Sociologists often talk about reality as a social construction. What they mean is that truth and knowledge are discovered, communicated, reinforced, and changed by members of society. Truth doesn’t just fall from the sky and hit us on the head. What is considered truth or knowledge is specific to a given culture. All cultures have specific rules for determining what counts as good and right and true. As social beings, we respond to our interpretations and definitions of situations, not to the situations themselves. We learn what sorts of interpretations are expected and reasonable from our cultural environment. Thus, we make sense of situations and events in our lives by applying culturally shared definitions and interpretations. In this way, we distinguish fact from fantasy, truth from fiction, myth from reality. This process of interpretation or “meaning making” is tied to interpersonal interaction, group membership, culture, history, power, economics, and politics.

Discovering truth and determining useful knowledge are the goals of any academic discipline. The purpose of an academic field such as sociology is to provide the public with useful and relevant information about how society works. This task is typically accomplished through systematic social research—experiments, field research, unobtrusive observation, and surveys. But gathering trustworthy data can be difficult. People sometimes lie or have difficulty recalling past events in their lives. Sometimes the simple fact of observing people’s behavior changes that behavior. And, as you saw in the previous chapter, sometimes the information needed to answer questions about important, controversial issues is hard to obtain without raising ethical issues.

Moreover, sometimes the very characteristics and phenomena we’re interested in understanding are difficult to observe and measure. Unlike other disciplines in, say, the natural sciences, sociologists deal with concepts that can’t be seen and touched. In “Concepts, Indicators, and Reality,” Earl Babbie gives us a brief introduction to some of the problems researchers face when they try to transform important, but abstract, concepts into indicators (things that researchers can systematically quantify so they can generate statistical information). In so doing, he shows us that although sociologists provide us with useful empirical findings about the world in which we live, an understanding of the measurement difficulties they face will provide us with the critical eye of an informed consumer as we go about digesting research information.

In a similar vein, Howard Schuman, in “Sense and Nonsense About Surveys,” provides us with some guidance about how to interpret the information the media provide us everyday. In particular, he focuses on the pitfalls of poor sampling and the dilemmas raised by poorly worded survey questions. In doing so, Schuman provides us with useful guidance on how to interpret the information that most people unquestioningly accept as truth. As a result, we become more informed consumers of socially constructed reality.
Something to Consider as You Read

Babbie’s and Schuman’s comments remind us that even scientists must make decisions about how to interpret information. Thus, scientists, working within academic communities, define truth and knowledge. This knowledge is often significant and useful, but we need to remember that it is the construction of a group of people following particular rules, not something that is just “out there.” As you read these selections, think about the kind of information you would need or would want that might convince you to question some truth that you have always taken for granted.
Measurement is one of the fundamental aspects of social research. When we describe science as logical/empirical, we mean that scientific conclusions should (1) make sense and (2) correspond to what we can observe. It is the second of these characteristics I want to explore in this essay.

Suppose we are interested in learning whether education really reduces prejudice. To do that, we must be able to measure both prejudice and education. Once we’ve distinguished prejudiced people from unprejudiced people and educated people from uneducated people, we’ll be in a position to find out whether the two variables are related.

Social scientific measurement operates in accordance with the following implicit model:

- Prejudice exists as a variable: some people are more prejudiced than others.
- There are numerous indicators of prejudice.
- None of the indicators provides a perfect reflection of prejudice as it “really” is, but they can point to it at least approximately.
- We should try to find better and better indicators of prejudice—indicators that come ever closer to the “real thing.”

This model applies to all of the variables social scientists study. Take a minute to look through the following list of variables commonly examined in social research.

<table>
<thead>
<tr>
<th>Concepts</th>
<th>Indicators</th>
<th>Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arms race</td>
<td>Tolerance</td>
<td></td>
</tr>
<tr>
<td>Religiosity</td>
<td>Fascism</td>
<td></td>
</tr>
<tr>
<td>Urbanism</td>
<td>Parochialism</td>
<td></td>
</tr>
<tr>
<td>TV watching</td>
<td>Maturity</td>
<td></td>
</tr>
<tr>
<td>Susceptibility</td>
<td>Solidarity</td>
<td></td>
</tr>
<tr>
<td>Stereotyping</td>
<td>Instability</td>
<td></td>
</tr>
<tr>
<td>Anti-Semitism</td>
<td>Education</td>
<td></td>
</tr>
<tr>
<td>Voting</td>
<td>Liberalism</td>
<td></td>
</tr>
<tr>
<td>Dissonance</td>
<td>Authoritarianism</td>
<td></td>
</tr>
<tr>
<td>Pessimism</td>
<td>Race</td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>Happiness</td>
<td></td>
</tr>
<tr>
<td>Revolution</td>
<td>Powerlessness</td>
<td></td>
</tr>
<tr>
<td>Alienation</td>
<td>Mobility</td>
<td></td>
</tr>
<tr>
<td>Social class</td>
<td>Consistency</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Delinquency</td>
<td></td>
</tr>
<tr>
<td>Self-esteem</td>
<td>Compassion</td>
<td></td>
</tr>
<tr>
<td>Idealism</td>
<td>Democracy</td>
<td></td>
</tr>
<tr>
<td>Prestige</td>
<td>Influence</td>
<td></td>
</tr>
</tbody>
</table>
Even if you’ve never taken a course in social science, many of these terms are at least somewhat familiar to you. Social scientists study things that are of general interest to everyone. The nuclear arms race affects us all, for example, and it is a special concern for many of us. Differences in religiosity (some of us are more religious than others) are also of special interest to some people. As our country has evolved from small towns to large cities, we’ve all thought and talked more about urbanism—the good and bad associated with city life. Similar interests can be identified for all of the other terms.

My point is that you’ve probably thought about many of the variables mentioned in the list. Those you are familiar with undoubtedly have the quality of reality for you: that is, you know they exist. Religiosity, for example, is real. Regardless of whether you’re in favor of it, opposed to it, or don’t care much one way or the other, you at least know that religiosity exists. Or does it?

This is a particularly interesting question for me, since my first book, *To Comfort and to Challenge* (with Charles Glock and Benjamin Ringer), was about this subject. In particular, we wanted to know why some people were more religious than others (the sources of religiosity) and what impact differences in religiosity had on other aspects of life (the consequences of religiosity). Looking for the sources and consequences of a particular variable is a conventional social scientific undertaking; the first step is to develop a measure of that variable. We had to develop methods for distinguishing religious people, nonreligious people, and those somewhere in between.

The question we faced was, if religiosity is real, how do we know that? How do we distinguish religious people from nonreligious people? For most contemporary Americans, a number of answers come readily to mind. Religious people go to church, for example. They believe in the tenets of their faith. They pray. They read religious materials, such as the Bible, and they participate in religious organizations.

Not all religious people do all of these things, of course, and a great deal depends on their particular religious affiliation, if any. Christians believe in the divinity of Jesus; Jews do not. Moslems believe Mohammed’s teachings are sacred; Jews and Christians do not. Some signs of religiosity are to be found in seemingly secular realms. Orthodox Jews, for example, refrain from eating pork; Seventh-Day Adventists don’t drink alcohol.

In our study, we were interested in religiosity among a very specific group: Episcopal church-members in America. To simplify our present discussion, let’s look at that much narrower question: How can you distinguish religious from nonreligious Episcopalians in America?

As I’ve indicated above, we are likely to say that religious people attend church, whereas nonreligious people do not. Thus, if we know someone who attends church every week, we’re likely to think of that person as religious; indeed, religious people joke about church-members who only attend services on Easter and at Christmas. The latter are presumed to be less religious.

Of course, we are speaking rather casually here, so let’s see whether church attendance would be an adequate measure of religiosity for Episcopalians and other mainstream American Christians. Would you be willing to equate religiosity with church attendance? That is, would you be willing to call religious everyone who attended church every week, let’s say, and call nonreligious everyone who did not?

I suspect that you would not consider equating church attendance with religiosity a wise policy. For example, consider a political figure who attends church every Sunday, sits in the front pew, puts a large contribution in the collection plate with a flourish, and by all other evidence seems only interested in being known as a religious person for the political advantage that may entail. Let’s add that the politician in question regularly lies and cheats, exhibits no Christian compassion toward others, and ridicules religion in private. You’d probably consider it inappropriate to classify that person as religious.
Now imagine someone confined to a hospital bed, who spends every waking minute reading in the Bible, leading other patients in prayer, raising money for missionary work abroad—but never going to church. Probably this would fit your image of a religious person.

These deviant cases illustrate that, while church attendance is somehow related to religiosity, it is not a sufficient indicator in and of itself. So how can we distinguish religious from nonreligious people?

Prayer is a possibility. Presumably, people who pray a lot are more religious than those who don’t. But wouldn’t it matter what they prayed for? Suppose they were only praying for money. How about the Moslem extremist praying daily for the extermination of the Jews? How about the athlete praying for an opponent to be hit by a truck? Like church attendance, prayer seems to have something to do with religiosity, but we can’t simply equate the two.

We might consider religious beliefs. Among Christians, for example, it would seem to make sense that a person who believes in God is more religious than one who does not. However, this would require that we consider the person who says, “I’ll believe anything they say just as long as I don’t rot in Hell” more religious than, say, a concerned theologian who completes a lifetime of concentrated and devoted study of humbly concluding that who or what God is cannot be known with certainty. We’d probably decide that this was a misclassification.

Without attempting to exhaust all the possible indicators of religiosity, I hope it’s clear that we would never find a single measure that will satisfy us as tapping the real essence of religiosity. In recognition of this, social researchers use a combination of indicators to create a composite measure—an index or a scale—of variables such as religiosity. Such a measure might include all of the indicators discussed so far: church attendance, prayer, and beliefs.

While composite measures are usually a good idea, they do not really solve the dilemma I’ve laid out. With a little thought, we could certainly imagine circumstances in which a “truly” religious person nonetheless didn’t attend church, pray, or believe, and we could likewise imagine a nonreligious person who did all of those things. In either event, we would have demonstrated the imperfection of the composite measure.

Recognition of this often leads people to conclude that variables like religiosity are simply beyond empirical measurement. This conclusion is true and false and even worse.

The conclusion is true in that we can make any measurement we want. For example, we can ask people if they attend church regularly and call that a measure of religiosity just as easily as Yankee Doodle called the feather in his hat macaroni. In our case, moreover, most people would say that what we’ve measured is by no means irrelevant to religiosity.

The conclusion is false in that no empirical measurement—single or composite—will satisfy all of us as having captured the essence of religiousness. Since that can never happen, we can never satisfactorily measure religiosity.

The situation is worse than either of these comments suggests in that the reason we can’t measure religiosity is that it doesn’t exist! Religiosity isn’t real. Neither is prejudice, love, alienation, or any of those other variables. Let’s see why.

There’s a very old puzzle I’m sure you’re familiar with: when a tree falls in the forest, does it make a sound if no one is there to hear it? High school and college students have struggled with that one for centuries. There’s no doubt that the unobserved falling tree will still crash through the branches of its neighbors, snap its own limbs into pieces, and slam against the ground. But would it make a sound?

If you’ve given this any thought before, you’ve probably come to the conclusion that the puzzle rests on the ambiguity of the word sound. Where does sound occur? In this example, does it occur in the falling tree, in the air, or in the ear of the beholder? We can be
reasonably certain that the falling tree generates turbulent waves in the air; if those waves in the air strike your ear, you will experience something we call hearing. We say you've heard a sound. But do the waves in the air per se qualify as sound?

The answer to this central question is necessarily arbitrary. We can have it be whichever way we want. The truth is that (1) a tree fell; (2) it created waves in the air; and (3) if the waves reached someone’s ear, they would cause an experience for that person. Humans created the idea of sound in the context of that whole process. Whenever waves in the air cause an experience by way of our ears, we use the term sound to identify that experience. We’re usually not too precise about where the sound happens: in the tree, in the air, or in our ears.

Our imprecise use of the term sound produces the apparent dilemma. So what’s the truth? What’s really the case? Does it make a sound or not? The truth is that (1) a tree fell; (2) it created waves in the air; and (3) if the waves reached someone’s ear, they would cause an experience for that person. That’s it. That’s the final and ultimate truth of the matter.

I’ve belabored this point, because it sets the stage for understanding a critical issue in social research—one that often confuses students. To move in the direction of that issue, let’s shift from sound to sight for a moment. Here’s a new puzzle for you: are the tree’s leaves green if no one is there to see them? Take a minute to think about that, and then continue reading.

Here’s how I’d answer the question. The tree’s leaves have a certain physical and chemical composition that affects the reflection of light rays off of them; specifically, they only reflect the green portion of the light spectrum. When rays from that portion of the light spectrum hit our eyes, they create an experience we call the color green.

“But are the leaves green if no one sees them?” you may ask. The answer to that is whatever we want it to be, since we haven’t specified where the color green exists: in the physical/chemical composition of the leaf, in the light rays reflected from the leaf, or in our eyes.

While we are free to specify what we mean by the color green in this sense, nothing we do can change the ultimate truth, the ultimate reality of the matter. The truth is that (1) the leaves have a certain physical and chemical composition; (2) they reflect only a portion of the light spectrum; and (3) that portion of the light spectrum causes an experience if it hits our eyes. That’s the ultimate truth of the universe in this matter.

By the same token, the truth about religiosity is that (1) some people go church more than others; (2) some pray more than others; (3) some believe more than others; and so forth. This is observably the case.

At some point, our ancestors noticed that the things we’re discussing were not completely independent of one another. People who went to church seemed to pray more, on the whole, than people who didn’t go to church. Moreover, those who went to church and prayed seemed to believe more of the church’s teachings than did those who neither went to church nor prayed. The observation of relationships such as these led them to conclude literally that “there is more here than meets the eye.” The term religiosity was created to represent the concept that all the concrete observables seemed to have in common. People gradually came to believe that the concepts were real and the “indicators” only pale reflections.

We can never find a “true” measure of religiosity, prejudice, alienation, love, compassion, or any other such concepts, since none of them exists except in our minds. Concepts are “figments of our imaginations.” I do not mean to suggest that concepts are useless or should be dispensed with. Life as we know it depends on the creation and use of concepts, and science would be impossible without them. Still, we should recognize that they are fictitious, then we can trade them in for more useful ones whenever appropriate.
THINKING ABOUT THE READING

Define the following terms: “poverty,” “happiness,” “academic effort,” “love.” Now consider what indicators you would use to determine people's levels of each of these concepts. The indicator must be something that will allow you to clearly determine whether or not someone is in a particular state (such as, poor or not poor; happy or not happy; in love or not in love). For example, you might decide that “blushing” in the presence of someone is one indicator of being “in love” or that the number of hours a person spends studying for a test is an indicator of “academic effort.” What’s wrong with simply asking people if they’re poor, if they’re in love, if they’re happy, or if they work hard? Consider the connection between how a concept is defined and how it can be measured. Is it possible that sociology sometimes uses concepts that seem meaningless because they are easier to “see” and measure?
Surveys draw on two human propensities that have served us well from ancient times. One is to gather information by asking questions. The first use of language around 100,000 years ago may have been to utter commands such as “Come here!” or “Wait!” Questions must have followed soon after: “Why?” or “What for?” From that point, it would have been only a short step to the use of interrogatives to learn where a fellow hominid had seen potential food, a dangerous animal, or something else of importance. Asking questions continues to be an effective way of acquiring information of all kinds, assuming of course that the person answering is able and willing to respond accurately.

The other inclination, learning about one’s environment by examining a small part of it, is the sampling aspect of surveys. A taste of something may or may not point to appetizing food. A first inquiry to a stranger, a first glance around a room, a first date—each is a sample of sorts, often used to decide whether it is wise to proceed further. As with questions, however, one must always be aware of the possibility that the sample may not prove adequate to the task.

Sampling: How Gallup Achieved Fame

Only within the past century—and especially in the 1930s and 1940s—were major improvements made in the sampling process that allowed the modern survey to develop and flourish. A crucial change involved recognition that the value of a sample comes not simply from its size but also from the way it is obtained. Every serious pursuit likes to have a morality tale that supports its basic beliefs: witness Eve and the apple in the Bible or Newton and his apple in legends about scientific discovery. Representative sampling has a marvelous morality tale also, with the additional advantage of its being true.

The story concerns the infamous Literary Digest poll prediction—based on 10 million questionnaires sent out and more than two million received back—that Roosevelt would lose decisively in the 1936 presidential election. At the same time, George Gallup, using many fewer cases but a much better method, made the more accurate prediction that FDR would win. Gallup used quotas in choosing respondents in order to represent different economic strata, whereas the Literary Digest had worked mainly from telephone and automobile ownership lists, which in 1936 were biased toward wealthy people apt to be opposed to Roosevelt. (There were other sources of bias as well.) As a result, the Literary Digest poll disappeared from the scene, and Gallup was on his way to becoming a household name.

Yet despite their intuitive grasp of the importance of representing the electorate accurately, Gallup and other commercial pollsters did not use the probability sampling methods that were being developed in the same decades and that are fundamental to social science surveys today. Probability sampling in its simplest form calls for each person in the population to have an equal chance of being selected. It can also be used in more complex applications where the chances are deliberately made to be unequal, for example, when oversampling a minority group in order
to study it more closely; however, the chances of being selected must still be known so that they can later be equalized when considering the entire population.

**Intuitions and Counterintuitions About Sample Size**

Probability sampling theory reveals a crucial but counterintuitive point about sample size: the size of a sample needed to accurately estimate a value for a population depends very little on the size of the population. For example, almost the same size sample is needed to estimate, with a given degree of precision, the proportion of left-handed people in the United States as is needed to make the same estimate for, say, Peoria, Illinois. In both cases a reasonably accurate estimate can be obtained with a sample size of around 1,000. (More cases are needed when extraordinary precision is called for, for example, in calculating unemployment rates, where even a tenth of a percent change may be regarded as important.)

The link between population size and sample size cuts both ways. Although huge samples are not needed for huge populations like those of the United States or China, a handful of cases is not sufficient simply because one’s interest is limited to Peoria. This implication is often missed by those trying to save time and money when sampling a small community.

Moreover, all of these statements depend on restricting your interest to overall population values. If you are concerned about, say, left-handedness among African Americans, then African Americans become your population, and you need much the same sample size as for Peoria or the United States.

**Who Is Missing?**

A good sample depends on more than probability sampling theory. Surveys vary greatly in their quality of implementation, and this variation is not captured by the “margin of error” plus/minus percentage figures that accompany most media reports of polls. Such percentages reflect the size of the final sample, but they do not reveal the sampling method or the extent to which the targeted individuals or households were actually included in the final sample. These details are at least as important as the sample size.

When targeted members of a population are not interviewed or do not respond to particular questions, the omissions are a serious problem if they are numerous and if those missed differ from those who are interviewed on the matters being studied. The latter difference can seldom be known with great confidence, so it is usually desirable to keep omissions to a minimum. For example, sampling from telephone directories is undesirable because it leaves out those with unlisted telephones, as well as those with no telephones at all. Many survey reports are based on such poor sampling procedures that they may not deserve to be taken seriously. This is especially true of reports based on “focus groups,” which offer lots of human interest but are subject to vast amounts of error. Internet surveys also cannot represent the general population adequately at present, though this is an area where some serious attempts are being made to compensate for the inherent difficulties.

The percentage of people who refuse to take part in a survey is particularly important. In some federal surveys, the percentage is small, within the range of 5 to 10 percent. For even the best nongovernment surveys, the refusal rate can reach 25 percent or more, and it can be far larger in the case of poorly executed surveys. Refusals have risen substantially from earlier days, becoming a major cause for concern among serious survey practitioners. Fortunately, in recent years research has shown that moderate amounts of nonresponse in an otherwise careful survey seem in most cases not to have a major effect on results. Indeed, even the *Literary Digest*, with its abysmal sampling and massive nonresponse rate, did well predicting elections before the dramatic
realignement of the electorate in 1936. The problem is that one can never be certain as to the effects of refusals and other forms of non-response, so obtaining a high response rate remains an important goal.

Questions About Questions

Since survey questions resemble the questions we ask in ordinary social interaction, they may seem less problematic than the counterintuitive and technical aspects of sampling. Yet survey results are every bit as dependent on the form, wording, and context of the questions asked as they are on the sample of people who answer them.

No classic morality tale like the Literary Digest fiasco highlights the question-answer process, but an example from the early days of surveys illustrates both the potential challenges of question writing and the practical solutions.

In 1940 Donald Rugg asked two slightly different questions to equivalent national samples about the general issue of freedom of speech:

- Do you think the United States should forbid public speeches against democracy?
- Do you think the United States should allow public speeches against democracy?

Taken literally, forbidding something and not allowing something have the same effect, but clearly the public did not view the questions as identical. Whereas 75 percent of the public would not allow such speeches, only 54 percent would forbid them, a difference of 21 percentage points. This finding was replicated several times in later years, not only in the United States but also (with appropriate translations) in Germany and the Netherlands. Such “survey-based experiments” call for administering different versions of a question to random subsamples of a larger sample. If the results between the subsamples differ by more than can be easily explained by chance, we infer that the difference is due to the variation in wording.

In addition, answers to survey questions always depend on the form in which a question is asked. If the interviewer presents a limited set of alternatives, most respondents will choose one, rather than offering a different alternative of their own. In one survey-based experiment, for example, we asked a national sample of Americans to name the most important problem facing the country. Then we asked a comparable sample a parallel question that provided a list of four problems from which to choose the most important; this list included none of the four problems mentioned most often by the first sample but instead provided four problems that had been mentioned by fewer than 3 percent of the earlier respondents. The list question also invited respondents to substitute a different problem if they wished (see table 1). Despite the invitation, the majority of respondents (60 percent) chose one of the rare problems offered, reflecting their reluctance to go outside the frame of reference provided by the question. The form of a question provides the “rules of the game” for respondents, and this must always be kept in mind when interpreting results.

Other difficulties occur with survey questions when issues are discussed quite generally, as though there is a single way of framing them and just two sides to the debate. For example, what is called “the abortion issue” really consists of different issues: the reasons for an abortion, the trimester involved, and so forth. In a recent General Social Survey, nearly 80 percent of the national sample supported legal abortion in the case of “a serious defect in the baby,” but only 44 percent supported it “if the family has a low income and cannot afford any more children.” Often what is thought to be a conflict in findings between two surveys is actually a difference in the aspects of the general issue that they queried. In still other cases an inconsistency reflects a type of illogical wish fulfillment in the public itself, as when majorities favor both a decrease in taxes and an increase in government services if the questions are asked separately.
Solutions to the Question Wording Problem

All these and still other difficulties (including the order in which questions are asked) suggest that responses to single survey questions on complex issues should be viewed with considerable skepticism. What to do then, other than to reject all survey data as unusable for serious purposes? One answer can be found from the replications of the forbid/allow experiment above: Although there was a 21 percentage points difference based on question wording in 1940 and a slightly larger difference (24 percentage points) when the experiment was repeated some 35 years later, both the forbid and the allow wordings registered similar declines in Americans’ intolerance of speeches against democracy. . . . No matter which question was used—as long as it was the same one at both times—the conclusion about the increase in civil libertarian sentiments was the same.

More generally, what has been called the “principle of form-resistant correlations” holds in most cases: if question wording (and meaning) is kept constant, differences over time, differences across educational levels, and most other careful comparisons are not seriously affected by specific question wording. Indeed, the distinction between results for single questions and results based on comparisons or associations holds even for simple factual inquiries. Consider, for example, a study of the number of rooms in American houses. No God-given rule states what to include when counting the rooms in a house (bathrooms? basements? hallways?); hence the average number reported for a particular place and time should not be treated as an absolute truth. What we can do, however, is try to apply the same definitions over time, across social divisions, even across nations. That way, we gain confidence in the comparisons we make—who has more rooms than who, for example. . . .

We still face the task of interpreting the meaning of questions and of associations among questions, but that is true in all types of research. Even an index constructed from a large number of questions on the basis of a sophisticated statistical calculation called factor analysis inevitably requires the investigator to interpret what it is that he or she has measured. There is no escaping this theoretical challenge, fundamental to all research, whether using surveys or other methods such as field observations.

Survey researchers should also ask several different questions about any important issue.

Table 1 Experimental Variation Between Open and Closed Questions

A. Open Question “What do you think is the most important problem facing this country today [1986]?”

B. Closed Question “Which of the following do you think is the most important problem facing this country today [1986]—the energy shortage, the quality of public schools, legalized abortion, or pollution—or, if you prefer, you may name a different problem as most important.”

1. Energy shortage
2. Quality of public schools
3. Legalized abortion
4. Pollution

In addition to combining questions to increase reliability, the different answers can be synthesized rather than depending on the angle of vision provided by any single question. A further safeguard is to carry out frequent experiments like that on the forbid/allow wordings. By varying the form, wording, and context of questions, researchers can gain insight into both the questions and the relevant issues. Sometimes variations turn out to make no difference, and that is also useful to learn. For example, I once expected support for legalized abortion to increase when a question substituted end pregnancy for the word abortion in the phrasing. Yet no difference was found.

Today, more and more researchers include survey-based experiments as part of their investigations, and readers should look for these sorts of safeguards when evaluating survey results.

The Need for Comparisons

To interpret surveys accurately, it’s important to use a framework of comparative data in evaluating the results. For example, teachers know that course evaluations can be interpreted best against the backdrop of evaluations from other similar courses: a 75 percent rating of lectures as “excellent” takes on a quite different meaning depending on whether the average for other lecture courses is 50 percent or 90 percent. Such comparisons are fundamental for all survey results, yet they are easily overlooked when one feels the urge to speak definitively about public reactions to a unique event.

Comparative analysis over time, along with survey-based experiments, can also help us understand responses to questions about socially sensitive subjects. Experiments have shown that expressions of racial attitudes can change substantially for both black and white Americans depending on the interviewer’s race. White respondents, for instance, are more likely to support racial intermarriage when speaking to a black than to a white interviewer.

Such self-censoring mirrors variations in cross-race conversations sense and nonsense about outside of surveys, reflecting not a methodological artifact of surveys but rather a fact of life about race relations in America. Still, if we consider time trends, with the race of interviewer kept constant, we can also see that white responses supporting intermarriage have clearly increased over the past half century (see table 2), that actual intermarriage rates have also risen (though from a much lower level) over recent years, and that the public visibility of cross-race marriage and dating has also increased. It would be foolish to assume that the survey data on racial attitudes reflect actions in any literal sense, but they do capture important trends in both norms and behavior.

Surveys remain our best tool for learning about large populations. One remarkable advantage surveys have over some other methods is the ability to identify their own limitations, as illustrated by the development of both probability theory in sampling and experiments in questioning. In the end, however, with surveys as with all research methods, there is no substitute for both care and intelligence in the way evidence is gathered and interpreted. What we learn about society is always mediated by the instruments we use, including our own eyes and ears. As Isaac Newton wrote long ago, error is not in the art but in the artificers.

Table 2  Percent of White Americans Approving or Disapproving of Racial Intermarriage, 1958–1997

<table>
<thead>
<tr>
<th>Year</th>
<th>Approve</th>
<th>Disapprove</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>4</td>
<td>96</td>
</tr>
<tr>
<td>1978</td>
<td>34</td>
<td>66</td>
</tr>
<tr>
<td>1997</td>
<td>67</td>
<td>33</td>
</tr>
</tbody>
</table>

SOURCE: Gallup Poll.
THINKING ABOUT THE READING

Schuman offers a few examples of how we all use sampling in our everyday lives. Come up with some examples of your own. For instance, how many students did you meet or talk with when considering whether to attend your current school? How confident are you about the ability of this “sample” to represent your school climate? What kinds of techniques and knowledge might make a social science survey more accurate than this kind of informal sampling? Schuman says that certain contemporary domains, such as the Internet, are difficult to survey representatively. What does he mean by this? What is the “wording problem”? In your opinion, do you think that people are inclined to give answers to survey questions even if they have never thought about the issue before? If so, what kind of information does this provide about human beliefs and motivations? What kinds of methods might be used to gather accurate information about human social life? What are some of the advantages and disadvantages of these methods?