Basic Influence Diagrams

Influence is a complicated concept, but in this section we assume it is simple.

Because we see so many things influencing other things in our daily lives, we believe that we understand what "influence" means. Eventually I am going to argue that this is not quite true. In fact, one of my chief reasons for writing this book is to try to explain how influence might be defined, so that scientists can use the concept in a consistent fashion. Happily, however, we can reason about influence before we actually understand it at a deep level, and that's what we will do in this chapter.

An important part of thinking about influence relationships is bound up with the idea of an influence diagram. Although I think there are limits beyond which diagrams are not actually very helpful, these limits are very far away, and so we can do quite a bit before we run into the ultimate complexities.

The simplest of all influence diagrams is shown in Figure 2.1. In the conventions of these diagrams, x and y are the names of two variables. The idea is that they are measured together over some set of *measurement opportunities*. They can both vary from one opportunity to the next, and they might vary together in some sense.

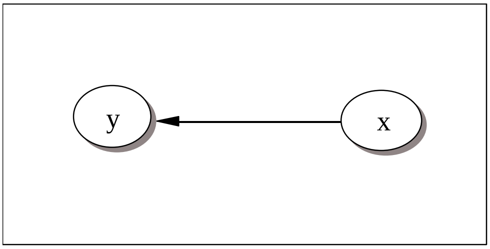

Fig. 2.1 In the simplest influence diagram, the arrow indicates that variable x has an (undefined) influence on the variable y.

The arrow that connects them asserts that x influences y. At the moment, we can take this to mean whatever we want, because this chapter is based on our intuitive notion of influence. If this seems to be too ambiguous, I would just point out that a rich and elaborate theory of *causal* influence has been constructed and studied for about a generation now, and it has no more to say about what the arrow in Fig. 2.1 means in causal terms than I do in more general cases. So the slippery ground we are standing on may be firmer than we think.

There is one very strong way x can influence y, and that is when y is a *function* of x. This means that there is some rule from which we can obtain the value of y knowing only the value of x. We call this an influence because whenever x takes on some particular value, then y must take on the value dictated by the function.

A simple example of a functional relationship is $y = \pi x^2$. This is a *logical* functional relationship when x is the radius and y the area of a circle. I say it is "logical" because it follows from the definition of area, using the logical operations of plane geometry. There is no point in designing experiments to "test" this functional relationship, because it is already known to hold. Since it is a logical relationship, it is also *universal.* This means that it holds for any circle, no matter how large or small. Moreover, if we took a particular circle, and steadily increased its radius, then the successive areas would still have the same functional relationship to the radii. The reason I am belaboring the obvious here is because it is precisely our general inability to think about *non-universal* relationships that makes our understanding of influence so difficult.

If y is a function of x, and if we change x systematically at different opportunities, then y will change correspondingly, so that it always satisfies the function. It does not make any difference *how* we change x, this correspondence of changes will always happen. Again, I am pointing out the obvious with a functional relationship, because a tight binding like this fails in many other influence relationships.

We now have at least two reasons to suspect that functional relationships like Fig. 2.1 will not happen very often. Influences are seldom *universal*, and in general *how* one changes x has an impact on y. Part of the reason for developing a theory of influence is to incorporate these two facts.

Another way that x can influence y happens when y is a *chance variable*. This means that y has a probability distribution. At each measurement opportunity, instead of being a determined quantity, there is variability in y that is in some sense unaccounted for. It is this excess variability that leads us to say that y has a probability distribution.

For an example, let x denote the measured radius of a circle. Suppose the measuring device that we use either gives the radius exactly (with probability 0.5), the radius plus 0.001 (with probability 0.3), or the radius minus 0.001 (with probability 0.2). The logical functional relationship for the area y is then $y = \pi(x\text{-}\varepsilon)^2$. Here ε is the error made by the measuring instrument, so that $x\text{-}\varepsilon$ is the true radius. If we make observations under these circumstances, then we will not see $y = \pi x^2$. Instead, we will see $y = \pi x^2$

with probability 0.5, $y = \pi(x-0.001)^2$ with probability 0.3, and $y = \pi(x+0.001)^2$ with probability 0.2.

In this example, y is a function of (x, ε). But because we cannot see ε, we will not see a universal relationship between y and x. Moreover, it is possible that if we change x very carefully, then it will not disturb the measurement process, but if we change x sloppily, the measurement process will give different probabilities to ε. I will eventually argue that things like this happen all the time, and cause considerable confusion. For the moment, however, I just want to point out that when this happens, then *how* one changes x has an impact on y. This example suggests that probabilistic influence might be a good deal more useful than functional influence.

Beyond functional and probabilistic influence, there is the notion of *causal* influence. It would be a great exaggeration to say that we have settled on good definitions of causal influence, but this does not seem to keep us from using the concept anyway. The reason is that we can imagine that causal influence, whatever it might be, has to behave according to certain rules. The rules alone can therefore give us some insight about causation.

2.1 Suppose that over five measurement opportunities, we see the following results:

Opportunity	x	y
1	0	-1
2	3	2
3	1	-1
4	3	3
5	2	2

Is y a function of x? Is x a function of y? Can you state a general test whether one variable is a function of another?

Several variables can influence another variable.

A more general notion of influence than Fig. 2.1 is that y might be a function of x and some other variables. Let z stand for a list of the "other variables". We then say that y is a function of (x,z). Of course, this means that when we know x and all the values of the variables in z, then we know y. Some would now argue that Fig. 2.1 is too simple, and that the other variables should be shown. Others would argue that it is alright to leave some variables out of an influence diagram. Both views are partly correct and partly incorrect.

Fig. 2.2 shows a situation in which y is influenced by three variables, x,z, and w. An example would be when $y = (x+z)^w$. Another example would be $y = x \ln(z/w)$. What is important about these is that in general all three variables are necessary to find the value of y. It is not possible to get y with only two of them. A probability example would be $y = zx + w + \varepsilon$, where ε is a chance variable (perhaps due to a measurement process). Here again y is a function of (x,z,w,ε), but because we cannot see ε we interpret the relationship between y and x, z, and w in probability terms.

It is possible to take a strict view of Fig 2.2, which says that all of the variables influencing y are shown (except perhaps for unobservable disturbances, like ε). In this case the diagram makes a powerful statement, that the variables influencing y, and only those variables, are shown. In this case the diagram is *complete*. We would then say that Fig. 2.1 is *incomplete*.

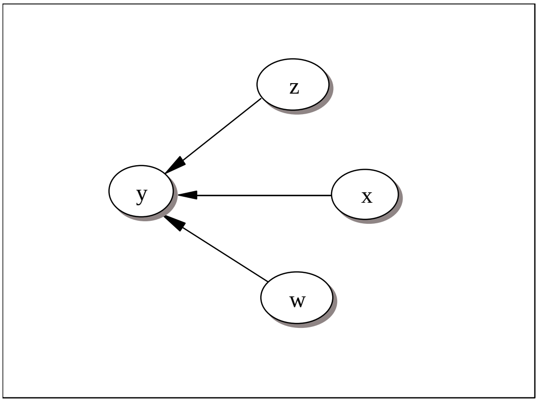

Fig 2.2. The variables x, z, and w all influence y, but the diagram does not show how.

This leaves us with an interesting dilemma, which we will have to work rather hard to solve. If Fig. 2.2 is the complete influence picture, then is Fig. 2.1 still a valid influence diagram? If we say yes, then we have to allow incomplete diagrams. If we say no, then we are in the absurd position of saying that although x is a part of a set of variables that influence y, it does not influence y itself.

I am going to maintain that influence diagrams provide us with useful tools for discussing and thinking about scientific relationships, but I am not going to maintain that they are perfect. This means that a diagram will depend on "something else", either in terms of formulas, or concepts, that may be difficult to depict in the diagram itself. In some cases, we will find that these "something elses" can be incorporated by adding decorations to the diagram, but in other cases this will be virtually impossible.

One such decorational device is to put bubbles around variables that can be observed, and to leave the bubble out when a variable cannot be observed. For example, in Fig. 2.3 the implication is that y is a function of (x,z,w,ε), and so the diagram is complete. Because ε has no bubble, we know that it cannot be observed. This diagram does not show, however, whether ε is a deterministic or chance variable. We could choose to enclose chance variables with rectangles, or diamond shapes, to distinguish them from ordinary variables, which would still be enclosed in bubbles. I am not so much interested in developing a set of conventions for this sort of thing as I am in opening up the possibilities.

2.2 In a circle with radius r, circumference c, and area A, it is always true that A = cr/2. Would you say that A is a function of (c,r)? Can you give a test whether one variable is a function of several others?

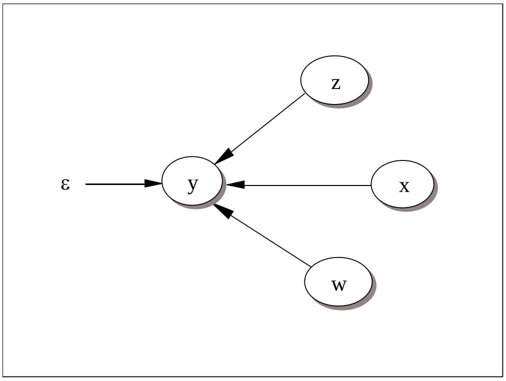

Fig 2.3. Unobservable variables can be indicated by leaving out the bubble.

Influences can have <u>strengths</u>, and in this case additive models are useful, though not very general.

An influence diagram like Fig. 2.3 shows *which* variables influence y, but it says nothing about how *strongly* each of them influences y. It would be very nice to be able to place measures of strength of influence onto the arrows in the diagram, because it is the arrows that indicate the influences. In general, it is rather hard to do this. There is, however, one special case in which it is easy – when y is an additive function of its influences. An example would be $y = 2.3x + 0.4z - 3.0w + \varepsilon$. Fig. 2.4 shows how we would decorate the diagram with the coefficients in this additive relationship. Note that there is no influence strength on the arrow from ε. This is again a sort of convention, that when no strength is indicated for a unobserved variable, the strength is assumed to be equal to 1. This may seem idiosyncratic, but there is actually a good reason for it, as we will see in a minute.

When we attach values to the arrows in an additive functional relationship, we need to keep in mind that these are expressed in terms of the units of measurement for all of the variables. For example, in Fig. 2.4 suppose that x were measured in inches. Now what if we ship the diagram off to our European colleagues, who would want x to be measured in centimeters? From our perspective, they want to think in terms of 2.54x, not in terms of x (inches). The fundamental point for doing the conversion is that 2.3x = (2.3/2.54)(2.54x). Since the latter term in parentheses is what the Europeans mean by "x", the European strength will be 2.3/2.54 = 0.91.

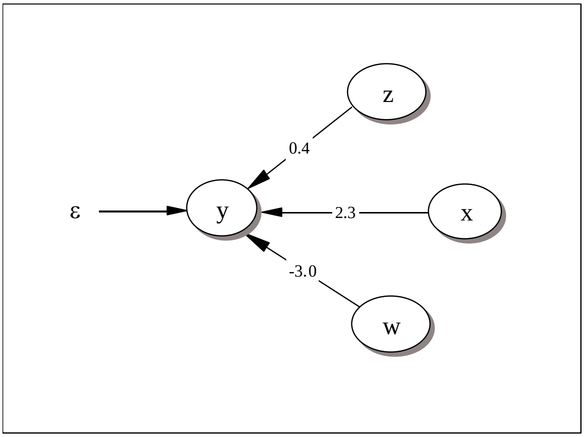

Fig. 2.4. In an additive functional relationship, the strengths of the influences of the variables can be associated with the arrows.

The strength measure for x is also affected by the units in which y is measured. Continuing the preceding example, suppose that we have measured y in pounds. Again, the Europeans would like to see y measured in kilograms. After we do the centimeter conversion for x, our additive relationship is $y = 0.91x + 0.4z - 3.0w + \varepsilon$. If we divide both sides of this equation by 2.2, then the left side will have been converted to kilograms, as the Europeans want. On the right side, 0.91 will become 0.41, 0.4 will become 0.18, and -3.0 will be come -1.36. Thus, changing the units for x, z, or w only changes the corresponding strength coefficient, but changing the units for y changes all three coefficients.

It is interesting here that we would not change the coefficient for ε. The reason is that since ε is not measured, it is ambiguous what units it would be measured in, and so we can simply carry the convention (coefficient of unmeasured variables are 1) from our equation to the European equation.

There is debate about whether one should somehow standardize units in additive influence diagrams, so they are free of the units used. This is usually done by expressing each chance variable in standard deviation units (that is, dividing it by its standard deviation) As usual, there are good arguments on both sides, and we will eventually examine them. For the moment, the important point is that in order to interpret influence strengths, one must bear in mind the units in which the variables are measured.

Additive models of this sort are very, very widely used. This is especially true in the social sciences, but it is often true elsewhere. There are several arguments in favor of this. First, influence diagrams faithfully represent the corresponding relationships when they are additive. That is, because additive relationships are so simple, we do not need to add excessive decoration to the influence diagrams that represent them. By simply looking at the coefficients next to the arrows in the diagram, we can immediately deduce the exact additive relationship. Secondly, in most practical cases the variables in an influence diagram are treated as if they were chance variables. This means that departures from exact functional relationships (additive or not) are regarded as being due to unobserved chance influences. Very often the strengths of the chance influences like ε dwarf the strengths of the observed-variable influences. In these cases, there is hard to find evidence favoring anything other than a simple additive relationship. Since additive relationships are easy to interpret, in the absence of compelling evidence for other relationships, they seem to be preferred. Note that this is a cultural argument, not a scientific argument. Thirdly, it is much easier to produce computer programs to fit data with additive models than it is with more general models. This is, of course, an argument from convenience, not from science.

We can note that all of the reasons for using additive functional relationships are advanced in order to make the life of the scientist less intellectually demanding. Non-additive relationships seem to abound in nature, in the sense that whenever sufficiently precise data can be obtained, non-additivity emerges. Perhaps the best attitude to take toward additive relationships is that they might give first-order approximations to the actual functional relationships, and that by examining and using them we might be led to better second-order approximations. Later on we will see many non-additive approaches that can be taken as second-order or subsequent approximations.

2.3 Suppose that x and z are variables that take values between 0 and 1, and that $y = x + z - xz$. Under what conditions will it be reasonable to approximate this by saying that y is an additive function of x and z?

Some influence pathways are <u>direct</u>, and some are <u>indirect</u>.

In Fig. 2.5, we see again the simplest influence relationship, that x influences y. The additional part of this figure shows that z influences x. The point that I want to make here is that there is a convention in influence diagrams, that an arrow between two variables stands for a *direct* influence. This can only be understood, however, if one has an example of an influence that is not direct, and that is what Fig. 2.5 provides. The idea is that z influences x, and this in turn influences y, because x influences y. Thus z has an influence on y due to its influence on x. This is called an *indirect* influence.

Here is a practical example of some importance. Let us suppose that y is some measure of a health risk (like high blood pressure, high serum cholesterol level, or lung damage). Further suppose that x measures some action that a person can take to lower the value of their risk variable (y). In the case of lung damage, x could be quitting smoking cigarettes. In case of blood pressure or cholesterol, x could be taking on an exercise program, or it could be taking one of the several drugs that influence these variables. Finally, let z denote an *indicator* of whether a person enters a health-improvement program of some kind (z=0 means they don't, z=1 means they do). If Fig. 2.5 correctly captures the influences in this case, then it says that z (the health program) influences y (the health risk) only because it influences x (a specific behavior), and this happens because x directly influences y.

In a case like this we say that x is a *mediator* variable. It mediates the influence of z on y. It is the medium by which z has its influence. It seems to be the case that when people claim that z influences y, but there is no obvious mediator variable, they have trouble believing that the influence is causal, but once a mediator variable is produced, they have much less problem with the causation issue. Somehow the existence of a mediator explains *why* z influences y. On the one hand, this is a psychological issue, but on the other, it seems as though much of science progresses through finding mediator variables, so perhaps our psychology is leading us in the right direction.

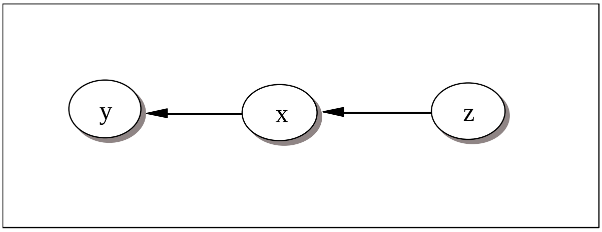

Fig. 2.5. x has a direct influence on y, but z has only an indirect influence on y, acting through x

A study that is designed to find mediator variables is called a *mechanistic* study. The idea behind the term is that x measures a mechanism through which z has its causal effect. What constitutes a "mechanism" depends on details of the scientific area under study, and since it is not general, I will not develop this idea very much in this section. But I do want to point out that it is very important.

In the case where influences are additive, there is an interesting computation that one can make in Fig. 2.5. If α measures the strength of x influencing y, and β measures the strength of z influencing x, then $\alpha\beta$ measures the strength of z influencing y. This is obviously an attractive result, because it simplifies the thinking and exposition of the indirect effect. As I will continue to argue, however, nature generally does not behave additively, and in particular nature has not arranged things to be attractive from our point of view. Let us always remember that additive models are often only first-order approximations.

Another situation that can arise is shown in Fig 2.6. In this figure we recognize that z has a direct influence on x, which has a direct influence on y, so that z has an indirect influence on y. But the arrow that connects z to y asserts that there is also a direct influence of z on y.

This diagram shows that we need to think in terms of influence *pathways*. There is one pathway from z to x to y, the indirect pathway. But the diagram also asserts that there is another pathway, a direct one through which z influences y. It is important to see that Fig. 2.6 is fundamentally different from Fig. 2.5.

Let's go back to the health-risk example I gave for Fig. 2.5. If Fig. 2.6 were valid, then this would say that putting someone on a health program (z) would have a direct effect on their health-risk (y), not mediated by their specific behavior (x). In the case where x was "quitting smoking", the health program (z) might involve not only specific quitting messages and strategies,

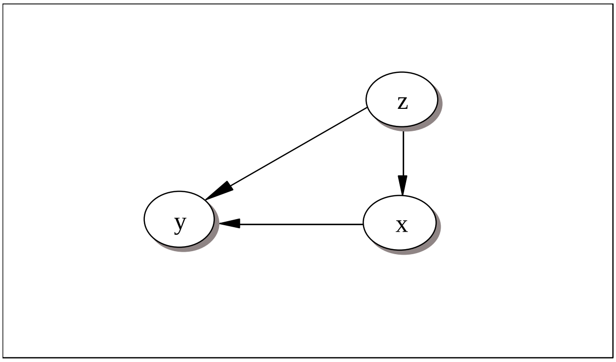

Fig. 2.6. A variable can have both direct and indirect influences.

but also more general information about healthy life-styles. It might, perhaps somewhat inadvertently, present information about the effect of exercise on health. This information might cause some people to start regular exercise programs, quite separately from whether they quit smoking or not. The health intervention (z) would then have had a direct effect on health risk (y) not mediated by quitting smoking (x), and this explains the additional arrow in Fig 2.6.

There is an enormous difference between Figures 2.5 and 2.6, and so it is surprising how seldom this shows up in practical research. To see the importance, imagine that we are in the "quit smoking" health program situation. If Fig 2.6 is right, but the researchers use Fig 2.5 as a guide for their analysis, then they will probably overestimate the effect of their health program due to quitting smoking, because some (maybe most) of the strength of influence might be through the "increase exercise" pathway. On the other hand, suppose that Fig 2.5 is valid, but Fig 2.6 is used to direct the analysis. In this case, the researchers will find that the direct influence of z on y is weak or nonexistent, and make the correct interpretation, that their health program has its effect through quitting smoking.

The lesson here seems to be that if one uses an influence diagram that is too simple to guide one's data analysis, then one may produce biased results. But if one uses an influence diagram that is more complex than it needs to be to describe reality, statistical tests will lead to deletion of extraneous influence arrows. Although I think this lesson is generally true, there are some substantial difficulties in using it in the naïve way that I have done here. We will attack this problem eventually.

One of the important lessons of this section is that it may make no sense to use adjectives like "direct influence" or "indirect influence" for individual variables. The term "direct" applies to a pathway, and similarly for "indirect". Thus, in Fig. 2.6 there are two influence pathways from z to y, one indirect acting through x, and one direct. A substantial amount of confusion in many scientific literatures has resulted from not making this distinction: direct/indirect pertains to pathways, not variables.

In a final comment on Fig 2.6, note that in my "quit smoking" example, the "direct" influence pathway from z to y was actually mediated by w = "adopting an exercise program", which did not appear in the diagram. This shows one of the caveats about the direct/indirect distinction: it is dependent on which variables we are willing to think about or measure. By leaving some variables out, we might promote indirect influences to direct influences. Perhaps every direct influence can be converted to an indirect influence, by a sufficiently diligent search for mediating variables. Thus, "direct" and "indirect" are relative terms, which depend on the level of detail (that is, variables) that can be measured in a particular study.

2.4 In the case of a circle of radius r, circumference c, and area A, certainly r influences c which influences A. Is there a direct influence of r on A that is not mediated by c?

2.5 Imagine a rectangle with sides x and z, diagonal $d = \sqrt{x^2 + z^2}$, and area A. It is clear that x influences d, and that d influences A. Is there a direct influence of x on A that is not mediated by d?

The concept of a <u>common influence</u> is central to evaluating influence relationships.

Look back at Fig. 2.1, which postulates a simple direct influence of x on y. If one collects data on x and y at different measurement opportunities, and one simply relates them in some statistical way, then ultimately the justification for this kind of analysis is that Fig. 2.1 is valid as an influence diagram. Are there ways that this can go wrong?

Fig. 2.7 shows that the answer is a redounding and disconcerting "yes". On both sides of the figure, we see that z influences both x and y. In this sense, z is a *common influence* of x and y. On the left side, we see a situation in which, despite z's influence on x and y, it is still true that x influences y. On the right side we see a different scenario, in which x does not directly influence y. The distinction between these two situations has created a lot of controversy.

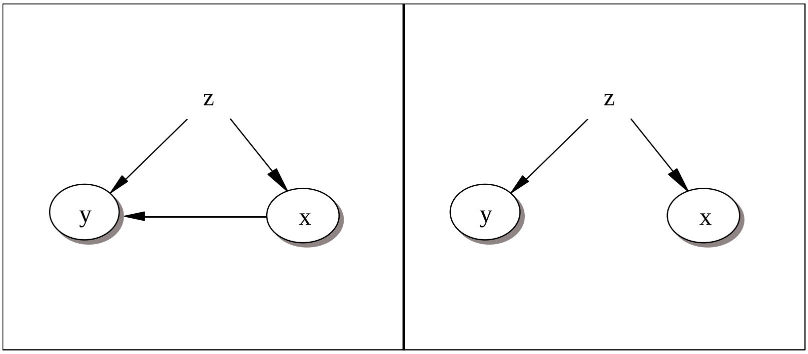

Fig. 2.7. A common influence (z) can happen in two different situations, where x influences y, and when it doesn't.

The fundamental problem is this. If the right panel in Fig.2.7 is valid, then ignoring z, Fig 2.1 could still be valid. So the story is this. In the right panel of Fig 2.7, x does not directly influence y. But z influences both x and y, so it is a common influence. Since we do not measure (or perhaps cannot measure) z, we revert to Fig. 2.1 for understanding influence. The point is that we may see an influence in Fig. 2.1, but it is not because x influences y, it is because they have a common influence, z.

This is one of the fundamental problems and controversies throughout all of science. If you can observe a relationship (functional, or more usually probabilistic) between x and y, then does this mean that x influences y? From the argument of the preceding paragraph, the answer is "potentially no". There may be another variable (z) that is a common influence of x and y, and when we exclude it (usually because we cannot measure it), then we see a statistical relationship between x and y, which is not indicative of an influence relationship.

In biomedicine this issue comes up most prominently in epidemiologic studies. Epidemiologists try to associate disease variables with other variables that might cause the disease. Their focus is on association, not influence, and not causation. When they find a "risk factor" (a potential cause variable) related to a "disease factor" (a variable measuring disease), they report it. They then acknowledge that there might have been a common influence, which would mean that the association was not an influence relationship, no less a causal relationship. Somehow in epidemiology it has come to be accepted that when the researchers confess the sin of not identifying and measuring the common influences, they are exonerated and their research is published nonetheless. (As an aside, we may note that epidemioligists use the term *confounder* for what I call a common influence.)

Almost all of the panic in the biomedical literature is that one might conclude that Fig. 2.1 is valid when in fact the right panel of Fig. 2.7 might actually be the real situation (for some hypothetical z). But this is not the only disturbing scenario. It is quite possible for the left panel of Fig. 2.7 to be valid, and when one just looks for a statistical relationship between x and y (ignoring z), none is to be found. In this case, the effect of the common cause is not to produce an artificial relationship between x and y, but to remove an actual relationship. This is no less possible than the case in which z produces an artificial relationship, but for some cultural reason scientists seem to be more wary of the production than the destruction of influence relationships due to common influences.

2.6 Again consider a rectangle with sides x and z, diagonal d, and area A. Consider an influence diagram that says that d is a common influence of A and x. Does x have a direct influence on A in this diagram?

One variable can <u>modulate</u> the influence of another variable.

I have suggested that influence diagrams can indicate influence relationships, but that sometimes the strength of the influence relationship is also important. This opens the possibility that one variable might influence the influential strength of another variable.

In Fig. 2.8, as usual x influences y. The strength of this influence is not shown, at least not overtly. The diagram says that z has an influence on the strength of the influence of x on y. In this case, z is called a *modulator*, and it is said to modulate the influence of x on y.

If, for convenience, we go back to additive models, then an example functional relationship that includes modulation is $y = (1-0.5z)x$. Here, the strength of x's influence on y is $(1-0.5z)$, showing explicitly that the value of z determines how strongly x influences y.

Modulation relationships can be extremely important in science, because ignoring them can lead to massive confusion. In the example, note that if $z>2$ then x influences y positively, while if $z<2$ then x influences y negatively, and if $z=2$ then x does not influence y at all. If one conducts experiments in which z is unobserved, then the results will consist of some cases where x influences y, some cases where x inversely influences y, and some cases where x does not influence y. What kind of influence relationship one concludes from data like these will depend on the distribution of values of the (unobserved) z value in the experiment. It is to be expected in cases like this, that different researchers doing different studies, with different distributions of z, will report different results. It is the failure to report consistent results across studies that makes conventional researchers so nervous about what to believe, and then one might look to an explanation in terms of modulators.

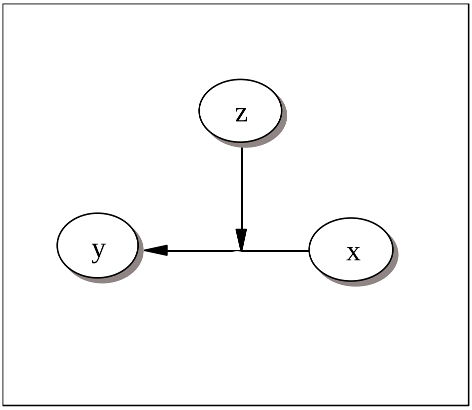

Fig. 2.8. One variable can modulate the relationship between two others.

One would think that these obvious facts would lead researchers to search for modulating variables. Oddly, this is sometimes true, but often false. In biochemical experiments, one often sees that potential modulatory variables are *controlled*, which means that they are fixed during experiments, or only permitted to vary in ways that the researchers determine. Biochemical scientists do not, however, generally admit that there may have been modulatory variables that they have not controlled, and failed to measure. Other biomedical scientists are often even more glaringly cavalier about modulatory variables. Overwhelmingly, the attitude is to ignore potential modulators. Almost always this attitude comes from statistical considerations. It is harder to find a statistically significant interaction, which would mean a modulation influence, than it is to find a statistically significant direct influence. In other words, statistical inference about modulatory influence is generally weaker than inference about direct influence. The decision to focus on the latter is, therefore, not driven by considerations of science, but by biases among scientists that are driven by statistical issues.

Even a glancing acquaintance with biology is enough to convince one that modulatory relationships happen very often. Because more general systems often mimic biological systems, we are justified in expecting modulatory influences to occur frequently. This suggests that Fig. 2.8 needs to be taken very seriously.

2.7. Produce and explain an influence diagram showing how vaccination works.

2.8 Suppose a researcher sees patients with favorable or unfavorable prognoses. Suppose he puts patients with more favorable prognoses on a new drug, and then compares the subsequent course of their disease with the remainder of patients who didn't get the drug. Produce and explain an influence diagram for this situation.

2.9 A physician tries to prevent heart disease in a patient by prescribing a cholesterol-lowering drug, but unfortunately it has the side effect of increasing the patient's blood pressure. Produce and explain the corresponding influence diagram.

2.10 Older people who become depressed often reduce their physical activity and stop taking certain medications, both of which can reduce the density of their bones, leading to fractures. Produce and explain the appropriate influence diagram.

2.11. Produce and explain a general influence diagram for the placebo effect.