2 METHODS OF DEVELOPING SOCIAL PSYCHOLOGICAL KNOWLEDGE

1964. The rape and stabbing death of 28-year-old Kitty Genovese in front of her Queens, New York, apartment building drew headlines. Although the attack went on for 30 minutes, complete with screams for help, none of the residents at the building reported the incident while it was ongoing.[1]

1983. A gang rape occurred at a bar, with a woman assaulted on a pool table in front of witnesses.[2]

2009. Outside a homecoming dance at a Richmond, California, high school, a 15-year-old was assaulted and raped while people walked by, some of whom even took pictures on their cell phones. Yet none of them reported the incident.

2016. Marques Gaines left a Chicago-area bar where he had been drinking and dancing. Assailants punched and knocked him unconscious into a crosswalk outside a 7-Eleven, robbing him of his cell phone and debit card. Pedestrians strolled along the sidewalk and lingered on the street corner, but no one attempted to lift him from the pavement, block the flow of traffic, or call for help until after a taxi ran him over, resulting in his death.

2021. A man approached a 65-year-old Asian American woman outside an apartment building near New York's Times Square. He kicked her aggressively to the sidewalk and stomped on her head and upper body, then walked away. Workers in the apartment lobby viewed the incident but did nothing.

In these five different incidents—years and miles apart, involving both male and female victims—bystanders observed the event and then hesitated or even failed to offer help in a timely fashion. In all cases, perpetrators of the crimes were ultimately punished. Nonetheless, you might ask, How can people just stand by and watch something horrific happen to another human being? That, indeed, was the research question posed by psychologists Bibb Latané and John Darley in the wake of the Kitty Genovese murder and sensationalized media accounts that accompanied it. The development of their classic work on bystander intervention in emergencies (Latané and Darley 1970), augmented by other scholars throughout the years, epitomizes the logic of scientific inquiry and the use of various methodologies to address substantive and theoretical questions of interest to social psychologists.

As the definition of social psychology offered in Chapter 1 emphasizes, to understand the relationship between the individual and society requires systematic study. Of course, we all have hunches about why people act the way they do. For instance, in the wake of the Kitty Genovese murder, some people attributed the commotion to a "lovers' quarrel" or their hesitancy to their

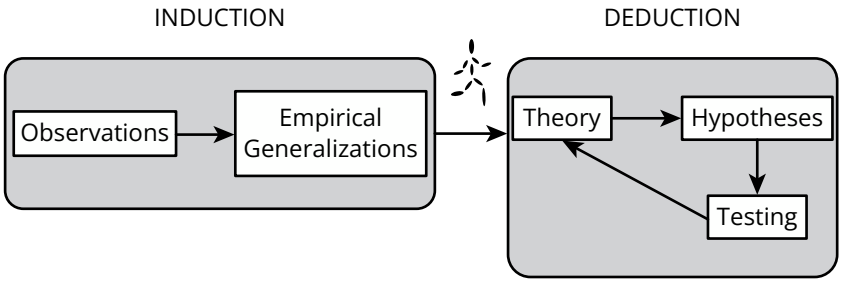
"fear of getting involved" or "fear of getting hurt." The first account implies that people did not perceive an emergency situation, whereas the others suggest that individuals weighed threats to themselves or just how much time they would lose if they had to answer police questions. Latané and Darley (1970), of course, considered these intuitive or commonsense explanations but went further by developing a *theoretical argument* and bringing *systematic evidence* to bear upon predictions derived from their argument about the conditions facilitating or hindering provision of interventions. Given what they had observed, they asked this: *Under what conditions* are people likely to help or not help in emergency situations?

In this chapter, we outline steps necessary for producing scientific evidence in social psychology. We apply the logic of such inquiry to the study of helping behavior generally. We use actual empirical research on bystander interventions and helping to illustrate the different methodologies that social psychologists use to investigate people's thoughts, feelings, and behaviors in social situations. We conclude by considering how social psychological research extends beyond commonsense knowledge and provides deeper insight, for example, than the musings of people at a party, the speculations of family members, or the conjectures of workers about why a friend, a cousin, or a boss acted in a particular way.

THE LOGIC OF SCIENTIFIC INQUIRY

Imagine reading the initial *New York Times* article on the Kitty Genovese murder or reading about any of the more recent incidents noted at the beginning of this chapter. You may feel stunned that (purportedly) so many people witnessed some element of the attack but did little to intervene. As you discuss the incident with your friends, someone might say, "Ah, do you remember the recent robbery at the convenience store? Two teenagers assaulted the cashier and then absconded with the cash in the register, running out of the store in full view of a dozen people getting off a bus. No one intervened." Someone else might raise another incident, such as the failure of anyone—despite the presence of an array of tourists along the embankment—to aid boaters whose vessel capsizes on the River Thames in London.[3] A third person could extend the discussion to why people don't contribute to charity fundraisers, where such help may assist a local hospital, food bank, or community center. As stories of similar incidents unfold (regardless of their veracity), you have taken the first step in any social scientific endeavor: making **observations**. Latané and Darley (1970) perhaps began with such a process before launching their systematic research program. Figure 2.1 illustrates the basic components of the logic of scientific inquiry.

FIGURE 2.1 ■ Components of the Logic of Scientific Inquiry



In the course of our daily lives, we make many observations, which we may later share with family, friends, or even acquaintances met while riding the bus. In effect, we are describing reality as we experienced it. Of course, our observations may be casual and flawed by our expectations and existing beliefs. For example, if you have heard that the required statistics course is the toughest course in your major, you may perceive after the first day that the instructor is hard-nosed and the homework arduous. While scientific inquiry begins with observations, it also requires a more *systematic description of reality* involving observations of similar incidents, across time, across locations.

Systematic observations attempt to exclude preconceived notions of the observers that may lead to selective observations, overgeneralization from events, or simple inaccuracy. To examine scientifically whether a professor or course is "tough," for example, might involve comparing the syllabi for that professor (or course) and others; identifying the discipline; and looking at the amount of reading, the number of assignments, the availability of tutorials, and the like. We could also compare professors' classroom policies on things like makeup exams, penalties for late assignments, and so forth as well as their demeanor on the first day of class (friendly? empathetic?) as a means of tapping into whether they are so rigid and inflexible as to deserve to be called hard-nosed. Plus, we could ask students to rate the toughness of their professor on a specific scale, not simply rely on the anecdotal evidence of some students. Moreover, we could extend our study beyond our own campus to look at what constitutes professorial toughness in other college settings. In other words, a scientific approach to describing reality ultimately involves gathering observations in a careful, orderly way, minimizing as much bias as possible. And in so doing, we may develop **empirical generalizations**, which refer to relationships between factors across time and across locations grounded in data. For example, we could look to see whether factors like discipline, type of institution, course topic, characteristics of the instructor, or level of the course influence whether professors appear to be tough. We might also see whether the same factors that affect perceptions of toughness at one university (say a large state school) likewise influence perceptions at a small liberal arts college. In so doing, we are looking for patterns that vary in terms of place or location. It might be further interesting to see if the impact of a particular factor changes over time as well. Establishing patterns of variation across time and place in the relationship between factors of interest provides the basis for empirical generalization but not the reasons for the observed relationships.

The logic of scientific inquiry captures two forms of logical reasoning—(1) induction and (2) deduction—that represent different stages in knowledge on a topic. **Induction** characterizes the process involving the movement from casual to systematic observations allowing the emergence of coherent patterns that provide the basis for theorizing (Babbie 2021). Inductive reasoning acts as a starting point in scientific inquiry typically when little is known about the phenomenon under study. This form of thinking stimulates research that may ultimately produce empirical generalizations. Recognition of a pattern, however, does not suffice to explain it.

The next step is for researchers to offer an **explanation** for *why* certain patterns exist. Karl Popper (1963), a renowned philosopher of science, suggests that the process of going from observations to abstract theorizing about a phenomenon constitutes what might be called a great leap of faith (as illustrated by the character in the figure leaping from empirical generalizations to theory!). In some ways, it does not matter where the theoretical ideas came from. What matters is the logic of the explanation, the ability to test predictions derived from it, and the way tests confirm or disconfirm the theoretical ideas. **Deduction**,

the specification of an abstract theory from which hypotheses deduced from it can be tested through further specifically focused observation in a research study, captures such reasoning (Bryman 2016). Several terms are central to understanding deduction.

A **theory** is a systematic explanation for the observations that link two (or more) aspects of social life (Babbie 2021). For example, a cost–benefit or social exchange approach to helping (see Dovidio et al. 2006) suggests that people are motivated to maximize their rewards (positive material or social outcomes) and minimize their costs (negative material or social outcomes) in social interactions. Accordingly, they are likely to weight the rewards and costs in their decision about whether or not to intervene in a helping situation. Such reasoning calls attention to several important concepts: rewards, costs, and intervention. At the abstract level, **concepts** represent aspects of social reality by naming groupings of observations that share similar features (Bryman 2016). The groupings may suggest underlying variation or distinctions from another grouping. For example, in trying to explain "interventions in emergency situations," a key concept is "intervention." An intervention occurs when an individual does something to change the ongoing situation by providing help to someone in need. Interventions vary in that some people intervene and some people do not in emergency situations. Similarly, the concepts of rewards and costs may vary from low to high. (How these concepts are "observed" for research purposes depends upon the nature of the situation or the nature of the study, as described more in the next section.)

Theories consist of propositions that may outline underlying motivations or identify how concepts are linked. These propositions form the core logic of the theoretical argument. From them, researchers may deduce a **hypothesis**, which represents a testable statement indicating how a set of factors (represented conceptually) are related. For instance, from the social exchange approach to helping, an investigator might hypothesize the following: People are more likely to intervene when rewards for doing so are high and costs are low. In addition, researchers may define the **scope** of their theory, or the conditions under which they expect their hypotheses to hold. A scope condition in assessing the social exchange helping hypothesis is that it may apply only when individuals are physically able to provide the assistance necessary in the situation (if they are not physically capable of doing so, rewards and costs may not matter at all).

Testing a deductive hypothesis requires bringing observations (of various types, through various methodologies) to confirm or disconfirm the hypothesis and, by implication, the explanation offered by the theory. If new observations (not those giving rise to the theoretical ideas in the first place!) provide no support for the hypotheses, researchers may first try to test the hypotheses under different conditions or using different methodologies. Continued lack of support, or falsification, however, inevitably requires reconsideration of the theoretical argument. Next, we use the development of an explanation for "why people fail to intervene in emergency situations" to illustrate movement from induction to deduction.

Bringing Data to Bear on a Hypothesized Relationship

Think about why people may or may not provide help in an emergency situation. What sort of factors does your list include? Are they factors focused on the individual's personality, friendliness, empathy, or the like? Or do the factors on your list describe characteristics of the situation, like the nature of the help that seems to be needed or the number of people present? As previously noted, Latané and Darley (1970) may have examined various

incidents about when people offered help and when they did not. They moved beyond simply suggesting that it is something about individuals—their apathy or lack of empathy (i.e., individual-level explanations)—by focusing on the situations in which bystanders are embedded and the decisions they make (i.e., contextual explanations). Their theoretical argument (Latané and Darley 1970) requires that bystanders (1) notice the situation, (2) interpret it as an emergency, (3) recognize their personal responsibility to help, (4) know how to help, and (5) make the decision to act in spite of potential negative consequences to themselves. While each step has been extensively investigated (see Dovidio et al. 2006), we focus on the third step to illustrate aspects of the logic of scientific inquiry, especially the development of a theory and subsequent testing.

Recognizing personal responsibility for helping (Step 3) is more likely when only one person witnesses the need for help. In thinking about the Kitty Genovese situation, many people were (ostensibly) present. Thus, no one individual may have felt personal responsibility for helping. Latané and Darley (1970) propose that when many people are present in an emergency situation, responsibility to provide help is diffused. That is, individuals believe that others will offer help and they themselves do not need to do so. When all individuals absolve themselves of personal responsibility, little bystander intervention emerges. With this theoretical reasoning in mind, Latané and Darley (1970) could hypothesize that diffusion of responsibility is negatively related to intervention in an emergency situation.

This hypothesis involves a testable relationship involving two concepts, each of which may vary from low to high or absent to present: (1) diffusion of responsibility and (2) intervention. When a **relationship** exists, variation in one concept coincides with variation in the other. Our diffusion of responsibility and intervention hypothesis also suggests a "direction" of the relationship between the two concepts. A *negative relationship* is one in which an increase in the variation represented by one concept is associated with a decrease in the variation represented by another concept. Thus, the more a situation allows for diffusion of responsibility, the less likely people are to intervene. In contrast, in a *positive relationship*, variation represented by one concept is associated with an increase in the variation represented by another concept. The following hypothesis suggests a positive relationship between feelings of empathy (experiencing the same emotion as others) and helping behavior: Individuals who are high in empathy are more likely to intervene in an emergency situation than those who are low in empathy (see Batson and Oleson 1991). Another way of stating that hypothesis is this: Empathy is positively related to interventions in an emergency situation. These hypotheses, of course, are posed at the conceptual level.

To test a hypothesis, researchers must **operationalize** the concepts, which involves translating them into what can be readily observed or measured. Operationalization allows movement from the abstract level of theorizing to the concrete level of measurement. For example, in assessing how diffusion of responsibility in a helping situation could unfold, researchers might measure how many people are present when emergency situations arise and track whether anyone intervenes with help. The more people present, the greater the diffusion of responsibility. Of course, the nature of intervention must also be operationalized; people might provide direct aid, call 911, find a police officer, or the like, depending upon the nature of the situation. As described more fully as follows, when investigators study helping behavior using experimental methods, operationalization pertains to what can be manipulated. For example, in early helping behavior studies, researchers created situations in which a target respondent observes someone in distress

when no one else is present, representing low diffusion of responsibility, or when five or more people are present, indicating high diffusion of responsibility.

In the hypothesis (and experiment) involving diffusion of responsibility operationalized as the number of people present, that concept and its operationalization create the **independent variable**, or presumed cause, which is assumed to affect a belief, feeling, or behavior. The presumed effect is labeled the **dependent variable**. Despite the language of presumed cause and effect, assessing **causality** between variables requires additional considerations. First, the two variables must be *statistically associated*, meaning that their relationship does not rest upon chance alone but rather signifies a real relationship. Second, the independent variable must precede in time the dependent variable to ensure appropriate *time order.* And third, the observed association between the independent and dependent variables must be *nonspurious*—that is, not caused by some other, perhaps unmeasured, variable to which the two other variables are related.

For example, an oft-cited empirical observation is that ice cream sales are positively related (statistically associated) to drownings. You might ponder why. Are people who drown failing to wait the hour after eating ice cream to return to their water activities? But why are ice cream eaters the only ones to violate the frequent parental dictum to wait that hour? Thinking further, you might ask, When are ice cream sales likely to be the highest? When are drownings likely to be the highest? Both are highly likely when the weather is hot. Thus, the observed positive relationship between ice cream sales and drownings is due to both variables being independently related to outside temperatures and not causally related to each other. The ice cream sales and drowning relationship is spurious and thus fails to confirm causality between eating ice cream and drowning. Generally, investigators may attempt to include and measure factors that they think may account for spuriousness between their focal concepts and associated measures in their analyses so that they can feel confident that their hypothesis tests are sound.

Thus, to bring data to bear upon a theoretically derived hypothesis requires translation from the abstract to the concrete and consideration of other processes that may impinge on the focal relationship. Importantly, hypothesis testing provides evidence that may confirm or disconfirm the underlying theoretical reasoning or the explanation offered for a particular phenomenon. Especially when hypotheses are disconfirmed (i.e., evidence fails to support the proposed relationship), researchers may reconsider their arguments and potentially alter their explanation or specifically identify conditions under which (through further testing) the proposed relationships are confirmed. Next, we examine the different methods of collecting data to test a hypothesis and in so doing consider how researchers employ various means to take into account other processes or factors that may impinge on their hypotheses.

SOCIAL RESEARCH METHODOLOGIES

In previous sections, you took the role of "researcher," beginning with your own observations of helping behavior, then developing an abstract theoretical argument to get at the conditions under which individuals might be likely to lend a hand. Armed with knowledge of previous patterns and with a potential understanding of *why* people might help, you need to determine how to collect further data to test your hypothesis. You would ask yourself, "What methodology might work best in trying to understand why people help others in emergencies?" Should you simply ask people, perhaps in a survey, how likely they

are to help others in need? Or maybe you could observe at a busy gathering place, waiting for someone to fall, and then watch to see who helps. Doing so might take a lot of time, so instead you may enlist a friend to act out a "falling scenario" and then observe people's responses. Researchers often ponder their methodological options before settling on a particular data collection strategy.

When Latané and Darley (1970) first began their investigations into helping behavior, they had to determine the most suitable way to collect data that would allow them to test their hypothesis. Indeed, the first step for any researcher is to ask, "Given my research question, what sort of data will allow me to address it?" To some extent, *the approach to data collection depends on how much knowledge exists on a topic*. As described previously, when scholars know little about a particular social process, they may begin with a series of observations in different types of situations. In contrast, once a theoretical explanation has been developed, they may opt to test derived hypotheses using methods that afford them the greatest amount of control to isolate the predicted relationship. By doing so, they can feel confident that they have minimized the influence of other factors on that focal relationship.

Social psychologists use a range of methods, including the experiment, written surveys, oral interviews, naturalistic observation, and archival investigations. All these methods may be used to produce **quantitative research**, which involves quantifying the collection and analysis of data—that is, using numbers! (Bryman 2016). For example, in a written survey, respondents might be asked to indicate on a numeric scale ranging from 1 to 7 how much they disagree or agree with a particular statement. Researchers may represent the extent of agreement with the statement by pointing to the average of all responses, where a 2.2 would suggest disagreement and a 6.7 strong agreement. The latter three methods, in different ways, may also be the means of developing **qualitative research**, which typically pertains to using words, rather than numbers, to capture and analyze the data (e.g., Glaser and Strauss 1967; Tavory and Timmermans 2014). For example, in an oral interview, researchers might probe how individuals who have one Black parent and one white parent see themselves (i.e., Do they identify as Black? White? Biracial?). A qualitative, compared with quantitative, research strategy may require greater interpretation of the data and construction, after data collection, of the nature of the social phenomena under study. Next, we address these fundamental methodologies. We use both quantitative and qualitative studies related to helping behavior to illustrate each methodology.

The Experiment

Imagine being in a situation in which you overhear an epileptic seizure. You know that you are alone and thus may be the only person who can provide aid. What do you do? Now imagine that four other people are likely to have overheard the same seizure in progress. What do you do? This scenario is the basis for one of the first bystander intervention studies conducted by Darley and Latané (1968). These researchers argue that despite humanitarian norms to help people in distress, *situational factors* weaken such norms by decreasing perceived personal responsibility to provide assistance. The number of bystanders, as previously noted, may allow diffusion of responsibility, which reduces the likelihood of intervention. Additionally, Darley and Latané argue that, with such diffusion, the punishment or blame for not acting is slight, and the inability of bystanders to observe each other makes it easy to assume someone else is taking action. These additional situational conditions allow observers to rationalize their own inaction.

Laboratory Experiments

To test their ideas, Darley and Latané (1968) designed a laboratory experiment. An **experiment** involves the exposure of study participants to specially designed situations and the systematic recording of their reactions. Those specifically designed situations represent manipulation of levels of the independent variable denoted by a hypothesis (while holding constant other factors), whereas the reactions capture the dependent variable. Experiments afford the researcher the means to assess causality because the manipulation of the independent variable unquestionably precedes the measurement of the dependent variable, thereby meeting the time order criterion.

Additionally, a fundamental characteristic of the experiment is **random assignment**, which involves the random or "by chance" (like flipping a coin) allocation of study participants to the experimental conditions constituting levels of the independent variable. Random assignment ensures that each participant has an equal chance of being assigned to any specific experimental condition. As a consequence, individual-level characteristics (such as empathy, shyness, or dominance) are evenly spread out across conditions, ensuring that the groups assigned to particular conditions are equivalent. Such equivalency helps to minimize the impact of factors that might influence measures of the dependent variable, thereby decreasing the likelihood that the observed relationship between the independent and dependent variables is spurious. For example, assume that individuals vary in terms of empathy. If all highly empathetic study participants were in the experimental condition of being the only bystander to an epileptic seizure, researchers would not know whether it was "empathy" or the nature of the condition that produced higher rates of helping than in other conditions. But, with random assignment, highly empathetic participants would be spread across all variations of "number of bystanders present."

Darley and Latané (1968) operationalized the concept of diffusion of responsibility by varying the number of people that the study participant thought could overhear the epileptic seizure of the ostensible victim in need of help.[4] Seventy-two male and female college students participated, and each was randomly assigned to one of three group size conditions: two (consisting of participant and victim); three (participant, victim, and one other); and six (including four others). The experimenter claimed that "he was interested in learning about the kinds of personal problems faced by normal college students in a high pressure, urban environment. He said to avoid possible embarrassment about discussing personal problems with strangers that several precautions had been taken" (Darley and Latané 1968:378). These "precautions" ensure creating a situation that circumscribes the situation in a manner consistent with researchers' theorizing. Thus, study participants remained anonymous within individual cubicles and had no communication with others, so they could not determine how others would respond to the epileptic seizure (the experimenter indicated that he would not be listening in during the initial discussion).

The researchers used prerecorded voices so that all study participants had a similar experience during the discussion. The "victim" spoke first, indicating that he was prone to seizures. After the study participants and the "others" presumed to be in the situation had a chance to talk, the victim began again, but this time, sounding incoherent, he noted that he was having a problem and needed help. At that point, the experimenter began to assess the focal dependent variable: time lapse before helping. Time was measured from the beginning of the victim's "seizure" until the study participant emerged from the cubicle to report the emergency to the experimenter.

The Darley and Latané (1968) study highlights the hallmarks of experimental methodology. The laboratory afforded the researchers a great deal of control. Such control allowed them to isolate theoretically relevant variables of interest and, through random assignment, examine the impact of group size on the likelihood of helping, without concern about how other extraneous factors (those not represented in the hypothesis) might impact the focal relationship. Thus, experiments provide a high degree of **internal validity**, defined as the extent to which findings regarding a causal relationship are likely to be sound. When experimental procedures capture researchers' intentions, without faulty manipulations, poor measurement, or contamination by extraneous variables, the findings produced are likely to be internally valid. Darley and Latané found clear support for their hypothesis, with 85% of participants in the "alone" condition providing help, 62% in the three-person group, and only 31% in the group of six. A similar pattern emerges in terms of how many seconds it took study participants to seek help. This effect of group size, as representing diffusion of responsibility, has emerged in many other studies as well. Replication of a study's results through repeated studies signals that findings are **reliable**.

While laboratory experiments such as Darley and Latané's (1968) with high internal validity and potential for replicability are an excellent way to examine causal hypotheses, several concerns do arise. First, one may be reluctant to make sweeping generalizations based on the responses of 72 undergraduates behaving under controlled circumstances. Experimental results tell us little directly about the natural world in which we live. They do, however, allow confirmation or disconfirmation of theoretical ideas, which have bearing upon understanding the social world. The theory acts as a bridge from the laboratory to the "real world" (Zelditch 1969). With this conception of the experiment and the theory driving it, criticisms that experimental situations are artificial grow moot (Webster and Sell 2014). Such artificiality allows incorporation of theoretically presumed causes while eliminating factors that may contaminate the relationship under study. Less artificial situations are more complex and may include other factors that enhance or minimize the impact of the presumed cause on the hypothesized effect.

A second concern is that even in well-designed experiments, threats to internal validity do arise. Study participants are human and thus may not interpret the manipulations in the manner that experimenters have planned. When that happens, results may not really indicate the effect of the independent variable on the dependent variable. Experimenters often include "manipulation checks" at the end of an experiment so they can assess the extent to which they have successfully operationalized their independent variables. Such checks may take the form of questions in a brief survey or asked by experimental assistants at the end of an experimental session. In the helping experiment, a manipulation check would be to ask the participants how many people were in their group to make sure they correctly could recall the size of their group.

Assuming that the manipulations are strong enough, other threats to internal validity include demand characteristics and experimenter effects. **Demand characteristics** take the form of activating study participants' preconceived notions of the study's purpose, which in turn lead them to behave in a certain way (thus, these are sometimes called subject effects). For example, if participants know that they are in a study explicitly about fairness in interpersonal relationships, that study label may entice them to act in what they imagine is the fairest way possible, regardless of the actual experimental conditions. Researchers are thus careful to eliminate any words or actions that may cue participants into particular behaviors. Indeed, experimenters must circumscribe their own behaviors

to avoid conveying to participants anything about what is expected of them to confirm the hypothesis. **Experimenter effects** are expectations for behavior unwittingly communicated by the people running the study. For example, researchers who nod far more to participants in the "alone" condition of a helping study compared to those in the "large group" condition when asked if they will be available during the experimental session are unconsciously conveying ease of helping to people in the condition expected to help more. That nodding contaminates the impact of the manipulated factor.

A third concern often voiced about experiments is the use of **deception**, which includes providing false information or withholding information with the goal of leading study participants to believe that something is true when it is not (Sieber 1992). The experimenter deliberately misleads the participants into believing something that is untrue. In the Darley and Latané (1968) experiment, study participants were led to believe that they were in groups of various sizes, when in reality each person was alone in the study. The recorded voices merely projected the image that others were involved in the discussion. When used as a means to manipulate independent variables, deception has scientific value. But the nature of the deception must be weighed in terms of the harm it might cause to study participants, as discussed further in a subsequent section (see Hegtvedt 2014). Researchers are responsible for anticipating and minimizing such harm. *Debriefing* at the end of an experimental session allows study participants to be fully informed of the nature of the deception and to decide whether to allow their data to be included in the study. Moreover, experimenters may use the debriefing to ensure that no harm has occurred or to make sure assistance is provided to those showing psychological distress. When experiments move outside of the laboratory, such debriefing grows less necessary because participants may not know that they are involved at all in a study.

Field Experiments

While the laboratory is the most common venue for experimental research, driven by theory testing, some experiments also may occur elsewhere. **Field experiments** involve the manipulation of factors within natural settings. Piliavin, Rodin, and Piliavin (1969) examined the response of bystanders to a person in distress in a New York subway. They employed the best experimental practices by controlling as many variables as possible. They staged the scene approximately a minute after the train passed the first station, with male victims of similar age and dressed in similar casual clothes, behaving in a similar fashion dropping to the floor in the same location. They manipulated whether the white victim appeared sober (carrying a cane) or drunk (without a cane).[5] In addition to the victims, three observers trained by the experimenter (called confederates) were present, located at strategic viewpoints in the car, to record the responses of subway riders. The experimental team ran 103 trials of the study, each time noting the race, sex, and location of all people on the train, the number who assisted the victim, and how long it took before assistance arrived. They also tried to record comments made by passengers. Unlike the patterns observed in laboratory studies, help was more frequent, often involving more than one person. In nearly 95% of the "sober, with cane" trials, observers provided help, whereas only 50% came to the drunk victim's assistance. Typically, males provided the initial assistance (perhaps because female observers believed that they did not have the physical strength to help a victim to a seat).

In this field study with its high degree of **ecological validity**, where study findings apply to behavior in everyday, natural settings (Bryman 2016), diffusion of responsibility

did not unfold. As a consequence, Piliavin and colleagues (1969) developed a model focused on emotional arousal (e.g., fear, disgust, or sympathy regarding victims, which may be shaped by victims' personal responsibility for their plight) coupled with an evaluation of cost and rewards. The desire to reduce the arousal spurs whether a person will offer help, considering the costs and rewards of each action. Help is most likely when the costs of doing so are low (e.g., no threat of physical harm, ease of effort) and the rewards (e.g., praise from others) are high. Helping also emerges when the costs of not helping (e.g., self-blame) are high and the rewards from not helping (e.g., being able to get on with one's own activities) are low. (Conversely, people may rationalize not helping when there are high rewards to do so and low costs!) In effect, Piliavin and colleagues used their field experiment results to elaborate on conditions under which diffusion of responsibility is likely to emerge. In doing so, they augmented reasoning offered by Darley and Latané (1968).

In field experiments, study participants are often unaware that their behavior is under scrutiny. Thus, they are less likely to respond to demand characteristics or experimenter effects. And the ecological validity of such studies may provide the opportunity to generalize findings to populations beyond those immediately present in the study. Moreover, sometimes field experiments can involve social changes that occur because of decisions by an organization or a government. In such cases, researchers do not actively manipulate their independent variable, but may examine behavior before and after the change. For example, to determine whether a "no littering law" has an effect, investigators could measure the amount of trash at a main intersection before and after the law goes into effect. Of course, such naturally occurring field experiments may involve a number of extraneous variables beyond the researchers' control.

In summary, the experimental method is ideal when scholars have a strong theoretical argument for their predictions and desire to test them under controlled circumstances. There are benefits to moving beyond the laboratory, though trade-offs between control and ecological validity emerge. While experiments may provide the surest way to assess causal relationships, they are inappropriate for answering particular research questions, where knowledge may not have yet extended to the development of a strong theoretical argument or where there is a need to assess patterns within a large population, perhaps pertaining to particular substantive or attitudinal issues. For such questions, survey approaches may be appropriate.

Survey Approaches

College students are rarely far from their mobile phones, computers, tablets, and the like, using them to facilitate communication with both near and distant family members, friends, and acquaintances. With the rise in technology-mediated communication, greater opportunity for hurtful online behaviors arises. Often such behavior may lead to harmful consequences for the victims and at the same time may be observed by countless others on YouTube or TikTok. Brody and Vangelisti (2015) ask the following: When do bystanders intervene in instances of cyberbullying? Their work builds on previous research examining assistance provided to victims of (offline) bullying, which suggests that in most cases other people are present (O'Connell, Pepler, and Craig 1999) and that assistance from bystanders, either directly as a means to stop the bullying or indirectly with provision of emotional support, can attenuate the ill consequences of suffering such mistreatment (Matsunaga 2010). And, of course, they draw from theoretical work of Latané and Darley (1970). But to tackle their research question, they used a written survey.

Survey research involves the collection of data by asking people questions. The questions may be written and self-administered either in paper form or via an online platform. In other instances, questions may be asked orally by an interviewer. Typically, researchers use written questionnaires to collect information quickly and efficiently, often involving a large number of respondents, resulting in a body of quantitative data. Interviews, on the other hand, usually involve fewer study participants but offer rich qualitative data because the researcher may ask follow-up questions and probe participants' responses to gain insight on a topic or process under investigation. We first focus on written surveys and then turn to interviews.

Written Surveys

A key characteristic of written survey methodology pertains to how researchers secure a sample of respondents. **Random sampling** involves the inclusion of a "unit" of a population of interest by chance alone. In other words, imagine that you wanted to know something about people's responses to bullying incidents in middle schools in your community. Your population of interest would be middle-school–aged children in schools within the districts in your community. Fortunately, you would not have to survey the entire population! Rather, you could select at random every 10th student (the exact size of the sample depends upon the size of the population and other factors). Such random selection ensures replication of population characteristics (e.g., percentage female or nonwhite), thus creating a **representative sample**.

To thoroughly ensure that such population characteristics are represented in a sample, researchers will perform random selection within each important category of the population. This is called **stratified random sampling**. Sometimes, however, researchers rely on a **convenience sample**, which is selected largely for its availability but may also provide a starting point in the development of knowledge about a particular topic. For example, if you wanted to know how college students at a particular institution view climate change, you could sit out in front of the student center and ask a set of students who come by on a particular day.

Finally, sometimes researchers use **snowball sampling**, which relies on using referrals of initial participants to obtain additional participants, when it is difficult to locate respondents because the desired sample characteristic is uncommon. For example, Khanna (2004) examined factors affecting the nature of the preferred racial identification of white/Asian college students. No list indicates multiracial people who identify as white/Asian individuals, and so they are not easy to locate. Thus, Khanna placed classified advertisements about her study in local university newspapers and some magazines serving multiracial readers and posted a call for participants on multiracial websites and mailing lists. The rest of the participants were found through word of mouth or referrals. Snowball samples are nonrandom, similar to convenience samples.

Random sampling, whether stratified or not, is a means of ensuring the **external validity** of study results, defined as the generalizability of findings beyond the sample to the population of interest or to other settings. Results from convenience and snowball samples, in contrast, cannot be generalized to the population, but as with experimental research, the theoretical argument underlying the hypotheses being tested may provide a bridge to the larger population.

For a first look at how bystanders might respond to cyberbullying incidents, Brody and Vangelisti (2015) used a convenience sample of 265 undergraduate students who

completed an online survey.[6] The survey instrument is how investigators measure their independent and dependent variables. As in previous laboratory research, the researchers hypothesized about the effects of the presence of other bystanders and perceived (visual) anonymity on reducing the likelihood of providing help and the positive impact of relationship closeness between the observer and the victim on help provision. Thus, through their survey questions, they needed to measure the independent variables of presence of additional bystanders, visual anonymity, and relationship closeness. They also operationalized helping, their dependent variable, in different ways. Other questions allowed them to control for factors that may influence how participants respond as bystanders, such as their personal experiences with cyberbullying, the extent that they perceived the victim to be hurt by the cyberbullying incident, and gender.

The researchers asked study participants to recall an incident, occurring in the past 6 months, that involved the online bullying via Facebook of someone they knew. Based on the recalled episode, a series of questions captured details and measured the independent variables. Responses to the question of "Approximately how many other people witnessed the bullying episode?" provided a means to measure the number of additional bystanders. Several questions captured visual anonymity. Participants indicated agreement on a seven-point Likert scale (where 1 represented "strongly disagree" and 7 "strongly agree") whether they thought that others (victim, perpetrator, additional observers) "were aware of me," "knew I was there," "recognized my presence," and "could see that I was online." **Likert scaled responses** are a frequently used means to measure attitudes or perceptions represented by degree of agreement with a series of statements. Multiple Likert items may be combined or added together to create a **composite scale** or score for a particular underlying concept. In this case, the researchers created a scale of visual anonymity (to do so, they had to reverse-code the items before combining so that higher numbers corresponded to higher perceived visual anonymity). Multiple-item scales may have greater **reliability**, indicating the degree to which the measure of a concept appears to be stable. Brody and Vangelisti (2015) also created a scale from the multiple items indicating closeness to the victim—for example, "How close are you to this person?" and "How important is your relationship with this person?" Responses could range from 1 = "not at all" to 7 = "very much."

Reported bystander behavior was measured through various questionnaire items. The researchers modified an existing scale consisting of Likert agreement items (Salmivalli and Voeten 2004). For example, respondents indicated their agreement with items such as "I told the perpetrator to stop their behavior," which suggests active defending of the victim, and items like "I stayed outside the situation" to capture passive responses. Additionally, they included items that tapped into different types of support that people might lend, such as emotional or network-based, by connecting the victim to people who could provide help.

When using survey questions to measure concepts, researchers take care to phrase items in a precise way to solicit the most accurate representations of what they are trying to study. Typically, it is easier for respondents to answer short, straightforward questions that *avoid jargon* rather than long questions—especially those using words that may be unfamiliar. For example, the visual anonymity scale includes clearly and simply worded items that pertain to a wide array of possible ways in which bystanders might perceive each other. Providing the option "were aware that I might be virtually present at the same time" is far more cumbersome than simply offering "were aware of me." Similarly, a single global

item asking "How likely did others know of your presence?" might produce results different from the scale of *multiple, focused items*. In addition, *response categories must "match" the questions*. For example, the relationship closeness items use a "not at all" to "very much" scale, rather than an agreement scale, to ensure a match with how the questions are worded. Variations in phrasing or response scales may affect the meaning of the question to respondents or create confusion, potentially generating results that are misleading or not replicated in other studies.

Survey researchers also may be sensitive to the level of measurement of each item (which affects the extent of variation potentially captured and the nature of the statistical analyses that might unfold). **Nominal measures** refer to variables having distinct, often mutually exclusive categories associated with them. Typical examples are political party affiliation, college major, and religious affiliation. In Brody and Vangelisti's (2015) study, gender is considered a nominal measure. **Ordinal measures** suggest more or less of a particular variable, such as agreement with a statement measuring political liberalness or support for a particular policy. Likert scales, noted previously, represent ordinal measures. The actual "distance" between marking a 2 versus a 5 on such a scale may not signify standardized units that carry the same meaning, but a 2 indicates far less agreement than a 5, which surely implies greater agreement. In contrast, the actual distance between variables that represent **interval measures** is meaningful, relying on standard intervals between responses. For example, measures of "age" that use "years" constitute an interval measure, where a 40-year-old respondent is twice as old as a 20-year-old one.

In development of their survey study, Brody and Vangelisti (2015) grappled with the best way to measure the concepts represented in their theoretical hypotheses. Although all concepts were measured at the same time, their theoretical argument indicates what factors are likely to affect helping. Their results confirmed their hypotheses that a higher number of bystanders and increased visual anonymity decreased active defending of the victim (and increased passive behaviors), whereas closeness to the victim increased such behaviors (and decreased passive responses). Visual anonymity decreased and closeness to the victim increased emotional and network support as well, but number of bystanders influenced only network support. In the support analyses, researchers included the control variables (e.g., gender, own experiences with cyberbullying) to rule out other potential influences on why respondents lent support.

While written survey approaches compared to experimental strategies are generally a weaker means by which to assess causal relationships, researchers may employ statistical techniques to assess theorized patterns of causality. Sometimes, such as when surveys are employed in the study of social problems, researchers will collect data at multiple points in time, involving the same study participants. Such **longitudinal surveys** provide greater opportunities for researchers to assess causal relationships by meeting the criterion of "time order." For example, an investigator interested in the relationship between self-esteem and juvenile delinquency might collect data at the beginning of high school and then yearly until graduation. That would mean that the level of self-esteem during the freshman year could be used to predict juvenile delinquency during subsequent years.

Of course, a study of juvenile delinquency or even of helping behavior raises a general issue about surveys: the extent to which individuals reply truthfully to questions. Typically, self-reports of one's attitudes and behaviors are relatively honest assessments—to the extent allowed by memory—unless the questions pertain to highly personal information, especially that which may be illegal or embarrassing. Faulty recall (even to simple

questions like "How frequently did you access Instagram in the last week?") and the tendency to provide socially desirable responses plague survey research. Yet the strengths of such a strategy often outweigh such weaknesses, especially when researchers anchor their studies in previous research and carefully construct their survey instruments. Survey research is ideal for efficiently collecting a large amount of data that allow for the description of a population (assuming a random sample) and patterns regarding attitudes and behaviors within that population. The use of oral questions within a subset of a population may allow for probing of patterns revealed by large-scale surveys or the creation of knowledge about underlying processes.

Oral Interviews

This chapter opens with several actual situations in which bystanders failed to provide assistance to individuals who were victims of horrendous assaults. If you wanted to delve into bystanders' perceptions of why they did not assist and how those reasons might stem from their belief systems or previous experience, a strategy would be to employ in-depth **interviews** (e.g., Glesne 2015; Gubrium et al. 2012). Essentially conversations between two people, interviews involve a researcher asking a series of questions to a member of a group of interest. The questions may be structured so that all interviewees are asked the same open-ended questions, or semistructured such that while most questions are asked of all study participants, the researcher may ask additional questions to probe answers given to initial queries or to pursue insights arising in the course of the interview.

For example, Einolf (2011) drew on narrative interviews with 88 respondents from a larger study, Midlife in the United States (MIDUS), to examine how religion influences the likelihood that people will engage in helping acts. By closely reading the narratives, he identified six themes characterizing individuals' religious beliefs and values: (1) the importance of religion as an aspect of one's own identity, (2) increased commitment to religious life during adulthood, (3) perceptions of the relationship between religion and helping, (4) inspiration found in Jesus' example and sacrifice, (5) religious definitions of morality, and (6) belief that it is one's life mission to carry out God's mission. While these themes emerged from the body of narratives, each interviewee did not necessarily talk about every one. Einolf was curious as to whether talking about a particular aspect of religious life (represented in the various themes) would impact how frequently individuals provided help in the form of volunteering and contributing to religious and secular charities. Thus, he combined data coded from the narratives for each respondent with other data available from responses to MIDUS surveys on volunteering and donating. His correlational analyses indicated that the first four themes were positively related to helping behavior. He then returned to the narratives to characterize the most highly prosocial individuals in the sample. For those individuals, religious themes were very prominent. From his qualitative analyses, he showed how people use language to learn and internalize religious values and how individuals come to act on those values by providing help of various kinds.

Narratives of the sort that Einolf (2011) used typically involve general questions that allow the respondent to tell a story. The interviewer guides the story through additional prompts. In such cases, researchers are interested in how people create meaning in their lives. Of course, in-depth interviewing may pertain to a variety of research questions. For example, the method is well suited to addressing issues of identity (e.g., Fields 2014; Khanna 2011), stigma (e.g., Pfaffendorf 2019; Snow and Anderson 1987), and emotions (e.g., Lois 2003; Wingfield 2010). In such cases, the interviewer begins with a research question, which

helps to direct development of the set of open-ended questions that will be posed to study participants. Yet the interviewer is not constrained by the questions on the interview guide. Indeed, flexibility is a fundamental characteristic of this methodology, allowing interviewers to ask follow-up questions and probe responses further. Typically, interviews are audio recorded and later transcribed. The transcriptions provide the raw data for analysis.

Researchers may rely upon electronic tools (e.g., computer programs like MAXQDA, NVivo, or ATLAS.ti) to help organize information derived from the transcriptions. These tools help identify themes or words that frequently emerge in study participants' responses. Often, researchers will begin by "looking for" particular themes based on their research questions, but they also allow themes to emerge organically from the data (see Glesne 2015 regarding coding, organizing, and analyzing interview data). For example, Einolf (2011) may have presumed that something about Jesus' teachings would be relevant to helping behavior among Christian respondents, so he may have first looked for such references. In contrast, it may be that the commitment to religion over one's adult life emerged as a theme unanticipated by Einolf. Organization of themes provides the basis for the important activity of the qualitative researcher: interpretation of the meanings within interviews and across interviews. The goal is to achieve understanding about why people behave in a certain way (e.g., why they help, how they defend an identity) or why they hold a particular set of beliefs, values, or identities. In other words, in-depth interviewing allows researchers to get at processes underlying actions and thoughts.

While the flexibility in asking questions and the depth of analysis exceeds that possibility when using survey or experimental methodologies, interviews are typically limited to small samples (often achieved through snowball sampling). The intent is not to generalize the results from these samples but to provide insight that may contribute to building theories and expanding understanding of a phenomenon. Interviewers must be careful, like experimenters, not to influence the responses of their study participants. A key concern is how the position of the interviewer vis-à-vis the respondent may influence how forthright and honest responses may be.

The **positionality** of researchers collecting qualitative data refers to the investigators' reflection on their own placements in the social contexts and power or status structures relevant in the situation and how that placement may engender subjective viewpoints, which may affect data collection and interpretation (England 1994). For example, a white researcher interviewing a Black study participant about racial identity might be conscious of being more privileged compared to the person being interviewed; in turn, the interviewee may view the white researcher with skepticism and distrust. Interviewers should recognize the potential impact of their positionality both during and after the interview. Establishing trust with the study participant while maintaining a degree of detachment so as not to cue the respondent in some way is the hallmark of a thoughtful interviewer. Investigators aware of their own subjectivities are likely to be careful to recognize how they may enter into interpretations of the data as well. And, as with written surveys, interviewees may be more inclined to provide responses that put them in a favorable light. In studying behavior, one means of avoiding the social desirability that may plague responses to written and oral questions may be to directly observe behavior.

Observation

When we were young sociologists, one of the most startling things that we realized as we were sitting in the lobby of the hotel hosting our annual professional meetings and

observing the dynamics of people around us was that we were not alone in the hazard of our trade: Others were observing us! We all "people watch" from time to time—in parks, airport waiting areas, bars, malls, and the like. These casual observations are usually just a way to pass the time and are rarely informed by preconceived notions of what might unfold. Sometimes, however, what people observe through casual observations might provide the basis for further investigation.

A key distinction between observational and survey research is that the former focuses on actual behavior, whereas the latter relies on self-reported behavior, which may be influenced by faulty recall or social desirability. Of course, not all behaviors are observable (e.g., criminal activities, sexual acts). As a consequence, observational approaches are limited to what can be seen, typically in public places or places to which an investigator has been given special access (e.g., a work organization, a unique religious group, high school classrooms).

The **observation strategies** of social psychologists involve gathering data directly by observing behaviors of interest (e.g., lending assistance), coding them into categories (e.g., direct help by administering to someone in need, indirect help by calling others to help), and classifying characteristics of the surrounding situation (e.g., number of people present, gender of victim and helper, time of day, location). They may be highly structured or less so, with noninvolvement or involvement by the researcher in the group under observation. Structured observations, described first in the next paragraphs, typically involve **nonparticipant** observers who are not engaging in the social situation and behaviors under scrutiny. The usual goal of such endeavors is to produce quantitative data regarding the behaviors observed. In contrast, less structured observations may be more likely when a researcher acts as a **participant**, engaged in the activities of the group being observed, often with the intent to develop qualitative data.

Structured, Nonparticipant Observation

Jiobu and Knowles (1974) extend Latané and Darley's (1970) interest in the impact of norms on helping, controlling for situational factors. In their study, rather than focusing on providing assistance to a victim in need, they focus on the prosocial activity of making charitable contributions as a form of helping. They argue that social expectations regarding helping are likely to be stronger around nearness to holidays characterized by giving (such as Christmas) than at other times of the year. To empirically examine this prediction about the strength of norms on people's contributions to charity, they devised a **structured observation** study, using a set of predetermined categories to code the behavior observed in a natural setting. On the four Wednesdays between Thanksgiving and Christmas from 9:30 a.m. to 9:00 p.m., they witnessed and coded people's willingness to "give" to the Salvation Army kettle in the central business district of a small midwestern city (population about 88,000).[7]

Two similarly aged and dressed Salvation Army workers took turns sitting in the booth where the kettle was located, steadily ringing a handbell but not engaging in any interaction with people passing by. The kettle sat on the sidewalk, near a department store, with plenty of room for pedestrian traffic. This setup ensured that the researchers controlled for a number of situational factors that could affect contributions: day of week, nature of solicitations, ease of accessing the kettle, and nature of stores near the kettle. Thus, they could focus on the increased influence of the norm about helping closer to Christmas (their independent variable) on giving behavior (their dependent variable) measured by

two unobtrusive observers. One observer recorded donor characteristics (e.g., gender, age), and another coded the number of people passing by. With this information, the researchers calculated (1) the percentage of pedestrians who gave (conceptualized as the "extensity of giving") and (2) the average contribution on each day ("intensity of giving").

Results provided mixed support for their hypothesis. While the average number of people giving dropped off as Christmas approached (perhaps because some of the people had already given), the average contribution was the highest on the last Wednesday before Christmas. In other analyses, the researchers noted that situational factors, such as the wind chill index and volume of pedestrians present, also affected the percentage of donors, with numbers decreasing with colder weather and more people present. The latter finding is consistent with previous evidence for diffusion of responsibility. Thus, results from this observational study dovetail with those from laboratory and survey research.

Nonparticipant observational studies focused on securing quantitative data typically employ structured coding schemes, with predetermined categories, to track behaviors of interest. Such coding schemes specify the behaviors to be observed and include instructions about which behaviors are allocated to particular categories (Bryman 2016). They must have a clear focus and be easy to use, or observers may get overwhelmed with trying to keep up with all that is going on in a situation! Yet even with a structured coding scheme, observers ultimately use their judgment to determine what fits and does not fit into a category.

Observers make decisions about what to include and what to exclude in terms of their observations. These decisions extend beyond the actual observation process to include considerations of when and where observations will take place and who will be included. Because these are, to some extent, judgment calls, researchers must be aware of how their own biases may influence them and try to minimize their effects. Often, observers are trained to follow the outlined procedures and to understand what fits within each category of a coding scheme. Observers may be assessed to make sure that their own coding is consistent over time, ensuring **intraobserver reliability**. Another strategy for minimizing potential biases emerging in structured observations is to include two or more observers, focused on the same actions, and to assess consistency across observers. Such consistency captures **interobserver reliability**. Assessment of such reliability presumes use of a structured coding scheme, which may, at least initially, be absent in other forms of observational approaches.

Less Structured, Participant Observation

When the researchers are participants in the group under observation, they engage, to various degrees, in the same activities of study participants and communicate with them informally and sometimes formally through in-depth interviews. Participant observers may immerse themselves in an organization or community for a long period of time, getting to know the people and flow of their interactions. Such placement may be characterized as **field research**. Rather than the structured observation of the donation study, these researchers rely upon less structured observations, often detailed in **field notes** chronicling the activities of the people observed, their conversations, responses to informal questions, and the overall processes unfolding. Investigators may attempt to reveal the meanings that the participants attribute to their behaviors. An **ethnography** is the written account organizing the observations and interpretations.

For example, Adler and Adler (2004) draw upon 8 years of participant observation and interviews in their examination of workers in five luxury resorts in Hawaii. In their

book, *Paradise Laborers*, they seek to understand how the global economy impacts the lives of people whose occupation is to provide services to vacationers (a form of help, you might say, though not in response to an emergency situation!). They observed characteristics of the workforce and, with permission of resort management, began with informal conversations with the workers about what they do, their views on their jobs and future opportunities, and factors affecting whether they remain in resort employment. A small number of workers functioned as **informants**, individuals occupying particular positions to provide background information and help the researchers to both organize their work and understand their findings. Adler and Adler also completed extensive interviews with 90 workers. Four ethnicity- and gender-diverse groups, each with different approaches to the services that they render as well as their futures, emerge: new immigrants holding unskilled positions with little pay; Hawaiian locals, often young but with few occupational alternatives in their island homeland; college-educated middle-class (mostly male) managers, whose stints at the resorts may be stepping-stones; and young, largely white adults from the mainland seeking an adventure-filled interlude in their lives. Through their observations and both short and long interviews, Adler and Adler reveal the different occupational cultures, worker lifestyles, and mobility patterns of the groups.

Such in-depth study often begins with an orienting question. Adler and Adler (2004) initially believed that the nature of the ownership of the resorts—by U.S. or international corporations—might affect the dynamics. Yet once their study was under way, they came to see that such a contrast was far less impactful than the nature of the employee groups themselves. Thus, especially in comparison to quantitatively focused research methodologies, ethnographies allow for greater flexibility in the data collection phase of a project. Sifting through relevant information to identify patterns of behavior remains one of the biggest challenges. And, as with structured observations and oral interviews, the positionality and biases of the researchers in shaping the behaviors they observe, the people with whom they talk, and so forth are a concern.

Concerns About Observational Strategies

As with demand characteristics in experimental research, investigators using observational strategies must be aware of how their presence, if known, may affect the behavior of the people and the interactions that they are studying. This "reactivity" of those observed may then represent atypical behaviors, which raise concerns about whether researchers are investigating what they had conceptualized. Such reactivity may emerge early in a study, but often study participants become acclimated to the presence of an observer and busy with other activities, and thus their behavior grows typical (McCall 1984). Nonetheless, researchers using observation strategies must be sensitive to the potential for reactivity.

The benefits of observation of actual behaviors must be balanced with the drawbacks of the influence of biases of researchers and the reactivity of those being observed. Also, as with experimental research, often the observation involves a narrow field of study participants, which leads to concerns about the generalizability of the results. Again, researchers must consider why observation is the most effective means of addressing the research question of interest. Through observation, they may reveal patterns of behaviors. While structured observations rarely can reveal the intentions underlying behaviors, less structured ethnographic work provides one means of delving into such intentions when observation is also accompanied by other methods, as evidenced in *Paradise Laborers* (Adler and Adler 2004). Indeed, qualitative researchers often may draw upon multiple sources

of data so that they may cross-check their findings. Such a process is called **triangulation**. Adler and Adler (2004) also had access to work policies and job descriptions and procedures developed by resort management. Review of such policies, descriptions, and procedures, which may influence social dynamics and workers' mobility, falls under the rubric of archival research.

Archival Methods

Unlike experimental, survey, or observation research in which study participants may respond, consciously or unconsciously, to the process of being involved in the research endeavor, archived materials as a source of data are not "reactive" to the fact that they are being examined. **Archival methods** refer to various strategies of analyzing (or reanalyzing) existing information, including documents, texts, pictures or videos, songs, websites, or statistics, collected or produced by others and often for purposes other than research. As with other methodologies, the goal is to reveal patterns of behavior currently or historically. To achieve that end, investigators rely on their research questions to direct them to certain forms of information and develop coding schemes (much as in observation research) to capture representations of variables of interest. Such endeavors are unobtrusive because they ensure that the methods of studying social behavior will not affect the behavior under scrutiny (Babbie 2021).

Recall the description of what Latané and Darley (1970) might have done in the wake of the news stories about the Kitty Genovese murder: look for other recorded incidents of whether bystanders provided help to victims. One means of doing so would have been to search newspapers for published accounts of assaults and murders, categorizing the content of each by aspects of the situation as reported. **Content analysis** involves a detailed review of the documents, physical or digital, written or visual. You may have heard recent references to analysis of "Big Data," which may involve extensive scaping of documents of all sorts to garner data relevant to a research question. In effect, researchers are using different types of tools both to collect and to analyze emergent patterns (see e.g., Goldenstein and Poschmann 2019). Whether researchers use innovative text-mining tools or read and code the text "by hand," they follow similar procedures.

To the extent that researchers attempt to quantify content of documents (broadly defined) in a systematic and replicable manner by coding information into predefined categories, their goal is quantitatively oriented. Their interpretations, however, may be considered qualitative. Those categories depend upon the research question and relevant theoretical ideas about the social behavior under examination as well as the operational definitions of central variables. They may identify the actors present, various types of behaviors, references to particular themes, or even the use of specific words (or images). Categories should be mutually exclusive (i.e., no overlaps between categories representing different concepts or variables) and exhaustive (i.e., all possible categories represented in the coding scheme; Bryman 2016).

To perform a content analysis, researchers must select a sample of the documents to be examined. (It would be impossible to review the contents of every newspaper published in the United States, across all the years of publication!) The sampling frame might specify which media sources, the time period covered, or the unit of analysis (e.g., individual behavior or rates of behavior aggregated over a larger unit such as a county). Once the sample has been determined, coders trained in the designed categorization scheme review the information provided in the selected documents to allow quantification of the behavior at

issue. And, as in the case of structured observation research, employing multiple coders and assessing how consistently they use the coding scheme ensures intercoder reliability.

Young, Miles, and Alhabash (2018) performed a content analysis to determine how bystanders respond to cyberbullying on a social media site. Guided by research questions pertaining to the strategies (virtual) bystanders use in the context of cyberbullying, they identified an appropriate website (ask.fm, where the content is public), then searched for comments illustrating aggression and responses. In effect, they captured a "snapshot" of instances of cyberbullying available at one point in time, indicating dyadic interactions between a victim of bullying (i.e., the "profile owner") and a perpetrator for a longer period of time. For each dyad, they scrutinized bystanders' remarks. The corpus of their data thus consisted of 352 initial aggressive posts and their responses, which constituted 641 question and answer sets.

A pilot study was previously conducted to help define coding categories for aggressive posts (negative or derogatory comments about a profile owner's popularity, sexual orientation, attractiveness, weight, or the like), victim responses, and bystander responses. Initial codes for bystander or "second person" posts indicated whether they were neutral (i.e., no reference to the aggressive attack), supportive of the profile owner, or aggressive. More specific codes capturing bystanders' strategies of helping included admonishing the victim to ignore the aggressive post, offering care and support to the victim, attacking the perpetrator with threatening questions or statements, and calling out the anonymity of the aggressive poster. The researchers trained two coders and assessed the reliability of their coding by having them both code posts from 20 dyads randomly selected from the sample. Once they were assured of strong intercoder reliability, each then coded half of the remaining posts independently.

In these instances of cyberbullying, the majority (86.5%) of bystanders posted anonymously. Results further showed that nearly half of the bystander responses were neutral, and of the remainder, about half were supportive of the profile owner and half the aggressor. Of the (144) supportive posts, 20.8% stressed their care for the victim, 19.4% provided indirect support by attacking the original perpetrator, and another 17.4% stressed ignoring the original aggressive post. The 182 aggressive responses from bystanders, however, seemed to constitute additional bullying as many of these posts (26.9%) maligned the social status of the victim and 20.9% suggested suicide. Yet other findings suggest what is necessary to cultivate positive bystander support. For instance, when profile owners launched counterattacks, they were more likely to receive bystander support. Moreover, the first bystander to respond often set the tone for subsequent posts. For example, if the first bystander response was supportive, others also added supportive comments. The researchers highlight this finding as a potential means to stall online aggression and to cultivate online civility. Young et al. (2018) clearly illustrate how content analysis of archival documents (i.e., social media posts) can shed light on helping behavior.

More generally, content analysis affords the researcher the opportunity to study an issue over time in an economical fashion. The method is well suited to identifying descriptive patterns and revealing relevant factors that may influence a behavior, though it is less suitable for explaining why a pattern emerges. Unlike other methodologies, it allows readily for the correction of errors. When manipulations in an experiment fail or when survey questions are poorly phrased, the researcher will need to relaunch the study with appropriate corrections. But with content analyses, researchers can go back and recode their documents. The completeness of coding schemes and the clarity of coding instructions also ensure the feasibility of replication and follow-up studies.

Assuming that such an archival method is appropriate for the research question, investigators do, however, need to be aware of the extent to which the documents they have sampled are complete and created by individuals who themselves were not biased in their recordkeeping. For example, the content of newspaper articles or the postings attracted by some social media platforms may be colored by the political leanings of the publisher or platform owner. Or, in the case of historical documents, there may be gaps in recording owing to degradation of materials over time or even destruction of some periods of data due to fire or flood (really!). And the coding scheme created by a researcher relies on interpretation and may also introduce biases owing to what the investigator chooses to code or not code. Thus, though content analysis focuses on nonreactive records, the method is not without potential contamination by biases of various sorts.

Another form of archival methods involves the compilation and **analysis of existing statistics**, a decidedly quantitative undertaking. Government agencies, corporations, religious organizations, nonprofit groups, and the like often collect routine data about a variety of things (e.g., characteristics of populations, rates of crime, number of dwellings with more than one bathroom, levels of contributions, disease rates). Researchers relying on existing statistics often bring together different sources to answer distinct questions. Those questions rarely pertain to specific individual behaviors but may capture what individuals, at an aggregated level, may do.

For example, Whitehead (2014) looks at how patterns of volunteering with charitable organizations depend on population density. He argues that people who live in cities, with higher population density (the number of people living in a designated unit of area like a square mile), often experience high sensory overload and thus spend less time on each activity and would therefore be less likely to volunteer than residents of more rural areas. To empirically examine this idea, he brought together statistics from the same time period regarding the number of residents per square mile in each state in 2010 (from the U.S. Census Bureau) and the volunteer rate by state (compiled by the Corporation for National and Community Service for 2011). His statistical analysis reveals a negative correlation between state-level population density and measures of volunteerism, confirming his expectation (even he cannot account for whether population density *caused* the levels of different forms of helping). The study highlights the importance of recognizing different population-level antecedents to types of helping behavior (at the state level).

With use of existing statistics, researchers must be wary of whether they accurately represent what they claim to report (Babbie 2021). While Whitehead's (2014) population density measure may have been accurate, the nature of the volunteer rate may depend upon what the Corporation for National and Community Service "counted" as relevant forms of volunteerism and whether they relied upon data from nonprofit groups or perhaps some sort of survey that they conducted. And, although Whitehead's study involves a snapshot in time, research using existing statistics needs to consider if and how recordkeeping systems may have changed over time. Nonetheless, given the availability of statistics from a wide variety of sources, this archival method may be useful in answering particular questions about social behaviors at the aggregated level, especially over time or even cross-culturally (assuming similar measures can be obtained from sources in different countries). Researchers, however, must recognize the potential for using their results in a manner that constitutes an **ecological fallacy**, or inferences about individual behaviors from aggregated data. For example, while Whitehead can make claims about patterns of population density and volunteering rates in states, he cannot say that individuals who live

in high-rise apartment buildings are less likely to engage in volunteer activities than those in single-family dwellings. To address that issue, researchers might survey people who live in different types of residences about their volunteer activities.

Archival methods, like the previously discussed experimental, survey, and observation strategies, have advantages (e.g., nonreactivity of the data) and disadvantages (e.g., potential for the ecological fallacy). As stressed earlier in the chapter, which method researchers choose depends upon the nature of their knowledge of the topic and the research question at issue. And regardless of the methodology, all researchers must be sensitive to how their approach ensures protections for the rights and well-being of their study participants. More generally, no researchers should go into the field without considering the ethics of their approach.

ETHICAL ISSUES

In most of the helping studies used to illustrate different methodological approaches to studying social psychological processes, investigators had either active contact with their study participants by inviting them into a laboratory or to complete a questionnaire, or passive contact through observing their behavior in public places. The contact between researchers and their study participants can be described as a form of interaction. And, importantly, as initiators of the interaction, scholars, throughout all stages of their research endeavor—from design, to data collection, to data handling and storage, to dissemination of results through publications—must be cognizant of treating their study participants in an ethical, respectful manner, demonstrating regard for their rights and well-being. Professional associations prescribe ethical procedures in research, as do various governmental organizations charged with enforcing federal regulations about the ethical conduct of research.[8] Here, we describe key features of research that ensure ethical treatment of study participants (see Babbie 2021; Bryman 2016; Hegtvedt 2014).

Regardless of the methodology chosen to address a research question, researchers must keep in mind that their study participants are not merely "research material" or objects to be examined. Such objectification may allow investigators to grow insensitive to the potential harms that may be produced, even unintentionally, from their studies. Instead, researchers should be mindful that study participants are human beings who think, feel, and behave just as they do. Additionally, to the extent that researchers have something that study participants desire (e.g., money, access to a particular treatment or grade), they may be in a position to coerce or exploit participants. As with objectification, the potential to exploit may lead to callous treatment of study participants. With these considerations in mind, investigators should anticipate and minimize the variety of **potential harms** that may befall those who join in their studies.

Harm may be of various types and vary in both intensity (how bad is it?) and likelihood of occurring (is it rare or highly likely?). Typically, rare in social psychological research is physical harm, such as injury. More common are psychological harms, often in the form of some sort of embarrassment, anxiety, distress, or loss of self-esteem or confidence. For example, failing to help a victim in need, even if others are present, may challenge individuals' views of themselves as good people, leading to feelings of guilt for not providing assistance or loss of self-esteem. Social harms may also arise when studies ask participants to reveal private facts about themselves that could potentially compromise their reputation or even threaten their freedom. For example, revelation of responses to surveys of

self-reported criminal activities might lead to inquiries by law enforcement officials. It is incumbent upon researchers to design study procedures that minimize the intensity and likelihood of anticipated harms.

Several common procedures built in during the design stage of a study and implemented during data collection help to lower risks of harms owing to participation. First, researchers should ensure that participation is **voluntary**, which includes both agreement to be involved in the study and the right to stop participation at any time during the study. No one should feel obligated to be in a study for fear of repercussions from the researcher or the institution conducting the study. For example, if students enrolled in introductory psychology, sociology, or social psychology classes are required to be part of a "subject pool" and participate in studies as part of their grade, voluntariness is jeopardized unless students are given a choice of whether to be part of a subject pool and a choice of studies in which to participate. (Departments that run subject pools are required to give students alternative assignments that require a similar amount of time as, say, a particular experiment that they may choose instead.)

Second, participants must offer **informed consent**, which is the process by which a researcher conveys information about the study (who is conducting it and for what purpose, what participation involves, how long it takes, potential harms and benefits, means to ensure anonymity or confidentiality, any compensation, etc.; Dunn and Chadwick 2012). Typically, informed consent occurs before the study begins, but in some cases, informed consent continues to be sought in the middle or at the end of the study. And while elements of the information conveyed are consistent across methodologies involving active contact between researcher and participants, the nature of the process varies depending upon the study population, the substantive content of the data being gathered, and the means of data collection. For example, in the typical laboratory experiment, the researcher will offer a written document describing the previously noted information about the study and ask the participant to sign the document to signal consent. In many surveys, participants read a description of the study, and the actual completion of the questionnaire (online or on paper) signals their consent. For collection of interview data, especially with what might be considered a vulnerable population (e.g., gang members, HIV-positive individuals), study descriptions may be offered orally and oral consent accepted.

Third, in any type of consent procedure, researchers must make efforts to protect the **privacy** of their study participants as a means of warding off psychological or social harms. One way to ensure privacy is to indicate that the data collected will be **anonymous**—no names will be associated in any way with the data provided. Assurances of anonymity are common in much survey and structured observation research in public settings. Oftentimes, however, researchers will know the names of their study participants (indeed, they may have signed a consent document!). And knowing the names is important for collecting follow-up data or triangulating with other data sources. In such cases, investigators may create identification codes so that they can link sources of data (e.g., surveys of high school students and archival data on their academic performance). When investigators can link names to the data collected, it is incumbent upon them to hold these records **confidentially**. Doing so focuses on handling of data in a manner that avoids breaches in confidentiality, which is a major threat for psychological and especially social harms (Citro, Ilgen, and Marrett 2003; Sieber 1992). In addition to creating identification codes, researchers might provide separate storage for and limited access to master files that link codes with names, use password-protected computers and computer files, and only

present data in aggregate form so that the responses of no one person are singled out. For qualitative endeavors using quotes from interviews, pseudonyms help protect respondents' privacy.

Although such precautions guard the well-being of study participants, additional concerns arise in experimental research that involves deception. Recall that in the Darley and Latané (1968) study, participants heard voices, presumably of other students involved in the study, and one of those ostensibly suffered a seizure in the course of the experimental session. Both of these study procedures were designed intentionally to create variation in group size and the need for help, respectively. And yet, the prerecorded voices were of **confederates**, research assistants instructed to act in a particular way by the experimenter and unbeknownst to the study participants, including sounding as if a seizure was in process. Darley and Latané deceived their study participants by providing false information, with the intention of misleading them to believe something that was not true (no one was suffering a seizure!). Many scholars have debated the potential harms to study participants and researchers, as well as professions that may be wrought by the use of deception (e.g., Cook and Yamagishi 2008; Hertwig and Ortmann 2008; Korn 1997; see Hegtvedt 2014 for a summary).

In effect, deception deprives study participants of their right to determine their willingness to participate in full knowledge of what to expect in the situation because informed consent procedures omit initial mention of deceptive practices. Such full disclosure prior to the participation would jeopardize the validity of the findings regarding the impact of independent variables created through deception. In such cases, further informed consent is sought at the end of the experiment, after the study participant has been fully informed of the intentions of the study and debriefed about the deception. At that point, assessment of whether the participant feels any distress or embarrassment at being "duped" is necessary, with researchers providing remedies for such distress and allowing participants to drop out of the study as well. Beyond the individual level, there is concern that the use of deception may stimulate suspicion about and negative inclinations toward the research endeavor, which may spill over to nondeceptive studies and thereby threaten the viability of social research in general. Little evidence, however, confirms this potential threat to the credibility of social research.

Typically, the benefits of deception must be weighed in terms of harms to individual study participants or society more generally. Deception should be used only when no alternative procedures are feasible to ensure a study's validity or address a particular issue. In addition, study designs must minimize risk of harm to study participants. Many deceptions are relatively harmless (e.g., creating false expectations about procedures), whereas others may involve false feedback that results in damage to individuals' self-evaluations. In the latter type of situation, researchers must provide a means of restoring a positive sense of self.

Beyond the ethical obligations that researchers have to their study participants, they must consider concerns arising after data are collected, in the process of analyzing and publishing their findings. For both of these steps in the research endeavor, investigators must signal the extent to which they are being honest about how they treated the data, interpreted the data, and conducted the statistical analyses that reveal the results that appear in published articles. Dishonesty may take the form of excluding particular cases because they do not "fit" the pattern expected, adding or changing participants' responses, or failing to include potential confounding variables in their analysis (which

may reduce the observed impact of focal factors). The bottom line is this: Researchers should not fabricate their data or their results!

Additionally, they should report negative findings—those that do not support the hypothesized relationships—and reveal shortcomings and limitations of their own work. Within psychology, debate continues about the extent to which findings from studies can be replicated (e.g., Maxwell, Lau, and Howard 2015; Simmons, Nelson, and Simonsohn 2011). When new studies fail to produce patterns of findings consistent with existing published work, the veracity of the findings and the soundness of the data collection procedures and analysis may be called into question. Such challenges may compromise scientific knowledge and faith in the community that produces it. Thus, it is important that scholars clearly report their data collection practices and procedures, as well as their statistical techniques, and even share their data so that others may attempt to reproduce or replicate the findings, thereby strengthening confidence in existing scientific knowledge. The movement to preregister studies (Logg and Dorison 2021) reinforces the importance of replicability and reproducibility in scientific endeavors. Conforming to the ethical standards of the scientific community signals researchers' integrity and ensures the public's confidence in the scientific endeavor.

SEGUE: FROM THE MEANS OF COLLECTING DATA TO THE PATTERNS REVEALED

As illustrated in Chapter 1, social psychologists pursue a wide variety of topics related to the understanding of social behavior. Their pursuit, however, differs from that of the casual observer of interaction dynamics insofar as they attempt to bring systematic evidence to bear upon the issues that they study. Using research in the domain of helping behavior, we have provided an overview of the logic of scientific inquiry and the means by which investigators collect data. The use of any one strategy—experiment, survey, interview, structured/less structured observation, or archival materials—depends upon the existing level of knowledge on a particular topic, the purpose of the research itself (i.e., to explore patterns in a particular place, at a particular time, or to test hypotheses derived from deductive theory), and the particular question under scrutiny. Strengths and weaknesses of each method shape which method is chosen for which research question. There is no "one size fits all" in social psychological research.

Subsequent chapters of this book present theories, supported by empirical research, that provide a basis for understanding how individuals influence and are influenced by their social worlds. Throughout, you will encounter studies that further illustrate the methodological tools of social psychologists. To move beyond commonsense understandings, feel free to employ the refrain "show me the evidence." Our intent is to present such evidence.

END NOTES

1. The headline for the original *New York Times* story read, "37 Who Saw Murder Didn't Call the Police" (Martin Gansberg, March 27, 1964). Yet further inquiries revealed inconsistencies: Few may have seen the attack, though many may have heard it; two people claim to have called the police in the pre-911 era, but no record exists; and one woman went to Kitty's aid as she died. Stephanie Merry raised such issues in "How Everyone

Got the Infamous Kitty Genovese Case All Wrong" (*Washington Post*, June 29, 2016). The 2015 documentary, *The Witness*, directed by James B. Solomon for Five More Minutes Productions, illustrates William Genovese's attempt to set the record straight about his sister's murder and raises questions about the veracity of the original reporting. Nonetheless, the original reporting spurred social psychological research.

2. The 1988 film *The Accused*, starring Jodie Foster and directed by Jonathan Kaplan for Paramount Pictures, depicts this incident in New Bedford, Massachusetts.
3. These are fictitious incidents.
4. The researchers also created variations in the composition of the three-person group, with the other present bystander either a male, a female, or a male who said he was premed.
5. They also included conditions in which the victim was Black. Plus, they were prepared to have one of the study team members act as a model, providing help if no one else did after a particular interval. The models, however, rarely had to act.
6. Brody and Vangelisti (2015) report two studies. Here, we describe Study 1. Their second study combines elements of experimental design by creating vignettes in which the three central independent variables of Study 1 (few/many bystanders, high/low visual anonymity, close/acquaintance relationship with victim) are manipulated to create eight versions of an online bullying incident. Each of 379 respondents in a convenience sample read one scenario and completed measures of the same dependent variables as in Study 1.
7. Originating in 1891, the Salvation Army kettle, bright red in color, was a means to collect small contributions to provide holiday meals to individuals in need. Today, contributions during the holiday season help to allow the Salvation Army to continue its assistance to people in need through the entire year.
8. Codes of ethics are available from the American Psychological Association (http://www.apa.org/ethics/code/index.aspx) and the American Sociological Association (https://www.asanet.org/about/ethics/?hilite=code+ethics). The U.S. Office for Human Research Protections (https://www.hhs.gov/ohrp) is responsible for overseeing the operation of university-based Institutional Review Boards, which implement procedures for the review of research involving human study participants, and for ensuring compliance with the *Code of Federal Regulations*, Title 45, Part 46.