

Causation to Association II: Conditional Association

1000: Introduction

1100: Overview

The previous module focussed on how causal graphs explain and predict patterns of unconditional association and unconditional independence among a set of variables. In this module we go further into the story by explaining how causal graphs predict patterns of conditional association and conditional independence.

You might want to review the modules on Independence and Conditional Independence if you are not clear on the difference between unconditional and conditional independence. To refresh your memory, the figure below defines both.

Independence	A ⊥ B	$Fr(A) = Fr(A B)$
Conditional Independence	A ⊥ B C	$Fr(A C) = Fr(A C, B)$

FIGURE 1100-1

For an example of how a causal theory makes predictions both about unconditional and conditional independence, consider the causal graph below concerning chicken pox.

EXPOSURE to other children with chicken pox causes **INFECTION**, which in turn causes an itchy rash (**SYMPTOMS**).

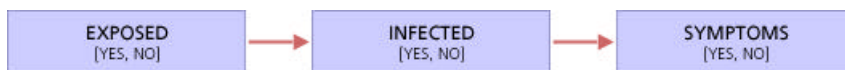


FIGURE 1100-2

From the last module, we know that this causal theory predicts that every pair of variables will be unconditionally associated. It also predicts, however, that **EXPOSURE** and **SYMPTOMS** will be independent **conditional** on **INFECTION**. This prediction about conditional independence deepens the scientific connection between causal theories and statistical data we established in the last module. Understanding how the graph above and other causal structures predict conditional independence or conditional associations is the topic of this module.

When you finish this module, you should be able to identify the pieces of causal graphs (like causal chains and common causes) that produce patterns of conditional association and conditional independence.

1200: The Causality Lab

The theory that takes causal graphs and outputs the predictions they make about independence relations is programmed into the Causality Lab. You can construct a causal graph in the Causality Lab, and then ask it to compute which pairs of variables are predicted to be associated, which pairs independent, and which pairs are conditionally associated or conditionally independent.

Before you proceed with this module, or at any time during the module, we invite you to use this feature of the Lab. See if you can figure out the general theory before we present it. Exactly what characteristics must a causal graph have in order for it to predict that **X** and **Y** are associated? What characteristics must it have for it to predict that **X** and **Y** are independent? Experiment until you have a theory. Then read the module and see if you got it right.

< [A link to a Java applet in the interactive version of this module.](#) >

2000: Unconditional Association

2100: Review of Causal Connection

Two variables **X** and **Y** in a causal graph are predicted to be unconditionally associated just in case they are causally connected in the graph. **X** and **Y** are **causally connected** in the graph just in case either

- + X is a cause of Y, or
- + Y is a cause of X, or
- + there is a common cause of X and Y

Variables can be causally connected in multiple ways, and each causal connection produces association. For example, in the graph below, a child's exposure to lead and his or her IQ score are causally connected in two ways, each of which produces association between Lead Exposure and IQ.

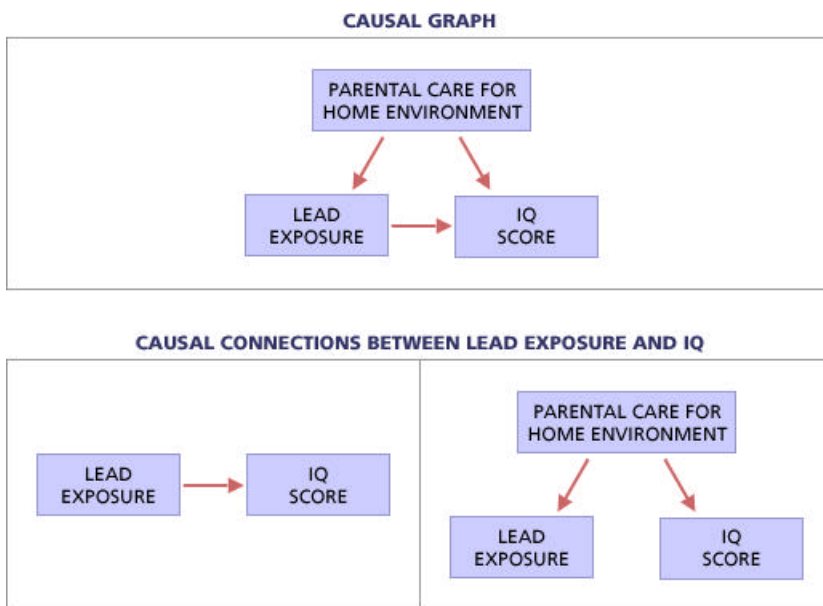


FIGURE 2100-1

Although it is possible, we assume that when multiple causal connections exist between X and Y, the overall association between X and Y is not zero. That is, if X and Y are causally connected, we predict they are associated, and if X and Y are not causally connected, then we predict that they are independent.

< [A link to exercises in the interactive version of this module.](#) >

2200: Mediators, Common Causes and Common Effects

Consider again the causal graph you just analyzed in the previous question:

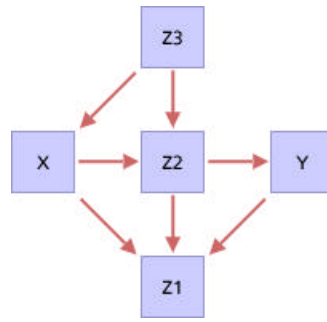


FIGURE 2200-1

We created your choices A - E by listing **all** the paths from **X** to **Y**, **neglecting** the direction of the arrows. Paths that connect two variables -- regardless of the direction of each of the arrows on the path -- are called **undirected paths**. If all the arrows on a path point the same way, then the path is also a **directed path**. All directed paths are also undirected paths, but not vice versa. We will often refer to an undirected path as just a "path."

For example, in the figure below, there is an undirected path from **X** to **Y**, a directed path (which is also an undirected path) from **X** to **V**, and no path at all from **X** to **W**.

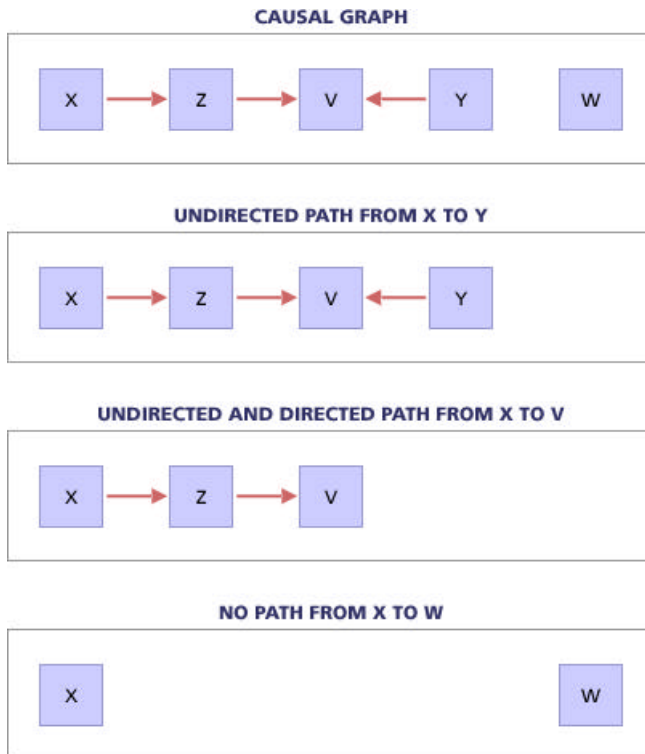


FIGURE 2200-2

< [A link to exercises in the interactive version of this module.](#) >

Here again is the graph and list of undirected paths from question in the previous section, with each path labelled as causal connection or not.

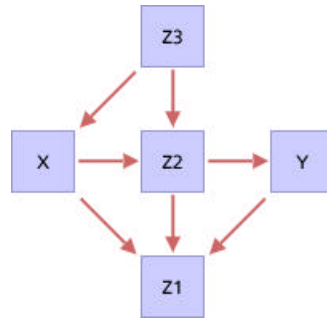


FIGURE 2200-3

TABLE 2200-1: CAUSAL PATHS

Instance	Undirected Path	Causal Connection Between X and Y
A	$X \leftarrow Z3 \rightarrow Z2 \rightarrow Y$	Yes
B	$X \rightarrow Z2 \rightarrow Y$	Yes
C	$X \rightarrow Z2 \rightarrow Z1 \leftarrow Y$	No
D	$X \rightarrow Z1 \leftarrow Y$	No
E	$X \rightarrow Z1 \leftarrow Z2 \rightarrow Y$	No
F	$X \leftarrow Z3 \rightarrow Z2 \rightarrow Z1 \leftarrow Y$	No

What is it about paths A and B that qualifies them as causal connections between X and Y?
 What is it about paths C, D, and E that disqualifies them as causal connections?

Consider undirected path A: $X \leftarrow Z3 \rightarrow Z2 \rightarrow Y$. The variable Z3 is a common cause on this undirected path. The variable Z2 a mediator on the path. In general, a variable is a mediator on an undirected path if one of the edges goes into it and the other goes out: $\rightarrow X \rightarrow$ or $\leftarrow X \leftarrow$.

We can classify every variable on an undirected path as one of the following:



FIGURE 2200-4

< A link to exercises in the interactive version of this module. >

What is it about an undirected path that qualifies it as a causal connection? Paths that are causal connections:

- + contain only common causes or mediators, and
- + do not contain any common effects.

< A link to exercises in the interactive version of this module. >

Again, mediators and common causes are the only kind of variable that occur on a causal connection. Common effects do not.

2300: Colliders and Non-colliders

What is the difference between mediators and common causes on the one hand, and common effects on the other? The arrows next to a common effect collide into it, but the arrows next to a common cause or mediator do not collide. Thus we divide variables on undirected paths into two types: colliders and non-colliders.

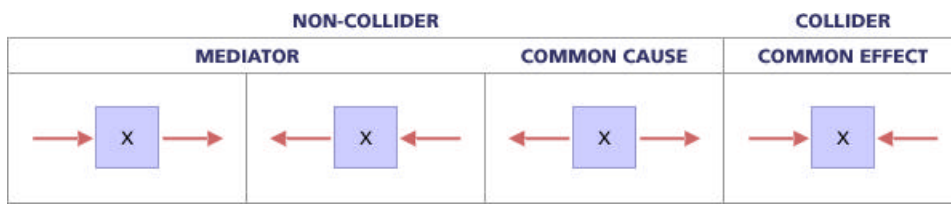


FIGURE 2300-1

As you will see, dividing variables on paths into colliders and non-colliders is perhaps the most fundamental insight connecting causal theories to statistical data, and it was only discovered around 1985.

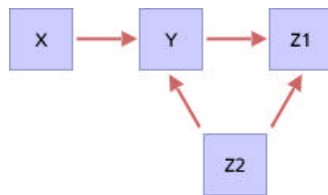


FIGURE 2300-2

By itself, a variable is neither a collider nor a non-collider. It is only a collider or a non-collider on an undirected path. For example, consider the variable Y in the figure above.

< A link to exercises in the interactive version of this module. >

As you can see, **Y** is a non-collider on one path, and a collider on the other, so a variable's status depends on the path being considered.

< [A link to exercises in the interactive version of this module.](#) >

3000: Conditioning on Mediators

3100: Chicken Pox

Consider 3-year-olds and chicken pox again:

TABLE 3100-1: VARIABLES FOR CHICKEN POX AND 3-YEAR-OLDS

Variables	Values
EXPOSED	[Yes, No]
INFECTED	[Yes, No]
SYMPTOMS	[Yes, No]

where **EXPOSED** = yes means that the 3-year-old has been in close proximity to another child with chicken pox in the past week, **INFECTED** = yes means that the child has the virus in his or her bloodstream, and **SYMPTOMS** = yes means that he or she has the typical chicken pox rash.

The causal graph among these variables is clear:

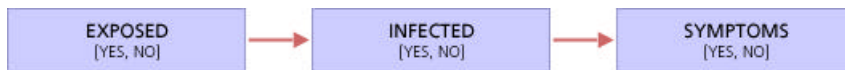


FIGURE 3100-1

EXPOSED is a direct cause of **INFECTED**, which is a direct cause of **SYMPTOMS**, but **EXPOSED** is not a direct cause of **SYMPTOMS**.

< [A link to exercises in the interactive version of this module.](#) >

Using common sense, it is clear that **EXPOSURE** and **SYMPTOMS** are associated, but independent **conditional** on **INFECTED** among 3-year olds. Once I know that a 3-year old is infected with the chicken pox virus, telling me that he or she was exposed in the last week doesn't tell me anything about whether he or she will develop symptoms.

In the causal graph, **INFECTED** is a mediator between **EXPOSURE** and **SYMPTOMS**. Whereas the causal path from **EXPOSURE** to **SYMPTOMS** produces association, conditioning on **INFECTION** (the mediator) blocks the path, thus preventing it from producing association conditional on the mediator.

Lets consider in detail how this happens. Like we did in the module on Causation to Unconditional Association, lets suppose that this system is pseudo-indeterministic because of hidden variables. Suppose, when we reveal the hidden variables in the graph below, the system is fully deterministic:

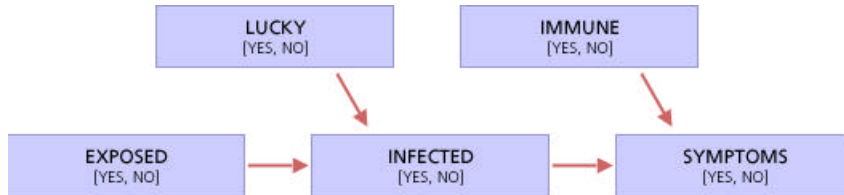


FIGURE 3100-2

Only children who are exposed and unlucky become infected, and only children who are infected and are not immune to chicken pox become symptomatic. We established in the module on Causation to Unconditional Association that **EXPOSED** and **SYMPTOMS** are associated in this system. Now lets establish that they are independent conditional on **INFECTED**.

< [A link to exercises in the interactive version of this module.](#) >

Below is a simulation of this causal system. To run a single trial in an experiment click the OK button in the "single trial" row and complete the steps that follow. To run multiple trial click on the OK button in the "multiple trials" row and complete the steps that follow. The results of each trial will be displayed as histograms. Run at least 200 trials.

< [A simulation in the interactive version of this module.](#) >

< [A link to exercises in the interactive version of this module.](#) >

In general, conditioning on a mediator prevents the causal connection it is on from producing association. The length of the causal path makes no difference. Consider the following causal graph.

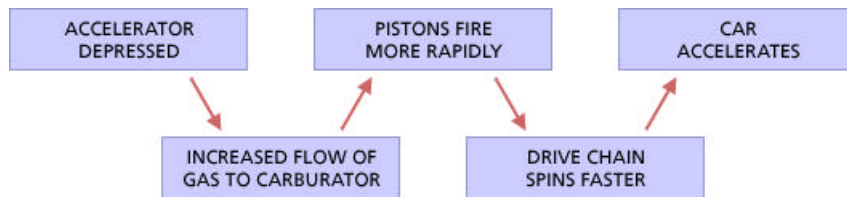


FIGURE 3100-3

The causal path produces association between **ACCELERATOR DEPRESSED** and **CAR ACCELERATES**. Conditioning on any subset of the mediators prevents the path from producing association, however. This graph, therefore, predicts that all of the following independence relations will hold:

- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **GAS FLOW**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **PISTONS**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **DRIVE TRAIN**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **GAS FLOW, PISTONS**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **GAS FLOW, DRIVE TRAIN**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **PISTONS, DRIVE TRAIN**
- + **ACCELERATOR DEPRESSED** $\perp\!\!\!\perp$ **CAR ACCELERATES** | **GAS FLOW, PISTONS, DRIVE TRAIN**

Blocking any subset of mediators blocks the whole path from producing association.

3200: Multiple Connections

In the chicken pox case there is only one causal connection between **EXPOSURE** and **SYMPTOMS**. By conditioning on **INFECTION**, we prevented it from producing association. We know that every causal connection produces association, however, so we now turn to cases in which two variables are multiply causally connected.

Consider the two causal graphs below:

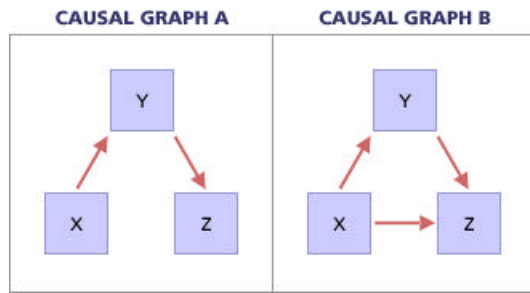


FIGURE 3200-1

< [A link to exercises in the interactive version of this module.](#) >

As you can see, in both graphs every pair of variables is causally connected. Thus both graphs predict that every pair of variables is associated.

< [A link to exercises in the interactive version of this module.](#) >

Both graphs, however, do not lead to the same predictions. Although they both lead to the same predictions about **unconditional** association -- they do not lead to the same predictions about **conditional** association.

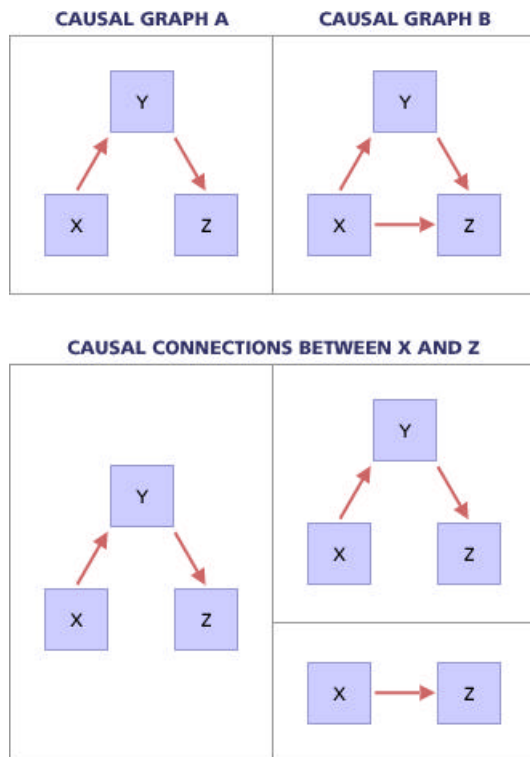


FIGURE 3200-2

Graph A has the same structure as the chicken pox example. There is only one causal connection between **X** and **Z**, a causal path from **X** to **Z** that goes through one mediator, **Y**. By conditioning on **Y**, we block this causal connection from producing association, and thus Graph A predicts that $X \perp\!\!\!\perp Z \mid Y$.

Graph B has two causal connections between **X** and **Z**. When we condition on **Y**, the first causal connection ($X \rightarrow Y \rightarrow Z$) is prevented from producing association between **X** and **Z**, but the second causal connection (the direct cause $X \rightarrow Z$) is still active. That is, it is not blocked from producing association. Thus Graph B predicts that **X** is **not** independent of **Z** | **Y**.

In general, all of the causal connections between **X** and **Y** must be blocked by a set **Z** for a graph to predict that $X \perp\!\!\!\perp Y \mid Z$.

3300: Exercise and Body Weight

Consider the following causal graph

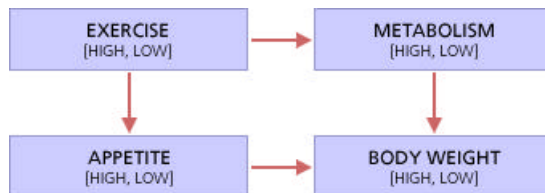


FIGURE 3300-1

For example, in the causal graph above, there are two causal paths from **EXERCISE** to **BODY WEIGHT**:

- 1 **EXERCISE** → **METABOLISM** → **BODY WEIGHT**
- 2 **EXERCISE** → **APPETITE** → **BODY WEIGHT**

< [A link to exercises in the interactive version of this module.](#) >

Since there are two paths from **EXERCISE** to **BODY WEIGHT**, each produces association, and thus this graph predicts that they are unconditionally dependent. Each path must be blocked if we are to make **EXERCISE** and **BODY WEIGHT** conditionally independent.

If we condition on **METABOLISM**, then we prevent path 1 from producing association, but path 2 will still make **EXERCISE** and **BODY WEIGHT** associated. If we condition on **APPETITE**, then we prevent path 2 from producing association, but path 1 will still make **EXERCISE** and **BODY WEIGHT** associated. The only way to make **EXERCISE** and **BODY WEIGHT** conditionally independent is to condition on **both METABOLISM and APPETITE**.

Thus, this graph predicts that **EXERCISE** $\perp\!\!\!\perp$ **BODY WEIGHT** | **METABOLISM, APPETITE**

4000: Conditioning on Common Causes

4100: Hereditary Baldness

Consider the variables:

TABLE 4100-1: VARIABLES FOR BALDNESS

Variables	Values
BALD	[Yes, No]
BALD BROTHER	[Yes, No]
MOTHER HAS BALDNESS GENE	[Yes, No]

We will abbreviate **MOTHER HAS BALDNESS GENE** with **BALDNESS GENE**. The causal graph relating these variables is:

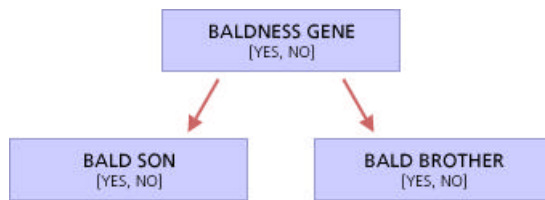


FIGURE 4100-1

[< A link to exercises in the interactive version of this module. >](#)

Although the story is of course more complicated, let's assume that the causal graph above is basically correct in describing the causal relations governing baldness. We know that the variables **BALD** and **BALD BROTHER** are associated because **BALDNESS GENE** is a common cause of both. Brothers with a mother who carries the baldness gene are both likely to be bald, and brothers with a mother who doesn't carry the gene are both likely not to be bald.

What about the independence of **BALD** and **BALD BROTHER** conditional on **BALDNESS GENE**. Are being bald and having a bald brother independent among brothers with a mother who carries the gene? Are being bald and having a bald brother independent among brothers with a mother who does not carry the gene? Put another way, suppose I tell you that my mother carries the baldness gene. What are the chances that I am bald? Now suppose I tell you an extra piece of information: my brother is bald. Does that change the chances that I am bald? The answer is no -- it adds no new information I didn't already have after I learned that my mother had the baldness gene.

How does this work in detail? Consider again a pseudo-indeterministic system of the sort we described in the module on Causation to Unconditional Association.

The causal graph for baldness above fills out into a deterministic graph as so:

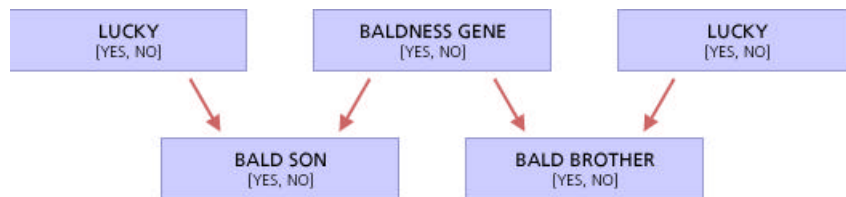


FIGURE 4100-2

The following simulation embodies this system. Once again, explore what happens when you manipulate the common cause: **BALDNESS GENE**. To run a single trial in an experiment click the OK button in the "single trial" row and complete the steps that follow. To run multiple trial click on the OK button in the "multiple trials" row and complete the steps that follow. The results of each trial will be displayed as histograms. Produce at least 200 trials, and study the histograms that capture the association between **BALD SON** and **BALD BROTHER**.

< [A simulation in the interactive version of this module.](#) >

< [A link to exercises in the interactive version of this module.](#) >

4200: Smoking, Yellow Fingers and Lung Cancer

Consider the variables:

TABLE 4200-1: SMOKES, YELLOW FINGERS AND LUNG CANCER

Variables	Values
SMOKES	[Yes, No]
YELLOW FINGERS	[Yes, No]
LUNG CANCER	[Yes, No]

The causal graph relating these variables is clear:

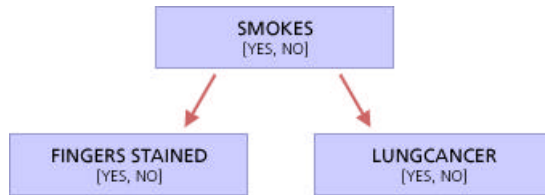


FIGURE 4200-1

< A link to exercises in the interactive version of this module. >

Clearly **YELLOW FINGERS** and **LUNG CANCER** are predicted to be associated by this graph. That is, an individual is more likely to get lung cancer given that he or she has yellow fingers than if not. Are **YELLOW FINGERS** and **LUNG CANCER** independent conditional on **SMOKES**, however?

Suppose that I tell you that I am 55 and that I smoke. Now suppose I add that I have yellow fingers. Did the bit about yellow fingers add any new information about whether I will get lung cancer? No. Among smokers, yellow fingers and lung cancer are independent, and among non-smokers they are as well.

In general, conditioning on a common cause blocks the causal connection it is on from producing association.

5000: Blocking a Causal Connection Generally

So far, we have seen that if two variables are causally connected because one is a cause of the other, and we condition on a mediator, then we prevent the causal connection from producing association.

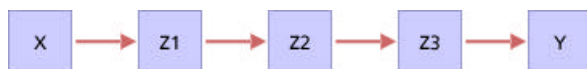


FIGURE 5000-1

< A link to exercises in the interactive version of this module. >

Conditioning on **any** mediator blocks a causal path from producing association.

A causal connection that is produced by a common cause is really two causal paths -- one that goes from the common cause to one effect, and another that goes from the common cause to the other effect.

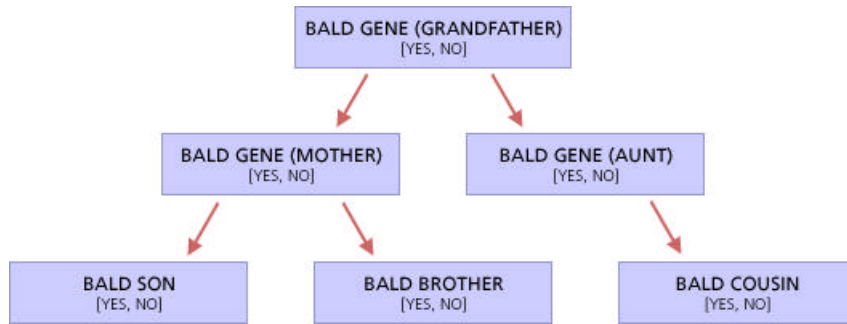


FIGURE 5000-2

For example, in the graph above, consider the causal connection between **BALD SON** and **BALD COUSIN**. Its really two causal paths, one from **BALD GENE** (Grandfather) to **BALD SON**, and another from **BALD GENE** (Grandfather) to **BALD COUSIN**:

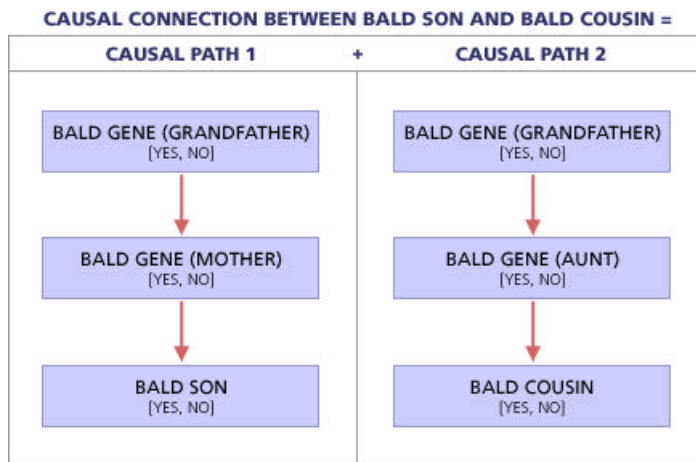


FIGURE 5000-3

If we block either of these causal paths from producing association, then we block the entire causal connection composed of them from producing association. So in this case, if we condition on **BALD GENE** (Mother), then we prevent Causal Path 1 from producing association, and thus the entire causal connection between **BALD SON** and **BALD COUSIN** from producing association. If we condition on **BALD GENE** (Aunt), then we prevent Causal Path 2 from producing association, etc.

< A link to exercises in the interactive version of this module. >

Remember that, in general, all it takes is one causal connection to produce association between a pair of variables. If we want to make a pair of variables conditionally independent, then we at least need to block all of the causal connections between them.

< A link to exercises in the interactive version of this module. >

6000: Conditioning on a Common Effect

In order for a causal graph to predict that a pair of variables **X** and **Y** are conditionally independent given a set of variables **Z**, it is **necessary** that each causal connection between **X** and **Y** be blocked by **Z**. Amazingly, that **Z** blocks all of the causal connections between **X** and **Y** is **not sufficient** to predict that **X** and **Y** are conditionally independent on **Z**. Paths that are not causal connections cannot produce unconditional association, but they can produce association conditional on **Z**! That is the subject to which we now turn.

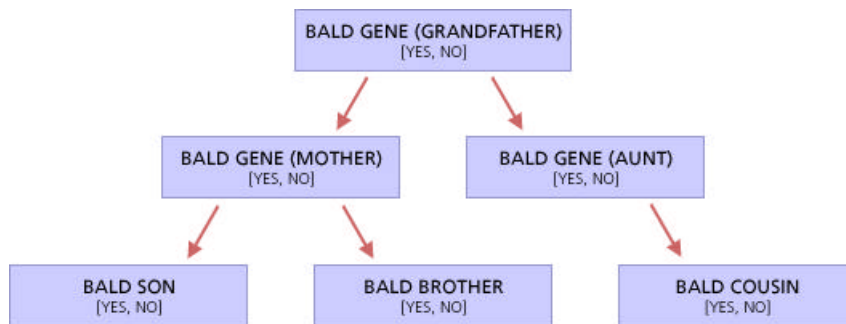


FIGURE 6000-1

In the graph of baldness heritability above, **BALD SON** and **BALD COUSIN** are predicted to be unconditionally associated, but they are predicted to be independent conditional on any set containing **BALD GENE** (Grandfather), **BALD GENE** (Mother), or **BALD GENE** (Aunt).

In general, conditioning on any variable that is on a causal connection blocks the causal connection from producing association. What sort of variable appears on a causal connection? A non-collider. What sort do not? A collider, or common effect. What happens when we condition on a common effect?

Consider a causal system that concerns starting your car in the morning.

TABLE 6000-1: VARIABLES FOR FOR STARTING A CAR

Variables	Values
BATTERY	[Charged, Uncharged]
CAR STARTS	[Yes, No]
GAS TANK	[Empty, Not empty]
EXHAUST PIPE	[Emitting, Not emitting]

Lets assume that the following graph depicts the causal relations among the first three variables.

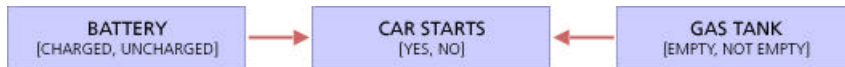


FIGURE 6000-2

< A link to exercises in the interactive version of this module. >

BATTERY and **GAS TANK** are predicted to be unconditionally independent because there is no causal connection between them. The frequency of charged batteries is the same as the frequency of charged batteries given the gas tank is empty. Are they independent conditional on **CAR STARTS**, however?

< A link to exercises in the interactive version of this module. >

Knowing nothing about whether the car started, informing us that the gas tank isn't empty tells us nothing about whether the battery is charged, i.e., **BATTERY** and **GAS TANK** are unconditionally independent. But, once we know that the car won't start, telling us that the gas tank isn't empty tells us the battery must not be charged! So **BATTERY** and **GAS TANK** are conditionally associated given their common effect: **CAR STARTS**!

In general, if **X** and **Y** are direct causes of **Z**, then **X** and **Y** are conditionally associated given **Z**, i.e., $X \not\perp Y \mid Z$

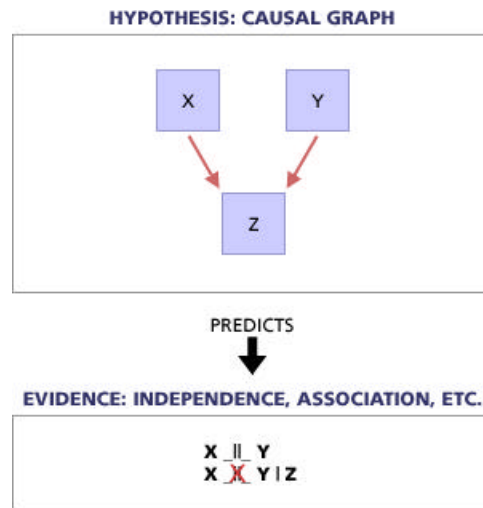


FIGURE 6000-3

Conditioning on common effects extends beyond the direct case. Consider the following graph, involving all four variables we listed as relevant to starting your car in the morning.

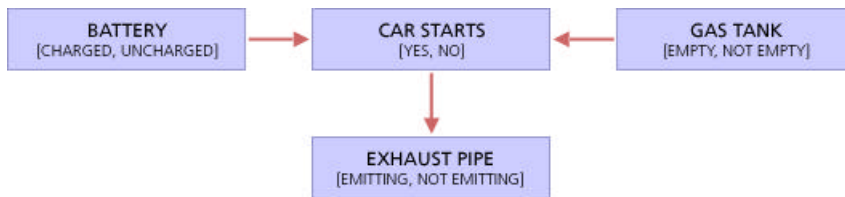


FIGURE 6000-4

We know that **BATTERY** and **GAS TANK** are associated conditional on **CAR STARTS**, because once we know that the car won't start, telling us that it wasn't the battery tells us it must be the gas tank, or vice versa. Similar reasoning works for the **EXHAUST PIPE**. If we know that the exhaust pipe is not emitting, then we know the car hasn't started. So if we are told that the the exhaust pipe is not emitting any gas, and we are told that the battery is charged, then we know that the gas tank must be empty.

In general, if **X** and **Y** are direct or indirect causes of every variable in **Z**, then **X** and **Y** are conditionally associated given **Z**, i.e., $X \not\perp\!\!\!\perp Y \mid Z$

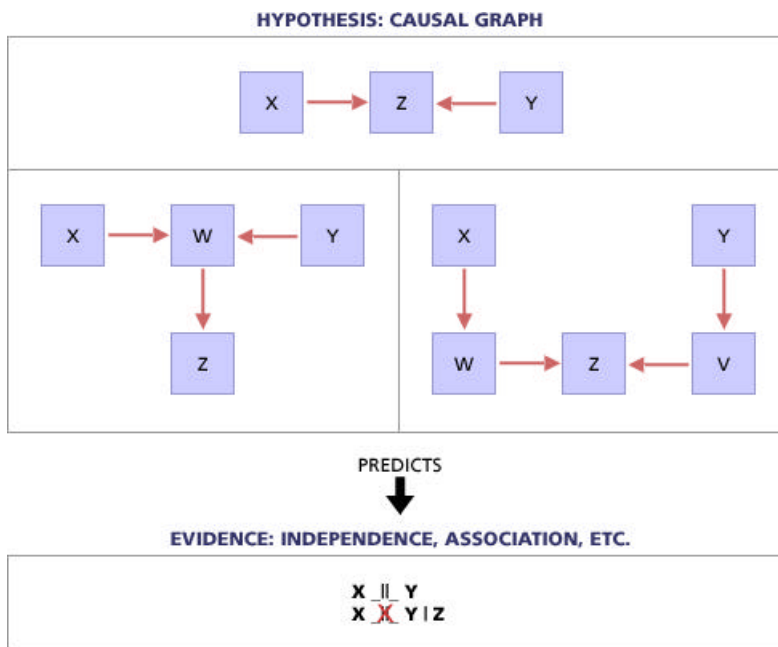


FIGURE 6000-5

7000: Summary

Causal Connection

Every causal connection between **X** and **Y** produces association. Thus, **X** and **Y** are unconditionally associated just in case they are causally connected. **X** and **Y** are causally connected just in case either **X** is a cause of **Y**, **Y** is a cause of **X**, or there is a common cause of **X** and **Y**. Every variable on a causal connection is a non-collider

Blocking Causal Connections

Conditioning on any variable on a causal connection blocks it from producing association.

Conditioning on Colliders (common effects)

Conditioning on only variables **Z** that are effects of both **X** and **Y** produces conditional association -- **X** and **Y** are associated conditional on **Z**.