

Preface

GOALS AND PHILOSOPHY

In 1999, when the first edition of this book came out, the mixed methods movement was just getting underway in the social sciences and the book's subtitle, *Qualitative and Quantitative Approaches*, announced something new. How new? The first citation of "mixed methods" in the *Social Science Citation Index* was in 1993 and by the end of 1999 there were a whopping 21 of them. We are happy to report that mixed methods is no longer innovative. There is a successful *Journal of Mixed Methods Research* (<http://mmr.sagepub.com>), and text books on designing and carrying out mixed-methods research (Creswell and Plano Clark 2018; Peltó 2017)—and over 45,000 citations to mixed methods in the *Social Science Citation Index* as of this writing.

We can hardly contain our enthusiasm for all this. There is, of course, an irreducible difference between those of us for whom the first principle of inquiry is that reality is constructed uniquely by each person and those of us who start from the principle that external reality awaits our discovery through a series of approximations. There is also an important (though not incompatible) difference between those of us who seek to *understand* human phenomena in relation to differences in beliefs and values and those of us who seek to *explain* human thought and behavior as the consequence of external forces.

But, while the boundaries between the disciplines remain strong, those boundaries are no longer about methods—if they ever were. Whatever our epistemological differences, the actual methods by which we collect and analyze our data belong to everyone across the social sciences.

In short, all the methods belong to all of us.

WHAT'S IN THIS BOOK?

There are 17 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 at length, in several chapters. Look for references to all those discussions under "ethics" in the index.

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 feel strongly that students 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 on scaling: how to use scales and how to build them. A single question on a questionnaire is technically a scale if it lets you assign the people you're studying to categories of a variable. Many interesting variables in social science, however, are complex and can't easily be assessed with single indicators. Chapter 5 covers methods for developing and testing composite measures of complex concepts. Methods covered include Guttman scales, Likert scales, and semantic differential scales.

Chapter 6 is about the basics of probability sampling, including the central limit theorem, and determining the size of representative samples. Chapter 7 is on nonprobability sampling, including quota sampling, snowball and respondent driven sampling, purposive sampling, and convenience sampling. Together, Chapters 6 and 7 address the question: Given that my findings are valid, how far can I generalize them beyond the people (or countries, or court cases) I actually studied?

Chapters 8, 9, 10, and 11 are on the real how-to of collecting social science data. Chapter 8 deals with unstructured and semistructured interviewing, including 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 8 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 9 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 10 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 deal with some of the special techniques for analyzing network data in Chapter 16.

Chapter 11 is on methods for field research: direct observation 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.

Chapters 12 and 13 cover the analysis of texts. 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, newspapers, artifacts, still and moving images, advertisements. . . . Chapter 12 is about the inductive method of grounded theory and the deductive methods of content analysis. Both methods involve coding, so this chapter includes a method for testing intercoder reliability. Chapter 13 covers conversation analysis, narrative analysis, phenomenology, language in use, and critical discourse analysis.

Chapters 14, 15, and 16 are on quantitative data analysis. Chapter 14 deals with univariate analysis—that is, statistics that describe a *single variable*, without making any comparisons among variables—and bivariate analysis—statistics that describe *relationships in pairs variables*. Chapter 15 is an introduction to multivariate analysis—statistics that describe *relationships in larger sets of variables* and that let you test hypotheses about what causes what. Chapter 16 is about analyzing network data. In addition to the specialized techniques required for analyzing network composition and network structure, this involves multidimensional scaling, cluster analysis, and cultural consensus analysis, so these methods of analysis are explained here, rather than in the chapter on multivariate statistics.

If you want to become comfortable with statistical analysis, you need more than a basic course; you need a course in regression and applied multivariate analysis and a course (or a lot of hands-on practice) in the use of statistical software like R, SPSS®, SAS®, STATA®, and SYSTAT®. 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 the materials in Chapters 14 and 15, 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, Chapter 17 is about the last big piece of the research puzzle: writing up the results of all your work and getting it published.

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 and all these terms are listed in the index as well.
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.* Finally, 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.

WHAT'S NEW IN THIS EDITION?

There are two authors of this edition. In boxes labeled “Amber’s Corner” and “Russ’s Corner” we report on our own experiences in doing research. There are also boxes on “Tips & Tricks” and on “Deep Dives”—pedagogical devices we’ve found helpful in our teaching of research methods.

We’ve replaced outdated references with more current ones and updated the literature where needed. There is new material on choosing a nonprobability sample and detailed instructions on how to write research proposals.

The Further Reading section at the end of each chapter is expanded to include recent material. You’ll also see reference to really, really old stuff because we want students to know about many of the classics.

HRB and AW
Oct. 22, 2025

The Foundations of Social Research

1

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, journalism, 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 16th century, modern science—which, unlike other ways of knowing, relies on systematic real-world observations—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 will 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).

Consider life insurance: You bet the insurance company that you will die within the next, say, 20 years. They ask you a few questions: How old are you? Do you smoke? What do you do for a living? Do you fly a private plane? Then, depending on the answers, they tell you that the bet is your \$435 every year for the next 20 years, against their promise to pay your heirs \$100,000 if you die at any time within those 20 years. If you *lose the bet* and stay alive for 20 years, they keep your $\$435 \times 20 = \$8,700$ and reset the yearly rate. This is simply spectacular human engineering at work, and it’s all based on scientifically developed knowledge about risk assessment.

Insurance also includes all the public health care delivery systems in the world—like Medicare in the United States and the National Health Service in England, in addition to all the private insurers—and all the pension plans in the world, like Social Security in the United States and Old Age Security in Canada.

In 1889, Otto von Bismarck came up with a pension plan for retired German workers. Based on sound social science data, Bismarck’s minister of finance suggested that 70 would be just the right age for retirement. At that time, the average life expectancy in Germany was closer to 50, and just 30% of children born then could expect to live to 70. Germany lowered the retirement age to 65 in 1916, by which time, life expectancy had edged up a bit—to around 55 (McCarthy Institute 2002). In 1935, when the Social Security system was signed into law in the United States, Germany’s magic number 65 was adopted as the age of retirement. White children born that year in the United States had an average life expectancy of about 63; for Black children it was about 51 (SAUS 1947:Table 88).

Today, life expectancy at birth in the highly industrialized nations is over 80—fully 30 years longer than 100 years ago—and 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 a medical system that caters to the needs of 80-somethings when birth rates are low and there are fewer working adults to support the retirement of the elderly? The answers to these questions must be guided by our values, but real-world 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 18th century, every phenomenon to which the scientific method has been systematically applied, over a sustained period, 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-19th 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 attacks. 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 what were then the latest supercomputers, found that if someone bought disposable diapers at 5 p.m., the next thing they were 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. 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. More than 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, validity, reliability, cause and effect, and theory. When you finish this section, you should understand the crucial role of measurement in science and 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, 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), 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 fifth graders, etc.).

Social research, whether it's based on questionnaires, field observations, or experiments, starts with 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.

Dimensions of Variables: Simple or Complex?

Variables can be **unidimensional** or **multidimensional**. The distance from Chicago to Albuquerque can be expressed in driving time or in miles, but no matter how you measure it, distance is expressed as a straight line and straight lines are one dimensional. You can see this in Figure 1.1.

FIGURE 1.1

Two Ways to Measure Distance

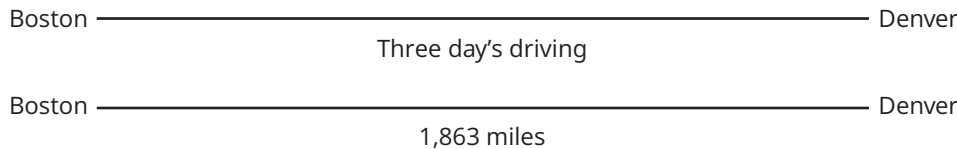
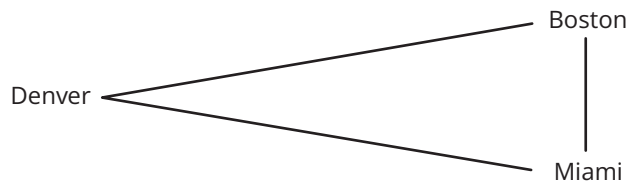


FIGURE 1.2

Three Points Create Two Dimensions



If we add Miami, we have three distances: Chicago–Miami, Chicago–Albuquerque, Albuquerque–Miami. One dimension isn't enough to express the relation among three cities. We have to use two dimensions. Look at Figure 1.2.

The two dimensions in Figure 1.2 are up–down and right–left, or North–South and East–West. If we add Nairobi to the exercise, we'd have to add a third dimension (straight through the paper at a slight downward angle from Albuquerque) or do what Gerardus Mercator (1512–1594) did to force a three-dimensional object (the Earth, a sphere) into a two-dimensional picture. The cost of doing this was distortion at the edges. This is why, on a Mercator Projection map of the world, Greenland (an island of 840,000 square miles), looks the same size as China (a land mass of about 3.7 million square miles).

Unidimensional variables, like height, weight, birth order, age, and marital status, are relatively easy to measure. By contrast, stress, wealth, and political orientation are all multidimensional and more difficult 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.

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.

We'll discuss the building of scales and how to test for unidimensionality of variables in Chapter 5 on scaling.

Dependent and Independent Variables

Recall the life insurance problem: 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* for, say, 20 years. You take the bet. If you lose (and remain alive), the company takes everything you've paid and banks it. If you win the bet (and die), the company pays your beneficiary.

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

Longevity, then, is the **dependent variable**, because it *depends on* sex, education, occupation, and so on. 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 need 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).

For example, across the developing world, when women have access to well-paying jobs outside the home, they tend to lower their fertility. However, a minimal level of education by itself doesn't cause women to attain economic self-sufficiency. Women in low- and middle-income countries who acquired basic literacy and numeracy skill in the 1980s got entry-level factory work. Once that happened, they encouraged their daughters to stay in school longer (Handwerker 1989). In other words, jobs for women in one generation caused increased education in the next, which, in turn, led to better-paying jobs. (More about this in Chapter 3 on the role of theory in developing testable hypotheses.)

CONCEPTS AND MEASUREMENT

Variables are measured by their **indicators**, and indicators are defined by their **values**. Some variables, and their indicators, are easily observed and measured. Others are more conceptual. The difference is important.

Consider the variable “gender.” Historically, to *measure* gender, many researchers have simply looked at a person and decided whether they are female or male. **Measurement** is *deciding which value to record*. And that decision is obviously prone to error. There’s no foolproof way to assign a gender category based visible expressions of gender norms, like someone’s clothes, hair length, or name. Any of these indicators may lead you into making the wrong measurement—incorrectly marking down woman or men, transgender or cisgender, and the like. Many people are intersex, and many more are genderqueer. Today, there are 105+ recognized gender categories in English alone. Good luck visually distinguishing between whether someone is agender, ambigender, androgyne, or androgynous—and that’s just the As.

Improving measurement in science means lowering the probability of and the amount of error. In national surveys, about 0.06% (that’s six tenths of one percent) of Americans identified as transgender (Herman et al. 2022). When survey questions omit transgender as a response category, they will make, on average, 60 errors in identifying transgender people of a random sample of 10,000 Americans. Whether that’s an acceptable risk of error is a research design problem—and an ethical question about who should be ignored and invisible in social science research.

Gender is a **concept**, or mental construction. We have to make it up to study it. As our thinking about a concept evolves, we update it: The newer concept of gender/sex (van Anders 2015) improves on the overly simplistic gender concept. All variables are concepts, but some concepts, like height and weight, are easier to measure than others. Concepts like race/ethnicity, religious intensity, dedication to public service, willingness to accept new agricultural technologies, tolerance for foreign fieldwork, desire for an academic job, compassion, and jealousy are complex and much more difficult to measure.

Complex concepts are often called **constructs**. We are led to defining constructs by our experience: Some people just seem more religiously intense than others, more jealous than others, more tolerant of living and working overseas than others, and so on. We verify our intuition about conceptual variables by measuring them, or by measuring their results.

Suppose you put an ad in the paper that says: “Roommate wanted. Easy-going, non-smoker preferred.” When people answer the ad, you can look at their fingers and smell their clothes to see if they smoke. But you’d have to ask people a series of *indicator questions* to gauge their easy-goingness.

Similarly, to predict who among a group of prisoners is predisposed to return to crime after release, you will want to measure that predisposition with a *series of indicators*. The indicators can be answers to questions on formal tests, or answers to open-ended questions about plans for the future, or even directly observable facts—like whether someone has a supportive family to which they will return after release from prison.

It may be easier to measure some concepts than others, but the fact is, all measurement is difficult. People have worked for centuries to develop good instruments for measuring things like temperature. And if it’s difficult to measure temperature (a concept, after all, backed up by time-tested theories), how do you measure worker alienation or sexism? Measuring variables like these, which lack concrete indicators, is one of our biggest challenges in social science because these variables are mostly what we’re interested in.

One of the most famous variables in social science is socioeconomic status (SES). It is typically measured using three indicators: income, education, and occupation. You can use income as a single indicator, but it is too simplistic to fully capture SES. For example, there are many wealthy people who have low SES—the so-called *nouveau riche*, such as Real

Housewives who have high incomes but relatively low education and occupational prestige. And some people with low incomes can 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, but still have high education and social prestige. You can add level of education to income as an indicator, but that still won't be enough in most societies of the world to get at something as multidimensional as SES. You can add occupation, father's occupation, number of generations in a community, and so on, depending on the group you are studying, and you still might wind up dissatisfied with the result if your measure fails to predict some dependent variable of interest.

And indicators of any concept may vary from culture to culture. This doesn't mean that measurement is impossible. It means that if you decide to use a scale that was developed and tested in one culture on a new culture, you need to adapt and test the scale for the new culture. More about testing scales in Chapter 5.

CONCEPTUAL AND OPERATIONAL DEFINITIONS

While most of the interesting variables in social science are concepts, some of our most important concepts are not variables. The concept of "love" is not a variable, but the concept of "being in love or not" *is* one. The concept of "culture" is not a variable, but the concept of "intensity of feeling of belonging to a particular culture" *is* one. The concept of attitude is not a variable, but the concept of "supporting or not supporting the right to abortion in the United States" implies a variable with at least two values: support and nonsupport.

Conceptual Definitions

There are two ways to define variables—conceptually and operationally. **Conceptual definitions** are abstractions, articulated in words, that facilitate understanding. They are the sort of definitions we see in dictionaries, and we use them in everyday conversation to tell people what we mean by some term or phrase. **Operational definitions** consist of a set of instructions, like a recipe, on how to measure a variable that has been conceptually defined.

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, and this means agreeing with your friend on exactly what to measure: Does the screened-in porch and the garage count as part of "being spacious" or just the interior living space? Are you counting the square footage or the cubic footage? It's the cubic footage, after all, that creates the "feeling of spaciousness" that comes from high ceilings. Do we measure the square footage of open space before or after the furniture and appliances go in? 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.

FIGURE 1.3

An Instrument for Measuring What People Think “Middle Age” Means

1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Here is a line that represents age. Obviously, a person 1 year of age is a baby, and a person 100 years of age is old. Put a mark on the line where you think middle age begins and another mark where you think middle age ends.

To get a feel for how complementary the two styles can be, ask some 50 year olds and some 20 year olds to tell you how old you have to be to be middle aged. You’ll see immediately how volatile the conceptual definition of “middle age” is. If you ask people about what it *means* to be middle aged, you’ll get plenty of material for an interesting paper on the subject. If you want to *measure* the differences between men and women and between older and younger people on this variable, you’ll have to do more than just ask them. Figure 1.3 shows an instrument for measuring this variable.

Many concepts that we use in social research have volatile definitions: “power,” “social class,” “machismo,” “alienation,” “willingness to change,” and “fear of retribution.” If we are to talk sensibly about such things, we need clear, **intersubjective** definitions of them. In other words, although there can be no objective definition of middle age, we can at least agree on what we mean by middle age for a particular study and on how to measure the concept.

Complex variables are conceptually defined by reducing them to a series of simpler variables. The concept of “ethnic identity” is very complex. But if you state clearly that you mean to measure: (1) varying levels of overt expression of pride in ethnic heritage; (2) varying levels of knowledge about ethnic foods; and (3) varying levels of financial commitment to participation in ethnic heritage activities in your conceptual definition, then at least others will understand what you’re talking about when you say that people are “high” or “low” on ethnic identity.

Similarly, “machismo” might be characterized by “a general feeling of male superiority,” accompanied by “insecure behavior in relationships with women.” Intelligence might be conceptually defined as “the ability to think in abstractions and to generalize from cases.” These definitions have something important in common: They have no external reality against which to test their truth value.

Conceptual definitions are at their most powerful when they are linked together to build theories that explain research results. When the United Nations was founded in 1945, the hope was that trade between industrialized and non-industrialized countries of the world would result in economic development for everyone. The economies of the developed countries would expand, and the benefits of an expanding economy would be seen in the underdeveloped countries. A decade later, it was obvious that this wasn’t happening. The rich countries were getting richer and the poor countries were getting poorer.

Raúl Prebisch, an Argentinian economist who worked at the UN, argued that under colonialism, rich countries were importing raw materials from poor countries to produce manufactured goods and that poor countries had come to depend economically on the rich countries. Prebisch’s dependency theory links the concept of “control of capital” with the concepts of “mutual security” and “economic dependency.” It is that linkage that helps explain why economic development often results in some groups winding up with less access to capital than they had before a development program (1984, 1994).

Conceptual definitions are at their weakest in the conduct of research itself because concepts have no empirical basis. (When we say “empirical,” we mean data or knowledge about the real world gained through observation.) To repeat: We have to make concepts up to study them. There is nothing wrong with this. 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 because, while they point us toward measurement, they don’t really give us any recipe for measurement. Without measurement, we cannot make useful comparisons. We cannot tell whether Spanish people are more emotive than the British people, or whether Catholicism is more authoritarian than Buddhism. We cannot evaluate one community’s anger level over police misconduct or compare it to the anger found in another community in another city.

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, assuming the family doesn’t need the money; if he says “no,” then give him a score of 1, and if he says “yes,” then score him 0. Ask him if he thinks women and men

Box 1.1 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 (Scott 1977). 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. 2023). Her expertise gave us the boost we needed to develop an operational definition. We create a code system—using words, not numbers—to operationalize *moral economies for water*.

We’re still 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.

should have the same sexual freedom before marriage; if he says “no,” score 1 and score 0 for “yes.” Ask him if the birth of a male child is more important than a female child; if he says “no,” score 0; score 1 for “yes.” Add the scores. A man who scores 3 has more machismo than a man who scores 2, and a man who scores 2 has more machismo than a man who scores 1.

3. Ethnic identity: Ask a sample of third-generation Chinese Americans who were born in San Francisco if they speak the language of their grandparents fluently. If “yes,” score 1. If “no,” score 0. Ask them if they eat at non-Chinese restaurants regularly when they eat out. Score 1 for “no,” and 0 for “yes.” Ask them eight other questions of this type and give them a score of 1 for each answer that signifies self-identification with their parents’ heritage. Anyone who scores at least 6 out of 10 is an “identifier.”

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 tell if doctors have higher intelligence scores than lawyers.
2. We can tell if men from Bolivia have higher machismo scores than men from Paraguay.
3. We can tell 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 tell 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, 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.

Operationism is often crude and simplistic, and that, too, can be a strength. Robert Wuthnow (1976) operationalized the concept of religiosity in 43 countries using UNESCO data on the number of books published in those countries and the fraction of those books classified as religious literature. Now *that’s* crude. Still, Wuthnow’s measure of “average religiosity” correlates with seven of eight indicators of modernity in societies across the world. For example, the higher the literacy rate in 1952, the lower the religiosity in 1972. And, going

Box 1.2 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.

forward, the higher the literacy rate the lower the suicide rate in 20 developed countries through 1989 (Cutright and Fernquist 2000, 2001).

Following up Wuthnow's work with more refined measurements—to test hypotheses about the societal conditions that support or weaken religiosity—is a lot more exciting than dismissing it because it was so audaciously crude.

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, sea temperature is different if you measure it from a satellite (you get an answer based on radiation) or with a thermometer (you get an answer based on a column of mercury). Intelligence is different if you measure it with a Stanford-Binet test or the Wechsler scales. If you ask a person in any of the industrialized nations “How old are you?” or “How many birthdays have you had?” you will probably retrieve the same number. But the very concept of age in the two cases is different because different “instruments” (queries are instruments) were used to measure it.

This principle was articulated in 1927 by Percy Bridgman in *The Logic of Modern Physics* and has become the source of an enduring controversy. 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—as scholars have done with gender/sex, as they improve on the gender concept and its flawed measurement.

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. It is obvious that “future orientation” is more than asking someone: “Do you buy large or small boxes of soap?” The problem is, *you* might not include that question in *your* interview of the same respondent, unless we specify that we asked that question in that particular in *our* interview.

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*. What’s your name? In what country were you born? Are you healthy? On the whole, do you think the economy is in good shape? Is Mexico in Latin America? Is Switzerland a rich country? What is your gender?

For gender, you can assign the numeral 1 to men, 2 to women, and 3 to non-binary people, but gender will still be a qualitative, nominal variable. The number 2 happens to be twice as

big as the number 1, but this fact is meaningless with nominal variables. You can't add up all the 1s and 2s and 3s and calculate average gender any more than you can add up all the telephone numbers in Chicago and get the average phone number.

Assigning numbers to things does make it easier to do certain kinds of statistical analysis on qualitative data—more on coding text in Chapter 12 and more on statistical analysis in Chapter 14.

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 *exhaustive* and *mutually exclusive*. The famous “other” category in nominal variables makes the list exhaustive—that is, all possible categories have been named in the list—and the instruction to “check one” makes the list mutually exclusive. (More on this in Chapter 9 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. We say generally because life is complicated and variables that seem mutually exclusive may not be. Some citizens of Lebanon have one Catholic and one Muslim parent and may think of themselves as both Muslim and Catholic.

Most people in Western countries think of themselves as a member of one so-called race or another, but more and more people think of themselves as belonging to two or more races. In the 2010 U.S. Census, a total of 9.1 million Americans checked off more than one so-called race from six choices: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian and other Pacific islander, and Other. In the 2020 Census, a total of 33.8 million people identified as multiracial—about 10% of the total U.S. population and an increase of 264% over the figures from 2010 (United States Census 2023:Table 4).

And when it comes to ethnicity, the requirement for mutual exclusivity is just hopeless. There are Chicano African Americans, Chinese Cuban Americans, Filipino Cherokees, and so on. This just reflects the complexity of real life, but it does make analyzing data more complicated because each *combination of attributes* has to be treated as a separate category of the variable “ethnicity” or collapsed into one of the larger categories. More about this in Chapters 14 and 15 when we get to data analysis.

Occupation is a nominal variable, but lots of people have more than one occupation. People can be pediatric oncology nurses and Uber drivers at the same time. A list of occupations is a measuring instrument at the nominal level: You hold each person up against the

list and see which occupation(s) they have.

The attributes of nominal variables can change over time. In the 1970s, surveys done in the United States typically asked people: “Are you (check one): Protestant, Catholic, Jewish, other religion, no religion.” In the 1970s, very few people in the United States reported themselves Muslims. By 2008, about 1.3 million people in the United States described themselves as Muslims (SAUS 2010:Table 75). By 2020, about 4.5 million people in the United States identified themselves as Muslims (Mohamed 2018).

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**.

The difference between 30° Celsius and 40° is the same 10° as the difference between 70° and 80°, and the difference between an IQ score of 90 and 100 is (assumed to be) the same as the difference between one of 130 and 140. On the other hand, 80° Fahrenheit is not twice as hot as 40°, and an IQ score of 150 is not 50% higher than a person with an IQ score of 100.

Ratio variables are interval variables that have a true **zero point**—that is, a 0 that measures the absence of the phenomenon being measured. The Kelvin scale of temperature has a true zero: It identifies the absence of molecular movement, or heat.

The consequence of a true zero point is that measures have ratio properties. A person who is 40 years old is 10 years older than a person who is 30, and a person who is 20 is 10 years older than a person who is 10. The 10-year intervals between the attributes (years are the attributes of age) are identical. That much is true of an interval variable. In addition, however, 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. These, then, are true ratios.

Although temperatures (in Fahrenheit or Celsius) and IQ are non-ratio, interval variables, most interval-level variables in the social sciences are also ratio variables. 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.

Some examples of ratio variables include: age, number of years of education, number of times a person has changed residence, income in dollars or other currency, years married, years spent migrating, population size, distance in meters from a house to a well, number of violent crimes per hundred thousand population, number of dentists per million population, number of months since last employment, number of kilograms of fish caught per week, number of hours per week spent in food preparation. Number of years of education is usually treated as a ratio variable, even though a year of grade school is hardly worth the same as a year of graduate school.

In general, concepts (like alienation, political orientation, level of assimilation) are measured at the ordinal level. People get a high score for being “very assimilated,” a low score for being “unassimilated,” and a medium score for being “somewhat assimilated.” When a concept variable like intelligence is measured at the interval level, it is likely to be the focus of a lot of controversy regarding the validity of the measuring instrument.

Concrete observables—things you can actually see—are often measured at the interval level. But not always. Observing whether a woman has a job outside her home is nominal, **qualitative measurement** based on direct observation.

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 \$150,000, between \$150,000 and \$300,000, or more than \$300,000.” If you really want to know how much education people have had, ask them how many years they went to school. Don't ask “Have you completed grade school, high school, some college, four years of college?”

This kind of packaging just throws away information by turning interval-level variables into ordinal ones. As we'll see in Chapter 9, survey questions are pretested before going into a questionnaire. If people won't give you straight answers to straight questions, you can back off and try an ordinal scale. 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 data together into ordinal or nominal categories. If you know the ages of your respondents on a survey, you can divide them into old and young; if you know the number of calories consumed per week for each family in a study, you can divide the data into low, medium, and high. But you cannot do this trick the other way around. If you collect data on income by asking people whether they earn “up to \$75,000 per year” or “more than \$75,000 per year,” you cannot go back and assign actual numbers of dollars to each respondent.

Notice that “up to \$75,000” and “more than \$75,000” is an ordinal variable that *looks like* a nominal variable because there are only two attributes. If the attributes are rankable, then the variable is ordinal. “A lot of fish” is more than “a small amount of fish,” and “highly educated” is greater than “poorly educated.” Ordinal variables can have any number of ranks. For purposes of statistical analysis, though, ordinal scales with five or more ranks are often treated as if they were interval level variables. More about this in Chapter 14 on univariate and bivariate analysis.

VALIDITY, RELIABILITY, ACCURACY, AND PRECISION

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 the data to be accurate to, say, within 5 minutes on, say, 90% of occasions, then the instrument is almost certainly not valid because people just can’t dredge up the information you want at that level of accuracy.

Is the question “Do you practice polytheistic fetishism?” a valid instrument for measuring religious practices? If people don’t understand the question—if you’re asking them to think in categories that are alien to their 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. That conclusion is 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 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. Reliability is one of the biggest problems with using Generative AI in research. Since LLMs (Large Language Models) learn and evolve over time, they do not necessarily produce the same results when experiments or analyses are repeated.

Instruments can be things like thermometers and scales, or they can be questions that you ask people. Like all other kinds of instruments, some survey 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 intercoder reliability—where more than one observer records a measurement—in Chapter 12.

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 was 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. You could stand on it 10 times in a row, without eating or doing exercise in between. That way, you’d be measuring the same thing 10 different times with the same instrument. If the reading were always the same, then the instrument would at least be reliable, even though it wasn’t accurate. Suppose it turned out that your scale was always incorrectly lower by 5 pounds.

This is called **systematic bias**. Then, a simple correction formula would be all you’d need to feel confident that the data from the instrument were pretty close to the truth. The formula would be:

$$\text{True Weight} = \text{Your Scale Weight} + 5 \text{ pounds}$$

The scale might be off in more complicated ways, however. It might be that for every 10 pounds of weight put on the scale, an additional half pound correction has to be made. Then the **recalibration** formula would be

$$\text{True Weight} = (\text{Your Scale Weight}) + (\text{Scale Weight} / 10)(.5)$$

or

$$(\text{Your scale weight}) \times (1.05)$$

That is, take the scale weight, divide by 10, multiply by half a pound, and add the result to the reading on your scale.

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. If it is precise enough for your research and reliable but inaccurate in known ways, then a formula can be applied to correct for the inaccuracy.

The real problem is when instruments are inaccurate in unknown ways. The bad news is that this happens a lot. If you ask people how long it takes them to drive to work, they'll tell you. If you ask people what they ate for breakfast, they'll tell you that, too. Answers to both questions may be dead on target, or they may bear no useful resemblance to the truth. The good news is that respondent accuracy is one of the methodological questions that social scientists have been investigating for years and on which real progress continues to be made (**Further Reading:** respondent and informant 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 about what it's like to break up after a relationship that's lasted more than a year. 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 12 (on grounded theory and content analysis) about how to calculate that score.

2. **Test–retest reliability.** A reliable test of, say, ability in math or of interest in a particular occupation 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. Another test for reliability is the split-half test. This is used in the development of scales, about which *much* more in Chapter 5.

IS MY MEASURE ANY GOOD? DETERMINING VALIDITY

You may have noticed a few paragraphs back that we casually slipped in the statement that some 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 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, asking people “How old were you when you were toilet trained?” is not a valid way to get at this kind of information. A paper-and-pencil test about the rules of the road is not, on the face of it, a valid indicator of whether someone knows how to drive a car. But the paper-and-pencil test is probably a valid test for determining if an applicant for a driver’s license can read road signs. These different instruments—the road test and the paper-and-pencil test—have face validity for measuring different things.

Face validity is based on consensus among researchers: If everyone agrees that asking people “How old are you” is a valid instrument for measuring age, then, until proven otherwise, that question is a valid instrument for measuring age.

Content Validity

Content validity is achieved when an instrument has appropriate content for measuring a complex concept, or construct. 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. Achievement tests—for assessing whether pilots are ready to fly solo and for assessing whether family therapists are ready to be licensed—are judged on their content validity.

Content validity is very, very tough to achieve, particularly for complex, multidimensional constructs. 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.

Religion: Mexican Americans are mostly Roman Catholic, but a growing number of Mexicans are now Protestants. In 2010, 67% of Latinos in the United States identified as Catholics, 12% identified as Protestant or Evangelical, and 10% reported being unaffiliated with any religion. Those numbers are changing fast. In 2013, 55% of Latinos in the United States identified as Catholic, 16% identified as Protestant, and 18% identified as unaffiliated (Pew Research Center 2014).

Language: Some second-generation Mexican Americans speak almost no Spanish; others are completely bilingual. Some use Spanish only in the home; others use it with their friends and business associates.

Socioeconomic status: Many Mexican Americans are poor (about 34% of Hispanic households in the United States have incomes below \$35,000 a year), but many others are well off (about 32% have incomes above \$75,000 a year). People with radically different incomes tend to have different political and economic values (United States Census 2017).

Sense of history: Some so-called Mexican Americans have roots that go back to before the British Pilgrims landed at Plymouth Rock. The Hispanos (as they are known) of New Mexico were Spaniards who came north from the Spanish colony of Mexico. Their self-described ethnic identity is quite different from recent immigrants from Mexico.

Gastronomy: The last refuge of ethnicity is food. When language is gone (Spanish, Yiddish, Polish, Gaelic, Greek, Chinese . . .), and when ties to the “old country” are gone, burritos, bagels, pirogis, corned beef, moussaka, and lo mein remain. For some second-generation Mexican Americans, cuisine is practically synonymous with identity; for others, it’s just part of a much larger complex of traits.

A valid measure of ethnic identity, then, has to get at all these areas. People’s use of Spanish inside and outside the home and their preference for Mexican or Mexican American 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.

“Life satisfaction” is another very complex variable, composed of several concepts like “having sufficient income,” “a general feeling of well-being,” and “satisfaction with level of personal control over one’s life.” In fact, most of the really interesting things that social scientists study are complex constructs—things like “quality of life,” “socioeconomic class,” “ability of teenagers to resist peer pressure to smoke,” and so on (**Further Reading:** measurement in the social sciences).

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—intelligence, ethnicity, machismo, self-esteem, gender, community resilience, food security, alienation, acculturation, liberal—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 is the **gold standard** test, the best one we have.

A tape measure, for example, is known to be an excellent instrument for measuring height. If you knew that a man in the United States wore shirts with 35" sleeves, and pants with 34" cuffs, you could bet that he was over six feet tall and be right more than 95% of the time. On the other hand, you might ask: “Why should I measure his cuff length and sleeve length to know *most of the time, in general*, how tall he is when I could use a tape measure and know *all of the time, precisely* how tall he is?”

Indeed. If you want to measure someone’s height, then use a tape measure. Don’t substitute a lot of proxy variables for something that’s directly measurable by known, valid indicators. But if you want to measure things like quality of life and socioeconomic class—things that don’t have well-understood, valid indicators—then a complex measure will just have to do until something simpler comes along.

Box 1.3 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.”

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 members and on Conservative Political Action Coalition members. Democratic Socialists of America members should get high “left” scores, and Conservative Political Action Coalition 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

Box 1.4 Amber's Corner

OPERATIONALIZING HOUSEHOLD WATER INSECURITY

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. 2024; Wutich 2009b). That finding—along with re-conceptualization 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.

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 really good *measure* of stress, then, ought to predict the likelihood of getting sick.

Remember the life insurance problem? You want to predict whether someone is likely to die over some time period in order to know how much to charge them in premiums. Age and sex tell you a lot. But if you know their weight, whether they smoke, whether they exercise regularly, what their blood pressure is, whether they have ever had any of a list of diseases, and whether they test fly experimental aircraft for a living, then you can predict—with a higher and higher degree of accuracy—whether they will die within any given time period. Each piece of data—each component of a construct you might call “lifestyle”—adds to your ability to predict something of interest.

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. 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, force ranks way up there (after all, it's got several hundred years of theory and experimentation behind it), while 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, acculturation 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

operationalize	intersubjective	accuracy
variable	level of measurement	recalibration
unidimensional	nominal variable	interobserver reliability
multidimensional	ordinal variable	test–retest reliability
dependent variable	interval variable	face validity
independent variable	quantitative measurement	content validity
educational model of social change	ratio variable	construct validity
indicators	zero point	criterion validity
values	qualitative measurement	gold standard
measurement	instrument validity	parsimony
concepts	data validity	Ockham’s razor
constructs	conclusion validity	known group comparison
conceptual definition	reliability	predictive validity
operational definition	precision	

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 measure 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. 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?

2. Common wisdom has it that as people grow older, their idea of how old you have to be to be middle aged changes. To test this hypothesis, produce a copy of Figure 1.3 on a blank sheet of paper and use that instrument to collect some data from people of different ages.
3. (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?
4. 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.

5. Discuss the difference between validity and reliability. Why is it so hard to establish validity?
6. 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.

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 of 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 research. Dingwall et al. (2017), Fisher and Anushko (2008), Hollingshead et al. (2022).

Operationism. Bridgman (1927), Chang (2019), Lundberg (1942, 1964), Vessonen (2021). For a critique: Lincoln and Guba (1985).

Respondent and informant accuracy. An (2022), Bernard, Killworth et al. (1984), Brewer and Yang (1994), Freeman et al. (1987), Godoy et al. (1998), Homburg et al. (2012), McCarty (2021), Roberts et al. (2005),

Romney et al. (1986), Schwarz (1999), Vadez et al. (2003), West and Kreuter (2014).

Validity. Campbell and Fiske (1959), Drost (2011), Fiske (1982).

Measurement in the social sciences. Bandalos (2018), Blalock (1974), Bluemke and Rammstedt (2020), Coombs (1964), Miller and Salkind (2002), Nunnally (1978), Stevens (1946), Walford et al. (2020).

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.

Box 2.1 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.

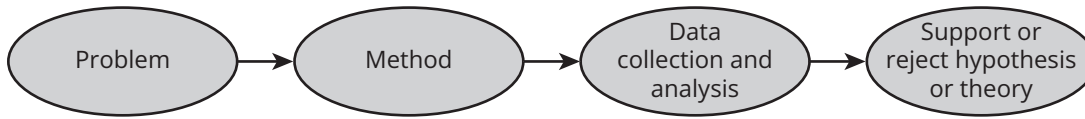
Gender inequality in pay is a social problem that *can* be the basis of many research problems. For example, women earn more in some jobs than men

and gender-diverse people do. What characteristics of individuals (like their gender, age, education, etc.) and the jobs they occupy (indoor vs. 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?

Similarly, bullying in schools is a social problem. Some research questions associated with this topic include: Under what circumstances do kids engage in bullying? What factors account for bullying or for being bullied?

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

Box 2.2 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 coworkers 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 coed 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).

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. Some nursing researchers enjoy working in the maternity ward, while others are drawn to pediatric oncology. The maternity ward is filled with children being born, and the oncology ward with children who are facing death. It doesn't take much to imagine that some people who love working around newborns are going to be less than enthusiastic about doing research in pediatric oncology.

It doesn't matter whether you're going to do experiments, conduct a telephone survey, or do in-depth ethnographic interviews: 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 where you're going. It's really hard to conduct penetrating, in-depth interviews over a period of a 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 Nonsense

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: Does the Old Testament offer support for unequal pay for women today? This is simply not answerable by the scientific method. It is no more answerable than the question: Is Rachmaninoff's music better than Tchaikovsky's? Or: Is it morally correct to mainstream children with slight intellectual disabilities in grades K–6? Or: Is Britain's parliamentary system or the U.S.'s presidential system a better form of democracy? Or: Should the remaining hunting-and-gathering bands of the world be preserved just the way they are and kept from being spoiled by modern civilization? You can shift these into scientifically testable questions—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 (preparing for surgery, putting on makeup, setting the table for Thanksgiving), or of events (weddings, football games, art shows), or of systems of nomenclature (kinds of trucks, ways to avoid getting HIV) 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 12.

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. Gwendolyn Dordick (1996) spent three months studying a homeless shelter for 700 men in New York City. She visited the shelter four times a week for three hours or more each time, and spent four days at the shelter from morning until lights-out at 10 p.m. This was enough time for her to understand a great deal about life in the shelter, including how a group of just 15 men had coalesced into a ruling elite and how some men had formed faux marriages (that could, but did not necessarily, involve sex) to protect themselves and their few possessions from violence and thievery.

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 days or 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

Many things come under the umbrella of money. Equipment is essentially a money issue, as is salary or subsistence for you and other people involved in the research. Funds for assistants, supplies, and travel all have to be calculated before you can actually conduct a major research project. 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.

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. (More about this in Chapter 3.)

Grants for Master’s-level research 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 10x the rate of social science, or more. Increasingly, Principal Investigators (PIs, or study leads) are expected to integrate social science methods into studies of, say, the social impact of a new technology or treatment on marginalized communities. 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, whether cheaper methods might work just as well. If your research requires the comparison of two panels of respondents using face-to-face interviews and you have only enough money to do telephone interviews, ask yourself if you can accomplish your research goal by using telephone interviews. If you can’t, then can you accomplish it by cutting out the comparison and running the more expensive interviews on just one group? 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,” find another topic.

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 competent people 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. 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 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).

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 (i.e., 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 (70%—almost identical to Milgram’s 65% 30 years earlier) 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 Srebrenica—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).

Box 2.3 A Virtual Reality Replication of Milgram's Study

Mel Slater and colleagues (2006) created an immersive virtual reality (VR) experiment to test Milgram's results. The 34 participants sat at a VR console (wearing 3D stereo glasses) and went through all the stages of the original Milgram experiment with a female VR learner. In 23 repetitions of the experiment, the participant saw a virtual avatar of the so-called learner—seeing and hearing all the grunts and screams, right up to that high-voltage shock where the virtual learner said she never signed up for all this and wanted to stop. In the other 11 reps, participants communicated with learners by text.

Just as with the original Milgram experiment, (1) the participants all agreed to take part in this one; (2) knew that they would get paid; and (3) knew that they

could withdraw at any time with no penalty and without giving any reason. In the end, all 11 of the learners who communicated with the learner by text gave all 20 of the possible shocks, but only 17 of the 23 who saw and heard the avatar gave her all the shocks. The difference in the proportions who gave all the shocks is statistically very significant, meaning that, even though all the participants knew that neither the learner nor the shocks were real, the ones who saw and heard the avatar respond to the shocks reacted to her in more-or-less the same way as if she were real.

Besides supporting many of the conclusions of the original Milgram study, this one opens the possibility of conducting experiments on things that it would be impossible to study empirically today.

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. Research using AI is raising ethical questions that once seemed resolved in the social sciences—about deception, data fabrication, and other once-forbidden research practices—leading us to re-examine our established ethical principles.

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 for vaginal or anal sex 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 their 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 Ritchie [2020] 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. Consider the problem of the relation between skin color and various measures of life success, including wealth, health, and longevity (Monk 2015). Would you, personally, be willing to participate in a study of this problem?

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 training, 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. Your PI should teach you the ins and outs of how to fill out the paperwork for your institution's ethics committee. There are also lots of helpful online guides that teach the general steps to getting ethics committee approval for research.
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. You wouldn't say that you will do a survey or experiment and let it go at that. Explain how you will recruit respondents, how and where you will collect data, and so on. Getting the details down 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 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. Getting this can take time—years even—so it is 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 contact the correct agencies and online trainings. An added bonus: Working collaboratively with local scholars will benefit the research community more than simply removing the data and publishing the findings as a lone-wolf.
6. The IRB can review minor changes in a previously approved protocol if you make the change while your project approval is still active. Many social research projects involve an inductive, exploratory phase and a second phase in which hypotheses are tested. If you plan on doing a multi-phase study that includes a survey, you can submit best guesses as to the questions you'll include in the survey

to your IRB. Then, you only have to submit the updated questions for review when you get beyond the inductive phase and into the hypothesis testing phase of your research. Another option is to simply break your multi-stage proposal into multiple IRB applications. This can be more time consuming, but it means you can start immediately with the early stages of the project that you have fully planned.

7. Request permission to recruit and interview more people than you think you'll need in your study to cover all contingencies. What contingencies? Sample attrition is one, but you'll also find that you need to be flexible in recruiting people for in-depth interviews. This may mean changing not only the number of people you have to recruit, but the way you recruit, as well. 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 of payment (like cash or gift cards) 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 recontact 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.

Box 2.4 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 nine 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 United States, 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-step-parent folk tales are based on more than fantasy. Or are they? A lot depends on the paradigm you start with.

Box 2.5 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’s students (McCarty et al. 2021) 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 as they studied new dynamics of gender transitions and detransitions 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.

Alternative Paradigms for Building Theories

One explanation is 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 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**—popularly known as **Darwinism**. (There are several varieties of this, which you'll see under label of evolutionary anthropology or behavioral ecology or evolutionary psychology.)

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.

Evolutionary theory, 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. Evolutionary theory 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 or centuries). The evolutionary explanation for the abuse 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 step-children is not inevitable, so an evolutionary explanation can't account for why some step-parents hurt their children and others don't. A materialist explanation is more productive for addressing that question. Some step-parents who bring a lot of resources to a second marriage become personally frustrated by the possibility of having their wealth raided and diluted by their new spouse's children. The reaction would be strongest for step-parents who have competing obligations to support their biological children who are living with yet another family. These frustrations will cause *some* people to become violent, but not others.

But the materialist explanation doesn't tell us why a particular step-parent is supportive or unsupportive of their 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 step-parents who bring a lot of resources to a second marriage become personally frustrated by the possibility of having their wealth raided and diluted by their new spouse's children. Maybe the step-parent has competing obligations to biological children who are now with yet another family. Perhaps these frustrations cause some people to become violent.

Is there an evolutionary basis for powerful spouses to abuse powerless ones? Or is this something that gets stimulated by material conditions, like poverty? Lots more research is needed on this fascinating question, but 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.

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?

Box 2.6 Case Study

INTIMATE PARTNER VIOLENCE IN 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 man. 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 abused powerless women and the children of powerless women, and powerless women abused their own children.

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 countries to have fewer children? Educate them about the importance of small families. Want to lower the rate of infectious disease in developing countries? Educate people about the importance of good hygiene. Want to get adolescents in Boston or Seattle or wherever to stop having high-risk sex? Educate them about the importance of abstinence or, if that fails, about how to take protective measures against sexually transmitted disease. 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 cultural ideas (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 your carbon footprint 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 the 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. The following examples showcase the contributions of both idealist and materialist perspectives as well as the importance of both idiographic and nomothetic theory.

Big Early Theories in the Social Sciences

Efforts to build nomothetic theory in the social and behavioral sciences are well known. In psychology, Sigmund Freud (1962 [1905]) based his theory that psychosexual development

Box 2.7 Deep Dive

THE ORIGINS OF IDIOGRAPHIC AND NOMOTHEIC THEORIES

The distinction was first made by Wilhelm Windelband, a philosopher of science, in an 1894 lecture at the University of Strasburg (Windelband 1998 [1900]). By the late 1800s, Wilhelm Dilthey's distinction between the *Naturwissenschaften* and *Geisteswissenschaften*—the sciences of nature and the sciences of the mind—had become quite popular. The problem with Dilthey's distinction, said Windelband, was that it couldn't accommodate the then brand-new science of psychology. The subject matter made psychology a *Geisteswissenschaft*, but the discipline relied on the experimental method, and this made it a *Naturwissenschaft*.

What to do? Yes, said Windelband, the search for reliable knowledge is, indeed, of two kinds: the sciences of law and the sciences of events, or, in a memorable turn of phrase, the study of what always is and the study of what once was (Windelband 1998 [1900]:13). Windelband coined the terms *idiographic* and *nomothetic* to replace Dilthey's *Naturwissenschaft* and *Geisteswissenschaften*.

Organic evolution is governed by laws, Windelband observed, but the sequence of organisms on this planet is an event that is not likely to be repeated on any other planet. Languages are gov-

erned by laws, but any given language at any one time is an event in human linguistic life. The goal of the idiographic, or historical sciences, then, is to deliver "portraits of humans and human life with all the richness of their unique forms" (Windelband 1998 [1900]:16).

Windelband went further. Every causal explanation of an event—every idiographic analysis, in other words—requires some idea of how things happen at all. No matter how vague the idea, there must be nomothetic principles guiding idiographic analysis.

Windelband's formulation is a perfect description of what all natural scientists—vulcanologists, ornithologists, astronomers, ethnographers—do all the time. They describe things; they develop deep understanding of the cases they study; and they produce explanations for individual cases based on nomothetic rules. The study of a volcanic eruption, of a species' nesting habits, of a star's death is no more likely to produce new nomothetic knowledge than is the study of a society's adaptation to new circumstances. But the idiographic effort, based on the application of nomothetic rules, is required equally across all the sciences if induction is to be applied and greater nomothetic knowledge achieved.

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).

And in the mid-to-late 19th 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 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 stage (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 19th-century theories of human evolution (Comte 1853).

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. In the mid-20th century, anthropologists and sociologists like Leslie White (1949), Julian Steward (1955), and Gerhard Lenski (1970) advanced more nuanced theories about how cultures evolve.

The Gender Gap in Wages

Across the industrialized world, women tend to earn less than men do for the same work, but there is wide variation in the size of gender wage gap. In the United States and Canada, for example, women earned about 82 cents in 2022 for every dollar men earned. In Korea and Japan, the figure was only 71 and 75 cents on the dollar. In Belgium, Denmark, Greece, and Italy, women earn 94–96 cents on the dollar (OECD 2022).

A theory that explains this gap for the United States is an idiographic theory. A theory that explains why women in most industrial societies earned less than men do for the same work (controlling for currency differences, and differences in cost of living across countries) is more nomothetic. But suppose we are involved in developing legislation on the matter in, say, Sweden or Chile. Then, an idiographic theory—one that takes account of the political and economic realities of a particular country—is what we need.

The Gender Gap in Voting

In 1920, when women got the vote in the United States, politicians were afraid that women would swamp the polls and vote for things like child-support programs. For decades, neither of those fears materialized. By 1954, women were still only 34% of voters and they were not voting in blocs for so-called women's issues. Since 1980, though, the percentage of eligible female voters who have gone to the polls in U.S. presidential elections has exceeded the percentage of eligible male voters who turned out. In this voting gender gap, a greater proportion of women voters than men went for the Democratic candidate (CAWP 2018, 2022; England et al. 2023).

Jeff Manza and Clem Brooks (1998) analyzed data from 11 presidential elections, from 1952 to 1992, to measure and explain what was then the emerging gender gap. Their theory was that since: (1) women are disadvantaged in the labor force (earning less than men do, hitting that glass ceiling in management, and so on); and (2) women depend more on public sector jobs than men do; and (3) women need more help with child care and with welfare than men do; then (4) women's increasing participation in the labor force would naturally drive them toward the Democratic Party and away from the Republican Party in national elections.

It turns out that the gender gap may be more complicated. Kellstedt et al. (2010) analyzed U.S. presidential election data and public policy data from 1980 through 2004 and found that the general tendency is for the U.S. electorate to become more conservative during periods of liberal policy and vice versa. This goes for both men and women, but men respond more quickly and in greater numbers than women do in making this shift. The result: The gender gap increases when public policy becomes more liberal because men become more conservative faster than do women.

And conversely: The gender gap decreases when public policy turns conservative—again, because men react more quickly than do women in moving to the left. This is just what happened in the 2016 presidential election in the United States and in the 2018 mid-term election: More liberal policies during President Obama's term in office, from 2009 to 2016 resulted in men voting more conservatively, and more conservative policies in President Trump's first two years in office produced a large turnout among more liberal women voters.

Manza and Brooks's theory about the rise of the gender gap in U.S. presidential elections doesn't explain why it took women in the United States so long to use their political power in presidential elections, and it doesn't tell us why women didn't put more Democrats into the U.S. House of Representatives prior to 2016. But Manza and Brooks's theory accounted for the facts on the ground in the case they dealt with and that's enough for a theory to do. In any science, a lot of the best work is at the idiographic level of theory making.

The Second Demographic Transition

Demographic transition theory accounts for variations in the average number of children born to women in a society. The first demographic transition happened at the end of the Paleolithic Era, about 12,000 years ago, when people swapped agriculture for hunting and gathering as the main means of production. During the Paleolithic (from about 2.5 million to about 12,000 years ago), population growth was very, very slow. But across the world, as people switched from hunting and gathering to agriculture, children—who could be put to work in agriculture and caring for animals—became a source of wealth accumulation and human populations exploded.

The second demographic transition—in which children ceased to be a direct source of wealth accumulation and the number of children born per woman in a country drops—began in the late 18th century in Europe with industrialization and has been spreading around the world ever since. Today, Japan, Germany, Italy, and other highly industrialized countries have total fertility rates (TFRs) in the neighborhood of 1.5 to 1.2. That's 29% to 43% below the 2.1 TFR needed in any country just to replace the current population.

Demographic transition theory—explaining the link between economic development and lower TFR—is highly nomothetic. It accounts for why Japan, a fully industrialized nation, has such a low TFR. But it doesn't predict what the consequences of that low TFR

will be. For the time being, at least (until even bigger nomothetic theories are developed), we still need an idiographic theory for this.

Japan has about 126 million people—about 40% of the population of the United States—living in an area the size of Montana. Japan has the world's third-largest economy (after the United States and China), and in 2018, the Japanese enjoyed a gross domestic product per capita (GDPPP) of about \$33,000 (IMF 2024). That relatively high level of GDPPP in Japan (it was about \$65,000 for the United States and about \$10,000 for China in 2018) is based on manufacturing products for export—everything from cars to chemicals to oil tankers. Most of the fuel (coal, oil, natural gas) needed to run the factories that produce all those exports has to be imported and so does a lot of the food to feed all those people who are working in the factories.

In 2017, about 27% of Japan's population was over 65 (World Bank 2017), and the percentage of elderly will be closer to 40% by 2050 (NIPSSR 2012:Table 1). That, plus a TFR of 1.4 makes it easy to predict that Japan's industries need to find lots of new workers to maintain productivity—and the lifestyle supported by that productivity.

Belgium and Italy—two other countries with low TFRs—solved this problem by opening their borders to people from the formerly communist countries of eastern Europe and by increasing female participation in the labor force. There was strong resistance to these solutions in Japan, but in 1990, the need for workers prevailed. Japan began offering Brazilians and Peruvians of Japanese descent special visas to resettle on the theory that it would be easy for people who looked Japanese to assimilate. Some 236,000 Brazilians and Peruvians took the offer, but in 2009, after nearly two decades of economic stagnation and the fact that the South Americans weren't becoming Japanese, the Japanese government was offering those hundreds of thousands of immigrants money to go home (Ducanes and Abella 2008:18; Tabuchi 2009).

Japan's economy has begun to recover, and the need for workers has only gotten stronger. This once again challenged the culture of ethnic nationalism in Japan. In 2018, only 23% of Japanese said they thought that more immigrants should be allowed in (Stokes and Devlin 2018), but demography prevailed and in 2019 Japan opened its doors to 345,000 new foreign workers.

The Japanese case shows that lower TFR in industrialized countries doesn't always lead to the same response—at least not in the short run. A nomothetic theory of how industrialized countries react to lower TFR requires a longer time frame, as does one that posits a change in culture (like ideas about ethnic nationalism) in accommodation to changes in the economy, in technology, and in fertility.

Dowry Deaths

In 1977, the New Delhi police reported 311 dowry deaths—deaths by kitchen fires of women, mostly young brides who were killed because their families had not delivered a promised dowry to the groom's family (Claiborne 1984). By 2016, the government of India reported 7,621 (NCRB 2017) such dowry deaths of young women, but this may be an underestimate (Babu and Babu 2011:38; Sanghavi et al. 2009). Even if the incidence were a fraction of what's reported, the phenomenon demands an explanation.

Daniel Gross (1992) theorized that the phenomenon was the consequence of female hypergamy (marrying up) and dowry. Families that can raise a large dowry in India can marry off their daughter to someone of greater means. This created a bidding war as the

families of wealthier sons demanded more and more for the privilege of marrying those sons, and many families of daughters went into debt to accumulate the dowries. When they couldn't pay off the debt, some of the families of grooms murdered the brides in faked "kitchen accidents," where kerosene stoves would purportedly blow up. This would give the grooms' families a chance to get another bride whose families could deliver (Lakhani 2005:254).

Gross's explanation for the kitchen fires in India doesn't explain why other societies that have escalating dowry costs don't have kitchen fires. Nor does it tell us why dowry persists in India despite its being outlawed since 1961, or why dowry—which, after all, only occurs in 7.5% of the world's societies—exists in the first place. But Gross's theory deals effectively with the facts of the case.

There are many lists of *topics* for social research—things like immigration, consumer behavior, racism, social media, gender roles, globalization ... But there is no list of research *questions*. We've discussed just two research questions: What causes dowry deaths? and How does the second demographic transition impact the economy of Japan? 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."

A GUIDE TO FINDING RESEARCH QUESTIONS, ANYWAY

There may not be a *list* of research questions, but there are some useful guidelines for *finding* research questions that might interest you. First of all, there are very few big-theory issues—we call them research arenas—in all of social science. Here are four of them: (1) the nature–nurture problem; (2) the evolution problem; (3) the internal–external problem; and (4) the superorganic, social facts problem.

1. **The nature–nurture problem.** This is an age-old question: How much of our personality and behavior is determined by our genes and how much by our exposure to different environments? Many diseases (cystic fibrosis, Tay-Sachs, sickle-cell anemia) are highly determined by our genes, but others (heart disease, diabetes, asthma, high blood pressure) are at least partly the result of our cultural and physical environment (see Dressler [2004] for a review of cultural risks of disease).

Schizophrenia, for example, may be a genetically inherited disease, but its expression, its treatment, and its outcomes are heavily influenced by our cultural environment. Hallucinations are commonly associated with schizophrenia, but when Robert Edgerton (1966) asked over 500 people in four East African tribes to list the behavior of people who are severely mentally ill, less than 1% of them mentioned hallucinations (see also Edgerton and Cohen 1994; Jenkins and Barrett 2004).

Research on the extent to which differences in cognitive functions of men and women—functions like short- and long-term memory, verbal and quantitative skills, etc.—are the consequence of environmental factors (nurture) or genetic factors (nature) or the interaction between those factors is also part of this research arena (Caplan et al. 1997; Coluccia and Louse 2004; Geary 2020).

2. **The evolution problem.** Studies of how groups change through time from one *kind* of thing to another kind of thing are in this arena. Societies change very slowly through time, but at some point we say that a village has changed into a town or a town into a city or that a society has changed from a feudal to an industrial economy (Katz 1993). All studies of the differences between small societies—*Gemeinschaften*—and big societies—*Gesellschaften*—are in this arena (Miner 1952; Redfield 1947; Tönnies (1999 [1887])). So are studies of inexorable bureaucratization as organizations grow (Meyer 1987; Pegan 2017).
3. **The internal–external problem.** Studies of the way in which behavior is influenced by societal values and by physical environmental conditions are in this arena. Studies of **response effects** in surveys (how people respond differently to the same question asked by a woman or by a man, for example) are in this arena, (Sudman and Bradburn 1974) as are studies of how the weather affects our moods and our behavior. (For more on response effects, see Chapter 8)
4. **The social facts, or emergent properties, problem.** A great deal of social research is based on the assumption that people are influenced by forces that *emerge* from the actions and interactions of humans but that transcend individuals. The name for this problem comes from Emile Durkheim’s (1933 [1893]) seminal observation that social facts exist outside of individuals and are not reducible to individual facts.

For example, the percentage of people in a region who profess a particular religion is a property of the region, not of any individual. Does that regional property exert its own force on people’s beliefs and behavior? The willingness to trust people you don’t know varies across individuals by age, education, race, and religion, but in the United States, at least, it also varies by the “religious ecology” of a geographic region. Specifically, the percent of evangelical Protestants in a city or county predicts a lack of individual-level trust in others—and this holds even for the nonevangelicals in the area, even after taking age, education, and so forth into account (Marshall and Olson 2018:488–89).

GENERATING TYPES OF STUDIES

Now look at Figure 2.2. We have divided research topics (not arenas) into classes, based on the relation among five kinds of variables.

1. **Internal states variables.** These include attitudes, beliefs, values, and perceptions. Cognition is an internal state.
2. **External states variables.** These include characteristics of people, such as age, wealth, health status, height, weight, gender, and so on.

FIGURE 2.2
Types of Studies

	Internal states	External states	Reported behavior	Observed behavior	Artifacts	Environment
Internal states	I	II	IIIa	IIIb	IV	V
External states		VI	VIIa	VIIb	VIII	IX
Reported behavior			Xa	Xb	XIa	XIIa
Observed behavior				Xc	XIb	XIIb
Artifacts					XIII	XIV
Environment						XV

3. **Behavior variables.** This covers what people eat, who they communicate with, how much they work and play—in short, everything that people do and much of what social scientists are interested in understanding.
4. **Artifact variables.** This includes all the physical residue from human behavior—radioactive waste, tomato slicers, sneakers, arrowheads, computer disks, Viagra, skyscrapers—everything.
5. **Environmental variables.** This includes physical, biological and social environmental characteristics. Physical and biological features that influence human thought and behavior include the amount of rainfall, the amount of biomass per square kilometer, location on a river or ocean front. Social environmental characteristics include things like living under a democratic versus an authoritarian régime or working in an organization that tolerates or does not tolerate sexual harassment.

Category (3), behavior, includes both reported and actual behavior. A great deal of research has shown that about a third to a half of everything people report about their behavior is not

Box 2.8 Biological Variables

There is a sixth kind of variable: **biological variables** or indicators, like blood pressure, heart rate, and body mass index. We won't cover this kind of variable here, but biocultural or biosocial research—the interaction among biological, cultural, social, and environmental factors in shaping human thought and human behavior—is a rapidly growing field across the social sciences. For example, Portnoy et al. (2019) found that

a low resting heart rate among female students at a private university was significantly associated with higher levels of self-reported academic dishonesty—but not associated with this for male students. For a review of biocultural research, see Dressler (2005) and see Scholtz (2019) for a review of work on the association of cortisol (in hair and saliva) on measures of stress in daily life.

true (Bernard and Killworth et al. 1984). If you ask children what they eat or how much they exercise, they'll tell you, but their report may have no useful resemblance to what they actually eat or how much they actually exercise (Johnson et al. 1996). If you ask people how many times a year they go to church, you're likely to get data that do not reflect actual behavior (Hadaway and Marler 2005).

Some of the difference between what people say they do and what they do is the result of out-and-out lying. Most of the difference, though, is because people can't hang on to the level of detail about their behavior that is called for when they are confronted by social scientists asking them how often they go to church, or trade stocks, or whatever. What people *think* about their behavior may be precisely what you're interested in, but that's a different matter.

Matching Kinds of Variables and Kinds of Problems

Most social research focuses on internal states and on reported behavior. But the study of humanity can be much richer, once you get the hang of putting together these five kinds of variables and conjuring up potential relations. Here are some examples of possible studies for each of the cells in Figure 2.2.

Cell I: The interaction of internal states, like perceptions, attitudes, beliefs, values, and moods.

Strong beliefs in masculine honor are positively associated with support for war and negatively associated with support for peace-making (Saucier et al. 2018).

Authoritarian personality is associated with higher levels of support for gun control among women and lower levels of support among men (Lizotte 2019).

This cell is also filled with studies that compare internal states across groups. See, for example, Cooke's (2004) study of attitudes toward gun control among American, British, and Australian youth and Yarrow et al.'s (2006) study comparing the early development of implicit racial prejudice in rural Japan and urban United States.

Cell II: The interaction of internal states (perceptions, beliefs, moods, etc.) and external states (completed education, health status, organizational conditions).

Firms that have demographically diverse boards of directors (boards with four or more seats held by women, racial minorities and/or the disabled) are more likely to offer employees progressive programs (like exceptional work-life benefits and cash profit-sharing) and have higher employee satisfaction (Creek et al. 2019).

Variations in organizational structure correlate with employee satisfaction (Gregory et al. 2009).

Attitudes about the price of food among women of different socioeconomic and health statuses (Bowman 2006).

Cell IIIa: The interaction between *reported* behavior and internal states.

Perception of how well the economy is doing and reported voting behavior (Kwon 2010).

In Sardinia, Italy, people who experience urban stress—from pollution, traffic—and who feel a strong identity with their city are much more likely to report environment-friendly behaviors—like not littering, using public transport, not wasting water (Meloni et al. 2019). (And see Bamberg and Moeser [2007] for more on psychological determinants of pro-environmental behavior).

Reported rate of alcohol consumption and attitudes toward gender roles (Christie-Mizell and Peralta 2009).

Cell IIIb: The interaction between *observed* behavior and internal states.

How attitudes and beliefs about resources are related to actual behavior in the control of a household thermostat (Kempton 1987).

Behavioral indicators of nervousness among men at an STI clinic (Lichtenstein 2004).

Cell IV: The interaction of material artifacts and internal states.

People who own guns in the United States are significantly less likely than nongun owners to attribute shootings to the availability of guns and to say that mass shootings do not impact their support for gun laws (Joslyn and Haider-Markel 2017).

The effects on Holocaust Museum staff in Washington, DC, of working with the physical reminders of the Holocaust (McCarroll et al. 1995).

How young children in New Zealand learn gender roles from pictures in early school readers (Jackson and Gee 2005).

Cell V: The interaction of social and physical environmental factors and internal states.

As homes become more crowded, parents are less responsive to their children (Evans et al. 2010).

Children in Africa who witness violence are likely to develop symptoms of stress, but a positive school climate reduces the effect (O'Donnell et al. 2011).

The design of memorials can help people recover from trauma and loss (Watkins et al. 2010).

Cell VI: How the interaction among external states relates to outcomes, like longevity or financial success.

Immigrants in 11 Western European countries who obtained a college degree in a non-Western country wind up with significantly lower wages than immigrants who get their degree in a Western country (Lancee and Bol 2017). Among children in the United States living in poverty, the role of income on birth weight decreases as the as the degree of poverty decreases. The effect of income inequality on birth weight, however, increases (Olson et al. 2010).

Cell VIIa: The relation between external states and *reported* behavior.

The impact of gender, area of residence, and religious affiliation on the likelihood of attending church as people get older (Schwadel 2010).

Factors affecting self-reported suicidal behavior among adolescents in Ireland (McMahon et al. 2010).

In the Netherlands, people of Turkish and Moroccan origin consumed higher quality diets and at the lowest cost compared to native Dutch respondents (Mackenbach et al. 2019).

Cell VIIb: The relation between external states and *observed* behavior.

Ethnicity of clientele, gender of bartender, and other factors associated with smoking in bars, despite laws against smoking (Moore et al. 2009).

Across 33,000 observations in four grocery stores in Sydney and Adelaide, Australia, 17% of shoppers were accompanied by children. Shoppers with children spent between 5 and 10 minutes less than shoppers without children to buy the same number of items (Page et al. 2018).

Cell VIII: The relation of physical artifacts and external states.

How age and gender differences relate to acceptance or disapproval of talking on a cell phone at a family dinner acceptable in five countries: the United States, Sweden, Italy, Japan, and South Korea (Baron and Campbell 2012).

How age and gender differences relate to cherished possessions among children and adolescents from 6 to 18 years of age (Dyl and Wapner 1996).

Cell IX: The relation of external states and environmental conditions.

The effect of neighborhood street culture on violence among adolescents, beyond that predicted by individual values (Stewart and Simons 2010).

How poor physical and social conditions of poor neighborhoods contribute to bad health and early death (Cohen et al. 2003)

Cell Xa: The relation between behaviors, as *reported* by people to researchers.

The relation of self-reported level of church attendance and self-reported “hooking up” among American college women (Burdette et al. 2009).

Adolescents are more likely to drink alcohol regularly if they report that their same-sex twin or their friends are drinkers than if their parents report regular drinking (Scholte et al. 2008).

Cell Xb: The relation between *reported* behaviors and behaviors *observed* by researchers.

Self-reported recycling behaviors predict observed recycling behaviors weakly ($r = .22$) (Huffman et al. 2014:266)

Almost 90% of Latino/a migrant workers on tobacco farms in North Carolina report washing their hands before eating, compared to 17% observed doing so (Walton et al. 2016:278–79).

Cell Xc: The relation among two or more directly observed behaviors

Ordering cocktails in rounds and drinking cocktails with a straw both increase the amount of time people in Queensland, Australia spent consuming alcohol (Buyucek et al. 2017).

In both New York City and Flagstaff, Arizona, pedestrians who didn’t have to wait at an intersection for a “walk” signal were less likely to text than were those who had to wait even one second. But pedestrians in New York City were less likely than those in Arizona to start walking during the flashing “don’t walk” phase of a crossing signal (Russo et al 2018:8–9).

Box 2.9 Direct Observation

Direct observation and comparison of behaviors is used in many fields, including psychology (in the assessment of behavioral disorders), in education (in assessing learning disabilities), in nursing (in assessing patients’ dietary habits), in political science

(in assessing the bellicosity of nations toward one another), in occupational sociology (in assessing ergonomics and performance), and so on. We’ll discuss direct observation at some length in Chapter 11, on fieldwork.

Cell XIa: The relation of observed behavior to specific physical artifacts.

Content analysis of top-grossing films from 1950 to 2006 shows that the portrayal of tobacco use declined proportionate to the actual decline of smoking in the population (Jamieson and Romer 2010).

People who respond to telephone interviews by cell phone speak slower and take longer to complete the interview than people who respond on a landline (Timbrook et al. 2018:566–67).

Cell XIb: The relation of reported behavior to specific physical artifacts.

People who are employed view prized possessions as symbols of their own personal history, while people who are unemployed see prized possessions as having utilitarian value (Dittmar 1991).

Cell XIIIa: The relation of reported behavior to factors in the social or physical environment.

The relation of compulsive consumer behavior in young adults and whether they were raised in intact or disrupted families (Rindfleisch et al. 1997).

Cell XIIb: The relation of observed behavior to factors in the social or physical environment.

In 2016, across eight countries in the European Union, and despite legislation against smoking in public places, smoking continued in local bars, especially those that a counter with barstools, slot machines, a lack of outside seating and no food service (van Beek et al. 2018).

The influence of environmental factors (one-way vs. two-way traffic, the presence or absence of a specific pedestrian signal, number of lanes in a road, etc.) on pedestrians obeying a traffic signal in Montreal (Cambon de Lavalette et al. 2009).

Spend more or less time in a store and spend more or less money, depending on factors in the store environment (Sherman et al. 1997).

Cell XIII: The association of physical artifacts to one another and what this predicts about human thought or behavior.

Comparing the favorite possessions of urban Indians (in India) and Indian immigrants to the United States to see whether certain sets of possessions remain meaningful among immigrants (Mehta and Belk 1991).

This is also an example of Cell IV. Note the difference between expressed *preferences* across artifacts and the coexistence of artifacts across places or times.

Cell XIV: The probability that certain artifacts (relating, for example, to subsistence) will be found in certain physical or social environments (rain forests, deserts, shoreline communities). This area of research is mostly the province of archeology.

Cell XV: How features of the social and physical environment interact and affect human behavioral and cognitive outcomes.

Environmental features of offices, like amount cubicle privacy, lighting and noise, affect job satisfaction and worker performance (Goins et al. 2010; Newsham et al. 2009).

Social and physical environmental features of retail stores interact to affect the buying behavior of consumers (Baker et al. 1992).

The above list is only meant to give you an idea of how to think about potential covariations and, consequently, about potential research topics. There are two things to keep in mind about all this: (1) *covariation does not mean cause*. Covariation can be spurious, the result of an antecedent or an intervening variable. (See Chapter 3 for a discussion of causality, spurious relations, and antecedent variables.) And (2) many of the examples in the list above are statements about possible bivariate correlations—that is, they are about possible covariation between two things. Social phenomena being the complex sorts of things they are, a lot of research involves multivariate relations—that is, covariation among three or more things at the same time.

For example, it's well known that people who call themselves religious conservatives in the United States are likely to support the National Rifle Association's gun policies (Cell I). But the association between the two variables (religious beliefs and attitudes toward guns) is by no means perfect and is affected by many intervening variables.

We'll tell you about testing for bivariate relations in Chapter 14 and about testing for multivariate relations in Chapter 15. As in so many other things, you crawl before you run, and you run before you fly.

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 until you know what everyone else has done. You can't delegate your literature search to AI either—research questions, literature reviews, and even text summaries generated by AI have a risk of turning your work into “junk science,” riddled with errors and derivative ideas. 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 is 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 citations, 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 get back a bibliometric network (co-citation map), plus a list of backward and forward citations. Many university libraries have lists of the latest tools, with reviews of how well they perform. As these tools are still being refined, it's best to use them in combination with an old-fashioned literature review.

All you need is a few key references to get started. Don't worry about the key references being out of date. Google Scholar, the *Web of Science*, and other bibliography search tools eliminate the problem of obsolescence in bibliographies. Read on to see why ...

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*. This set of indexes, available at most university libraries and in many small college libraries, covers about 10,000 journals, including about 3,500 in the social sciences. The title, author, journal, year, and page numbers for every article goes into the database, along with the email address of the corresponding author, when it's available.

Other Documentation Databases

These days, documentation is a robust business, and there are many indexing and abstracting resources. Besides Google Scholar and the citation indexes, some important resources for social scientists are: PsycINFO; Sociological Abstracts; ERIC; NTIS and FDsys; LEXIS NEXIS OCLC and WorldCat, PubMed and MEDLINE. 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.

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.

Box 2.10 Tips & Tricks

CONTACTING RESEARCHERS

Another tip: Contact people on listservs and networking groups that deal with your research topic. If there are central figures in the field, contact them by email and request a time when you can call them. Yes, by phone. Or video call. Email 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.

Sociological Abstracts and Dissertation Abstracts

Sociological Abstracts (SA) is one of many databases offered by ProQuest, including dissertation and theses abstracts. SA indexes and abstracts about 1,800 journals dating from 1952, with excellent coverage of research methods, the sociology of language, occupations and professions, health, family violence, poverty, and social control. It covers the sociology of knowledge and the sociology of science as well as the sociology of the arts, religion, and education, and also includes Social Services Abstracts (which focuses on literature about social work).

Much of the best and most up-to-date research is done by graduate students. If your institution subscribes to this database, be sure to check it out when you do the background reading for your own project.

ERIC

The Educational Resources Information Center (ERIC) is available free at <http://www.eric.ed.gov/>. It covers literature since 1966 of interest to researchers in education, but many of the over 1,000 journals in the database are of interest to all social scientists. The ERIC database includes a lot of grey literature—government reports and reports from private foundations and industries that contain useful information but can be tough to find. The ERIC database contained 1.6 million records in 2019 and is continually updated.

NTIS and FDsys

The National Technical Information Service (NTIS) indexes and abstracts federally funded research reports in all areas of science. It's available free at <http://www.ntis.gov/>. If you have a research contract with a U.S. government agency, you'll be asked to file regular technical reports on the work you do as you go along. Those technical reports get logged in to the NTIS. Many technical reports later get published as articles.

But many don't. Some of the reports aren't published because they are too preliminary, but lots of technical reports don't get published because they contain huge tables of basic data. That's not the stuff that journals can publish, but it may be pure treasure for another researcher. It used to be that reports on government contracts were filed and then shelved, never to be heard from again. But with the NTIS database, the public can now easily locate all that information.

The NTIS has technical reports about alcohol and drug abuse, violence against women, Native American health, public and private prisons, new designs for low-cost housing, as well as laboratory experiments on how much people might be willing to pay for gasoline, reports of archeological digs, natural experiments to test how long people can stay in a submerged submarine without going crazy, and so on. If the federal government has funded research under contract, there's probably a technical report about it.

These reports can be accessed through the Federal Digital System (FDsys) available through the main U.S. government information site: <https://www.govinfo.gov/>.

LEXIS/NEXIS

If your library has LEXIS/NEXIS, don't consider any literature search complete until you've used this database. The system began in 1973 as a way to help lawyers find information on cases. Today, the database contains the searchable text of billions of documents, including the major English-language newspapers in the world.

OCLC and WorldCat

The OCLC (Online Computer Library Center) is the world's largest library database. Over 17,000 member libraries in 123 countries catalog their holdings, in 490 languages, in OCLC's catalog, called WorldCat. The system had 550 million bibliographic records in 2023. If you find a book or a journal article in the Social Science Citation Index or PsycINFO, and so forth, and your library doesn't have it, then OCLC will tell you which library *does* have it.

Interlibrary loans depend on OCLC. In addition, OCLC publishes a database called ArticleFirst. This leviathan, which is updated daily, covers 16,000 journals in all fields, including many in the social sciences and contained over 34 million records in 2023.

PubMed and MEDLINE

PubMed is a product of the National Library of Medicine (National Institutes of Health) and is available free at <http://www.ncbi.nlm.nih.gov/pubmed/>. This database covers about 5,400 journals in the medical sciences, including the medical social sciences. It contained over 35 million citations in 2023 and is continually updated.

If you are working on anything that has to do with health care, PubMed is a must. In 2023, we asked PubMed for articles on “high-risk sexual behavior and adolescents” and it returned a list of 4,092 items.

HRAF: The Human Relations Area Files

Founded in 1949, eHRAF World Cultures is an online archive of publications and reports covering 2,500 world cultures. The records are pre-coded using hundreds of entries from the Outline of Cultural Materials. If, for example, you want to find ethnographic descriptions of tattooing from every world culture, you can find this in minutes in the eHRAF. Every entry is an ethnographic publication you can cite and use for backward and forward citation.

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? That is, do people who get psychotherapy benefit, compared to people who have the same problems and who don't get psychotherapy? 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 data-

base from 1967 on for articles that contained *any* of the terms “sexual abuse,” “child abuse,” “rape,” “sex,” and “juror.” They also posted requests on 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. The 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

research problem	theoretical paradigms	external states variables
social problem	levels of analysis	behavior variables
research process	time scale	artifact variables
ethical research	educational model of social change	environmental variables
code of ethics	idiographic theory	biological variables
cultural relativism	nomothetic theory	backward citation
moral relativism	the nature–nurture problem	forward citation
Institutional Review Boards	the evolution problem	literature mapping
theoretical contribution	the internal–external problem	meta-analysis
evolutionary theory (Darwinism)	the social facts, or emergent properties, problem	
idealist theory	internal states variables	
materialist theory		

SUMMARY

- The process of 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 Milgram's experiments on obedience would be funded today, yet the lessons from his experiments continue to provide guidance on the responsibility of the individual for their 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.
 - The three main paradigms for explanation are evolutionary theory, idealism, and materialism. Evolutionary theory (popularly called 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.
- 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 1,685 (yes, 1,685) and 2,011. 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. Use Figure 2.2 to think up some research problems. Think about how you would operationalize the variables for each study you think up. Go to the library and see if you can find any studies on the research problems you come up with.
3. 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?

4. 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 of taken by the authors.

FURTHER READING

Milgram's obedience experiments. Benjamin and Simpson (2009), Bridgman and Cummings (2023), Burger (2009), Kaposi (2021), Packer (2008), Perlstadt (2013).

Ethical issues in social science. Boruch and Cecil (1983), Bosk (2004), Burgess (1989), Citro et al. (2003), Fluehr-Lobban (1996), Hammersley (2009), Herrera (1996), Hosseini et al. (2022), Keith-Spiegel and Koocher (2005), Lyman (1989), Mertens and Ginsberg (2009),

van den Hoonaard (2002), Weinhardt (2021), Weisstub (1998), Weisstub and Diaz Pintos (2007).

Meta-analysis. Andrade (2020), Cook et al. (1992), Cooper et al. (2009), Farley and Lehman (1986), Glass (1976), Guzzo et al. (1987), Hansen et al. (2022), Hedges and Olkin (1985), Hunt (1997), Hunter and Schmidt (2004), Matt and Navarro (1997), Pan (2008), Rosenthal (1984), Wolf (1986).