Science and the Scientific Method

The main goals of this chapter will be to define what we mean by science (and distinguish it from other means of obtaining knowledge), to describe the steps of the scientific method, and to outline some of the major types of research.

What is Science?

Science is one approach to epistemology, the branch of philosophy that deals with how we know things. It is a combination of two ancient Greek philosophical approaches to epistemology: **empiricism** – a reliance on evidence from the senses – and **rationalism** – a reliance on logical reasoning. Science uses logic to create an argument which makes predictions about evidence. To better understand what science is, we will contrast it with two other epistemologies: authority and intuition. All three are useful under some circumstances, and all three have limitations.

Authority

When you use authority to obtain knowledge, you rely on information provided to you by someone whom you assume has some expertise and whom you assume is trustworthy (not intentionally trying to manipulate you). Given the degree of specialization in industrialized societies, it makes a lot of sense to rely on others' expertise. For example, rather than conduct your own study on toothpaste, you might listen to what the American Dental Association has to say. However, expertise does not guarantee accuracy. An expert may be wrong. In addition, experts may provide information in the hopes of furthering their own interests. When relying on authority as the source of your knowledge, it is important to remember the following: 1) authorities can be wrong (or deliberately misleading), 2) authorities testifying outside of their area of expertise are "just as dumb as the next guy," and 3) experts who cannot back up their position with evidence are relying on their intuition, to which we turn next.

Intuition

When you draw on intuition, you rely on all the impressions and automatic associations you've formed through your experiences. For example, in using your intuition to decide whether affirmative action policies are worthwhile, you would likely draw on associations you have about minority groups in the U.S. and the role of government. The advantages of intuition over science are speed and effort: Intuition can provide you with information quickly and easily. If your experience in a domain is considerable, your intuition may be a useful resource. The intuition of experienced doctors, for example, might lead to a more accurate diagnosis than the intuition of medical students. However, intuition at any level of expertise suffers from several shortcomings. Intuition is based on the experiences of a single person. For example, a White man trying to intuit what it would be like to be denied opportunities because of his race or gender is unlikely to gain an accurate perspective. In addition, there are several well-documented biases in human reasoning that can lead well-intentioned people astray. For example, the **confirmation bias** is the tendency for people to search only for information that confirms what they already suspect and to neglect information that disconfirms their suspicions. Because of these biases in reasoning and the limitations of a single perspective, intuition can be an unreliable source of knowledge.

Science

A scientific epistemology demands that statements of fact be tested through systematic observation that is designed to reduce bias. We will discuss the elements of the scientific method below, but for now it is important to point out two essential features of science: **falsifiability** and **replication**.

Falsifiability. Falsifiable means "able to be proven false." It may sound strange to be focusing on proving something false, but that is what science tries to do: it tests statements by attempting to prove them wrong. Repeated failures to disprove a statement increase the confidence that can be placed in the statement. **Karl Popper**, a philosopher of science, argues that the essence of falsifiability is the ability to make **risky predictions** – predictions that could be checked against evidence. For example, a theory of the solar system might lead a researcher to predict a solar eclipse on a given day. If the day comes and goes without an eclipse, then the theory is discredited. Falsifiability is essential to science because it makes science self-correcting: theories that make accurate predictions survive, while theories that make inaccurate predictions are discarded. Otherwise, inaccurate theories would accumulate and there would be no way to distinguish them from accurate theories.

Replication. The second essential feature of science is replication. Scientific evidence must be repeatable under the same conditions. If a group of scientists claim that they have produced nuclear fusion in their laboratory and hundreds of other scientists are unable to duplicate the results under similar conditions, then the claims of the first scientists are discredited. One consequence of this principle is that scientific reports must contain enough information about methods so that other researchers could repeat the experiment. Like falsifiability, replication is designed to protect the integrity of scientific knowledge by screening out false claims. Replication illustrates the role of the scientific *community* in protecting the integrity of scientific knowledge. Ideally, inaccuracies generated by one researcher will be identified and corrected when the study is repeated by other researchers. However, in practice, it is unusual for a researcher to attempt to duplicate the findings of another researcher. Such community policing is more common when a result is controversial, contradicts accepted wisdom, or when two or more researchers have rival explanations for the same phenomenon.

Authority and intuition in science. It would be incorrect to think that authority and intuition are incompatible with science. In fact, both are commonly used in science. Levels of expertise in particular areas are often acknowledged within a scientific community, and the statements of these experts are often given more weight. However, this expertise often consists of the ability to integrate and summarize vast amounts of scientific evidence, not simply on a person's status. Intuition is also used in science. The generation of new research ideas is one area where intuition is useful. The interpretation of unusual results is another application of intuition.

In conclusion, the scientific principles of falsifiability and replication are designed to increase our confidence in the conclusions we reach and to prevent the accumulation of errors in the scientific literature.

The Scientific Method

Although there is no single recipe for a study to be called "scientific," science generally proceeds in five steps: question, review, hypothesize, test, theorize.

Question

Research questions often come from observations in everyday life, but they are also often generated as unanswered questions from previous research or as implications from theory. Some tips for generating your own research question: 1) become an active observer of your social world and take note of occasions when you "wonder if…"; 2) legal issues are fertile ground for research: the way evidence is presented, characteristics of the defendant or plaintiff, etc.; 3) social problems are another source of research questions: racism, sexism, etc.; 4) consider a change you would like to make in your environment (e.g., more opportunities for first-year students to meet people) and design a plan to test whether your change has the desired effect; 5) look over your introductory psychology text for studies that you thought were interesting and think of a way to carry that research forward one step.

Review

Once you have a research question, it's important to review the scientific literature to: 1) find out if someone has already answered your question satisfactorily, 2) gain a better understanding of previous theory and findings relevant to your topic, and 3) discover methods that worked and did not work in the pursuit of your question. We will discuss how to go about searching the scientific literature in a later section.

Hypothesize

Third, researchers must translate a research question (e.g., "How does hypnotic pain control work?") into a **hypothesis** – a *testable prediction*. To be testable, a hypothesis must involve conditions that can be observed repeatedly. For example, “Hypnosis works by reducing the amount of *attention* focused on pain” might be tested if a researcher were clever enough. However, “Hypnosis works through invisible purple dragons that have no mass” is probably not testable. This is the main reason that fundamental religious questions, such as the existence of a non-material spirit world, are difficult to address scientifically: It is unclear how one would make a testable prediction about them.

Test

Fourth, researchers design a set of procedures to *test* the hypothesis. As mentioned above, it is essential that this test be designed so that it is *possible* for the hypothesis to be either supported or discredited. If you look at a design and think “No matter how the evidence comes out, it would support the hypothesis,” then the design is not a true test; a test must involve a *risky* prediction.

If the results of the test support the hypothesis, then the hypothesis is only “supported”; it is not “proven.” Tests never “prove” anything – they merely increase or decrease our confidence in a hypothesis. (One consequence of this is that you should **never use the word “prove”** or “proof” or “truth” in a scientific paper unless you are discussing formal logic. Instead, use “supports,” “is consistent with,” “suggests,” “indicates,” etc.) If ten attempts to disprove a hypothesis all fail, it does not mean the hypothesis is correct, but it should increase our confidence in the hypothesis.

Researchers have a wide selection of strategies they can use to test a hypothesis. We will talk about the six strategies most commonly used in psychological research.

Laboratory experiments. In a laboratory experiment, the researcher attempts to control conditions so that only a single factor (or more factors, but we’ll get to that later) is allowed to vary. For example, in an experiment on the effects of heat on aggression, a researcher may have participants come into a room where the temperature is carefully controlled: for some participants it is 90°, and for other participants it is only 70°. All other features of the room and the situation are held constant. The logic of the laboratory experiment is this: If people in the hot room are more aggressive than people in the cool room *and the only systematic difference between the people is the room temperature*, then we can have some confidence that the room temperature is responsible for causing the aggression. In summary, the key feature of laboratory experiments is control over the situation and the major advantage is more confidence in a cause-and-effect relation. The major criticism of laboratory experiments is that the setting may be too artificial – people may act differently in “real-life” contexts. The next research strategy is designed to address this criticism.

Field experiments. In a field experiment, researchers attempt to conduct an experiment outside of the laboratory in a more natural setting. For example, researchers studying the effects of rewards on performance might go to General Motor’s corporate headquarters. Typically, researchers manipulate one variable (e.g., give a pay raise to some employees but not others) and observe its effect on another variable (e.g., performance, as measured by supervisor job evaluations). This research strategy is less artificial than laboratory experiments, but generally it involves a sacrifice in the amount of control that researchers can exert over the setting. For example, employees may talk to one another and discover the unequal treatment.

Naturalistic observation. In naturalistic observation, researchers again collect data in a natural setting, but this time there is little to no interference with participants. For example, researchers may observe the number of people who buy ice cream under different outside temperatures. The primary advantages of naturalistic observation are that it offers a look at more “unguarded” behavior (because participants do not feel as if they are in a study) and that there is generally less effort required by the researcher in terms of generating laboratory procedures or materials. The primary disadvantage of naturalistic observation is that researchers can have little confidence in cause-and-effect relationships. We will discuss the reason for this later in the course when we get to internal validity, but the basic reason is that unless we *control* temperature, it is always possible that some other variable besides temperature (vacation time, seasonality, humidity, etc.) is really causing changes in ice cream sales.

Questionnaire / Survey. A questionnaire is a set of questions that are usually designed to measure the same thing – an attitude, belief, personality characteristic, etc. Questionnaires are typically distributed on paper, but they could be delivered over the phone or on the web. A questionnaire that is distributed to a large but usually well-defined group is often called a **survey**. The advantage of questionnaires is that they can reach a large group of people at relatively little cost to the researcher. If the group is carefully selected, conclusions can be drawn about the attitudes or beliefs of a population. There are two disadvantages to questionnaires. First, researchers must rely on participants' responses. When questions have a clearly socially desirable answer, participants may be motivated to portray themselves in the best possible light. In addition, people's reports of *why* they behaved as they did are not always trustworthy. In an article titled "Telling more than we can know," Nisbett and Wilson (1977) report one study in which participants were asked which of four nylon stockings was the best quality. Unbeknownst to participants, the stockings were identical. Early results indicated that the pair on the right was chosen over the pair on the left by a four-to-one ratio. However, no participant ever mentioned the position of the stockings as a factor in their choice. This did not stop participants from generating reasons for their choice. When the position of the stockings was suggested as a factor, "virtually all subjects denied it, usually with a worried glance at the interviewer suggesting that they felt either that they had misunderstood the question or were dealing with a madman" (p. 244). In summary, people can easily manufacture an explanation for their behavior, but this explanation may be independent of the real causes. The second disadvantage is that cause-and-effect conclusions with questionnaires are difficult. For example, the Program on International Policy Attitudes randomly sampled over 3,000 U.S. adults between June and September 2003 and found that, compared to people who got most of their news from PBS, people who got most of their news from Fox News were significantly more likely to believe that Saddam Hussein was working closely with al Qaeda, that the U.S. had found weapons of mass destruction in Iraq, and that the rest of the world favored the U.S. going to war against Iraq (all three of these beliefs are false). From this study, it is unclear whether Fox News caused people to become more poorly informed, poorly informed people tend to watch Fox News, or some third factor caused people to both be poorly informed and to watch Fox News. Any of those three are equally possible (and none speaks well of Fox News).

Archival research. Researchers conducting archival research use pre-existing records of behavior as their data source. Examples include crime statistics (e.g., the FBI's annual Uniform Crime Report), diaries of famous people, or accounts of a murder in different newspapers. The advantage of archival research is that the records are often much easier to obtain than if the researcher had conducted a study to collect the data. Some records (e.g., U.S. census information) may offer researchers a much larger and more representative sample of behavior than they could have obtained through their own data collection. The major disadvantages of archival research are a reliance on records generated by others, which may be of questionable reliability, and the cause-effect problem discussed above.

Case study. A case study is an in-depth analysis of a particular person (e.g., Winston Churchill), organization (e.g., Enron Corporation), or event (e.g., the stock market crash of 1929). Case studies are often used in psychology to examine a condition or event that is rare (e.g., NASA's disastrous decision to launch the space shuttle *Challenger* despite warnings from engineers) or the result of an event that could not ethically be replicated in an experiment (e.g., brain injury). The primary advantages of case studies is that they usually offer great richness of detail and are often the only strategy available. The richness of detail is especially valuable early in a research program, when researchers are still learning basic information about a phenomenon. The primary disadvantages of case studies are that cause-effect conclusions are difficult and that the case under investigation may not be representative of the phenomenon generally.

Multi-method approach. Because each research strategy has different advantages and limitations, researchers investigating a particular phenomenon should conduct multiple studies with multiple methods. For example, Richard Nisbett and Dov Cohen (1995) studied North-South differences in aggression using archival research, laboratory experiments, questionnaires and field experiments. Typically, methods with less control (e.g., case studies, archival research) are used early in a research program, and methods with more control (e.g., experiments) are used once a phenomenon is at least partially understood.

Choosing a strategy. When choosing a research strategy, consider what you want to optimize in your study. If you want confidence in cause-and-effect conclusions, select a laboratory experiment. If you want to estimate population values (e.g., campus attitudes toward gun control), select a survey. When planning your study, you should also keep in mind the resources you will have available: participants are generally available for only a single 50-minute period, you may be responsible for providing all your own supplies (e.g., props, questionnaires), and using actors to create an artificial social situation can be very time-consuming (and ethically problematic, as we'll see in the ethics section).

Between-subjects, within-subjects, and mixed designs. There is one final distinction you should be familiar with as you begin to consider how you will design your own study. In a **between-subjects design**, different people are placed into different conditions. One group of people might be assigned to view sexist advertisements while another group is assigned to view non-sexist advertisements. The two groups are then measured in some way (e.g., their attitudes toward women in leadership positions) and their averages are compared. In a **within-subjects design**, the same people are in several conditions. For example, each participant could be interviewed by both a Black and a White interviewer and their nonverbal behavior toward each one could be compared. We won't come back to within-subjects designs for awhile, but it is worth mentioning them now because they are *much more statistically powerful* than between-subjects designs: you will be able to detect differences with a within-subjects design that would be undetectable with a between-subjects design. The reason for this is covered in the section on within-subjects designs. For now, if the phenomenon you are investigating is fairly subtle and might be difficult to detect, you should consider a within-subjects design. The disadvantage of within-subjects designs is that participants are more likely to discover the hypothesis and then change their behavior. If you viewed several advertisements and then answered questions about female leadership, and then viewed several other advertisements that seemed to differ in how sexist they were and again answered questions about female leadership, you'd probably have a pretty good idea of what the researchers were studying. Finally, a **mixed design** is a combination of a between-subjects and a within-subjects design. In a mixed design, people are assigned to different conditions but also measured more than once – often before and after they receive some kind of treatment. For example, you could evaluate the effectiveness of a new stop-smoking message by assigning smokers to two groups, exposing one to the new message, and measuring both groups both before and after exposure to the message. If the message group showed a greater decrease in smoking over the two time points than the non-message group, you would have evidence that the message is effective. A mixed design has the statistical power of the within-subjects design without as much risk that the participants will discover the hypothesis.

Theorize

A theory is an organized set of principles designed to explain a phenomenon. For example, the heliocentric theory specifies that the sun is at the center of the solar system and that the earth revolves around it. A theory is considered strong to the degree that it: 1) is supported by evidence, 2) summarizes existing research, and 3) offers questions for future research.

According to scientific historian **Thomas Kuhn** (1962), every theory will inevitably be discarded in favor of a new and better theory during a process he called a **scientific revolution**. According to Kuhn, this process begins with the discovery of **anomalies** – scientific observations that cannot be explained by the old theory. Eventually, enough anomalies are documented that scientists grow dissatisfied with the old theory's inability to account for them. One group of scientists then initiates a scientific revolution to overthrow the old theory in favor of a theory that can explain the new as well as the old evidence. For example, the heliocentric theory of the solar system was able to explain several astronomical anomalies, such as the tendency for Mars to appear to double-back on its path across the night sky and for Venus to have phases. Sometimes revolutions occur when a new theory does not directly contradict an older theory but instead offers a simpler explanation or one that accounts for a wider range of phenomena. For example, James Maxwell's theory of electromagnetism in the 1860s unified and expanded current theories of electricity and magnetism. One consequence of Kuhn's theory of scientific revolutions is that all theories should be regarded only as the latest drafts in the evolution of knowledge.

The word "theory" is often used misunderstood to mean that it is based on speculation rather than evidence. As discussed above, theories vary in how well they are supported by evidence. Some theories, such as the theory of electromagnetism and the theory of evolution by natural selection, have accumulated so much support that they can be relied upon with great confidence. The fact that better theories may be developed in the future should not make you cynical about current theories, but it should make you skeptical about treating them as "truth." Rather than thinking about a theory in terms of whether it is "true" or "false," it may be more useful to think in terms of confidence: based on the evidence, how much confidence is appropriate for a particular theory?