no	female	0.202


$$\text{boosting factor } \frac{1}{P(X = \text{yes})} = 2$$


Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101

$$P(X = \text{yes}) = 0.500$$

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = \text{yes})$ operation, representing a deliberate decision to administer the drug to the same population.

To calculate the distribution of weights in this population, we need to compute the factor $P(X = \text{yes} | Z = z)$ for each z , which, according to **Table 3.3**, is given by

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = yes)$ operation, representing a deliberate decision to administer the drug to the same population.

To calculate the distribution of weights in this population, we need to compute the factor $P(X = yes|Z = z)$ for each z , which, according to **Table 3.3**, is given by

$$P(X = yes|Z = male) = \frac{0.116 + 0.009}{0.116 + 0.009 + 0.334 + 0.051} = 0.245$$

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	no	male	0.009
no	yes	male	0.334
no	no	male	0.051

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = yes)$ operation, representing a deliberate decision to administer the drug to the same population.

To calculate the distribution of weights in this population, we need to compute the factor $P(X = yes|Z = z)$ for each z , which, according to **Table 3.3**, is given by

$$P(X = yes|Z = male) = \frac{0.116 + 0.009}{0.116 + 0.009 + 0.334 + 0.051} = 0.245$$

$$P(X = yes|Z = female) =$$

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = yes)$ operation, representing a deliberate decision to administer the drug to the same population.

To calculate the distribution of weights in this population, we need to compute the factor $P(X = yes|Z = z)$ for each z , which, according to **Table 3.3**, is given by

$$P(X = yes|Z = male) = \frac{0.116 + 0.009}{0.116 + 0.009 + 0.334 + 0.051} = 0.245$$

$$P(X = yes|Z = female) = \frac{0.274 + 0.101}{0.274 + 0.101 + 0.079 + 0.036} = 0.765$$

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	female	0.274
yes	no	female	0.101
no	yes	female	0.079
no	no	female	0.036

3.6 INVERSE PROBABILITY WEIGHING

Let us now examine the population created by the $do(X = yes)$ operation, representing a deliberate decision to administer the drug to the same population.

To calculate the distribution of weights in this population, we need to compute the factor $P(X = yes|Z = z)$ for each z , which, according to **Table 3.3**, is given by

$$P(X = yes|Z = male) = \frac{0.116 + 0.009}{0.116 + 0.009 + 0.334 + 0.051} = 0.245$$

$$P(X = yes|Z = female) = \frac{0.274 + 0.101}{0.274 + 0.101 + 0.079 + 0.036} = 0.765$$

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

3.6 INVERSE PROBABILITY WEIGHING

Multiplying the gender-matching rows by $1/0.245$ and $1/0.765$, respectively, we obtain **Table 3.5**, which represents the postintervention distribution of the population of **Table 3.3**.

$$\frac{1}{P(X = \textit{yes}|Z = \textit{male})} = \frac{1}{0.245}$$

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \textit{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.473
yes	no	male	0.037

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.116
yes	no	male	0.009

3.6 INVERSE PROBABILITY WEIGHING

Multiplying the gender-matching rows by $1/0.245$ and $1/0.765$, respectively, we obtain **Table 3.5**, which represents the postintervention distribution of the population of **Table 3.3**.

$$\frac{1}{P(X = \text{yes} | Z = \text{female})} = \frac{1}{0.765}$$

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	female	0.358
yes	no	female	0.132

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	female	0.274
yes	no	female	0.101

3.6 INVERSE PROBABILITY WEIGHING

Multiplying the gender-matching rows by $1/0.245$ and $1/0.765$, respectively, we obtain **Table 3.5**, which represents the postintervention distribution of the population of **Table 3.3**.

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.473
yes	yes	female	0.358
yes	no	male	0.037
yes	no	female	0.132

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

3.6 INVERSE PROBABILITY WEIGHING

The probability of recovery in this distribution can now be computed directly from the data (**Table 3.5**), by summing the first two rows:

$$P(Y = \text{yes} | do(X = \text{yes})) = 0.473 + 0.358 = 0.832$$

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $do(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.473
yes	yes	female	0.358

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

- 1) the redistribution of weight is no longer proportional but quite discriminatory.
 - Row #1, for instance, boosted its weight from 0.116 to 0.473, a factor of 4.08,

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.473
yes	yes	female	0.358
yes	no	male	0.037
yes	no	female	0.132

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

- 1) the redistribution of weight is no longer proportional but quite discriminatory.
 - Row #1, for instance, boosted its weight from 0.116 to 0.473, a factor of 4.08,

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.116

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	male	0.473

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

- 1) the redistribution of weight is no longer proportional but quite discriminatory.
 - Row #1, for instance, boosted its weight from 0.116 to 0.473, a factor of 4.08,
 - Row #2 is boosted from 0.274 to 0.358, a factor of only 1.307.

Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	female	0.274

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

<i>X</i>	<i>Y</i>	<i>Z</i>	<i>% population</i>
yes	yes	female	0.358

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

- 1) the redistribution of weight is no longer proportional but quite discriminatory.
 - Row #1, for instance, boosted its weight from 0.116 to 0.473, a factor of 4.08,
 - Row #2 is boosted from 0.274 to 0.358, a factor of only 1.307.

This redistribution renders X independent of Z , as in a randomized trial (**Figure 3.4**).

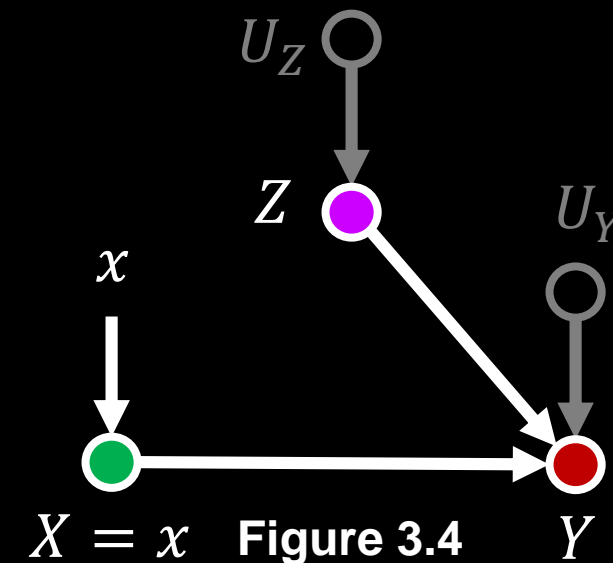


Table 3.3 Joint probability distribution $P(X, Y, Z)$ for the drug gender-recovery story of Part 1 (Table 1.1)

X	Y	Z	% population
yes	yes	male	0.116
yes	yes	female	0.274
yes	no	male	0.009
yes	no	female	0.101
no	yes	male	0.334
no	yes	female	0.079
no	no	male	0.051
no	no	female	0.036

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

X	Y	Z	% population
yes	yes	male	0.473
yes	yes	female	0.358
yes	no	male	0.037
yes	no	female	0.132

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

2) an astute reader would notice that in this example no computational savings were realized; to estimate

$$P(Y = \text{yes} | do(X = \text{yes}))$$

we still needed to sum over all values of Z , males and females.

Indeed, the savings become significant when the number of Z values is in the thousands or millions, and the sample size is in the hundreds.

In such cases, the number of Z values that the inverse probability method would encounter is equal to the number of samples available, not to the number of possible Z values, which is prohibitive.

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $do(X = \text{yes})$, determined via the inverse probability method

X	Y	Z	% population
yes	yes	male	0.473
yes	yes	female	0.358
yes	no	male	0.037
yes	no	female	0.132

3.6 INVERSE PROBABILITY WEIGHING

Three points are worth noting about this procedure:

3) an important word of caution. The method of inverse probability weighing is only valid when the set Z entering the factor

$$\frac{1}{P(X = x|Z = z)}$$

satisfies the backdoor criterion.

Lacking this assurance, the method may actually introduce more bias than the one obtained through naive conditioning, which produces **Table 3.4** and the absurdities of Simpson's paradox.

Table 3.4 Conditional probability distribution $P(Y, Z | X)$ for drug users ($X = \text{yes}$) in the population of Table 3.3

X	Y	Z	% population
yes	yes	male	0.232
yes	yes	female	0.548
yes	no	male	0.018
yes	no	female	0.202

Table 3.5 Probability distribution for the population of Table 3.3 under the intervention $\text{do}(X = \text{yes})$, determined via the inverse probability method

X	Y	Z	% population
yes	yes	male	0.473
yes	yes	female	0.358
yes	no	male	0.037
yes	no	female	0.132

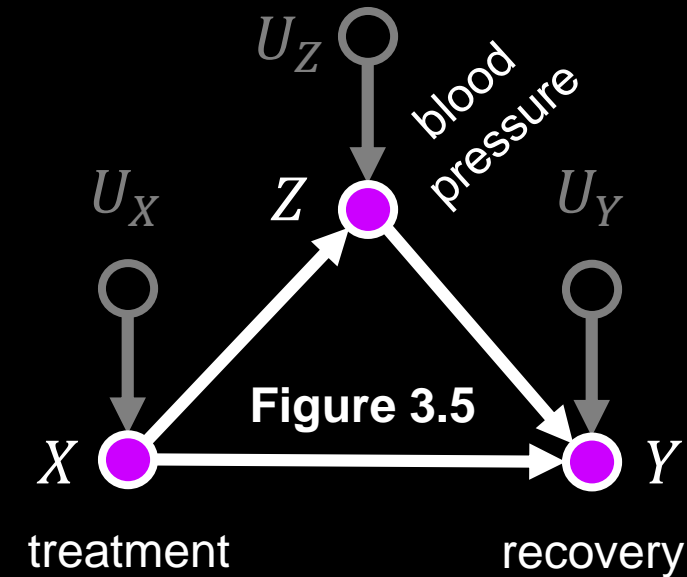
3.7 MEDIATION

Often, when one variable causes another, it does so both directly and indirectly, through a set of mediating variables.

For instance, in our blood pressure (Z) / treatment (X) / recovery (Y) example of Simpson's paradox, treatment is both a direct (negative) cause of recovery, and an indirect (positive) cause, through the mediator of blood pressure—treatment decreases blood pressure, which increases recovery.

In many cases, it is useful to know

- how much of variable X 's effect on variable Y is direct and



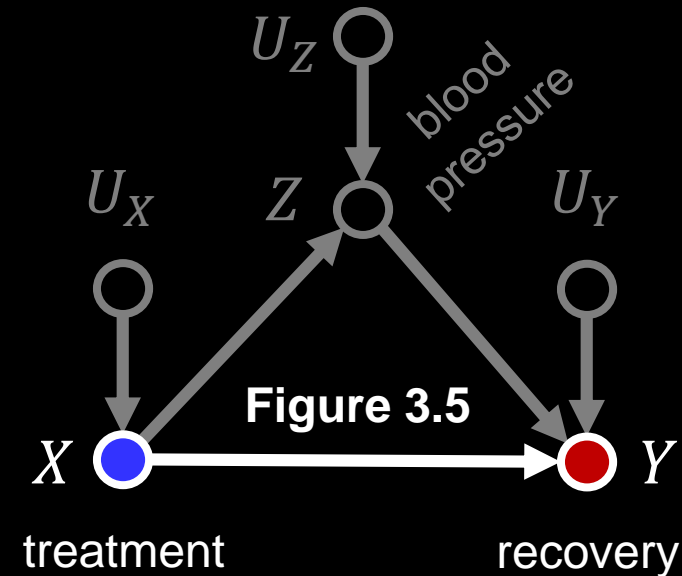
3.7 MEDIATION

Often, when one variable causes another, it does so both directly and indirectly, through a set of mediating variables.

For instance, in our blood pressure (Z) / treatment (X) / recovery (Y) example of Simpson's paradox, treatment is both a direct (negative) cause of recovery, and an indirect (positive) cause, through the mediator of blood pressure—treatment decreases blood pressure, which increases recovery.

In many cases, it is useful to know

- how much of variable X 's effect on variable Y is direct and



3.7 MEDIATION

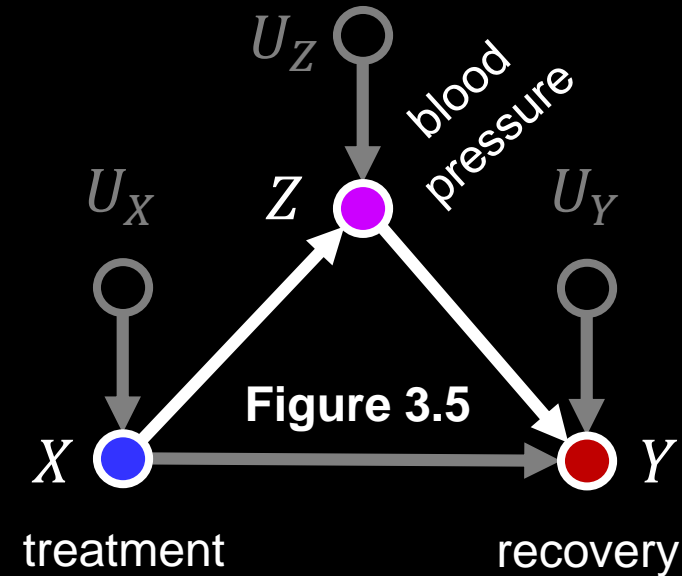
Often, when one variable causes another, it does so both directly and indirectly, through a set of mediating variables.

For instance, in our blood pressure (Z) / treatment (X) / recovery (Y) example of Simpson's paradox, treatment is both a direct (negative) cause of recovery, and an indirect (positive) cause, through the mediator of blood pressure—treatment decreases blood pressure, which increases recovery.

In many cases, it is useful to know

- how much of variable X 's effect on variable Y is direct and
- how much is mediated.

In practice, however, separating these two avenues of causation has proved difficult.

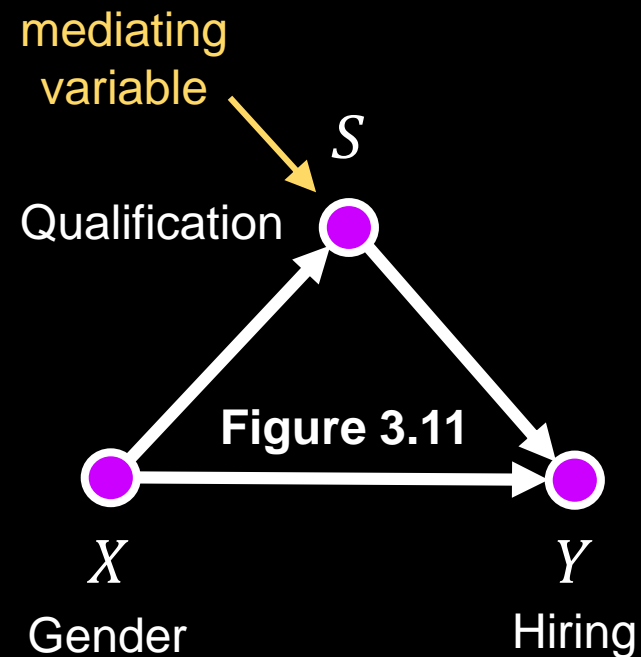


3.7 MEDIATION

Suppose, for example, we want to know whether and to what degree a company discriminates by Gender (X) in its Hiring practices (Y). Such discrimination would constitute a direct effect of gender on hiring, which is illegal in many cases.

However, Gender (X) also affects Hiring (Y) practices in other ways; often, for instance, women are more or less likely to go into a particular field than men, or to have achieved advanced degrees in that field.

So Gender (X) may also have an indirect effect on Hiring (Y) through the **mediating variable** of Qualification (S).



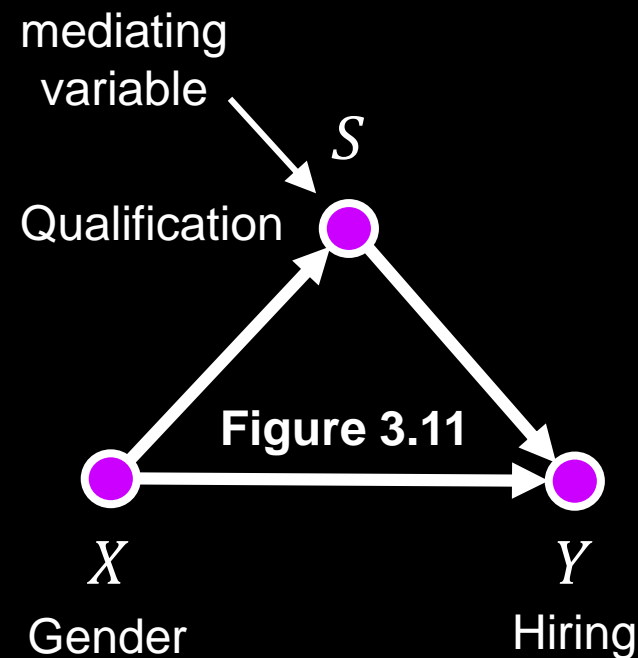
3.7 MEDIATION

Suppose, for example, we want to know whether and to what degree a company discriminates by Gender (X) in its Hiring practices (Y). Such discrimination would constitute a direct effect of gender on hiring, which is illegal in many cases.

However, Gender (X) also affects Hiring (Y) practices in other ways; often, for instance, women are more or less likely to go into a particular field than men, or to have achieved advanced degrees in that field.

So Gender (X) may also have an indirect effect on Hiring (Y) through the **mediating variable** of Qualification (S).

In order to find the direct effect of Gender (X) on Hiring (Y), we need to somehow hold Qualification (S) steady, and measure the remaining relationship between Gender (X) and Hiring (Y); with Qualification (S) unchanging, any change in Hiring (Y) would have to be due to Gender (X) alone.



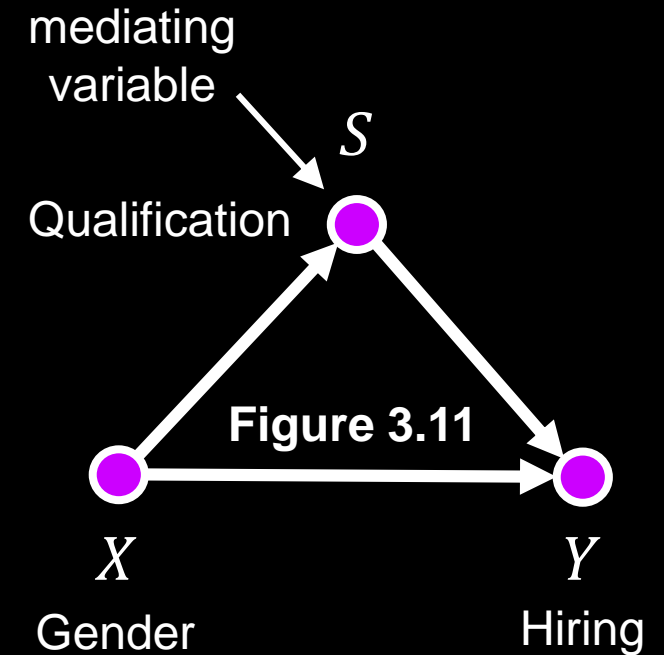
3.7 MEDIATION

Traditionally, this has been done by conditioning on the mediating variable (Qualification (S)).

So if

$$P(\text{Hired} \mid \text{Female, Highly Qualified}) \neq P(\text{Hired} \mid \text{Male, Highly Qualified}),$$

the reasoning goes, then there is a direct effect of Gender (X) on Hiring (Y).



In order to find the direct effect of Gender (X) on Hiring (Y), we need to somehow hold Qualification (S) steady, and measure the remaining relationship between Gender (X) and Hiring (Y); with Qualification (S) unchanging, any change in Hiring (Y) would have to be due to Gender (X) alone.

3.7 MEDIATION

Traditionally, this has been done by conditioning on the mediating variable (Qualification (S)).

So if

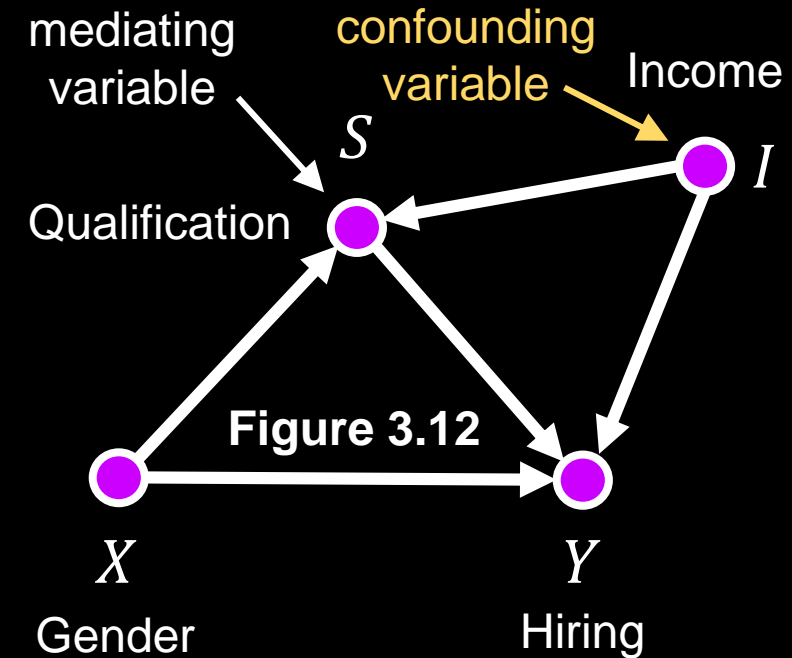
$$P(\text{Hired} \mid \text{Female, Highly Qualified}) \neq P(\text{Hired} \mid \text{Male, Highly Qualified}),$$

the reasoning goes, then there is a direct effect of Gender (X) on Hiring (Y).

In the example in **Figure 3.11**, this is correct.

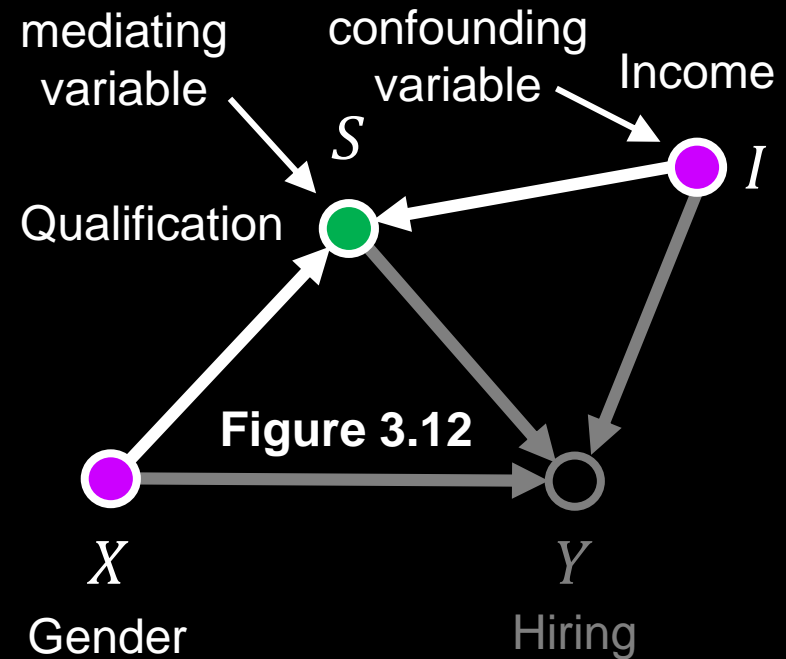
But consider what happens if there are confounders of the mediating variable and the outcome variable.

- For instance, **Income** (I): People from higher income backgrounds are more likely to have gone to college and more likely to have connections that would help them get hired.



3.7 MEDIATION

Now, if we condition on Qualification (S), we are conditioning on a collider.

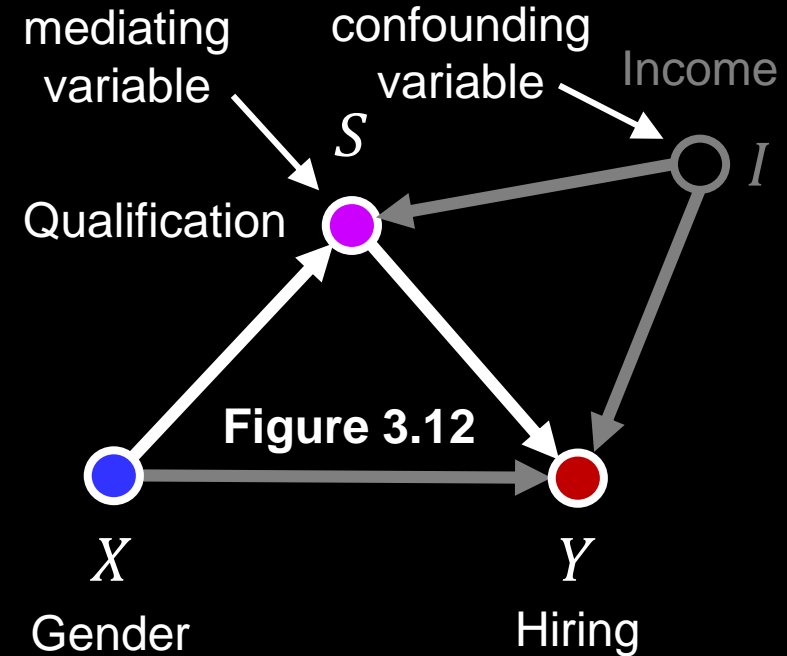


3.7 MEDIATION

Now, if we condition on Qualification (S), we are conditioning on a collider.

- if we don't condition on Qualification (S), indirect dependence can pass from Gender (X) to Hiring (Y) through the path

Gender \rightarrow Qualification \rightarrow Hiring



3.7 MEDIATION

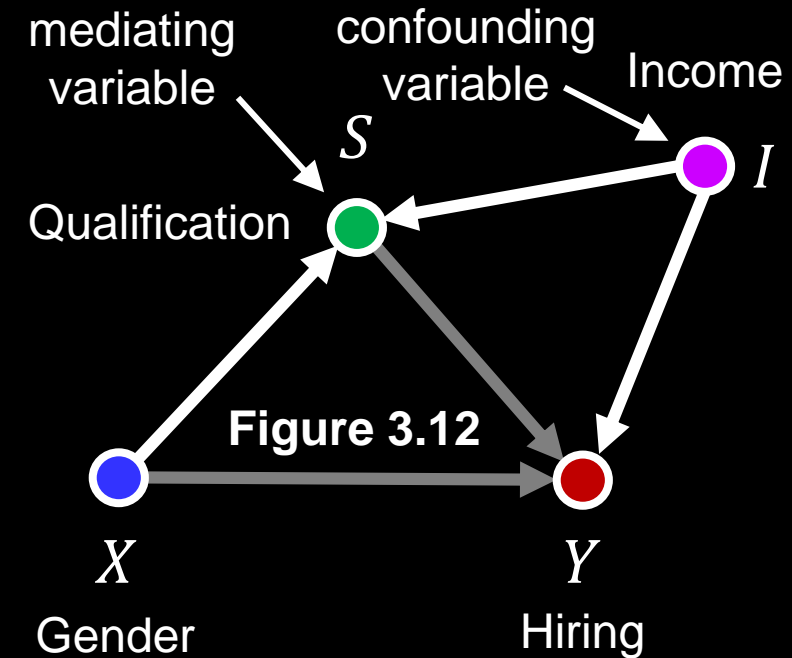
Now, if we condition on Qualification (S), we are conditioning on a collider.

- if we don't condition on Qualification (S), indirect dependence can pass from Gender (X) to Hiring (Y) through the path

Gender \rightarrow Qualification \rightarrow Hiring

- if we do condition on Qualification (S), indirect dependence can pass from gender to hiring through the path

Gender \rightarrow **Qualification** \leftarrow Income \rightarrow Hiring

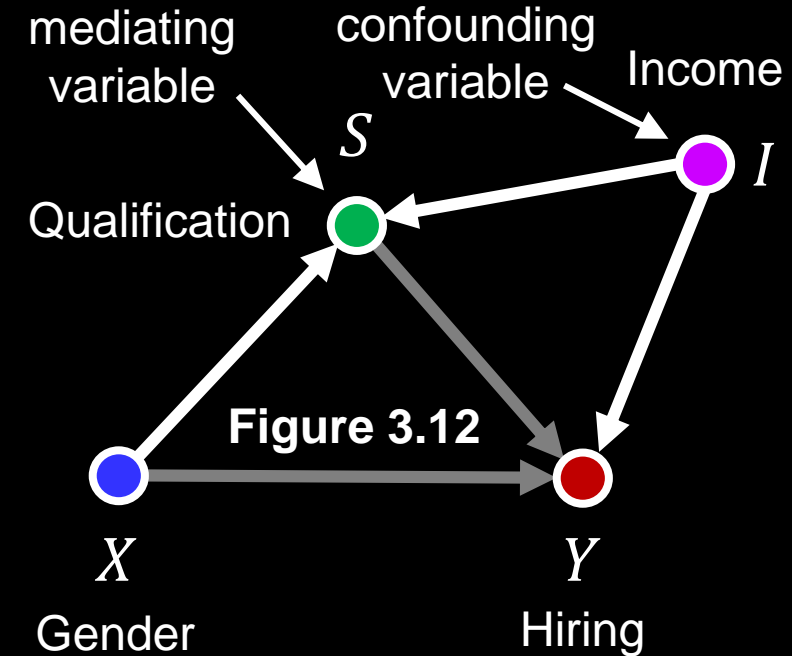


3.7 MEDIATION

To understand the problem intuitively, note that by conditioning on Qualification (S), we will be comparing men and women at different levels of income, because income must change to keep qualification constant.

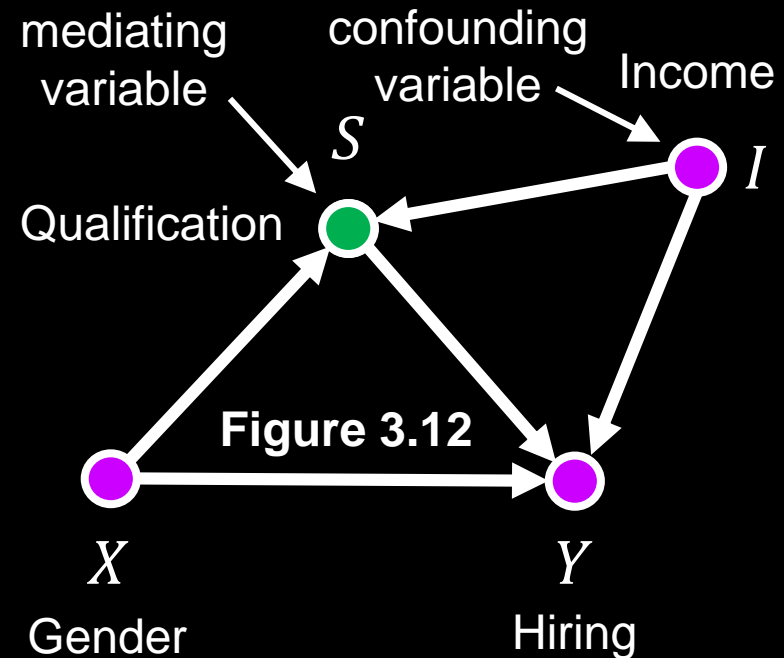
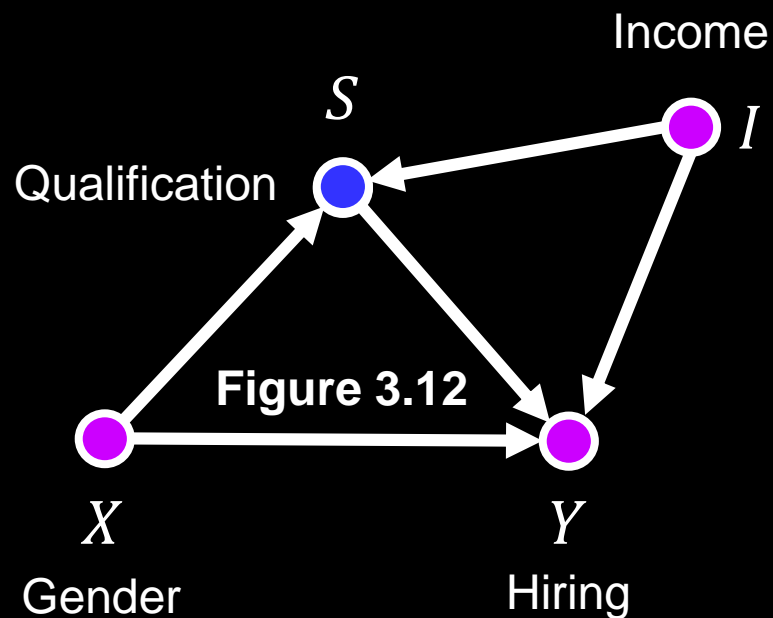
No matter how you look at it, we're not getting the true direct effect of Gender (X) on Hiring (Y).

Traditionally, therefore, statistics has had to abandon a huge class of potential mediation problems, where the concept of “direct effect” could not be defined, let alone estimated.



3.7 MEDIATION

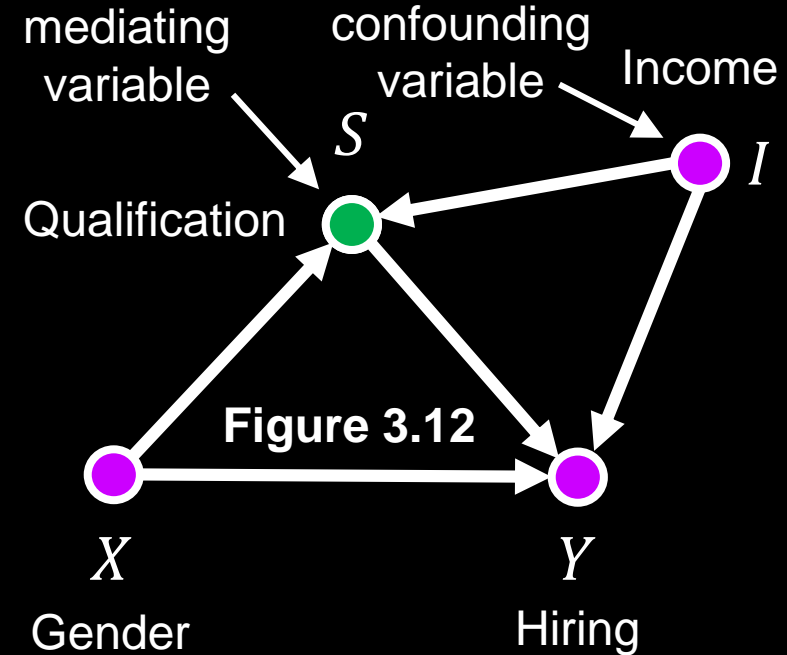
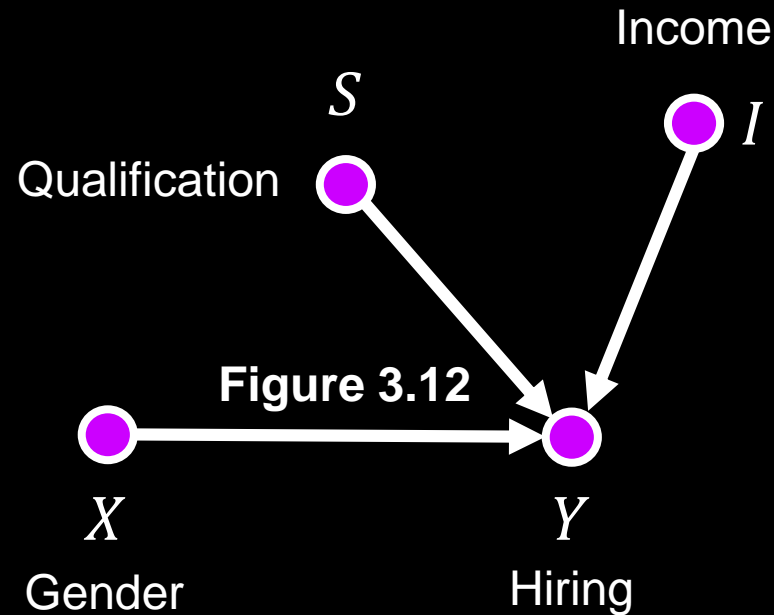
Luckily, we now have a conceptual way of holding the mediating variable steady without conditioning on it: We can intervene on it.



If, instead of conditioning, we fix the Qualification (S), the arrow between Gender (X) and Qualification (S) (and the one between Income (I) and Qualification (S)) disappears, and no spurious dependence can pass through it.

3.7 MEDIATION

Luckily, we now have a conceptual way of holding the mediating variable steady without conditioning on it: We can intervene on it.



If, instead of conditioning, we fix the Qualification (S), the arrow between Gender (X) and Qualification (S) (and the one between Income (T) and Qualification (S)) disappears, and no spurious dependence can pass through it.

Of course, it would be impossible for us to literally change the Qualification (S) of applicants, but recall, this is a theoretical intervention of the kind discussed in the previous section, accomplished by choosing a proper adjustment.

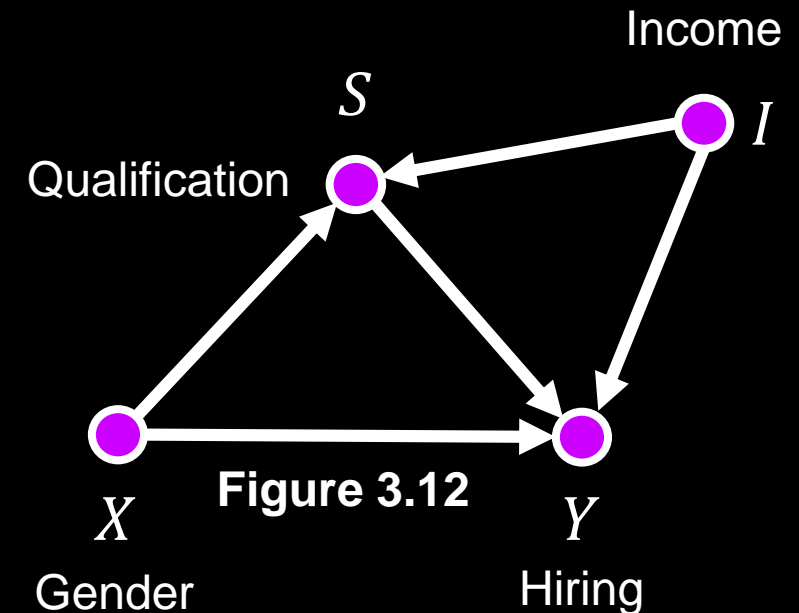
3.7 MEDIATION

So for any three variables X , Y , and S , where S is a mediator between X and Y , the **controlled direct effect (CDE)** on Y of changing the value of X from x to x' is defined as

$$CDE = P(Y = y | do(X = x), do(S = s)) - P(Y = y | do(X = x'), do(S = s))$$

The obvious advantage of this definition over the one based on conditioning is its generality;

- it captures the intent of “**keeping S constant**” even in cases where the $S \rightarrow Y$ relationship is confounded (the same goes for the $X \rightarrow S$ and $X \rightarrow Y$ relationships).
- practically, this definition assures us that in any case where the intervened probabilities are identifiable from the observed probabilities, we can estimate the direct effect of X on Y .



3.7 MEDIATION

Note that the direct effect may differ for different values of S ; for instance, it may be that

- hiring practices discriminate against women in jobs with high qualification requirements,
- hiring practices discriminate against men in jobs with low qualification.

Therefore, to get the full picture of the direct effect, we'll have to perform the calculation for every relevant value s of S .

(In linear models, this will not be necessary; for more information, see **Section 3.8**.)

How do we estimate the direct effect when its expression contains two do-operators?

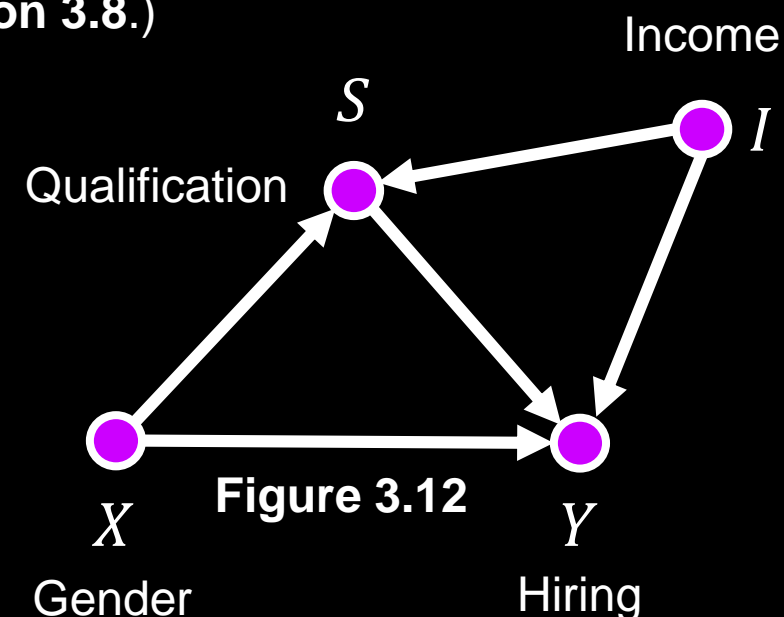


Figure 3.12

3.7 MEDIATION

The technique is more or less the same as the one employed in **Section 3.2**, where we dealt with a single do-operator by adjustment.

In our example of **Figure 3.12**, we first notice that

- there is no backdoor path from X to Y in the model, hence we can replace $do(x)$ with simply conditioning on x (this essentially amounts to adjusting for all confounders).

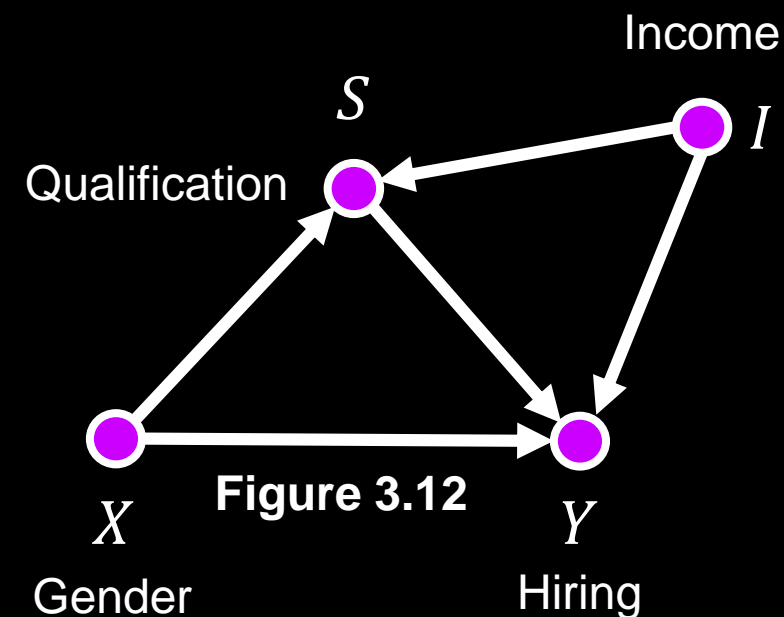


Figure 3.12

3.7 MEDIATION

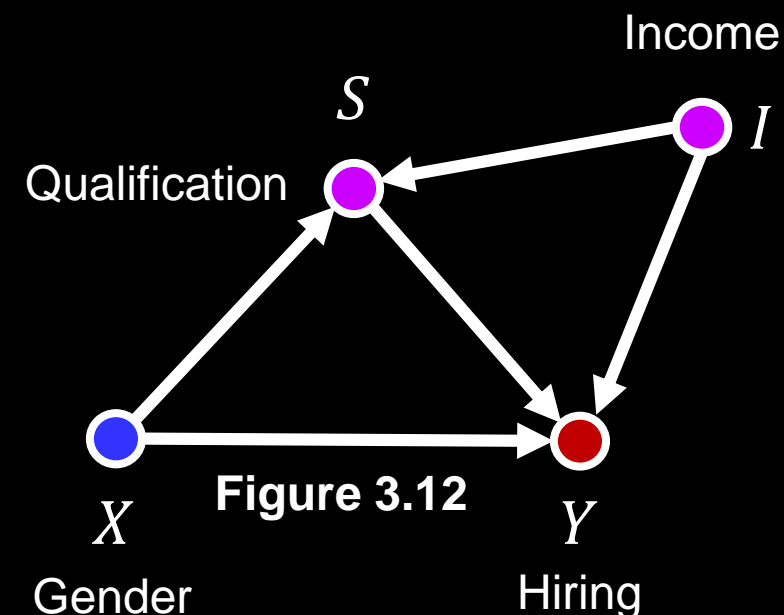
The technique is more or less the same as the one employed in **Section 3.2**, where we dealt with a single do-operator by adjustment.

In our example of **Figure 3.12**, we first notice that

- there is no backdoor path from X to Y in the model, hence we can replace $do(x)$ with simply conditioning on x (this essentially amounts to adjusting for all confounders).

Definition 3.3.1 (The Backdoor Criterion)

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



3.7 MEDIATION

The technique is more or less the same as the one employed in **Section 3.2**, where we dealt with a single do-operator by adjustment.

In our example of **Figure 3.12**, we first notice that

- there is no backdoor path from X to Y in the model, hence we can replace $do(x)$ with simply conditioning on x (this essentially amounts to adjusting for all confounders).

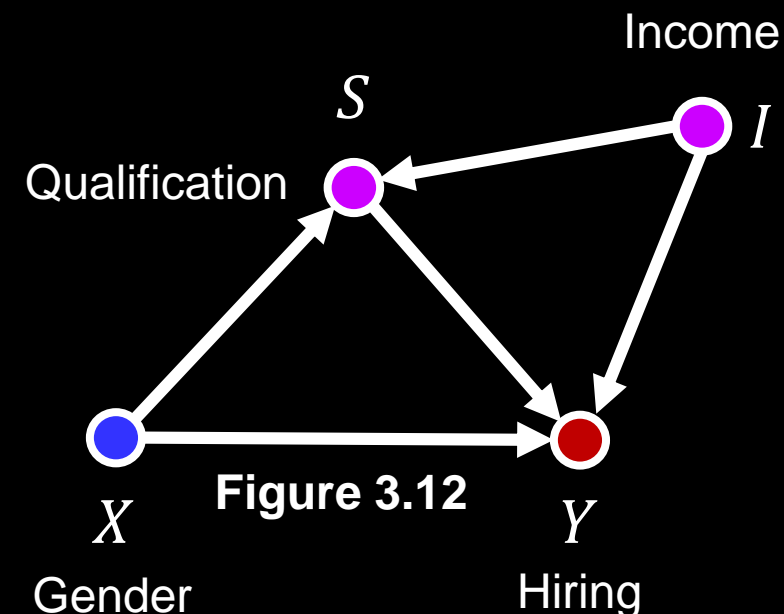
This means that

$$Z = \{\emptyset\}$$

and thus, we do not need to adjust.

This results in

$$P(Y = y|X = x, do(S = s)) - P(Y = y|X = x', do(S = s))$$

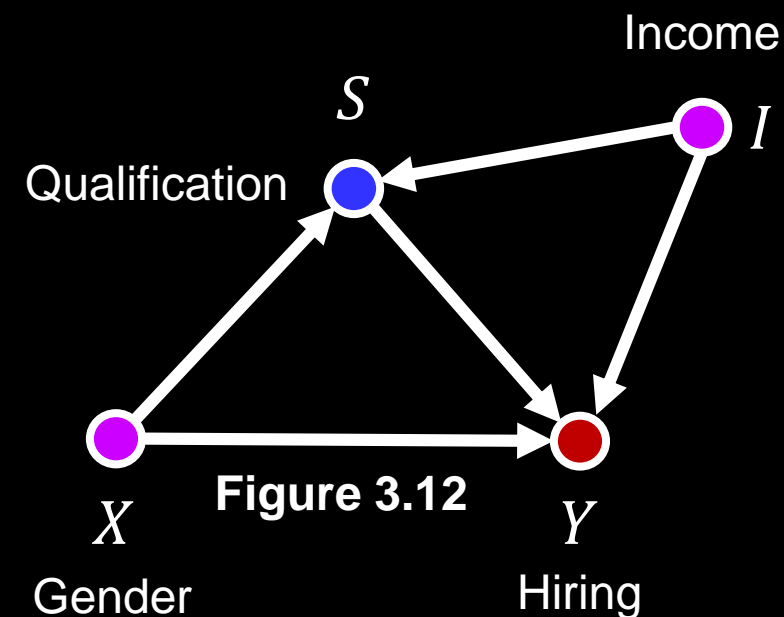


3.7 MEDIATION

Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

Definition 3.3.1 (The Backdoor Criterion)

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



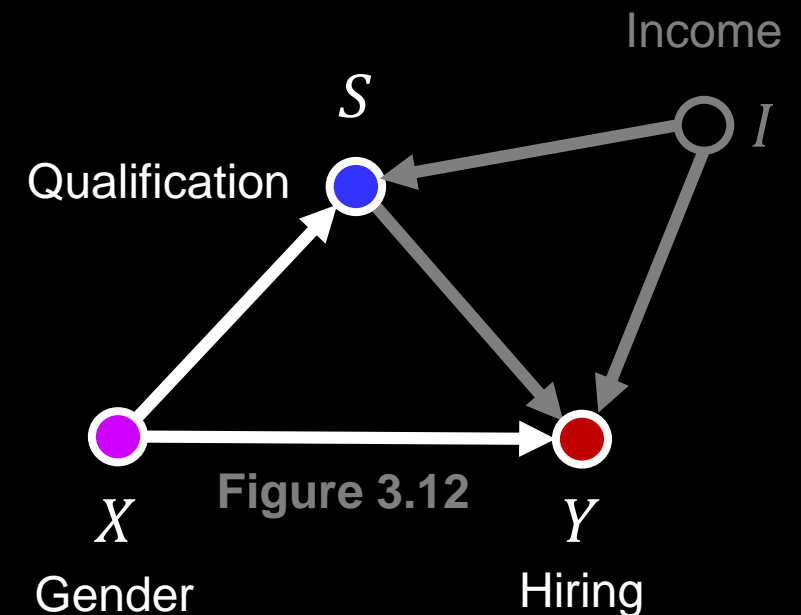
3.7 MEDIATION

Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

- one through X (Qualification \leftarrow Gender \rightarrow Hiring)

Definition 3.3.1 (The Backdoor Criterion)

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



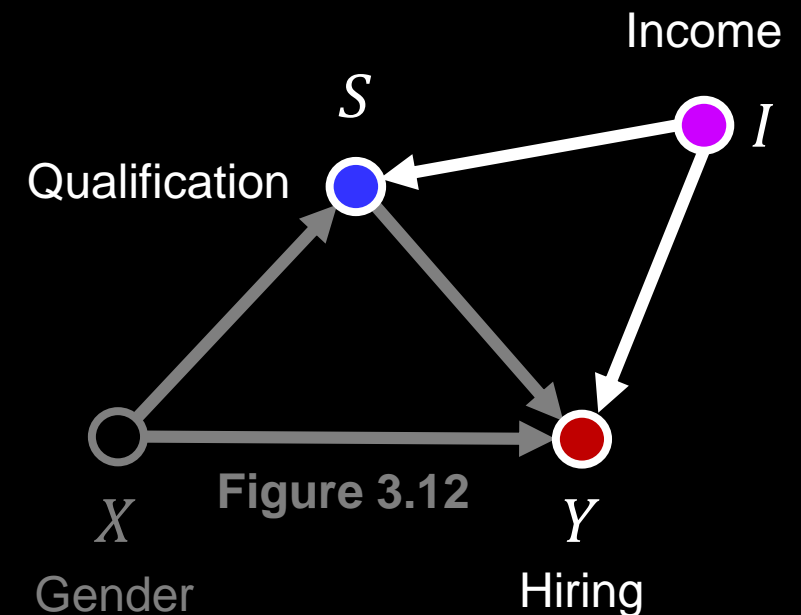
3.7 MEDIATION

Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

- one through X (Qualification \leftarrow Gender \rightarrow Hiring)
- one through I (Qualification \leftarrow Income \rightarrow Hiring)

Definition 3.3.1 (The Backdoor Criterion)

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



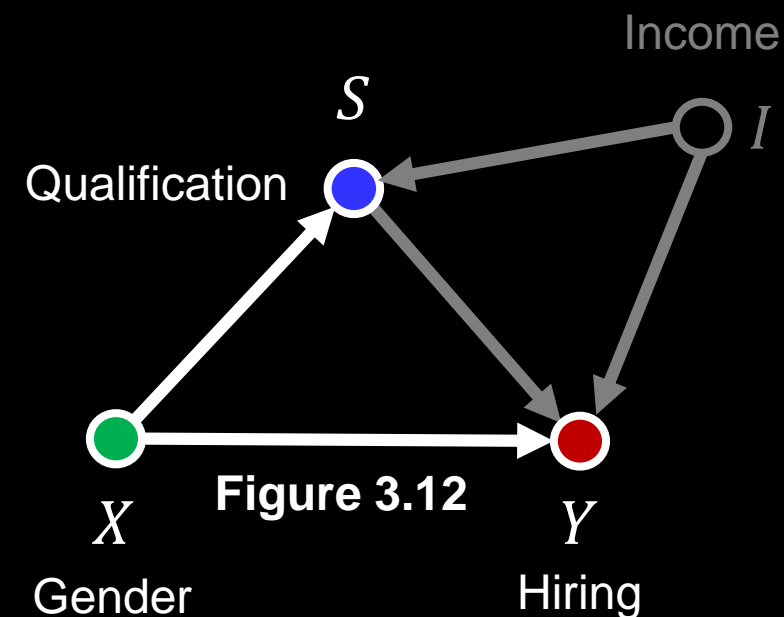
3.7 MEDIATION

Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

- one through X (Qualification \leftarrow Gender \rightarrow Hiring) **blocked (since X is conditioned on)**
- one through I (Qualification \leftarrow Income \rightarrow Hiring)

Definition 3.3.1 (The Backdoor Criterion)

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



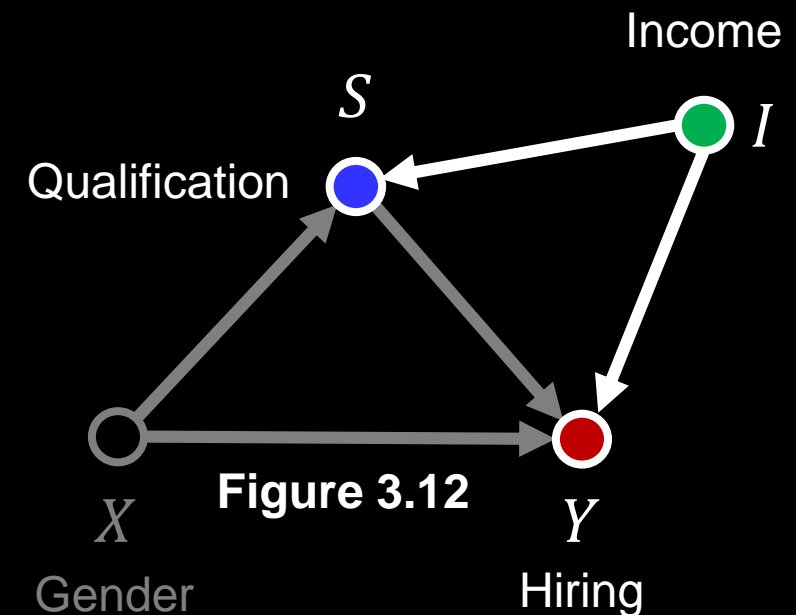
3.7 MEDIATION

Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

- one through X (Qualification \leftarrow Gender \rightarrow Hiring) blocked (since X is conditioned on)
- one through I (Qualification \leftarrow Income \rightarrow Hiring) **blocked if we adjust for I**

Definition 3.3.1 (The Backdoor Criterion)

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



3.7 MEDIATION

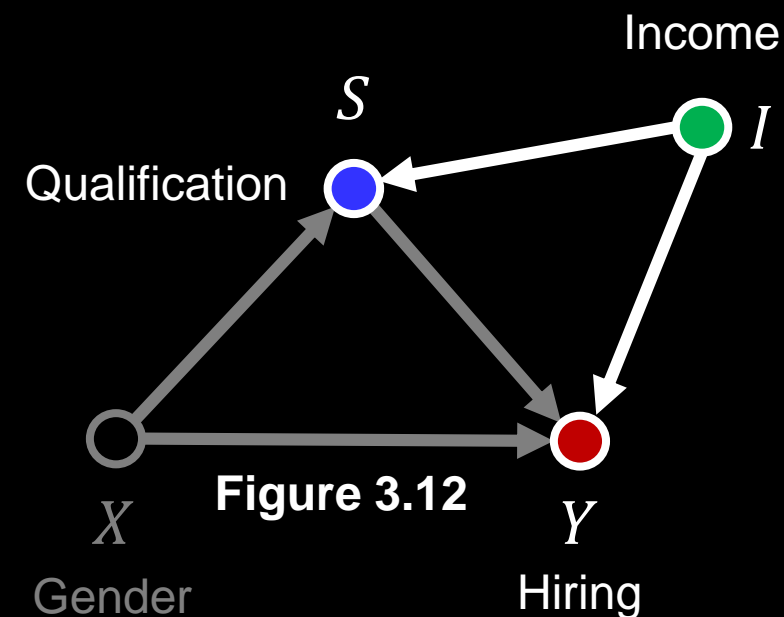
Next, we attempt to remove the $do(s)$ term and notice that two backdoor paths exist from S to Y ,

- one through X (Qualification \leftarrow Gender \rightarrow Hiring) blocked (since X is conditioned on)
- one through I (Qualification \leftarrow Income \rightarrow Hiring) blocked if we adjust for I

This gives

$$\sum_i \{P(Y = y|X = x, S = s, I = i) - P(Y = y|X = x', S = s, I = i)\} P(I = i)$$

The last formula is do-free, which means it can be estimated from nonexperimental data (i.e., observational data).



3.7 MEDIATION

In general, the CDE of X on Y , mediated by Z , is identifiable if the following two properties hold:

1. There exists a set S_1 of variables that blocks all backdoor paths from Z to Y .
2. There exists a set S_2 of variables that blocks all backdoor paths from X to Y , after deleting all arrows entering Z .

If these two properties hold in a model M , then we can determine

$$P(Y = y | do(X = x), do(Z = z))$$

from the data set by adjusting for the appropriate variables, and estimating the conditional probabilities that ensue.

3.7 MEDIATION

In general, the CDE of X on Y , mediated by Z , is identifiable if the following two properties hold:

1. There exists a set S_1 of variables that blocks all backdoor paths from Z to Y .
2. There exists a set S_2 of variables that blocks all backdoor paths from X to Y , after deleting all arrows entering Z .

Note: condition 2) is not necessary in randomized trials, because randomizing X renders X parentless.

The same is true in cases where X is judged to be exogenous (i.e., “as if” randomized), as in the aforementioned gender discrimination example.

It is even trickier to determine the indirect effect than the direct effect, because there is simply no way to condition away the direct effect of X on Y .

It's easy enough to find the total effect and the direct effect, so some may argue that the indirect effect should just be the difference between those two.

This may be true in linear systems, but in nonlinear systems, differences don't mean much; the change in Y might, for instance, depend on some interaction between X and Z —if, as we posited above, women are discriminated against in high-qualification jobs and men in low-qualification jobs, subtracting the direct effect from the total effect would tell us very little about the effect of gender on hiring as mediated by qualifications.

Clearly, we need a definition of indirect effect that does not depend on the total or direct effects.

We will show in **Part 4** that these difficulties can be overcome through the use of **counterfactuals**, a more refined type of intervention that applies at the individual level and can be computed from structural models.