PREFACE

Why This Book?

There are many intro texts on methods for social research. So, why this one? The title says it all: These are the essential methods of social research—the methods that we believe every undergraduate needs to know, whatever their major or career goals are. Students who master the methods in this book will have foundational skills in the design and conduct of research on human thought and human behavior. They can cite these skills when they apply for jobs or for graduate schools. We also love methods. As methodologists, we think there's nothing more exciting than inventing, innovating, and testing methods. Methods are the heart of science—and we hope you can feel our passion and playfulness about methods as you read this book.

What's in This Book?

There are 13 chapters in this book, covering research design, data collection, data analysis, and write-up. There is no chapter devoted to ethics in this book because ethics is too important to deal with that way. It's part of *everything* we do in social science, and we deal with this all-important topic throughout the book.

The first three chapters are about what goes on before we collect or analyze any data: Chapter 1 is on the language of social research—the basic vocabulary, including levels of measurement and tests for reliability and validity of measurements. Chapter 2 is on setting up a research project: defining a research question, and doing the kind of literature search required. And Chapter 3 is on research design and on writing proposals for research. We want students to practice the craft of proposal writing as part of their basic instruction on the methods of social research.

Chapter 4 introduces students to the logic of the experimental method: formulating a hypothesis, measuring the independent and dependent variables, and controlling for threats to validity. Students who never conduct an experiment need to control this logic in order to design and conduct convincing research.

Chapter 5 is about the basics of probability sampling, including simple and systematic random sampling, stratified random sampling and cluster sampling, and determining the size of representative samples. Chapter 6 is on nonprobability sampling, including theoretical sampling, quota sampling, purposive sampling, snowball sampling, respondent-driven sampling, case-control sampling, and convenience sampling. Nonprobability sampling is becoming increasingly popular, and not just for studies based on qualitative data. A key feature of Chapter 6 is the section on minimum sample sizes for 10 qualitative analysis methods.

Chapter 7 deals with interviewing and focus groups. In unstructured interviewing, the idea is to get people to open up and to let them express themselves in their own terms and at their own pace. Semistructured interviewing follows a written list of questions and topics that need to be covered in a particular order. Focus groups—a kind of semistructured interview—are recruited to discuss a particular topic, like people's reaction to a television commercial or their attitudes toward a social service program. Survey researchers may use focus groups to help design a questionnaire and also to help interpret the results of surveys.

Response effects are measurable differences in interview data that are predictable from characteristics of respondents, interviewers, and environments. They are a problem in all interviewing, and Chapter 7 ends with a review of this important problem.

In fully structured interviews, like questionnaires, people are asked to respond to as nearly identical a set of stimuli as possible. Chapter 8 covers questionnaire design, improving response rates, asking questions about sensitive topics, using interviewers in team research, and translating questionnaires from one language to another.

Chapter 9 is about collecting social network data. Most social science data are about characteristics of individuals: their gender, their age, their income, their answers to questions about their behavior and feelings. … Network data are about relations among people: who they like and who they interact with. We'll introduce you to some of the special techniques for analyzing network data in Chapter 12.

Chapter 10 is on methods for field research: direct, indirect, and participant observation. Direct observation involves watching people and recording their behavior, using methods like continuous monitoring, spot sampling, and experience sampling. Direct observation is reactive when people know that you are watching them (people can play to the observer), so some researchers use unobtrusive observation. This involves deception, which raises obvious ethical issues.

Participant observation turns fieldworkers into instruments of data collection and data analysis. This requires certain skills, which include learning the local language, dialect, or jargon; developing explicit awareness; building memory; maintaining naiveté; learning to hang out and build rapport; maintaining objectivity; and learning to write clearly.

Most of the recoverable information about human thought and behavior is in naturally occurring text—diaries, property transactions, recipes, correspondence, song lyrics, billboards, books, magazines, transcripts of TV and radio shows, newspapers, artifacts, still and moving images, advertisements. Chapter 11 covers methods for analyzing those data: discourse analysis, critical discourse analysis, conversation analysis, narrative analysis, content analysis, language in use, and grounded theory.

Chapter 12 introduces methods for analyzing quantitative data, including (1) univariate analysis—that is, statistics that describe a *single variable*, without making any comparisons among variables; (2) bivariate analysis—statistics that describe *relationships in pairs of variables*; and (3) multivariate analysis—statistics that describe *relationships in larger sets of variables* and that let you test hypotheses about what causes what.

Neither the material in this book nor a course in the use of statistical software is a replacement for taking statistics from professional instructors of that subject. Nevertheless, after

working through Chapter 12, students should be able to use basic statistics to describe their data and be able to take their data to a professional statistical consultant—and understand what she or he suggests.

Finally, in Chapter 13 we take students through nine rules for writing up research for publication. Many students are conducting novel and publishable research, but aren't yet thinking about getting published. There is a world of difference in what passes for writing a good term paper and what is required for getting past professional reviewers—this chapter can help as you begin the transition from student to published researcher.

Some Features of the Book

1. *Opening objectives and introductions to the chapters*: Each chapter begins with a page called "In This Chapter" that lays out all the major headings. This page repeats the material in the detailed Table of Contents for each chapter and makes it easier to understand, at a glance, what's coming.

2. *Key concepts*: All new terms appear in bold-faced **type like this**—but only the first time they appear. Then, as a study aid, there is a page at the end of each chapter titled "Key Concepts in This Chapter." That page has all the bold-faced terms in the chapter.

3. *Summary and exercises*: Following the Key Concepts page, there are two more study aids: a summary of the main points of the chapter and a series of exercises for illustrating and nailing down some of the main points.

4. *Further reading*: Each chapter ends with recommendations for further reading. Some of these recommendations are classics in the literature on social science methods; others are recent examples that further illustrate some of the key points in each chapter. All the extra readings are pointed to in the text with a bold reference to **Further Reading.**

5. *Boxes labeled: "Amber's Corner," "Russ's Corner." "Tips & Tricks," and "Deep Dives."* In these pedagogical devices, we report on our own experiences in doing research and on things we've found helpful in our teaching of research methods.

HRB and AW

February 20, 2026

Chapter 1

What is Social Science Research?

WHAT'S IN THIS CHAPTER

Introduction

Let's get right to the point: Anyone studying the causes or consequences of human thought or human behavior is doing social science research—and it's everywhere. Our public schools are scenes of one social science experiment after another, as we search for better ways to help children learn. Our cities are scenes of hundreds of programs designed with social research to help people develop their employment skills or gain access to health care or find shelter. Hospitals across the world are scenes of surveys and experiments and direct observations designed to bring down the rate of iatrogenic (doctor-caused) and nosocomial (hospital-caused) illnesses. We are bombarded with ads to buy this or that thing, to vote for this or that candidate, to give to this or that charity.

Traditionally, there were a handful of social sciences and they were initially distinguished by the methods they used. Psychologists and social psychologists did experiments; sociologists and political scientists did surveys; anthropologists did field research; geographers made maps; economists built and tested mathematical models. All those traditional disciplines are still thriving, but today, hundreds of thousands of people in dozens of fields—communications, criminology, marketing, international studies, nursing, epidemiology, gender studies, urban studies, and on and on—do social science research and no discipline owns any of the methods.

And social science methods aren't just for professional researchers. Anyone in industry or government called on to write a report on worker productivity or the cost of some public works project will be using one or more of the methods in this book.

No one is expert in all the methods available for research. But seasoned social scientists all know about the array of methods available to them for collecting and analyzing data. By the time you get through this book, you'll have a good idea of the range of methods used in the social sciences and what kinds of research problems are best addressed by the various methods.

What Is Social Science Research?

We are always on the lookout for ways to extend and make more comfortable our own lives and the lives of our families. In the absence of any hard information about how to do that, humans quite naturally mystify the forces that make some people rich and some poor, make some people sick and others healthy, and make some people die young and others live a long time. From its beginnings in the sixteenth century, modern science has been demystifying those forces. Science, including social science, is about the systematic creation of knowledge that provides us with the kind of control over nature—from the weather to disease to our own buying habits—that humans have always sought.

Some people are uncomfortable with this "control over nature" metaphor. When all is said and done, though, few people would give up the material benefits of science. Try getting middle-class people in Florida or Arizona to give up air conditioning for just one day in the summer or asking people to reject taking antibiotics when they have pneumonia and you'll find out in a hurry about the weakness of ideology compared to the power of creature comforts. If running air conditioners pollutes the air or uses up fossil fuel and if overuse of antibiotics leads to new strains of drug-resistant bacteria, most people rely on *more* science (new sources of energy and new antibiotics, we hope) to solve those problems.

Air conditioning and antibiotics, like cell phones, self-driving cars and artificial limbs are some of the *products* of physical and biological science. Products of social science include marketing, all forms of insurance, polling, management, and public health programs. In fact, every industry in the world that relies on risk management is a product of fundamental knowledge produced by social science. One of those industries is state lotteries—taxes on people who are bad at math (Petty 1899 [1690]:64).

Today, social science data are being used more than ever in the development of public policy. How much leisure time should we have? What kinds of tax structures are needed to support our retirement and medical systems? How can economic transitions to sustainability be managed? The answers to these questions must be guided by our values, but empirical data can help us consider the feasibility, costs, and benefits of different options.

Ethics and Social Science

The biggest problem in conducting a science of human behavior is *not* selecting the right sample size or making the right measurement. It's doing those things ethically, so you can live with the

consequences of your actions. We're not exaggerating about this. Ethics is part of method in science, just as it is in medicine or business, or any other part of life. For while scholars discuss the fine points about whether a true science of human behavior is really possible, effective social science is being done all the time and with rather spectacular, if sometimes disturbing, success.

The scientific method is barely 400 years old and its systematic application to human thought and behavior is less than half that. But since the eighteenth century, every phenomenon to which the scientific method has been systematically applied, over a sustained period of time, by a large number of researchers, has yielded its secrets, and the knowledge has been turned into more effective human control of events.

And that includes human thought and behavior. When a science of human affairs was founded in the mid-nineteenth century, no one could predict the outcome of elections, or help people through phobias with behavior modification, or engineer the increased consumption of a particular brand of cigarettes. We may question the wisdom of engineering cigarette purchases in the first place, but the fact remains, we *can* do these things, we *are* doing these things, and we're getting better and better at it all the time.

It hardly needs to be pointed out that the increasing effectiveness of science over the past few centuries has also given human beings the ability to cause greater environmental degradation, to spread tyranny, to entrench racist social structures, and even to cause nuclear war. This makes a science of humanity even more important now than it has ever been before. (**Further Reading:** ethics and social science)

Consider this: In 1992, marketers in a midwestern city, using the latest supercomputers, found that if someone bought disposable diapers at 5 p.m., the next thing he or she was likely to buy was a six-pack of beer. So they set up a display of chips next to the disposable diapers and increased snack sales by 17% (Wilke 1992). At the time, 35 years ago, that was a breakthrough in the monitoring of consumer behavior. Now, market researchers simply track all your buying behavior across websites, and develop ads that are targeted just for you.

We need to turn our skills in the production of such effective knowledge to important problems: hunger, disease, poverty, war, environmental pollution, family and ethnic violence, and racism, among others. Social scientists can play important roles in social change by predicting the consequences of ethically mandated programs and by refuting false notions (such as various forms of racism) that are inherent in most popular ethical systems.

Don't get us wrong here. The people who discovered that fact about the six packs and the diapers were darned good social scientists, as are the people today who design all those automated data-collection mechanisms for monitoring your behavior on the Internet. We're not calling for rules to make all those scientists work on problems that we think are important. Scientists choose to study the things that industry and government pay for, and those things change from country to country and from time to time in the same country. Science has to earn its support by producing useful knowledge. What "useful" means, however, changes from time to time even in the same society, depending on all sorts of historical circumstances.

Suppose we agreed that "useful" meant to save lives. Over three times as many people died in motor vehicle accidents in the United States in 2024 as died of ovarian cancer—about 40,000 and 13,000 respectively. Should we spend three times more money teaching safe driving than we do researching the causes of ovarian cancer?

We think the answer is pretty clear. In a democracy, researchers and activists want the freedom to put their skills and energies to work on what they think is important. That's just how it is—and, personally, we hope it stays that way. In the rest of this book, we deal with some of the methods we can use to make useful contributions. But *you* have to decide what those contributions will be, and for whom they will be useful.

The Language and Logic of Social Research

We move next to the key concepts in social research and the vocabulary for those concepts. The fundamental concepts of social research include: variables, measurement, reliability, and validity. When you finish this section, you should understand the crucial role of **measurement** in science and begin to appreciate the mutually supportive roles of data and ideas in the development of theory.

You should also have a new skill: You should be able to **operationalize** any complex human phenomenon, like "being modern" or "anomie" or "alienation" or "readiness to learn research methods." You should, in other words, be able to reduce any complex variable to a set of measurable traits. And you should also become very critical of your new ability at operationalizing. Just because you *can* make up measurements doesn't guarantee that they'll be useful or meaningful. The better you become at concocting clever measurements for complex things, the more critical you'll become of your own concoctions and those of others.

Variables: The Joy of Measurement

A **variable** is something that can take more than one **value**, and those *values can be words or numbers*. If you ask a woman how old she was at her first pregnancy, the answer will be a number (16 or 40, or whatever), but if you ask her about her religion, the answer will be a word ("Muslim" or "Methodist").

The most common variables in social research are age, gender/sex, race/ethnicity, education, income, marital status, and occupation. Others that you might see include blood pressure (in medical social science), number of children (lots of studies use this one), number of times married, distance from an airport (or a hospital, or a welfare agency, or a public library, or a bus stop), and level of support for various causes (a woman's right to an abortion, the distribution of clean needles to IV drug users, sex education for 5th graders, etc.).

Social research, whether it's based on questionnaires, field observations, or experiments, is based on defining variables, looking for associations among them, and trying to understand whether—and how—variation in one thing causes variation in another. Research affects people, though, so all research has an ethical component. We'll have more to say about this throughout the book. (**Further Reading:** measurement in the social sciences)

Dimensions of Variables: Simple or Complex?

Variables can be **unidimensional** or **multidimensional**. Unidimensional variables, like height, weight, birth order, age, and marital status, are relatively easy to measure. We often talk about

political orientation as if it were unidimensional, with people lying somewhere along a line between conservative and liberal. But if you think about it, people can be liberal about some dimensions of life and conservative about others. For example, you might agree strongly with the statement that "Men and women should get equal pay for equal work" and also with the statement that "A strong military is necessary to defend freedom." These statements test political orientation about domestic economic policy and foreign policy—two of the many dimensions of political orientation.

Multidimensional variables—like stress, wealth, and political orientation—are more difficult to measure. Even something as seemingly straightforward as income is multidimensional. To measure the annual income of various Americans or Canadians, for example—cab drivers in St. Louis, sous chefs in New York, retired people in Tampa, sociologists in Toronto—you may have to account for salaries, social security, private pension funds, gifts, gambling winnings, tax credits, interest on savings, wages paid entirely in cash (including tips), food stamps, contributions from extended kin, and so on.

Complex variables can sometimes be reduced to a series of simpler variables. We'll discuss the building of scales and how to test for unidimensionality in Chapter 12.

Dependent and Independent Variables

Consider life insurance: The company predicts how long you will live, given your sex, age, education, weight, blood pressure, and a few other variables. They bet that you will *not die* this year. You take the bet. If you lose (and remain alive), the company takes your annual premium and banks it. If you win the bet (and die), the company pays your beneficiary.

For insurance companies to turn a profit, they have to win more bets than they lose. They can make mistakes at the individual level, but in the *aggregate* (that is, averaging over all people) they have to predict longevity from things they can measure.

Longevity, then, is the **dependent variable**, because it *depends on* sex, education, occupation, etc. These are called **independent variables** because they are logically prior to—and therefore, independent of—the dependent variable of longevity. How long you live doesn't have any effect on your sex.

It's not always easy to tell whether a variable is independent or dependent. Do adolescent girls get pregnant because they are poor, or is it the other way around? Does the demand for cosmetic surgery stimulate the production of plastic surgeons, or is it the other way around?

A lot of mischief is caused by failure to understand which of two variables depends on the other. In a classic study, Oscar Lewis (1959, 1966), described what he called a "culture of poverty" in five Mexican families. People who live in a culture of poverty, said Lewis, are not very future oriented. This plays out, he said, in their shopping for food every day and in never buying large economy sizes of anything. Lewis's point was that truly poor people can't invest in soap futures by buying large boxes of it. He saw a low level of expressed orientation toward the future, then, as the dependent variable and poverty as the independent variable.

Many people interpreted Lewis's work as meaning exactly the opposite: that poverty is caused by a low level of future orientation. According to this topsy-turvy, victim-blaming

reasoning, if poor people—in rural Appalachia, slums in Delhi, *favelas* in Rio de Janeiro—would just learn to save their money and invest in the future, then they could break the poverty cycle.

This **educational model of social change** is the basis for one of the world's biggest industries—social change and development. But the model is mostly ineffective because behavioral change (the supposed dependent variable) doesn't usually depend on education (the supposed independent variable). Getting the relationship between independent and dependent variables right, then, isn't just a matter of good science—it's a matter of good ethics too. (More about this in Chapter 3 on the role of theory in developing testable hypotheses.)

Concepts and Variables

One of the most famous variables in social science is socioeconomic status (SES). SES is a **concept**, or mental construction. All variables are concepts, but some concepts, like height and weight, are easier to measure. Concepts like SES, race/ethnicity, or religiosity are complex and much more difficult to measure. These complex concepts are often called **constructs.** Whether simple or complex, variables are measured by their **indicators**, and indicators are defined by their values.

To measure SES, you can use income as a single indicator. But there are many wealthy people who have low SES, like the so-called nouveau riche. And many relatively low-income people who have high SES—think of those down-at-the-heels nobles in England who have to open their castles to tourists to make ends meet. You can add level of education to income as an indicator, but that still might not be enough to get at something as multidimensional as SES. You can add parents' occupations, and so on, and you still might wind up dissatisfied with the result, but you'll get closer.

Improving measurement in science means lowering the probability of and the amount of error. It may be easier to measure some concepts than others, but the fact is, all measurement is difficult. Measuring complex variables, which lack concrete indicators, is one of our biggest challenges in social science because these variables are mostly what we're interested in.

Conceptual and Operational Definitions

There are two ways to define variables—conceptually and operationally. **Conceptual definitions** are like dictionary definitions—abstractions, told in words, that facilitate understanding. **Operational definitions** consist of a set of instructions, like a recipe, on how to measure a variable that has been conceptually defined.

Conceptual Definitions

Suppose a friend tells you that "Parker and Pat just moved to a spacious house." Nice concept. You ask the friend: "What do you mean by 'spacious'?" and they say: "You know, big rooms, lots of room for all their furniture, high ceilings."

If that isn't enough for you, you'll have to move from a conceptual definition of "spacious" to an operational one. This means agreeing with your friend on exactly what to measure: Are you counting the square footage or the cubic footage? Does the screened-in porch count? Are you including the area under the high ceilings? If we had to agree on things like this for every concept, ordinary human discourse would come to a grinding halt.

Science is not ordinary human discourse, however, and this is the most important difference between the humanistic and the scientific approaches to social science. Humanistic researchers seek to maintain the essential feel of human discourse. Positivists focus more on specific measurement. These two styles are not inimical to one another. They are complementary.

Conceptual definitions are at their most powerful when they are linked together to build theories that explain research results. There are three things one wants to do in basic science: (1) describe a phenomenon of interest; (2) explain what causes it; (3) predict what it causes. The existence of a conceptual variable is inferred from what it predicts—how well it makes theoretical sense out of a lot of data.

Operational Definitions

Conceptual definitions are limited. They point us toward measurement, but they don't really give us any recipe for measurement. Without measurement, we cannot make useful comparisons.

AMBER'S CORNER

FROM CONCEPT TO MEASUREMENT: 25 YEARS AND COUNTING

As a newly minted Ph.D., I was asked in the talk for my first job: "How are you so confident you can measure all these concepts that nobody has measured before?" My answer? "Methodologists! We're all brash like this." I was trying to be funny, but my ongoing struggle to define a classic concept, *moral economy*, was no joke.

The idea had been around since the 1970s to explain why people riot over food shortages (Thompson 1971, Scott 1976). After Bolivia's Water War of 2000, I started thinking about how this might apply to water. It took me about 10 years to come up with a decent conceptual definition of a moral economy for water (Wutich 2011). And then I got stuck.

I lucked out a decade later, when economic anthropologist Melissa Beresford took the lead on our collaborative team of water scholars from economics, geography, and anthropology (Beresford et al. 2022). Her expertise gave us the boost we needed to develop an operational definition. We create a code system using **qualitative measurement**—words, not numbers—to operationalize *moral economies for water*.

We're still a long way from quantitatively measuring moral economies for water. But we're getting closer. And the more people who work on refining the conceptual and operational definitions, the faster we make progress—together.

Operational definitions specify exactly what you have to do to measure something that has been defined conceptually. Here are four examples of operational definitions:

1. Intelligence: Take the Wechsler Adults Intelligence Scale (WAIS) and administer it to a person. Count up the score. Whatever score the person gets is their intelligence.
2. Machismo: Ask a man if he approves of women working outside the home; if he says "no," he scores 1, and if he says "yes," he scores 0. Ask if he thinks women and men should have the same sexual freedom; score 1 for "no," and 0 for "yes." Ask if the birth of male children is more important than female children; score 0 for "no," score 1 for "yes." Add the scores. The higher the score, the more machismo a man has.
3. Ethnic identity: Ask a sample of third-generation Chinese Americans born in San Francisco if they speak their grandparents' language fluently. If "yes," score 1. If "no," score 0. Ask if they eat at non-Chinese restaurants at least once a week. Score 1 for "no," and 0 for "yes." Ask them eight other questions like this, and score 1 for each answer that signifies self-identification with their grandparents' heritage. Higher scores indicate greater ethnic identification.

These definitions sound pretty boring, but think about this: If we all use the same definitions for variables, *and if we stick to those definitions in making measurements*, then our data are strictly comparable:

1. We can test if doctors have higher intelligence scores than lawyers.
2. We can test if men from Bolivia have higher machismo scores than men from Paraguay.
3. We can test if Chinese Americans in San Francisco have higher cultural identity scores than Chinese Americans in New York City.

The ability to make such comparisons isn't boring to us. In fact, it sounds pretty exciting. But did you notice that we *never* said anything in those comparisons about ethnic identity per se, or intelligence per se, or machismo per se. In each case, all we said was that we could test if the *scores* were bigger or smaller. (**Further Reading:** operationism)

What's So Good about Operationism?

Operational definitions are *strictly limited to the content of the operations specified*. That's why we also didn't say anything about whether it was a good idea or a stupid one to make any of these measurements or comparisons. *If the content of an operational definition is bad, then so are all conclusions you draw from using it for measurement*.

This is *not* an argument against operationism in science. Just the opposite. Operationism is the best way to expose bad measurement. By defining measurements operationally, we can tell if one measurement is better than another. If the operational measurement of, say, machismo,

RUSS'S CORNER

IMPROVING MEASUREMENT

Adhering to bad measurements is bad science and can have some bad consequences for people. In the 1960s, I was a consultant on a project that was supposed to help Chicano high schoolers develop good career aspirations. Studies had been conducted in which Chicano and Anglo high schoolers were asked what they wanted to be when they reached 30 years of age. Chicanos expressed, on average, a lower occupational aspiration than did Anglos. This led some social scientists to advise policymakers that Chicano youth needed reinforcement of career aspirations at home. (There's that educational model again.)

Contrary to survey findings, ethnographic research showed that Chicano parents had very high aspirations for their children. The parents were frustrated by two things: (1) despair over the cost of sending their children to college; and (2) high-school counselors who systematically encouraged Chicana girls to become housewives and Chicano boys to learn a trade or go into the armed services.

The presumed relation between the dependent variable (level of career aspiration) and the independent variable (level of aspiration by parents for the careers of their children) was backwards. The parents' level of career aspiration for their children didn't cause the children to have low aspirations. The children were driven to low aspirations by structural features of their environment. The parents of those children reflected this reality in order—they said explicitly to interviewers who bothered to ask—not to give their children false hopes.

The operational definition of the variable "parents' career aspirations for their children" was useless. Here's the operational definition that should have been used in the study of Chicano parents' aspirations for their children's careers:

> Go to the homes of the respondents. Use the preferred language of the respondents (e.g., Spanish or English), and talk to parents about what they want their high school age children to be doing in 10 years. Explore each answer in depth and find out why parents give each answer.
>
> Ask specifically if the parents are telling you what they think their children *will* be doing or what they *want* their children to be doing. If parents hesitate, say: "Suppose nothing stood in the way of your [son] [daughter] becoming anything they wanted to be. What would you like them to be doing 10 years from now?"
>
> Write down what the parents say and code it for the following possible scores: 1 = unambivalently in favor of children going into high-status occupations; 2 = ambivalent about children going into high-status occupations; 3 = unambivalently in favor of children going into low- or middle-status occupations.
>
> Use the Nam-Powers-Boyd occupation scale (Nam and Boyd 2004) to decide whether the occupations selected by parents as fitting for their children are high, middle, or low status. Be sure to take and keep notes on what parents say are the reasons for their selections of occupations.

Notice in this example that taking a qualitative approach did not stop us from being operational.

seems silly or offensive, it may be because the concept is not very useful to begin with. No amount of measurement or operationism bails out bad concepts. The act of trying, though, usually *exposes* bad concepts and helps you jettison them.

The Problem with Operationism

Strict operationism creates a knotty philosophical problem. Measurement turns abstractions (concepts) into reality. Since there are many ways to measure the same abstraction, the reality of any concept hinges on the device you use to measure it.

So, intelligence is different if you measure it with a Stanford-Binet test or the Wechsler scales. The very concept of intelligence in the two cases is different because different "instruments" (queries are instruments) were used to measure it.

The bottom line on strict operational definitions is this: No matter how much you insist that intelligence is really more than what is measured by an intelligence test, that's all it can ever be. Whatever you think intelligence is, it is exactly and only what you measure with an intelligence test and nothing more.

If you don't like the results of your measurement, then build a better test, where better means that the outcomes are more useful in building theory, in making predictions, and in engineering behavior. Or if it's really unsalvageable, build a better concept.

We see no reason to waffle about this, or to look for philosophically palatable ways to soften the principle here. The science that emerges from a strict operational approach to understanding variables is much too powerful to water down with backpedaling.

Levels of Measurement

Whenever you define a variable operationally, you do so at some **level of measurement**. Most social scientists recognize four levels of measurement, in ascending order of amount of information in each: nominal, ordinal, interval, and ratio. The general principle in research is: *Always use the highest level of measurement that you can*. (This principle will be clear by the time you get through the next couple of pages.)

Nominal Variables

A variable is something that can take more than one value. The values of a **nominal variable** comprise a list of names. You can list religions, occupations, and ethnic groups; you can also list fruits, emotions, body parts, things to do on the weekend, baseball teams, rock stars . . . the list of things you can list is endless.

Think of nominal variables as *yes-no questions, the answers to which tell you nothing about degree or amount*. The following survey item is an operationalization of the nominal variable called "religious affiliation":

26a. Do you identify with any religion? (check one)

☐ Yes ☐ No

If you checked "yes," then please answer question 26b.

26b. What is your religion? (check **one**)

☐ Protestant

☐ Catholic

☐ Jewish

☐ Muslim

☐ Other religion

This operationalization of the variable religious affiliation has two important characteristics: It is *mutually exclusive* and *exhaustive*. The instruction to "check one" makes the list mutually exclusive. The famous "other" category in nominal variables makes the list exhaustive—that is, all possible categories have been named in the list. (More on this in Chapter 8 when we discuss questionnaire design.)

Mutually exclusive means that things can't belong to more than one category of a nominal variable at a time. We assume, for example, that people who say they are Catholic generally don't say they are Muslim. But some citizens of Lebanon have one Catholic and one Muslim parent and think of themselves as both. When it comes to some variables, like ethnicity, the requirement for mutual exclusivity is just hopeless.

You can assign numbers to nominal variables. Assign 0 if Catholic, 1 if Muslim, and so forth. But you can't add the numbers up in any meaningful way. Still, assigning numbers to things does make it easier to do certain kinds of statistical analysis on qualitative data—more on this in Chapters 11 (on coding text) and 12 (on regression).

Ordinal Variables

Like nominal-level variables, **ordinal variables** are generally exhaustive and mutually exclusive, but they have one additional property: Their values can be rank ordered. Any variable measured as high, medium, or low, like socioeconomic class, is ordinal. Scales of opinion—like the familiar "strongly agree," "agree," "neutral," "disagree," "strongly disagree" found on so many surveys—are ordinal measures. They measure an internal state, agreement, in terms of *less* and *more*.

What ordinal variables do not tell us is *how much* higher or lower one attribute of a variable is from any other. A person who is middle class might be twice as wealthy and three times as educated as a person who is lower class. Or they might be three times as wealthy and four times as educated. A person who "agrees strongly" with a statement may agree twice as much as someone who says they "agree"—or eight times as much, or half again as much. There is no way to tell.

Interval and Ratio Variables

Interval variables have all the properties of nominal and ordinal variables. They are an exhaustive and mutually exclusive list of attributes, and the attributes have a rank-order structure. They have one additional property, as well: The distances between the attributes are meaningful. Interval variables, then, involve true **quantitative measurement.**

Ratio variables are interval variables that have a true **zero point**—that is, a 0 that measures the absence of the phenomenon being measured. Age is one example. The 10-year intervals between people who are 10, 20, and 30 years old are identical. That much is true of any interval variable. In addition, however, age has a true zero. A person who is 20 is twice as old as a person who is 10; and a person who is 40 is twice as old as a person who is 20. This is how ratio variables work.

Most interval-level variables in the social sciences are also ratio variables. Some examples of ratio variables include: income in dollars, hours worked weekly, and population size. In fact, it has become common practice in the social sciences to refer to ratio-level variables as interval variables and vice versa. This is not technically pure, but the confusion of the terms "interval" and "ratio" doesn't cause much real damage.

A Rule: Always Measure at the Highest Level Possible

Remember this rule: *Always measure things at the highest level of measurement possible. Don't measure things at the ordinal level if you can measure them as ratio variables.*

If you really want to know the price that people paid for their homes, then ask the price. Don't ask them whether they paid "less than $400,000, between $400,000 and $600,000, or more than $600,000." This kind of question just throws away information by turning interval-level variables into ordinal ones.

If people won't give you straight answers to straight questions, you can back off and try an ordinal scale. (As we'll see in Chapter 8, survey questions are pretested before going into a questionnaire.) But why start out hobbling a perfectly good interval-scale question by making it ordinal when you don't know that you have to?

During data analysis, you can lump interval-level housing data together into ordinal (high, medium, low) or nominal (rent, buy) categories. But you cannot do this trick the other way around. If you collect ordinal data on house prices, you cannot go back and assign actual numbers of dollars to each respondent.

Notice that "less than $400,000" and "more than $400,000" is an ordinal variable that *looks like* a nominal variable because there are only two attributes. Ordinal variables can have any number of ranks. If the attributes are rankable, then the variable is ordinal. Unlike interval variables, however, ordinal data cannot be analyzed using a wide range of statistical analyses. More about this in Chapter 12 on bivariate analysis.

Validity, Reliability, Accuracy, and Precision

Now that we've covered how we do measurement in social science, we can consider how we evaluate if these measures are any good. **Validity** refers to the accuracy and trustworthiness of instruments, data, and findings in research. Nothing in research is more important than validity.

The Validity of Instruments and Data

Does the question "How long does it take you to drive to work each day?" have **instrument validity**? In other words, is it a valid instrument for measuring the amount of time it takes people to drive to work each day?

That depends on how accurate you want the data to be. If you want the data to be accurate to within, say, 20 minutes on, say 70% of occasions, then the instrument is probably valid. If you want 90–100% accuracy, then the instrument is almost certainly not valid. People just can't dredge up the information you want at that level of accuracy.

Is the question "Do you experience commuting spillover?" a valid instrument for measuring people's experiences at work (Calderwood and Mitropoulos 2021)? If people don't understand the question—if you're asking them to think in categories that are alien to their language or culture—then it's not a valid instrument for measuring anything.

Data validity is tied to instrument validity. If questions asking people to recall their behavior are not valid instruments for measuring that behavior, then the data retrieved by those instruments are not valid, either.

The Validity of Conclusions

Once we have valid instruments and valid data, we can ask about **conclusion validity**. Asian Americans (including Chinese Americans, Japanese Americans, and Vietnamese Americans) get higher average scores on the math part of the Scholastic Aptitude Tests (SAT) than do other ethnic groups in the United States—626 versus 508 for all ethnic groups combined in 2023 (College Board 2023).

Suppose that the SAT math test is a valid instrument for measuring the general math ability of 18 year-olds in the United States. Is it valid to conclude that "Asians are better at math" than other people are? No, it isn't. Such a conclusion is likely reached by invoking an unfounded, racist assumption about the influence of certain genes on the ability of people to do math. The measurement may be valid, but the conclusion certainly isn't.

Reliability

Reliability refers to whether or not you get the same answer by using an instrument to measure something more than once. If you insert a thermometer into boiling water at sea level, it should register 212° Fahrenheit each and every time. Instruments can be things like thermometers and scales, or they can be questions that you ask people.

Like all other kinds of instruments, some questions are more reliable for retrieving information than others. "How many brothers and sisters do you have?" is a pretty reliable instrument—you almost always get the same response when you ask a person that question a second time as you get the first time. "How much is your parents' house worth?" is much less reliable. And "How old were you when you were toilet trained?" is just hopeless.

We'll take up the measurement of interrater reliability—where more than one observer records a measurement—in Chapter 11.

Precision

Precision is about the number of decimal points in a measurement. Digital bathroom scales typically have read-outs in half pounds. So, if your true weight, to the nearest thousandth of a pound is 156.625 pounds, a typical scale might say that you weigh 156.5 pounds or 157 pounds. In this case, you might not be too concerned about the error introduced by lack of precision.

Whether you care or not depends on the needs you have for the data. If you are just monitoring your weight, you're probably not going to worry too much about the fact that your scale is only precise to the nearest half a pound. But if you're measuring the weights of pharmaceuticals, and someone's life depends on your getting the precise amounts into a compound, that's another matter.

Accuracy

Finally, **accuracy**. Assume that you are satisfied with the level of precision of the scale, but there was an error at the factory where the scale were built and the scales were off?

Now we have the following interesting situation: The data from this instrument are valid (it has already been determined that the scale is measuring weight—exactly what you think it's measuring); they are reliable (you get the same answer every time you step on it); and they are precise enough for your purposes. But they are not *accurate*. What next?

You could see if the scale were always inaccurate in the same way. Suppose it turned out that your scale was always incorrectly lower by 5 pounds. This is called **systematic bias**.

If an instrument is sufficiently precise and reliable, but inaccurate *in known ways*, then a formula can be applied to correct for the inaccuracy. In this case, a simple +5 pounds correction formula is all you'd need to feel confident about the data.

But the scale might be off in more complicated ways that are harder to correct. If an instrument is not accurate enough for what you want to do with the data, then you simply have to build a more accurate one. There is no way out.

The bad news is that instruments are often inaccurate in unknown ways. When your instrument is a question—asking how long it takes to drive to work, for example, or what was for dinner last night—people will give you an answer. The answers may be dead on target, or they may bear no useful resemblance to the truth.

The good news is that scientists have been studying respondent accuracy for years. The methodological literature has lots of guidance on how to ask questions to elicit more accurate answers. And real progress continues to be made. (**Further Reading:** respondent accuracy)

Tests for Reliability

There are several tests of reliability:

1. **Interobserver reliability.** Suppose you set up an experiment to see whether 5-year-old children act more aggressively or more cooperatively in same-sex play groups or in mixed-sex play groups. You'll have several observers code the behavior of the children and you'll want them to achieve consistency in what they see and write down.

Or suppose you have a set of open-ended interviews with childcare providers about what it's like to supervise same-sex and mixed-sex play groups. You'll want several people coding those interviews. In both cases—whether you're coding behavior or the content of text—you'll want your coders to achieve a high **interobserver reliability** score. We'll have more to say in Chapter 11 (on text analysis) about how to calculate that score, which is called inter-rater reliability (or inter-coder reliability) in text analysis.

2. **Test–retest reliability.** A reliable test of, say, ability in math should give you more or less the same results each time you use it on the same person. This is known as **test–retest reliability**. When tests are developed, they are typically tested for reliability by giving them to a group of people then calling back those same people a week later to take the test again.

Many standardized tests have two parallel test forms. When both forms are given to the same person, they should produce more or less the same results. "More or less" here means at least 0.80.

Is My Measure Any Good? Determining Validity

You may have noticed a few paragraphs back that we casually slipped in that our weight scale had *already been determined* to be a valid instrument. How do we know that a scale is measuring weight? Maybe it's measuring something else. How can we be sure? We have to make concepts up in order to study them, so there is no direct way to evaluate the validity of an instrument for measuring a concept. Ultimately, we are left to decide, on the basis of our best judgment, whether an instrument is valid or not.

We are helped in making that judgment by some tests for face validity, content validity, construct validity, and criterion validity.

Face Validity

Establishing **face validity** involves simply looking at the operational indicators of a concept and deciding whether or not, *on the face of it*, the indicators make sense. The indicators might be items on an opinion survey, or they might be tests of knowledge and ability.

On the face of it, a paper-and-pencil test about the rules of the road is not a valid indicator of whether someone can actually drive. But it is probably a valid test for determining if a driver's license applicant can read road signs. Different instruments—a real road test and a paper-and-pencil test on rules of the road—have face validity for measuring different things.

Face validity is based on consensus among researchers: If everyone agrees that asking people, "Can you drive?" is a valid instrument, then, until proven otherwise, that question is a valid instrument for measuring driving ability.

Content Validity

Content validity is achieved when an instrument has appropriate content for measuring a complex concept, or construct. Achievement tests—for assessing whether pilots are ready to fly solo

and whether family therapists are ready to be licensed—are judged on their content validity. If you walk out of a test and feel that it was unfair because it tapped too narrow a band of knowledge, your complaint is that the test lacked content validity.

A test to measure the strength of "ethnic identity" among, say, second-generation Mexican Americans has to have content that deals with religion, language, political and economic values, sense of history, and gastronomy. A valid measure of ethnic identity, then, has to get at *all* these areas. People's use of Spanish and their preference for certain foods are good measures of *some* of the content of Mexican American ethnicity. But if these are the only questions you ask, then your measure of ethnicity has low content validity.

Most of the really interesting things that social scientists study are complex constructs—things like "quality of life," "socioeconomic class," and so on. Content validity is very, very tough to achieve, particularly for complex, multidimensional constructs.

Construct Validity

An instrument has high **construct validity** if there is a close fit between the construct it supposedly measures and actual observations made with the instrument. An instrument has high construct validity, in other words, if it allows you to infer that a unit of analysis (a person, a country, whatever) has a particular trait and if it supports predictions that are made from theory.

Asking people "How old are you?" has so much face validity that you hardly need to ask whether the instrument has construct validity—whether it gets at the construct of chronological age. Giving people an IQ test, by contrast, is controversial because there is so much disagreement about what the construct of intelligence is.

Lots of constructs in which we're interested—ethnicity, self-esteem, gender, community resilience, food security, alienation, acculturation, political orientation (liberal vs. conservative)—are controversial and so are the measures for them. Getting people to agree that a particular *measure* has high construct validity requires that they agree that the construct is valid in the first place.

Criterion Validity

An instrument has high **criterion validity** if there is a close fit between the measures it produces and the measures produced by some other instrument that is known to be valid. This known measure test of criterion validity is the **gold standard**, the best we have. But for most complex social science variables, we actually don't have a gold standard measure.

TIPS & TRICKS
THE PRINCIPLE OF OCKHAM'S RAZOR

The preference in science for simpler explanations and measures over more complicated ones is called the principle of **parsimony**. It is also known as **Ockham's razor**, after William of Ockham (1285–1349), a medieval philosopher who argued *entia non sunt multiplicanda praeter necessitatem*, or "Don't make things more complicated than they need to be."

AMBER'S CORNER

USING KNOWN GROUP COMPARISONS TO TEST A NOVEL WATER INSECURITY SCALE

I developed a new operationalization for *household water insecurity* for use in informal settlements where mostly Quechua and Aymara migrants lived in Cochabamba, Bolivia. It had never been measured before, so it was impossible to conduct a gold standard test of criterion validity against a known measure.

But it was well documented in the literature that the burdens of household water insecurity fell disproportionately on women. I conducted a known-group comparison of women and men's scores on my new household water insecurity scale (Hadley and Wutich 2009). As predicted, women scored higher than men. This, along with seasonal tests that confirmed water insecurity scores were higher in the dry season than the wet season, was good evidence of criterion validity.

Interestingly, though, we've since found that men are more sensitive to household water insecurity than we once thought (Nébié et al. 2023; Wutich 2009b, Ilboudo Nébié et al. 2023). That finding—along re-conceptualizations and new operationalizations of gender that include transgender and gender-diverse populations (DuBois and Shattuck-Heidorn 2021)—brings into question my approach to criterion validity. If I had to conduct a known-group comparison of household water insecurity scales today, I wouldn't be as confident as I once was in running a comparison of men versus women.

You can tap the power of criterion validity for complex constructs with the **known group comparison** technique. If you develop a scale to measure political ideology in the United States, you could try it out on Democratic Socialists of America (DSA) members and on Conservative Political Action Coalition (CPAC) members. DSA members should get high "left" scores, and CPAC members should score high "right" scores. If they don't, then there's probably something wrong with the scale. In other words, the known-group scores are the criteria for the validity of your instrument.

A particularly strong form of criterion validity is **predictive validity**—whether an instrument lets you predict accurately something else you're interested in. "Stress" is a complex construct. It occurs when people interpret events as threatening to their lives. Some people interpret a bad grade on an exam as a threat to their whole life, while others just blow it off. Now, stress is widely thought to produce a lowered immune response and increase the chances of getting sick. A good *measure* of stress, then, ought to predict the likelihood of getting sick.

The Problem with Validity

If you suspect that there is something deeply, desperately wrong with all this, you're right. The whole argument for the validity (indeed, the very existence) of something like

intelligence is, frankly, circular: How do you know that intelligence exists? Because you see its effects in achievement. And how do you account for achievement? By saying that someone has achieved highly because they're intelligent. And how do you know machismo exists? Because men dominate women in some societies. And how do you account for dominance behavior, like men's violence against women? By saying men's violence is acting out their machismo.

In the hierarchy of construct reality, physicists' work on force ranks way up there—after all, it's got several hundred years of theory and experimentation behind it. Things like intelligence and machismo are pretty weak by comparison. And yet, as we made clear at the beginning of this chapter, the social and behavioral sciences are roaring successes, on a par with the physical sciences in terms of the effects they have on our lives every day.

Just think of all the ordinary things in our lives that are driven by social science data, methods, and analysis: Polling in the conduct of elections; choosing the site for a new school or hospital or green space; the building and roll-out of social media sites; choosing where to advertise any product and how much to spend on that advertising; building stock portfolios for retirement; deciding how much to charge in premiums for insurance; deciding where to concentrate vaccination programs; managing supply chains—all of this and much more is possible because social scientists have refined and tested many useful concepts and measurements for those concepts.

The Bottom Line

The bottom line on all this is that although various forms of validity can be demonstrated, Truth, with a capital T, is never final. We are never dead sure of anything in science. We try to get closer and closer to the truth by better and better measurement.

All of science relies on concepts whose existence must ultimately be demonstrated by their effects. You can ram a car against a cement wall at 50 miles an hour and account for the amount of crumpling done to the radiator by referring to a concept called "force." You can't see force, but you can sure see its effects. The greater the force, the more crumpled the radiator. (**Further Reading:** validity)

Ultimately, the validity of any concept—force in physics, the self in psychology, modernization in sociology and political science, culture in anthropology—depends on two things: (1) the utility of the device that measures it and (2) the collective judgment of the scientific community that a concept and its measure are valid. In the end, we are left to deal with the effects of our judgments, which is just as it should be. Valid measurement makes valid data, but validity itself depends on the collective opinion of researchers.

KEY CONCEPTS IN THIS CHAPTER

accuracy
concept
conceptual definitions
conclusion validity
construct
construct validity

content validity
criterion validity
data validity
dependent variable
educational model of social change
face validity
gold standard
independent variable
indicators
instrument validity
interobserver reliability
interval variables
known group comparison
level of measurement
measurement
multidimensional
nominal variable
Ockham's razor
operationalize
operational definitions
ordinal variables
parsimony
precision
predictive validity
qualitative measurement
quantitative measurement
ratio variables
reliability
systematic bias
test–retest reliability
unidimensional
validity
value
variable
zero point

SUMMARY

- Social research is about variables—that is, about characteristics of people, countries, organizations, or other units of analysis and how variables are related to one another. A controversial foundation of modern social research is operationism, which involves making absolutely explicit how variables are measured.
- The main advantage of operationism is that researchers can replicate one another's work and build cumulative knowledge.
    - The main disadvantage of operationism is that it forces us to measure complex variables—like compassion, religiosity, political orientation, and the like—using simple tools. There is, then, the risk of trivializing the process of research.
    - On balance, more social scientists rely on operationism than criticize it, but this varies across the social science disciplines.
- Measurement in the social sciences can be at the nominal, ordinal, or interval/ratio level.
    - Ratio-level variables have a true zero point.
    - The rule is always to measure at the highest level of measurement possible. You can turn a variable measured at the ratio level into an ordinal or a nominal variable, but you can't go the other way.
- In developing measures for variables, researchers are concerned with the problems of reliability, validity, precision, and accuracy.

- Reliability is a necessary but insufficient condition for validity.
- Validity is never proven absolutely, but is a goal toward which we strive.

EXERCISES

1. (a) Here is a list of adjectives, each of which is a concept that is of interest to social researchers. Try to conceptualize and operationalize these concepts that describe characteristics of individual people: poor, religious, macho, affluent, abusive. (b) Suppose we want to array the countries of the world according to how much economic freedom and how much political freedom their citizens have. How can we conceptualize and operationalize economic and political freedom?

2. Using concrete examples, explain the characteristics of nominal, ordinal, and interval/ratio measurements. How would you measure each of the following, using at least two different levels of measurement? (a) age; (b) income; (c) family size.

3. Ask 20 people a question that requires a self-report of behavior. For example, ask "How many times during the last month have you skipped class?" Then, ask the same people "How many times during the last week have you skipped class?" At the end of the interview, ask people to explain how they figured out what to say in answering your question. Did people think about and count up the actual incidents, or did they estimate the number of incidents? If they estimated, then ask them how they did that and try to understand the rules of inference they used.

Calculate the per-week average for the one-month question and the average for the one-week question. Are the two averages the same?

4. Discuss the difference between validity and reliability. Why is it so hard to establish validity?

5. Ask some people if they consider themselves to be politically liberal or politically conservative. Some people will find the question unanswerable, but some people will answer the question. Repeat this until you have data from 20 people. This question will create data at the nominal level of measurement. Next, ask those same 20 people who answered your first question whether they consider themselves very liberal (or very conservative), somewhat liberal (conservative), or mildly liberal (conservative). This will create some data at the ordinal level of measurement.

6. Finally, ask those same 20 people a series of focused issue questions. Here are a few examples: (a) Are you in favor of an adult woman's right to an abortion, entirely at her own discretion, or are you opposed to women having that right? (b) Are you in favor of the death penalty for premeditated murder or are you opposed to the death penalty for that crime? (c) Are you in favor of sending our troops to [fill in whatever country is in the news at the moment] to fight a ground war or are you opposed to sending troops?

These questions, or questions like them, will create data at the nominal level of measurement. Keep careful track of the reactions of your respondents and write up your findings. Among other things, you may find that people who say they are very conservative or very liberal fail to answer specific issue questions in way you might expect. Why is that?

FURTHER READING

Ethics and social science Bosk (2004), Boruch and Cecil (1983), Bosk (2004), Burgess (1989), Citro et al. (2003), Fluehr-Lobban (1996), Hammersley (2009), Herrera (1996), Keith-Spiegel and Koocher (2005), Lyman (1989), Mertens and Ginsberg (2009), van den Hoonaard (2002), Weisstub (1998), Weisstub and Diaz Pintos (2007).

Measurement in the social sciences. Blalock (1974), Coombs (1964), Lester and Bishop (1997), D. C. Miller and Salkind (2002), Nunnally (1978), Stevens (1946).

Operationism. Bridgman (1927), Lundberg (1942, 1964), For a critique: Lincoln and Guba (1985).

Respondent and informant accuracy. Bernard, Killworth et al. (1984), Brewer (2000), Brewer and Yang (1994), Freeman et al. (1987), Frenk et al. (2011), Godoy et al. (1998), Jackson and Nuttall (1997), Poulin (2010), Ricci et al. (1995), Roberts et al. (2005), Romney et al. (1986), **Tanur (1992)**, Schwarz (1999), Sudman et al. (1996), Vadez et al. (2003), **Wutich (2009)**. **Validity**. Campbell and Fiske (1959), Fiske (1982).

Chapter 2

Preparing for Research

WHAT'S IN THIS CHAPTER

Setting Things Up

This chapter is about some of the things that go on *before* data are collected and analyzed. First, we'll take you through the ideal research process and compare that to how research really gets done. Next, we'll discuss the problem of choosing research problems—how do I know what to study?—and we'll give you some pointers on how to scour the literature so you can benefit from the work of others when you start a research project. Finally, we'll discuss ethical issues in social research, including how to work with your local Institutional Review Board to gain permission to conduct your research.

The Ideal Research Process

Despite all the myths about how research is done, the **research process** is actually a messy process that's cleaned up in the reporting of results. Figure 2.1 shows how the research process is supposed to work in the ideal world:

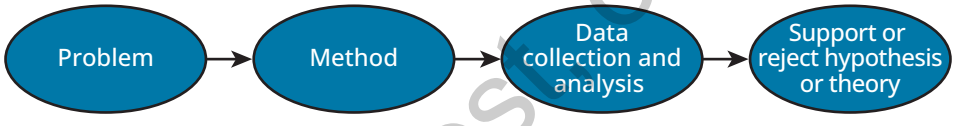
1. First, a theoretical problem is formulated.
2. Next, an appropriate site and method are selected.

TIPS & TRICKS
RESEARCH PROBLEMS AND SOCIAL PROBLEMS

As you think about choosing a project, be sure to distinguish between research problems and social problems. The fact that women earn less than men do in many comparable jobs is a social problem but, being a social problem doesn't necessarily make it a research problem. A **research problem** can tell us something new about how the world works. Social problems, though, often have causes that are well understood but difficult to solve.

Gender inequality in pay is a social problem that *can* be the basis of many research problems. For example, women earn more in some jobs than men and gender-diverse people do. What characteristics of individuals (like their gender, age, education, etc.) and the jobs they occupy (indoor versus outdoor, for example, or the amount of training in technology required) account for differences in the pay that women get for doing the same job as men and gender-diverse people?

FIGURE 2.1 ■ How Research is Supposed to Work



3. Then, data are collected and analyzed.
4. Finally, the theoretical proposition with which the research was launched is either challenged or supported.

In fact, all kinds of practical and intellectual issues get in the way of this neat scheme. In the end, research papers are written so that the chaotic aspects of research are not emphasized, and the orderly inputs and outcomes are.

There's nothing wrong with this: It would be a monumental waste of precious space in books and journals to describe the *real* research process for every project that's reported. Besides, every seasoned researcher knows just how messy it all is, anyway. On the other hand, you shouldn't have to become a highly experienced researcher before you're let into the secret of how it's really done.

A Realistic Approach

There are five questions to ask yourself about every research question you are thinking about pursuing. Most of these can also be asked about potential research sites and research methods. If you answer these questions honestly (at least to yourself), chances are you'll do good research every time. If you cheat on this test, even a teeny bit, chances are you'll regret it. The questions are:

1. Does this topic (or research site, or data collection method) really interest me?
2. Is this a problem that is amenable to scientific inquiry?
3. Are adequate resources available to investigate this topic? To study this population at this particular research site? To use this particular data collection method?
4. Will my research question, or the methods I want to use, lead to unresolvable ethical problems?
5. Is the topic of theoretical and/or practical interest?

Personal Interest

The first thing to ask about any potential research question is: Am I really excited about this? Researchers do their best work when they are genuinely having fun, so don't do boring research when you can choose any topic you like.

We've seen many students doing research for term projects, M.A. theses, and even doctoral dissertations simply out of convenience and with no enthusiasm for the topic. If you are not interested in a research question, then no matter how important other people tell you it is, don't bother with it. If others are so sure that it's a dynamite topic of great theoretical significance, let *them* study it.

The same goes for research populations. If you select a topic of interest, and then try to test it on a population in which you have no interest, your research will probably suffer. Enthusiasm counts for a lot in research. Federal prisons and Wall Street banking firms are both complex organizations. But they are very, very different kinds of places to spend time in, so if you are going to study a complex organization, check your gut first and make sure you're excited about

RUSS'S CORNER

YOU CAN'T ALWAYS CHOOSE TOPICS YOU LIKE

Of course, you can't always choose any topic you like. In contract research, you sometimes have to take on a research question that a client finds interesting but that you find deadly dull. The most boring research I've ever done was on a contract where my co-workers and I combined ethnographic and survey research of rural homeowners' knowledge of fire prevention and their attitudes toward volunteer fire departments. This was in 1973. I had young children at home and the research contract paid me a summer salary. It was honest work and I delivered a good product to the agency that supported the project. But I never wrote up the results for publication.

By comparison, that same year I did some contract research for the U.S. Bureau of Prisons on the effects of co-ed prisons on same-sex intimacy among male and female inmates. I was very interested in that study and it was much easier to spend the extra time and effort polishing the contract reports for publication (Killworth and Bernard 1974).

where you're going. It's really hard to conduct penetrating, in-depth interviews over a period of several weeks to a year if you aren't interested in the lives of the people you're studying.

And if you think it's tough to run personal interviews on a topic you're bored with, try making up and administering a 10-page questionnaire on a topic of no interest to you. Or try designing an experiment in which you have to run subjects for months on end and where you have no personal stake in the results. It's not just deadly dull, it's a recipe for bad research.

You don't need any justification for your interest in studying a particular group of people or a particular topic. Personal interest is . . . well, personal. So, ask yourself: Will my interest be sustained there? If the answer is "no," then reconsider. Accessibility of a research site or the availability of funds for the conduct of a survey are pluses, but by themselves they're not enough to make good research happen.

Science Versus Nonscience

If you're really excited about a research topic, then the next question is: Is this a topic that can be studied by the methods of **science**? If the answer is "no," then no matter how much fun it is, and no matter how important it seems, don't even try to make a scientific study of it. Either let someone else do it or use a different approach.

Consider this empirical question: How often do derogatory references to women occur in the Old Testament? If you can come up with a good, operational definition of "derogatory," then you can answer this question by looking through the corpus of data and counting the instances that turn up. Pretty straightforward, descriptive science.

But consider this question: Is the death penalty morally correct? This is simply not answerable by the scientific method. It is no more answerable than the question: Is Britain's parliamentary system or the U.S.'s presidential system a better form of democracy? You can shift these questions into scientifically testable ones—for example, by specifying a metric, and asking under which condition that metric performs better—but questions about what's morally correct or generally better just aren't scientifically testable.

Whether or not a study is a scientific one depends first on the nature of the question being asked and *then* on the methods used.

And when we're talking about using the scientific method we're *not* talking about numbers. In science, whenever a research problem can be investigated with valid quantitative measurement, numbers are more than just desirable, they're required. On the other hand, there are many intellectual problems for which quantitative measures are not yet available. Those problems require qualitative measurement.

First-pass descriptions of processes (e.g., preparing for surgery), or of events (e.g., weddings), or of systems of nomenclature (e.g., kinds of trucks) require words, not numbers. Dorothy Holland and Debra Skinner (1987) asked some university women to list the kinds of guys there are. They got a list of words like "creep," "hunk," "nerd," "jerk," "sweetie pie," and so on. Then they asked some women, for each kind: "Is this someone you'd like to date?" The yes–no answers are nominal—that is, qualitative—measurement.

We'll get back to this kind of systematic collection of qualitative data in Chapter 11.

Resources

The next question to ask is whether adequate resources are available for you to conduct your study. There are three major kinds of resources: time, money, and people. What may be adequate for some projects may be inadequate for others. Be totally honest with yourself about this issue.

Time

Some social research projects can be completed in just a few days, while others take years. Kirsten Hunt-Howard spent two years studying drug users and dealers on the streets of Los Angeles. It took her that long (and 38 in-depth interviews) to understand what a typical day was like for sellers and users of methamphetamines and other drugs, and to understand how "drug consumption and distribution rituals facilitated social bonding within friendship networks" (Hunt-Howard 2017:12). A lot of focused ethnography, however, can be done much more quickly.

Some experiments in social psychology can also take months or years to set up, especially in evaluation studies. By contrast, the data-collection phase of surveys and of some types of experiments might be completed in a matter of weeks.

If you are doing research for a term project, the topic has to be something you can look at in a matter of a few months—and squeezing the research into a schedule of other classes, at that. It makes no sense to select a topic that requires two semesters' work when you have one semester in which to do the research. This effort to cram 10 gallons of water into a 5-gallon can is futile and quite common. Don't do it.

Money

Naturally, most researchers do not have the money it takes to mount a major research effort. That's why there are granting agencies. Writing proposals is a special craft. It pays to learn it early.

Grants for Master's-level research in the social sciences are typically between $1,000 and $5,000. Grants for doctoral research are typically between $10,000 and $40,000. If you spend 100 hours working on a grant proposal that brings you $10,000 to do your research, that's $100/hr for your time. If you get turned down and spend another 100 hours rewriting the proposal, that's still $50 an hour for your time if you're successful. Pretty good pay for interesting work.

Collaborative grants are a great way to secure research funding. Research in engineering and medicine, for example, is funded at 10 times the rate of social science, or more. Building your research into a sub-study on a larger funded grant could be the best way to secure the research funds you need.

Consider, too, if cheaper methods might work just as well. Ask yourself whether it's worthwhile pursuing your research if it has to be scaled down to fit available resources. If the answer is "no," then find another topic. No matter how interesting it is to you, and no matter how important it may seem theoretically, if you haven't got the resources to use the right methods, skip it for now.

People

"People" includes you and others involved in the research, as well as those whom you are studying. Does the research require that you do logistic regression? If it does, then do you have—or are you prepared to acquire—that skill? Does the research require access to or acceptance by a particular group of people, like ambulance paramedics? Do you have access to that group?

Does the research require that you speak Haitian Creole? If so, are you willing to put in the time and effort to learn that language? If the research can be done with interpreters, are skilled workers available at a cost that you can handle?

Will the research require that you interview elite members of the society you are studying—like medical malpractice lawyers, plastic surgeons, Lutheran priests? Do you have access to these populations? Will you be able to gain their cooperation? Or will they tell you to get lost or, even worse, provide you with perfunctory answers to your questions. Better not do the study in the first place than wind up with useless data.

Ethics of Social Research

We wish we could give you a list of criteria against which you could measure the "ethicalness" of every research idea you'll ever come up with. Unfortunately, it's not so simple. The fact is, what is popularly **ethical research** today may become popularly unethical tomorrow, and vice versa and what's ethical in one culture may be unethical in another. (This does *not* mean that all ethics are relative. But more on that later.)

For centuries, slavery was legal and considered ethical in many societies. Today, while human trafficking hasn't ended, it is outlawed everywhere and almost universally considered unethical. Physician- assisted suicide was once widely considered unethical but is becoming ethically acceptable to many doctors. During World War II, many social scientists in the United States worked for what would today be called the Department of Defense and they were applauded as patriots for lending their expertise to the war effort.

Twenty-five years later, during the Vietnam War, social scientists in the United States, who worked for the Department of Defense were excoriated. From 2007 to 2014, social scientists participated in U.S. military programs for studying local culture in the battle zones of Iraq and Afghanistan (McFate and Laurence 2015). This, too, produced intense debate about the proper role, if any, of social scientists in military and intelligence operations (Fluehr-Lobban 2008; Forte 2011). (**Further Reading:** social science in the military and in intelligence)

Milgram's Obedience Experiment

It's because popular ethics change that Stanley Milgram was able to conduct his famous experiment on obedience in the 1960s. Milgram (1963, 1965) told people that they were taking part in an experiment on how well human beings learn under conditions of punishment. Some participants in the study would be teachers and others would be learners. At least, that's what they were

told. The real idea was to see how obedient people would be—how much electrical shock they would administer to a "learner"—all of whom were Milgram's accomplices—when told to do so by someone in authority, like an experimenter in a lab coat.

Milgram varied the conditions in his experiments to test for gender differences, for differences in locale (the Yale University campus vs. a run-down building in downtown Bridgeport, Connecticut), and for differences in the proximity of the experimenter and the victim to the subject (in the same room, in different rooms), but in all the experiments, the basics were the same. The subjects sat at a panel of 30 switches, labeled from 15 volts to 450 volts. There was a label every fourth switch (that is, every 60 volts), from "Slight Shock" (15 volts) all the way up to "Danger: Severe Shock" (375 volts) and "XXX" (435 volts and 450 volts). Each time the "learner"—remember, one of Milgram's accomplices—made a mistake on a word-recall test, the subject was told to give the learner a bigger shock.

Milgram paid each participant $4.50 up front (about $50 in 2025), to make them feel obligated to go through with the experiment. He also gave them a little test shock—45 volts (the second lever on the 30-lever panel)—to make them believe that the punishment they'd be delivering to the so-called learners was for real.

In many of the experiments, the learner grunted at 75 volts. The reaction escalated as the putative voltage increased. At 150 volts, learners began pleading to be let out of the experiment. At 285 volts, the learner's response, as Milgram reported it, could "only be described as an agonizing scream" (1974:4). All those reactions by the learners were actually played back from tape so that subjects would hear exactly the same things. The experimenter, in an official-looking lab coat, kept telling the subject to administer the shocks—saying things like: "You have no choice. You must go on."

Most of those who dropped out of Milgram's experiments did so after administering 150-volt shocks. But 65% of the subjects in the original experiment (where, by the way, the learner didn't complain until the subject supposedly gave him a 300-volt shock and then pounded on the wall of the room separating him from the subject), obeyed orders and administered what they thought were shocks beyond the XXX level. Many subjects protested but were convinced by the researchers that it was all right to follow orders.

Milgram's full experiment wouldn't get by any committee for the protection of human subjects now, but Jerry Burger (2009) was able to replicate Milgram's original experiment up to the crucial 150-volt limit. The bottom line: 28 of Burger's 40 subjects agreed to continue after the 150-volt limit.

Were Milgram's experiments unethical? Did Milgram cause his subjects emotional harm when they thought about what they'd done? If you were among Milgram's subjects who obeyed to the end, would you be haunted by this? The literature on this is mixed (see Murray [1980] and Herrera [2001] for contrasting views), but we do know this: Milgram's make-believe experiment was less costly and more ethical than the natural experiments carried out at My Lai, and Shatila, and Srebenica—the Vietnamese village (in 1968), the Lebanese refugee camp (in 1982), and the Bosnian village (in 1995)—whose civilian inhabitants were wiped out by American, Lebanese, and Serbian soldiers, respectively, "under orders."

Those experiments, too, showed what ordinary people are capable of doing—except that in those cases, real people really got killed. Until Milgram did his experiments, it had been easy to

scoff at Nazi war criminals whose defense was that they were "just following orders." In 1979, Milgram was asked on CBS's show, *Sixty Minutes*, if that sort of thing could happen again. His answer: "Having observed a thousand people in the experiment . . . if a system of death camps were set up in the United States of the sort we had seen in Nazi Germany, one would be able to find sufficient personnel for those camps in any medium-sized American town" (quoted in Blass 1999:955). (**Further Reading:** Milgram and obedience experiments)

What Does It All Mean?

Just because times, and ethics, seem to change, does not mean that anything goes. Everyone agrees that scholars have ethical responsibilities, but not everyone agrees on what those responsibilities are. All the major scholarly societies have published their own **code of ethics**—all variations on the same theme, but all variations, nonetheless. We've listed the Internet addresses for several of these codes of ethics in Appendix E.

These documents are not perfect, but they cover a lot of ground, and are based on the accumulated experience of thousands of researchers who have grappled with ethical dilemmas over the past 70 years. Look at those codes of ethics regularly during the course of any research project, both to get some of the wisdom that has gone into them and to develop your own ideas about how the documents might be improved.

Don't get trapped into nihilistic relativism. **Cultural relativism** (the unassailable fact that people's ideas about what is good and beautiful are shaped by their culture) is a great antidote for overdeveloped ethnocentrism. **Moral relativism** (that all moral and ethical systems are equally good since they are all cultural products) is something else entirely (Jarvie 1983; Salmon 1997).

Can you imagine defending the human rights violations of Nazi Germany as just another expression of the richness of culture? Would you feel comfortable defending, on the basis of cultural relativism, the so-called ethnic cleansing in the 1990s of Bosnians and Kosovar Albanians by Serbs in the former Yugoslavia? Or the slaughter of Tutsis by Hutus in 1994 in Rwanda? Or of American Indians by immigrant Europeans in the 19th century? We can't.

There is no value-free science. Everything that interests you as a potential research focus comes fully equipped with risks to you and to the people you study. Should social scientists do social marketing for a state lottery, knowing that poor people will be squandering their meager resources on false hopes of sudden riches? Or is social marketing only for getting people to use condoms and to wash their hands before preparing food?

How about working on projects that raise worker productivity in poverty zones if that means some workers will become redundant and lose their jobs? In each case, all you can do (and *must* do) is assess the potential human costs and the potential benefits. And when we say "potential benefits," we mean to you, personally, not just to humanity through the accumulation of knowledge.

Don't hide from the fact that you are interested in your own glory, your own career, your own advancement. It's a safe bet that your colleagues are interested in their career advancement, too. We have all heard of cases in which a scientist put his or her own career aggrandizement above the health and well-being of others. This is devastating to science, and to scientists, but

it happens when otherwise good, ethical people (1) convince themselves that they are doing something noble for humanity, rather than for themselves; and (2) consequently fool themselves into thinking that *that* justifies their hurting others. (See Judson [2004] for more on fraud in science.)

When you make these assessments of costs and benefits, be prepared to come to decisions that may not be shared by all your colleagues. Would you participate in studies of the relation between darkness of skin color and various measures of life success (including wealth, health, and longevity)? Suppose the study was likely to show that a small, but significant percentage of the variation in earning power in the United States was predictable from (*not* caused by) darkness of skin color. Some would argue that this would be useful evidence in the fight against racism and would jump at the chance to do the investigation. Others would argue that the evidence would be used by racists to do further damage in our society, so the study should simply not be done lest the information it produces fall into the wrong hands.

There is no answer to this dilemma. Above all, be honest with yourself. Ask yourself: Is this ethical? If the answer is "no," then skip it; find another topic. Once again, there are plenty of interesting research questions that won't put you into a moral bind. (**Further Reading:** ethical issues in social science)

Research and Institutional Review Boards

The key ethical issue in the conduct of all social research is whether those being studied are placed at risk by those doing the studying. This goes for field research—including surveys, ethnographies, and naturalistic experiments—as much as it does for laboratory studies. All universities in the United States have long had **Institutional Review Boards**, or IRBs. These are internal agencies whose members review and pass judgment on the ethical issues associated with all research on people, including biomedical and psychosocial.

With regard to the protection of human subjects, most social research in the United States is covered by the Code of Federal Regulations, Title 45, Part 46, from the Department of Health and Human Services. (See http://www.nsf.gov/bfa/dias/policy/human.jsp and https://grants.nih.gov/policy/humansubjects.htm) This set of regulations has been adopted by many agencies, including the National Science Foundation and the National Institutes of Health and is known as the Common Rule. Many American Indian tribes have their own IRBs; research with each tribe's members or on its own land is governed by its IRBs.

Here are some tips for getting your IRB application approved:

1. Learn the rules and learn them thoroughly. You'll have to complete trainings, like earning a Collaborative Institutional Training Initiative (CITI) certificate in responsible research conduct. As a student, you'll need a professor to be listed as the Principal Investigator (PI) on your project. The PI promises to be responsible for the research team's ethical research conduct.

2. If you are using unstructured exploratory methods like participant-observation, explain exactly what you are going to do. Exploratory methods require planning, just like experiments or archival research or survey research. Explain how you will recruit

respondents, how and where you will collect data, etc. Getting down the details will come in handy later when you write up your publications.

3. Get your IRB protocol in long, long before you plan to start your research. Don't make a lack of planning on your part anyone else's emergency. Most funding agencies can't release money to your institution, even if funding is approved, until you have the IRB approval. Getting a proposal for master's or doctoral-level funding six months before you want to do your work is not a burden. It's just part of time management.

4. If you are working overseas and using a language other than English, you'll be required to translate and back-translate your informed consent and data collection protocols—into French or Swahili or Urdu or whatever language you'll be using in the field. That means at least two people will need to work on the translations. It's easiest if you and a colleague can swap translation on proposals, but also consider that you may need to hire a translator.

5. Many countries and sovereign tribal nations require ethics clearances and research permissions, in addition to whatever ethical approval is required at your university. This can take time—sometimes years even—so it's important to plan ahead and get advice. Working with professors in your research community can help a lot. Local professors can explain the rules—and help you make sure the research benefits the research community.

6. The IRB can review minor changes in a previously approved protocol if you make the change provided your project approval is still active. If your research has multiple phases, you can submit your IRB with best-guesses as to the questions you'll include in each phase. Then, you only have to submit the updated questions for review when you solidify the questions in each phase.

7. Request permission to recruit and interview more people than you think you'll need in your study to cover all contingencies. What contingencies? Drop-outs, over/under-sampling categories of respondents, switching recruitment sites or strategies. And this means putting *all* the potential methods for recruiting into your proposal—from hanging out on campus, to advertising on social media, to sending IRB-approved letters to colleagues asking them for referrals. If you intend to pay respondents, then include all the details on methods and amounts of payments.

8. Don't settle for a short-term gain from the IRB that may hog-tie your research agenda later. Don't agree, for example, not to collect names if you may need them later. Getting permission to collect names or re-contact respondents takes more work in an IRB protocol, but if your research depends on doing follow-up interviews with some respondents, then that extra work is simply one of the costs of doing the research.

9. Treat all this as part of normal training in how to do research. This will help make the relationship between researchers and those whose job it is to protect human subjects of social research collaborative rather than confrontational.

AMBER'S CORNER

DON'T DAWDLE IN GETTING IRB APPROVAL

Many social scientists have IRB horror stories. Mine is about not knowing the IRB existed. My first research was in China, the culmination of years of Mandarin language study. My study was about how people in Xi'an were using the Internet, as it had just become publicly available. Over 9 months, I did hundreds of hours of participant-observation in internet cafes and interviewed 30 internet users. My faculty advisor had warned me I needed something called "IRB approval" from my university, but I assumed I could get that at any point before I graduated.

I was back in the U.S., ready to analyze and write up my thesis, when my advisor asked: "Where's the IRB approval?" I admitted my mistake, and he broke the bad news: The data weren't usable without prior approval and consent. I'd have to start my thesis over entirely—now on a new, more difficult topic that didn't require data collection. I never again dawdled in getting my IRB protocol under review.

Theory—Explanation and Prediction

All research is specific. Whether you conduct ethnographic or questionnaire research, do content analysis or run an experiment, the first thing you do is *describe a process* or *investigate a relation* among some variables in a population. **Description** is essential, but to get from description to theory is a big leap. It involves asking: "What causes the phenomenon to exist in the first place?" and "What does this phenomenon cause?" **Theory**, then, is about explaining and predicting things.

It may seem odd to talk about theory in a textbook on research methods, but you can't design research until you choose a research question and research questions depend crucially on theory. A good way to understand what theory is about is to pick a phenomenon that begs to be explained and to look at competing explanations for it. See which explanation you like best. Do that for a few phenomena and you'll quickly discover which paradigm you identify with. That will make it easier to pick research problems and to develop hypotheses that you can go off and test.

Here is an example of something that begs to be explained: Everywhere in the world, there is a very small chance that children will be killed or maimed by their parents. However, the chance that a child is killed by a parent is much higher if a child has one or more nonbiological parents than if the child has two biological parents (Archer 2013; Daly and Wilson 1998, 2005; Lightcap et al. 1982). This "Cinderella effect," as it's known, means that those evil-stepparent folk tales are based on more than fantasy. Or are they? A lot depends on the paradigm you start with.

Alternative Paradigms for Building Theories

One explanation is that it's biological—in the genes, as it were. Male gorillas are known to kill off the offspring of new females they bring into their harem. Humans, the reasoning goes, have

DEEP DIVE
HOW TO MAKE A THEORETICAL CONTRIBUTION

Early-career scholars often wonder: What is a **theoretical contribution**, exactly, and how do I know when I've made one? It's a good question, and there's an easy answer. A theory is just a good idea about how the world works. Making a theoretical contribution simply means adding something new to what we already know about how the world works. To do this, you need to *first read everything* that's already known about your idea. Next, you have to *design and do research* that builds on what's already known. That's really it!

There are three main ways to make theoretical contributions in social science:

1. *Use an old theory to study a new phenomenon.* This approach is probably the easiest, because you can build on a large existing literature about a theory, and even borrow some ideas from older research designs. For example, Erving Goffman's 1963 theory of stigma helped Amber study globalizing fat stigma 50 years later.
2. *Use a new theory to study an old phenomenon.* This approach is a bit trickier, because you have to convince your colleagues that you've come up with a new way to conceptualize or explain something social scientists have been studying for a long time. An example of this is when, in the 1980s, Russ helped develop egocentric network analysis to study the social dynamics around a single person, building on the older field of social networks (which began in the 1930s).
3. *Study a new phenomenon with a new theory.* This is the hardest one—you'll have to understand a new social phenomenon at the same time you are creating a new theory or concepts. An example of this is how, in the 2020s, scholars developed novel theories of gender/sex as they studied new dynamics of gender transitions across the life span.

We usually recommend that people try approach 1 or 2 when they are just starting out in research. It's a lot easier to build on work others have done, especially when there are existing journals and scholarly communities to support your work. But approach 3 is the high risk/high reward strategy: probably the most exciting, if you can make it work.

a bit of that instinct in them, too. They fight the impulse, and culture usually trumps biology, but over millions of cases, biology is bound to come out sometimes. This is an explanation based on assumptions from **evolutionary theory**, or Darwinism. (There are several varieties of this, which you'll see under the label of evolutionary anthropology or behavioral ecology or evolutionary psychology or sociobiology.)

Another explanation is that it's cultural. This is called an **idealist theory** (or ideational theory) because it is based on what people think—on their ideas. Yes, it's more common for children to be killed or hurt by nonbiological than by biological parents, but this kind of mayhem is more common in some cultures than in others. Also, the deaths of some children at the hand of their biological parents may go unnoticed and unreported simply because we don't expect that, while the deaths of children at the hands of nonbiological parents get more notice simply because we're on the lookout for it (Crume et al. 2002; Parrish et al. 2017). And, although killing and maiming of children is rare everywhere, in some cultures mothers are more likely than

fathers to be the culprits, even when the woman's partner is the stepfather (Alexandre et al. 2010). Women and men learn different gender roles in different societies, so, the theory goes, we have to look at cultural differences for a true explanation of the phenomenon.

Yet another explanation is that, when adult men and women bring children to a second marriage, they know that their assets are going to be diluted by the claims the spouse's children have on those assets—immediate claims and claims of inheritance. This leads some of those people to harm their spouse's children from the former marriage. In a few cases, this causes death. This is a **materialist theory**, as is the idea that women who have children from a previous marriage may, on average, be forced to marry men who carry a higher risk of being abusive.

Darwinism, idealism, and materialism are not theories. They are **theoretical paradigms** or theoretical perspectives. They contain a few basic *rules for finding theories* that explain observed events. Darwinism stresses the primacy of evolutionary, biological features of humans as the basis for human behavior. Idealism stresses the importance of internal states—attitudes, preferences, ideas, beliefs, values—as the basis for human behavior. And materialism stresses structural and infrastructural forces—like the economy, the technology of production and reproduction, demography, and environmental conditions—as causes of human behavior.

When you want to explain a specific phenomenon, you apply the principles of your favorite paradigm and come up with a specific explanation—a theory.

Why do women everywhere in the world tend to have nurturing roles? If you think that biology rules here, then you'll be inclined to support evolutionary theories about other phenomena as well. If you think that culture—people's values—is of paramount importance, then you'll tend to apply the idealist perspective to come up with explanations. If you think economic and political forces cause both values and behavior, then you'll be inclined to apply the materialist perspective in your search for explanations in general.

The different paradigms are not so much in competition as they are complementary, for different **levels of analysis** (e.g., individual or community) and analysis at different **time scales** (e.g., a day or a decade). The evolutionary explanation for the battering of nonbiological children is appealing for aggregate, evolutionary phenomena—the big, big picture. An evolutionary explanation addresses the question: What is the reproductive advantage of this behavior happening at all?

We know that the behavior of hurting or killing stepchildren is not inevitable, so an evolutionary explanation can't account for why some stepparents hurt their children and others don't. A materialist explanation is more productive for addressing that question. Some stepparents who bring a lot of resources to a second marriage become motivated to reduce the possibility of having their wealth raided and diluted by their new spouse's children. The reaction would be strongest for stepparents who have competing obligations to support their biological children who are living with yet another family. These motivations will cause *some* people to become violent, but not others.

But the materialist explanation doesn't tell us why a particular stepparent is supportive or unsupportive of his or her nonbiological children. At this level of analysis, we need a processual and psychological explanation, one that takes into account the particular historical facts of the particular case. We might theorize that stepparents who bring a lot of resources to a second

marriage become personally angered and frustrated by the possibility of having their wealth raided and diluted by their new spouse's children. Perhaps these frustrations cause some people to become violent.

There is a long list of things that beg to be explained in the social world: Why does total fertility (the number of children born to women of childbearing age) decrease when societies move from agricultural to industrial production? And why do poor women in industrial societies have more children than middle-class women in those societies? Why does modernization result in a lower ratio of ascribed to achieved statuses? Why does romantic love become the basis for marriage in economically advanced societies? What causes the expansion or the decrease in economic inequality across countries of the world? Why do states go to war? And why do democracies rarely go to war with each other?

The Consequences of Paradigms

Differences in theoretical paradigms have profound consequences. If you think that beliefs and attitudes are what make people behave as they do, then if you want to change people's behavior, the obvious thing to do is change their attitudes. This is the basis of the educational model of social change that we mentioned in Chapter 1—the runaway best-seller model for change in modern societies.

Do you want to get students in American high schools to achieve more? Educate them about the importance of taking the most challenging courses. Want to get women in developing

CASE STUDY

INTIMATE PARTNER VIOLENCE ON BARBADOS

Handwerker (1996b) found that stepparents in Barbados were, overall, no more likely to treat children violently than were biological parents. But the presence of a *stepfather* increased the likelihood that women battered their daughters and decreased the likelihood that women battered their sons. In homes with stepparents, women saw their daughters as potential competitors for resources available from their partner and they saw sons as potential sources of physical protection and income.

And there was more. Powerful women (those with their own sources of income) protected their children from violence, treated them affectionately, and elicited affection for them from their male partner. The probability that a son experienced an affectionate relationship with a biological father rose with the length of time the two lived together, but only for sons who had powerful mothers. Men battered powerless women and the children of powerless women, and powerless women battered their own children.

Is there a sociobiological basis for powerful spouses to batter powerless ones? Or is this all something that gets stimulated by material conditions, like poverty? Lots more research is needed on this fascinating question, but the two points here are clear: (1) different paradigms produce different answers to the same question; and (2) a lot of really interesting questions may have answers that are generated from several paradigms.

countries to have fewer children? Educate them about the importance of small families. Want to get people in the United States to use their cars less? Educate them about car-pooling.

These kinds of programs often fail—but they do work sometimes. The closer a behavior is to the culture (or superstructure) of society, the easier it is to intervene culturally. Brand preferences are cultural artifacts, so advertising works to get people to switch brands—to change their behavior. But if people's behavior is rooted in the economic structure or physical infrastructure of society, then forget about changing their behavior by educating them to have better attitudes.

For example, if you need a car because the only affordable housing is 30 miles from your job, no amount of rhetoric about the importance of reducing our carbon footprint in order to protect the environment will convince you to ride a bike. One more example: In poor countries, having many children may be the only security people have in their old age. You can educate people (through social advertising) about using IUDs or birth control pill as opposed to less-effective methods of birth control, once people have decided to lower their fertility, but educational rhetoric doesn't influence the number of children that people want in the first place.

Idiographic and Nomothetic Theories

Theory comes in two basic sizes: elemental or **idiographic theory** and generalizing or **nomothetic theory**. An idiographic, or elemental, theory accounts for the facts in a single case. A nomothetic theory accounts for the facts in many cases. The more cases that a theory accounts for, the more nomothetic it is. When you first run into these concepts, it's easy to suppose that nomothetic is better than idiographic, but idiographic theories are often more immediately useful.

DEEP DIVE
BIG EARLY THEORIES IN THE SOCIAL SCIENCES

Efforts to develop nomothetic theories in the social and behavioral sciences are well known. In psychology, Sigmund Freud (1962/1905) based his theory that psychosexual development in children proceeds in five stages—oral, anal, phallic, latency, and genital—on just a few cases. Jean Piaget (1952) did the same in developing his theory of four stages of cognitive development in children (sensorimotor, preoperational, concrete operational, formal operational).

In the mid-to-late nineteenth century, sociologists and anthropologists made brave, if ill-fated efforts to devise nomothetic theories of social and cultural evolution—all on the basis of the cases at hand. Auguste Comte, for example, argued that human society had progressed through three stages: a theological stage in which humans explained natural phenomena (tides, fires, illness) as being caused by supernatural forces and beings; a metaphysical stage in which natural phenomena were explained by natural-but-mystical forces, like spiritual energy or consciousness; and a positive stages (from which the word *positivism* derives), in which natural phenomena are explained by the amassing of data from observation and experimentation.

The idea that the evolution of human societies involved *progress* (from undesirable to desirable) was a key element of all the 19nineteenth century theories of human evolution.

Lewis Henry Morgan (1877), for example, argued that societies evolved through three stages: savagery, barbarism, and finally, civilization, with Western societies as the pinnacle of modern civilization. The unilineal theories of human social evolution advanced by the 19th-century social scientists were offensive and wrong, but the effort to produce nomothetic theory was *not* wrong. Half a century later, anthropologists and sociologists like Leslie White (1949), Julian Steward (1955), and Gerhard Lenski (1970) advanced more nuanced theories about how cultures evolve.

There Is No List of Research Questions

There are many lists of **research topics** for social scientists—things like immigration, consumer behavior, racism, social media, gender roles, globalization. … But there is no list of **research questions**. The reason there is no list of research questions is because any such list would be, for all practical purposes, endless. Are there health effects for running a high-traffic personal account on social media? Why do people continue to buy vinyl records? Why do some communities bounce back quickly after a natural disaster, like a hurricane or flood, while others don't?

Use your imagination and your curiosity about how things work and follow your hunches. Above all, never take anything at face value. Every time you read an article, ask yourself: "What would a study look like that would test whether the major assertions and conclusions of this article were really correct?" If someone says: "The only things students care about are sex, drugs, and social media," the proper response is: "We can test that."

The Literature Search

The first thing to do after you get an idea for a piece of research is to find out what has already been done on it. Don't neglect this part of the research process: You can't argue your contribution is new until you know what everyone else has done. And *never* say "little is known about ..." any topic in any research paper or grant proposal you write.

First of all, it's almost never true. And if it were true, it'd mean hardly anyone is interested in reviewing and supporting your work. Building on past research—no matter how old or seemingly tangential—always strengthens (and never diminishes) your own contribution. Second, and most chillingly, any paper you submit for publication and any proposal you submit for funding will inevitably be read by exactly the people who've written about that ""little known"" topic. The bottom line on this that you need to make a heroic effort to uncover the literature for any topic you study. Fortunately, heroic efforts are easier than ever, what with all the documentation resources available for scouring the literature.

Getting Started: Backward and Forward Citation

You just need *one name* to start. Ask a professor or colleague: Who's the most important person to read on X topic? It doesn't matter if the answer is a good one, or just so-so. Either way, this name can unlock the history of scholarship on your topic.

First, take that name and find their most-cited relevant paper in Google Scholar. Go to the list of authors they cited in the paper's bibliography—this starts the **backward citation**. Keep pulling names and following the thread backward in time until you find the people who founded this field. Now you have a list. From this list, identify the scholars and papers that are most cited *and* the ones that interest you the most.

Next, search them in Google Scholar, and then press the "cited by" button: This is **forward citation**. It leads you to the people who are building on the original citation. Keep pulling names and following the thread forward in time until you find the people who are leading the field today. Once you've traced the backward and forward citation, you can map the field, locate everyone who ever worked in it, and find the best places for you to make your own contribution.

Want a shortcut? Another way to start the work of backward and forward citation is with volumes of the *Annual Review*. There are *Annual Review* volumes for psychology (every year since 1950), anthropology (since 1959), sociology (since 1975), public health (since 1997), and political science (since 1998). Authors who are invited to publish in these volumes are experts in their fields; they have digested a lot of information and have packaged it in a way that gets you right into the middle of a topic in a hurry. And their backward citations should be impeccable.

Today, there are plenty of online tools to help with **literature mapping**. On literature mapping websites, you can enter just one key paper—and it get back a bibliometric network (co-citation map), plus a list of backward and forward citations. Many university libraries have lists of the latest literature mapping tools, with reviews of how well they perform. As these tools are still being refined, it's best to use them in combination with old-fashioned literature review.

Documentation and Database Resources

These days, documentation is a robust business, and there are many indexing and abstracting resources. Besides Google Scholar, we cover some important databases here. And if you are connected to a library database when you search these databases, you'll have access to the actual articles, not just to the citations and abstracts.

TIPS & HOW-TO
CONTACTING RESEARCHERS

Another tip: Contact people on listservs and networking groups that deal with your dissertation research topic. If there are central figures in the field, contact them by e-mail and request a time when you can call them. Yes, by phone. Or video call. E-mail and texting may be convenient for *you*, but many scholars are too busy to respond to requests for lists of articles and books. On the other hand, many scholars, especially older ones, *will* talk to you if they think they can really help. And established scholars love to meet up-and-coming new colleagues.

The Web of Science

The Clarivate Analytics *Web of Science* contains the *Science Citation Index*, the *Social Sciences Citation Index*, and the *Arts and Humanities Citation Index*, covering about 10,000 journals.

PsycINFO

PsycINFO is a product of the American Psychological Association. The Jurassic version of this database goes back to the early 1800s. It indexes and abstracts about 2,500 journals in the behavioral and social sciences and contains over 3 million records.

Sociological Abstracts and Dissertation Abstracts

Sociological Abstracts (SA) is one of many databases offered by ProQuest, including dissertation and theses abstracts with about 1,800 journals dating from 1952.

ERIC

The Educational Resources Information Center (ERIC) is available free online. It covers education journals and a lot of grey literature—government reports and reports from private foundations and industries that contain useful information but can be tough to find.

NTIS and FDsys

If the federal government has funded research on a topic, there's probably a technical report about it. The National Technical Information Service (NTIS), indexes and abstracts federally funded research reports in all areas of science.

LEXIS/NEXIS

LEXIS/NEXIS began in 1973 to help lawyers find case information. Today, it has 5 billion documents from some 60,000 legal sources and 40,000 news sources, including world's major English-language newspapers.

OCLC and WorldCat

The OCLC (Online Computer Library Center) is the world's largest library database, with 17,000 member libraries in 123 countries, in 484 languages, in the WorldCat catlog. This database helps run the "interlibrary loan" system, telling your library where to borrow the books and articles you need.

PubMed and MEDLINE

PubMed is a product of the National Library of Medicine (National Institutes of Health), covering about 5,400 journals in the medical sciences. If you are working on anything that has to do with health care, PubMed is a must.

HRAF: The Human Relations Area Files

The eHRAF World Cultures database has 100+ years of ethnography covering 2,500 world cultures, pre-coded for topics on everything from marriage customs to tattooing.

Meta-Analysis

Meta-analysis involves piling up all the studies ever done on a particular topic to assess what is known about the size of the effect. It is, as Hunt (1997) put it, how science takes stock. The foundational work on meta-analysis (M. L. Smith and Glass 1977) addressed the question: Does psychotherapy make a difference? Since then, there have been tens of thousands of meta-analyses on everything from gender differences in performance on math tests (Lindberg et al. 2010) to the influence of the Internet on citizen participation in public policy making (Yang and Zhiyong 2010).

Meta-analysis forces you to become familiar with the literature on a particular topic and it makes you aware of the research holes that need to be filled. Schutte and Hosch (1997), for example, did a meta-analysis of mock jury studies about rape or child sexual abuse. In a mock jury study, participants are shown evidence of a defendant's guilt and innocence in a particular crime. The jury deliberates and renders a verdict. It's an attractive method because it mimics a real-world situation and because you can manipulate the experimental treatment—the crime, the various kinds of evidence for the defendant's guilt or innocence, the demographics of the jurors, and so on.

Schutte and Hosch scoured the literature. They began by searching the PsycINFO database from 1967 on for articles that contained *any* of the terms "sexual abuse," "child abuse," "rape," "sex," and "juror." They also posted requests an Internet discussion group for people interested in law and psychology. They then used the bibliographies from the articles they turned up to hunt for further references and kept on doing this iterative search until no new studies turned up that fit their criteria.

And what were the criteria? First, they only used reports that were based on studies of jury-eligible people. That meant excluding studies of people under 18 years of age and excluding studies of non-U.S. citizens (so all studies of Canadians and Britons, for example, were excluded). Second, they excluded studies in which respondents (mock jurors) were asked to rate a mock defendant's guilt on a Likert-type of scale of, say, 1-to-5. In real jury cases, defendants are judged guilty or not guilty, not "somewhat guilty" or "very guilty."

Schutte and Hosch wound up with 36 studies, 19 involving accusations of rape, and 17 involving accusations of child sexual abuse. All these studies together comprise 9,813 participants (51% of whom were women) and a mean of 271 participants per study. This points to one of the strengths of meta-analysis: Even though the number of *studies* in such an analysis might be low, the number of *people* represented in those studies can be huge.

Across the 36 studies, women jurors were far more likely to vote for conviction than were men (58.5% compared to 41.5%). This was hardly surprising, but the study did turn up something very interesting: 29 out of the 36 studies involved female victims and male defendants. Of the seven studies in which women were the accused, every case was about child sexual abuse, and three of the seven studies reported no difference in the probability that male or female mock jurors would vote to convict. This is just the sort of finding that sharp-eyed researchers latch on to when they're out shopping for interesting research gaps to fill. In fact, Schutte and Hosch's 1997 study was part of a wave of mock-jury studies on the effects of gender of both victim and defendant (Lynch et al. 2019; (Pozzulo et al. 2010; Quas et al. 2002).

Today, there are published guidelines for meta-analysis like the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). Your approach can be quantitative or qualitative. It can be a comprehensive review, or a scoping review that takes stock of an emerging field. (**Further Reading** meta-analysis)

KEY CONCEPTS IN THIS CHAPTER

Backward citation
Code of ethics
Cultural relativism
Description
Ethical research
Evolutionary theory
Forward citation
Idealist theory
Ideographic theory
Institutional Review Boards
Levels of analysis
Literature mapping
Materialist theory
Meta-analysis
Moral relativism
Nomothetic theory
Research problem
Research process
Research questions
Research topics
Science
Theory
Theoretical contribution
Theoretical paradigms
Time scales

SUMMARY

- Research is idealized, but in the end, it gets done the way most things get done: by doing the best we can and by trying to do better next time.
    - Researchers choose their problems for many reasons, including personal interest, availability of research funds, contractual obligations, and to build sound explanations for social and behavioral phenomena.
- The ethics dilemma in social research is profound. The operational test of whether a particular piece of research is ethical is whether social norms tolerate it.
    - This relativistic position, however, does not encourage absolute moral judgments. Ultimately, the choice is left to researchers, and the researchers are responsible for the consequences of their actions.
    - It is unlikely that either Milgram's experiments on obedience would be funded today, yet the lessons from their experiments continue to provide guidance on the responsibility of the individual for her or his actions.
- There are quite different approaches, or paradigms, to theory building in the social sciences. These paradigms guide us to search for different *kinds* of answers—biological, ideational, and material—to the same question.

- All research projects begin with a literature search. The bibliographic tools available today make it much easier than in the past to cover the literature thoroughly.
    - The Social Science Citation Index, ERIC, NTIS and FDsys, PubMed, Sociological Abstracts, LEXIS-NEXIS, and OCLC are some of the documentation resources available.
    - Many topics of research have been the subject of meta-analysis. Begin your assessment of the literature by reading any meta-analyses that may be available.

EXERCISES

1. Building a database of references for a research topic of your choice is the best way to learn how to use the bibliographic tools in your college library. Choose any topic you like and try to make the literature search exhaustive. This is a great way to learn about narrowing down your research *interests* into manageable research *problems*.

   If you're interested in gender differences, for example, the initial search for the string "gender differences" in PsycINFO returns about 25,000 items between 1685 (yes, 1685) and 2011. Better focus it more. Asking for "gender differences" and "test taking" returns about 40 items, with the earliest at 1987. Asking for "human sex differences" and "test taking" returns about 90 items, beginning with 1966.

2. After reading this chapter, you should have more to say about the concepts of value-free science and research. The examples, though, have been experiments, not research based on questionnaires or on participant observation ethnography.

   Does questionnaire research done over the telephone pose any ethical problems? How about online? How about ethnographic research? Use the bibliographic tools in your library to find articles on these issues in social research ethics. Look up the problem of informed consent and learn about the different requirements in social research and medical research. If you stop someone on the street to administer a questionnaire and they answer you, does that imply consent?

3. Use the bibliographic tools in your college library to find at least one example of social research that is based explicitly on the idealist paradigm. Then find an example of research based on the materialist paradigm and another based on the evolutionary paradigm. Be sure that the three articles are reports of research, not a theoretical discussion. Write a brief report describing the articles and then discuss the different approaches taken by the authors.

FURTHER READING

Social science in the military and in intelligence. For opposing views of social scientists' involvement in the wars in Iraq and Afghanistan, see Price (2003) and McFate (2005); see

also González (2007) and Kilcullen (2007). For a summary of this debate, see Fluehr-Lobban (2008) and Forte (2011).

Milgram and obedience experiments. Benjamin and Simpson (2009), Blass (2004), Bocciaro and Zimbardo (2010), Packer (2008), Slater et al. (2006).

Ethical issues in social science. Bosk (2004), Boruch and Cecil (1983), Burgess (1989), Citro et al. (2003), Fluehr-Lobban (1996), Hammersley (2009), Herrera (1996), Keith-Spiegel and Koocher (2005), Lyman (1989), Mertens and Ginsberg (2009), van den Hoonaard (2002), Weisstub (1998), Weisstub and Diaz Pintos (2007).

Meta-analysis. Cook et al. (1992), Cooper et al. (2009), Farley and Lehman (1986), Glass (1976), Guzzo et al. (1987), Hedges and Olkin (1985), Hunt (1997), Hunter and Schmidt (2004), Matt and Navarro (1997), Pan (2008), Rosenthal (1984), Wolf (1986).