Experiments, Political Science

Rose McDermott
University of California, Santa Barbara, California, USA

Glossary

artifact An exogenous variable that is created by some aspect of the experimental manipulation and which can vary with the independent variable. The effects of these variables can then become confounded with the experimental findings.

between-subjects design An experimental design in which each person is exposed to only one of the experimental manipulations. Differences are then measured between groups.

control condition The condition of subjects in an experiment who receive no treatment or manipulation.

direct replication Repeating an experiment as closely as possible to determine the reliability of results.

experiment A method of investigation with an independent variable, a dependent variable, high levels of control on the part of the investigator, and the random assignment of subjects to treatment conditions.

experimental condition The condition of subjects in an experiment who are exposed to the manipulated variable or treatment.

interaction effect When the outcome or effect of one variable varies with the outcomes or effects of a second variable. This occurs in factorial design experiments.

systematic replication Repeating an experiment in such a way that certain aspects, measures, or protocols of the original experiment are systematically varied. These replications serve to clarify or extend aspects of the original experiment.

within-subjects design An experimental design in which each person is exposed to all treatment conditions. Thus, each person serves as his or her own control and differences are found within each individual.

Experiments constitute the gold standard of many forms of scientific inquiry in a variety of fields, including hard sciences, such as biology and physics, as well as social sciences, such as psychology and behavioral economics. Because of the methodological demands of experimental control and the randomization of subjects, the experimental method offers unparalleled leverage in determining causation. Experiments can be used to great effect in uncovering causal relationships in political science as well, although they have not yet been as widely employed as other forms of methodological inquiry, such as formal or quantitative work. This article explores why experiments offer a potentially useful form of social measurement and investigation. Various basic aspects of experimental design will be covered, along with potential sources of error. The advantages of experimental design, as well as nonexperimental alternatives, will be discussed. Finally, the ethics of experimentation require mention.

Why Experiments? Advantages of Experimentation

Experiments offer the best scientific method available for finding out whether or not one thing causes another thing. All other methods can offer interesting, important, and persuasive arguments about causality, but only experiments can actually prove causality. For example, many people knew that smoking caused lung cancer as a result of all the correlational data showing the relationship between smoking and lung cancer. However, the tobacco companies could still engage in plausible deniability because these relationships were correlational and they could argue that these associations were spurious and that other factors, such as genetic predisposition, could cause both smoking and lung cancer. However, once experimental work at the cellular level took place and demonstrated that tar, nicotine, and other components of smoke caused lung tissue to die under controlled conditions, the causal link was established and it became much more difficult for the tobacco companies to argue that the relationship was spurious in origin. This causal link remains critical in addressing many questions of great importance in political and social life as well.

There are at least three ways in which experiments differ from other forms of social measurement. First, many political scientists begin their inquiries by looking for naturally occurring phenomenon in which they are interested and then seeking to examine them in a variety of ways. Thus, if a scholar was interested in military coups, he could find states where they occurred and then study them in some systematic fashion. In these cases, methods of inquiry can include field work, interviews, surveys, archival exploration, and so on. On the other hand, an experimentalist tends not to wait for events to occur. Rather, an experimentalist goes about creating the conditions that produce the events of interest.

Thus, if a scholar is interested in how leaders make decisions under conditions of stress, he can design an experiment in which he asks subjects to make various decisions and then imposes stress on them while they are completing these tasks. Experimentally induced stress can take many forms, depending on the interest of the investigator. One can turn up the heat in the room, impose a tight time limit, give an impossible task with false-negative feedback, make the room too loud or too bright, or simply create a crisis in the midst of the task, such as a fire or medical emergency. Obviously, some issues and questions of concern lend themselves more readily to experimental creation and manipulation than others, but clever experimentalists can gain a great deal of control by creating their own conditions of interest. The real advantage of this opportunity lies in the experimentalists' ability to measure and capture exactly what they are interested in studying. Because of this, experimenters are not easily led astray by irrelevant forces. In other words, experimentalists do not need to test or replicate every aspect of the real world in order to investigate their areas of interest. Rather, they need only examine the key relationships between variables that they suspect have causal impact on the outcomes of interest. In this way, an experimentalist draws on theoretical ideas and concepts to derive hypotheses and then designs a test, or multiple tests, to explore the variables and relationships that appear to be causal in nature. Therefore, experimentalists do not need to re-create the external world in their laboratories in order to achieve meaningful results; by carefully restricting their observation and analysis to those variables deemed central to the relationships of interest, they can test competing alternative causal hypotheses within controlled conditions.

Second, and related, experimentalists can control and systematically manipulate the treatment conditions to which their subjects are exposed. This ability to control the events and measures allows an investigator to expose subjects under different conditions to exactly the same situation except for the variable being tested and manipulated. Therefore, if the scholar observes a systematic difference between groups, he knows exactly what caused the change because only one thing varied between conditions. In less controlled conditions in the real world, it can often prove extremely difficult, or even impossible, to find events that vary systematically enough to allow for such comparisons to be made. Two different dictators, for example, certainly differ in more ways than they are similar and thus systematic comparisons may prove difficult and dangerous. Yet analogies between them remain accessible and powerful nonetheless, as when Saddam Hussein was compared with Adolf Hitler by members of the Bush administration. Although such comparisons can prove strongly evocative, problems arise because real systematic comparisons between actors remain impossible.

Third, experimenters can decide who gets exposed to what treatment condition. The ability to randomize subjects to manipulations provides the single most powerful tool available to experimenters. Random assignment to condition ensures that observed differences do not result from preexisting systematic differences between groups or individuals. Because of this procedure, subjects cannot self-select into a particular condition due to personal proclivities or systematic environmental pressures. With random assignment, condition should be determined by chance alone. The best experimenters tend to use random number tables or some other systematic method to ensure truly random assignment of subject to condition. When this happens, any preexisting systematic differences between subjects should cancel each other out. As Aronson *et al*. (1990, p. 38) explain: "In an experiment, the experimenter both has control over what treatment each subject receives and determines that treatment by assigning subjects to conditions at random. If the subject receives a treatment that has truly been determined at random, it is virtually impossible for a third variable to be associated with that treatment. Consequently, such a third variable cannot affect the dependent variable."

These elements of control and randomization, along with the ability to create the experimental environment, allow investigators to determine whether certain factors really do cause other ones or are merely correlated with them. The null hypothesis in these circumstances typically assumes that the independent variable or variables of interest have no effect on the dependent variables of interest. When systematic observed differences are found between conditions, investigators can feel confident that these discrepancies were actually caused by the manipulations they created. In this way, the primary benefit of experiments lies in their unparalleled ability to uncover causal relationships. As Aronson *et al*. (1990, p. 39) sum up, "the major advantage of an experimental enquiry is that it provides us with unequivocal evidence about causation. Second, it gives us better control over extraneous variables. Finally, it allows us to explore the dimensions and parameters of a complex variable."

Experimental Design and Potential Sources of Error

Variables

As Aronson *et al*. (1990, p. 114) write, "experimental design refers to the selection and arrangement of conditions." Experimentalists typically structure particular experimental protocols that outline what subjects are expected to do and indicate which measures will be used to record outcomes of interest. Typically, as in many other methods of social inquiry, experimenters start with a hypothesis about the relationship between two or more variables. These hypotheses, as in other methods of inquiry, can come from a variety of different sources. They can derive from theoretical understandings, problems in the real world, previous research, and so on. The experimenter then seeks to determine whether the independent variable, or putative cause, influences the dependent variable, or effect, in the manner expected. Experiments then seek to explore how and in what way these variables are related. As Underwood (1957, p. 87) noted, the most basic goal of experimentation is to "design the experiment so that the effects of the independent variables can be evaluated unambiguously." The design of experiments is often extremely time-consuming because it can be hard to figure out how to operationalize the conceptual variables of interest into an engaging and coherent activity for subjects. This process of transforming ideas into concrete events is called empirical realization.

Conditions

Several specific aspects of experimental design deserve mention. First, perhaps the most important aspect of experimental design relates to the creation of the experimental conditions themselves. Experiments require at least two conditions in which to measure the independent variable in order to see whether it has had any effect on the dependent variable; obviously, one condition simply constitutes a constant and no effect can be demonstrated. Typically, an experiment is designed with a control condition and an experimental one. This means that in one treatment group, the control condition, subjects engage in some benign but related task, which should take a similar amount of time, require the same kind and amount of information, involve the same experimenters, and engage a similar amount of interest as the experimental condition. In particular, it remains critical that subjects in the control condition receive the same background information as those in the experimental condition, so that ostensible findings do not simply result from differences in the information provided to the two groups at the outset. In practice, this equivalency of experience can often be very difficult to achieve. In the other group, the experimental condition, subjects receive the experimental manipulation. In practice, the control and experimental conditions often translate into the presence or absence of the independent variable. The control condition does not have the independent variable of interest, but the experimental condition does. Recall that the null hypothesis usually assumes that no relationship between independent and dependent variables exists, until proven otherwise.

Although random assignment of subjects to conditions allows the investigator to rid himself of systematic sources of error, random error, or background noise, can still pose a problem. In order to avoid this problem, experimenters often hold as many variables as possible constant. Another option for investigators is to allow certain variables, hopefully the less important ones, to vary randomly. The third alternative is to introduce systematic variation by adding a second (or more) independent variable for examination. When this occurs, two experimental treatment conditions exist, rather than one control and one experimental condition.

This strategy of using two experimental conditions, rather than one treatment and one control condition, occurs, for example, in medical experiments, where no treatment would constitute an unethical option. The control condition then becomes the standard treatment of care, compared with the experimental condition provided by a new medication or procedure. Similarly, in political science, many settings would not lend themselves to ethical experimentation. For example, it would not be ethical to try to examine the impact of torture on prisoners of war by re-creating such techniques in a laboratory. Similar examples of unethical experimentation could be imagined in field settings, such as investigating the impact of suicide bombing on policy preferences. More will be said about field experiments later.

Between- and Within-Subjects Designs

Experiments can take the form of within- or between-subjects designs. Both types offer certain advantages and disadvantages. There is no one right way to conduct an experiment; investigators decide whether to use a within- or between-subjects design depending on the circumstances, based on how and what they want to study. In between-subjects designs, each subject is exposed to only one condition and the differences between groups are then examined. Because each subject is exposed to only one condition in this design method, observed differences exist between groups or individual subjects. Randomization should prevent the introduction of systematic bias into the subject pool, but there remains a very small possibility that some systematic personality differences underlying each population can introduce some kind of systematic error. In within-subjects designs, each subject serves in effect as his or her own control. In this design, each subject encounters every condition in sequence. Thus, the same group of people constitute the entire experimental variance.

The advantage of a within-subjects design is that it prevents the introduction of any underlying systematic individual differences from contaminating the experimental findings. The problem lies in the risk of potential contamination of experimental results in one condition from the subject's prior experience of another condition in the same experiment. These within-subjects designs prevail in experiments that ask subjects to make judgments about hypothetical scenarios or individuals. They remain a very cost-effective option for collecting large amounts of data relatively quickly.

The disadvantages of within-subjects designs lie in their potential for interexperimental contamination. When an investigator examines an issue or topic that has great impact, because it appears quite vivid, concrete, or emotional in nature, within-subjects designs risk the threat of a so-called carryover effect, in which exposure to one condition unduly affects or contaminates a subject's response to the other condition.

Overall, if an investigator is interested in examining how people make judgments in particular situations or contexts, a within-subjects design can be ideal. However, as Greenwood warns, if a scholar wishes to see how a particular person, event, or condition affects people in isolation, certain precautions should be used to make sure that within-subjects designs remain valid. For example, making sure that stimuli are counterbalanced in their order of presentation will help avoid systematic contrast effects that might produce artifacts in the experimental results. In other words, if an experimenter wants to examine the effect of fear versus anger on decision-making, some subjects should be presented with the fear condition first, while others receive the anger condition first. This strategy ensures that the contrast between the most recent mood state does not systematically contrast with, and affect, the subject's performance on the dependent measures in a way that produces systematic, but confounded, results.

In addition, some experimenters will combine within- and between-subjects designs to avoid the pitfalls inherent in either one. A final technique that can be used to help overcome the contamination effects of a within-subjects design without invoking the difficulties of potential underlying personal similarities inherent in a between subjects design is called matching. In matching, experimenters try to closely match subjects based on particular criteria, such as age, race, sex, and then expose one subject to the experimental condition and one subject to the control condition. This strategy works best when the dimension on which subjects are matched is similar to the dependent variable. Intelligence, education, and profession provide examples of such categories in which people can be matched because these factors can influence a wide variety of other behaviors and outcomes in systematic ways. Sometimes, particularly in medicine, matching can take place based on the dependent variable. If someone wants to examine factors that led to a certain disease, he might match individuals who are similar in terms of age, sex, race, education, and occupation, but who differ in their disease status, where one might be infected with hepatitis and the other not, to examine the differences between them. To be clear, matching in no way substitutes for the random assignment of subjects to condition; indeed, in matching, subjects are still assigned to treatment conditions at random within paired blocks. For example, one could not take groups of people who naturally differ on some dimension such as race, construct two new groups based on party identification, and then treat one group as an experimental group and another as a control group. Rather, matching provides an additional control, typically used in within-subjects designs, that can further protect against the impact of systematic personal variations on experimental results.

Interaction Effects

In an experiment, investigators typically look for main effects, that is, the direct effect of the independent variable on the dependent variable. One of the potential methodological advantages of introducing more than one experimental condition at a time, however, lies in the ability to uncover interaction effects between variables that might have been hidden if only one variable were examined at a time. Interaction effects occur when one independent variable has different effects depending on the impact of a second independent variable. For example, an interaction effect would be discovered if one found that party identification had a different impact on voting behavior depending on race.

Obviously, many variables can affect social and political behavior. Therefore, depending on the topic under investigation, sometimes an experimenter will hold certain things constant, sometimes he will allow certain variables to vary at random, and sometimes he will introduce a systematic variation in independent variables. When such a systematic variation exists, precautions must be taken to avoid artifactual findings, those results that may appear to look like real interaction effects, but in fact result from systematic sources of error that occur when other exogenous variables shift with the independent variables of interest. When this happens, it can appear that those exogenous factors influence outcomes in a systematic way, but in fact those findings remain spurious. Such variables can confound the results of the experiment.

Two kinds of artifacts pose relatively consistent potential problems for experimenters in general. The first concerns demand characteristics, whereby subjects change their behavior either because they know they are in an experiment and are being watched or because they want to please or impress the experimenter in some way. This can bias results in systematic, but irrelevant, ways. Second, experimenter bias can produce another kind of systematic artifact. This occurs when subjects change their behavior either because they think they are given certain cues about what the experimenter wants or because the experiment is designed in such a way that they actually are given such cues. This creates particularly problematic effects when these subtle pressures actually encourage subjects to behave in inauthentic ways that nonetheless support the experimenter's hypotheses.

Replication

Obviously, repeating experiments constitutes a fundamental part of the scientific method in order to determine the reliability of results. Replication also means that results remain falsifiable. However, the reasons for, and forms of, replication shift depending on the purpose.

Replication means only that an investigator repeats an experiment. In a direct replication, the experimenter tries very hard to reproduce closely as many elements of the original experiment as possible. Similar results typically confirm the reliability of the findings. These kinds of replications are unusual, however. More common is a replication of some elements of the original experiment coupled with some kind of extension into a new and different aspect of the original work. Typically, this occurs when some aspect of the original experiment stimulates some additional hypotheses about the phenomenon under investigation.

Another type of repetition takes the form of systematic replication. In this situation, an experimenter systematically varies some aspect of the original experiment. Most commonly, this takes places to clarify some unexpected or unexplained aspects of the original experiment or to seek some additional information about some previously unconsidered feature of the original study. This can occur, for example, when a new alternative explanation, which did not occur to anyone beforehand, presents itself after the first experiment is complete. Even though it can seem time-consuming and expensive at the time, pilot testing almost always ends up saving time, money, and enormous amounts of aggravation by helping to reduce these kinds of last-minute complications. Sometimes, further experimental replications are planned from the outset because many variables of interest exist, but any one experiment that includes every factor would be too complex or confusing. As a result, a sequence of experiments may be planned in order to address each variable systematically in relation to other variables of interest.

Cross-cultural replications no doubt present the most difficult challenge to experimenters. The difficulty with cross-cultural replications lies in the fact that exact replications do not always translate directly into another culture, either linguistically or socially. Although cross-cultural work is becoming increasingly important and common, it can be difficult to translate stimulus materials from one culture to another without introducing tremendous variations. Because subjects in different cultures may not interpret materials the same way, identical manipulations may prove meaningless because experimenters cannot assume that what would constitute, say, self-determination in the United States would mean the same thing to the Kurds in northern Iraq and southern Turkey. However, changing experimental materials, even with simple language translations, can profoundly alter the experimental question, technique, and findings. The question becomes: are differences found true differences—do they result from different interpretations of the same materials, the same interpretation of different materials, or, most likely, an entirely different understanding of the meaning and purpose of experimentation itself?

This does not mean that cross-cultural research should not be pursued. Rather, experimentalists should remain cautious in designing materials. They should try as best as possible to ensure that materials used in different cultures stay conceptually similar in terms of the perspective of the subjects themselves. The subjects' experience should be as similar as possible, even if the stimulus materials differ somewhat. Furthermore, cross-cultural research might benefit from the knowledge of comparativists who can help translate materials across cultural boundaries in conceptually similar ways. Finally, experimentalists can feel more confident about their results to the extent that they can replicate their studies both within and across different cultures. Similar results from different studies in the same culture add credibility to the findings, as do consistent cross-cultural differences across experiments. Wide variation in results may indicate that the differences in materials and techniques simply overwhelm any true differences that might exist.

Subjectivity in Design

The best and most clever experimentalists put a lot of time and energy into designing successful experiments.

At least two factors really enhance the ability of an experimentalist to design experiments of high quality and impact. First, strong experimentalists do their best to put themselves in the place of their subjects. They try to imagine what it would be like to enter their experimental condition blind, for the first time, without a clear sense of the purpose or the working hypothesis. In undertaking this fantasy, it can be helpful to assume that although subjects may be cooperative and want to help, they will nonetheless most probably be trying to figure out what the experimenter wants. A further step in this process involves careful debriefing of subjects after early or pilot testing. They should be asked what they experienced, what made sense to them, and what was confusing. It should be ascertained whether they were able to determine the purpose, if they were not told outright, or, if not, was there any pattern to what they thought was going on that differed from the real intention of the experiment?

A second way to help improve experimental design is to try to imagine how a different potential pattern of responses might appear. If the results were not what was wanted or expected, why might that be the case? What other potential patterns might emerge? How can new tests be designed to examine these alternative hypotheses? Being able to anticipate alternatives prior to actual large-scale testing can really help clarify and improve experimental designs beforehand.

Nonexperimental Alternatives

Why?

In many cases, experiments may not be the optimal form of investigation in political science. There may be other forms of inquiry that are better suited to the problem under investigation. However, when this happens, there may be other forms of nonexperimental research that nonetheless remain related to experimental work that might be worth considering.

There are several reasons that researchers might want to seek nonexperimental alternatives to experimental work. The first and most important lies in the fact that many questions of interest are not causal questions. For example, in trying to predict the outcome of a particular election, pollsters may not need to know why certain people vote in particular ways; they just want to be able to survey people in such a way as to obtain an accurate prediction. Other issues that are important to political scientists also fall outside the causal realm. For example, sometimes investigators merely want to demonstrate that a particular phenomenon exists, such as incipient democracy in a formerly communist country.

Second, there are many situations in which experimental work would be impossible to conduct or unethical to perform and thus a researcher will look to alternative methods. Sometimes it is simply impossible to impose the variable of interest on people randomly, as it would be with race or sex. Other times, it would be immoral and unethical to do so. For example, no one would want to start a war or induce a famine or plague in order to examine the effect of these phenomena on economic structures. People can study these things after a naturally occurring event, of course, but, again, such work will not constitute experimental work.

Finally, nonexperimental work can supplement experimental work and vice versa. In this way, nonexperimental work does not constitute a substitute but rather an addition to experimental work. This occurs frequently in behavioral economics, for instance, when formal models are tested with experiments and then further refined. There is no reason that similar kinds of interplay cannot take place between game theoretic models in political science and experimental validation and refinement of such models.

Correlational Work

Correlational work does not involve the administration of any experimental treatment manipulation or condition. As noted previously, correlational work can offer useful, interesting, and important hypotheses about the relationship between variables, but it cannot demonstrate causality. Partly this results from the possibility of a third spurious cause influencing both factors. On the other hand, correlational work can establish that no causal relationship exists. If there is no correlation between variables, there is no causation. Occasionally, some very complicated relationship that depends on several mediating factors may exist, but at least a direct or simple causal relationship is ruled out if no correlation exists. In other words, it cannot prove, but it can disprove, a causal relationship between variables.

A second problem with correlational work results from confusion about directionality. In correlational work, it can be difficult to determine what the cause is and what the effect is in many circumstances. Some techniques that seek to partially ameliorate the problem of directionality exist. The cross-lagged panel technique collects correlational data on two separate occasions. Thus, in other words, investigators obtain information on variables 1 and 2 at times 1 and 2. By looking at the same variables in the same populations over time, some element of directionality might be illuminated.

Pseudo-Experimental Designs

Pseudo-experimental designs represent nonexperimental designs in which the investigator nonetheless maintains some control over the manipulation and the collection of data. In other words, an investigator does create and administer an independent variable to subjects and then measures its effect on some dependent variable of interest. However, because it is either not possible or not desirable to have a control group, proper comparisons across conditions are not possible. Without such appropriate comparison groups, any one of a wide array of factors could have led to the observed results. Because of this, researchers are not able to fully determine the effects of their intervention.

Several types of pseudo-experimental designs exist. The one-shot case study design, often referred to as the posttest-only design, examines a single group on a single occasion. These kind of tests appear common after particular disasters, in order to examine their effects on a particular population. Such tests might be conducted after an earthquake, a terrorist event such as 9/11, or other man-made disasters such as the explosion of the space shuttle *Columbia*.

Other types of pseudo-experimental designs include one-group pretest—posttest design studies that examine a group of people both before and after a particular event. As Campbell and Stanley warn, the trouble with this design is that it still poses a number of threats to the internal validity of the study, including problems presented by preexisting history, maturation, testing, instrumentation, and mortality. The static group comparison design constitutes a final type of pseudo-experimental design. In this design, researchers compare two different groups on the same variables at the same time. This design also poses selection threats to the internal validity of the study.

Quasi-Experimental Designs

A quasi-experiment allows an investigator to assign treatment conditions to subjects and measure particular outcomes, but the researcher either does not or cannot assign subjects randomly to those conditions. To be clear, in pseudo-experimental design, the study lacks a control condition, whereas in quasi-experimental design, the researcher does not or cannot assign subjects to treatment conditions at random. This feature actually makes quasi-experiments much easier to use and administer in field and applied settings outside of the laboratory. However, what is gained in flexibility and external validity may be lost in being able to make unequivocal arguments about causality. However, quasi-experiments do allow scholars to make some causal inferences and interpretations, just not fully dependable arguments about causality.

In general, two types of quasi-experimental designs predominate: the interrupted time series design and the nonequivalent control group design. In the former, a within-subjects design is employed to examine the effects of particular independent variables on the same group of subjects over time. Typically, subjects are measured both before and after some kind of experimental treatment is administered. In the latter, a between-subjects design is invoked to measure the impact of the independent variable on different groups of subjects. What remains common to both types of quasi-experiments is the fact that investigators do not or cannot assign subjects to treatment condition at random.

Field Experiments

Field experiments are those that take place outside a laboratory, in a real-world setting. Goznell published the first field experiment in political science in the *American Political Science Review* in 1927. In conducting a field experiment, an investigator typically sacrifices control in order to achieve increased generalizability. In psychology, nonexperimental field testing early in a research program often helps generate novel hypotheses. Once these ideas are tested and refined in laboratory settings, psychologists often then return to field testing to validate their findings. In this way, field studies can prove extremely helpful in specifying the level of generalizability of laboratory results to the real world.

Two important points should be made about field experiments. First, successful field experiments rarely constitute a simple transfer of a laboratory study to a field setting. Investigators have less control in the field; they cannot always control the nature of the treatment, the comparability of subjects, or the impact of unplanned, extraneous factors on outcomes of interest. Most importantly, often subjects cannot be assigned to treatment conditions at random. Second, results from a single field study, no matter how large the population, can rarely be taken on their own, because there are often so many extraneous and potentially confounding factors occurring during the completion of the field study. Figuring out what is actually causing the outcome of interest can be challenging. Such findings typically need to be interpreted within the context of similar studies conducted using other formats, including laboratory experiments, interviews, surveys, or other methods. Results that demonstrate convergent findings across methods allow for greater confidence in accuracy.

Despite these concerns, field experiments can prove quite beneficial for a variety of reasons. Often, behaviors of interest cannot be induced in a laboratory, but can easily be observed in the real world. In addition, experimental findings that replicate in the field increase people's sense of confidence in the generalizability of the results. Obviously, trade-offs between laboratory and field experiments exist and the nature of the research question of interest often determines the appropriate setting for a study.

Rather than creating a setting as experimenters do in a laboratory setting, those who conduct field experiments seek to find the best environment in which to examine their issue of concern. Experimenters should seek out settings that possess the variables and relationships of interest, but as few extraneous factors as possible. Effective field experiments usually share certain characteristics. Settings should involve events or processes that subjects find meaningful and impactful. Experimenters should be able to intervene in some way to test and measure the phenomenon of interest. These settings should contain the elements of coherence, simplicity, and control to the greatest extent possible. Coherence requires that experiments be defined and limited to a specific time and place or event. Focusing on a single main event often increases the likelihood that the setting is both coherent and as simple as possible; in this way, the impact of extraneous factors can be reduced. Field experimenters can often benefit by using a naturally occurring event as a variable of interest. This can happen, for example, by taking advantage of a change in the law to examine before-and-after effects of processes such as segregation, busing, and affirmative action. Often, field researchers who are not able to assign subjects to condition at random can select comparable control groups to try to extricate the impact of extraneous factors on outcomes. When this occurs, investigators need to be careful to ensure the comparability of the control and the treatment groups.

A critical factor to remember in field experiments is that if the experimenter is not able to intervene in such a way as to assign subjects randomly to conditions, then the study is no longer a true experiment, but rather a field study. In these cases, conclusions about causal relationships cannot be justified and results should be interpreted as suggestive at best.

Experimental Ethics

When conducting any experimental work, ethics remain an important consideration. The U.S. Department of Health and Human Services imposes certain guidelines on the ethical treatment of human subjects. Most universities have human subjects institutional review boards to oversee the administration and implementation of these guidelines. These guidelines encompass four aspects.

First, subjects must be able to give their informed consent before participating in an experiment. Experimenters should provide subjects with a clearly written, simple statement explaining the potential risks and expected gains from their participation in the experiment. They should be told that they can stop their participation at any time without penalty. And they should be given contact information about who to go to in case they have any concerns about the ethics of experiments in which they participate.

Second, experimenters are required to take every reasonable precaution to avoid harm or risk to their subjects as a result of their participation. Third, experimenters should provide a debriefing opportunity to all subjects, in which they are told as much as possible about the experiment in which they just participated. In particular, subjects should be told that their information will be kept confidential and that no identifying information will be released without their prior written consent. Often, subjects are told how they can receive copies of the results at the conclusion of the experiment if they are interested in doing so.

Finally, the issue of deception remains a controversial topic. Psychologists continue to employ deception more than behavioral economists. Deception may prove necessary in those instances in which a subject's prior knowledge of the working hypotheses would influence his or her behavior in systematic or inauthentic ways. This bias can hinder the discovery of important processes and dynamics. However, when investigators employ deceptive techniques, institutional review boards remain particularly vigilant to ensure that the use and value of such experiments are carefully monitored.

Conclusion

Experiments provide a valuable tool for the measurement of social variables and processes. They provide unequaled purchase on causal inference through experimental control and the random assignment of subjects to treatment conditions. And careful design can ensure the ethical treatment of subjects during the experimental process, so that the benefits of discovery continue to outweigh the risks posed to subjects.

Experiments provide an unparalleled ability to clarify causality in ways that can reduce confusion about important processes and relationships. By showing the true direction of the casual link, human beings can learn something important about themselves, and possibly, take steps to change environments and institutions that can cause ill. In his famous experiment, Stanley Milgram attempted to try to understand what it was about the German national character that would lead ordinary citizens to become complicit in the atrocities surrounding the Holocaust. Before he tested his hypotheses about the nature of obedience to authority in Germany and Japan, where he expected high levels of compliance, he ran his control group of presumed individualist Americans at Yale. All the experts agreed that less than 1% of subjects would shocks "learners" to their assumed death. Yet, the majority of subjects did so. As the films show, these subjects were not obvious sadists nor did they delight in being cruel to their fellow man. They did not easily or readily comply; they argued, cried, walked around the room, tried to talk their way out of the situation, and were inordinately relieved when they found out their partner was not dead, but they obeyed nonetheless. After Milgram's experimental findings, those observers who discarded Nazi defenses that claimed that they were "just following orders" were forced to confront results that proved that the power of the situation could overcome personal dispositions. One may not be comfortable knowing this, but this insight teaches one the importance of deviance when morality is compromised, the critical significance of opposition to inappropriate authority, and the transcendent knowledge that removing oneself from the situation can provide the best defense against self-destruction.

See Also the Following Articles

Correlations • Cross-Cultural Data Applicability and Comparisons • Experiments, Overview • Explore, Explain, Design • Field Experimentation • Political Science • Quasi-Experiment

Further Reading

Aronson, E., Ellsworth, P., Carlsmith, J. M., and Gonzales, M. (1990). *Methods of Research in Social Psychology.* McGraw-Hill, New York.

Brody, R., and Brownstein, C. (1975). Experimentation and simulation. In *Handbook of Political Science* (F. Greenstein and N. Polsby, eds.), Vol. 7, pp. 211–263. Addison-Wesley, Reading, MA.

Campbell, D., and Stanley, J. (1966). *Experimental and Quasi-Experimental Designs for Research.* Rand McNally, Chicago, IL.

Greenwald, A. (1976). Within-subjects design: To use or not to use? *Psychol. Bull.* **82,** 1–20.

Iyengar, S., and McGuire, W. *Explorations in Political Psychology*. Duke University Press, Durham, NC.

Kagel, J., and Roth, A. (1995). *Handbook of Experimental Economics.* Princeton University Press, Princeton, NJ.

Kinder, D., and Palfrey, T. (1992). *Foundations of an Experimental Political Science.* University of Michigan Press, Ann Arbor, MI.

McConahay, J. (1973). Experimental research. In *Handbook of Political Psychology* (J. Knudson, ed.), pp. 356–382. Jossey-Bass, San Francisco, CA.

McDermott, R. (2002). Experimental methodology in political science. *Polit. Anal.* **10,** 325–342.

McDermott, R. (2002). Experimental methods in political science. *Annu. Rev. Polit. Sci.* **5,** 31–61.

McGraw, K. (1996). Political methodology: Research design and experimental methods. In *A New Handbook of Political Science* (R. Goodin and H. Klingeman, eds.), pp. 769–786. Oxford University Press, Oxford, UK.

McGraw, K., and Hoekstra, V. (1994). Experimentation in political science: Historical trends and future directions. *Res. Micropolit.* **4,** 3–29.

Palfrey, T. (1991). *Laboratory Research in Political Economy.* University of Michigan Press, Ann Arbor, MI.

Plott, C. (1979). The application of laboratory experimental methods to public choice. In *The Application of Laboratory Experimental Methods to Public Choice* (C. Russell, ed.), pp. 14–52. Johns Hopkins Press, Baltimore, MD.

Sniderman, P., Brody, R., and Tetlock, P. (1991). *Reasoning and Choice: Explorations in Political Psychology.* Cambridge University Press, New York.

Underwood, B. (1957). *Psychological Research.* Appleton-Century-Crofts, New York.