Consider the following questions as you read Chapter 3

- How do psychologists observe behavior?
- What are some common techniques for observing and recording behavior (i.e., collecting data) in different situations?
- How do psychologists use observations to learn about behavior?
- What questions about behavior do the different research methods allow psychologists to answer?
- Which research method is best when asking about the cause of behavior?

Imagine that you work in a busy corporate office. One day your boss comes to you for advice. A report found that productivity in the office tends to decline later in the afternoon and he or she wants to find a way to increase productivity during this time period. Your boss’s suggestion is that having a cappuccino machine in the office lunchroom may cause workers to drink more coffee after lunch, in turn giving them more energy and productivity in the afternoon. You are asked to use your knowledge of research methods to find out if the suggestion is a good one. How would you conduct a study to provide the advice your boss is looking for?

In this chapter, we discuss the methods psychologists use to learn about behavior. This chapter provides an overview of some of the main data collection techniques and research designs used in psychological research to illustrate how psychologists apply the scientific methods described in Chapter 1 to the study of behavior. Figure 3.1 illustrates the steps involved in the research process while designing the study. The choice of data collection technique and research design is made by the researcher. This choice depends on the type...
of behavior that is of interest and what kinds of questions one wants to answer about the behavior. External and internal validities also play a role in these choices. External validity is the degree to which the behavior observed in the study is realistic, would occur naturally, and can be generalized beyond the boundaries of the study to other individuals and situations. How much external validity a study has is important because the goal of research is to gain knowledge about behavior that applies to a large group of individuals in their everyday lives, not just to the individual study participants with any situational restrictions the study included. In other words, the conclusions need to generalize beyond the study itself. Some of the observation techniques and research designs that psychologists use tend to allow for higher external validity than others. However, in many cases the higher the external validity in a study, the lower the internal validity.

**Internal validity** is the degree to which a study provides a good test of a causal hypothesis, where alternative explanations of the data can be ruled out. A study with high internal validity provides causal information about behavior. To increase the internal validity of a study, a researcher controls for extraneous factors that can affect the observations. With more control over the factors in a study, internal validity increases, but behavior may become more artificial and lower the external validity of that study. As we discuss the different types of studies that can be conducted by psychologists, you will see that internal and external validities are important concepts in research.

Another important issue in observing behavior is making certain that the observations are **reliable**. For some observation techniques, this means making sure that observers are categorizing behaviors the same way. For other techniques, this means making certain that different items on a survey or questionnaire designed to measure a specific behavior all provoke similar responses from the research participants. In other words, reliability in a survey means that the participants’ responses are similar from item to item or from one time they complete the survey to the next. Thus, reliability is important to consider when you design a study and choose an observation technique, but how you increase reliability depends on the observation technique you are using.

---

**External Validity**: the degree to which the results of a study apply to individuals and realistic behaviors outside the study

**Internal Validity**: the degree to which a study provides causal information about behavior

**Reliability**: the degree to which the results of a study can be replicated under similar conditions

---

**DATA COLLECTION TECHNIQUES**

How do psychologists observe behavior? When researchers are planning to observe a behavior, they must first decide how they are going to define that behavior. This is the **operational definition** of the behavior. An operational definition of an abstract concept (depression, memory ability, etc.) allows the researcher to define the concept for the purpose of measurement and data collection. Thus, an operational definition is a definition of an abstract concept that makes it concrete for the purpose of the research study. If social
behavior is of interest, the researcher may define this as “the number of people who spend time in groups,” “the number of times an individual approaches another individual to interact with him or her,” or “the score on a questionnaire that asks about social behaviors.” There are clearly many ways to define a particular behavior for data collection. Researchers choose an operational definition for behavior that they expect will provide the best (i.e., most valid) method of learning about the behavior they are interested in. In other words, they define the concept so that it can be measured by a specific behavior (or set of behaviors). See Table 3.1
for some examples of operational definitions of different concepts. The techniques described in the following sections are different ways in which psychologists collect data about behavior. As you will see, the choice of technique is linked to the operational definition the researcher uses. See Table 3.2 for a comparison of these techniques.

**Naturalistic Observation**

A recent popular show on the Animal Planet cable network called *Meerkat Manor* follows packs of meerkats as they interact with each other within the pack, with other packs of meerkats, and with the environment they live in. The show is narrated to make it more entertaining, but the animal’s natural behaviors displayed on the show are observed and exhibited as they occur in their normal environment. In other words, the researchers for the show use naturalistic observation to learn about meerkat behaviors. Naturalistic observation is used when a researcher wants to learn about behavior that naturally occurs for an individual without influencing the behavior. The goal of naturalistic observation is to be unobtrusive so that the researcher does not affect the observed individuals’ behavior. Naturalistic observation is often used by researchers interested in observing the behavior of animals in their natural environment or children who are in specific environments (both natural and contrived).

---

**Table 3.1 Examples of Operational Definitions**

<table>
<thead>
<tr>
<th>Concept</th>
<th>Possible Operational Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>Score on a mood questionnaire</td>
</tr>
<tr>
<td></td>
<td>Number of times someone has thought about suicide in the last month</td>
</tr>
<tr>
<td></td>
<td>Measure of certain neurotransmitters in areas of the brain</td>
</tr>
<tr>
<td>Problem-Solving Ability</td>
<td>Amount of time it takes to complete a puzzle</td>
</tr>
<tr>
<td></td>
<td>Number of problems solved correctly</td>
</tr>
<tr>
<td></td>
<td>Score on a standardized test</td>
</tr>
<tr>
<td>Learning</td>
<td>Difference in score from pretest to posttest</td>
</tr>
<tr>
<td></td>
<td>Change in time to complete a problem or test</td>
</tr>
<tr>
<td></td>
<td>Change in confidence ratings to perform a skill</td>
</tr>
</tbody>
</table>
To understand observed behaviors using naturalistic observation, a researcher must develop a coding scheme to categorize the exhibited behaviors of the participants. This allows the researcher to quantify (i.e., count) and summarize the behaviors for the group of individuals observed. The coding scheme depends on the operational definition of the behavior the researcher is using. In other words, the behaviors the researcher indicates are part of the operational definition become categories of behaviors in the coding scheme. For example, if researchers are studying helping behaviors with naturalistic observations, they may define helping behavior as approaching someone who appears to need help (e.g., has dropped something, has a broken-down car, is unable to open a door, etc.) and offering help or acting to help the person. Thus, observers may sit on a university quad and count behaviors they observe that fit into categories such as “asking someone if he or she needs help,” “helping someone pick something up,” “giving someone directions when asked.” This allows researchers to quantify the different behaviors they see. The researchers can then describe helping behaviors according to their coding scheme to indicate the frequency of helping behaviors overall for this situation and group of individuals, and which types of behaviors occur more often than others.

Developing a coding scheme generally involves defining categories of behavior that fit the type of behavior being observed. The operational definition of the behavior should guide the selection of categories of behaviors that qualify. It is important to clearly define

<table>
<thead>
<tr>
<th>Technique</th>
<th>Definition</th>
<th>Realism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalistic observation</td>
<td>Observing individuals’ behavior in their normal environments</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decreases if observers fail to be unobtrusive</td>
</tr>
<tr>
<td>Surveys/questionnaires</td>
<td>Individuals respond to items in written form or on the Internet</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Decreases with self-monitoring, phrasing of the questions and the scale used to control responses</td>
</tr>
<tr>
<td>Systematic observation</td>
<td>Collection of systematic behaviors in controlled tasks</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Due to control of task situation</td>
</tr>
<tr>
<td>Archival data</td>
<td>Using available records to collect observations</td>
<td>Varies with type of record</td>
</tr>
</tbody>
</table>
these categories so that observers are clear in what behaviors they are looking for when they observe the individuals of interest in a study. Clear categories also help multiple observers be more consistent in how they classify the behaviors they are looking for (more on this issue below). Finally, the coding scheme can involve counting the number of certain types of behaviors and/or the amount of time individuals engage in the defined behaviors either in set time intervals or across the entire span of the observation period.

The primary advantage in using naturalistic observation to study behavior is that the behavior is likely to be more realistic compared to some of the other techniques. This can increase the external validity of a study. However, this technique has its disadvantages. It can sometimes be difficult to be unobtrusive. The presence of an observer can easily change the behavior of the individuals being observed. This is an issue that has come up for the *Meerkat Manor* show. Camera crews follow the meerkats closely all the time to record their behaviors for the show, and their presence may have affected the behaviors they observe. Thus, researchers using this technique must take great care to ensure that they are not influencing the behavior they are observing simply by being present in an individual’s environment. Another drawback to naturalistic observation is that it can be very time-consuming. The observers must wait for the behavior they are interested in to be exhibited by the participants. Thus, this technique can be more time intensive and consume more resources than other observation techniques. A third disadvantage is that multiple observers (observing the same or different individuals) may not be recording the behaviors they observe in the same way. To deal with this problem, most studies that involve naturalistic observations include training of the observers to ensure that they are collecting the data in the same way. In fact, the similarity in coding of the data is typically measured and reported in such studies. This is known as interrater reliability (how similarly the observers are coding the data). A measure of interrater reliability is usually reported based on the percentage overlap in the way the observations are classified across multiple observers. To illustrate this concept, consider the study described above that looked at helping behaviors. In this study, it is likely that more than one person would observe on the quad (either at the same time or at different times) to allow enough helping behaviors to occur and be observed. If the observers code the behaviors differently (e.g., one observer counts bending over to help as “helping someone pick something up,” whereas another observer only counts this behavior if someone actually picks up something someone dropped), the internal validity of the study decreases because the observers will have different operational definitions of the behaviors.

Chiang (2008) provides a recent example of a study that used naturalistic observations. In this study, 32 children with autism were observed to investigate aspects of their spontaneous communication. The children were videotaped in their natural environments (classrooms at their school), while they completed normal, everyday activities (lunch, free time, academic activities, etc.). In the article reporting the study (published in the journal *Autism*), Chiang described the coding schemes developed to summarize and understand the communication behaviors that were seen in the tapes viewed by the observers (speech or writing, vocalizations that were not identified as words, eye contact, common gestures such as hugging, waving, or nodding, etc.). The interrater reliability (above 80%) is also reported to provide evidence that the observers were coding the data in a similar manner. Chiang concluded that children with
autism exhibit a range of communicative behaviors across the different settings and suggested a model for future studies of spontaneous communication in autistic individuals. The issue of intrusiveness can be considered for this study. How much did the presence of the video camera affect the participants’ behavior? Were participants aware of the video camera? If so, were they more self-conscious or uncomfortable about their behavior because of its presence? The issue of obtrusiveness should be considered whenever naturalistic observations are used in a study.

Surveys/Questionnaires

The other data collection techniques we will discuss are more obtrusive than the naturalistic observation technique because they involve some type of interaction with the research participants. One of these techniques commonly used in psychological research is a survey in which individuals are asked about specific behaviors. (Although the terms survey and questionnaire are sometimes used in different contexts in research, in this text I will use these terms interchangeably.) Survey research is often conducted to measure mood, attitudes about a topic, or frequency of certain behaviors through self-reports from the participants. Typically, surveys contain a number of questions that ask the research participant to rate the presence or frequency of his or her own thoughts or behaviors. When surveys are used, participants are often asked to use a response scale (e.g., 1 to 5 or 1 to 7) or response category (e.g., often, sometimes, not very often, never) that matches how they feel about a behavior or how likely they are to exhibit the behavior. This means that the participants are limited in the types of responses they can make to the survey items. In other words, the survey uses a closed-ended response scale because only certain responses are valid responses to the items. The scale in Table 3.3 was designed to assess how likely one is to discuss one’s emotions and disclose personal problems to others. Scores on this scale have been shown to be related to higher self-esteem and general satisfaction with one’s life (Kahn & Hessling, 2001). The Distress Disclosure Index provides an example of a closed-ended response scale, as a 5-point scale is given for responses.

Another way to design a survey is to ask participants to respond to questions on an open-ended response scale. In other words, they can respond in whatever way they wish to the questions you asked them. Analyzing the data from an open-ended response scale also requires the development of a coding scheme, because the responses are qualitative rather than quantitative. Such coding schemes are developed by researchers for some validated surveys used frequently in certain types of research. Using a closed-ended

---

**Naturalistic Observation:** a data collection technique involving noninvasive observation of individuals in their natural environments

**Interrater Reliability:** a measure of the degree to which different observers rate behaviors in similar ways

**Survey Research:** a research study that uses the survey observational technique to measure behavior

**Closed-Ended Response Scale:** participants respond to survey questions according to the response options provided by the researcher
### Table 3.3 Distress Disclosure Index

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>When I feel upset, I usually confide in my friends.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>2</td>
<td>I prefer not to talk about my problems.</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>3</td>
<td>When something unpleasant happens to me, I often look for someone to talk to.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>4</td>
<td>I typically don’t discuss things that upset me.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>5</td>
<td>When I feel depressed or sad, I tend to keep those feelings to myself.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>6</td>
<td>I try to find people to talk with about my problems.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>7</td>
<td>When I am in a bad mood, I talk about it with my friends.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>8</td>
<td>If I have a bad day, the last thing I want to do is talk about it.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
<tr>
<td>9</td>
<td>I rarely look for people to talk with when I am having a problem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Strongly agree</td>
</tr>
<tr>
<td></td>
<td>Strongly disagree</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Strongly disagree</td>
</tr>
</tbody>
</table>
response scale allows the researcher to collect quantitative responses (i.e., numerical responses), so no coding scheme is needed for closed-ended scales. Surveys are often administered using pencil and paper or (as is becoming more frequent) over the Internet via website.

Researchers are often interested in testing the validity and reliability of the surveys and questionnaires they use. Checking the validity of a survey means making sure that the questions asked are really about the behavior the researcher is interested in. If a survey is designed to measure someone’s level of anxiety, the questions have to be written to ask about behaviors that are related to anxiety, or the survey is not measuring what it is designed to measure. In other words, does the survey actually measure the construct it was designed to measure? Checking the reliability of a survey means making certain that the responses you get from an individual are similar either at different points in time or to similar items on the questionnaire. If an individual’s responses change drastically from time to time or across similar items, even though the attitude or behavior being measured by the survey does not change, you will not get an accurate measure of that attitude or behavior. A more detailed discussion of the validity and reliability of surveys is presented in Chapter 9.

Typically, using a validated survey gives a researcher observations of behavior that are a step ahead of the other techniques discussed because the validity and reliability of a survey will already have been tested and the survey revised (if necessary) to maximize its accuracy. The primary disadvantage of using surveys to collect data is that the observations are considered self-reports, which means that they may not be correct representations of a person’s behavior. Individuals do not always view their behavior accurately and may report who they think they are on a survey, not who they actually are. Participants may also want to portray themselves more positively to the researcher and intentionally respond in a way that achieves that goal (i.e., they self-monitor). This is called social desirability and it can bias the results of a survey or questionnaire. Thus, researchers must be careful in interpreting behaviors observed with this technique, as they may not be accurate representations of individuals’ behaviors.
The Beck Depression Inventory–II (BDI–II; Beck, Steer, & Brown, 1996) and Beck Anxiety Inventory (BAI; Beck & Steer, 1993) are two commonly used surveys in psychological research on mood. These respective surveys contain items that ask individuals about the intensity of certain feelings and the intensity of specific behaviors related to depression and anxiety. For example, the BDI–II contains 21 items and asks respondents about feelings of sadness, being punished, and lack of interest in sex and behaviors such as difficulty in sleeping and changes in eating habits. Many studies use the BDI–II to measure depression or the BAI to measure anxiety, and the reliability and validity of these surveys have been frequently tested.

Interviews. Surveys can also be administered as **interviews** such that individuals respond to questions orally. Interviews can be done face-to-face, over the phone, or in focus groups. Like naturalistic observations, observing behaviors with interview data requires the researcher to develop a coding scheme to understand the behaviors described or exhibited in the interview. One advantage of using interviews is that you can ask about a specific behavior instead of waiting for the individual to exhibit the behavior spontaneously (as in naturalistic observations). Another advantage is that if the interview is structured to allow flexibility, different questions can be asked depending on the response that is given. For example, if a participant responds that a question particularly applies to him or her, the interviewer can follow up that response with additional questions on the topic tailored to the type of response made.

Focus groups are becoming a popular way to conduct interviews to learn about individuals’ attitudes toward a societal issue, political candidate, or consumer product. Interviewing people in groups uses fewer resources and can sometimes elicit responses from individuals who may be more reluctant to voice an opinion when they are asked on their own. When reluctant individuals hear that others have an opinion similar to their own, they may be more likely to voice their opinion. However, this can also be a limitation to the use of interviews: If they are conducted in groups, individuals may go along with the group rather than voice an opinion that differs from others (Ashe, 1955). In other words, conformity of responses occurs. Interviewees may also self-monitor during interviews, meaning that they can respond according to how they wish to appear to others instead of how they actually are (i.e., the social desirability bias can occur).

Another drawback to the use of interviews is that the way a question is asked can affect the response that is given. Thus, great care must be taken in writing questions for interviews.

A recent study by Creasey and Ladd (2005) used interviews of individuals to investigate the relationship between parental attachment and conflict resolution behaviors in current romantic
relationships. To understand attachment style to their parents, research participants were interviewed about their relationships with their parents. Responses were then coded to categorize individuals according to different types of attachments that children have with their parents. Creasey and Ladd reported that the success of conflict resolution strategies with a romantic partner depended on the type of attachment individuals had with their parents.

**Systematic Observation**

*Systematic observations* are typically used when the researchers want to exert the highest amount of control over behavior. In other words, the observations are obtrusive and can often affect the behavior being studied. Thus, systematic observation is often used to study behaviors that are least likely to be affected by the process of measuring them. Examples of these behaviors are often cognitive or biological in nature (e.g., memory accuracy, problem-solving speed, firing of a neuron, activity in a particular brain area). However, systematic observation can be used to study behaviors in other areas of psychology as well. A recently developed method for studying automatic social attitudes called the Implicit Association Test, or IAT (Greenwald, McGhee, & Schwartz, 1998), uses systematic observation of reaction times of participant responses to sets of word or picture stimuli to measure one’s unconscious prejudicial attitudes. In the IAT, the speed with which people respond to items after certain associations have been formed is recorded to determine if some judgments take longer than others. The IAT works by associating a specific response to different dimensions of a concept (e.g., press the right key when you see a female name and the left key when you see a male name; press the right key when you see items a teacher might use and the left key when you see items a firefighter might use). The concepts are then combined (e.g., press the right key when you see a female name or an item used by a firefighter and the left key when you see a male name or an item used by a teacher). The assumption is that longer responses with one combination of concepts (e.g., female names and firefighter items both requiring a right-key response) than the other combination (e.g., female names and teacher items both requiring a right-key press) may reveal unconscious social biases people may have that they either are unaware of or consciously suppress when asked explicit questions about their beliefs (e.g., Is it appropriate for a woman to become a firefighter?). Figure 3.2 illustrates the IAT procedure. You can also try out the IAT procedure for yourself at http://implicit.harvard.edu.

Because a high degree of control can be exerted on the measurements of behaviors observed using systematic observations, they typically add to the internal validity of a study. The situation in which the behaviors are measured is typically controlled to eliminate influences on the behaviors that are not the focus of the study. Thus, systematic observations are often collected in a laboratory setting, where distractions of normal life are minimized and tasks are presented and completed on a computer to maximize accuracy. The drawback of this level of control is that the behaviors being studied may be artificial. In other words, external validity can be lower for systematic observations than other data collection methods, though these observations may have better internal validity.
Using Archival Data

Sometimes when researchers have questions about behavior, they find that those behaviors they are interested in have already been observed. In other words, the data they wish to analyze to answer their research question already exist. Someone else has collected them. For example, if researchers are interested in health-related behaviors, they may wish to use existing hospital records as their observations. An example of this type of study was done in Pueblo, Colorado, a few years ago (Bartecchi et al., 2006). Pueblo is a small town with two hospitals where residents of the town and surrounding area receive medical care. After the town passed a smoking ban, researchers decided to look at hospital records to compare the number of hospitalizations for heart attacks during the year and a half before the smoking ban began (as a way to determine the number of heart-related illnesses that occurred when people were allowed to smoke in public places) with the number of hospitalizations for heart attacks that occurred during the year and a half after the smoking ban started. They found that the number of hospitalizations for heart attacks decreased significantly during the year and a half after the smoking ban and concluded that the decrease in public smoking was related to this decrease (by comparing heart attack hospitalization rate change for comparable areas without a smoking ban over the same period of time). The use of hospital records in this study is an example of how researchers use archival data as an observation technique.

---

Figure 3.2  Simplified IAT Procedures

<table>
<thead>
<tr>
<th>Response-Dimension Pairing</th>
<th>Gender-Key Association</th>
<th>Occupation-Key Association</th>
<th>Common Associations Provide Baseline Condition</th>
<th>Test of Biased Association—Slower Responses Than Baseline Indicate Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (left key) Female (right key)</td>
<td>Teacher (left key) Firefighter (right key)</td>
<td>Male/paramedic Female/paramedic (right key)</td>
<td>Male/teacher (left key) Female/firefighter (right key)</td>
</tr>
</tbody>
</table>

**Stimuli (With Correct Response)**
- John (left)
- Mary (right)
- Angela (right)
- Steve (left)

- Ladder (right)
- Books (left)
- Hose (right)
- Pencil (left)

- Henry (left)
- Desk (right)
- Truck (left)
- Heather (right)

- Rick (left)
- Book (left)
- Axe (right)
- Judy (right)

**Source:** Greenwald, McGhee, and Schwartz (1998).

**Notes:**
1. Conditions with keys (right, left) reversed (e.g., male with right key, female with left key) are also used.
2. Longer response times for bias test trials indicate unconscious social biases.
Many archival data sets are collected by agencies on a periodic basis. A quick web search will show summary results for many of these observations. For example, one can find data related to presidential approval ratings, consumer confidence, consumer spending, and opinion polls. Figure 3.3 shows approval ratings (percentage of people who approve of the job the president is doing) for President Bush over a 2-year period near the end of his presidency. Currently, a web search will yield periodic ratings for President Obama. Presidential approval ratings are collected and published frequently each year by many news agencies. Many of these data sets are collected by governmental agencies and are available to researchers who wish to analyze the data on their own. Corporations may also make archival data sets available to researchers who wish to study workplace behaviors such as work productivity and absenteeism.

Archival data offer researchers a means of collecting data quickly. Few resources are needed, as the data are collected by another agency or institution. However, archival data offer the researcher no control over the circumstances under which the data are collected, the sampling technique used, or the measures used to observe behavior. Researchers using archival data also have no control over how the data are coded, which can make comparisons difficult across groups or areas if data are coded differently by different organizations.

Archival Data: a data collection technique that involves analysis of preexisting data

Figure 3.3 Approval Ratings for President George W. Bush

![Figure 3.3 Approval Ratings for President George W. Bush](source: ABC/Washington Post Data retrieved February 17, 2008, from www.pollingreport.com/BushJob.htm.)
Content Analysis. Content analysis is a specific type of archival data observation technique that involves analysis of what someone has said (as in a speech or interview) or written (as in a book or article). This may involve analyzing the transcript of a speech someone gave, a written document, or a recorded statement. In content analysis, a researcher is analyzing a record of what someone has said or written in the past; thus, no interaction takes place between the research participant and the researcher. A coding scheme must be developed by the researcher to determine which behaviors will be considered in the analysis. This type of analysis can be resource intensive and time-consuming. Table 3.4 contains some examples of behaviors that psychologists might want to observe. For each behavior, consider how psychologists might use each of the observation techniques described in this chapter to measure the behavior.

### Table 3.4  Thinking About Observations

For each behavior listed below, consider how you might observe the behavior using (a) naturalistic observation, (b) surveys and questionnaires, or (c) systematic observation. Be sure to operationally define the behavior before you describe how the observations would be collected. Also consider the limitations of each observation method as you choose one for each behavior.

1. How do humans (or animals) solve a problem?
2. How do people react to bad news?
3. What types of people are most likely to disclose personal problems to others?
4. How do groups of children organize themselves to complete a task?
5. What behaviors characterize people with attention deficit hyperactivity disorder (ADHD)?
6. What types of brain activity result when one consumes caffeine?
response variable. Dependent variables are measured in every research study. For some designs, only a single dependent variable is measured and the behavior is examined descriptively or causally (if the researcher is interested in a causal relationship and uses an experiment to study this relationship). Other designs examine relationships between multiple dependent variables. As a result, how a dependent variable is measured depends on the data collection technique used and what is learned about a dependent variable depends on the type of research design used.

Each of the data collection techniques described earlier can be used in any of the major research designs; however, practically speaking some techniques are more common in certain designs than in others. As each design is discussed, examples of the most common techniques used in that design will be described. In addition, some of the more common designs are discussed in further detail in Part II of this book. Note also that you may see the term research design applied to many aspects of a design. Here the term applies to the major categories of research designs that are used by psychologists to answer different types of research questions. What follows is a description of the major research designs used in psychological research with some examples of these designs.

**Case Studies**

In 1970, a woman walked into a welfare office in Southern California with her daughter Genie. After the woman was interviewed, it became clear that although Genie appeared to be about 6 or 7 years old, she was in fact 13 years old and had no language abilities. Her parents had kept her locked in her bedroom every day since she was very young. Genie did not attend school and had not been exposed to enough language from her family to learn how to speak. After Genie’s situation was discovered, psychologists became interested in her case and hoped to learn from her about the development of language and whether it can occur at such a late age in a child (Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974). Genie became the subject of intensive study by a number of individuals. From the case study of Genie, evidence was gained for a critical period of language development because Genie was raised with little language interaction with others and had difficulty learning language after she was rescued.

The goal of a case study is to gain insight into and understanding of a single individual’s (or just a couple of individuals’) behavior. Case studies can also be conducted for a group of individuals, such as an agency or institution. Typically, a case study involves intensive observation of an individual’s naturalistic behavior or set of behaviors. Thus, researchers often use naturalistic observations, interviews, or archival data (especially in the case of a famous individual) to learn about the individual’s behavior, although some case studies have also included systematic observations and surveys to learn about an individual. Case studies are often exploratory studies, wherein a researcher can learn about a behavior when little is known about it (e.g., unusual symptoms of a new disorder). When a researcher
is interested in testing theories about how behavior works or attempting to find a treatment that will help an individual or small set of individuals with a problem behavior, a small n design is typically used instead of a case study. The primary difference between case studies and small n designs is the goal of the researcher. In addition, small n designs are typically classified as experiments (see section on Experiments later in this chapter). Small n designs are further discussed in Chapter 13.

Some of the more well-known case studies conducted by psychologists have been with individuals who suffered a brain injury. Researchers study such individuals after their brain injury both while they are still alive (if possible) and after their death. Behaviors these individuals exhibit are then connected to the brain injury they suffered, which can be explored more extensively after their death. Physiological psychologists have gained a lot of important knowledge about brain functions through case studies of these individuals. A famous case is that of H. M. (typically these patients are identified by initials only or a pseudonym to keep the findings confidential for these individuals). H. M. suffered damage to a brain area known as the hippocampus during a surgery in 1953 that was done to help reduce his epileptic seizures (Hilts, 1996). As a result of the surgery, H. M. could no longer remember new experiences for longer than a few minutes. Through extensive study of this subject, psychologists learned that the hippocampus is an important structure for encoding or retrieving memories of events because H. M. lost the ability to retrieve memories of things that happened after his surgery (even though he still had access to memories of events he experienced before his surgery). From case studies of H. M. came a new set of questions about the hippocampus and exactly what function it serves in memory. In addition, researchers found that H. M. could improve on tasks over time, indicating that certain types of new memories were still available to him. For example, a recent study (Bohbot & Corkin, 2007) found that his ability to learn a spatial memory task was quite good, despite having severe amnesia (his memory for things he experienced lasts only a few minutes). Case studies of H.M. are still conducted today. Although he died in 2008, H. M.'s brain will continue to be studied to further our understanding of the importance of specific brain areas in cognitive functioning.

Other case studies have used archival data or content analysis of a document to better understand an individual’s behavior. For example, Lester (2006) recently examined the diaries of a man who committed suicide, in an attempt to identify specific events in his life that may have been connected to his choice to kill himself. Abramson (1984) conducted a case study of the sexual behaviors of a woman who called herself “Sarah,” after receiving a letter from her. Sarah had been abused (physically and sexually) as a child and was recovering from those traumas at the time she wrote to Abramson, a psychologist at the University of California, Los Angeles. Abramson then conducted interviews with Sarah to better understand how individuals recover from such traumatic experiences. Abramson learned a good deal about psychological resilience and recovery in an individual despite traumatic childhood events.

Case studies can even be done after a person has died. When Albert Einstein died in 1955, portions of his brain were preserved to allow it to be studied. Since then scientists have examined these brain sections to look for ways in which Einstein’s brain may differ from brains of other people. One difference that has been found is that Einstein’s brain contains many more neuron cells in a section of his brain’s cortex than in the brains of
control participants that were used as a comparison (Anderson & Harvey, 1996). In addition, a section of Einstein’s brain appears to have been improperly formed, allowing connections between areas of the brain that are not connected in other people’s brains (Witelson, Kigar, & Harvey, 1999). These differences may account for the intellectual abilities Einstein possessed, but simply examining Einstein’s brain does not allow researchers to make such strong conclusions. Instead, these studies allow researchers to start hypothesizing about which brain areas or characteristics may be important for advanced intellectual abilities, and these hypotheses can then be tested in further studies (see Figure 3.4).

Figure 3.4  Case Studies Can Provide a Unique View of an Individual

Due to their exploratory nature, case studies often focus on rare or unusual cases to gain some information about a behavior that is not exhibited by all individuals. This means that the behaviors examined in case studies often cannot be generalized to all individuals. Because of the reduced generality of the behaviors, case studies do not allow for strong tests of the cause of the behavior (experiments are best for that—as explained in a later section), and researchers must be careful about drawing conclusions about the causes of the behaviors they are studying. However, case studies can give researchers a starting place for investigations of a behavior or a set of behaviors. Thus, they serve an important purpose in psychological research, drawing attention to new research questions that can be further explored with additional studies.

**Case Study:** a research design that involves intensive study of particular individuals and their behaviors

**Small n Design:** an experiment conducted with one or a few participants to better understand the behavior of those individuals

---

**Correlational Studies**

Is insomnia related to depression? Do students who watch more TV have lower grade point averages (GPAs)? Do children who play violent video games behave more violently? Each of these questions can be explored in a correlational study (correlation means relationship). Correlational studies allow a researcher to examine relationships between variables and, if a relationship is found, predict values for one variable from values on the other variable(s). If a predictive relationship is examined, the variable that is used to make the prediction is called the **predictor variable**, and the variable that is being predicted is called the **outcome variable**. Therefore, the goal of a correlational study is to determine if different behaviors are connected and occur together. This type of study, however, still does not allow us to determine if one variable causes another to occur (again, only well-designed experiments allow researchers to test causal relationships). All we can learn from a correlational study is if two variables covary (i.e., change together—up, down, or in opposite directions from each other). Then researchers may be able to predict one variable from another.

If a relationship is found in a correlational study between two variables, it can take one of two forms: a **positive relationship** or a **negative relationship**. A positive relationship means that the values on the variables change in the same direction (up and down) at the same time. A negative relationship indicates that as values on one variable increase, the values on the other variable decrease. Figure 3.5 illustrates each type of relationship that might exist between GPA and the number of hours an individual watches TV per week. The graphs in Figure 3.5 are called **scatterplots**. In a scatterplot, one variable is placed on the x-axis, and the other variable is placed on the y-axis. The data points in the graph represent the scores on the two variables for each individual (horizontally for the first variable and vertically for the second variable). Thus, the data point that is circled in Panel B shows that one individual in this study reported watching 25 hours of TV per week and had a GPA of 3.25. You may also notice that there are very few data points below 1.0 on the GPA scale in the graph. This may be because very few students in college have GPAs lower than 1.0 (if they do, they do not remain in college very long).
Figure 3.5  Positive (A) and Negative (B) Relationships Between Number of Hours of TV Watched and GPA
Remember that a relationship between these variables does not mean that watching TV will cause one to have a higher or lower GPA or that having a higher or lower GPA will cause one to watch more TV. It is quite possible that a third variable (e.g., lack of interest in academic topics, a poor academic environment while growing up, good background in academic topics that are covered in TV shows) is causing the number of hours of TV watched and GPA to change. This is called the third-variable problem and it is the reason that researchers cannot determine causation from correlational studies.

Experiments

Have you ever wondered why it took so long for scientists to learn that smoking causes cancer? People have been smoking for hundreds of years, and medical science has been studying the effects of smoking on health for many decades. Yet warnings on cigarette packages did not appear until 1966, and claims about smoking causing lung cancer were tentative for many years. Scientists could show that smoking and cancer were linked but could not show that smoking caused cancer in humans in a definitive way. The reason these findings were tentative for so long is that ethically it is very difficult (if not impossible) to conduct an experiment to examine whether smoking causes cancer in humans. An experiment involves manipulating the presumed causal variable known as the independent variable. This means that in the smoking experiment, one group of people would be assigned to smoke and a similar group of people (similar in age, weight, health, diet, etc.) is not allowed to smoke. Obviously, researchers cannot force people to smoke (especially, if they have a hypothesis that it is harmful to them) or force people to never smoke. Thus, it has been very difficult (and has taken many correlational studies and animal experiments) to show that smoking is a cause of lung cancer, but this example illustrates what is required to test causal hypotheses: experiments.

As mentioned above, a key aspect of experiments that allows tests of causal relationships is the manipulation of the independent variable (or independent variables—an experiment can contain more than one independent variable). The independent variable is the factor that the researcher thinks may affect the observed behavior. Thus, data are collected from
the participants in an experiment under at least two different conditions. The data from these conditions are then compared with one another to see if there are differences caused by the independent variable manipulation. These different conditions make up the levels of the independent variable. For example, an experiment may involve subjecting one group of participants to treatment for depression, while another group receives no treatment. In this case, the treated group is the experimental group, because they receive the treatment condition, whereas the nontreated group is the control group, because they receive no treatment. Comparison of depression scores (the dependent variable in this experiment) for the two groups would allow the researcher to determine if the treatment was helpful (scores for the treatment group are higher than the control group’s), harmful (scores for the treatment group are lower than the control group’s), or makes no difference (scores for the treatment and control groups are similar). Another way to conduct this experiment would be to use two different treatments (e.g., talk therapy versus drug therapy) for the two groups. In this case, there is no control group; both groups of participants receive some type of treatment. Comparison of the scores for the two groups in this experiment would indicate which of the two therapies was more helpful or if they are similar. In fact, this comparison is a key feature of an experiment. The goal in an experiment is to examine the effect of the independent variable on the dependent variable. We do that by comparing the data observed in the different levels of the independent variable; this comparison tells us if the independent variable has an effect. In other words, the comparison across the levels of the independent variable provides the test of the hypothesis.

Another key feature of experiments is the control of factors that could affect the results but are not part of the manipulation of the independent variable(s). These extraneous factors are called confounding variables. If confounding variables are not controlled, the causal relationship between the independent and dependent variables will be unclear. For example, in the depression examples discussed earlier, the severity of the depression the participants have before the experiment could be a confounding variable because participants with more severe depression...
may show less improvement with therapy than participants with less severe depression. If the participants with more severe depression are inadvertently assigned to the therapy group in the first example or to one of the therapy groups in the second example, they may show no effect of therapy, even if an effect does exist. In other words, the severity of the depression may mask the causal relationship between therapy group and depression at the end of the study. Thus, the researcher should control for this factor in the experiment, perhaps by randomly assigning participants to groups. The control aspect of experiments is discussed further in Chapter 11.

Systematic observations and surveys are the most common observation techniques used in experiments. These observation techniques allow for more control over the measurement of behavior than the other techniques (see Table 3.2). The IAT procedure described earlier (and in Figure 3.2) provides an example of the use of systematic observation in an experiment. An independent variable in an experiment using the IAT procedure is the type of word (or image) pairing the participant is responding to. The dependent variable is often the amount of time the participants take to respond to the word pairings.

**Quasi-Experiments**

In some cases, researchers want to compare the behavior of groups of individuals but are unable to manipulate the characteristic on which the groups differ. For example, suppose you wanted to compare the behavior for older and younger adults. You cannot randomly assign individuals to be of a specific age, so age cannot be a manipulated variable. This means that you lose some control over alternative explanations of the data, because participants are not randomly assigned to the groups. In this example, if younger and older adults differ on the measured behavior, age may have caused the difference or something that varies with age (e.g., experiences of different generations with technology) may have caused the difference. Thus, conclusions from quasi-experiments must be more tentative than conclusions from experiments, where a true independent variable is manipulated. However, quasi-experiments involve group comparisons just as experiments do, and data from quasi-experiments are often analyzed the same way as they are analyzed in experiments (see Chapter 14 for more on data analysis). The key difference between experiments and quasi-experiments is that the random assignment of participants to groups is missing in a quasi-experiment (Shadish, Cook, & Campbell, 2002).

Some studies also have less control over conditions under which data are collected in a repeated-measures design. For example, suppose a researcher is interested in the change in attitude regarding trust of politicians after taking a political science class. In this study, attitude is measured before and after the class in what is called a pretest-posttest design. If attitude changes from pretest to posttest (getting either better or worse), this change may be either because of the class or because of other events that occurred in the time between the
tests (e.g., a political scandal may have occurred in this time). Thus, the causal relationship between the political science class and attitude change is less clear. For this reason, pretest-posttest designs (that do not include a control group) are considered quasi-experiments. Other types of quasi-experimental designs are considered in Chapter 12.

CHAPTER SUMMARY

Reconsider the questions from the beginning of the chapter:

- How do psychologists observe behavior? There are some common techniques used by psychologists to observe behavior described in this chapter.
- What are some common techniques for observing and recording behavior in different situations? The common techniques used by psychologists to observe and record behavior are naturalistic observations, surveys/questionnaires, systematic observations, and archival data.
- How do psychologists use observations to learn about behavior? Each technique can be used in different research designs to allow psychologists to answer different types of questions about behavior.
- What questions about behavior do the different research methods allow psychologists to answer? Different research designs (e.g., case studies, correlational studies, experiments, and quasi-experiments) allow researchers to ask different questions about behavior. Case studies allow descriptive questions to be answered for a single individual or institution. Correlational studies allow descriptive and predictive questions to be answered about behavior. Quasi-experiments and experiments allow comparisons among groups, with experiments answering causal questions about behavior.
- Which research method is best when asking about the cause of behavior? Experiments are the best method to use when asking causal questions about behavior.

THINKING ABOUT RESEARCH

A summary of a research study in psychology is given below. As you read the summary, think about the following questions:

1. The study below is described as an experiment. What aspects of the method indicate that the researchers used an experimental research design?

2. Describe the variables in this study (independent variable and levels, dependent variable).

3. Based on the description, which observation technique do you think the researchers used in this study? In what way did their design avoid some of the disadvantages of this technique? What disadvantages of this technique are possibly affecting their results in this study?

4. What types of extraneous variables should the researchers be concerned about in this study?

5. How would you judge the external validity of this study?
Purpose of the Study. The researchers examined preferences for beer under conditions that varied in terms of when information about an ingredient of one of the beers was given: before tasting, after tasting but before preferences were indicated, and never (no information was given to one group about the ingredients). The ingredient given is one that most people think should make the beer taste worse. The research question was whether the timing of the ingredient information would affect the preference for the beer by influencing one’s expectation of taste of the beer. Preference for the beer with the undesired ingredient should be lower in any condition where the information influences the preference.

Method of the Study. Pub patrons in Massachusetts were asked to participate in a taste test of two types of beer labeled “regular beer” and “MIT brew.” The “MIT brew” contained a few drops of balsamic vinegar (the vinegar apparently changed the flavor of the beer very little). Participants were randomly assigned to one of the three groups that differed according to when information was given: blind group (no information given), before-tasting group (information given before tasting), and after-tasting group (information given after tasting but before preference was indicated). All participants were given a small sample of each beer to taste. They were asked to indicate which of the two beers they preferred.

Results of the Study. In the blind condition, the “MIT brew” was preferred more often (about 60% of the group) than the before condition (only about 50% of the group), indicating that ingredient information had an effect before tasting. However, the “MIT brew” was also preferred more often in the after condition (just over 50% of the group) than in the before condition and was not preferred less often than the blind condition, indicating that when ingredient information is given after tasting, it does not affect preference. Figure 3.6 presents the means of the three groups.

Figure 3.6 Results of the Beer-Tasting Study

SOURCE: Results from Lee, Frederick, and Ariely’s (2006) study.
Conclusions of the Study. The researchers concluded that the timing of information about a beer-drinking experience affects preference for the beer. Their results indicated that when information about the beer ingredient was given before the participants tasted the beer, it affected their tasting experience (and their preferences), but when information was given after the participants tasted the beer, it did not affect their experience or their preference. More generally, this study showed that our expectations of our perceptual experiences affect how we judge those experiences.

COMMON PITFALLS AND HOW TO AVOID THEM

Problem: Concluding cause without manipulation—nonexperimental studies are sometimes interpreted as providing causal information.

Solution: To test causal relationships, an experiment with a manipulated independent variable is needed.

Problem: External validity versus internal validity—some view external validity as more important than internal validity in all psychological research studies.

Solution: Be aware that high internal validity is required to test causal relationships but that internal and external validity typically trade off in psychological studies.

Problem: Use of control groups in experiments—students sometimes incorrectly believe that all experiments require a control group.

Solution: Remember that control groups are important in experiments if a single treatment is being evaluated or the effectiveness of multiple treatments is being tested, but control groups are not required in every type of experiment.

USING RESEARCH

In 1998, a group of British doctors published a paper suggesting that there may be a link between vaccines for measles, mumps, and rubella diseases and incidence of autism (Wakefield et al., 1998). This means the researchers were suggesting there is a positive relationship between these vaccinations and occurrence of autism in children that were tested. Their findings started a controversy over the causes of autism. Media sources in the UK reported the findings of this paper suggesting that having children vaccinated can cause them to develop autism. As a result, many parents in the UK and the U.S. refused to have their children vaccinated out of fear of their children becoming autistic, and the incidence of childhood diseases, such as measles, increased in the UK in the years after the paper was published. Based on the information you read in this chapter, why should parents hesitate to conclude from the findings of this paper that vaccinations can cause autism? What error did the media sources make in publicizing the study’s findings? Why should the public refrain from making important decisions (such as refusing vaccinations) based on the findings of a single study? If this last question is a bit difficult to answer, consider what has happened more recently in this story: In February 2010, the journal that published the original
paper describing this study retracted the paper because the study was found to have serious methodological flaws that were not entirely evident in the original report of the research. Consider what type of study would be necessary to show that vaccinations cause autism in children. How would such a study need to be done? Is this a realistic study to conduct? Why or why not?

**Test Yourself**

Match each research design below with its appropriate description.

1. Experiment  (a) Design that focuses on observing just one or a few individuals
2. Case study  (b) Design that will allow one to look for relationships between measured variables
3. Correlational study  (c) Best design for testing a causal relationship
4. Quasi-experiment  (d) Design that allows comparison of groups but does not contain a true independent variable

5. Reread the scenario described at the beginning of the chapter about increasing work productivity. Use the concepts you read about in this chapter to design a study to test the boss's suggestion. Be sure to operationally define work productivity and decide what technique you will use to observe this behavior. Choose the research design that best fits the question you are trying to answer.

6. Suppose you are interested in testing the hypothesis “The herb ginkgo biloba causes one to have better memory.”
   (a) What is the best research design to test this hypothesis? Why?
   (b) Describe the study you would conduct to test this hypothesis.

7. ________ validity indicates that a study’s results can be generalized to other individuals and real-life situations. ________ validity indicates that a study’s results provide causal information about the variables tested.

8. For each description below, indicate which data collection technique was used:
   (a) Medical records of patients with depression are examined to determine how often these patients attempt suicide based on what type of treatment they received
   (b) Participants are asked to perform a task on a computer where they must unscramble sentences as quickly as possible—the amount of time it takes to complete the task on each trial is recorded
   (c) A series of statements regarding their alcohol consumption behaviors is presented to participants—they are asked to rate their agreement with each statement on a scale from 1 to 5
   (d) Students in a college class are observed to record the number of times they exhibit behaviors indicating lack of attention to the lecture
9. If a study finds that as self-esteem goes up, symptoms of depression decrease, this study has found a _________________ relationship.

10. __________ research designs are typically used when a researcher wants to explore the behavior of an individual or group of individuals to better understand unusual or atypical behaviors.

**STUDENT STUDY SITE**

Visit www.sagepub.com/mcbride2e open access student study site for the following resources:

- Web Quizzes
- Chapter Summary
- “Learning From SAGE Journal and Encyclopedia Articles”
- Thinking About Research
- eFlashcards

**Answers:** (1) c, (2) a, (3) b, (4) d, (5) There are lots of ways to test this. One example is to use a measurable work behavior (such as number of reports written in a week or number of clients served per day) as the operational definition of work productivity. The study could compare work productivity change (difference between work productivity before the study starts and work productivity during the study) for workers on different floors of the office (where one floor has the cappuccino machine and the other floor does not). Or work productivity could be measured for a week without the machine and then measured for a week with the machine and compared for the 2 weeks. Both designs involve a quasi-experiment. A true experiment would be difficult to conduct in the workplace. (6a) This is a causal hypothesis; thus, an experiment is the best design. (6b) A manipulated independent variable that includes random assignment of participants to a ginkgo biloba group and a control group that receives a placebo should be included. (7) External, Internal (8a) archival data, (8b) systematic observation, (8c) survey/questionnaire, (8d) naturalistic observation (9) Negative (10) Case study.