Chapter 1

Why Do Experiments?

Murray Webster, Jr.
University of North Carolina–Charlotte, Charlotte, North Carolina

Jane Sell
Texas A&M University, College Station, Texas

I A BRIEF HISTORY OF EXPERIMENTS

Many social scientists, and most physicists, chemists, and biologists, see experimental methods as one of the defining characteristics of scientific inquiry. Although the experiment is far from being the only research method available to the social sciences, its usage has grown remarkably in the years since World War II. Many historical changes are associated with the growth in experimental methods, of which two kinds of changes are especially important: new topics and new technology.

In early decades of the 20th century, sociologists were largely occupied with classifying types and growth of societies or with development of different parts of cities. Following that war, many social scientists became interested in phenomena that can be studied experimentally. In sociology and social psychology, for instance, topics such as interpersonal influence, distortions of judgment, and conformity processes seemed more pressing than they had seemed before authoritarian and repressive societies were common topics. With the new topics came new theories, many of them amenable to experimentation. Economists began to conceptualize strategic game-playing and became interested in behavioral economics; political scientists developed rational choice theories of voting choice; communications scientists began to understand influence processes; sociologists had new theories of social exchanges; and psychologists, whose discipline had used experiments from its beginnings, expanded its study of effects of social factors that appear in the presence of one or more other individuals. New topics and new theory were both congenial to the development of experimental methods in social science.

The second factor was new technology. Starting at a few universities, experimental laboratories were built, followed by laboratories in government and at private research firms. New laboratories required and facilitated development

Laboratory Experiments in the Social Sciences. http://dx.doi.org/10.1016/B978-0-12-404681-8.00001-7

of many kinds of technological advance: coding schemes to record discussion groups, one-way mirrors and, later, television and computers to observe and control communication among researchers and experimental participants, sound and video recorders, and many other elements of contemporary experiments began developing in the years following World War II.

Although experiments are a recognized part of today's social science research techniques, for many social scientists they are still not well understood. Training in laboratory experimentation is still not part of the graduate training of the majority of social scientists (psychology may be the exception). That is unfortunate for many reasons. Those who might wish to conduct experimental research may not feel confident enough in their skills to approach this method. Social scientists who use other methods, such as survey researchers in sociology, may misunderstand the goals and uses of experiments. Because every science relies on peer review of research, misunderstandings can slow the accumulation of knowledge; good experiments may be criticized on inappropriate grounds, and real flaws in an experiment may be overlooked.

With the continuing growth and development of experimental methods in social science, it will not be long before understanding experiments is an important part of every social scientist's professional skills. We and the other authors in this book hope to contribute to that understanding. For new experimenters, we offer suggestions to improve the quality of their work; for those who read and wish to assess experimental research, we describe techniques and offer guidelines. In these ways, we hope to contribute to the growing quality of experimental research in social science. A poorly designed experiment will either produce no results or, worse, will produce results that are not what an experimenter thinks they are. This book brings together "best practices" by several of today's outstanding experimental researchers. The chapters can be read as "how to" manuals for developing one's own experiments or as sources of criteria to judge and improve the quality of experimental research by practitioners and by the professional audience. All of the chapters contain background to their individual topics that explicitly address common and some uncommon points crucial for understanding this method.

Experimental research is one kind of intellectual activity. A good way to approach experiments, either those one plans to conduct or those conducted by others, is to ask what they contribute to knowledge. What do we know as a result of an experiment or what do we hope to learn from a contemplated experiment? As will be clear in several of the chapters, the central issue in experimental research, as well as in other kinds of research, is how the research can contribute to knowledge of social processes and social structures.

We begin with some terminology on research design. All research is about how things are related. In describing a research design, often it is convenient to distinguish *independent* and *dependent variables*, and we use those terms in describing what we mean by *experiments*. A *variable* is anything that takes on different values, something that can be measured. In research, variables are tied to measurement operations; for instance, the variable *socioeconomic status* or SES may be measured by a person's or a family's income in dollars. Beginning students must become accustomed to the idea that so-called *independent variables* are usually controlled in some way by investigators, whereas *dependent variables* are left free to vary; they are controlled only by nature. Thus, a study might control educational level statistically by partitioning a sample of individuals into those whose education ended after eighth grade, after some high school, with a high school diploma, etc., to determine education's effect on SES. Education is the independent variable in this study, and SES is the dependent variable. The design described uses survey methods.

As we use the term, a study is an *experiment* only when a particular ordering occurs: when an investigator controls the level of an independent variable(s) *before* measuring the level of a dependent variable(s). In the preceding hypothetical survey design, presumably data were collected all at the same time on the respondent's schooling and income. The independent variable was partitioned afterwards, and the interest was in how SES divided after education was so divided. If the design had been an experiment, the investigator might have begun with a large group of children and placed them into different groups determining how many years of education they would receive. When members of the last group had completed their education, average income levels of the different groups might be compared.

That hypothetical example illustrates two more points about research designs. First, not every research design is an experiment. Informal usage sometimes describes any study as an experiment, but if we are going to focus on creating good research designs, we need to be clear what type of designs we work on. Although many of the criteria of good scientific research apply to all kinds of research, there are specific criteria for good surveys that are quite different from the criteria of good experiments. For instance, sample representativeness is a major concern in many surveys and less so (for reasons discussed in several chapters) for experiments. Random allocation of respondents to conditions is crucial for most experimental designs, but it is often impossible in surveys. Second, for many reasons, not every research question can be studied experimentally. Moral and practical considerations make the preceding hypothetical experimental design—deciding ahead of time how much education every person can attain—ridiculous. Knowing what we can study experimentally is important at the very earliest stages of research design, when an investigator selects a type of design to develop.

Experimental studies came into social sciences in approximately 1900, beginning in psychology with studies of biological responses (particularly saliva production in dogs) conducted by the Russian physiologist Ivan Petrovich Pavlov. Pavlov's training in the biological laboratory was demonstrated in his turning to an experimental design, as well as in his use of physiological terms such as "stimuli" to name independent variables in the work. American psychologists such as Edward L. Thorndike (1905) at Harvard University and

J. B. Watson (1913) at Johns Hopkins University adopted and developed the method for many studies of individual differences and interpersonal influences, and the methods spread across the social sciences.

Social psychologists Solomon Asch (1951), Muzafer Sherif (1948), and Leon Festinger and J. Merrill Carlsmith (1959) developed experimental methods beginning in the 1940s, approximately the same time the economist E. H. Chamberlin (1948) began to study markets experimentally. Mathematically trained social scientists, including Siegel and Fouraker (1960) and Von Neumann and Morgenstern (1944), analyzed rational choices, negotiations, and games, providing the foundations for many contemporary theories in sociology, political science, communications, and economics. In the 1950s, Robert Freed Bales and his colleagues and students at Harvard University began studying discussion groups using techniques and technology of an experimental laboratory. Kurt Lewin and Dorwin Cartwright founded the Research Center for Group Dynamics at Massachusetts Institute of Technology in 1945, and Cartwright moved it to the University of Michigan upon Lewin's death in 1947 (Cartwright & Zander, 1953). The experiment has been a significant part of all the social sciences for more than half a century, and its particular advantages continue to attract researchers across the social sciences.

The history of experimental methods shows diffusion across disciplines. Starting with techniques Pavlov learned in biology, through widespread diffusion of Bales' studies of discussion groups (e.g., into contemporary "focus groups" in communications and business), imaginative researchers have built on what worked in other projects and other disciplines. This is all to the good. When investigators in one discipline come up with useful designs, other disciplines benefit from using those when appropriate. Designs and technology do not have to be reinvented constantly, and investigators can focus their attention more appropriately on research questions. Experimental methodologies are not tied to any particular discipline; it makes no sense to argue, for instance, that economic experiments are fundamentally different from experiments in psychology or sociology.

There are different theoretical concerns in different disciplines, and different disciplines have developed different typical designs and even display different "tastes" in design. For instance, some experiments in social psychology use deception to create independent variables, something most economists do not do and often do not approve of. Importantly, however, there definitely are criteria for well-designed experiments and reliable experimental results, and those do not differ across disciplines. Quality experimentation is furthered by ecumenism.

II COMPARISONS TO OTHER DESIGNS USED IN SOCIAL SCIENCE RESEARCH

Social scientists have a wide range of data collection methods. They include the following:

- Unstructured observation. This method is typically used when a social scientist is present at some unexpected event, such as an accident, natural disaster, or violent attack. The scientist can record relevant data on presumed causal factors and outcomes, often also identifying intervening factors and their effects. Other observers, such as from news media or ordinary citizens, can also record unstructured observations, although in many cases the social scientist's training helps him or her to choose what to observe and record as being theoretically important.

- Structured observation. This method is used when a social scientist begins with a coding scheme of what to observe, perhaps also how to record it. In contrast to unstructured observation, here the observation site is predictable and well enough known so that observations can be chosen ahead of time. For instance, a sociologist might go to a courtroom to code jurors' facial expressions and other behaviors to determine how those may be related to verdicts. The observations here are limited and focused, and it is possible to assess reliability if two observers record the same data from the situation.
- Historical archival research. This method relies on documents as data for answering research questions. For instance, a researcher may compare recorded lynchings to the price of cotton in the old Confederacy states. Documents may be written, video, or audio. As with observational studies, archival documentary research does not exert control over factors, except through statistical techniques.
- Participant observation. This may be structured or unstructured; its defining characteristic is that the observer has privileged access to the setting by virtue of being a member of the group studied. The natural settings are extremely important for such studies; their explication is as high a priority as the interaction itself. So, for example, the rules of the sheltered workshop, the structure of the asylum, or the city in which a gang operates are important actors in their own right. In many ways, observation research of all three types is at the other end of a methodological continuum from experiments: the setting is natural, and the outgoing processes in actual settings are valued. Control and randomization, so highly prized in experimental research, are not usually possible in such studies.[1]
- Survey research. Surveys are usually defined as observations on or about individuals or groups. The U.S. Census is a famous survey used by literally millions of researchers. It asks respondents about many behaviors or characteristics, and then analysis proceeds by controlling some factors to determine how that affects outcome variables. Survey research is generally interested in generalizing from a sample to a particular empirical population. Interestingly, an incorrect presumption that generalization proceeds from experiments and from survey samples in the same way can lead to survey researchers criticizing experiments for being "artificial." We consider

1. There are exceptions. For example, Milton Rokeach's (1981) *The Three Christs of Ypsilanti* involved a kind of control because all three men claiming to be Christ were in the same setting and talking to the same psychologist (Rokeach).

artificiality in greater detail later, as well as the differences between sample generalization in surveys and theoretical generalization from experiments.

- Combining different methods is a defining characteristic of case studies. Case studies often involve documentary historical data and may also include surveys and participant observation.[2]

III ADVANTAGES AND DISADVANTAGES OF EXPERIMENTS

The benefits of any research method cannot be assessed independently of the questions the method is designed to answer. A beautiful research design cannot compensate for a flawed research question. This is especially true for experiments because they are designed to determine how specific kinds of independent variables and antecedent conditions affect dependent variables (or consequents). If a research hypothesis, for example, is derived from faulty assumptions or premises, even the most elegant research—experimental or otherwise—cannot save the study. What a good design can do, however, is help the investigator identify the part of the theory that is faulty. Disconfirmed predictions, when they have been derived from an explicit theory, are valuable because they can show which assumptions or conditions in the theory need improvement.

The greatest benefits of experiments reside in the fact that they are artificial. That is, experiments allow observation in a situation that has been designed and created by investigators rather than one that occurs in nature. Artificiality means that a well-designed experiment can incorporate all the theoretically presumed causes of certain phenomena while eliminating or minimizing factors that have not been theoretically identified as causal. In the language of research design, an experiment offers an opportunity to include the independent variables of theoretical interest while excluding irrelevant or confounding variables. An experiment (like a theory) is always simpler than a natural setting; because of that, it offers the possibility to incorporate factors of interest and limit extraneous factors so that the theoretical principle being examined is isolated. Experiments permit direct comparisons: most often the comparison is between conditions in which a factor is present (an experimental condition) or in which the factor is absent (a baseline condition).[3] In this way, the effects of a factor in the experimental condition can be gauged. Such direct testing of a theoretical prediction is much more difficult in other, less artificial settings. Because they

2. Robert K. Yin (1984) argues that case studies can share some of the characteristics of experiments. In particular, he argues that multiple case studies can be chosen such that they represent variation on the independent variable. Then measurements can be taken on the dependent variables. So, for example, if the researcher had a theory concerning how land use compared between government-mandated programs and voluntary programs, she could randomly choose (or perhaps choose based on specific criteria) mandated and voluntary programs, gather data, and examine the dependent variables.

3. Sometimes comparison may be between an empirical condition and a theoretical prediction rather than between two empirical conditions.

are less artificial, they are more complicated and contain a myriad of factors that could interfere with, magnify, or dilute the effects of the particular factor being investigated.

Another technique that experiments often use to ensure comparison is random assignment. The power of randomization is the power assured by probability theory: if extraneous influences (errors) are distributed randomly, they sum to zero. When individuals are randomly assigned to different experimental conditions, different effects observed in different conditions are not due to uncontrolled factors, such as personal traits of the individuals studied, because those factors have been evenly distributed across conditions. So, for example, in drug studies, there is often a control condition or treatment in which individuals receive only a sugar pill and an experimental treatment in which individuals receive only a treatment drug. When individuals have been randomly assigned to either the control or the treatment condition of the experiment, comparing outcomes in the control or baseline (sugar pill) condition with the treatment (drug) condition permits a gauge of the effect of the drug treatment. Random assignment of individuals to conditions assures a researcher that there is nothing unanticipated about the individuals that might have led to an effect. So, for example, if there is no random sampling, there is the possibility that, by some chance, healthier people or people with better nutrition or taller people ended up in the same condition. With random assignment, we can eliminate the explanation that differences in conditions result from characteristics of the individuals or group treated.

Artificiality also allows experiments to provide settings that are difficult, if not impossible, to find. For instance, some studies of voting, such as the experiment of Wilson (see Chapter 14), allow individuals to record dozens of votes within a single experimental session to assess a theory of how individuals reach equilibrium points where each person's interests are represented as closely as possible. The situation is analogous to repeated voting on bills in Congress or to repeated voting among condominium residents. However, because Wilson studied it experimentally, he was able to assess the process not in months or years but within a couple of hours per group.

Because experiments are artificial and controlled, they also invite and enable clear replication by other investigators and comparison across different settings. Findings from an experiment can be assessed by someone else who replicates the experiment. In contrast, findings that occurred in some natural setting can be impossible to replicate because it can be very difficult or impossible to find another natural setting enough like the first one that the investigator can be confident it truly replicates the important features of the setting. Also, experiments are well suited to cumulative research programs that develop and test theories sequentially because they can be evaluated under consistent conditions. The researcher can be assured that differences do not arise over different settings or different operationalizations but, rather, because of the different theoretical factors being tested. Such consistency can be crucial for theoretical cumulation—results for one study can be used for subsequent studies. For instance, the experimental or manipulated condition in a study at time one can become the baseline condition for study two.

Temporal or theoretical ordering also can be examined in experimental settings, brought by, again, artificiality. In interactions that occur in natural settings, it is often difficult to disentangle cause and effect, antecedent and consequent conditions. Thus, for example, it is difficult to disentangle how experience within a group and the statuses of individuals within a group (often derived from their experience) affect individuals' behavior. In experiments, not only can experience and status be separated but also they can be manipulated prior to the measurement of behavior. In this manner, the antecedent and the consequent can clearly be distinguished. In natural settings, this is much more difficult, even given longitudinal designs, because individuals possess many different characteristics and experiences, those factors may change dramatically over time, groups may be involved in different kinds of tasks, and group composition is constantly changing.

Although enormous benefits accrue from artificiality, experiments receive a great deal of criticism for this defining property (e.g., see Babbie, 1989). The criticisms center around generalizability: because experiments are artificial, they do not mirror any real setting and they are not representative of a particular empirical population. Those criticisms are correct, as far as they go. Furthermore, they might be considered disadvantages of experiments. We argue that is true only insofar as experiments are not appropriate for certain kinds of questions. Basically, if the goal is to study properties of some natural setting itself, an experiment is not particularly appropriate. For studying theories that have abstracted some properties of natural settings, however, experiments can be ideally suited.

Experiments cannot attempt to simulate all the complexities of particular settings. They cannot reproduce, for example, elementary school classrooms in Texas or all features of the work environment of a particular corporation such as Apple. However, experiments can produce abstract features of school classrooms (e.g., giving someone chances to perform and then evaluating performances), and they can produce authority structures such as might also occur in a corporation such as Apple. So, rather than reproduction, experiments are designed to match the characteristics of theories composed of precisely defined, abstract concepts. In Chapter 8, Zelditch develops this argument in detail in his analysis of external validity.

These concepts and, consequently, the theories composed of these concepts are not defined by a particular time or place. Rather, they are abstracted from particular times and places. For instance, social scientists from several disciplines have studied "public goods," a line of research described by Sell and Reese in Chapter 10. Naturally occurring instances of public goods might include parks in a large city, listener-supported radio, or a free shuttle bus. Those instances have certain features in common. Individuals need not pay money in order to enjoy their benefits—anyone may be allowed to use a park or a playground in the city. However, there always are costs, such as taxes foregone by keeping businesses out of the park, and cost of maintenance crews. Public goods dilemmas revolve around keeping enough people motivated to contribute a fair share for the things they can actually use without paying.

The term "public good" is an abstract concept meant to capture key features of concrete things such as buses and parks. The term is independent of time and place: whereas a shuttle bus may be an instance of a public good at the present, the term public good will apply in any culture at any time where other concrete things may meet that definition. A theory using the term public good must include an abstract definition telling someone how to find instances of it. The definitions of abstract terms can be much more precise than definitions of concrete terms. Although someone might argue whether a particular piece of land is or is not a public park, for instance, if the theorist has done his or her job well, there will be no doubt whether something is a public good.[4] In Chapter 3, Thye provides more detail on abstract definitions and the parts of a theory. Our point here is that experiments are ideally suited to developing and testing abstract theories, and this is their most appropriate and most valuable use. If there is no good theory, it is too early to think about doing experimental research. The investigator would be better advised to concentrate on observation of natural settings for a while.

Because the theories are abstract and the experiments are artificial, the question arises: how can the results and general principles be applied to the settings in which we really are interested—natural settings such as those elementary school classrooms in Texas? The answer rests with the chain of activities in scientific research and application. At the first stage, an investigator decides on a particular kind of research issue, problem, or theoretical dilemma. The issue could arise from observations of some phenomenon that, for whatever reason, she wishes to understand better.

However, observation is not a necessary prerequisite. The research issue could be suggested by elaborations of other theories or by purely deductive implications of a formal system. The investigator conceptualizes the phenomenon abstractly to develop propositions about how it functions under different conditions. In other words, she comes to develop a theory. Next, she may create an experiment to test her developing theory in exactly the kind of situation that illuminates the parts she is most concerned about. If the experiment receives experimental confirmation, she comes to believe in her developing theory and is ready to apply it to understand other natural settings. She might even, if she is unusually energetic and skillful, decide to use her theory to devise interventions to produce desirable changes in some natural setting that fits the relevant conditions specified in the theory.

4. Logicians say such definitions are "exact class," and they obey the law of the excluded middle. That means that for an adequate definition, it must be clear whether something does or does not meet that definition; there is no middle ground where it "sort of" meets the definition.

Note that the theory is the bridge between the laboratory experiment and natural settings. The experiment tests the theory, and the theory can be applied to natural settings as well as to the experiment.

Thus, for example, suppose that a theory predicted that norms for cooperation in dyads would strongly increase cooperation of a whole group (composed of many dyads) in a public good context. Further suppose that a scope condition for this theory was that the initial dyadic cooperation occurs in settings in which there are no authority figures endorsing the cooperation. Such a theory could be applied to classroom settings as long as no individual (e.g., a teacher) required dyadic cooperation.

IV STEPS IN CONDUCTING EXPERIMENTAL RESEARCH

Experimental work takes place in four large blocks that we call *foundations*, *abstract design, operations*, and *interpretations*. *Foundations* include theory and hypotheses behind an experiment; they are the intellectual reasons to conduct the experiment. General questions at this stage are "What do I want to learn and how can an experiment answer those questions?" *Abstract design* refers to the "plan" of an experiment, including independent and dependent variables, measures, interaction conditions, and all other things that go into making up an experiment. The main question here is "How can I design a situation to answer the previous research questions?" *Operations* are the way things actually look to someone watching or participating in an experiment. Variables become operational measures; interaction conditions become instructions or words used by participants to communicate with each other. An important question to consider here is "How will this look in actual experience to someone in the experiment? Is that the same as I intended in my design?" *Interpretation* is what experimental results mean. The general question at this stage is "What did we learn from the experiment?" The answer may be in terms of the theory or hypotheses stated at the foundations stage or in terms of some unrelated general issues such as methodological advance.

All four kinds of activity—and all four questions—must occur before an experiment is actually conducted. Although that might be obvious for the first three blocks (*foundations*, *design*, and *operations*), it is equally true for *interpretations*. That is, an experimenter should consider what results of an experiment could mean before beginning the work. Experiments that work out—that is, that support research hypotheses—have fairly straightforward interpretations, but experiments resulting in disconfirmed hypotheses also mean something. We expand this point later. Let us consider in greater detail what each of the four kinds of activities entails.

Foundations of an experiment are the reasons for doing it. An experimenter has some questions to answer or a set of ideas to assess. Ideally, the experimenter begins with a developed theory and rigorously derived hypotheses, although often theoretical and hypothesis development are incomplete. However, by understanding the ideal case, often it is possible for an investigator to assess how damaging departures from that ideal may be—that is, whether the work is ready for experimental investigation or whether more time needs to be spent on the theoretical foundations of the experiment.

Theoretical foundations have different elements that affect experimental design. We consider scope and initial conditions and derived consequences or hypotheses. (Other essential parts are the abstract propositions that constitute the actual theory and the logical or mathematical calculus used to combine propositions to yield derivations or hypotheses. However, scope and initial conditions and hypotheses are crucially linked to experimental design, and we focus on those parts here.)

All theories have limited scope. Scope conditions describe classes of situations to which a particular theory claims to apply. Newton's laws of acceleration of falling bodies (the general propositions) famously apply only in the absence of factors such as air resistance or magnetic deflection; in other words, they take as scope conditions the absence of those factors. Einstein's special theory of relativity treats the speed of light as a constant; it does not use scope limitations of air resistance or deflection. Sociological theories of status processes treated in Chapter12 take as scope conditions situations in which individuals are task-focused and collectively oriented—that is, situations in which individuals view solving problems as their main reason for interacting (rather than, for example, enjoying each other's company) and in which they believe it is important to let everyone contribute to problem solving (as opposed to taking tests without help).

Scope conditions tell the kinds of situations in which a theory claims it can describe what happens. If a situation meets its scope conditions, the theory should be able to predict accurately. If a situation is outside that scope, the theory makes no claim to being able to make predictions. Given that scope conditions are met, confirmation of prediction increases confidence in the theory and disconfirmation indicates something wrong with the theory (or the test, if methodology could be a problem). Thus, confirmation and disconfirmation both have meaning *as long as a situation is within the theory's scope.* Outside the theory's scope, any results are irrelevant for judging the theory. This means that an experimenter must be careful to design and operationalize situations within a theory's scope.

It is unfortunately true that many social science theories have been offered without any explicit scope conditions. They are presented as if they applied to every conceivable kind of social situation. If you pressed a theorist about that, he or she might admit that the theory probably does not apply everywhere, but many theorists appear not to have considered just where their theories do and do not apply. To design an experiment to assess a theory that is not explicitly scope-limited, an experimenter must infer (i.e., guess) what kinds of scope limits the theorist had in mind. In other words, suppose a particular natural situation or experiment produced results different from what the theory would expect. Would the theorist say, "Well of course I never intended the theory to apply to such cases?" If so, that is an implicit scope condition. All theories have scope limitations. Although it is really a theorist's job to tell what those are, if a theorist does not do that, the experimenter must because there is no way to design a relevant experiment until issues of scope have been settled.

Initial conditions are what prompt the theoretical process; they are the instigators of change in a situation. Presuming a situation meets a theory's scope conditions, then initial conditions describe the setup for what will happen. For example, in theories of status processes mentioned previously, one situation of interest is what happens when two individuals occupying different status positions interact. Interactants having different statuses (however their society may define status) is an initial condition in this case. In that theory, if individuals have different status positions, that fact will become salient to their interaction, they will form differential performance expectations for each other, and they will treat each other differently in specific ways, all related to power and prestige inequality. From the theory, we could derive predictions or hypotheses that if a high-status person interacts with a low-status person, the former person will talk more and listen less, and the latter's discussion behavior will be the inverse. The initial condition to be created in an experimental test is status inequality.

Initial conditions often define the independent variables of experimental design. The preceding predictions might be tested in an experiment designed so that two individuals of different status positions will interact in a task-focused collectively oriented situation. To be clear, the theory states that if a status difference exists, then certain behavioral outcomes will occur, given a situation meeting the scope conditions. An experiment to assess that prediction must, therefore, create a situation in which individuals of different statuses encounter each other. The design is governed by initial conditions specified in the theory and by its predictions. Independent variables must reflect initial conditions, and dependent variables that reflect predicted theoretical outcomes must be possible.

Experimental operations are what actually will happen in an experiment. They are the instances an experiment offers of the abstractly defined independent and dependent variables. In the previous example, creating a situation in which two individuals differing in status interact might be accomplished by creating a team consisting of a teenager and a 35-year-old woman. In order to use that operation, we would need to know that age constitutes a status characteristic for people in our society because the theory and the independent variable speak of "status difference," not "age difference." The theory predicts effects of status processes; it is definitely not a theory of age (or of gender or skin color or any other actual characteristic). The theory states that *if* age is a status characteristic for people in a particular society, then it will have certain interaction effects. The operational challenge is to instantiate status difference, which in our society can be done using age difference.

For a theorist, the distinction of the theoretical term "status" and the operational term "age" is important because societies change. Status characteristics are created and atrophy all the time, but the same theoretical processes apply to whatever may be the status distinctions existing at a particular time. The theoretical status process, not the historical fact of what happens to be status, is of primary interest here.

For an experimenter, the distinction of theoretical terms and operational terms is also crucial because a well-designed experiment must instantiate "status difference" in order to test that theory. In other words, many challenges at the stage of operational design involve creating situations that are instances of cases described theoretically. Partly that means understanding the culture in which the experimenter operates—for instance, what constitutes a status characteristic for this population.

However, operational challenges extend far beyond what an experimenter knows of society. He or she must also translate crucial concepts into terms meaningful for the experimental participants. For instance, if a situation must be task-focused, as one of the scope conditions requires, that must be created in any experimental test of that theory. Yet experimental participants do not ordinarily use a term like "task-focused," and it would not communicate much to tell them to try to simulate being task-focused. The experimenter must create a situation that individuals will accept as task-focused. In other words, experimenters must understand the world of participants well enough to instantiate situations they understand in the same ways as the experimenter does, even though the experimenter thinks in abstract, theoretical terms and participants think in concrete actual realities. Going back and forth between the theoretical world and the worlds of participants is one of the most difficult challenges facing experimenters.

Operations include introductions and instructions—whatever participants are told when they appear for an experiment. Of course, those elements should be as standard as possible because in a well-designed experiment, participants in different conditions experience most of the same situation, with the exception of differences specified by independent variables and initial conditions. Reading instructions, for instance, is preferable to talking about a situation informally because the former kind of administration is much more uniform.

The final part of operations is dependent measures—operational translations of dependent variables from the guiding hypotheses. Here again, translation is not straightforward. An experimenter has to think about what can be done technically and given the limits of human observers. Limits include technology, such as what can be seen, and human limits such as fatigue and inattention. In our status example, the hypothesis predicted more talking and less listening from high status people. How can "more talking" be assessed? By recording their interaction and counting words? Maybe sentences would be a better measure, or maybe complete thoughts. How can an experimenter measure amount of listening? Is silence enough, or does the hypothesis entail some sort of paying attention, and if the latter, how can that be measured?

None of the steps in creating operational measures is likely to be simple, and yet the strength of any experimental design depends on doing all of them well. The general advice here is for the experimenter to think through absolutely every step in the operations as thoroughly as possible and then pretest everything to be sure it works as intended.

A crucial decision in experimental operations involves what data will be collected and how they will be used. In an experiment, it is possible to assess behavior of various sorts, to distribute questionnaires, and to interview participants. Selecting measures to use and carefully developing them are significant steps in experimental design. Just the simple case of testing a prediction about who will talk more, for instance, requires making decisions about how much difference matters. If an experimenter counted words, would a 10% difference in average words across high-status and low-status conditions, for instance, be considered confirmation? Could we put the decision in the hands of statisticians so that any statistically significant difference counts as success? In addition to mean differences across conditions, many other statistics are available that have potential value. For instance, variability across participants may give clues to both theoretical and methodological questions.

Presuming an experimenter has reached thoughtful decisions about what sorts of data to collect and what will constitute confirmation and disconfirmation of hypotheses, other questions arise. Usually, there are many hypotheses possible from a theory under test, and in the nature of things, not all of the test results will point in the same direction. That is, it is reasonable to expect some confirmations and some disconfirmations. An experimenter who gives some thought to what different possible patterns might mean before collecting the data is in a better position to collect additional information, such as other measures, when disconfirmations or indeterminate outcomes appear.

Even the purest cases—total confirmation and total disconfirmation—deserve some advance thought. Total confirmation of all hypotheses greatly strengthens confidence in the theoretical foundation, of course. This happy state should lead to the question "What's next?" An experimenter who has even a sketch of an idea what questions to investigate following the present experiment is primed to move quickly once the data are in.

What is not often appreciated is that disconfirmation is also informative. It might, of course, mean an entire theory, or at least the parts of it under test, was wrong. However, that outcome is superficial and it is never likely. Did the experimenter really begin with a theory so weak that he or she is not surprised by getting total disconfirmation? That seems unlikely. Most often, experimenters believe the theories they work with, and disconfirmation does not usually mean a theory should be abandoned. Abandoning a theory without a substitute theory really says "I have no idea at all what's going on here." Although that might be true in rare instances, for most social situations, we have at least vague intuitions about what is going to happen. (The work of an experimenter, as of a theorist, is to make vague intuitions explicit and testable.) Disconfirmation invites further thought. What parts of the theory were used in the disconfirmed prediction? How could those parts be modified to account for the outcomes of the experiment? What would an independent test of that reformulation look like? In a well-designed experiment, disconfirmation is at least as useful information as confirmation.

Much of this section identifies challenges and potential problems with good experimental design. That reflects the world: good experiments are difficult to design and conduct, and only a foolish person thinks otherwise. Fortunately, some of the best experimenters in social science have prepared guidance for all steps of the process in the succeeding chapters of this book.

V THE PLACE OF EXPERIMENTS IN SOCIAL SCIENCE

Experimental methods are one kind of data collection that may be used to assess theoretical knowledge. They are certainly not the only method, for many others are available: surveys, content analyses, structured and unstructured observation, and others. As noted previously, all research searches for relations between concepts. All good research depends on being able reliably to infer that things are related in the ways a theory thought they would be—or in understanding what parts of the predictive apparatus needs revision.

The preceding discussion suggests that experiments are not well suited to study in the absence of any theoretical foundation—for instance, just "to see what will happen." Anytime someone sets up a situation and collects observations, we know that "something will happen." However, the real questions are what we learned from it. Did something happen for a reason we can specify, or was it just a chance occurrence? If the latter, it is not interesting scientifically. If there is a reason, that is the beginning of a theoretical understanding. A scientist will usually want to work out the theoretical understanding before moving to empirical research. The reasons why are more interesting and important than the simple facts of what happened.

In fact, a reader may have realized another feature of experimentation from the discussions in this chapter: experimental results themselves are not really interesting except as they bear on a theory. In other words, who cares if two people in our status example talk different amounts? That becomes an interesting fact if it shows us something about how status operates, for we will encounter thousands of other situations of status inequality and with the theory, we have some way to understand what is likely to occur in those new situations also. However, the simple numbers from experimental data hold little interest without a theory. (For other kinds of research, such as unstructured observations of police interacting with drug dealers, the observations themselves are likely to be interesting; the cases may be described as "socially important.") For experimental research, the theory is the interesting part; the experimental results are interesting insofar as they tell us something about the theory.

Currently, some phenomena do not lend themselves well to experimental research. Understanding how different patterns of parent–child interaction affect children's success in school, for instance, may require several years of observation in natural settings. However, sometimes the apparent inapplicability of experimental methods becomes less serious upon closer examination. Many years ago, Morris Zelditch, Jr. (one of the authors of this volume) wrote an article titled, "Can you really study an army in the laboratory?" (Zelditch, 1969) and answered in the affirmative. Although you cannot take an army into a laboratory, you certainly can study important theoretical features of armies, such as authority structures, power exercise, and organizational efficiency.

As the social science become more theoretical, experimental methods are likely to become increasingly important. Experiments generally offer the most convincing evidence for success or weakness of theoretical explanations. No other kind of research method produces data so directly relevant to a theory or suggests causality so conclusively as do experiments.

VI HOW THIS BOOK CAN HELP

Most students first learn about experimental methods from coursework in colleges, but there is a long way to go from knowing experiments exist to being able to actually do one and get some useful results. It is unreasonable to expect a student who has studied chemistry to be able to go into a laboratory and create a drug to block the replication of HIV or build a safe cleaning solution for home use. Nobody would just give a biology student a preserved animal and ask her to figure out why it died. It is the same in the social sciences. There are some theoretical ideas about how social structures and social processes work, but those theories do not tell enough for someone to be able to study those things experimentally.

Laboratory experimentation is taught in all fields through apprenticeship programs. Students, from new undergraduates to postdoctoral trainees, work in laboratories with established scholars. At the early level, students learn some very basic facts about experimental methods, such as Mills' canons of difference and similarity, the importance of developing theoretical concepts and tests, and how to use and interpret statistical tests. At the most advanced level, a student learns advanced techniques, how to use and interpret very sensitive instrumentation, an extensive body of experiential knowledge about likely outcomes and what they mean, and other esoteric knowledge usually associated with a narrow specialization.

The odd fact is that for the vast area between the extremes of expertise, there is little in the way of reliable information. This means that someone who has never designed an experiment is unlikely to find a good source of the information needed to get useful outcomes from it. Even an experimenter in one specialty is unlikely to know how to devise and conduct useful experiments in another area. Because social science experimentation is newer and less widely known than natural science experimentation, many of us do not have a good understanding of why someone would want to do experiments, what they are good for, what problems can arise and how to deal with them, and what some exemplars of good experimentation look like in the different social sciences. This is the gap we address in this book.

REFERENCES

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

Babbie, E. (1989). *The practice of social research*. Belmont, CA: Wadsworth.

Cartwright, D., & Zander, A. F. (1953). *Group dynamics: Research and theory*. New York: Harper & Row.

Chamberlin, E. H. (1948). An experimental imperfect market. *Journal of Political Economy*, *56*, 95–108.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Rokeach, M. (1981). *The three Christs of Ypsilanti: A psychological study*. New York: Columbia University Press.

Sherif, M. (1948). *An outline of social psychology*. New York: Harper & Brothers.

Siegel, S., & Fouraker, L. E. (1960). *Bargaining and group decision making*. New York: McGraw-Hill.

Thorndike, E. L. (1905). *The elements of psychology*. New York: Seiler.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*, 158–177.

Yin, R. K. (1984). *Case study research: Design and methods*. Newbury Park, CA: Sage.

Zelditch, M., Jr. (1969). Can you really study an army in the laboratory? In A. Etzioni (Ed.), *Complex organizations* (pp. 528–539) (2nd ed.). New York: Holt, Rinehart, & Winston.