

# Experimental Design

In the past twenty years, experimental research designs have become increasingly popular in disciplines in the social sciences, such as political science, that do not have a longstanding tradition of experimentation. Advocates of experiments argue that these methods are the gold standard in assessing causality: whether a variable of interest is actually caused by another variable. However, critics push back that causality in the context of an experiment does not always translate well to causal processes in the real world.

There is no single research design that is perfectly suited to test every hypothesis a social scientist might assert: the best research matches the most effective research design to the particular puzzle to be explored. The purpose of this module is to help you understand what differentiates experiments from other forms of research designs to facilitate your informed evaluation of the benefits and drawbacks of experimentation, in order to assess whether an experiment is well-suited for your research question.

By the end of this module, you should be comfortable with the following terminology: fundamental problem of causal inference, treatment, control, random assignment, internal validity, external validity, and convenience sample. You should also be aware of the avenues open to you to conduct experiments for an independent research project.

## 1. What is an Experiment?

In common parlance, people often say “experiment” when they mean “study.” All experiments are studies, but not all studies (in fact the vast majority of studies) are not experiments. In social science research methods, labeling a study as an experiment means something very precise.

Morton and Williams (2010) write that an experiment is conducted when “a researcher intervenes in the data generating process by purposefully manipulating elements of that process” (p. 42). Researchers want as much control as possible over their manipulations to minimize the interference of confounding factors. The most common way of controlling confounding factors is to compare experimental results to a baseline where all observable conditions are identical except for the presence of the manipulation. In within-subjects designs, subjects serve as their own baseline: all facets of the study are held constant and the same subjects experience all manipulations.

However, between-subjects designs are more versatile and more common, and the selection of an appropriate baseline comparison, a control group, is essential.

Consequently, when most political scientists use the term “experiment,” they refer to a study where subjects (or whatever entity we are studying) are randomly assigned to different treatments, where the treatments are causal interventions. Druckman et al. (2011) write “In contrast to modes of research that address descriptive or interpretive questions, researchers design experiments to address causal questions. A causal question invites a comparison between two states of the world: one in which some sort of intervention is administered (a treated state, i.e. exposing a subject to a stimulus) and another in which it is not (an untreated state)” (pg. 16).

## 2. When are Experiments Most Appropriate?

When is an experiment the optimal design for a research question? First, if you are interested in studying causation, an experiment is the cleanest way to show a causal relationship between two variables. The biggest problem plaguing many observation studies is assessing whether one variable (the independent variable) actually causes another variable (the dependent variable). For example, does television news actually influence public opinion? It could be the case that people likely to watch television news differ in important ways from those people who do not, so in an observational study, we cannot pinpoint the role of television news in actually changing opinion (Morton and Williams 2010, p. 13). This kind of question is well-suited to an experimental design.

Second, if you are interested in human behavior, experiments often provide the most straightforward way to understand the mechanisms that explain why people do what they do. Demonstrating a relationship between two variables in observational data does not tell you how that relationship actually operates, while experimental techniques can isolate the process of influence.

Finally, some research puzzles are more conducive to experiments than others. We would never want to randomly assign individuals to experience civil war, nor would we want to randomly assign some people to be disenfranchised, for example. However, a special kind of study, natural experiments (described more below) can help researchers leverage the concept of random assignment even if they are not the force actually assigning the treatment. A creative experimentalist can design a treatment to ethically test a wide variety of political and social phenomenon. In recent years, there have been papers published in political science journals utilizing a wide variety of experimental

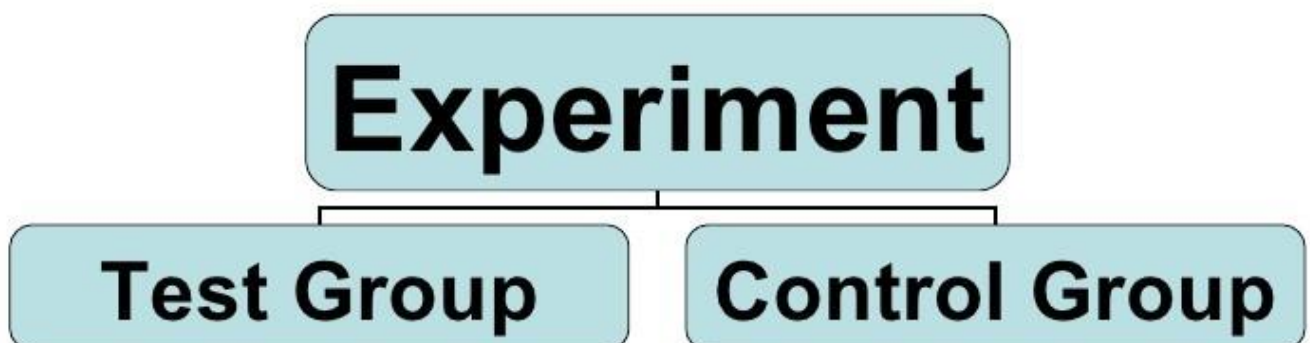
stimuli to test questions like a “get out the vote” experiment on 61-million Facebook users.

## Causal Inference: Key Concepts

### 3. Fundamental Problem of Causal Inference

If causal questions involve comparisons between two states of the world, the fundamental problem of causal inference declares that we cannot simultaneously observe those two states of the world, one in which an individual is treated and one in which the individual is not. We cannot directly overcome this reality but we can work around it by using random assignment. Random assignment means that “each entity being studied has an equal chance to be in a particular treatment condition” (Druckman et al. 2011, p. 16-17). While we expect there to be random sampling variation (we don’t expect the two treatment groups to be exactly identical on all characteristics), this variation is due to chance alone, meaning it is uncorrelated with the treatment subjects receive. Except for the intervention itself, we can assume that randomly assigned groups are probabilistically equivalent. In other words, we can assume that the control group acts like the treatment group would have if the treatment group had not received the treatment, allowing us to assess the average treatment effect.

### 4. The Meaning of “Control”



Morton and Williams (2010) define control in experiments as a researcher fixing or holding constant elements of the data generating process in an experiment to better measure the effects of the manipulations (treatments) (p. 44). This control is done to minimize the effect of observable and unobservable confounding factors in an experiment. Researchers have the most control in a laboratory environment, where they are able to measure more factors to make them observable and hold constant more of the factors influencing how subjects receive the treatments.

## 5. Random Assignment vs. Random Selection

Although these two terms sound similar, they are distinct concepts. Random sampling refers to the process by which subjects are selected from a population such that every potential participant has an equal chance of being selected for the study. An experiment does not require a random sample, although its external validity can be increased if a random sample is used. Random assignment refers to the procedure for determining how subjects in an experimental study are assigned to treatment groups.

**Within and Between Subject Designs:** In a between-subjects design, each subject is randomly assigned to a single treatment, and the researcher compares the treatment groups to each other or to the control group. In a within subjects design, individuals serve as a control to themselves, and a subject is observed before and after receiving the treatment. While within-subjects designs are appropriate in many instances, the concern is that there can be many confounding factors with the delivery of the treatment itself, making it harder to establish a causal connection between the treatment and the outcome variable.

## 6. Internal vs. External Validity

Experiments offer very strong internal validity—the ability to assess the degree to which the independent variable (the treatment you are manipulating) actually causes your dependent variable. The best experimental research designs include appropriate baseline or control groups, so that researchers can make the most precise causal inferences possible. Conversely, experiments are weaker in their external validity, or the degree to which the causal relationship demonstrated in the experiment could be replicated in other contexts, with different people, or with different operationalizations of the treatment or outcome variables.

## 7. Types of Experiments

**Survey Experiments:** The term survey experiment refers to a broad category of stimuli where an individual decision-making experiment is embedded within a survey (Morton and Williams 2010, p. 79). These can take many forms. In the most basic form of survey experiments, researchers can vary question wording or question ordering to assess whether these factors affect respondents' answers. More complicated survey experiments involve vignettes, where subjects read about hypothetical scenarios in which researchers manipulate various features of the situation. Increasingly common are list experiments, a form of a permissive design (Druckman et al. 2011, p. 107) that removes pressure favoring one response over another, allowing respondents to more honestly convey their behaviors or attitudes. In a list experiment, subjects report the number of statements with which they agree, not the statements themselves. Subjects are randomly assigned to one of two groups, a baseline condition with a list of statements of length  $n$ , and a treatment condition with a list of identical statements but where an additional, potentially objectionable statement is inserted (length  $n+1$ ). By comparing the mean number of items agreed with between the two groups, a researcher can estimate the proportion of people who agreed with the objectionable statement.

**Lab Experiments:** The basic requirements for a lab experiment are that subjects are recruited to a common physical location where the study takes place, and a researcher directs the behavior of the subjects. Lab experiments in political science have their root in the experimental techniques developed in other disciplines. Experimental and social psychology have had an enormous influence on the field of political psychology, and political psychology experiments typically adhere to the norms established in these disciplines. Subjects are paid a flat fee (or given class credit if student convenience samples are used) and researchers assume that subjects will behave sincerely. Researchers often go to great length to make their studies seem naturalistic and may use deception. Conversely, in experiments modeled after techniques used in behavioral economics, subjects are paid based on the choices they make in the experiment under the assumption that the payment motivates subjects to behave in a more natural fashion. These experiments tend to be abstracted away from reality and are used to test theories and formal models. Deception is rarely used. Finally, "lab in the field" experiments occur when researchers bring the controlled environment of the lab into a field environment, in an effort to conduct experiments on a wider variety of populations and to increase the external validity of the results of the experiment.

**Field Experiments:** Field experiments occur when a researcher's intervention takes place in subjects' natural environment. One of the most common types of field experiments in political science is "get out the vote" experiments, where registered voters are

randomly assigned to receive an encouragement to vote in some form (mailing, door hanger, phone call, etc.) and the turnout rate of the treatment group is compared to a group that did not receive the encouragement. However, there are many other clever and interesting manipulations that take place in the natural environment.

**“Natural” Experiments:** The term natural experiment is applied to studies where researchers take advantage of random assignments that occur naturally in the world. One of the most common examples of this is a lottery. Researchers often label their study as a natural experiment even if the events that happen to some people but not others are not determined in a truly random way, arguing that the process that led to the assignment of the “treatment” was administered as if it were random. While these are not true natural experiments, they can be leveraged in meaningful ways to help researchers understand a wider variety of phenomena.

## Implementing an Experiment

### 8. Finding Subjects

All research designs pose unique challenges. One of the biggest in experimental designs is the need to recruit subjects to take your study. In an ideal world, every experiment would be conducted on a random sample of the population of interest to achieve both strong internal and external validity. But one of the main benefits of experimentation is that you can achieve strong internal validity on a convenience sample. Therefore, experiments disproportionately rely on samples that are not randomly selected and researchers make no claim that their sample is representative.

Although reliance on non-probability convenience samples is widely accepted, it is still a considerable amount of effort to identify and recruit subjects to participate in your study. To assist you in this process, the SSRMC implements the Omnibus Project every semester. The project coordinates and streamlines the development of a student subject pool (a convenience sample) for faculty and student research. Project coordinators collect data for a set of common variables, such as demographic information and political covariates, which are provided to all researchers submitting proposals. Individual researchers submit their survey questions and customize their own portion of the survey instrument. The subject pool is comprised of students enrolled in departmental classes; some instructors will require participation or offer extra credit to students to participate in the study.



Another cost-effective option is to use Amazon’s Mechanical Turk (MTurk) service, a website that allows researchers to publish tasks (HITs or Human Intelligence Tasks) and provide payment to subjects who choose to participate. Those who request a task can limit the availability of the task to respondents who meet certain qualifications, such as age or location. Studies using samples from Mechanical Turk have been published in the top journals in political science and have been found to replicate important experimental findings in psychology. While MTurk is still a convenience sample, it is more representative of adult populations than undergraduate samples or samples populated from those who respond to web advertisements.

If you have used the above samples to collect pilot results that are promising tests of your hypotheses, with the support of a faculty member, you can put together a proposal for the Time Share Experiments in the Social Sciences. This project, sponsored by the National Science Foundation, allows researchers to submit experimental proposals for consideration to be fielded on a representative sample of American adults on an Internet survey platform.

## 9. Delivering the Treatments

# Institutional Review Board

First and foremost, studies involving human subjects always require approval from the Institutional Review Board (IRB). Applications to the IRB must include documentation that you have successfully completed the ethics training mandated by the federal government. Depending on the nature of your experiment, your study may be exempt from full review of the board, but that is a decision made by the chair of the IRB, not by

you as the researcher. You should allow at least three weeks for your study to be approved before you can collect data.

Political scientists tend to conduct three main types of experiments: survey experiments, lab experiments (including lab-in-the-field experiments), and field experiments.

If you want to do a survey experiment, the simplest option is Qualtrics, a software program for which the College maintains a subscription for both computer-based and mobile platforms (for experiments you want to do remotely). Qualtrics can be programmed in very sophisticated ways to randomly assign subjects to different treatments.

One of the drawbacks of survey experiments that subjects take in the comfort of their own environment is a lack of control on the part of the researcher in controlling that environment. For example, if you are conducting your survey on an Internet sample, subjects are able to browse the Internet or walk away from the computer while they are taking your study. This is most problematic if there is reason to think that some facet of your treatment might make subjects more likely to get distracted or visit other webpages, for example if your treatment is very long or if you ask post-test questions about political knowledge, where subjects might feel inclined to seek out the correct answers online. Programs like Qualtrics have some built in functionalities to be able to detect this.

However, in some instances, you may want to conduct a survey experiment in the laboratory, either to exert more control over the experiment or because you want to deliver a treatment that is not well suited to the online or phone format. In that instance you may want to use the research lab facilities of the SSRMC. If you are doing a lab experiment and you are interested in delivering media (images, audio, or videos) to subjects in a laboratory environment, one of the most popular software programs to do so is SuperLab. The SSRMC currently has licenses for this program on one computer. An example honors thesis that relies on SuperLab for stimulus delivery can be found [here](#).

If you are interested in doing an experiment in the field, academics frequently partner with outside organizations—such as campaigns or advocacy organizations—because it is often difficult to get access to the large subject pools necessary to conduct these studies. William & Mary undergraduates have successfully done this, and this honors thesis is a great example.



## 10. Analyzing the Data

There are more sophisticated ways of analyzing the data from experiments, and methodologists are constantly developing new ways to extract less biased estimates of the causal effects in a study. While these more advanced approaches are beyond the scope of this introductory module, the resources listed at the end of this module contain more in-depth information.

## 11. Considerations and Cautions



Studies involving human subjects must go through ethics review for good reason, and this is especially important in the case of experiments, where researchers manipulate the environment or stimuli to which participants are exposed.

Different experimental traditions in the social sciences have different norms. One of the biggest differences is in the instructions that researchers give participants about the nature of the experimental tasks. Experiments rooted in psychology often allow researchers to use mild deception in their instructions if the researcher thinks that knowing the true purpose of the study would alter the way that participants behave. The SSRMC allows deception in studies, as long as that deception is approved by the Institutional Review Board. However, deception is almost always avoided in experiments rooted in economics. A second major difference between economic and psychology experiments is whether (or how) subjects are incentivized for their participation.

The classic tradeoff in experimental design is between internal validity and external validity. While experiments have high internal validity, to varying degrees, they may lack external validity, the ability of a researcher to make claims about how the results of the

study would generalize and hold up in different contexts. One particularly common generalizability concern stems from differences in the sample used in your experiment compared the population to which you want to generalize. When is it problematic to generalize the findings of an experiment conducted on a convenience (often, student) population? First, it is important to know how the student population differs from a more representative population. The obvious answers are in age, education level, and geographic location.

But other factors can matter as well. The key question to ask is “how else are college students different in a way that should affect the strength or direction of the causal relationship I am testing?” If your convenience sample is different in a way that makes it harder to find the relationship you observe, then you can assert that your study likely underestimates the relationship between the variables in a more representative population (a testable proposition!) However, if your sample makes it easier to find effects, then generalizability concerns become more serious. Sometimes, these concerns are very large. For example, college students are particularly susceptible to conformity (Sears, 1986), which could be important depending on the nature of your study. The Omnibus Project draws from courses in the government and international relations program, suggesting that our participants have a greater interest and level of knowledge about politics. These factors may help or hurt your ability to make claims about how general your findings are.

More broadly speaking, researchers have commented on the abundance of WEIRD subjects in experimental studies: subjects that are Western, educated and come from countries that are industrialized, rich and democratic. This has been written about considerably in the popular media and has been addressed in academic research as well.

## 12. Resources

There are many excellent resources available online, through SWEM library, and through the SSRMC research methods collection. The module below draws on information from these sources, but there is much more detail available in the original sources.

### **Experimental Design**

- Druckman, James N. et al. 2011. *Cambridge Handbook of Experimental Political Science*.
- Rebecca Morton and Kenneth Williams. 2010. *From Nature to the Lab: Experimental Political Science and the Study of Causality*
- McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5:31-61.

- McDermott, Rose. 2002. "Experimental Methodology in Political Science." *Political Analysis* 10(4):325-342
- Druckman, James, Donald Green, James Kuklinski, and Arthur Lupia. 2006. "The Growth and Development of Experimental Research in Political Science," *American Political Science Review* 100:627-635
- Gaines, Brian et al. 2006. "The Logic of the Survey Experiment Reexamined." *Political Analysis*
- Imai, Kosuke et al. 2011. "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review*

•  
**Field Experiments**

- Green and Gerber's Field Experiments book
- Yales ISPS Field Experiments Initiative as well as there Data Archive