The Fundamentals of

POLITICAL SCIENCE RESEARCH

Paul M. Kellstedt
Guy D. Whitten

SECOND EDITION

dealing with race and political participation, at first it appeared that race might be causally related to participation rates, with Anglos participating more than those of other races. But, we argued, in this particular case, the first glance was potentially quite misleading.

Why? Because what appeared to be the straightforward comparisons between three groups – participation rates between Anglos, Latinos, and African Americans – ended up being far from simple. On some very important factors, our different groupings for our independent variable X were far from equal. That is, people of different racial groupings (X) had differing socio-economic statuses (Z), which are correlated with race (X) and also affected their levels of participation (Y). As convincing as those bivariate comparisons might have been, they would likely be misleading.

Comparisons are at the heart of science. If we are evaluating a theory about the relationship between some X and some Y, the scientist's job is to do everything possible to make sure that no other influences (Z) interfere with the comparisons that we will rely on to make our inferences about a possible causal relationship between X and Y.

The obstacles to causal inference that we described in Chapter 3 are substantial, but surmountable. We don't know whether, in reality, X causes Y. We may be armed with a theory that suggests that X does, indeed, cause Y, but theories can be (and often are) wrong or incomplete. So how do scientists generally, and political scientists in particular, go about testing whether X causes Y? There are several strategies, or **research designs**, that researchers can use toward that end. The goal of all types of research designs is to help us evaluate how well a theory fares as it makes its way over the four causal hurdles – that is, to answer as conclusively as is possible the question about whether X causes Y. In the next two sections we focus on the two strategies that political scientists use most commonly and effectively: **experiments** and **observational studies**.[1]

4.2 EXPERIMENTAL RESEARCH DESIGNS

Suppose that you were a candidate for political office locked in what seems to be a tight race. Your campaign budget has money for the end of the campaign, and you're deciding whether or not to make some television ad buys for a spot that sharply contrast your record with your opponent's – what some will surely call a negative, attack ad. The campaign manager has had a public relations firm craft the spot, and has shown it to you in

[1] Throughout this book, we will use the term "experiment" in the same way that researchers in medical science use the term "randomized control trial."

your strategy meetings. You like it, but you look to your staff and ask the bottom-line question: "Will the ad work with the voters?" In effect, you have two choices: run the attack ad, or do nothing.

We hope that you're becoming accustomed to spotting the causal questions embedded in this scenario: Exposure to a candidate's negative ad (X) may, or may not, affect a voter's likelihood of voting for that candidate (Y). And it is important to add here that the causal claim has a particular directional component to it; that is, exposure to the advertisement will *increase* the chances that a voter will choose that candidate.[2]

How might researchers in the social sciences evaluate such a causal claim? Those of you who are campaign junkies are probably thinking that your campaign would run a focus group to see how some voters react to the ad. And that's not a bad idea. Let's informally define a focus group as a group of subjects selected to expose to some idea (like a new kitchen knife or a candidate's TV ad), and to try to gather the subjects' responses to the idea. There's a problem with the focus group, though, particularly in the case at hand of the candidate's TV ad: What would the subjects have said about the candidate had they *not* been exposed to the ad? There's nothing to use as a basis for comparison.

It is very important, and not at all surprising, to realize that voters may vote either for or against you for a variety of reasons (Z's) that have nothing to do with exposure to the advertisements – varying socio-economic statuses, varying ideologies, and party identifications can all cause voters to favor one candidate over another. So how can we establish whether, among these other influences (Z), the advertisement (X) also causes voters to be more likely to vote for you (Y)?

Can we do better than the focus group? What would a more scientific approach look like? As the introduction to this chapter highlights, we will need a comparison of some kind, and we will want that comparison to isolate any potentially different effects that the ad has on a person's likelihood of voting for you.

The standard approach to a situation like this in the physical and medical sciences is that we would need to conduct an experiment. Because the word "experiment" has such common usage, its scientific meaning is frequently misunderstood. An experiment is *not* simply any kind of analysis that is quantitative in nature; neither is it exclusively the domain of laboratories and white-coated scientists with pocket protectors. We define an

[2] There is a substantial literature in political science about the effects that negative advertisements have on both voter turnout and vote choice. For contrasting views on the effects of negative ads, see Ansolabehere and Iyengar (1997), Wattenberg and Brian (1999), and Geer (2006).

experiment as follows: *An experiment is a research design in which the researcher both controls and randomly assigns values of the independent variable to the participants.*

Notice the twin components of the definition of the experiment: that the researcher both *controls* values of the independent variable – or X, as we have called it – as well as *randomly assigns* those values to the participants in the experiment. Together, these two features form a complete definition of an experiment, which means that there are no other essential features of an experiment beside these two.

What does it mean to say that a researcher "controls" the value of the independent variable that the participants receive? It means, most importantly, that the values of the independent variable that the participants receive are *not* determined either by the participants themselves or by nature. In our example of the campaign's TV ad, this requirement means that we cannot compare people who, by their own choice, already have chosen to expose themselves to the TV ad (perhaps because they're political junkies and watch a lot of cable news programs, where such ads are likely to air). It means that we, the researchers, have to decide which of our experimental participants will see the ads and which ones will not.

But the definition of an experiment has one other essential component as well: We, the researchers, must not only control the values of the independent variable, but *we must also assign those values to participants randomly*. In the context of our campaign ad example, this means that we must toss coins, draw numbers out of a hat, use a random-number generator, or some other such mechanism to divide our participants into a **treatment group** (who will see the negative ad) and a **control group** (who will not see the ad, but will instead watch something innocuous, in a social science parallel to a **placebo**).

What's the big deal here? Why is randomly assigning subjects to treatment groups important? What scientific benefits arise from the random assignment of people to treatment groups? To see why this is so crucial, recall that we have emphasized that all science is about comparisons and also that every interesting phenomenon worth exploring – every interesting dependent variable – is caused by many factors, not just one. Random assignment to treatment groups ensures that the comparison we make between the treatment group and the control group is as pure as possible and that some other cause (Z) of the dependent variable will not pollute that comparison. By first taking a group of participants and then randomly splitting them into two groups on the basis of a coin flip, what we have ensured is that the participants will not be systematically different from one another. Indeed, provided that the participant pool is reasonably large, randomly assigning participants to treatment groups ensures that the groups,

as a whole, are *identical*. If the two groups are identical, save for the coin flip, then we can be certain that any differences we observe in the groups must be because of the independent variable that we have assigned to them.

Return to our campaign advertising example. An experiment involving our new ad would involve finding a group of people – however obtained – and then randomly assigning them to view either our new ad or something that is not related to the campaign (like a cartoon or a public service announcement). We fully realize that there are other causes of people's voting behaviors and that our experiment does not negate those factors. In fact, our experiment will have nothing whatsoever to say about those other causes. What it *will* do, and do well, is to determine whether our advertisement had a positive or negative effect, or none at all, on voter preferences.

Contrast the comparison that results from an experiment with a comparison that arises from a non-experiment. (We'll discuss non-experimental designs in the next section.) Suppose that we don't do an experiment and just run the ad, and then spend our campaign money conducting a survey asking people if they've seen our ad, and for whom they plan to vote. Let's even assume that, in conducting our survey, we obtain a random sample of citizens in the district where the election will take place. If we analyze the results of the survey and discover that, as hoped, the people who say that they have seen our ad are more likely to vote for us than people who say they have not seen our ad, does that mean that the ad *caused* – see that word again? – people's opinions to shift in our favor? No. Why not? Because people who saw our ad and people who did not see our ad might be *systematically different* from one another. What does that mean? It means that people who voluntarily watch a lot of politics on TV are (of course) more interested in politics than those who watch the rest of what appears on TV. In this case, a person's level of interest in politics could be an important Z variable. Interest in politics could very well be associated with a person's likelihood to vote for you. What this means is that the simple comparison in a non-experiment between those who do and do not see the ad is potentially misleading because it is confounded by other factors like interest in politics. So is the higher support for you the result of the advertisement, or is it the result of the fact that people likely to see the ad in the first place are people with higher interest in politics? Because this particular non-experimental research design does not answer that question, it does not clear our fourth causal hurdle. It is impossible to know whether it was the ad that caused the voters to support you. In this non-experimental design just described, because there are other factors that influence support for a candidate – and, critically, because these factors are also related to whether or not people will see the advertisement – it is very difficult to say conclusively that

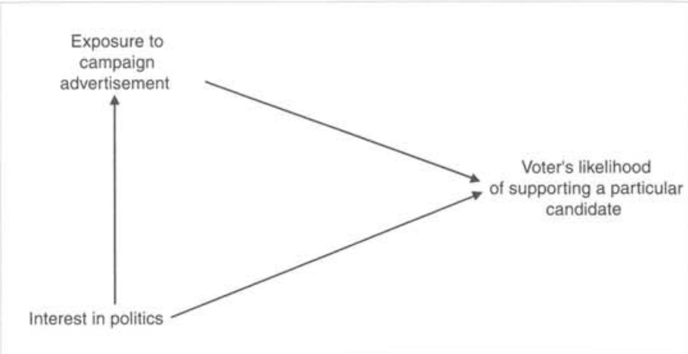

Figure 4.1. The possibly confounding effects of political interest in the advertisement viewing–vote intention relationship.

the independent variable (ad exposure) causes the dependent variable (vote intention). Figure 4.1 shows this graphically.

Here is where experiments differ so drastically from any other kind of research design. What experimental research designs accomplish by way of random assignment to treatment groups, then, is to decontaminate the comparison between the treatment and control group of all other influences. Before any stimulus (like a treatment or placebo) is administered, all of the participants are in the same pool. Researchers divide them by using some random factor like a coin flip, and that difference is the only difference between the two groups.

Think of it another way. The way that the confounding variables in Figure 4.1 are correlated with the independent variable is highly improbable in an experiment. Why? Because if X is determined by randomness, like a coin flip, then (by the very definition of randomness) it is exceedingly unlikely to be correlated with anything (including confounding variables Z). When researchers control and assign values of X randomly, the comparison between the different groups will not be affected by the fact that other factors certainly do cause Y, the dependent variable. In an experiment, then, because X is only caused by randomness, it means that we can erase the connection between Z and X in Figure 4.1. And, recalling our definition of a confounding variable, if Z is not correlated with X, it cannot confound the relationship between X and Y.

Connect this back to our discussion from Chapter 3 about how researchers attempt to cross four hurdles in their efforts to establish whether some X causes Y. As we will see, experiments are not the only method that help researchers cross the four causal hurdles, but they are uniquely

capable in accomplishing important parts of that task. Consider each hurdle in turn. First, we should evaluate whether there is a credible causal mechanism before we decide to run the experiment. It is worth noting that the crossing of this causal hurdle is neither easier nor harder in experiments than in non-experiments. Coming up with a credible causal scenario that links X to Y heightens our dependence on theory, not on data or research design.

Second, in an experiment, it is impossible for Y to cause X – the second causal hurdle – for two reasons. First, assigning X occurs in time before Y is measured, which makes it impossible for Y to cause X. More importantly, though, as previously noted, if X is generated by randomness alone, then nothing (including Y) can cause it. So, in Figure 4.1, we could eliminate any possible reverse-causal arrow flowing from Y to X.

Establishing, third, whether X and Y are correlated is similarly easy regardless of chosen research design, experimental or non-experimental (as we will see in Chapter 7). What about our fourth causal hurdle? Have we controlled for all confounding variables Z that might make the association between X and Y spurious? Experiments are uniquely well equipped to help us answer this question definitively. An experiment does not, in any way, eliminate the possibility that a variety of other variables (that we call Z) might also affect Y (as well as X). What the experiment does, through the process of randomly assigning subjects to different values of X, is to equate the treatment and control groups on all possible factors. On every possible variable, whether or not it is related to X, or to Y, or to both, or to neither, the treatment and control groups should, in theory, be identical. That makes the comparison between the two values of X unpolluted by any possible Z variables because we expect the groups to be equivalent on all values of Z.

Remarkably, the experimental ability to control for the effects of outside variables (Z) applies to *all* possible confounding variables, regardless of whether we, the researchers, are aware of them. Let's make the example downright preposterous. Let's say that, 20 years from now, another team of scientists discovers that having attached (as opposed to detached) earlobes causes people to have different voting behaviors. Does that possibility threaten the inference that we draw from our experiment about our campaign ad? No, not at all. Why not? Because, whether or not we are aware of it, the random assignment of participants to treatment groups means that, whether we are paying attention to it or not, we would expect our treatment and control groups to have equal numbers of people with attached earlobes, and for both groups to have equal numbers of people with detached earlobes. The key element of an experimental research design – randomly assigning subjects to different values of X, the independent variable – controls for every Z in the universe, whether or not we are aware of that Z.

In summary, if we think back to the causal hurdles scorecard from the previous chapter, all properly set-up experiments start out with a scorecard reading [? y ? y]. The ability of experimental designs to cleanly and definitively answer "yes" to the fourth hurdle question – Have we controlled for all confounding variables Z that might make the association between X and Y spurious? – is a massive advantage.[3] All that remains for establishing a causal relationship is the answers to clear the first hurdle – Is there a credible causal mechanism that connects X to Y? – and hurdle three – Is there covariation between X and Y? The difficulty of clearing hurdle one is unchanged, but the third hurdle is much easier because we need only to make a statistical evaluation of the relationship between X and Y. As we will see in Chapter 7, such evaluations are pretty straightforward, especially when compared to statistical tests that involve controlling for other variables (Z).

Together, all of this means that experiments bring with them a particularly strong confidence in the causal inferences drawn from the analysis. In scientific parlance, this is called **internal validity**. If a research design produces high levels of confidence in the conclusions about causality, it is said to have high internal validity. Conversely, research designs that do not allow for particularly definitive conclusions about whether X causes Y are said to have low degrees of internal validity.

4.2.1 "Random Assignment" versus "Random Sampling"

It is critical that you do not confuse *the experimental process of randomly assigning subjects to treatment groups*, on the one hand, with *the process of randomly sampling subjects for participation*, on the other hand. They are entirely different, and in fact have nothing more in common than that six-letter word "random." They are, however, quite often confused for one another. **Random assignment** to treatment and control groups occurs when the participants for an experiment are assigned randomly to one of several possible values of X, the independent variable. Importantly, this definition says nothing at all about how the subjects were selected for participation. But **random sampling** is, at its very heart, about how researchers select cases for inclusion in a study – they are selected at random, which means that every member of the underlying **population** has an equal probability of being selected. (This is common in survey research, for example.)

[3] After all, even the best designed and executed non-experimental designs must remain open to the possibility that, somewhere out there, there is a Z variable that has not yet been considered and controlled for.

Mixing up these two critical concepts will produce a good bit of confusion. In particular, confusing random sampling with random assignment to treatment groups will mean that the distinction between experiments and non-experiments has been lost, and this difference is among the more important ones in all of science. To understand how science works, keep these two very important concepts separate from one another.

4.2.2 Varieties of Experiments and Near-Experiments

Not all experiments take place in a laboratory with scientists wearing white lab coats. Some experiments in the social sciences are conducted by surveys that do use random samples (see above). Since 1990 or so, there has been a growing movement in the field of survey research – which has traditionally used random samples of the population – to use computers in the interviewing process that includes experimental randomization of variations in survey questions, in a technique called a **survey experiment**. Such designs are intended to reap the benefits of both random assignment to treatment groups, and hence have high internal validity, as well as the benefits of a random sample, and hence have high **external validity**.[4] Survey experiments may be conducted over the phone or, increasingly, over the internet.

Another setting for an experiment is out in the natural world. A **field experiment** is one that occurs in the natural setting where the subjects normally lead their lives. Random assignment to treatment groups has enabled researchers in the social sciences to study subjects that seemed beyond the reach of experimentation. Economists have long sought conclusive evidence about the effectiveness (or the lack thereof) of economic development policies. For example, do government fertilizer subsidies (X) affect agricultural output (Y)? Duflo, Kremer, and Robinson (2011) report the results of an experiment in a region in Western Kenya in which a subsidy of free delivery of fertilizer was offered only to randomly chosen farmers, but not to others.

Field experiments can also take place in public policy settings, sometimes with understandable controversy. Does the police officer's decision whether or not to arrest the male at a domestic violence call (X) affect the incidence of repeat violence at the same address in the subsequent months (Y)? Sherman and Berk (1984) conducted a field experiment in Minneapolis, randomizing whether or not the male in the household would automatically (or not) be arrested when police arrived at the house.

On occasion, situations in nature that are not properly defined as experiments – because the values of X have not been controlled and assigned

[4] See Piazza, Sniderman, and Tetlock (1990) and Sniderman and Piazza (1993).