1 What is Social Science Research and Why Would We Want to Evaluate It?

What is Social Science Research?

If we knew what is was we were doing, it would not be called research, would it?
—Albert Einstein (1879-1955)

What is social science research? This is not an easy question and there are no simple, authoritative answers since social science research has been defined in many, sometimes conflicting, ways. But it is an important one, because whatever one says about social science research must be premised on this delineation. Social science research—although social inquiry would be a more appropriate term—as used throughout this book, refers to *any scientific study of human action and interaction focusing on elements of thought and behavior that are in some sense social*. As such, social scientists aspire to science. They intend to study human action and interaction and thought and behavior in a systematic, rigorous, evidence-based, generalizing, replicable, and cumulative fashion. Such research is, or can be, of great importance to human affairs. Even though some social scientists would dispute this definition, it is, nonetheless, neither too precise nor too general and therefore sufficient to define the practice of social science research in a lucid, cogent way.

That being said, social science research is, more specifically, a truth-seeking activity aimed at contributing to existing knowledge, generating new knowledge, or for application to some specific problem related to human action and interaction. Truth seeking is the search or investigation of or for a body of real things, events, or facts. In the social sciences truth seeking is normally the process of applying a scientific method to social inquiry. A scientific method is the set of principles and procedures used by social scientists for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. In many instances, this process involves formulating or testing a specific theory or hypothesis, in the broadest sense, where theory is defined as "a set of interrelated constructs, definitions, and propositions that present a systematic view of phenomena by specifying relations among variables, with the purpose of explaining and predicting phenomena" (Kerlinger, 1986, p. 9).

Types of Questions Investigated by Social Scientists

Often, social scientists contribute to existing knowledge and generate new knowledge by systematically investigating one or more of three types of questions: (1) descriptive questions; (2) relational questions; and (3) causal questions (Kline, 2008):

1. *Descriptive*: Descriptive questions are the most rudimentary types of questions that social scientists seek to answer. Such questions involve the simple account of a set of observations on a set of variables of interest. Few questions of interest to social scientists are exclusively descriptive.
2. *Relational*: Relational questions are among the most common types of questions of concern to social scientists. They involve the most basic assumptions investigated by social scientists, such as whether a relationship between two or more phenomena exists at all. More often than not such questions are not only concerned with whether a relationship exists between two or more variables, but more specifically the direction and magnitude of covariation.
3. *Causal*: Causal questions are concerned with whether or how one or more independent variables effect one or more dependent variables. Causal questions can be relatively simple (i.e., causal description) or more sophisticated (i.e., causal explanation). In general, descriptive causal

> questions are those in which social scientists inquire as to whether consequences attributable to varying an independent variable can be established, whereas questions about causal explanation are those in which social scientists seek to identify the mechanisms through which and the conditions under which causal relationships hold.

Put simply, to describe involves representing or giving an account of. A social psychologist might simply be interested in describing how juveniles who commit violent crimes are handled in a state's justice system. Are they incarcerated? Are they sentenced to public service? Or, are they handled in some other way? Any one of these would constitute a description. Although descriptive research sometimes is dismissed as overly simplistic, such inquiry is fundamental to the scientific endeavor and sometimes can inform public policy decisions, has added immeasurably to basic knowledge claims, and often forms the basis for investigating relational and causal questions. For example, questions such as who is sentenced, to what degree, and how are they sentenced, might arise.

The same social psychologist in the example above might then try to determine whether incarceration or public service is related to future behaviors of violent juvenile offenders, such as whether there is a relationship between incarceration and the likelihood of committing violent crimes after release, and if such a relationship exists, the direction and magnitude of that relationship. That is, does a relationship between sentencing and future violent acts exist? Do the two vary together? Here, the social psychologist simply is interested only in establishing the existence of a relationship rather than inferring that the observed relationship is causal, which requires meeting additional assumptions.

Social scientists also attempt to explain some aspect of human action and interaction, and the social world, through their research. To explain is to give the reason for or cause of. In this case, the same social psychologist might seek to explain why some juveniles are more likely to commit violent crimes than others. Such explanations can be very general (causal description) or very specific (causal explanation). One general hypothesis might be that juveniles commit violent crimes because their parents hit, slapped, or spanked them as children. Or, more specifically, the social psychologist might hypothesize that, through a complex process of social learning, juveniles whose parents (or other caretakers) modeled violent behaviors internalize the observed behaviors then themselves perform the same behaviors, and through a contingency process of reinforcement commit increasingly violent acts in adolescence. Continuing the example then, a simple causal description (examining the whole rather than its component parts) might be that juvenile offenders who are sentenced to public service are less likely than those sentenced to incarceration to commit future acts of violence. In this case, the social psychologist is interested only in whether sentencing to public service can be causally associated with decreased future acts of violent crime as contrasted with sentencing to incarceration. That is, does one condition cause another? The social psychologist may also inquire into more specific explanations for the causal effect. For instance, does public service cause changes in empathy toward others which in turn reduces the likelihood of committing violent crime? Here the question is whether the relationship between public service and reduced incidence of violent crime is explained by empathy. Social scientists studying such questions seek to determine how rather than when effects will occur by accounting for the relationship between two variables by one or more additional variables.

The general concepts of causal description and causal explanation (though they also apply to relational types of questions and provide information that simply describes relationships) are conceptually illustrated in Figure 1.1 using the random variables X, Y, and Z. In 1.1 A, the effect of variable X on variable Z is direct ($X \rightarrow Z$), whereas in 1.1 B, the effect of X on Z is mediated through Y ($X \rightarrow Y \rightarrow Z$). In 1.1 C, the effect of X on Z is both direct ($X \rightarrow Y$ and $X \rightarrow Z$) and indirect ($X \rightarrow Y \rightarrow Z$). To illustrate, in case 1.1 A, Z is assumed to covary with X. That is, to what degree (and in what direction) does incarceration or public service (X) predict, or account for, the likelihood of committing violent crimes after release (Z)? In the second case, 1.1 B, does sentencing to public service versus incarceration (X) cause changes in empathy (Y) toward others, which in turn reduces the likelihood of committing violent crime (Z)? In the third case, 1.1 C, the question becomes does sentencing to public service versus incarceration (X) directly cause changes in empathy (*Y*) toward others, which in turn (indirectly through *Y*) reduces the likelihood of committing violent crime (*Z*) or does *X* directly cause *Z* in the (presumed) absence or presence of *Y*?

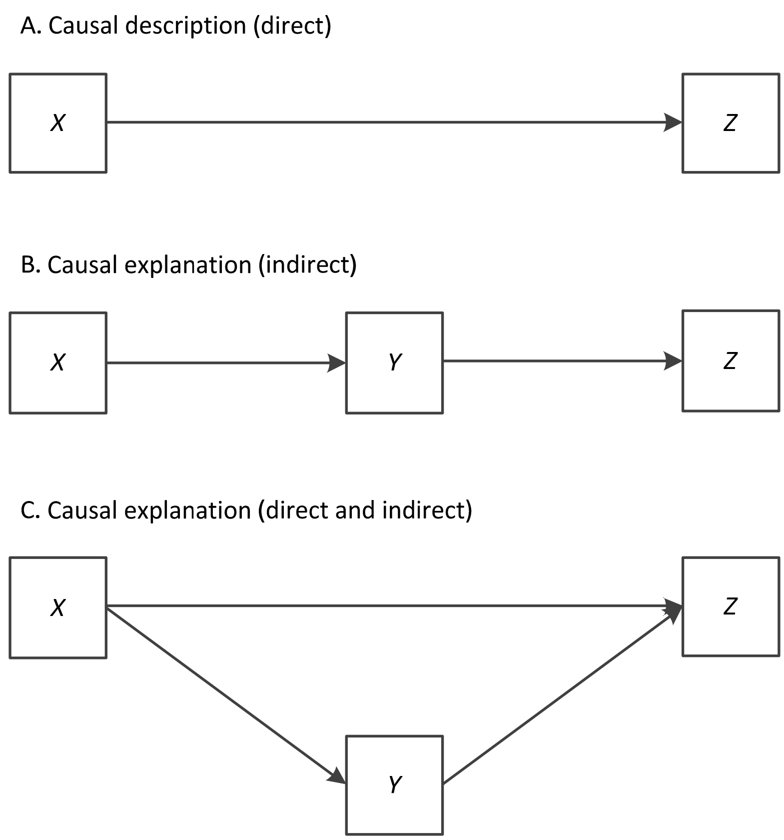

Figure 1.1 Conceptual Illustration of Causal Description and Causal Explanation

As shown in Figure 1.2, such relationships also can be described in terms of whether they moderate (1.2 A) and/or mediate (1.2 B) one another (Baron & Kenny, 1986); and, both can be combined. In 1.2 A, the effect of *X* on *Z* is moderated by A. That is, *Z* (ignoring *Y* in 1.2 B) varies as a function of the A × B interaction. Here then, if A were gender (i.e., male or female), the effect (*Z*) differs over different levels of A (gender). That is, the likelihood of recidivism (i.e., reincarceration) is different over different levels of gender (i.e., *Z* varies over A). Hypothetically then, the effect of *Z* could be greater than (or less than) for females (one level of A) than for males (the other level of A). In the mediator model (1.2 B), if empathy (*Y*) explains recidivism (*Z*), then the indirect effect from *X* to *Y* (path A) through *Y* to *Z* (path B), combined, should be greater than 0, whereas path C (*X* →*Z*) should statistically equal zero if the effect was truly through *Y*. Both are causal explanations given that they explain causal relationships between one or more independent variables and one or more dependent variables.

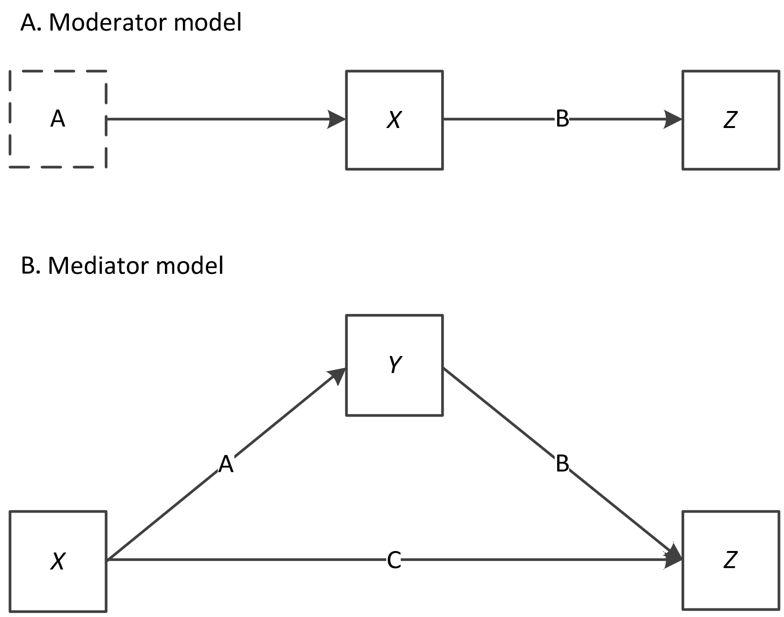

Figure 1.2 Conceptual Illustration of Moderating and Mediating Relationships

Even so, most causal reasoning in the social sciences is probabilistic (i.e., frequentist; cf., Kline, 2004, 2008; Thompson, 2006) rather than deterministic (i.e., general laws or principles—such as the law of gravitational fields used in many of the physical sciences that describes the relationship between force and mass and that has been used in classical mechanics as well as quantum physics), excluding Bayesian principles of cause and effect. That is, most social scientists assume that the probability of a particular outcome, for example, being female decreases the probability of recidivism following incarceration, rather than assumptions or absolute statements or claims that being female always and invariably results in recidivism following sentencing to public service rather than imprisonment, is a statistical, or other type of, regularity (e.g., is *likely* to occur) rather than an absolute (i.e., *always* occurs; e.g., an apple always falls to the ground once it drops from a tree due to gravitational forces).

Knowledge Accumulation

Social science research, in principle, is intended to be cumulative. It is intended to be cumulative in that a body of research on a particular problem, question, or phenomenon should produce facts or knowledge that is in some sense generalizable or replicable. Here then, the assumption is that the social sciences are capable of responding to socially important questions that can be answered in reasonably definitive ways. Also assumed is that empirical and theoretical structures build on one another in a way that permits results of current studies to extend results from prior studies. Generalizability and replicability, therefore, are two fundamental characteristics or properties of the scientific enterprise and more importantly, cumulative knowledge. In most cases, however, replicability has a more central role in the social sciences than generalizability given that characteristics of generalizations and the most appropriate methods for arriving at them are controversial. For instance, both experimental studies and ethnographies are highly localized, so they are often criticized for lack of generalizability, as are single-subject studies. Because validity is a property of knowledge claims, not methods, Shadish (1995) outlined five principles that social scientists might apply to claims about generalization generated by any method:

1. *Proximal (surface) similarity*: Apparent similarities between study operations and the prototypical characteristics of the target generalization.
2. *Heterogeneity of (ruling out) irrelevancies*: Ruling out those things that are irrelevant because they do not change a generalization.
3. *Discriminant validity (making discriminations)*: Clarification of those things that limit generalization.
4. *Empirical interpolation and extrapolation*: Interpolations to unsampled values within the range of sampled instances and extrapolation beyond the sampled range.
5. *(Causal) Explanation*: Theories about the pattern of effects, causes, and mediational processes that are essential to the transfer of a causal relationship.

Even so, few social science investigations, or theories, however, produce results that can be generalized beyond the particular instance, case, or sample studied. Therefore, great care must be taken when considering whether or not the results of one or more studies on a particular phenomenon can or should be generalized to similar or dissimilar groups or individuals, conditions, settings, locations, or periods of time. Unfortunately, such errors or fallacies are not only common in our everyday lives but also in scientific practice. Numerous personality studies, for example, have claimed that women, in general, are more agreeable than men, who are prone to be more disagreeable. Conclusions such as "this particular woman must too be agreeable, because she is a woman," therefore, is an example of such a fallacy. Only rarely, very rarely, can group traits or characteristics, or conversely, individual traits or characteristics, be used to make conclusions about one thing based on information about another. In reference to individuals, such fallacies are sometimes referred to as a fundamental attribution error. Here, the tendency to perceive dispositional or trait-based explanations for observed behaviors or attributes of others is predominantly driven by the cognitive presumption that the traits, characteristics, or actions of others are indicative of the kind of person they are, rather than the kind of situations that compel their behavior.

Replicability, on the other hand, is a hallmark of mainstream science, distinguishing it from pseudoscience, and a powerful criterion for testing scientific theories. Unfortunately, there is a much stronger tradition of replication in the natural sciences than in the social sciences. Replicability is premised on the principle that results should be reproducible by other researchers. Without it, facts and knowledge are ambiguous at best. Nonetheless, results of different studies on the same problem, question, or phenomenon are often conflicting. Are workers more productive when they are satisfied with their jobs? Studies do not agree. Do students learn more when class sizes are smaller? Studies do not agree. Does participative decision making in management increase productivity? Does job enlargement increase job satisfaction and output? Does psychotherapy really help people? The studies do not agree. Meta-analyses of studies ranging from racial integration in schools, crime, juvenile delinquency, employment discrimination, and drug use, among others, have found that almost all conflicting results are largely due to sampling error (Hunter & Schmidt, 1996). What is more, social scientists often lose interest in a particular problem, question, or phenomenon before any true cumulative knowledge can be acquired.

Specific Aims and Purposes of Social Science Research

Social science research can also be described, in a very general sense, in terms of its specific aims or purposes. Three types of social science research are the foci of this book, although there are many others. The first, basic research, sometimes called pure or fundamental research, is research undertaken with the intent of advancing knowledge or theoretical understanding. A sociologist testing a new theory or hypothesis of social isolation in low-income urban neighborhoods is conducting basic research. Such research usually does not have immediately recognizable practical applications and is often guided by a researcher's curiosity, although it sometimes provides the foundation for applied research. The second type, applied research, is research undertaken with the intent of applying research results to some specific problem. An educational psychologist who develops a mastery test to screen elementary students' English language proficiency for admission to an English and Spanish language immersion school is conducting applied research. This research serves a specific purpose. That is, should the school admit a particular student based on his or her score on the assessment or not? The third type, evaluation research, is research undertaken to determine the merit (synonymous with quality), worth (synonymous with value), or significance (synonymous with importance) of something, such as a program, policy, or consumer product (Stufflebeam & Coryn, 2011). An evaluator investigating whether Head Start is effective in promoting school readiness or an economist determining whether the desired effects of Head Start are worth what they cost to produce relative to alternatives is conducting evaluation research. Consumer Reports® studies of end-user products such as automobiles, to determine their comparative quality and value according to ratings on reliability, safety, miles per gallon, costs, and other criteria, as well as performance testing, is also an example of evaluation research. Whereas most social scientists endeavor to investigate facts and knowledge in an objective way, evaluation researchers explicitly seek to render judgments as to something's quality or value, which requires integrating both objective facts and systematically validated values (Scriven, 1991).

Characteristics Common to Most Social Science Research

Social science research can be simultaneously descriptive, relational, or explanatory, as well as basic, applied, or evaluative. In fact, nearly all social science research involves evaluation (e.g., of data, of hypotheses, of theories, of claims and conclusions), which is a major premise of this book. Large parts of social science research also can be characterized by its design, which consist of the structural elements of the plan a researcher will follow when conducting a study and, by extension, from which facts, evidence, and inferences and conclusions are drawn. Structural elements of design include, but are not limited to, samples, conditions, methods of assignment to conditions, observations, and the timing or scheduling of measurement or when a treatment begins or ends, among many others. The simplest means for classifying most, but not all, social science research designs are as experimental, quasi-experimental, or nonexperimental. These categories, however, are not mutually exclusive because research can be designed with elements from each or such designs can be integrated within others.

Experimental designs are those in which the researcher has control over some (or all) of the conditions of the study and control over some aspects of an independent variable (i.e., the suspected cause) being studied. Experimental designs, sometimes referred to as randomized experiments or randomized controlled trials, are those in which independent variables are manipulated rather than only observed, and it is their effect on the dependent variable (i.e., the effect), or variables, that is of primary interest. Random assignment of subjects (e.g., people, animals, schools) to comparison, control, and experimental or treatment conditions is a necessary criterion for a randomized experiment. Such designs are prospective, in that groups are assigned to a treatment or intervention or other condition prior to receiving the treatment or intervention.

Quasi-experimental designs, sometimes called field experiments or natural experiments, among many others, are experiments that lack random assignment to conditions, but that otherwise have similar attributes to randomized experiments. These two general types of designs are most often used in cause-probing studies. Research investigating whether a new reading curriculum (treatment) is more effective than the standard curriculum (control) for teaching 4th grade students to read or whether one anti-retroviral drug (treatment) is better than another (comparison) at increasing the 5-year life expectancy for persons with AIDS is cause-probing research. That is, do 4th grade students exposed to the new reading curriculum acquire the ability to read at a greater rate than for students exposed to the standard curriculum? Or, is the new anti-retroviral drug more effective than the existing one at increasing the 5-year life expectancy for persons with AIDS? Such studies seek to determine whether one thing causes or produces a change in another. And, as Lipsey, Rossi, and Freeman (2003) note, such experiments are hardly a recent invention:

> One of the earliest "social experiments" took place in the 1700s when a British naval captain observed the lack of scurvy among sailors serving on the ships of Mediterranean countries where citrus fruit was part of the rations. Thereupon he made half his crew consume limes while the other half continued with their regular diet...the intervention worked and British seamen were compelled to consume citrus fruit regularly, a practice that gave rise to the still-popular label limeys. (pp. 2-3)

In contrast, nonexperimental designs are those in which the researcher neither alters nor controls the research setting. Nonexperimental research designs are also sometimes known as correlational, cross-sectional, passive, non-interactive, naturalistic, and observational. In comparison to experimental and some quasi-experimental research, such research is retrospective in nature given that one or more suspected causal agents have already occurred and, therefore, cannot be manipulated. In nonexperimental studies, independent variables are often called predictor or explanatory variables and dependent variables are often called criterion or outcome variables, if such distinctions are made at all. Although these designs too are sometimes used to investigate cause-effect relationships, it is often difficult to ascertain what is a cause and what is an effect, given that both are sometimes simultaneously observed. Even so, such designs are a necessary and reasonable procedure in many cases given that random assignment is often infeasible in many domains of interest to social scientists and that many causes simply are not manipulable (e.g., race, gender).

Some social scientists mistakenly conclude that experimental and quasi-experimental designs are exclusively quantitative and that non-quantitative (i.e., qualitative) research is always nonexperimental. However, such a distinction is hard to maintain and justify. With few exceptions, virtually every major research design can be used to gather qualitative or quantitative information, or both. In fact, in the last few decades, it has become increasingly common for researchers to integrate qualitative designs or methods (i.e., mixed method research) into what have traditionally been quantitative studies (e.g., evaluation research). In some types of randomized controlled trials, particularly in health and medicine, researchers are increasingly investigating not only treatment effectiveness, but also its implementation, including treatment delivery, treatment receipt, and treatment adherence. Even so, treatment implementation can be studied qualitatively, quantitatively, or both.

Some Distinctions Between Social and Natural Sciences

Social science research is conducted across and within a wide range of disciplines including, but not limited to, anthropology, economics, education, health, political science, psychology, and sociology, and increasingly is becoming a multidisciplinary activity (i.e., drawing on numerous disciplinary traditions and perspectives to investigate the same phenomena). The social science disciplines differ from the arts and humanities in that they tend to emphasize the use of a scientific method in their study of the world. Disciplines within the arts and humanities range from modern languages and English literature to the creative and performing arts. Often, research conducted in the arts and humanities may entail hearing it, viewing it, reading it, or experiencing it in ways that differ considerably from social science research, such as through performance on a stage or in some other public forum. In contrast to the natural sciences, such as mathematics, taxonomy, biology, chemistry, or physics, for example, which often have predetermined methods of inquiry, the social sciences are largely pluralistic in its methods. A physicist studying the movement of particles from regions of higher density to regions of lower density has few methods choices, whereas a social scientist interested in investigating the factors that explain the relationship between out-group threat, prejudice, and exclusion of out-group members from social justice considerations has numerous methodological options. As such, questions of interest to social scientists may be examined from a multitude of perspectives using a wide variety of methods.

Consider now a familiar instance experienced by many social scientists. It has been debated both by those inside and outside of science that social science is different than the natural sciences. The implication is that social phenomena are different than those found in natural science. This is a strong assertion because it points to a logical difference, not just one of different techniques and different methods. If this is agreed to, then a different logic of inquiry may be needed. Therefore, science possibly cannot be characterized as having one common logic of justification, but several, or else it might be concluded that social science is not science. One reason for this view is that it is widely believed that social science research is less replicable than research conducted in the natural sciences. Even so, a meta-analysis of studies in 13 areas in particle physics and studies in 13 areas in psychology have found that there is as much variability in the results in the former as there are in the latter (Hedges, 1987). Therefore, such assertions are sometimes unfounded, but even so, are still widely accepted both inside and outside of scientific circles.

Another distinction that can be, and often is, made between the social and natural sciences is the notion of constructs (Cronbach & Meehl, 1955). In the social sciences, a construct is a label used to describe something that is believed to systematically vary but which cannot be directly observed. Characteristics measured by many tests and other types of measuring devices are sometimes referred to as constructs because they are not usually directly observable, but rather are described as “constructed.” While a characteristic such as empathy does not exist in a physical sense, unlike height or weight, it is a characteristic on which people are believed or hypothesized to vary along some continuous dimension, with some people behaving regularly in ways regarded as more compassionate, and others behaving regularly in ways regarded as more apathetic. Therefore, we "construct," rather than directly observe, these regularities and describe them as empathy for the purpose of both measuring them and communicating about them. A large majority of phenomena of interest to social scientists are constructions (sometimes referred to as latent variables) rather than observed phenomena (sometimes referred to as manifest, observed, or indicator variables). Only rarely do researchers working in the natural sciences disagree about the simple definitions of "constructs," such as what constitutes a planet (i.e., an object that orbits the sun and is large enough to have become round due to the force of its own gravity). Recent examples, such as the former planet Pluto, would, however, contradict conventional beliefs about agreements of constructs in the natural sciences.

What is more, and although the social science disciplines share many common features, such as their study of human action and interaction, they also differ in a number of important ways. These differences are often exhibited and manifest in terms of philosophical and ideological disagreements and, more generally, perspectives on truth and knowledge. That is, what is known and what is knowable. Among most social scientists, these philosophical and ideological views are commonly referred to as paradigms. Although the term was never intended to apply to the social sciences, it has, nonetheless, been widely accepted and acknowledged as a reasonable descriptor of the social scientific endeavor since its introduction by Kuhn in the early 1960s (1962/1996). The contemporary meaning of a paradigm is the set of practices that define a scientific discipline during a particular period of time, or more generally, a knowledge schema. Very often, a scientific paradigm determines what is to be observed and studied, the kind of questions that are supposed to be asked and probed for answers, how research questions or hypotheses are to be structured, and how results of scientific investigations should be interpreted. Therefore, prevailing paradigms often represent a specific way of viewing reality and what can be known rather than the much more general scientific method. In the social sciences, many of the most widely accepted paradigms are considered incommensurable.

A Social Science Perspective on the Nature of Knowledge, Knowing, and Truth

> Even among self-identified social scientists, we apologize for the thinness of our data, the crudeness of our methods, and the inexactitude of our findings. We are despondent about the state of research, and the lack of resolution to longstanding problems. We remark upon how much we do not know.
>
> —Gerring (2001)

In the social sciences competing paradigms and, inevitably, philosophical and ideological disagreements between quantitatively-oriented social scientists and qualitatively-oriented social scientists, form the basis for what are sometimes profound and irreconcilable differences. Ultimately, these disagreements stem from views about the nature of knowledge and what can and cannot be known, both of which are intimately tied to epistemology and ontology. In simple terms, epistemology is the philosophy of knowledge or how we come to know, and is intimately related to ontology and methodology. Ontology involves the philosophy of reality, including questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences, whereas epistemology addresses how we come to know that reality. Methodology identifies the particular practices used to attain knowledge.

Ideological Perspectives and Orientations

Although a complete treatment exceeds the scope of this book, two paradigms in particular—positivism and post-positivism and naturalistic and constructivism and/or constructionism—are of greatest relevance to understanding contemporary thinking and practice in many of the social science disciplines. Other knowledge schemas held by social scientists include postmodernism, functionalism, interactionism, relativism, realism, and critical theory. Nevertheless, most fields of study are multi-paradigm disciplines given that there are several competing ways to understand disciplines and their problems. Very simply, in the positivist and post-positivist paradigms the object of study is independent of the researcher. Knowledge is discovered and verified through direct observations or measurements of phenomena. Facts are established by decomposing or reducing a phenomenon to examine its component parts. According to the positivist epistemology, science is seen as the way to get at truth, to understand the world well enough so that it might be predicted and controlled. The world and the universe are deterministic, they operate by laws of cause and effect that are capable of being discerned if a systematic scientific method is applied. Thus, research is largely a mechanistic process. In this view, deductive reasoning is used to postulate theories that can be empirically tested.

An alternative view, the one held by many naturalist or constructivist/constructionist social scientists, is that knowledge is established through the meanings attached to the phenomena studied. These researchers interact with the objects under study to obtain data. As such, inquiry changes both researcher and object, and knowledge is context and time dependent and only can be truly known by the knower. Therefore, inductive reasoning, where the researcher moves from observations to the development of hypotheses, and ultimately, theory, can be said to be one of the foremost features of the naturalist or constructivist schema.

Shown in Figure 1.3 are the essential features of and distinctions between deductive and inductive reasoning as well as their hypothetical relationship.

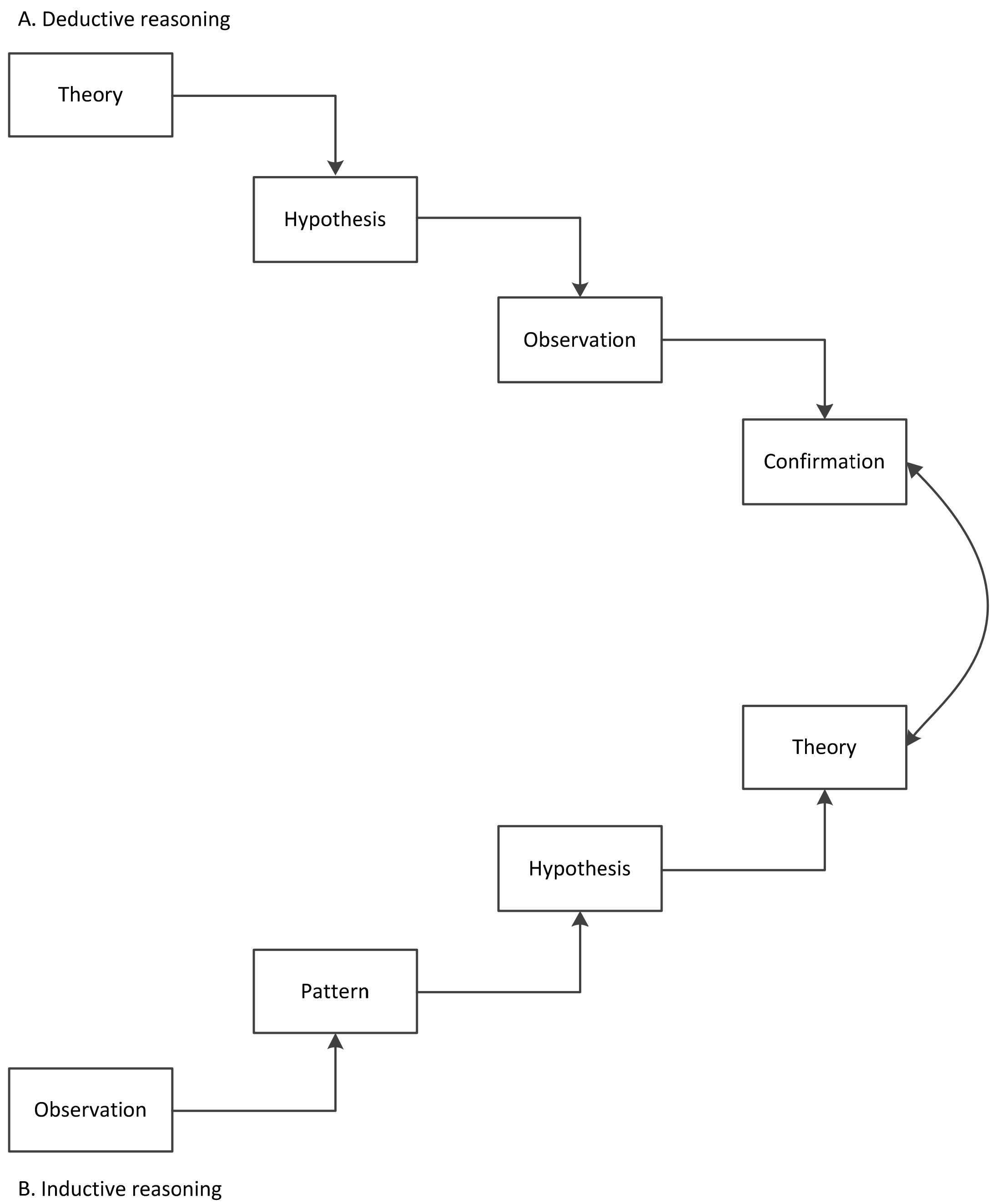

Figure 1.3 Essential Features of Deductive and Inductive Reasoning

Understanding the differences in epistemology among social science research paradigms begins primarily as an ideological endeavor. The question of whether there is one knowable reality or that there are multiple realities of which some individual knowledge can be acquired is largely a function of the researcher's theoretical lens, which plays an important role in the choice of methods because the underlying ontological assumptions of the researcher largely determines the choice of method. At the heart of most scientific paradigms such differences are primarily philosophical, not methodological. Philosophical assumptions about the nature of reality are crucial to understanding the overall perspective from which a study is designed and carried out. A theoretical paradigm is thus the identification of the underlying basis that is used to construct a scientific investigation, or a loose collection of assumptions, concepts, and propositions that orientates thinking and research. Ultimately then, a paradigm can be defined as the basic belief system or world view that guides an investigation.

Scientific Method

> Methodologists are the preachers of science. Armed with canons of correct procedure, they have the power to castigate and exhort. They can instruct us to have clearly defined objectives and explicit frames of reference, to base our studies on good theories…the process of science does not work from rules to practice but from attempt to attempt…one good piece of research influences research practice more than many methodology textbooks.
>
> —Przeworski (1987)

Scientific method, in one form or another, is nearly inseparable from science itself and has been the subject of intense and recurring debate throughout the history and philosophy of science. As stated earlier in this chapter, a scientific method is the set of principles and procedures used by social scientists for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. The essential elements of a scientific method, applying mostly to experimental sciences, are iterations, recursions, and orderings of the following:

1. *Characterizations*: A scientific method depends upon increasingly sophisticated characterizations of subjects of investigation. Subjects can also be called lists of unsolved problems or unknowns. While seeking the pertinent properties of subjects, this may also entail definitions and observations, where observations often demand careful measurements.
2. *Hypotheses*: A hypothesis is a suggested description, explanation, or predication of the subject. Sometimes, but not always, they can also be formulated as existential statements, stating that some particular instance of the phenomenon being studied has some characteristic or causal explanations, which have the general form of universal statements, stating that every instance of the phenomenon has a particular characteristic. In general scientists tend to look for theories that are 'elegant' or 'beautiful.' However, if a model is mathematically complex, it is often difficult to deduce any prediction.
3. *Predictions*: Any useful hypothesis will enable predictions, by reasoning, including deductive reasoning. It might predict the outcome of an experiment in a laboratory setting or the observation of a phenomenon in a natural setting. The prediction can also be statistical and only refer to probabilities. It is essential that the outcome be currently unknown. If the outcome is already known, it is called a consequence and should have already been considered when formulating the hypothesis. If predictions are not accessible by observation or experience, the hypothesis is not yet useful for the method, and must wait for others who might come afterward, and perhaps rekindle its line of reasoning.
4. *Experiments*: Once predictions are made, they can be tested by experiments. If test results contradict predictions, then hypotheses are called into question and explanations may be sought. Sometimes experiments are conducted incorrectly and are at fault. If results confirm predictions, then hypotheses are considered likely to be correct but might still be wrong and are subject to further testing. Depending on the predictions, experiments can have different shapes. It could be a classical experiment in a laboratory setting, a double-blind study, or an archeological excavation. Scientists assume an attitude of openness and accountability on the part of those conducting an experiment. Detailed recordkeeping is essential to aid in

> recording and reporting on experimental results, and providing evidence of the effectiveness and integrity of the procedure. They also assist in reproducing experimental results.

A linearized, pragmatic scheme of the four essential elements of the standard scientific method is sometimes offered as guidance and a guideline for scientists in conducting research in the form of:

1. Define the question
2. Gather information and resources
3. Form hypothesis
4. Perform experiment and collect data
5. Analyze data
6. Interpret data and draw conclusions that serve as a starting point for new hypotheses
7. Publish results

In most social science and natural science disciplines, this schema is widely accepted as standard scientific method. However, many philosophers, historians, and sociologists of science claim that it has little relation to the way that science is actually practiced, nor does it represent alternative notions of scientific method. If experiments by independent investigators replicate results, then a hypothesis may be regarded as a theory. If experiments fail to support a hypothesis, then it must be rejected or modified. Central to this description of scientific method of inquiry is the predictive power of a theory. Theories and hypotheses can never be proved, only disproved, for it is logically impossible to conduct all observations across all time and space to provide all possible replications. Therefore, the next observation may disprove a widely accepted theory. Some anthropologists, for example, have even been known to state that “A theory is only a theory if it is disputed. Otherwise, it is merely a hypothesis.”

In a scientific method one of the primary tasks is identifying and defining a research question. Without questions, science would cease to exist. Therefore, it is of utmost importance that social scientists identify, define, and explicitly state the problem under investigation, including the particular question or questions of interest related to that problem and the specific strategy employed for resolving, partially resolving, or simply studying the problem. What are the questions to be investigated? How will those questions be answered? These, and many others, are the essential concerns that should be identified at the outset of any social inquiry. What is more, the process of identifying and defining research questions, in conjunction with gathering information and resources (e.g., existing knowledge about the phenomenon being investigated), typically serves as a guide to formulating one or more specific research hypotheses.

For the majority of social science researchers formulating hypotheses is the sine qua non (i.e., something that is essential or necessary) of all social inquiry. In essence, a research hypothesis is a deductive guess that states the expected outcome a study. When formulating hypotheses, the researcher deduces, through a literature review process, experience, or observation, an anticipated result. Research hypotheses can be expressed in numerous ways, but typically are formulated first as a null or nil (literally meaning zero difference or zero relationship) hypothesis, then as either an alternative non-directional hypothesis (two-tailed, two-sided) or an alternative directional hypothesis (one-tailed, one-sided). Alternative, non-directional hypotheses imply that a difference is anticipated, but does not express the direction of that difference. Directional hypotheses, however, state the expected direction of an expected difference. Each of these types of hypotheses are presented and defined in Table 1.1. In general, most social scientists are interested in one of the alternative hypotheses, whether directional or non-directional, not the null or nil hypothesis (though the null or nil nearly always serves as the basis for the majority of statistical tests).

Table 1.1

Common Types of Hypotheses in Social Science Research

Type of Hypothesis	Definition
Null or nil hypothesis	States that no difference is expected
Non-directional hypothesis	States that a difference is expected but does not state the direction of the expected difference
Directional hypothesis	States that a difference is expected and the direction of the expected difference

Null or nil hypotheses are implicit in nearly all forms of social science research, whether explicitly stated or not. And, nearly all statistical tests are tests of the null or nil hypotheses rather than tests of the alternative hypothesis. Such hypotheses can be expressed in numerous ways, and the methods for doing so vary by disciplinary traditions, norms, and standards. Some biostatisticians, for example, refer to these types of tests or hypotheses as tests of equivalence (e.g., Is a new drug *equally* effective as an old drug?) or superiority (e.g., Is a 500 mg dose *more* effective than a 250 mg dose of the new drug?). Even so, it often can be difficult to locate either an explicit research question or hypothesis in much of the literature published in the social sciences. If an epidemiologist, for instance, were interested in determining whether the average adult body temperature in the United States is actually 98.6° Fahrenheit, the epidemiologist might express the research question (i.e., What is the average adult body temperature in the United States?) in the form of a null and alternative hypothesis. In notational form, where H_0 is the null hypothesis, H_A is the alternative hypothesis (where A represents alternative) and μ is the population mean, this hypothesis would be represented as:

$$H_0: \mu = 98.6°$$
$$H_A: \mu \neq 98.6°$$

Using the same hypothesis, the epidemiologist might formulate a directional alternative hypothesis rather than a hypothesis simply suggesting that the population mean, μ, does not equal 98.6° Fahrenheit, such as the average adult body temperature in the United States is less than 98.6° Fahrenheit. This directional hypothesis would be expressed as:

$$H_0: \mu = 98.6°$$
$$H_A: \mu < 98.6°$$

Two concepts are important considerations for understanding the practice of null hypothesis significance testing: Type I and Type II errors. A Type I error is the conditional prior probability of rejecting H_0 when it is true, where this probability is typically expressed as alpha (α). Alpha is a prior probability because it is specified before data are collected, and it is a conditional prior probability, *p*, because H_0 is assumed to be true. This conditional prior probability is usually expressed as

$$\alpha = p\ (\text{Reject } H_0 \mid H_0 \text{ true})$$

where | means assuming or given. Both *p* and α are derived from the same sampling distribution and are interpreted as long-run, relative-frequency probabilities. Unlike α, however, *p* is not the conditional prior probability of a Type I error (often referred to as a false-positive) because it is estimated for a particular sample result. Conventional levels of α are either .05 or .01 in most of the social sciences (Cohen, 1994). Alpha sets the risk of a Type I error rate, akin to a false-positive because the evidence is incorrectly taken to support the hypothesis, for a single hypothesis only (sometimes referred to as a primary or focal outcome). When multiple statistical tests are conducted, there is also a familywise probability of Type I error (sometimes referred to as multiplicity), which is the likelihood of making one or more Type I errors across a set of statistical tests. If each test is conducted at the same level of α, then

$$\alpha_{FWE} = 1 - (1 - \alpha)^c$$

where *c* is the number of tests performed, each at a specified α level. In this equation, the term $(1 - \alpha)$ is the probability of not making a Type I error for any individual test, $(1 - \alpha)^c$ is the probability of making no Type I errors across all tests, and the whole expression represents the probability of making at least one Type I error among all tests. So, for example, if 10 statistical tests were performed, each at $\alpha = .05$, the familywise Type I error rate would be

$$\alpha_{FWE} = 1 - (1 - \alpha)^{10} = .40$$

Thus, the Type I error rate across all 10 statistical tests would be 40%. This result indicates the probability of committing one or more Type I errors, but does not indicate how many errors have been committed or which specific statistical test, or tests, the error occurred in.

There are two basic ways to control familywise Type I error. Either reduce the number of tests (or only test the primary or focal outcome) or lower α to a tolerable rate for each test. The former reduces the total number of tests to those with the greatest substantive meaning, whereas the latter can be determined by a number of methods, including the Bonferroni correction. The Bonferroni correction simply requires dividing the target value of α_{FWE} by the number of tests, and setting the corrected level of statistical significance at α_B where

$$\alpha_B = \frac{\alpha_{FWE}}{c}$$

If 10 statistical tests were conducted and the tolerable Type I error rate was 5%, then $\alpha_B = .05/10 = .005$ for each individual test.

Although formal tests of statistical significance largely originated from the works of Fisher (1925) and Neyman and Pearson (1933), statistical power, and the concept of Type II error, however, is largely derived from the work of Cohen (1969, 1980, 1988). Power is the conditional prior probability of making the correct decision to reject H_0 when it is actually false, where

$$\text{Power} = p\,(\text{Reject } H_0 \mid H_0 \text{ false})$$

A Type II error (often referred to as a false-negative) occurs when the sample result leads to the failure to reject H_0 when it is actually false. The probability of a Type II error is usually represented by β, and it is also a conditional prior probability:

$$\beta = p\ (\text{Fail to reject } H_0 \mid H_0 \text{ false})$$

Because power and β are complimentary:

$$\text{Power} + \beta = 1.00$$

Therefore, whatever increases power decreases the probability of a Type II error and vice versa. Several factors affect statistical power, including α levels, sample size, score reliability, design elements (e.g., within-subject designs, covariates), and the magnitude of an effect, among many others (Cohen, 1988; Lipsey & Hurley, 2009). By lowering α, for example, statistical power is lost, thus reducing the likelihood of a Type I error, which simultaneously increases the probability of a Type II error. Conversely, increasing sample size generally increases power. The relationship between Type I and Type II decision errors arising from statistical hypothesis testing is summarized in Table 1.2.

Table 1.2

The Accept-Reject Dichotomy and Decisions for Hypotheses

	H_0 true	H_0 false
Fail to Reject	Correct decision $1 - \alpha$	Type II error β
Fail to Accept	Type I error α	Correct decision $1 - \beta$

Null and nil hypothesis significance testing, in the social sciences and many other disciplines, has been widely misused and misinterpreted (e.g., a *p*-value is the probability that a result is due to sampling error, a *p*-value is the probability that a decision is wrong). The correct interpretation of *p*-values, for $p < .05$, essentially includes only the following (Kline, 2008):

1. The odds are less than 1 in 19 of getting a result from a random sample even more extreme than the observed sample when H_0 is true.
2. Less than 5% of test statistics are further away from the mean of the sampling distribution under H_0 than the one for the observed result.
3. Assuming H_0 is true and the study is repeated many times, less than 5% of these results will be even more inconsistent with H_0 than the observed result.

Despite clear evidence of a multitude of both conceptual and statistical flaws (e.g., *p*-values for test statistics are estimated in sampling distributions that assume random sampling from known populations, most statistical tests are premised on particular distributional assumptions [e.g., the sphericity requirement of the dependent-samples *F* test], and so forth; neither of which are met in most social science research) associated with this practice, it is unlikely that they will be abandoned anytime in the foreseeable future (Kline 2004, 2008; Thompson, 2006; Wilkinson and the Taskforce on Statistical Inference, 1999). In recognition of better statistical practice, however, many social scientists and social science journals now report effect sizes and confidence intervals in addition to *p*-values (Kline, 2004; Thompson, 2006).

Inference, Evidence, Warrants, and Backings

> Evolution has given us quite powerful intellectual tools for processing vast amounts of information with accuracy and dispatch, and our questionable beliefs derive primarily from the misapplication or overutilization of generally valid and effective strategies for knowing.
>
> —Gilovich (1993)

Inference is the act or process of deriving a logical conclusion from premises, based on evidence, warrants, and backings. Inferences can be either valid or invalid, but not both. Independently and combined six features are common to most forms of inquiry, each of which work together to support and justify conclusions resulting from a process of inquiry:

1. *Claims*: That which is taken as acceptable and legitimate.
2. *Evidence*: The facts forming the basis for the claim.
3. *Warrants*: That which legitimate inferences drawn from evidence by appeal to some authority.
4. *Backings*: That which supports the warrant by appeal to some more general authority.
5. *Conditions of exception*: That which points to conditions or circumstances when a warrant will not hold.
6. *Qualifiers*: That which identifies the strength of a claim.

Each of these features and their interdependencies, central to all forms of systematic inquiry, and the process of inference beginning with evidence and ending with knowledge claims, are illustrated in Figure 1.4.

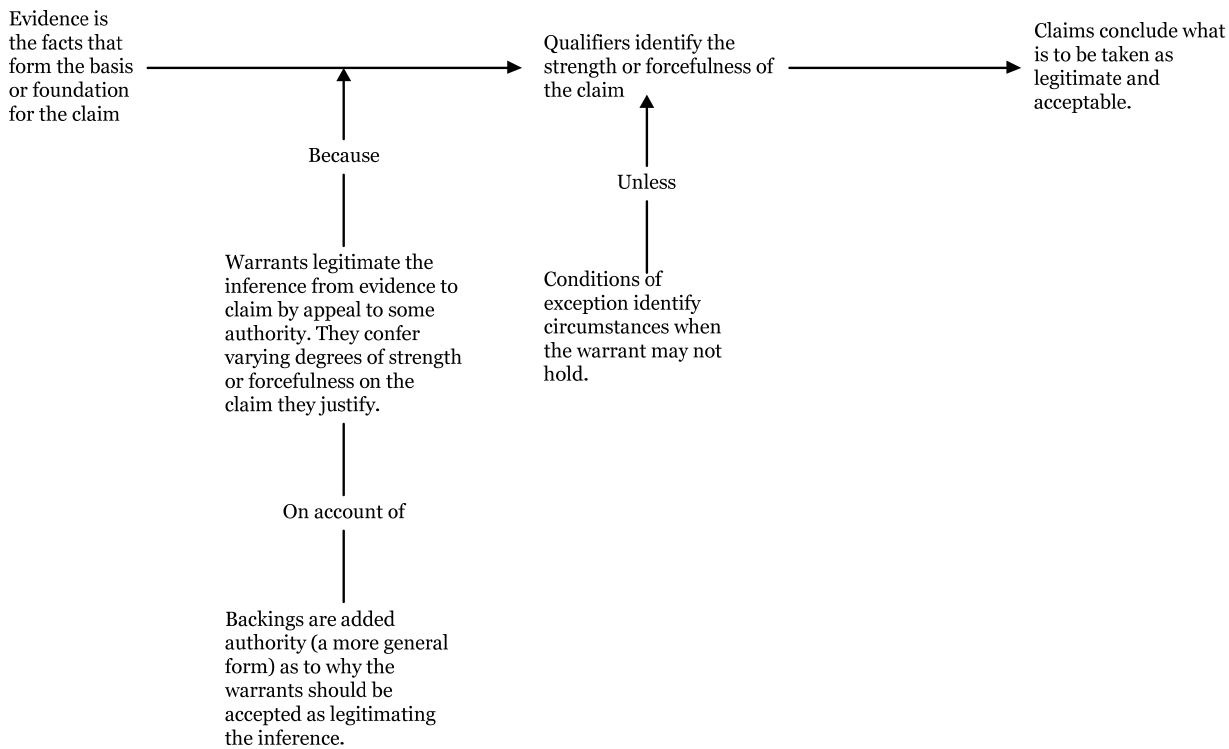

Figure 1.4 Six Main Logical Features Common to all Types of Inquiry

Source: Fournier, D. M. (2008). Establishing evaluative conclusions: A distinction between general logic and working logic. In D. M. Fournier (Ed.), *Reasoning in evaluation: Inferential links and leaps* (pp. 15-32). *New Directions in Evaluation, 68*. San Francisco, CA: Jossey-Bass, p. 24.

This logic of inquiry represents a general pattern of reasoning or basic logic that guides and informs the practice of systematic inquiry. This basic logic provides researchers with rules for constructing and testing claims, and it specifies the conditions under which rationally motivated argumentation can take place. That is, it justifies how someone would reason to justify his or her claims. Without a basic logic circumscribing the inquiry process, there is only a loose set of activities.

Reasons and Motives for Evaluating Social Science Research

> Evaluation is one of the key analytic processes in all disciplined intellectual and practical endeavors…it is said to be one of the most powerful of the 'transdisciplines' that apply across broad ranges of the human investigative and creative effort…
>
> —Scriven (1991)

Evaluation is an essential characteristic of the human condition and is perhaps the single most important and sophisticated cognitive process in the repertoire of human reasoning and logic. Without such processes there is simply no means for distinguishing the bad from the good, the worthwhile from the worthless, or the important from the trivial (Coryn, 2007a). Simply stated, what's good is good, what's bad is bad, and it's the evaluator's job to determine which is which. When deciding which car to buy (or selecting a spouse), whether using Consumer Reports®, the opinions of friends and relatives, or personal experience with similar cars or the same manufacturer, one of the primary reasons why one evaluates becomes self-evident. That is, to determine the best course of action from among competing alternatives.

Similarly, the reasons and motives for evaluating social science research are many. These can range from mundane purposes, such as evaluating the existing literature on a particular social problem or phenomenon, to more sophisticated purposes, such as evaluating a body of social science research on a particular phenomenon to locate studies for use in a meta-analysis or to make decisions regarding effective policies, programs, or practices (e.g., "what works"). As noted in the Preface, this book is about being a critical consumer of social science research. Therefore, the reasons and motives for evaluating social science research as discussed throughout this book are:

1. To ascertain the methodological quality of research found in the social science research literature.
2. To ascertain the inferential basis for claims and conclusions found in the social science research literature.
3. Constructing, translating, and testing new and existing theories generated through evaluations of research found in the social science literature.
4. To design and execute better research by carefully attending to purposes #1, #2, and #3..

It is not, however, intended for high stakes evaluations of research, mentioned above, such as those conducted by the Campbell Collaboration, Cochrane Collaboration, What Works Clearinghouse, and similar repositories concerned with evidence-based research for purposes of policy making and identification of best practices in education, crime and justice, healthcare, and social welfare, for example (Donaldson, Christie, & Mark, 2008); though they too might benefit from the approach discussed in this book.

Chapter Summary

Social science research is the scientific study of human action and interaction focusing on elements of thought and behavior that are in some sense social. Social scientists typically study human action and interaction and thought and behavior in a systematic, rigorous, evidence-based, generalizing, replicable, and cumulative fashion. The principles for generating knowledge claims in the social sciences vary widely according to the particular world views held by social scientists. The legitimacy of knowledge claims in the social sciences are largely a function of methodological quality and the inferential basis for supporting those claims.

Chapter Exercises

These exercises are designed to reinforce understanding of the concepts discussed in this chapter. Answers to the chapter exercises are located in Appendix A.

1. What is social science research?
2. What types of questions do most social scientists investigate?
3. In what ways do positivism and post-positivism differ from constructivism and constructionism?
4. What are the central features of scientific method?
5. What are the differences between Type I and Type II errors?

Recommended Readings

Gerring (2001) provides descriptions of the central characteristics of social science research. Kline (2008) outlines the central principles for conducting high quality social science research. Miller and Salkind (2002), Shadish, Cook, and Campbell (2002), and Trochim and Donnelly (2006) present detailed accounts of a variety of research methods. Vogt (2005) provides an accessible reference source for terminology found in the social sciences.

Gerring, J. (2001). *Social science methodology: A criterial framework*. Cambridge, MA: Cambridge University Press.

Kline, R. B. (2008). *Becoming a behavioral science researcher: A guide to producing research that matters*. New York, NY: Guilford.

Miller, D. C., & Salkind, N. J. (2002). *Handbook of research design & social measurement* (6th ed.). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Trochim, W. M. K., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Mason, OH: Atomic Dog.

Vogt, W. P. (2005). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. Thousand Oaks, CA: Sage.