Chapter 2
Overview of the Research Process

Scientific research is the process of (1) developing an empirically answerable question, (2) deriving a *falsifiable* hypothesis derived from a theory that purports to answer the question, (3) collecting (or finding) and analyzing empirical data to test the hypothesis, (4) rejecting or failing to reject the hypothesis, and (5) relating the results of the analyses back to the theory from which the question was drawn. This last step usually involves revising the original theory to handle discrepancies between what the empirical data show and what the original theory posited, although the findings of only a single study usually are not sufficient to warrant major revisions of a theory. Nonetheless, a scientific research study, no matter how small the contribution, *must* make a new contribution to be considered original scientific research. In other words, a research study adds to our knowledge base. This requirement distinguishes a research paper from a report.

For the most part, the scientific understanding of a topic changes slowly, in part because science is about falsifying theories—and not proving them—and in part because a well-done study can generally only address a single, very narrow research question or hypothesis. Occasionally, some research leads to major changes in theory or a major change in an entire orientation that gives rise to theory (a "paradigm shift"; see Kuhn 1962), but generally, science progresses very slowly, with theories being "fine-tuned" with each additional study.

2.1 What Is NOT Research

Based on the above definition of scientific research, there are at least three things we can say do *not* constitute scientific research. I discuss these here, because they are commonly misconceived as constituting research, but they are not. The remainder of this book is geared toward discussing the role of statistics in actual research, and not in report writing.

S.M. Lynch, *Using Statistics in Social Research: A Concise Approach*,
DOI 10.1007/978-1-4614-8573-5_2,

1. A literature review alone is not research.

   Although reviewing and synthesizing literature is an important part of the research process, it is not, in-and-of-itself, research. The process of reviewing literature does not add to our knowledge base. In the "hard sciences," for example, writing a paper that delineates the history of Newtonian physics and the emergence of Einsteinian quantum physics does not constitute research. Any physicist knows this history without engaging in any experimental or nonexperimental work. Similarly, in social science research, writing a review of the state of the literature on race differences in educational attainment, for example, does not constitute original research.
2. Theory construction or forging links between theories or perspectives in the absence of empirical data is not research.

   Although the process of modifying and perhaps linking theories is an important part of research, it is not research by itself. For example, forging links between Newtonian physics and Einsteinian physics is not research. Any physicist should be able to derive such links analytically or at least conceptually. Similarly, in sociology, relationships between Durkheim's concept of anomie (see Durkheim 1997) and Marx's concept of alienation (see Marx 1988) are derivable without reference to any empirical data. In other words, theory construction and linking theories are logical, but not empirical processes. Furthermore, while it may contribute to our understanding of anomie and alienation, it does not contribute anything new to our understanding of the world.
3. Collection and analysis of data without theoretical grounding (i.e., having no research question) and offering no explanation of the data is not research.

   Data collection is an incredibly important—central—part of research; without data, no research could be done. However, data collection alone is not research, because there is no theory-driven question being asked nor answered. For example, the Census Bureau collects data on the size of the US population every decade (as well as additional factors, like age structure, racial composition, etc.), and it usually produces numerous reports on the various features of the population. This data collection and reporting is essential to social science research, because social scientists often use Census data in their own research, but the Census Bureau's collection and reporting process itself is not research, because there is no research question driving the collection and reporting. In other word, there is no direct contribution to our knowledge base regarding a topic of scientific interest.

2.2 Replication as Research

One activity that may not appear to be research, but in fact is, is replication. Replication is the process of repeating an experiment or study to verify the original findings. Technically speaking, true replication involves mimicking a previous study exactly; that is, using the same data collection and analysis method, and observing whether the original research results hold. As a simple example of replication, take the construction of a volcano for a science fair project in junior high school. The typical such exercise involves building something that looks somewhat like a volcano (usually, this is where students concentrate most of their effort, although this is not the actual "science" part of the process) and mixing vinegar and baking soda in a container at the volcano's peak. Mixing these two substances produces a chemical reaction that causes "lava" to spew out of the top and pour down the sides. If the student hypothesized that the mixture of an acid (vinegar) with a base (baking soda) will produce a chemical reaction that generates a gas and something else, and the student verified this with the empirical data (the "eruption"), then this process could be considered replication of a test of some centuries-old chemistry hypothesis.

Replication is an important part of the scientific process, although it receives very little attention in contemporary social science literature. Frankly, it is very difficult to get such work published. Usually, seemingly replicative research is not truly replicative; instead, it may differ in data and method very slightly from the original work. For example, it may involve analyzing data from a previous study in a slightly different way. Technically, this does not constitute replication, but it is research. More importantly, this type of work constitutes the bulk of social science research: most social science research elaborates on extant research by expanding it in relatively minor ways. For example, one study may find that stress is related to mortality risk in the general population. A second study—using the same data, perhaps—may find that this relationship holds for persons over age 45, but not for persons under 45. This finding then leads to a revision of the theory that produced the hypothesis that stress affects mortality, but it does not invalidate it—it simply qualifies it.

2.3 Stages of Research and Scientific Paper Structure

Although I have defined the overall process of research as involving five primary steps, the process of research can be decomposed more specifically into a number of sub-steps. Table 2.1 shows this decomposition and its relationship to the parts of scientific research papers seen in contemporary research journals.

These ten steps in the research process can be subdivided into additional steps, and some of these steps will be the focus of the material covered in this book. We will only briefly review the first five steps—corresponding to the introduction and theory sections of a research paper—in this chapter; steps 6–10 will constitute the focus of the remainder of the book. I note at the outset that these research

The Research Process	Paper Section
(1) Start with a Perspective (2) Select a Theory (3) Derive a Proposition (4) Ask a Question (5) Derive Hypotheses	Introduction & Lit. Review/Theory
(6) Find or collect data (7) Analyze data	Data & Methods
(8) Report results & answer question	Results
(9) Interpret results in terms of theory (10) Draw implications for theory	Discussion & Conclusions

Table 2.1. Correspondence of the parts of the research process with the parts of a scientific paper.

components as I have spelled-them out may not be universally agreed-upon. For example, there are several ways to define a proposition and a hypothesis. As another example, I have seen some definitions of a proposition and a research question that equate them; their only difference in that case being whether one poses it as a question (research question) or a statement (proposition).

2.3.1 What Is a Perspective?

A perspective is a general orientation toward the world. Perspectives are ultimately untestable but simply frame the world in a particular way. Therefore, there really shouldn't be much room for serious debate about them, and research ultimately can't confirm nor disconfirm them. For example: are you politically liberal or conservative? Do you have a generally optimistic or pessimistic view of the world? Do you view human nature as inherently good or inherently evil?

As a more specific example of perspectives, within the social sciences (sociology in particular) the two chief perspectives are the functionalist and conflict

perspectives. The functionalist perspective views the social world as a system of components that function together to maintain the whole. The conflict perspective views the social world as a system of antagonistic components held together (perhaps) by force. From the functionalist perspective, a behavior like crime can be viewed either as (a) a dysfunction that is corrected or held in check over time via institutions like the criminal justice system, or (b) a phenomenon that actually serves a function—like demarcating and clarifying boundaries between acceptable and unacceptable behaviors when it is punished. A conflict perspective, on the other hand, might view crime as a response to the oppression of one social class (the poor) by another (the wealthy), or it might view it as one more way the wealthy control the poor (e.g., consider the punishment for blue collar versus white collar crime).

These larger orientations toward the world can not be confirmed nor disconfirmed by data for several reasons. First, they are simply too broad: even if one *could* find evidence that supported, say, a contemporary liberal position on a given topic, the evidence could not be construed as validating liberalism as a whole. For example, there is little doubt that Social Security—a liberal policy—improved the quality of life for elders in the US. That fact, however, does not imply that the government always has the best (or even a) solution to any social problem.

Second, and relatedly, general perspectives may seem right in some contexts and wrong in others. Consider, for example, the statement that "the government should stay out of our lives." This view is held by both political liberals and conservatives, but in different contexts. When liberals make this statement, they mean that the government should not be able to restrict behaviors, like recreational drug use and homosexual marriage. When conservatives make this statement, they mean that the government should not be able to levy heavy taxes or regulate businesses. Very few hold views that are both socially and economically liberal (or conservative).

Third, and perhaps most problematic, almost any empirical data that are observed can be interpreted as being consistent with any perspective. In general, perspectives are so broad that they have more than one way—and possibly contradictory ways—to predict or explain any data. For example, suppose I believe that humans are selfish by nature. I might predict, then, that no one would jump in front of a bus, risking certain death, to save a child from being hit. Yet, certainly this sort of behavior happens, and it does fairly often. Does this occurrence invalidate my perspective? No: there could be numerous followup explanations that make the data consistent with—or irrelevant to—the perspective. One might argue that the individual was ultimately being selfish, because s/he simply wanted recognition for being a hero and possibly miscalculated the risk. One might argue that the individual may have been temporarily insane, not making a rational calculation at all (either selfish or altruistic). One could argue—if the person the individual saved were a relative—that the act was a selfish attempt to further the family genes. If the person were unrelated, one could argue that saving the child helps preserve the species for at least another generation.

In short, a perspective cannot be rejected. Data can either be interpreted directly to support the perspective, or a question-begging (i.e., circular) argument can be

made to oppose evidence to the contrary: Altruistic behavior cannot occur, because I can interpret any behavior that seems altruistic in a fashion that indicates it isn't.

If perspectives cannot be tested, then why are we discussing them here as part of the scientific process? I list starting with a perspective as part of the process, because we need to be aware that we all start with particular orientations toward the world. We are not *tabula rasa,* that is, blank slates that simply digest data as it emerges. Instead, we all view the world through particular lenses, and these lenses affect how we interpret data. Put another way, data are not unambiguous; they do not "speak for themselves." To guard against letting our own biases get the better of us while we are conducting research, we need to first recognize that we have them. One way to reduce the influence of our biases is to work at a "lower" level than that of a perspective; that is, to derive smaller, more manageable, testable (and falsifiable) ideas.

2.3.2 What Is a Theory?

A theory is a systematic description of how the world (or part of it) "works." Generally, a theory is too broad to be tested in its entirety in a single study, but it offers a framework for understanding how things in the world operate. Therefore, theory enables prediction, which we will discuss shortly (as hypotheses). Within sociology, Durkheim's theory of modernization seeks to explain why modern societies "hold together" as a unit rather than fragment rapidly into anarchy or chaos, and what the consequences of maintaining a society are (Durkheim 1997).

A short version of this theory is as follows. Pre-modern societies maintain stability because people within them are very similar to one another—everyone knows how to (and does) perform the basic tasks necessary for survival, everyone shares similar understandings of events like why the sun comes up, why it rains, etc. With such similarities, there is little to disagree about, and so the society is held together by a force that Durkheim called "mechanical solidarity," where "mechanical" refers to the automatous nature of people in the society. However, as societies develop the ability to produce surpluses (especially of food, e.g., via agricultural revolutions), a division of labor begins to emerge. That is, people have the ability to specialize in particular types of labor once the basic necessities of life for all of a society can be produced by a few (e.g., in the US, approximately 2 % of the population is engaged directly in food production; thus, the other 98 % can do something else). However, with the division of labor come two significant changes. First, people become interdependent; they are no longer self-sufficient. Second, the commonality of societal beliefs begins to fragment. Regarding the first change—the emergence of interdependence—Durkheim argued that societies still hold together, but the "glue" changes. Society no longer evidences mechanical solidarity, but rather "organic solidarity," where organic solidarity is the interdependence of the components of society (think of the organ systems of the body—each has its own function, and none can exist without the others). The second change—the erosion

of common beliefs—stems from the dissimilarity in the day-to-day lives of the members of the society. For example, consider the different lives of a scientist and a preacher, and consider the different sets of beliefs held by the two (e.g., consider evolution vs. creationism). Eventually, Durkheim argued that the dissimilarity of the members of a society would lead to "anomie," a state of normlessness in which society's members would be uncertain about how to behave. At the individual level, this condition of anomie would produce angst.

Notice how this theory is fairly broad and encompasses a number of steps. No single research project could possibly address all the components of the theory.

2.3.3 What Is a Proposition?

A proposition is a *single* potentially testable component of a theory. The theory described above is a condensed version of Durkheim's original theory, but yet a number of propositions can be derived from it. One proposition, for example, might be that societal development of a division of labor generates anomie.

Notice that this proposition follows from the original theory, but it does not encapsulate ALL that the theory entails. It does not, for example, say anything about any prior stage of the theory, in terms what leads to the production or expansion of a division of labor.

2.3.4 What Is a Research Question?

A research question may be considered a refined restatement of a proposition that is testable. For example, we can revise the proposition above as: Do societies with more extensive divisions of labor evidence greater levels of anomie? (or: Does the expansion of a division of labor produce anomie?) Notice that the research question is (1) framed as a question, and (2) makes the proposition testable by suggesting a way to evaluate it: societies may be compared on measures of division of labor and anomie in order to answer the question.

2.3.5 What Is a Hypothesis?

A hypothesis is a *falsifiable* statement that makes a prediction derived from the research question that suggests a relationship between variables. In other words, it is a statement *that can be disproven*. A single research question may generate several hypotheses. The refinement of a research question into a hypothesis typically includes clarifying exactly what some of the concepts in the proposition and question mean. For example, what is a division of labor? What is anomie? How do these concepts link with empirically-measureable phenomena?

Some methodologists say that hypotheses are statements about expected relationships between variables that "operationalize" (i.e., measure) the concepts discussed in the research question. Often, however, the hypotheses presented in scientific articles are presented in the second section of the paper (the literature review and theory section), while the actual operationalization of the concepts occurs in the data section. Regardless, hypotheses stem from the research question and are more specific. So, for example, regarding the research question above, a hypothesis might be: societies with a more heterogeneous occupational structure will have members with higher levels of anxiety. Here, we have refined the terms "division of labor" and "anomie" with the terms "heterogenous occupational structure" and "anxiety." We certainly need to clarify these terms as well, but this is generally done in the data section.

2.4 Some Summary and Clarification

The distinction between theory, propositions, research questions, and hypotheses is admittedly somewhat fuzzy. This is largely because there are different definitions for each across generations of researchers and across scientific disciplines. Within social science research there has been a trend over the last half-century away from the formal spelling-out of propositions and hypotheses in research articles. Instead, most articles now simply present a research question, followed by theory that purports to answer the question, followed by a test of part of the theory. Even the magnitude of a theory may differ across fields and time. For example, early work in sociology was concerned with developing "grand theories" that were large and attempted to explain all of human behavior. Over the last several decades, however, smaller "middle range" theories have emerged to explain more narrow forms of behavior (Merton 1968). These are often called "substantive theories" today.

The location of the research question as the fourth step in the research process is also tenuous. We generally begin with a question of interest (perhaps call this step 0). However, as we learn more about research in the topic area, more often than not, we find that our initial question is too broad and has to be refined. Thus, one could consider our original question as a "research question," but, based on our outline here, the true "research question" will be much narrower and more precise than our original question.

2.5 What Research Cannot Do: Proof

The process of moving from a theory to a proposition to a research question and hypotheses is a *deductive* one, that is, each stage follows from (is a *necessary consequence of*) the stage prior to it. A consequence of this deductive process is

Symbolic	Generic language	Example
$A \rightarrow B$	If A is true, then B is true	If it rains today at my home, my yard will be wet
B	B is true	My yard is wet
$\therefore$	Therefore,	Therefore,
A	A is true	It rained

Table 2.2. An example of the fallacy of affirming the consequent.

that theories cannot be proven. Put another way, *science is not a process of proving theories.* Some basic logic shows the problem with the notion of proof. Consider the argument structure presented in Table 2.2.

At first glance, the argument may seem reasonable. However, it is invalid—it is a "fallacy of affirming the consequent"—because it is possible that $C \rightarrow B$. In other words, B may be a consequence of A, but it may also be the consequence of some other premise (or theory), C. It is possible, for example, that 1,000 (or more) leprechauns urinated on my lawn while I was away. More realistically, my wife may have watered the lawn. In short, the data may be perfectly consistent with my original theory but data cannot, in fact, validate it.

More generally, if a theory makes a prediction regarding some pattern in data, the fact that real data may evidence the predicted pattern does not prove the theory true, because another theory may make the same prediction. To make this idea more concrete, suppose the data do in fact show that societies with heterogenous occupational structures have higher rates of anxiety. This does not suggest that Durkheim's theory is correct, because an alternate explanation might be Marx's theory of alienation resulting from capitalistic exploitation of labor (Marx 1988). As another example, in physics, most predictions about the physical world were consistent with classical, Newtonian physics. However, once better measurement capabilities emerged, physical reality was more consistent with quantum theory, leading to a shift in support for that theory. In sum, when we find evidence that is consistent with a hypothesis/theory, we can only say that the evidence is consistent with the hypothesis/theory. The evidence does not prove the theory true (see, for example, Hawking 1988).

Although we cannot prove a theory true through empirical research, we may be able to disprove one. Consider an alternative, but similar logical structure, as shown in Table 2.3. Although this argument structure looks like the one presented before, it is fundamentally different. It says that, if we find out that the consequent (B) of a conditional statement ($A \rightarrow B$) is untrue, then the original premise (A) cannot be true. In other words, if a theory predicts some pattern will be observed in empirical data, and that pattern is not observed, then the theory cannot be true. For that matter, no theory that makes the same prediction could be true. This argument structure is

Symbolic	Generic language	Example
$A \rightarrow B$	If A is true, then B is true	If it rains today at my home, my yard will be wet
$\neg B$	B is not true	My yard is not wet
$\therefore$	Therefore,	Therefore,
$\neg A$	A is not true	It did not rain

Table 2.3. An example of the logically valid argument structure of *modus tollens*.

logically valid and is called "modus tollens" (see Bonevac 2003). Scientific research is therefore geared toward falsifying theories—or hypotheses drawn from them—and not proving them true (see Popper 1992).

Given that the goal of science is to falsify theories, there is a considerable amount of confusion between the scientific conception of a theory and the common usage of the term. Consider, for example, the ongoing debate regarding evolutionary theory, in which opponents have denigrated evolution by claiming it is merely "a theory." In contrast to the common view of a theory as mere speculation, a scientific theory is *not* simply speculation. It is an explanation for how some aspect of the world works that is falsifiable and has "stood the test of time" by not being falsified by evidence in repeated studies (Coyne 2009). Given that it is logically impossible to prove a theory true, there is no greater status for an explanation to attain than as a "theory."

2.6 Conclusions

In this and the previous chapter, we discussed the process of scientific investigation and its relationship to the components of a research paper. This process—of science and scientific writing—is more or less the same for every discipline, from the so-called "hard" sciences to the social sciences. The key difference across disciplines is in the subject matter, and therefore, the type of data used and the methods for gathering it. We will discuss these issues in the next chapter before turning to the process of analyzing data in Chap. 4 and beyond.

2.7 Items for Review

Be familiar with the following concepts, terms, and items discussed in this and the previous chapter:

- Falsification/falsifiability
- Stereotype

- Hasty generalization
- Replication
- Scientific theory
- Proposition
- Research question
- Hypothesis
- Deduction
- Fallacy of affirming the consequent
- *Modus tollens*

2.8 Homework

Answer the following questions:

1. What differentiates the scientific definition of a theory from the common-use definition of the term?
2. Why can a theory not be proven?
3. Why is a perspective untestable?
4. Explain the difference between a research question and a hypothesis.
5. What does "falsifiable" mean?
6. Give three examples of the fallacy of affirming the consequent.
7. Give three examples of arguments that use *modus tollens.*
8. Explain how the *modus tollens* argument form is useful for science.

http://www.springer.com/978-1-4614-8572-8

Using Statistics in Social Research
A Concise Approach
Lynch, S.M.
2013, XXIII, 229 p. 44 illus., Hardcover
ISBN: 978-1-4614-8572-8