Causation and Counterfactuals

In law as in science, one thing we typically want to resolve are questions of causation. For example, in biology we might ask things like "does this drug reduce deaths from heart disease," in economics we might ask "does raising the minimum wage increase unemploment." In law, such questions will come up in any substantive area where the outcome of litigation depends on attributing factual responsibility. The most obvious area is torts: in a toxic tort case, for example, the plaintiff will need to be able to show that the poison the defendant dumped into the water supply actually caused the children to get sick.

The good thing about causation in the scientific sense is that it more or less maps to our ideas of causation in tort law. (Philosophers disagree about the details of all this stuff, but it's proven pretty workable for actual scientists.) In that tort-y sense, it's all about *counterfactuals*—that is, when we imagine what it might mean for X to have caused Y, we imagine a world which is completely identical to the real world, except that X didn't happen, and we want to know whether Y happened in that world.

- I also think our ideas of causation in tort law map to other legal contexts that aren't as explicitly about causation. For example, we can treat many kinds of intentional discrimination cases as causal questions: when we ask "did the prosecutor challenge the jurors because of their race," we mean "was it race or something else that the prosecutor was using as their reason to exclude them," and what that amounts to for practical purposes is "but for their race, would they have been excluded," i.e., the counterfactual question.

Sometimes, we can create the identical world where X didn't happen. This is what a **scientific experiment**, in its ideal form, does. If we want to know whether drug X helps people recover from heart disease, we can give it to people, and compare their outcomes to people who didn't get the drug. Assuming the people who get the drug are identical to the people who didn't get the drug (*foreshadowing:* this is a super-heavy assumption that will drive everything else we do in talking about causation), the results of the experiment will lead us to a belief about whether the drug caused improvements in outcomes in the patients who got it. And, supposing the patients are identical to the people who might need the drug out in the world, that lets us come to conclusions about whether the drug will cause health improvements if doctors give it to their patients—which is, ultimately, what we want to know.

We do experiments in the legal context as well. The most prominent example, which we'll focus on this week, is the practice of discrimination testing. In many contexts, but particularly in housing discrimination, litigants or potential litigants develop evidence using experimental methods.

- The basic setup in these contexts goes like this: someone claims that they've been discriminated against by a leasing company, e.g., on the

basis of race. An anti-discrimination organization sends people who look similar on paper (similar mannerisms, similar stories about employment and credit, etc.) to go to the leasing office and inquire about apartments. If the company tells everyone of one race "we have a bunch of apartments available, they're $500 a month, please take an application!" and tells the otherwise-identical people of another race "there aren't any apartments available... maybe you can get on the waiting list for an application, but all our apartments are $1000/month," then there's pretty good evidence that the difference in race caused the difference in behavior, that is, that the leasing company is engaging in illegal discrimination.

Variations on a theme of not identical I: changing the experimental subjects

All the work in what I said above is in the notion of "identical." There are lots of ways that the world of an experiment can be not identical to the world that we care about learning about.

First, we could have random variation just in terms of how the thing that's supposed to be doing the causing (the *treatment*) works. Maybe the heart disease drug only works for 10% of people and we didn't get those people in our study. (Or it only works for 10% of people and we got all of those in our study!)

- In general, the solution for this kind of random variation is *sample size*—as we discussed before, the genius of the law of large numbers is that when we sample from a population, we can expect the mean of the sample (here, the degree of susceptibility to the drug) to converge to the population mean, so the larger the group of people in our study (which will always be a sample from the larger population of people who might need a heart disease drug), the more confident we can be that the extent to which people in our study respond to the drug will be the same as the extent to which people in general do so.
  - Note that I'm ignoring some mathematical formalities here (like the difference between sampling with and without replacement, a bunch of stuff about the concept of mathematical expectation, etc). This will be good enough for present purposes.

Second, we could have *nonrandom* variation in terms of the match between our sample and the population. For a blunt example, suppose our study only had men participating. Maybe women respond differently to the drug than men? (This is a longstanding real-world problem in medical studies.) Then we would have strong reason to worry that maybe we were missing some effect of the drug linked to sex (such as an interaction with sex-linked hormone levels).

- In general, the solution to this problem is *random sampling* (although experimental researchers have created a bunch of fancy variations on random sampling for special situations, which we won't cover in this class).

- The intuition here is that if you don't randomly sample in selecting the people for your study, you're not really studying the whole population, you're studying some subset of the population, like just men.
- This is a pervasive problem in psychological research, for at least two reasons. First, many psychologists use opportunistic ("convenience") samples of undergraduates, often assigned to participate in studies as part of classes—and they might not be identical to the the general population. Second, there are cross-cultural differences in lots of psychological factors, so studies that focus on participants from only a small subset of countries (like Americans, because America has such a robust university system) can fail to generalize to people from other cultures.
- Experiments typically have treatment and control groups so we can compare them, and another source of this kind of nonrandom variation is if people aren't assigned randomly to treatment and control groups. For example, what happens if in our housing discrimination there are systematic differences other than race between the groups of people sent into the leasing office? Suppose, for example, that everyone in one group wore a tie, and nobody in the other group did. We won't know whether it's race that caused the different treatment, or whether the leasing agent just prefers people who dress up.

Variations on a theme of not identical II: changing the experimental context

Remember I said above that our ideal of causation involves identifying a world where X (our candidate cause) didn't happen, and seeing how things vary from a world where X did happen. Sometimes, however, we don't have the ability to manipulate X independent of manipulating other things in the world, so it's impossible to create truly identical worlds with and without X happening.

Here's an example: a lot of psychological experiments happen in artificial contexts. For example, they'll be trying to test some social phenomenon, like whether people with deeper voices are more convincing, so they'll invite a bunch of undergraduates into a lab and ask them to watch videos of people making political speeches. If the students who see the deeper-voiced speaker (saying identical content) are more convinced by the political position the speech espoused, that's evidence that voice pitch matters to persuasion.

The difficulty with that experiment is that real-world speeches don't happen in labs. And that might make a difference to the result. Continuing our hypothetical voice pitch experiment: suppose deeper voices are easier to hear over cheering crowds? Then deeper voices might be more convincing in the real world of political rallies (just because the audiences can hear what they say) but not in the lab. Because we can't vary X while holding everything else in the world constant, we might not be able to learn about that effect.

Psychologists use the term *external validity* to capture this class of worries: is the setting of the study similar enough to the real world to allow it to be used to make inferences about that real world? By contrast, many of the issues we discussed in the previous subsection often get described under the rubric of *internal validity*, that is, whether we genuinely believe that the effects we observe in an experiment are the result of our changing X, as opposed to some other source of variation created between treatment and control like having all men in the treatment group. (Psychologists talk about lots of other sub-types of validity too, but we won't bother with that here.)

Lab experiments vs field experiments

In social scientific context (and many experimental contexts in the law, like those housing testers I mentioned, are effectively social science contexts) there are two general kinds of experiments.

- In *lab experiments* the researcher controls the environment, so they have total control over which participants recieve the treatment (the deep-voiced speech, the heart disease drug) and which don't, they can make sure all other influences are held constant between the groups, etc.
- In *field experiments* the researcher applies the treatment to randomly selected people out in the real world, in the natural context where the treatment might be relevant.
  - Our housing discrimination tests are examples of field experiments—they don't take a bunch of leasing agents into the lab and show them videos of black and white prospective tenants; they send the black and white tenants out into the real environment where leasing decisions are made and see what happens.

(There are also more specialized kinds of experiments that draw on features of both. For example, many political scientists conduct what are known as *survey experiments* where variation is introduced in the questions asked on large, national, random surveys in order to test things like the extent to which the order of candidate names matters for who people say they favor.)

There's an inherent tradeoff between internal and external validity in lab vs field experiments.

- Lab experiments are likely to have problems with external validity, because the lab can never fully resemble the real-world environment where the phenomena we care about are actually occurring.
- Field experiments are likely to have problems with internal validity, because the world introduces a lot of variation that the experimenter can't control. For example, what happens to our housing discrimination test if a real tenant shows up and rents the last apartment midway through? Or what

happens if several of our white testers get sick and start coughing all over the leasing agent?

There is, in other words, no perfect research design.

Variations on a theme of not identical III: experimental vs observational studies

Not all research is experimental; indeed, there are many areas of research that can't be experimental. For example, my own Ph.D. field of political science has huge amounts of research that can't be studied effectively with experiments. If we want to know how having a war affects whether a state is democratic a few years later, we obviously don't have the power to start a war and see what happens, and even if we had the power, it would be egregiously unethical to do so.

In law, similarly, if we're trying to figure out whether the defendant's negligent dumping of the poison in the water supply caused the kids to get cancer, we obviously aren't going to be able to randomly select some control towns and some treatment towns, dump poison in the water supply of the treatment towns, and then wait a couple decades to see how many kids die. Also, in law we often want to borrow scientific methods to study behavior retrospectively—to come to some kind of conclusion about whether *existing* disparaties in salary were caused by discrimination *over an extended period of time*, for example—but we can't get in a time machine and vary gender in the defendant's workplace to figure that out.

In these contexts, we have to do *observational research* where instead of controlling who gets the treatment, we have to look at the world and see if there are differences in outcomes for who actually got the treatment. For example, look at employees in the workplace and see if there are differences in salary for who got "assigned" (by "nature," that is, not by some scientist) the "treatment" of being a woman.

Because we don't have the ability to create otherwise identical worlds in which people get or do not get the treatment, observational research can be very difficult. We'll spend some time over the rest of the course on this, but here is a brief a sampling of the difficulties:

- *Selection bias* is when the people whom nature assigns to treatment are systematically different than the people whom nature assigns to a control. For example, suppose we want to know whether a workplace discriminates against redheads, so we compare the salaries of redheads and the salaries of brunettes. But suppose that the employer has a reputation (deserved or undeserved) as being hostile to redheads, so only desperate readheads who can't get a job anywhere else end up working there. They might be systematically worse qualified than brunettes, within that workplace.

- *Omitted variables* are a problem when the omitted variable is causally related both to X, the thing you're trying to test, and Y, the thing you're trying to see if X causes. There are lots of ways this might happen—maybe Z, the omitted variable, causes X, which causes Y. Maybe X causes Y only in the presence of Z. There are entire research agendas organized around different ways causal inferences can go wrong in the presence of external variables. We'll discuss some of this later in the course.
- *Reverse causation* sometimes we think X causes Y, but actually Y causes X.

This week, we'll be talking about experimental research; for several weeks thereafter, we'll be talking about the various ways people get around the problems of doing observational research. For now, if you'd like to read more, here's a good summary of the basic causal issues in observational science.