

# General Principles

**H**aving some specific studies to refer to will help to clarify the discussion that follows. Described next, therefore, are two examples of research in developmental psychology. Both studies have been simplified somewhat to make the points drawn from them easier to follow.

Brownell, Svetlova, and Nichols (2009) were interested in very early forms of sharing behavior in young children. The participants for their study were in fact quite young: one group of 18-month-olds and a second group of 25-month-olds. Each child in the study first learned a simple task: pulling a lever in order to receive a snack. A series of trials followed on which the child had to choose between two levers. Pulling one lever resulted in a snack only for the child; pulling the other lever resulted in a snack for both the child and an adult experimenter. Choice of the second lever served as the measure of sharing. Note that the child received a snack whatever the choice; thus there was no cost involved in the decision to share.

The study included one other aspect. On half the trials, the adult recipient merely sat silently; on the other half, she verbalized her desire for the snack (e.g., “I like crackers. I want a cracker.”) prior to the child’s choice of a lever.

Table 2.1 shows the results. It can be seen that older children shared more than younger

children—only, however, when the adult verbalized. When the adult remained silent the two groups did not differ. The results can also be stated in a different way: Sharing was more likely when the adult verbalized—only, however, among the older children.

For the second example, we move to a different phase of the life cycle. The goal of a study by Kliegel and colleagues (Kliegel, Martin, McDaniel, & Phillips, 2007) was to compare planning behavior in young adults (mean age = 25.6) and older adults (mean age = 70.9). The task was to plan a series of errands (e.g., pay an electric bill, withdraw money from the bank, visit a friend in the hospital) so that they could be carried out in the most efficient possible way. Optimal performance required taking into account the locations of the various goals, figuring out the best sequence of actions (for example, paying the bill only after a trip to the bank to get the necessary money), and ignoring various irrelevant pieces of information that were provided (for example, various unneeded locations, the reason the friend was in the hospital).

This study also had one further aspect. For half the participants, the task was as just described: to carry out a series of familiar errands, working from a map of a familiar-looking town. For the other half, the task demands remained the same, but the context in which the errands had to be

Table 2.1

*Proportion of Trials in Which Children Shared in the Brownell et al. Study*

	Adult silent	Adult verbalized	Total
18-month-olds	.55	.50	.53
25-month-olds	.46	.66	.56
Total	.51	.58	

Note. Adapted from "To Share or Not to Share: When Do Toddlers Respond to Another's Needs?", 2007, by C. A. Brownell, M. Svetlova, and M. Nichols, 2009, *Infancy*, 14, 117–130.

Table 2.2

*Errand Planning Scores for Young and Old Adults in the Kliegel et al. Study*

	Familiar setting	Novel setting	Total
Young adults	6.98	8.51	7.75
Old adults	7.02	5.90	6.46
Total	7.00	7.21	

Note. The point totals are based on a scoring system that awarded points for the number of goals achieved and the avoidance of errors in achieving the goals. Higher scores indicate better performance. Adapted from "Adult Differences in Errand Planning: The Role of Task Familiarity and Cognitive Resources," by M. Kliegel, M. Martin, M. A. McDaniel, and L. H. Phillips, 2007, *Experimental Aging Research*, 33, 145–161.

completed was a novel and unfamiliar one. Thus participants in this group had to pay their taxes on Planet A, withdraw gold on Planet B, visit a politician on Planet C, and so forth Table 2.2 shows the results. As the rightmost column indicates, the young adults outperformed the older adults. All of the difference, however, came from performance in the novel setting. In the familiar setting the two groups did not differ.

## Variables

I begin the discussion of general principles with some terminology. Research in psychology involves variables and the relations that hold among variables. The variables are of two sorts: dependent and independent. **Dependent variables** are outcome variables—those measures whose values constitute the results of a study. In the first example, the dependent variable was the proportion of trials on which the child shared; in the second example, the dependent variable was the number of points earned in carrying out the errands. Such variables are dependent in the sense that variation in them follows from or depends on other factors. A central job for the researcher is to determine what these other factors are. They are variable necessarily: If there were no possibility of variation in the dependent measure, there would be no point in doing the study.

The dependent variable is something that the researcher measures but does not directly control. **Independent variables**, in contrast, are variables that are under the control of the researcher. The object of the study is to determine whether the particular independent variables chosen do in fact relate to variations in the dependent variable. The independent variables in the Brownell et al. (2009) study were the age of the child and the presence or absence of verbalization, whereas those in the Kliegel et al. (2007) study were the age of the participant and the context in which the errands had to be carried out. Such variables are independent in the sense that their values are decided on in advance rather than following as results of the study. The "variable" part is again necessary: If there were no variation in the independent variable, there would be no possibility of determining whether that factor has an effect. Variation and comparison are intrinsic parts of all research.

The description of research as divisible into independent and dependent variables is valid for many but not for all studies. Suppose, for example that you wish to know whether there is a relation between a child's IQ and how well that

child does in school. You might test a sample of grade-school children and collect two measures: performance on an IQ test and grades in school. Your interest would be in whether variations in one measure relate to variations in the other; for example, do children with high IQs tend to do well in school? A study like this does not have an independent variable whose values are under the experimenter's control; rather, IQ, grades, and the relation between them are all outcome variables in the study. "Correlational" research of this sort is discussed at length later. The point for now is simply that not all studies fit the independent variable–dependent variable mold.

The example studies can serve to illustrate a further point about independent variables. The contrasts that define an independent variable can be created in two ways. One way is through an experimental manipulation that literally creates the variable. This is what Brownell et al. (2009) did with their manipulation of the verbalization factor and what Kliegel et al. (2007) did with their construction of the familiar and novel contexts. This was not the approach, however, for the other independent variable in both studies: chronological age. Clearly, investigators cannot create an age contrast in the same way that they can create a familiar versus novel contrast. In the case of a variable like age, the control occurs not through manipulation but through *selection*: choosing people for study who are at the desired levels of the variable (e.g., 25 years old or 70 years old). Because selection is the only control possible, age and other "subject variables" can present special problems of interpretation—an issue to which I return later in the chapter.

A bit more terminology is necessary before proceeding. Independent variables are also referred to as **factors**, and the particular values that the variables take are referred to as **levels**. The Brownell et al. (2009) study, therefore, can be described as a  $2 \times 2$  factorial design—that is, an experiment with two factors, each of which has two levels. The Kliegel et al. (2007) study is also a  $2$  (age)  $\times$   $2$  (condition) factorial design. Note that symbolizing the design in this way serves to

tell us the number of distinct cells or groups in the experiment. For example, in the Kliegel et al. study there are four ( $2 \times 2$ ) distinct groups: young adults in the familiar-context condition, young adults in the novel-context condition, old adults in the familiar-context condition, and old adults in the novel-context condition. If sex had been included as a variable, the study would have had a  $2$  by  $2$  by  $2$  design with eight distinct cells (young men in the familiar-context condition, young women in the familiar-context condition, etc.).

## Validity

All research involves variables and the relations that hold among variables. When we wish to describe research, therefore, the construct of variables is central: What kinds of contrasts are being examined, and what forms do the examinations take? When we wish to move beyond description to *evaluation* of research, the central construct becomes that of **validity**. The question of validity is the question of accuracy: Has the study in fact demonstrated what it claims to demonstrate? All of the specific methodological points discussed throughout the book come down to this one basic question of the accuracy of the conclusions that we draw from research.

Various forms of validity can be distinguished (Shadish, Cook, & Campbell, 2002). In this chapter, I discuss three forms: *internal*, *external*, and *construct*. Chapter 9 adds a fourth form: *statistical conclusion validity*.

**Internal validity** applies within the context of the study itself. The issue in question is whether the independent variables really relate to the dependent variables in the manner claimed. Have we drawn the correct conclusions about the causal impact (or lack of causal impact) of one set of variables on the other set? Let us take the Kliegel et al. (2007) study as an example. Their conclusions are internally valid if young adults and older adults really are equivalent in planning ability in familiar contexts but pull apart when the context is unfamiliar,

with only the young adults able to maintain the same level of performance when coping with a novel environment. If there is a plausible alternative explanation for this pattern of findings, then the internal validity of the study is thrown in doubt. Suppose, for example, that the young adults were a more select group than the older adults—perhaps all college graduates at the younger age but a mixture of educational levels in the older group. If so, we would have an alternative explanation for the better performance by the young adults: The difference reflects not a natural change with age but rather a difference in ability level, with more capable participants more skilled at dealing with novelty. (I discuss this problem, labeled *selection bias*, more fully later.)

The question of **external validity** is the question of generalizability. It applies, therefore, once we move outside the immediate context of the study. The question now is whether we can generalize the findings of the study to other samples, situations, and behaviors—not just any samples, situations, and behaviors, of course, but those for which we wish the study to be predictive. In this case, let us take the Brownell et al. (2009) study as the example. Their findings would have external validity if they applied to toddlers in general and not just to the toddlers in the study, to sharing in general and not just to the particular form of sharing examined, and to contexts in general and not just to the lab environment that was the locus for the research. If any of the findings fails to generalize across these dimensions, then that finding lacks external validity. Perhaps, for example, the findings hold only within the confines of the laboratory and do not apply to sharing in the natural environment. If this limitation actually held (other research makes it doubtful that it does), then the study would have limited external validity.

Exactly what forms of generalizability are important varies to some extent across studies. Table 2.3 lists and briefly describes the most common dimensions that are relevant to external validity.

A satisfactory study must have both internal validity and external validity. As the classic treatment by D. T. Campbell and Stanley (1966) observes, “internal validity is the basic minimum without which any experiment is uninterpretable” (p. 5). Logically the internal validity question is the primary one, because findings can hardly be generalized if there are no valid findings in the first place. External validity is also critical, however. Internally valid conclusions do not mean much if they cannot be generalized beyond the study in which they occur.

Table 2.3  
*Dimensions of External Validity*

Dimension	Issue
Sample	Do the results generalize beyond the sample tested to some broader population of interest?
Setting	Do the results generalize beyond the setting used in the research (e.g., a structured laboratory environment) to the real-life settings of interest (e.g., behavior at home or at school)?
Researcher	Are the results specific to the research team that collects the data, or would the same results be obtained by any team of investigators?
Materials	Are the results specific to the particular materials used to represent the constructs of interest, or would the same results be obtained with any appropriate set of materials?
Time	Are the results specific to the particular time period during which the data were collected, in either a short-term sense (e.g., a measure administered in late afternoon) or a long-term sense (e.g., a measure affected by historical events)?

Internal validity is also a prerequisite for the third form of validity: **construct validity**. Construct validity has to do with theoretical accuracy: Have we arrived at the correct explanation for any cause-and-effect relations that the study has demonstrated? We assume, in other words, that we have internally valid conclusions; the question now is whether we know *why* the results have occurred.

Suppose, for example, that we are confident that the verbalization manipulation in the Brownell et al. (2009) study really did cause the difference in response between the older and younger toddlers. Why did the verbalization have this effect? The most obvious explanation—and the one favored by Brownell et al.—is that the verbalization was sufficient to alert the older children to the adult's desire for the snack, thus activating their nascent prosocial tendencies; the younger children, however, were not yet capable of using such a cue. But perhaps there is a different basis for the age difference. Perhaps, for example, the children interpreted the verbalization as a request for compliance, and the older children were more likely to comply with an adult's request than were the younger ones. If so, the study would really be assessing obedience, not sharing. (As Brownell et al., 2009, note, further considerations make this interpretation unlikely.) If plausible competing explanations for the results cannot be ruled out, then the study lacks construct validity.

The preceding discussion has been just a first pass at constructs that recur in various contexts throughout the book. For now, let us settle for one more point with respect to validity. It concerns the difficulty of simultaneously achieving the various forms of validity in the same study. This difficulty exists because often research decisions that maximize one form of validity work against another form. The trade-off is most obvious with regard to internal and external validity. In general, the more tightly controlled an experiment is, the greater its internal validity—that is, the more certain the experimenter can be that the variables really do relate in the manner hypothesized. At the same time, the artificiality

of a tightly controlled experiment may make generalization to the nonlaboratory world hazardous. Conversely, research conducted in natural settings with naturally occurring behaviors may pose little problem of generalizability, because the situations to which the researcher wishes to generalize are precisely those under study. The lack of experimental control, however, may make the establishment of valid relationships very difficult.

## Sampling

Decisions about variables have to do with the *what* of research: What independent variables am I going to manipulate, and what potential outcomes of these variables am I going to measure? Also important are decisions about *who*: With what sorts of participants am I going to explore these independent variable–dependent variable links?

The selection of participants for research is referred to as **sampling**. Sampling is important because of the constraints on the scope of research. With very rare exceptions, psychologists are not able to study all of the people in whom they are interested. The researcher of infancy, for example, is not going to test all of the world's babies, or even all those in the United States, or (probably) even all those in one specific geographical community. Instead, what researchers do is to test **samples**, from which they hope to generalize to the larger **population** of interest. The generalization is legitimate if the sample is *representative* of the larger population. This, clearly, is an issue of external validity.

How can researchers ensure that a sample is representative of the population to which they wish to generalize? A logical first step is to define what the population of interest is. It need not be as broad as all of the world's infants; more likely, perhaps, is something like “all full-term, healthy 3-month-olds growing up in the United States.” Once the desired population has been defined, the next step is **random sampling** from that population. As the term implies, random sampling

means that every member of the population has an equal chance of being selected for the research. If all members of the population really are equally likely to be selected, then the most probable outcome of the sampling process is that the characteristics of the sample will mirror those of the population. Note, however, that the likelihood that this desired outcome will in fact be achieved varies directly with the size of the sample. A random sample of 100 is a good deal more likely to be representative than a random sample of 10. This principle is just one of a number of arguments (we will encounter some others in Chapter 9) for using large rather than small sample sizes.

In some instances, researchers may use modified forms of random sampling, especially when the intended sample size is limited and pure random selection might therefore not produce the desired outcome. In **stratified sampling** researchers first identify the subgroups within the population that they want to be sure are represented in their correct proportions in the final sample. A researcher might want to be sure that males and females are represented equally, for example, or that different ethnic groups appear in proportions that match their numbers in the general population, or that freshmen are just as common as seniors in a college student sample. Samples are then drawn in the desired proportions from the identified subgroups—thus, equal numbers of males and females, 25% of the participants from each year in college, and so forth.

The goal of stratified sampling is to ensure that different members of the population are represented in their actual proportions in the sample selected. In contrast, with **oversampling** the researcher deliberately samples one or more subgroups at rates *greater* than their proportion in the target population, the goal being to achieve a sufficiently large sample of the subgroup to permit conclusions. Suppose, for example, that we plan to conduct a survey of high school students in which comparisons among ethnic groups are one of the issues of interest, and suppose also that Asian Americans constitute 3% of the high school population in the city in which we are working.

Even with a total sample of 1,000 students, a random sampling approach will give us only about 30 Asian American participants, which may not be enough to draw conclusions. If we deliberately oversample Asian Americans, however (say at a 6% rather than a 3% rate, thus giving 60 students total), we can end up with a sufficient subsample for analysis, while still achieving adequate numbers in the other groups of interest.

How often do psychologists in fact draw their samples in the textbook-perfect fashion just described? The answer is: not very often. Random sampling and its variants are occasionally found in psychological research—perhaps most commonly in large survey projects in which it is important that the sample match some target population. More generally, most researchers undoubtedly start with at least an implicit notion of the population to which they wish to generalize, and most would certainly avoid selecting a sample that is clearly nonrepresentative of this population. Nevertheless, true random sampling from some target population is rare. The most obvious and frequent deviation from randomness is geographical. Researchers tend to draw samples from the communities in which they themselves live and work. Often, moreover, they may sample from only one or a few of the available hospitals, day care centers, or schools within the community. Such selection of samples primarily on the basis of availability or cooperation is referred to as **convenience sampling**. Samples obtained in this way may not be representative of the broader population with respect to variables such as social class and race, and they *cannot* be completely representative with respect to variables like region of the country or size of the community. A recent survey of leading journals in developmental psychology indicated that 80% to 90% of the samples were convenience samples (Bornstein, Jager, & Putnick, 2013).

How important are these deviations from random sampling? There is no simple answer to this question; among the dimensions that are relevant are the topic under study; what the researcher wishes to conclude about the topic;

and, of course, just how nonrandom and potentially nonrepresentative the sample is. We revisit issues of sampling throughout the book in the context of particular kinds of research. For now, I settle for two pieces of advice, one directed to the reader of research reports and the other to the author of such reports.

The advice for the reader is to make a careful reading of the Participants section an important part of the critical evaluation of any research project. However satisfactory the other elements of a study may be, the results do not mean much if the sample is not representative of some larger population of interest. One question concerns the standing of the sample on the demographic characteristics that may affect response. At the least, these characteristics will include age, sex, and race; for particular studies additional dimensions (e.g., income level, geographical region, health status) may also be important. Another question concerns the method of recruitment. What was the initial pool from which participants were drawn, how many of these potential participants actually made it into the study, and (if there was any dropout) how many stayed in the study until the end? Finding a representative pool of potential participants is a good starting point for research, but it is not sufficient; the real question is how well the final sample reflects the starting point.

The advice for the author follows from the points just made. Readers cannot critically evaluate the samples for research if Participants sections do not tell them enough about the samples. It is the author's responsibility to make sure that all of the necessary information is conveyed to the reader. Helpful further sources with respect to what sorts of information to convey include the *APA Publication Manual* (APA, 2010b), Hartmann (2005), and Rosnow and Rosnow (2012).

## Control

The notion of control was touched on in each of the preceding sections. Recall that the independent

variable is defined as a variable that is under the control of the researcher. Control is central to the establishment of validity, especially internal validity. And selection of the right participants is one sort of control that a researcher must exercise. The purpose of the present section is to discuss the further sorts of control that become important once participants are in hand.

As Table 2.4 indicates, three forms of control are important in the execution of studies. The table summarizes the forms and gives examples of how each type applies or might apply to the illustrative studies.

One type of control concerns the exact form of the independent variable. If the interest, for example, is in the effects of a certain kind of reinforcement, then the researcher must be able to deliver exactly this kind of reinforcement to the participants. If any unintended deviations occur—in form, timing, consistency, or whatever—the researcher can no longer be certain what the independent variable is. Or consider again the Kliegel et al. (2007) examination of planning behavior. Because the researchers' interest was in possible effects of context, it was critical that they present the same novel-familiar contrast to all of the participants.

The point being made about this first form of control is hardly an esoteric one. It is simply that if one wants to study the possible effects of something, one must first be able to produce that something. Note, however, that doing so is not always as easy as in the Kliegel et al. (2007) study, in which the levels of the independent variables were defined simply by the different stimulus materials that were presented. When the experimental manipulation is more complicated, delivering the variable in the same form to all participants can become a challenge. The challenges, moreover, are often multiplied when children are the participants, a point to which I return later.

A second form of control has to do with factors in the experimental setting other than the independent variable. Independent variables do not occur in a vacuum; there must always be a

Table 2.4  
*Forms of Control in Experimental Research*

Type of control	Methods of achieving	Examples from illustrative studies
Over the independent variable	<ul style="list-style-type: none"> <li>• Make the critical elements of the experimental manipulation the same for all participants</li> </ul>	<ul style="list-style-type: none"> <li>• In Kliegel et al. (2007), present the instructions and stimuli for the familiar and novel conditions in the same way to all participants</li> </ul>
Over other potentially important factors in the experimental setting	<ul style="list-style-type: none"> <li>• Hold the factors constant for all participants</li> <li>• Disperse variations in the other factors randomly across participants</li> </ul>	<ul style="list-style-type: none"> <li>• In Brownell et al. (2009), use the same quiet testing room for all participants</li> <li>• In Kliegel et al., vary the time of testing randomly across participants</li> </ul>
Over preexisting differences among the participants	<ul style="list-style-type: none"> <li>• Randomly assign participants to experimental conditions</li> <li>• Match participants on potentially important attributes prior to experimental conditions</li> <li>• Test each participant under every experimental condition</li> </ul>	<ul style="list-style-type: none"> <li>• In Kliegel et al., randomly assign half of the participants at each age to the familiar condition and half to the novel condition</li> <li>• In Kliegel et al., measure the participants' IQs and assign equal-IQ participants to the different conditions (not actually done)</li> <li>• In Brownell et al., test every child in both the verbalization and no-verbalization conditions</li> </ul>

context for them, and it is the job of the researcher to determine exactly what this context will be. In the Kliegel et al. (2007) study, for example, the researchers had to decide not only what stimuli and instructions to use but also what the immediate environment for the testing would be. One easy decision in this particular case is to make the environment as quiet and clutter-free as possible, in order to minimize distractions. Once the experimenter has made this decision, it is then his or her job to ensure that each participant receives the same quiet environment.

Let us introduce some further terminology at this point. Differences in scores on the dependent variable are referred to as the *variance* of the study. Those differences that can be attributed to the independent variables are called **primary variance**; those that result from other factors are called *secondary variance* or *error variance*. By controlling the level of other potential variables,

experimenters attempt to maximize the proportion of primary variance in the study. Perhaps even more important, they attempt to make sure that other sources of variance are not systematically associated with any of the independent variables. Suppose, for example, that Kliegel et al. (2007) had tested all of their young adult participants in a quiet laboratory on campus but all of their older participants in a noisy room at a senior citizens' center. Clearly, in this case there would have been two independent variables—age and testing environment—when only one had been intended. Any such unintended conjunction of two potentially important variables is referred to as **confounding**. A major goal of good research design is to rule out confounding.

As Table 2.4 indicates, control of unwanted variables can take a couple of forms. Often it is possible to control the variable by making it the same for all participants. This is the case in

the errand planning study, in which the testing environment can be held constant for all participants. Sometimes, however, such literal equating is not practical. We can return to the Brownell et al. (2009) study for an example. In a study such as theirs, the time of day at which the testing occurs may be important. As any parent of a toddler knows, young children are more alert and responsive at some times of the day than at others. In addition, the desirability of the snack might well depend on how recently the child had eaten. Clearly, Brownell et al. would have introduced a potentially important confounding if they had tested all of the younger toddlers early in the day and all of the older toddlers late in the afternoon. One way to avoid this problem would be to test all of the children at the same point in the day, say at 10:00 in the morning. With this approach, however, most studies would take months to complete, and even then, only time of day and not day of the week or time of year (which also can be important) would be held constant. A sensible alternative would be to let the time of testing vary across children but to make sure that the variations are the same for the groups being compared—in the present example, younger and older toddlers. In this case, the control of the time-of-testing variable would lie not in its equation but in its randomization—that is, by dispersing differences in it equally across the groups of interest.

Shorn of certain specifics, the discussion thus far should have a familiar sound to it. What has been presented here is simply the classic scientific method: to determine the effects of some factor, systematically vary that factor (the first form of control) while holding other potentially important factors constant (the second form of control).

There is still a third form of control that is essential. Thus far, the “other potentially important factors” that have been discussed have been factors within the experimental setting—for example, the noise level of the testing room. Another important source of variance in any experiment stems from individual differences among the participants. Participants are not

identical at the start of an experiment, and differences among them contribute error variance to the final results. Because there is no way to rule out such differences, the method of control must again be through dispersion rather than equation. What the experimenter must ensure is that the differences are spread equally across the different treatment groups—or, to make the same point in different words, that the groups are equivalent prior to the application of the treatment. Doing so requires that the experimenter have control not only over the form of the treatment but also over who gets what treatment.

How can the experimenter assign people to groups in a way that will ensure that the groups all are initially equivalent? The answer is that although there is no way literally to ensure equivalence, there are ways to come as close as can reasonably be expected. The most common method is through **random assignment** of participants to the different groups. Random assignment means that each participant has an equal chance of being assigned to each group. If each participant has an equal chance of being assigned to each group, then the characteristics associated with each participant (IQ, sex, relevant past experience—whatever might affect the results) have an equal chance of falling in each group. It follows that the most probable outcome of the assignment process is that these characteristics will end up equally distributed in the different groups, a result that is, of course, the researcher’s goal. The logic of random assignment is clearly the same as the logic of random sampling, and the success of the process shows a similar dependence on sample size. One could not randomly divide eight participants into two groups and conclude with any confidence that the randomization had produced equivalent groups. With a sample of 80 participants, the odds are much better.

Random assignment is a much more frequent component of research than is true random sampling. Indeed, random assignment has been referred to as “the key defining attribute of the experimental method” (McCall & Green, 2004, p. 4).

Powerful though random assignment is, it does have a limitation. At best, random assignment makes it *probable* that the groups being compared are equivalent; it cannot guarantee this outcome. An obvious question follows: Why settle for probability? Why not identify the dimensions on which we wish the groups to be equal (e.g., intelligence, socioeconomic status, health status—the list will vary across studies) and then assign participants based on these dimensions—thus, the same proportion of high-intelligence participants in each group, the same proportion of middle-income participants in each group, and so forth? Why, in short, not do the assignment in a way that *ensures* equivalence?

The general answer to this question is that such matching is more difficult than might at first appear and that the attempt to achieve it can sometimes create more problems than it solves. A more specific answer is given in Chapter 3, when we return to the issue of selecting and assigning participants. Also discussed in Chapter 3 is the third general technique for achieving equivalence: testing every participant under each experimental condition.

## Subject Variables

### Manipulable Versus Nonmanipulable Variables

Thus far the discussion of experimental control has focused on the ideal situation for research: the case in which the researcher can systematically manipulate the independent variables of interest while holding all other variables constant and can assign participants to the different treatment groups either randomly or randomly within certain desired constraints. With many variables, such control is not only desirable but also quite feasible. We saw examples of this kind of control in both of the example studies.

The developmental psychologist's life is complicated, however, by the fact that not all variables of interest lend themselves to the kind of manipulation that good research design demands. Again, both of

the cited studies provide examples, and in this case it is the same example: chronological age. Clearly, age is not something that the researcher randomly assigns to people; rather it is a characteristic that people bring to the experimental setting. Age is just one example of what are called **subject** (or *classification* or *attribute*) **variables**: intrinsic properties of individuals that cannot be experimentally manipulated but must be taken as they naturally are. Other common examples are race and sex. The researcher who wishes to work with such characteristics as independent variables forgoes the possibility of control through manipulation. The only control possible in such cases is control through selection of people who already possess the characteristic.

A number of other variables of interest, although not literally nonmanipulable, are never in fact the subject of controlled experiments with humans. From a theoretical perspective, for example, it would be very interesting to know whether infants deprived of mothers develop in the same way as infants who have mothers. Needless to say, we do not have manipulative studies of this issue. Yet there has long been a literature on “maternal deprivation” and its effects on the child. What researchers have done is to identify situations in which infants have already been left motherless (usually in orphanages) and then take advantage of these “natural experiments” by studying how the infants develop. And there are numerous similar examples of psychologists’ ability to capitalize on naturally occurring events—studies of malnutrition in infancy, of father absence during childhood, of social isolation in old age, and so forth. In each case the independent variable is created through selection rather than experimental manipulation.

Research with nonmanipulated variables does not attain the status of the “true experiment,” because the controlled manipulation that constitutes the heart of an experiment is not possible. For this reason, such research is labeled as *preexperimental* in D. T. Campbell and Stanley’s (1966) influential discussion of experimental design. Because of the lack of control, such studies can never establish cause-and-effect conclusions with the certainty that is possible in a manipulative experiment.

What exactly are the limitations of such research? The problems are of two main sorts. First, it is impossible to assign participants randomly to groups. Because random assignment is impossible, there is no way to be sure that the groups under study are equivalent except for the variable of interest (e.g., presence or absence of mother), and therefore no way to be sure that any differences between groups are caused by that variable. This, in fact, was one criticism of the early maternal deprivation studies. Perhaps babies who grow up in orphanages are a nonrandom subset of the general population of babies, a subset that includes an unusually high proportion of genetic or organic problems. If so, then differences between orphanage babies and other babies could not be attributed with any confidence to the effects of the orphanage rearing. In a well-designed experiment, such confounding would be ruled out by random assignment. This, it should be clear, is a problem with internal validity: We cannot be certain that our independent variable is really the causal factor.

The other problem concerns the broad-scale and longstanding nature of most subject variables. Orphanage rearing, father absence, social isolation, growing up Black (or White), and growing up male (or female) all encompass a host of factors

that can affect an individual's development. Thus, even if we find a significant effect associated with a particular subject variable, we still do not know what the specific causal factors are. This, too, has been a problem in research on maternal deprivation. Although the damaging effects of certain kinds of orphanage rearing are not in dispute, there has long been debate about whether the effects result from lack of normal mothering or from more general cognitive-perceptual deprivation. Even if we could conclude that mothering *per se* is important, we still would not know which of the many things that mothers normally do with infants are critical to the effect. Again, there is a confounding of factors that a well-designed experiment would keep separate. A researcher with control over variables is unlikely to set up an independent variable that is so global that its effects cannot be interpreted. This, it should be clear, is a problem with construct validity: We do not know whether we have arrived at the correct theoretical interpretation of the results.

This discussion is not meant to suggest that there is no value in demonstrating that a variable like maternal deprivation or age or sex is associated with important outcomes in the child. But it should be realized that such a demonstration is merely the first step in a research program.

### BOX 2.1 CONTROLLED STUDIES WITH ANIMALS

A point may have occurred to you in reading the text's discussion of nonmanipulable variables. It is true that we cannot deliberately create bad environments in the study of human development—deprive infants of their mothers, expose fetuses to potentially damaging drugs, or whatever. But what about research with animals? Assuming that we *can* perform such manipulations with other species (a point to which I return), then animal research opens up possibilities that are not available in research with human participants.

In fact, research with animals has played an integral role in the development of psychology as a science, including developmental psychology (Carroll & Overmier, 2001; Huag & Whalen, 1999). Such research has taken a variety of forms, not all of which (or even most of which) have involved the study of deliberately negative experiences. Nevertheless, two considerations make the focus on negative experience an especially important subset of the animal literature. One is the importance of the issues involved. The other is the impossibility of examining these issues experimentally with humans.

(Continued)

(Continued)

Decades of animal research have explored a wide range of potentially important variations in an organism's early experience. Studies of dark rearing, for example, have been used to pull apart innate and experiential contributors to perceptual development. Studies of selective breeding provide a powerful tool for identifying the genetic underpinnings for development. Studies of cross-rearing, in which an infant is separated from the biological parents to be reared by other parents, speak to the same issue. Clearly, none of these forms of research is possible with humans.

The animal literature also includes extensive work on the two topics introduced in the opening paragraph. Research under the heading of **teratology** examines the potentially adverse effects of atypical experience during the prenatal period—exposure to drugs, for example, or radiation, or disease. Controlled studies of various potential teratogens with a range of species have provided a valuable complement to the study of naturally occurring instances with humans. Controlled studies with animals have also contributed to our understanding of the importance of early social experience, including the effects of maternal deprivation. The best known such work is that of Harry Harlow with rhesus monkeys (Harlow, 1958). The Harlow studies appeared at about the same time as the first studies of orphanage rearing in humans, and the convergence of evidence from the two lines of research was a powerful force in sensitizing the field to the importance of the early social environment.

Valuable though the work with other species has been, several limitations of such work are important to note. The most basic limitation concerns the difficulty of generalizing across species. Work in teratology provides one well-known example. The drug thalidomide, when taken at a certain point in pregnancy, resulted in the births of thousands of malformed infants in the late 1950s and early 1960s. Yet thalidomide had been tested on a range of animal species prior to its use with humans, and no adverse effects had been reported. Thus, what was true of one species turned out not to be true of another.

It is, of course, an empirical question whether a particular experience has a similar effect across species, and decades of careful research have provided a number of guidelines as to when animal models are likely to be informative with respect to human development (Gottlieb & Lickliter, 2004; Overmier, 1999). For many issues, however, there simply is no close animal analog. This point applies at both the independent variable and dependent variable end. Thus, there is no form of animal research that can inform us about the impact of divorce on children's development, or the effects of the internet on adolescent functioning, or the contribution of education to economic success. Nor can animal research tell us why some children do better in school than others, or about variations in speed of language learning, or about the sources of life satisfaction in old age. Many topics in human development are uniquely human topics.

A final limitation concerns the "can" question raised in the opening paragraph. We will see in Chapter 10 that various codes of ethical conduct now govern all research with human participants, and that many experiments that were carried out in the first 50 or so years of the discipline would therefore be impossible today. The same point applies to research with animals (Akins, Panicker, & Cunningham, 2005). Whether the deliberate imposition of painful or damaging experiences is ever justified in research is a difficult and much debated question. Today's ethical guidelines, however, ensure that such research is less common than was once the case. The result is some constraint on scientific possibility, but this is a price that most researchers are quite willing to pay.

## Age as a Variable

Because of its importance in developmental research, the variable of chronological age deserves a somewhat fuller consideration. Much research in developmental psychology has as one of its points a demonstration that participants of different ages either are or are not similar on the dependent variables being studied. Even studies with a single age group may have age comparisons at their core, for often the comparison is implicit rather than explicit. A researcher of neonates, for example, may not include a comparison group of older children in the study, but findings about how neonates function can nevertheless be interpreted in light of a large body of information about the functioning of older children. To take a simple example, one would hardly do research to determine whether young infants have color vision (e.g., Adams, 1995) unless one already knew that color vision is eventually part of the human competence.

Developmental psychologists are sometimes apologetic about the “merely age differences” nature of much research in developmental psychology. But the identification of genuine changes with age is clearly a valid part of a science of development. Not only is description a legitimate part of any science, but accurate description provides the phenomena to which explanatory models must speak. It is only when we know, for example, that young children do not understand conservation (Piaget & Szeminska, 1952) that we can begin to build a model of why this fact is so and of where eventual understanding comes from.

Although we may agree that the study of age changes is legitimate, it is important to be clear about exactly what is meant by a “genuine change with age.” What is *not* meant, certainly, is that chronological age in any direct sense causes the change. What *is* meant is that variables that are regularly and naturally associated with age produce the change. It is then the job of the researcher to determine which of the potentially important variables are in fact important.

The earlier discussion stressed that a primary goal of experimental control is the creation of

groups that are equivalent in every way except for the independent variable being examined. This goal takes on special meaning in the case of a broad subject variable like age. Imagine that you are interested in comparing 7-year-olds and 12-year-olds. If you wish to make the groups equivalent in every way except age, then you will have to find 7- and 12-year-olds whose levels of biological maturation are the same, who have been going to school for the same number of years, whose general experiences in the world are equivalent, and so forth. Clearly, such a goal is not only impossible but quite misguided. Variables like biological maturation, years of schooling, and general experience are among the variables that are “regularly and naturally associated with age.” As such, they are factors to be studied, not ruled out through experimental control.

On the other hand, there are other potentially important factors that must not be allowed to confound the age comparison. An obvious kind of confounding would occur if all of the 7-year-olds were boys and all of the 12-year-olds girls. Maleness is not an intrinsic part of being 7, nor is femaleness an intrinsic part of being 12; hence, this factor must not be allowed to covary with age. A somewhat less obvious confounding might occur if all of the 7-year-olds were drawn from one school and all of the 12-year-olds from another school. The mere fact of attending different schools is probably not important, and in any case this difference may be unavoidable for the particular age range studied. Nevertheless, it will be important for the researcher to select schools that are as comparable as possible on dimensions such as educational philosophy, geographical location, and socioeconomic status of the population served. If this criterion is not met, then an apparent age change may not in fact be genuine.

As these examples suggest, decisions about what to match and what not to match when comparing different ages are generally straightforward. As we will see, however, such decisions are not always straightforward, nor is it always easy to achieve whatever matching one has decided on. We will return to the issue of age comparisons in Chapter 3.

## Outcomes

Researchers manipulate independent variables in order to examine effects on dependent variables. But what are the possible effects? In a factorial study—that is, a study with two or more independent variables—the possible effects are of two sorts: main effects and interactions.

### Main Effects

A **main effect** is a direct effect of an independent variable on a dependent variable. It is what researchers examine when they compare the levels of a single independent variable independent of (or summed across) the other independent variables in the study.

The Kliegel et al. (2007) study provides an example of a main effect. Recall that the young adults in their sample outperformed the older adults. The means for this effect are shown in the rightmost column of Table 2.2; they are the values for all the young participants and all the old participants in the study, summed across the levels of the other independent variable (the familiar-novel contrast). The effect is a main effect because we are considering only a single independent variable—in this case, the age of the participants.

Because it included two independent variables, the Kliegel et al. (2007) study had a second potential main effect: that of experimental condition. The relevant values in this case are shown at the bottom of Table 2.2: the means for all participants in the familiar condition and all those in the novel condition, summed across the two levels of age. In this case, however, the difference between the means was too small to achieve statistical significance. The same conclusion applies to the two potential main effects in the Brownell et al. (2009) study: that of age (the .53 vs. .56 comparison in Table 2.1) and that of condition (the .51 vs. .58 comparison). (I discuss the notion of statistical significance in Chapter 9.)

## Interactions

A main effect is an effect of a single independent variable considered in isolation. An **interaction**, in contrast, becomes possible when we consider two or more independent variables simultaneously. An interaction occurs whenever the effect of one independent variable varies with the level of another independent variable.

Both of the example studies produced interactions. In the Brownell et al. (2009) study the effects of the adult's verbalization varied with the level of the age variable: no effect for the younger children, a positive effect for the older children. As with any interaction, the results can also be stated with the opposite emphasis. The effects of age varied with the level of the verbalization variable: no effect in the no-verbalization condition, a marked effect in the verbalization condition. Note that in the case of an interaction, in contrast to a main effect, it is the individual cell means that are relevant (thus .55, .50, etc.).

The interaction in the Brownell et al. (2009) study is an example of a two-way interaction—"two-way" because two independent variables are involved. Figure 2.1 presents the interaction graphically. The data are the same as those presented in Table 2.1; the figural presentation, however, makes the nature of the interaction more visible. Note, in particular, the nonparallel nature of the lines. Graphically, an interaction is always signaled by some deviation from parallelism—some spreading apart or crossing over of lines that reflects the differential effects of one variable across the levels of the other. Conversely, if there were no interaction the lines would be parallel or close to parallel, reflecting the fact that the effect of each variable is constant across the levels of the other variable.

The Kliegel et al. (2007) study also produced a two-way interaction. As we saw, the younger participants in their study outperformed the older participants—only, however, in the novel condition. Thus, the effect of age varied with the level of the experimental variable. Figure 2.2 presents this interaction graphically.

The interactions in the two example studies were between a subject variable and an experimentally manipulated variable. Interactions are not limited to such designs, however; rather, they can occur between independent variables of any sort. Interactions are possible, therefore, in any multiple-factor experiment. I add two further examples to illustrate this point: the first an interaction between two experimentally manipulated variables, and the second an interaction between two subject variables.

A study by Moore (2009) examined sharing behavior in 5-year-olds as a function of two factors. One was the recipient of the sharing: a friend, a nonfriend, or a stranger. The other was whether the decision to share entailed a sacrifice on the child's part (a condition labeled "sharing") or whether the sharing could be done with no cost to the child (a condition labeled "prosocial"). Figure 2.3 shows the results (the "1, 1" choice denotes sharing). When

sharing involved a cost, children shared most with a friend, next most with the nonfriend, and least with the stranger. In contrast, when no cost was involved, children were as generous with the stranger as they were with the friend; in this case only the nonfriend fared less well. The effect of one of the experimental variables in the study thus varied with the level of the other variable.

The focus of a study by Sumter and associates (Sumter, Bokhorst, Steinberg, & Westenberg, 2009) was a topic of considerable importance: adolescents' susceptibility to peer influence (a topic to which I return in Chapter 14). Susceptibility was measured via an instrument called the Resistance to Peer Influence Scale. Figure 2.4 shows response to the scale as a function of the participants' age and gender (higher numbers indicate greater resistance). It can be seen that girls reported more resistance than boys across the adolescence age span, but the difference

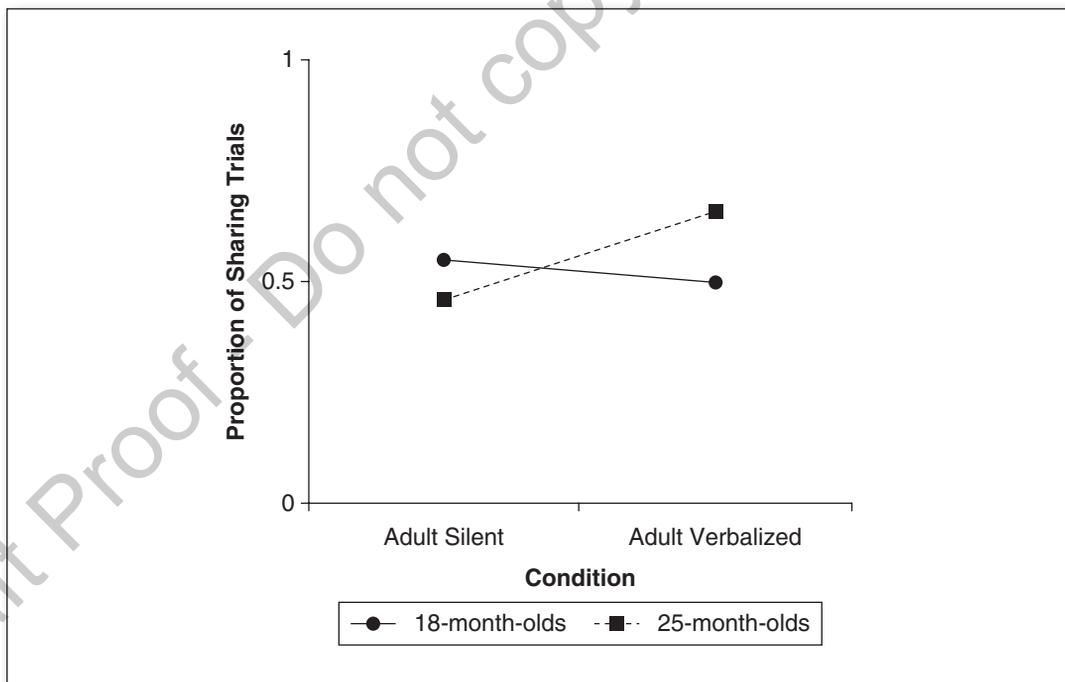
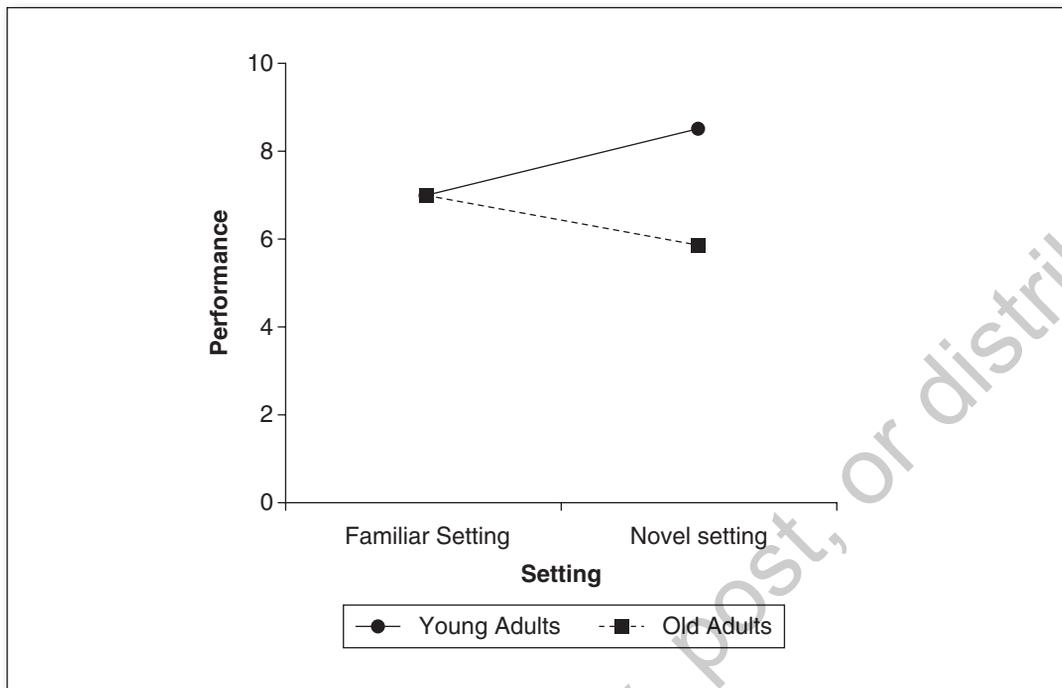


Figure 2.1 Interaction of age and condition in the Brownell et al. study.

Adapted from "To Share or Not to Share: When Do Toddlers Respond to Another's Needs?", 2007, by C. A. Brownell, M. Svetlova, and M. Nichols, 2009, *Infancy*, 14, 117–130.



**Figure 2.2** Interaction of age and condition in the Kliegel et al. study.

Adapted from "Adult Differences in Errand Planning: The Role of Task Familiarity and Cognitive Resources," by M. Kliegel, M. Martin, M. A. McDaniel, and L. H. Phillips, 2007, *Experimental Aging Research*, 33, 145–161.

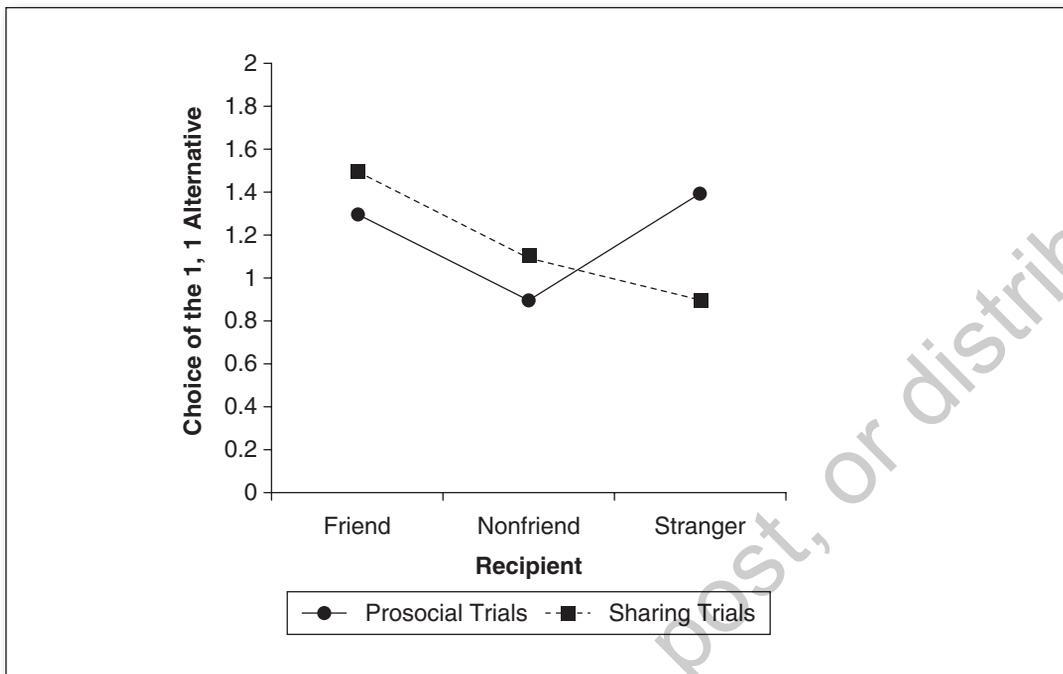
was greatest during the mid-adolescence years; indeed, it was only at this age that the gender difference achieved statistical significance. The effects of gender thus varied with the age of the child: little effect in early or late adolescence, a definite effect in mid adolescence.

As a comparison of the figures suggests, interactions can take a variety of forms. They can also become exceedingly complicated when a variable has several levels or when more than two independent variables are involved. Although some researchers try, it is seldom possible to make sense of a four- or five-way interaction.

Interpreting any sort of interaction can be a complex matter, both statistically and theoretically. I settle here for one basic point. The most general implication of a significant interaction between two variables is that interpretations of main effects involving those variables must be

made with caution. In the Kliegel et al. (2007) study, for example, there was a main effect of age; as Figure 2.2 reveals, however, the age effect was limited to the novel condition. In the Brownell et al. (2009) study, in contrast, the main effect of age was *not* significant, a finding that would suggest that this variable had no effect. The interaction, however, tells us that age did have an effect but only under one of the two experimental conditions. An interaction, then, is a signal that the world is more complicated than we might have expected. Studying an independent variable in isolation cannot give us a full picture of the way in which that variable operates.

Note that the point just made can also be put in the context of external validity. An interaction implies a limitation in the generality of conclusions about the independent variables that enter into the interaction. In the Kliegel et al. (2007) study,



**Figure 2.3** Interaction of experimental conditions in the Moore study.

Adapted from "Fairness in Children's Resource Allocation Depends on the Recipient," by C. Moore, 2009, *Psychological Science*, 20, 944–948.

for example, the effect of age did not generalize across the levels of the context variable, nor did the effects of context generalize across the levels of age. Conversely, the absence of an interaction is

evidence in support of the external validity of conclusions regarding the variables in question—at least across the particular dimensions and levels that are sampled.

### BOX 2.2 PSYCHOLOGY'S "REPLICATION CRISIS"

The term **replication** refers to a research project that duplicates the essential elements of some previous study, the attempt being to determine whether the original results can be obtained again. Successful replication is a necessary step in establishing any form of validity. It is only once we know for certain that a study's results are reproducible that we can go on to ask why the results occur or how broadly they apply.

The importance of replication is not a subject of debate. In the words of one author, "Replication is the gold standard by which scientific claims are evaluated" (Bonett, 2012, p. 410). Yet this same author goes on to state that "replication research is rare in psychology." It is rare because historically

(Continued)

(Continued)

there has been little reward for it. Those who evaluate research (e.g., journal editors, dissertation chairs) tend to place a premium on originality, and replications are therefore unlikely to find their way into the best journals in the field. One survey, dating back more than 100 years, found that only 1% of published psychology articles involved replication (Makel, Plucker, & Hegarty, 2012). Another tabulation, focusing on recent publications in developmental psychology, reported similarly low values (Duncan, Engel, Claessens, & Dowsett, 2014).

In just the last few years this situation has begun to change. It has begun to change in the wake of what has been labeled a “replication crisis” in psychology (Pashler & Wagenmakers, 2012). The crisis has multiple origins, including the failure of several well-known findings to replicate successfully; the documentation of questionable research practices that may inflate the probability of positive results; and, most seriously, a handful of instances of scientific fraud (C. Gross, 2016). The result has been an influx of commentaries (e.g., Maxwell, Lau, & Howard, 2015; Simonsohn, 2015) directed not just to the importance of replication but to ways to make replications most informative. I will note four suggestions from such writings. One is to couple direct replication with a theoretically interesting extension of the original research, thus providing new findings along with the test of old findings. Another emphasis is on the need for multiple replications of any phenomenon of interest, for a single attempt is unlikely to be definitive. A third emphasis is on the need to establish not just that an effect occurs but how large the effect is, something for which multiple studies are clearly more informative than a single study. This is a point that we will return to in Chapter 9, “Statistics.” A final suggestion is that investigators make both their procedures and their primary data fully available to others. Such transparency is necessary to ensure that replication efforts can be as fully informed as possible.

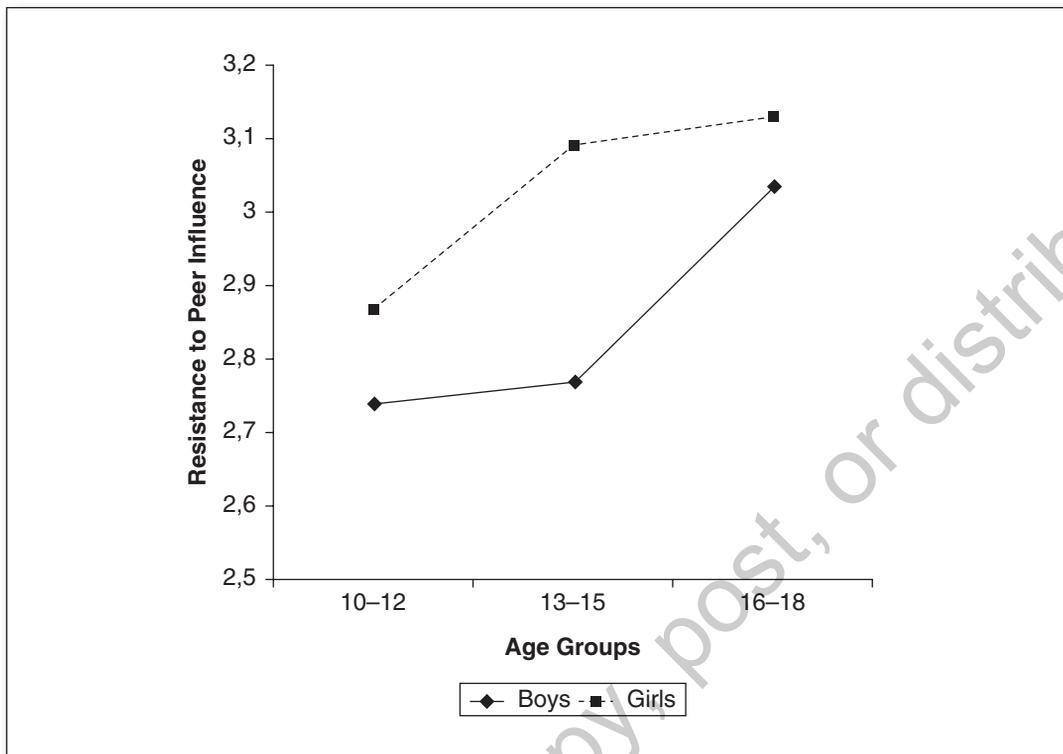
Of course, prescriptions about how to carry out replications will be of value only if investigators commit to such an endeavor and publication practices reward such commitment. Fortunately, recent years have seen the publication of several collaborative replication efforts devoted to replicating important findings in psychology (e.g., R. A. Klein et al., 2014; Open Science Collaboration, 2015). One of the field’s major organizations, the Association for Psychological Science, has also launched an initiative to encourage and to publish replications. You can track the results of this initiative at the organization’s website (<http://www.psychologicalscience.org/index.php/replication>) and in the journal *Perspectives on Psychological Science*.

## Threats to Validity

As we have seen, the ultimate goal in designing research is always to arrive at valid conclusions about the phenomena being studied. The converse to successful research design comes when there are threats to validity—uncertainties or limitations in what can be concluded that the design has failed to rule out. Several threats to

validity were touched on in this chapter, and many more are discussed in the coming chapters. It will be helpful for the coming discussion to have a brief overview of the factors to be considered—an overall list and a set of definitions that can be referred to as necessary. This is the purpose of Table 2.5.

Table 2.5 is derived from an influential monograph by D. T. Campbell and Stanley (1966) that



**Figure 2.4** Interaction of age and gender in the Sumter et al. study.

From "The Developmental Pattern of Resistance to Peer Influence in Adolescence: Will the Teenager Ever Be Able to Resist?" by S. R. Sumter, C. L. Bokhorst, L. Steinberg, and P. M. Westenberg, 2009, *Journal of Adolescence*, 32, p. 1015. Copyright 2009 by Elsevier. Reprinted with permission.

was subsequently elaborated by T. D. Cook and Campbell (1979) and Shadish et al. (2002). It does not provide an exhaustive list of things that can go wrong in research (Shadish et al. discuss 37 threats to validity!); it does, however, include many of the problems that are discussed later in the text. Again, there is no expectation that the table is completely self-explanatory; its purpose, rather, is as a preliminary guide to concepts that will receive further attention as we go along.

## Summary

This chapter begins with some basic terms and concepts. All research involves variables.

*Dependent variables* are the outcome variables in research—for example, the number of aggressive acts in a study of aggression. *Independent variables* are potential causal factors that are controlled by the researcher—for example, reinforcement for aggression. The goal of most research is to determine whether variations in the independent variable relate to variations in the dependent variable—for example, does aggression increase following reinforcement?

The basic issue with respect to all research is validity. *Validity* refers to the accuracy with which conclusions can be drawn from research. Three forms are discussed in this chapter: *internal validity*, which concerns the accuracy of cause-and-effect conclusions within the context

Table 2.5  
*Threats to Validity*

Source	Description
Selection bias	Assignment of initially nonequivalent participants to the groups being compared
Selective dropout	Nonrandom, systematically biased loss of participants in the course of the study
History	Potentially important events occurring between early and later measurements in addition to the independent variables being studied
Maturation	Naturally occurring changes in the participants as a function of the passage of time during the study
Testing	Effects of taking a test upon performance on a later test
Reactivity	Unintended effects of the experimental arrangements upon participants' responses
Instrumentation	Unintended changes in experimenters, observers, or measuring instruments in the course of the study
Statistical regression	Tendency of initially extreme scores to move toward the group mean upon retesting
Low reliability	Errors of measurement in the assessment of the dependent variable
Low statistical power	Low probability of detecting genuine effects because of characteristics of the design and statistical tests
Mono-operation bias	Use of a single operationalization of either the independent or dependent variable
Mono-method bias	Use of a single experimental method for examining possible relations between the independent and dependent variables
Compensatory rivalry	Reduction of the effects of an experimental treatment because of enhanced motivation and effort on the part of the untreated control group
Resentful demoralization	Magnification of the effects of an experimental treatment because of reduced motivation and effort on the part of the untreated control group

of the study; *external validity*, which concerns the generalizability of the conclusions beyond the study; and *construct validity*, which concerns the accuracy of the theoretical interpretation of the conclusions.

An important decision that the researcher must make concerns the participants for research. The goal in sampling participants is to obtain a *sample* that is representative of the larger *population* to which the researcher wishes to

generalize. The common prescription for achieving representativeness is to do *random sampling* from the target population. In fact, most research in developmental psychology employs sampling procedures that are less than totally random. In some instances, the deviations are intentional and systematic, the goal being to ensure that the sample possesses certain characteristics; *stratified sampling* and *oversampling* are examples. More commonly, the devia-

tions reflect the use of samples that are readily available, an approach known as *convenience sampling*. How important such departures from randomness are varies across different topics. Nevertheless, representativeness and external validity remain important questions to examine for any study.

The discussion turns next to the construct of control. Three kinds of control are important. A first is over the exact form of the independent variable. A second is over other potentially important factors in the situation. Two methods of achieving this second form of control are discussed: holding the other factors constant and randomly dispersing variations in them across participants. The third kind of control is over preexisting differences among participants. One method of achieving this form of control, random assignment, is discussed in the present chapter; two others (matching and within-subject testing) are deferred for later consideration.

In some kinds of research, the degree of control is limited by the nature of the variables. The term *subject variable* refers to preexisting differences among people that are not experimentally manipulable; examples include age, sex, and race. The only control possible with such variables is through selection, a point that applies also to situations (e.g., maternal deprivation) whose experimental induction would be unethical. Although such variables are often of great interest to the developmental psychologist, cause-and-effect conclusions are difficult to establish in the absence of experimental manipulation. Specifying the exact basis for an effect can be a problem with a broad and multifaceted variable; ruling out other possible causal factors can also be difficult.

Subject variables are often of special interest when they enter into interactions. An *interaction* occurs whenever the effects of one independent variable depend on the level of another variable. In contrast, a *main effect* refers to an effect of an independent variable that is independent of the other factors in the study. Interactions can occur

with independent variables of any sort, and they can take a variety of forms. Their most general message is that relations are complicated and that conclusions about any one variable must be made with caution.

The chapter concludes with a brief return to the concept of validity and an overview of some of the major threats to validity that are considered throughout the book.

## Key Terms

Confounding	Population
Construct validity	Primary variance
Convenience sampling	Random assignment
Dependent variables	Random sampling
External validity	Replication
Independent variables	Samples
Interaction	Sampling
Internal validity	Stratified sampling
Main effect	Subject variables
Oversampling	Teratology
	Validity

## Exercises

1. Find at least three recent summaries of developmental psychology research in the popular press (newspapers, magazines). For each, generate a list of possible threats to the validity of the research. If the description of the research is not complete enough for you to evaluate some forms of validity, specify what further information you would need.

2. Consider the task of recruiting research participants of the following ages: 6 months, 4 years, 12 years, 70 years. For each age group, generate a list of ways in which you might recruit prospective participants. For each method of sampling, discuss the likely representativeness of your final sample.

3. A particular construct can serve as either an independent or a dependent variable, depending on the way it is used in research. Consider the

following constructs: anxiety, activity level, and academic readiness. For each, generate a study in which the construct serves as (a) a dependent variable, (b) an experimentally manipulated independent variable, (c) a subject variable, and (d) a correlational variable.

4. Imagine a study with two independent variables, A and B, each of which has two levels—hence a  $2 \times 2$  design. The dependent variable, C, is measured on a scale that can range from 0 to 50. For each of the following outcomes,

draw a figure that illustrates one form that the result might take and then say in words what the outcome would mean.

- (a) significant main effects of A and B with no interaction
- (b) a significant interaction of A and B with no main effects
- (c) a significant main effect of A and a significant interaction of A and B

Draft Proof - Do not copy, post, or distribute