Recall from last time that experiments provide a very easy way of testing whether one variable has a causal influence on another variable. If you manipulate the potential cause of interest – while holding everything else either constant or equal on average – and find a change in the effect of interest, then you have evidence in favor of your cause-effect relationship. You can also learn quite a bit about causation from correlational studies, but it's nowhere near as easy.

One way to see how much harder it is to get causation from a correlational is to remember that a correlation always works in both directions. If you have found a correlation between, for example, anxiety and depression, then you have also found a correlation between depression and anxiety. Therefore, without further information, it is just as likely that depression causes anxiety as it is that anxiety caused depression.

This problem didn't come up for experiments because experiments always have one variable that is completely under the control of the experimenter – namely, the independent variable. If you find a relationship between lighting and memory in an experiment, it can't be the case that the subjects' memories caused the lighting to be different; the only direction of causation that is possible is that the lighting influenced memory. It's the lack of control of either variable in a correlational study that opens the door to what is often called "backwards" or "reverse causation" – the possibility that the variable that you think is the effect is actually the cause. In general, this is referred to as the "directionality problem" with correlations.

The second problem with trying to establish a causal relationship from a correlation comes from the fact that correlational studies are – by definition – always low on internal validity. If you use one of the fluffier definitions of internal validity – the extent to which there are no confounds – and then note that we make no effort to ensure that, for example, the subjects high in anxiety are the same, on average, as the subjects low in anxiety *in every way other than anxiety*, then you'll see that there are always myriad confounds in a correlational study. For example, maybe the subjects who are high in anxiety have lousier jobs, on average, than the subjects who are low in anxiety, and maybe that's why they're depressed. Not because they are anxious, but because their jobs stink. In fact, maybe it's the quality of their job that causes both their depression and their anxiety!

The possibility that some extraneous variable is responsible for the correlation that was found between the two variables of interest (because it caused both of them) is known as the "third-variable problem." In its most extreme form, the third variable is the entire cause of the correlation between the two variables of interest and there's no causal relationship between them at all. This would make the correlation between the two variables of interest "spurious" (in the exact same way that we used this word for experiments).