8. Arguments for causation

A causal claim explains *why* something is true by making a statement linking a cause to an effect. Here are some examples of causal claims:

> The car crash was caused by the driver using his mobile phone while driving.
> Jane is anxious because she has two medical exams tomorrow.
> The main cause of criminal behaviour among adolescents is inadequate parenting.
> AIDS is caused by the HIV virus.
> A diet high in fresh fruit and vegetables prevents heart disease.

Explaining why things are the way they are is central to every scientific field from physics and chemistry to psychology and medicine and so causal claims can be found in every science. Understanding causes is also important in politics, economics and social policy. If you do not know what causes crime, poverty or economic recessions, you will not be able to formulate effective policies to combat them.

Causal claims also play an important role in personal decision making. For example, if you accepted the causal claim in the last example above and you were concerned to reduce your risk of heart disease, you would probably want to make sure your diet included lots of fruit and vegetables. And of course, many people have been motivated to quit smoking because they accepted the causal claim that smoking causes lung cancer and other serious health problems. To give a different kind of example, many people attempt to treat various illnesses using a variety of 'alternative' therapies or medicines, because they believe that such treatments can cure those illnesses – a belief about their causal effectiveness.

Clearly it is very useful to be able to distinguish causal claims that are reasonable to believe from those that are not. This is the topic of the current chapter. What evidence do we need in order to accept a causal claim? What kind of arguments can there be for thinking there is a causal connection between two things? We will look at two types of causal claim: *specific* causal claims and causal *generalisations*. In both cases, we begin by looking at arguments that are often thought to establish a causal connection but which do not; these are 'fallacies of causation'. We then go on to look at what counts as *good* evidence for a causal claim.

Types of causal claim: specific and general

Causal claims are of two main kinds. Some are *specific* – they tell us that one particular event caused another particular event. The first two examples above are specific: the first tells us about the cause of a particular car crash, the second tells us the cause of Jane's anxiety on this particular occasion.

Other causal claims are *general* – they tell us that events of one kind tend to cause events of some other kind. The remaining examples above are all causal generalisations. They do not tell us the cause of any specific event. Instead they make a claim that is intended to cover a wide range of specific events.

Causal generalisations can be tricky to think about. Consider the claim that smoking causes lung cancer. This is a causal generalisation. Does it mean that *everyone* who smokes invariably gets lung cancer? No, it does not. Does it mean that everyone who gets lung cancer is a smoker? Again, no. You cannot refute the claim that smoking causes lung cancer by pointing out that some people who smoke never develop lung cancer, though that is certainly true. Nor can you refute it by pointing out that some people with lung cancer have never smoked, though that is also true.

What the claim means is that smoking *increases your risk* of developing lung cancer. Think of it like

this. Your risk of developing lung cancer is like the roll of a dice. If you roll a 1, you will develop lung cancer, otherwise you will not. If you start smoking however, that changes the game. If you smoke, then you will develop lung cancer if you roll a 1, 2 or 3. Smoking has increased your risk of developing lung cancer from 17% to 50%. Of course, those figures are not supposed to be accurate – they are just intended to illustrate the idea.

This point can be put like this. To say that *A causes B* does not mean that *A* is *sufficient* for B and it does not mean that *A* is *necessary* for *B*. *A* might not be sufficient for *B* because even though *A* plays a role, other conditions have to be met too. For example, striking a match causes it to light. But it is not true that whenever you strike a match it will light. In order for the match to light it must also be dry and there needs to be sufficient oxygen in the room. So striking the match is not by itself sufficient for the effect.

A might not be *necessary* for *B* because *A* need not be the *only* way for *B* to happen. Consider the example above about a diet high in fruit and vegetables. The claim is that this will reduce the risk of heart disease. That might be true even though that kind of diet is not the *only* way to reduce your risk of heart disease. Regular exercise might also reduce that risk. A cause does not have to be the only way for an effect to occur.

'Post hoc' reasoning: a very common fallacy

Causes always come before effects in time: if *A* caused *B*, then *A* must have happened first and *B* later.[1] So you can refute a causal claim by pointing out that *A* cannot have caused *B*, because *B* was already true, before *A* happened. A very common mistake in reasoning about causes is to suppose that the converse is also true: that if *A* happened before *B*, then *A* caused *B*. Here is an example of this fallacy:

> Mary spent all day yesterday painting her garden fence. Today she has a rash on her hands. So she must be allergic to some of the chemicals in the paint.

The conclusion is that the rash Mary has today was caused by the paint she used yesterday (notice that this is a specific, rather than general causal claim). But although it is natural to assume that the paint and the rash are connected, the mere fact that the rash followed after using the paint is not sufficient to show there is a causal connection. There are many other possible explanations for the rash which have nothing to do with the paint. Perhaps Mary was stung by an insect while she was painting the fence. (Try to think of some more possibilities: it's a useful exercise). We do not have good evidence for a causal link between the painting and the rash until we can rule out these other possible explanations.

One name for this fallacious form of reasoning is *post hoc ergo propter hoc* which is Latin for 'after it, therefore because of it'. It is probably the most common mistake in ordinary reasoning about specific causal claims.

The 'post hoc' fallacy

1. *A* happened and then *B* happened later
Therefore:
C. *A* caused *B*

This form of argument is fallacious because it does not rule out alternative explanations for *B*.

[1] If time travel is possible then an effect might happen before the event that caused it. But we can safely ignore this possibility here: it is not relevant to ordinary reasoning about causes either in science or everyday life.

The fallacy can be made vivid by thinking of examples where we already know that *A* does not cause *B*, even though *A* happened and then *B* happened later. For example, your alarm clock might go off just before the sun rises, but the alarm did not *cause* the sun to rise. This should all be very obvious and yet it is *very* natural to assume that if you do something with the aim of achieving a goal and then *do* achieve you goal, it must be that your action helped achieve that goal. But the mere succession of events is not sufficient to show this. You have to rule out other possible explanations of your success.

Because of its naturalness, 'post hoc' reasoning is very common and difficult to avoid. Consider this kind of example:

> The government's economic stimulus program is clearly working to reduce unemployment. Since its introduction, the unemployment rate has fallen by 15%.

This seems quite a convincing argument until you realise all we have been told is that *A* happened (the government introduced its policy) and then *B* happened later (unemployment fell). It *could* be that the fall in unemployment had something to do with the policy, but then again, the fall in unemployment might have happened anyway, for all sorts of other reasons. We have not been given sufficient evidence to establish a causal link. It is in fact very difficult to get good evidence for the success of a policy, but it is possible. What we need is some additional evidence that rules out other possible causes.

As an exercise, try to think of at least two examples where *you yourself* have reasoned in this fallacious way: *A* happened and then *B* happened afterwards, so *A* caused *B*. You should find it quite easy because (like all good fallacies) it is such a tempting and natural way to reason.

Arguments for specific causal claims

What does a good argument for a causal claim look like? It would consider *all* the likely explanations and then eliminate all but one of them. Here is an example:

> The lemon tree in my garden is dying. The most likely causes are insufficient water, poor soil or a disease. But the tree gets plenty of water and the soil is well suited for citrus trees. So it is probably a disease that is killing my lemon tree.

The first premise tells us that the most likely cause is either *A*, *B* or *C*. The second premise tells us that cause is not *A* and not *B*. It follows that *C* is the most likely cause. The premises in this argument provide very strong support for the conclusion: *if* they are true, we would have a sufficient reason to accept the conclusion.

But of course, the premises might not all be true. Firstly, the list of likely causes might not be complete; perhaps the lemon tree is dying because it is not getting enough sun or because its roots have been damaged. Or one of the other premises might be false; perhaps the tree is not really getting enough water for example. The premises in this argument would likely be supported by further evidence (sub-arguments). For example, the claim that the 'soil is well suited for citrus trees' might be supported by evidence from a chemical test of the soil. But again, the test might not be completely reliable or might not have been carried out correctly.

Nevertheless, this kind of argument certainly can be sufficient to establish beyond a reasonable doubt the cause of some event, provided that all the likely causes have been considered and the evidence which rules out all but one is sufficient. Consider again the case of Mary and her rash. If we added to that argument evidence that ruled out all the likely alternative explanations for the rash (such as an insect bite and so on) we would have a much stronger argument in support of the

conclusion that it was something in the paint that caused the rash.

The general pattern of argument involved could be represented like this:

1. The possible/most likely causes of X are A, B, C, D …
2. B is **not** the cause because …
3. C is **not** the cause because …
4. D is **not** the cause because …

…
Therefore:
C. The cause of X is A

This is exactly the kind of careful and painstaking reasoning that an accident investigator might go through in order to establish the cause of an accident; a coroner might go through in order to establish the cause of a person's death; a doctor might go through in order to diagnose a patient; or a mechanic or software engineer might go through in order to determine the cause of an engine problem or software bug.

For the argument to be strong, each premise would have to be backed up by sufficient evidence. The list of possible causes (premise 1) might be justified by previous experience, expert opinion or well supported theories about events similar to X. Possible causes can be eliminated by a variety of different arguments. For example, given that causes precede effects, one might eliminate a possible cause of X by establishing that it actually came *after* the event X. For example, I might argue that the recent cold weather is not the cause of my lemon tree's problems because the tree was already dying *before* the cold weather came along. Or one might appeal to arguments like this; if C were the cause, then S would be true. But S is not true, so C cannot be the cause. For example; if poor soil was the problem with my lemon tree, then the lime tree growing next to it in the same soil would also be dying. But the lime tree is not dying, so poor soil is not the problem.

Causal generalisations and correlations

We turn now to *causal generalisations*. These are claims such as 'smoking causes lung cancer' which do not tell us about the causes of a specific event, but instead tell us that in general, events or actions of type *A* are a causal factor in producing outcomes of type *B*.

One kind of evidence that is relevant, but as we shall see, *not* sufficient to establish a causal link between events of type *A* and outcomes of type *B* is that *A* and *B* are **correlated** (or 'associated') with each other. What does this mean? Here is a definition.

> **Correlation between attributes**
>
> There is a (positive) **correlation** between the attributes *A* and *B* when *B* is *more likely* if *A* is present than if *A* is not present.
>
> Another way of putting it is as follows:
>
> There is a (positive) correlation between *A* and *B* when the proportion of *A*s which are *B* is *greater* than the proportion of non-*A*s which are *B*.

For example, to say that there is a correlation between smoking and lung cancer is to say that lung cancer is *more likely* if a person smokes than if they do not smoke. The phrase 'more likely' refers *only* to probability or frequency. So another way of putting it is just to say that the *proportion* of smokers with lung cancer is higher than the proportion of non-smokers with lung cancer. If

smoking caused lung cancer, such a correlation is exactly what you'd expect to see, which is why correlation is relevant evidence for causation. But, as we shall see below, a correlation between *A* and *B* is not sufficient to show that *A* causes *B* because the correlation might have some other explanation.

What we have defined above is the concept of a **positive** correlation between *A* and *B*. We can also define the idea of a **negative** correlation.

Negative correlation between attributes

There is a **negative correlation** between *A* and *B* when *B* is *less likely* to occur if *A* is present than if *A* is not present.

Another way of putting it is as follows:

There is a negative correlation between *A* and *B* when the proportion of *A*s which are *B* is *smaller* than the proportion of non-*A*s which are *B*.

For example, to say that there is a negative correlation between a diet high in fresh fruit and vegetables and heart disease is to say that a person is *less likely* to develop heart disease if they have such a diet than if they do not. That *could* be because there is a causal link between diet and risk of developing heart disease, but again, it might not be: there could be other explanations for the correlation.

Correlations between quantities

In some cases, the variables *A* and *B* we are interested in do not represent attributes that an individual either has or does not have, but *numerical quantities*. Examples are a person's height or weight, their annual income, the number of mobile phones they own and so on. What does it mean to say that two numerical variables are correlated? Here is a definition:

Correlation between quantities

There is a **positive correlation** between the quantities *A* and *B* when individuals with higher values of *A* are more likely to also have higher values of *B*.

There is a **negative correlation** between the quantities *A* and *B* when individuals with higher values of *A* are more likely to have lower values of *B*.

For example, there is a positive correlation between height and weight. Suppose you picked two people at random and compared their heights. If the first person happens to be *taller* than the second person, then the first person is also *more likely* to be heavier than the second. This correlation is not perfect of course. Sometimes the taller person will not be heavier than the shorter person. To say that there is a positive correlation between height and weight is just to say that *on average*, it is more likely that the taller person will also be heavier.

Correlations between quantities are often represented on charts like the one shown below (Figure 1). Each red dot represents a person (this is real data by the way, not made up). The person's height (in inches) is represented on the horizontal axis – the further a dot is over to the right, the taller the person. Weight (in pounds) is represented on the vertical axis – the higher up a dot is, the heavier the person is. You can see from the chart that on average, the taller a person is, the more they weigh. You can also see that the correlation is not perfect. The dots do not form a perfect

straight line, but appear to cluster around an imaginary straight line (shown in blue). It is not always true that the taller of two people is also heavier, but it is true *most* of the time.

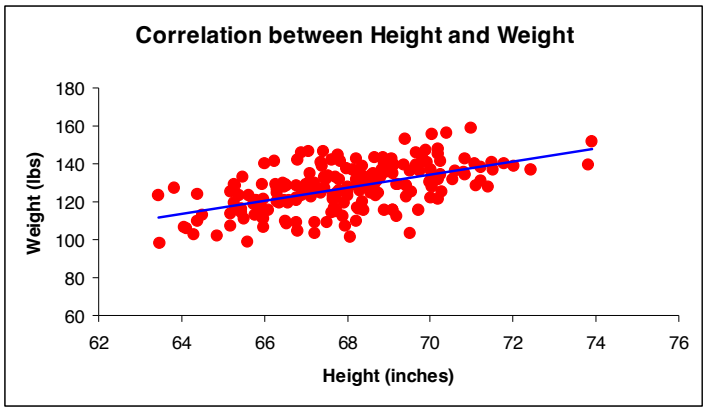

Figure 1

Strength of a correlation

Correlations between quantities and attributes come in *degrees of strength*. We have said that two attributes *A* and *B* are correlated if an individual is more likely to be *B* if they have attribute *A* than if they do not. *How much* more likely? A lot more likely or only a little? The answer to that question measures the strength of the correlation. Likewise, two quantities *A* and *B* are correlated if, individuals with higher values of *A* are more likely to also have higher values of *B*. Again, the strength of the correlation is a matter of how much more likely that is.

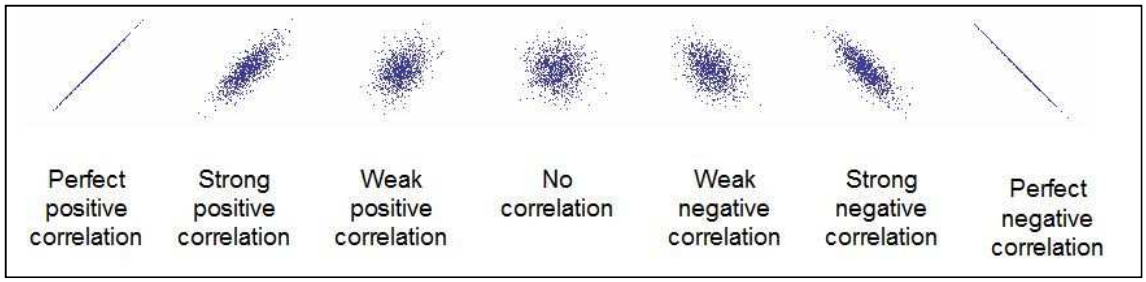

Figure 2

Illusory correlations

People often see correlations where no real correlation exists. Suppose that on many (or even most) occasions when *A* is present, *B* is also present. Does that mean there is a correlation between *A* and *B*? It does not. We also need to know something else - how often *B* is present when *A* is *not* present.

Suppose you have a friend who believes in the power of 'positive thinking'. "On many occasions," they tell you, "I have visualised myself achieving a goal before attempting it and then I've been successful." You might want to say to them, "OK, but how many times have you visualised achieving a goal but *not* been successful?" If success is just as likely when visualisation does not take place as when it does, then there is no correlation between the two.

Suppose your friend said that, "Well actually, I have been keeping records. On 50 different occasions I visualised myself succeeding. In 40 of those cases I did succeed; I failed in only 10". We can represent this information in a table, like this:

	Success	No success
Visualised achieving the goal	40	10

What we know from this table is that the likelihood of success is greater than the likelihood of failure when your friend visualises themselves succeeding. When they visualise success, they are more likely to succeed than fail. But even this is *not enough* to show that visualisation is correlated with success. We also need to know what happens when your friend does *not* visualise themselves succeeding. For there to be a correlation success must be more likely when the goal is visualised than when it is *not* visualised. (That is just the definition of correlation.) So we need to fill in the rest of the table:

	Success	No success
Visualised achieving the goal	40	10
Did not visualise achieving the goal	?	?

Suppose that on another 50 occasions, your friend did not visualise achieving the goal. In 45 of those cases, they succeeded and in only 5 did they fail. So the table looks like this:

	Success	No success
Visualised achieving the goal	40	10
Did not visualise achieving the goal	40	10

What we now know is that your friend is equally likely to succeed in their goals, whether they visualise themselves succeeding in advance or not. So their visualising themselves achieving their goal seems to have nothing to do with their success – visualisation and success are not correlated. On the other hand, if the table looked like this

	Success	No success
Visualised achieving the goal	40	10
Did not visualise achieving the goal	20	30

then there *would* be a positive correlation between visualisation and success. In this table, the frequency of success when your friend visualises is 80%, as before. But now the frequency of success when they do *not* visualise is only 40%. So they are more likely to succeed if they visualise than if they do not, at least in this sample of cases.

People have a tendency to decide whether *A* and *B* are correlated just by looking at what happens when *A* occurs. In effect they focus just on the top row of a table like the one above, or even just on the top left hand corner. If *A* and *B* happen together a lot, people tend to think that there is a correlation between them. They ignore or fail to see the relevance of the bottom row – what happens when *A* is not present.

For example, someone might believe that God answers prayers because on many occasions they have prayed and then got what they wanted. Or someone might believe that drivers of red cars drive more recklessly because they have often noticed red cars being driven recklessly. Or someone might believe that Geminis are more sympathetic than people of other star signs because most of the Geminis they know are sympathetic. And so on.

Correlation is not sufficient for causation

Even when there really is a correlation between *A* and *B* that does not necessarily mean that *A* is a *cause* of *B*. That is one possibility, but there are always other possible explanations for a correlation.

Consider the following example. People who use artificial sweeteners instead of sugar are more likely to be overweight than people who just use sugar. That is to say that there is a correlation between using artificial sweeteners and being overweight. Should you conclude that using artificial sweeteners somehow *causes* people to gain weight? A more likely explanation for this correlation is that people who are already overweight are more likely to use artificial sweeteners because they hope it will help them lose weight. If so, then the causal link is from being overweight to using artificial sweeteners, rather than the other way around.

Or consider another observed correlation: people who regularly attend church services tend to live longer than people who do not. Does this mean that going to church regularly *causes* longer life? Not necessarily. An alternative explanation is that people who attend regular church services are more likely (for religious reasons) to take better care of their health. They are more likely to avoid alcohol and cigarettes for example. It is these things that are causally responsible for their longer life, rather than attending church services.

In this example, although *A* (going to church regularly) is correlated with *B* (longer life) there is no causal link between *A* and *B* in either direction. *A* does not cause *B* and *B* does not cause *A* either. Instead, what's true is that regular church attendance is correlated with a third variable *X* (looking after your health) and it is *X* that causes *B*. The technical term for this, which you might come across, is **confounding**. We say that going to church regularly is 'confounded' with taking better care of your health and that is why we would not be justified in concluding that going to church causes longer life.

So although correlation is necessary for causation, it is not sufficient. It is worth emphasising this point because it is very important

Correlation alone is not sufficient to establish causation

Arguments that infer a causal claim from premises that only tell us about correlation are therefore fallacious. For example:

> **1.** People who use artificial sweeteners instead of sugar are more likely to be overweight than people who just use sugar.
> **Therefore:**
> **C.** Using artificial sweeteners can lead to weight gain.

The conclusion here is a causal claim – it says that artificial sweeteners can *cause* ('lead to') weight gain. But the premise only states a correlation and we have already seen that this correlation might be explained in other ways. So the premise does not support the conclusion. In order to make the argument cogent, we would need to find evidence that counts against the alternative explanations.

In this context, beware of the weasel word *linked* – it is ambiguous between 'causally linked' and 'correlated'. Suppose you read that "Researchers have discovered a **link** between insecticides and the collapse of honey bee colonies." This is lazy reporting. Have the researchers found that the two are causally linked, or only that there is a correlation? You cannot tell until you find out more about the research.

Possible explanations for a correlation

In general, if *A* and *B* are correlated, there are always the following possibilities to consider (see Figure 3 for a visual representation of these possibilities):

1. *A* and *B* are correlated because *A* causes *B*
2. *A* and *B* are correlated because *B* causes *A*
3. *A* and *B* are correlated because a third factor *X* causes both *A* and *B* (common cause)
4. *A* and *B* are correlated because a third factor *X* is correlated with *A* and *X* causes *B* (confounding)
5. The correlation is a coincidence – just due to chance.

Therefore, to establish that *A* causes *B* we need to first check to see if there is a correlation between *A* and *B*. If there is no real correlation between them, then *A* does not cause *B*. But if there is a correlation, we have to do more. We have to somehow rule out the alternatives 2-5. This can be done either by conducting some kind of **controlled experiment** or by looking for other evidence that can rule out the alternative possibilities. We will look at this in more detail below.

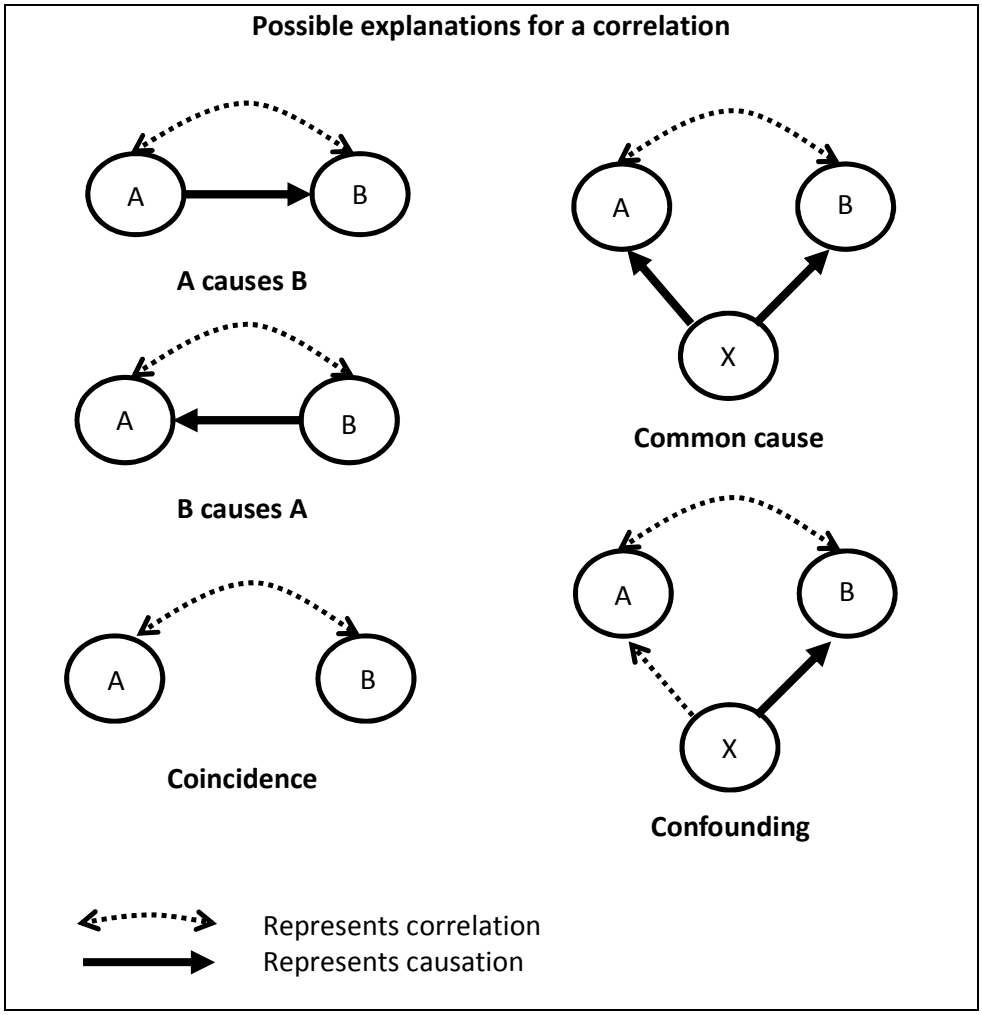

Figure 3

Causation and chance

Before getting to that it is worth spending some time thinking about the last possibility– that the correlation is just due to chance. This is the possibility most often ignored in everyday reasoning about causes. Consider again the visualisation example. Suppose your friend's record of success and failure looked like this:

	Success	No success
Visualised achieving the goal	2	1
Did not visualise achieving the goal	0	3

That is, on three occasions they visualised the goal and were successful twice and on a different three occasions they did not visualise the goal and they failed every time. This data shows a correlation between visualisation and success – your friend had a higher rate of success when visualising than when not visualising.

Nevertheless, since we only have six trials in our sample we cannot rule out *pure chance* or *blind luck* as a possible explanation. The sample size is just too small. Suppose what actually happened is this. As an experiment your friend tossed a coin six times, but before doing so decided that heads meant 'success' and tails meant 'failure'. For the first three tosses of the coin, they visualised the coin landing heads. Lo and behold it landed heads twice! For the next three tosses, they did not visualise the coin landing heads. And this time the coin landed tails three times in a row. All this might have happened just by chance – the outcome is not particular unlikely or improbable. On the other hand, if we had obtained a strong correlation with a sample of 1000 coin tosses that would be a different matter: then we could rule out chance as a possible explanation.

As you might guess, it is possible to make this a bit more rigorous using probability theory and statistics. You may see the term **statistically significant** used in this context. What it means to say that an outcome is 'statistically significant' is just that a calculation has been done which rules out pure chance as a possible explanation of the outcome. The details of such calculations are not important here. What matters is the way the calculation is used to rule out chance as a possible explanation. The standard procedure is to pick a 'cut-off' probability in advance: 1 in 20 or 5% is the standard choice. Then the calculation (which depends on the sample size and various other relevant factors) is carried out to determine the probability of the outcome. That is, you calculate how probable was this event, assuming it was just due to chance? If the result is less than the cut-off probability – less than 1 in 20 – then the outcome is classified as 'statistically significant'. On this (admittedly somewhat arbitrary) basis, chance is ruled out as a possible explanation for the outcome.

Ignoring the possibility of chance

Here is an interesting fact. A study of the incidence of kidney cancer in the 3,131 counties of the United States reveals a remarkable pattern. The counties in which the incidence of kidney cancer is the lowest are mostly rural, sparsely populated, and located in traditionally republican states in the Midwest, the South and the West. What might explain this pattern?

Lots of possible explanations typically occur to people. For instance, it might be that people in those kind of counties tend to have healthier lifestyles or better diets. The fact that the counties are mostly rural seems especially relevant.

Did you notice the part about the counties being *sparsely populated*? Most people do not notice that and do not consider it relevant. Nevertheless, it points to the most likely explanation of the pattern, which is *chance*. Remember the sample size principle: **atypical results are more likely with**

smaller samples. So if I take a small sample of people, an atypical incidence of kidney cancer (or anything else) is more likely. And that's exactly what we have in this case. Since the counties are sparsely populated, they have fewer people. So we would expect to see cancer rates which are lower than the national average more often in these smaller counties.

"But wait!" you might want to interject at this point. Why is the incidence of kidney cancer *lower* in these small counties? Wouldn't we also expect to see *higher* than average rates of cancer in small counties too, by the same reasoning? That's very clever of you and the answer is yes, we would. And that is exactly what we *do* find. For remarkably, it is also true that the counties in which the incidence of kidney cancer is the *highest* are mostly rural, sparsely populated, and located in traditionally republican states in the Midwest, the South and the West. Again, this is just what we would expect given the sample size principle: atypical (significantly above or below average) results are more likely with smaller samples.

We are all quite good at thinking of causal explanations for patterns that we observe; it is a natural human tendency. What is *not* natural is to consider the possibility that the pattern is a pure coincidence and has no causal explanation at all. Nevertheless, chance is an ever present possibility that must be taken into account if we want to explain and understand the way the world works.

Controlled experiments

A correlation between *A* and *B* is necessary, but not sufficient to establish that *A* causes *B*. If *A* causes *B*, there will certainly be a correlation between them, but a correlation between *A* and *B* is not always the result of a causal link between *A* and *B*.

This raises an important question; given that there is a correlation between *A* and *B*, what else do we need to get good evidence for causation? The answer is that *we need to rule out all the alternative explanations of the correlation.* How can this be done?

The strongest type of evidence for causation comes from a **controlled experiment.** In a controlled experiment, the experimenter makes changes in one variable *A*, *while holding all other relevant factors constant*. If *B* changes, this provides a strong case for thinking that *A* causes *B*. That's because holding all the possible relevant factors constant rules them out alternative explanations.

Consider again the correlation between church attendance (*A*) and longer life (*B*). One possible explanation is that people live longer because they attend church. But another explanation is that church attendance is correlated with a third variable (*X*) – taking better care of your health – and it is this that is causally responsible for longer life, rather than church attendance.

Suppose though that we conducted a controlled experiment. We would take two groups of people; one group that regularly attend church services and another group that does not. We would try to ensure that everyone in both groups takes equally good care of their health, so that the possible confounding variable *X* is kept constant. Then, if we find there is still a difference in lifespan between the two groups, we can rule out *X* out as an alternative explanation. It cannot be *X* that causes longer life because that was the same for both groups and yet there was still a difference in lifespan.

Randomized controlled trials

That's fine in theory, but the real problem is to work out what factors might be relevant and devise clever ways to keep them constant while just changing one other variable. Scientists have developed many techniques for this, perhaps the most sophisticated of which is the **randomized controlled trial**.

Suppose you are conducting a randomized controlled trial to determine whether *A* causes *B*. For example *A* might be a new medicine or therapy and *B* patients' health outcomes. Or *A* might be a new type of teaching method and *B* reading ability.

You form two groups of individuals and assign them *at random* to one of two groups; either a **treatment group** or a **control group**. In effect, for each individual you toss a fair coin; if the coin lands heads, you assign them to the treatment group, if it lands tails you assign them to the control group. This random assignment ensures that the two groups are equivalent at the start of the experiment.

You begin by measuring the variable *B* that you hope to change at the start of the experiment. Then you make some change in the variable *A* for the people in the treatment group. You give them the new medicine for example, or try out your new teaching method on them. The individuals in the control group do *not* receive the intervention, but you try to keep everything *else* the same for them, as far as possible. Afterwards you measure *B* again.

Now if *A* really does cause *B* to change, what you would expect to see is something like the situation shown in the diagram below. The outcome *B* will be different for both groups the second time you measure it (because measurements always vary) but the change in *B* will be *larger* for the treatment group than for the control group.

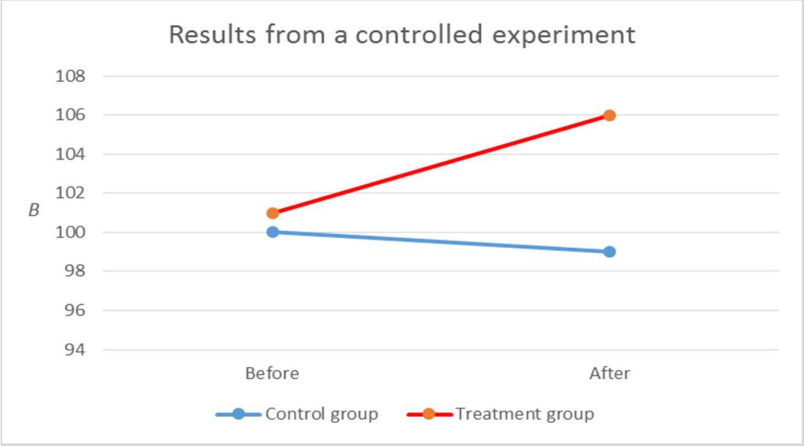

Figure 4

In the example above, we can see that *B* was approximately the same for both groups at the start of the experiment. At the end of the experiment *B* has increased for the treatment group (shown in red) but has not changed much for the control group (shown in blue). If the experiment was carried our properly, this is very good evidence that *A* causes *B*.

But how do we know the change in *B* was caused by *A* and not by any number of other possible factors? Age, gender, IQ, dietary habits, level of education and many other things are all possible variables that might influence *B*. By design, a randomized controlled trial can rule out these alternative explanations.

First, we know that all those potential confounding variables will be distributed in roughly the same way in both groups at the start of the experiment. How? Because we assigned people *at random* to the two groups. That means that each person has a fifty-fifty chance of being assigned to either the control group or the treatment group. So, for example, the number of males and females will be about the same in both groups. For the same reason, the number of people in each age group will also be about the same. In fact, the two groups will be similar with respect to *every* variable in the

population – even variables that you do not know anything about. This is guaranteed by the fact that every individual has an equal chance of ending up in either the control group or the treatment group.

Secondly, once the experiment begins, both groups are treated as far as possible in the same way, except for the fact that the treatment group receive the intervention. This is what the word 'controlled' refers to in the phrase 'randomized controlled trial'. You try to keep all other variables constant between the two groups. Therefore, if you see a larger change in *B* for the treatment group than for the control group, you can be quite sure this was due to *A* and not to any other variable. Why? Because all those other variables either did not change at all during the experiment or changed in the same way for both groups.

Summary

In a **randomized controlled trial** to determine if an intervention *A* causes a change in an outcome *B*, the procedure is as follows:

1. Compare two large groups of individuals, the **treatment group** and the **control group**.
2. Use **randomization** to ensure that the two groups are equivalent at the start of the experiment.
3. Measure *B* at the start of the experiment.
4. As far as possible, make sure that everything that happens to the treatment group also happens to the control group except that the treatment group receives the intervention *A*. (**Equal treatment** for both groups).
5. Measure *B* again at the end of the experiment and compare the two groups.
    - If *B* has changed in the treatment group more than it has changed in the control group, you have good evidence that *A* causes *B*.
    - If *B* has not changed more in the treatment group than in the control group, the experiment has not shown that *A* causes *B*.

Randomized double blind trials

Experimenters often go to great lengths to ensure step 4: that the treatment and control groups are treated in the same way, except for the fact that the treatment group receives the intervention. For example, in medical trials, the fascinating **placebo effect** needs to be taken into consideration. If a doctor in a white coat, wearing a stethoscope gives you a pill (or better yet, an injection) and tells you confidently that it will make you feel better, it probably will (at least for a while) even if the pill or injection has no active ingredient at all.

To control for this effect, people in both the control group and the treatment group should go through the same kind of process. Both should be given an identical looking pill by a doctor in a clinic for example. The difference will be that the pill does not contain the new medication for people in the control group. Instead it will contain something inert and harmless, but which looks and tastes about the same; a placebo pill. For this to work, it is crucial that none of the patients know whether they are in the control group or the treatment group until the end of the experiment. They must be "blind' to which group they are in.

But that is not all. If the *doctor* giving the person the pill knows whether the person is getting the real treatment or just the placebo pill they might inadvertently give this away. So it is also critical, in a properly conducted medical trial that the people administering the pills do not know whether they are giving out the placebo pill or the real pill either. That is, the doctors and nurses must also be 'blind' to which group the patient belongs to. A trial carried out under these stringent conditions

is therefore called a **double blind** randomized controlled trial. The phrase 'double blind' refers to the fact that neither the doctors nor the patients know until the end of the trial who is in the treatment group and who is in the control group.

Critical thinking about experiments

Researchers are not just interested in what caused what in a specific experiment – they want to know whether they can *generalize* the result of the experiment. For example, you would like to say that the vitamin supplement improves short term memory capacity in adults; not just that it happened to do so in your experiment. But caution is needed to generalize in this way, even from a very carefully conducted and successful experiment. You would be arguing like this: in a sample of 200 adults, we found that the vitamin supplement led to an improvement in short-term memory capacity; therefore the vitamin supplement would lead to a similar improvement in any group of adults. But whether this further inference is justified depends on whether your sample is representative of the general population you are interested in. Maybe the sample size is large, but how were the participants recruited into the experiment in the first place? If you repeated the experiment on a different group of individuals you might not find evidence for causation, or if you did, you might find that the difference between the two groups was smaller.

That is why researchers value **replication**. Before making a general inference about a population from a study, others should try to repeat the experiment on a different group of people and see if they get the same result. (This is also protects against researchers who fake or distort their results; sadly this does sometimes happen). There is a useful saying in research: *one positive study does not make a finding*. You might have found evidence for causation in your study, but you should be cautious about generalizing. That said, if many well conducted experiments on different groups of individuals all show evidence that this particular vitamin improves short-term memory, then it would be reasonable to accept that the results of the experiments can be generalized to the entire population.

Often what happens is that many experiments are done and some are carried out more carefully than others. Some of the experiments will show a big effect of the vitamin supplement, some will show a smaller effect and some will not find any evidence of causation at all – no effect. This is a situation ripe for **cherry-picking**. Someone who wants to endorse the theory that the vitamin supplement improves short-term memory can truthfully cite several experiments that show this. But someone who wants to deny it can equally truthfully cite several experiments that failed to find any effect or found only a very small effect.

Obviously in both cases, this is cheating. To come a rational conclusion, *all* the relevant studies, both positive and negative need to be taken into account. There are ways to combine the results of lots of different studies to come to an overall conclusion. This is a called a **meta-analysis.** In a meta-analysis, you try to find out about *all* the experiments or studies that have ever been done on a given topic and combine the results. This is not a simple matter of just averaging all the results however. You need to give more weight to studies which were more carefully conducted (better controlled) or which had larger samples for example. The **Cochrane Review** is an invaluable organization set up to do exactly this kind of work in the field of medical research. If you really want to know whether a diet high in fresh fruit and vegetables protects you against some forms of cancer for example, then the Cochrane Review website would be a good place to start. There will be hundreds of relevant studies on this, all with differing results. The Cochrane Review puts them all together and reports an overall evaluation of the claim. It has been an invaluable resource for medical science. In some cases its meta-analyses have found that a medical intervention that was thought to be beneficial was in fact quite dangerous. Careful review of *all* the evidence saves lives.

Non-experimental criteria for causation

Controlled experiments are not always possible. Consider the following causal hypotheses:

> The death penalty deters people from committing violent crimes.
> Living near electric power lines can cause cancer.
> The Measles, Mumps and Rubella vaccine causes autism.
> Global warming is caused by carbon dioxide in the atmosphere.

In these cases, we cannot, for practical or ethical reasons carry out a controlled experiment. We cannot force a selected group of people to go and live near electric power lines and compare their rates of cancer with an exactly similar group that we force to live well away from power lines. How then can we evaluate these claims? What kind of evidence would support or refute them?

Given a correlation between two variables *A* and *B*, what additional evidence helps to establish a causal link between *A* and *B*, when an experiment is not possible? Here are five additional criteria that add support to a causal hypothesis:

Non-experimental criteria for a causal link between *A* and *B*

1. The correlation between *A* and *B* is strong.
2. The correlation is consistent across different groups.
3. *A* comes before *B* in time.
4. Bigger differences in *A* are correlated with larger responses in *B*.
5. There is a plausible mechanism linking *A* to *B*.

As an example, consider the case of smoking and lung cancer. There is a known correlation between smoking and death from lung cancer. But there have been no controlled experiments on human beings establishing this causal link. Applying the above criteria though, we can establish that the evidence for a causal link is very strong indeed.

1. The correlation between smoking and lung cancer is very strong.
As noted above, statisticians have developed ways to measure the 'strength' of a correlation. We do not need to go into the details of how this is done. All we need to remember is that to say that *A* and *B* are correlated is to say that *B* is more likely if *A* is present than if it is not present. A measure of the *strength* of a correlation is simply a way of answering the question *how much* more likely? Measures of the strength of a correlation are usually given as a number between zero and one. Zero means 'no correlation', while 1 refers to a 'perfect' correlation. In the case of smoking and lung cancer, the correlation is not perfect, but it is very strong – typical values range from 0.91 to 0.98. That means you *a lot* more likely to develop lung cancer if you smoke than if you do not.

2. The correlation between smoking and lung cancer is consistent
Many different studies on different groups of people have found the same strong correlation between smoking and lung cancer. For example, the correlation is found across all age groups and in both males and females. It is found in groups from many different countries. That makes it unlikely that the correlation is due to chance or to some unknown common characteristic of the people studied.

3. Smoking comes before lung cancer
Lung cancer develops after years of smoking. The number of men dying of lung cancer rose as smoking became more common with a time-lag of about 30 years. Lung cancer was rare among women until women began to smoke. The incidence of lung cancer in women rose along with

smoking, again with a time-lag of about 30 years. This allows us to rule out an alternative hypothesis: that lung cancer somehow causes smoking.

4. Bigger differences in *A* are correlated with larger responses in *B*.
People who smoke more cigarettes per day or who smoke over a longer period are more likely to get lung cancer. People who stop smoking are less likely to get lung cancer.

5. There is a plausible mechanism linking smoking to lung cancer
Controlled experiments with animals show that tars from cigarette smoke cause cancer in those animals. So it is plausible that smoking causes lung cancer in much the same way in human beings.

All of this adds up to a very strong case that smoking causes lung cancer. Of course this does not mean that everyone who smokes gets lung cancer. And it does not mean that everyone who gets lung cancer is a smoker; smoking is not the *only* thing that can cause lung cancer. What this evidence suggests is that smoking is a significant causal factor in developing lung cancer, in the sense that smoking dramatically increases your risk of getting lung cancer. You are about 25 times more likely to develop lung cancer if you smoke than if you do not.

The five criteria set out above are sometimes known as *Hill's criteria* after the scientist Bradford Hill, who proposed them in epidemiology (the study of the spread of disease) for determining if a particular toxin or biological agent causes a disease. Hill's original set also included some specifically biological criteria which need not concern us here. The five listed above apply quite generally to any claim of the form '*A* causes *B*' not just medical ones. For example, you might consider how they apply to the hypothesis that carbon dioxide and other greenhouse gases released into the air by human activity are the cause of global warming.

Further reading

For more on evaluating causal claims and arguments (as well as further exercises on this topic) see:

> Alec Fisher: *Critical Thinking: an introduction* (2nd edition), Chapter 10, pp. 138-153.
>
> Jill LeBlanc, *Thinking Clearly*, chapter 10, pp. 270-297.

For more on the topic of causation and chance, see:

> Daniel Kahneman: *Thinking, fast and slow*, chapter 10 ('The law of small numbers'), pp. 109-118.

This wonderful book summarises in an accessible way, decades of fascinating research by two cognitive scientists (Daniel Kahneman and Amos Tversky) on the system biases in human thinking and decision making. If you are interested in learning more about cognitive biases and the 'dual process' or 'two systems' theory that has been developed to explain them, this is an excellent place to start. In relation to the topic of reasoning about causes chapters 16 ('causes trump statistics') and chapter 17 ('Regression to the mean') are also worth reading.

For more on how people are prone to see correlations where none exist ('illusory correlations') see:

Jonathan Baron, *Thinking and Deciding* (3rd edition), chapter 8 ('Judgement of correlation and contingency'), pp. 173-190.

Exercise 8.1

For each of the following arguments, say whether the conclusion makes a **causal claim**. If it does, how strong is the evidence for causation? Explain your answer.

(1) This herbal remedy helps you heal faster. This morning I cut my finger. I took the herbal remedy straightaway and now, just a few hours later, my finger has almost stopped bleeding!

(2) In August this year we began a campaign to get the government to abandon the proposed desalination plant. Today the government has announced that the plant will not go ahead. Our campaign was successful.

(3) The last five times I've been to this restaurant the food has been excellent. This time, the food is only average. They must have a new chef.

(4) A large scale study compared 250 schools across Australia, selected at random from a wide range of different socio-economic areas. It found that the schools which have the highest average student grades tend to be the smaller schools; the schools in the top 5% for student achievement are in the bottom 5% for size. Therefore, we should encourage smaller schools if we want to improve student performance.

(5) Signs posted around a university campus: "Improve your marks. Go to PASS. Students who attend more PASS sessions (Peer Assisted Study Sessions) get better marks."

(6) During the three months before and the three months after a major earthquake in California, students at a college there happened to be keeping a record of their dreams. After experiencing the earthquake, half of the students reported dreaming about earthquakes. During the same six months, a group of college students in Ontario who had never experienced an earthquake also recorded their dreams. Almost none of the students in Ontario reported dreaming about earthquakes. So it is clear that experiencing an earthquake can cause people to dream about earthquakes.

(7) Intelligent children are more likely to become vegetarians later in life, a study says. A Southampton University team found those who were vegetarian by 30 had recorded five IQ points more on average at the age of 10. The study of 8,179 was reported in the *British Medical Journal*. Twenty years after the IQ tests were carried out in 1970, 366 of the participants said they were vegetarian - although more than 100 reported eating either fish or chicken. Men who were vegetarian had an IQ score of 106, compared with 101 for non-vegetarians; while female vegetarians averaged 104, compared with 99 for non-vegetarians. (BBC news report, December 2006)

(8) Since the approximate number sense is essential for survival, it might be thought that all humans would have comparable abilities. In a 2008 paper, psychologists at John Hopkins University and the Kennedy Krieger Institute investigated whether or not this was the case among a group of 14-year-olds. The teenagers were shown varying numbers of yellow and blue dots together on a screen for 0.2 seconds and asked only whether there were more blue or yellow dots. The results astonished the researchers, since the scores showed an unexpectedly wide variation in performance. Some pupils could easily tell the difference between 9 blue dots and 10 yellow, but others had abilities

comparable to those of infants – hardly even able to say if 5 yellow dots beat 3 blue. (From Alex Bellos, *Alex's Adventures in Numberland*)

(9) An even more startling finding became apparent when the teenagers' dot-comparing scores were then compared to their maths scores in kindergarten. This study found a strong correlation between a talent at reckoning and success in formal maths. The better one's approximate number sense, the higher one's chance of getting good grades. This might have serious consequences for education. If a flair for estimation fosters mathematical aptitude, maybe maths classes should be less about time tables and more about honing skills at comparing sets of dots. (From Alex Bellos, *Alex's Adventures in Numberland*)

(10) A study investigated the association between hyperactivity and how much sugary food there is in a child's diet. Here is a table showing the results.

		Greater than recommended sugary food in child's diet	
		Yes	**No**
Hyperactivity	**Yes**	250	50
	No	50	10

Did this study find a correlation between sugar food in a child's diet and hyperactivity? If so, does that imply there is a causal link between the two? If not, what alternative explanations might there be of the correlation?

Exercise 8.2

The following examples are adapted from the book *How to think about weird things: critical thinking for a New Age*, T. Schick and L. Vaughn, McGraw Hill, 2002. In each case, some evidence is reported regarding a particular treatment or diet. In each case, say whether there is sufficient evidence for causation and explain your answer.

(1) Therapeutic Touch (TT) is an alternative medicine technique said to be used by over 40,000 nurses in North America alone. It is supported by major nursing organizations and gets favourable mention in the media. TT practitioners claim that an "energy field" unknown to science surrounds the human body and that practitioners can use their hands to detect and manipulate this field. (No physical touching of the body is involved, though.) In particular, they say that they can cure disease by "balancing" people's fields, which are said to extend four to eight inches from the surface of the skin. Many people do in fact report feeling better after TT treatment.

(2) A macrobiotic diet can be effective against cancer. Macrobiotics is a lifestyle and diet derived from Far Eastern ideas and promoted by many adherents. The diet is semi-vegetarian and low in fat. In recent years there have been many published accounts of people who say they have recovered from cancer because they ate a macrobiotic diet. There have also been several case reports. Attempts have been made to compare the outcomes of these cases to those of patients with comparable cases of cancer who did not follow a macrobiotic diet. These comparisons show that people on the macrobiotic diet often have better outcomes.

(3) Shark cartilage has been called to public attention by a CBS *60 Minutes* program focused on the theories of biochemist William I. Lane, Ph.D, author of *Sharks Don't Get Cancer*, Narrator Mike Wallace began by calling attention to the book and stating that Lane says that sharks don't get cancer. The program focused on a Cuban study of twenty nine "terminal" cancer patients who were given shark-cartilage preparations. Wallace visited the site of the experiment, filmed several of the patients doing exercise, and said that most of the patients felt better several weeks after the treatment had begun. Two American cancer specialists then said that the results were intriguing. One, who was aligned with the health-food industry, said that three of the patients appeared to have improved. The other, who appeared to be solidly scientific, noted that evaluation was difficult because many of the X-ray films were of poor quality, but he thought that a few tumours had gotten smaller.