4

MAKING IT HAPPEN

A Hands-On Guide to a First Research Project[1]

EVERYBODY DANCE NOW

So you think you can dance? Or maybe you know painfully well that you *can't* dance. Either way, you might be interested to know that conducting research is a lot like dancing. Like a dance, research has a very specific temporal order. Thus, data collection must precede data analysis. Like dancing, research often goes better when more than one person is involved in it. Good research is often collaborative. Further, just as dancing is usually technically superior when done by experts, the same is true of research. But just as you don't have to be a ballet dancer to get into a dance club, you don't have to have a PhD in psychological statistics to conduct psychological research. Furthermore, it is arguably a lot simpler to become a decent psychological researcher than it is to become a decent dancer. For example, if you are highly uncoordinated and have no sense of timing, you may always suck at dancing. But a lack of physical skills rarely gets in the way of doing good research. Further, if someone else is offering you step-by-step suggestions while you dance, your dancing is unlikely to impress anyone. But your first research project can be pretty impressive even if it took some hand holding to get you across the finish line. Research is more about the finished product than the flow, and it usually happens in months rather than minutes. This means that when conducting research, you sometimes have a chance to correct your mistakes—or even start all over again. Such relaxed rules rarely apply to dancing.

You might critique this analogy by noting that research is much more complicated than dancing. Point taken, Mr. Baryshnikov. The same goes for you, Ms. Knowles. But almost any research project can be broken down into six steps, and none of them involve brain surgery (unless perhaps you are engaging in behavioral neuroscience). In this chapter, we'll discuss the six basic steps involved in conducting a psychological research project. These include (1) hypothesis **generation** (coming up with an idea), (2) **operationalization** (translating your idea into a concrete design), (3) **permission** (getting IRB approval), (4) **execution** (collecting your data),

(5) **calculation** (statistical analysis), and (6) **communication** (e.g., writing up your findings in the form of a paper or poster presentation). In this chapter, we will very strongly emphasize the first two steps of the six steps, namely (1) how to generate a research idea and (2) how to convert an idea into an empirical study. The reason for this emphasis is twofold. First, these two steps come first. You can't get IRB permission, for example, before you have decided what you are going to do. Second, these first two steps are probably the two that give beginning researchers the greatest amount of trouble.

Before we discuss the issue of hypothesis generation, we should note that most experienced researchers seem to generate research ideas intuitively. They read a research paper or attend a research talk, and a great idea just pops into their heads. We suspect that this seemingly magical hypothesis generation process happens because seasoned researchers have been thinking about psychology for a very long time. They have become what K. Anders Ericsson (Ericsson & Pool, 2017) would call true experts. The problem with waiting to become a true expert in psychology before you effortlessly design a research study is that becoming an expert usually takes about eight to 10 years of intensive training (e.g., getting a bachelor's degree and then a PhD in psychology). We assume your research project in this course is due in eight to 10 weeks rather than eight to 10 years. With that in mind, how does a junior researcher *quickly* generate a good psychological research idea?

STEP 1: HYPOTHESIS GENERATION

You already got some exposure to where research hypotheses come from in the section of Chapter 2 that discussed *inductive* and *deductive* research generation techniques. In case you'd like to work with something a little easier to remember, and perhaps a little more user-friendly, let's review a simple *heuristic* (a handy set of rules) for generating research ideas. This is what we call the **IDEA heuristic**. Each letter of IDEA represents a different specific way of generating a research hypothesis. The four ways include ***i*ntegration**, ***d*issection**, ***e*xtension**, and ***a*pplication**.

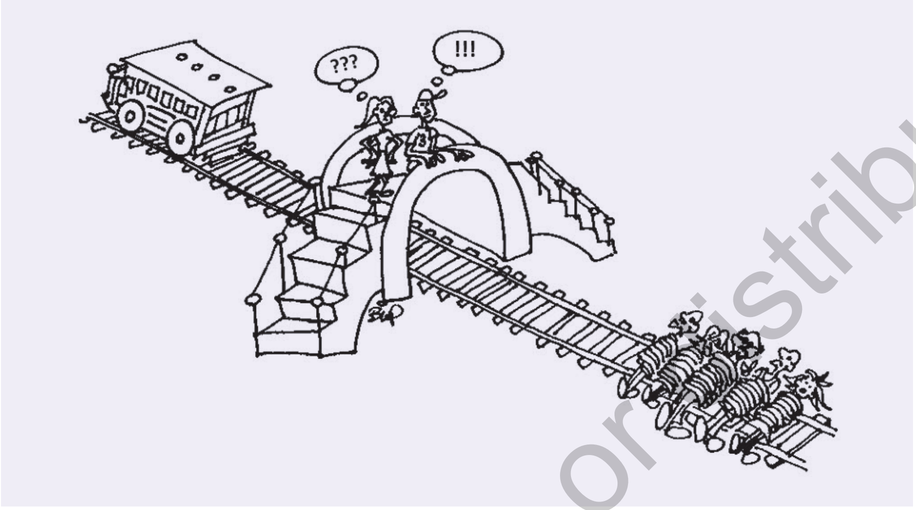
Integration

As you already know, hypotheses are derived from theories. But different theories often make different predictions. In fact, sometimes two theories make *opposing* predictions. Do people prefer predictability or novelty? Do toddlers deeply want to connect to their parents, or do they want to be independent? Some theories also fly in the face of common sense or traditional legal, moral, or economic thinking. In any of these cases, you may be able to generate a novel research hypothesis by beginning with a theory, finding, or common sense observation in one area of research, and *merging* it (thus "integration") with a theory, finding, or observation from somewhere else. For example, both common sense and the economic idea of self-interest suggest that human beings are inherently selfish. For example, the "cost-benefit" principle that is popular among economists suggests that people strive to minimize costs and maximize gains (see Pelham, 2018a, for a review). As folk rocker Todd Snider put it in the song "Easy Money," "Everybody wants the most they can possibly get for the least they can possibly do." In contrast to this idea, research on prosocial behavior suggests that under certain conditions people may behave *un*selfishly—by helping others even when it is costly to the self to do so. There are several reasons why people may not always behave selfishly. Let's consider two very similar reasons. First, as selfish as we human beings are, we also like things to be fair. In his work on *procedural justice*, Tyler (2003) found that in work settings, people will often report that they are happy getting less than others. The key is they have to believe they *deserve* less than others (e.g., because Brook works harder than you do or because Lincoln has more seniority). A similar principle known as *equity* has been demonstrated in both work and family situations. People who contribute very little to a team or a relationship do *not* usually expect to get as much out of the team or relationship as people who put in a great deal (Adams, 1963).

The list of possible ways of *integrating* the idea that we are selfish and the opposing idea that we are generous is as long as the imagination of the researcher who ponders the two opposing ideas. But let's begin with an evolutionary perspective. One of the most well-known ideas in evolutionary theory is that, to some degree, all organisms are programmed to be selfish. But experts in evolution have also argued that organisms (people included) can afford to be much nicer than usual to others who share some of their genes. Hamilton's (1964a, 1964b) concept of *kin selection* makes this prediction explicit, and kin selection has lots of empirical support. With this in mind, we could design a study in which everything was the same except that people had to behave selfishly or unselfishly with (a) those who shared genes with them or (b) those who did not. Consider a variation on the classic trolley problem.

As you can see in Figure 4.1, an evil genius (probably David Boninger), has tied five strangers to a trolley track. A runaway trolley will soon kill them all—unless you push a stranger in front of the trolley to bring it to a stop (Foot, 1967). If you knew for sure that this strategy would work, would you push the stranger? A lot of people say they couldn't bring themselves to do so—even though they know that it'd be better to see one person die than to see five die. Now consider a variation on the trolley problem that involves **altruism** (unselfish helping). This situation is identical except that no strangers are handy atop the bridge. Now you must throw *yourself* in front of the trolley to save five strangers. Would you do so? Not many people say that they would sacrifice their own life to save five strangers. But what if the five potential victims were not five strangers but your five closest family members? Or what about your five closest (genetically unrelated) friends? What about five fellow Chicago Cubs fans? By integrating what we know about human selfishness with what we know about human helpfulness, you could generate a lot of research designs.

Figure 4.1 The trolley problem. Do you push one stranger onto the tracks to save five other strangers?

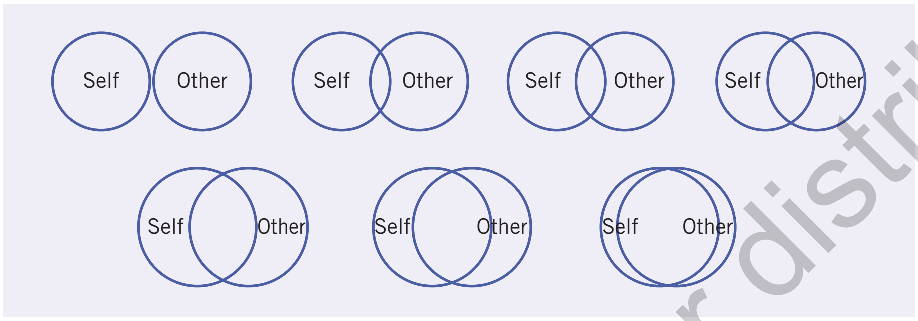


Researchers in evolutionary psychology, behavioral economics, social psychology, and personality psychology have created dozens of variations on the trolley problem.

Here's one such example. Swann and Gomez found that people who say their personal *identities* are closely *fused* to those of a large social group often say they would readily sacrifice their own lives to save five fellow group members (Swann et al., 2014). Swann and colleagues (2014) define **identity fusion** as "a visceral sense of oneness with the group." If you're wondering how you'd ever measure "oneness with the group," take a peek at Figure 4.2. The "other" in question could be your best friend, your rabbi, or (if your name is Inigo Montoya), your fellow Spaniards (a group of about 47 million). In a series of studies using variations on the trolley problem, Swann and colleagues (2014) found that Spaniards who said they were more *identity fused* with their fellow Spaniards (those who selected a highly overlapping pair of circles) were much more likely than those who were less identity fused to say they'd jump in front of the trolley to save five Spanish strangers. Sadly, none of them said they'd sacrifice their own lives to save any balding middle-aged American social psychologists.

Or consider a very different way to look at human selfishness. Consider the **ultimatum game**. In this game, one person is arbitrarily selected to be an offer-maker, and the other person is arbitrarily selected as a potential recipient. Offer-makers are given some money (say $10) and then asked to decide how to split the money between themselves and potential recipients. The recipients then have the chance to accept the offer or to reject it. If they reject it, *no one gets anything*. So if you were playing the game as a recipient, you might receive a note that says, "The offer-maker proposes to take $8 and offer you $2." Notice that you have two options. First, you may accept the

Figure 4.2 The Inclusion of Other in the Self (IOS) Scale. According to Aron, Aron, and Smollan (1992) and Swann and colleagues (2014), if you want to know if Hart truly identifies with his fellow psychologists, ask him to circle the image that best describes his relationship with his fellow psychologists (the "other").



Source: Aron et al. (1992). Reproduced with permission.

proposal and be $2 richer (making the proposer $8 richer). Second, you may reject the offer, in which case *no one gets anything*. *You* don't get the $2, and the selfish *proposer* doesn't get $8. Many subjects report feeling torn. Two dollars would be nice, but the proposer who offered a lopsided 8:2 split seems pretty greedy. So the part of you that just wants to maximize personal gains should accept the offer. But the part of you that is annoyed by extreme selfishness *in others* should *reject* the offer. Dozens of variations on this game have revealed a lot about how much people care about fairness (e.g., see Rand, Tarnita, Ohtsuki, & Nowak, 2013). As far as we know, though, no experiment has yet been conducted to pit a truly *extreme* version of simple pragmatism against the desire for fairness. So we'll have to consider it as a thought experiment.

Here's the *low stakes* version: The proposer offers you $1 and proposes to take $9. Research shows very clearly that very few people accept such a one-sided offer. Most people are happy to lose a dollar to teach the selfish proposer a lesson. Now consider a *high stakes* version: The proposer offers you $100 (out of $1,000) and proposes to take $900. Would you really pass up $100 to teach the selfish you-know-what a lesson? Good for you. You've got principles. For your information, both of your authors report that they'd do the same thing. But consider an *extremely high stakes* version: The proposer offers you $1 million out of $10 million and proposes to take $9 million for himself. If you are anything like your textbook authors, your principles just flew out the window. As much as we'd like to punish the selfish offer-maker, we'd much rather buy a beach house in Cancun. By the way, a critic of this study might argue that it would be better to pit two theories against one another rather than to pit an empirical finding against an intuition. We agree. But when it comes to integration, there is nothing wrong with following any lead that is available to see if you can erase or reverse a well-established finding.

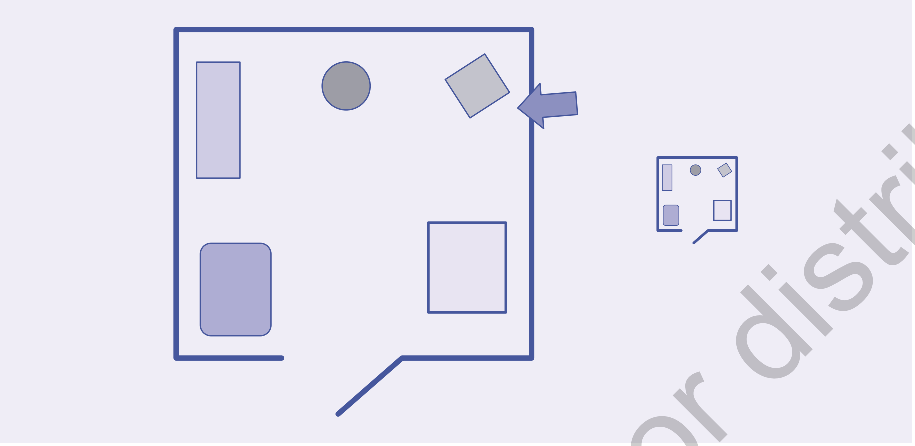
Dissection

So sometimes people generate research hypotheses that are designed to pit two opposing theories, ideas, or findings against one another. But at other times people do roughly the opposite—by trying to pull a single theory or a single empirical demonstration apart, to see exactly what it really means. In other words, sometimes people dig a little deeper than others have in the past to figure out more precisely *why* something is true. We refer to this as dissection. Dissection almost always happens when two different theories or ideas make the *same* prediction (rather than competing predictions). If you wished, you could also call dissection *de-confounding* because most often it's a tool for disentangling two very similar theories (two equally plausible accounts of why a specific research finding occurs). Let's dissect a couple of examples of methodological dissection.

A lot of research in developmental psychology examines the development of *abstract* or *symbolic* thinking. Much of human cognitive development involves going from being very concrete thinkers (who focus on the tangible things we can see, hear, and taste right now) to becoming abstract thinkers (who reason using ideas and arbitrary symbols). A classic demonstration of how we become abstract thinkers comes from work on toddlers who try to use maps or models. Most 2.5-year-old toddlers struggle to make sense of even the simplest maps. For example, take a look at the small room whose floor plan is shown in Figure 4.3. Imagine that an experimenter showed you that she was hiding a troll ("Big Terry") inside the small piece of furniture we've marked with an arrow. Now imagine that the experimenter took you to a *toy model* of this room (see the right hand portion of Figure 4.3). The experimenter tells you that she has hidden a tiny troll ("*Little* Terry") in just the same spot in the *toy* room where she had hidden Big Terry in the *real* room. Your job is simple: Find Little Terry. Now imagine searching in the little room as if you had no idea whatsoever where Little Terry might be. This might sound weird, but it's exactly what most 2.5 year olds do. They just don't seem to understand that one thing (a model of a room) can represent (can be a symbol of) something else. Only about one in five 2.5-year-olds who've been given this kind of test show clear evidence that they can use such a model. But just six months later, by the age of 3.0 years, the very large majority of older toddlers have no difficulty using the model to find Little Terry.

The traditional explanation of this striking age difference is that older toddlers can think *abstractly* in ways that younger toddlers cannot. The problem with this explanation is that older and younger toddlers differ in more than just one way. First, older toddlers have better *memories* than younger toddlers. Second, older toddlers have longer *attention spans* than younger toddlers. We have three competing

Figure 4.3 A floor plan for the real room (left), in which experimenters hid *Big* Terry, and the toy room (right), where experimenters hid *Little* Terry (adapted from DeLoache et al., 1997)

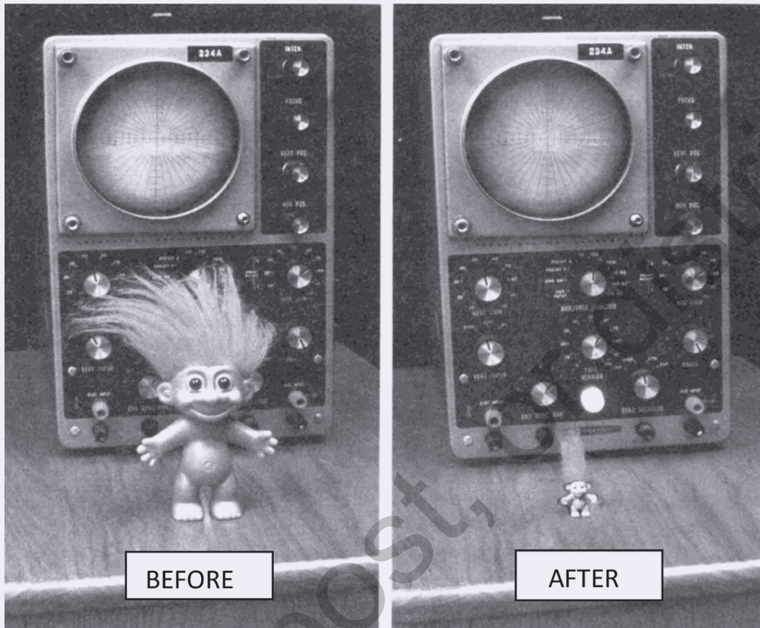


Source: Adapted from DeLoache and Rosengren (1997). Reproduced with permission.

reasons to expect that three-year-olds will outperform 2.5-year-olds in this test of abstract thinking. Only the explanation involving abstract thinking takes us very far beyond common sense.

So is there any way to *dissect* this finding? Yes, there is. In fact, the floor plans you saw in Figure 4.3 were adapted from DeLoache, Miller, and Rosengren (1997), who found that very few 2.5-year-olds but most 3.0-year-olds were able to pass the model test. To see if this age difference reflected a difference in abstract thinking *per se*, DeLoache et al. played a little trick on their participants. Specifically, they arranged it so that *some* kids did *not* have to think abstractly to find Little Terry. Figure 4.4 offers a clue about how they did so. That's right. They gave one group of 2.5-year-olds the usual version of this task, but they convinced a second group of 2.5-year-olds that they possessed a *shrinking machine*. They began by showing that the machine could shrink a troll doll. Kids had no reason to be suspicious, by the way, because their parents dutifully played along with the story. DeLoache et al. then hid Big Terry and set the machine up to shrink *the whole room*. Of course, everyone had to evacuate. A few minutes later, the kids returned to find that, sure enough, the whole room had become very tiny. Because these kids believed the model room *was* the real room (rather than merely representing it), they didn't have to think abstractly to find Little Terry. And when they didn't have to think abstractly, about 80 percent of the 2.5-year-olds had no trouble finding Little Terry. In fact, they performed as well as three-year-olds. So we can now conclude with much greater confidence than we could prior to the publication of this clever study that a form of *abstract thinking* emerges between the ages of 2.5 and 3.0.

Figure 4.4 The apparent work of the shrinking machine that DeLoache et al. (1997) used to convince some toddlers that they could shrink a troll doll (or an entire room)



Source: DeLoache et al. (1996).

Another clever example of dissection comes from classic research on judgment and decision-making. More specifically, it comes from a study of the *anchoring and adjustment heuristic* (Tversky & Kahneman, 1974). The gist of this judgmental rule of thumb is that when making guesses about frequency or magnitude (e.g., how much should that car cost? How many nations has Donald Trump pissed off this week?), we sometimes stick too close to arbitrary *anchors* (starting points) to which we were exposed prior to offering our judgments. In other words, when a person adjusts an initial guess or starting point, the amount of adjustment is usually insufficient.

In an early demonstration of judgmental anchoring, Quattrone and colleagues (1984) asked Stanford University students to estimate the average yearly temperature in San Francisco, California (which is only a few miles from the Stanford campus). However, before students provided a final answer, Quattrone and colleagues gave them a high or low *anchor*. Some students first guessed whether the average temperature in San Francisco was higher or lower than 80° F. Others first guessed whether the average temperature in San Francisco was higher or lower than 50° F. Students who began with the higher anchor estimated San Francisco to be about 10 degrees warmer than students who began with the low anchor.

Critics of this study argued that there could be *two very different reasons* why people who began with the higher anchor might offered larger final guesses about how hot it usually is in San Francisco. First, it *might* well be anchoring followed by insufficient adjustment, which was what Quattrone and colleagues believed. Second, it might be simple politeness. From a social rather than a cognitive perspective, wouldn't most reasonable judges assume that experimenters usually give them anchors because the anchors are helpful? I mean, why would he have asked me if the answer is higher or lower than 80° if 80° isn't somewhere *near* the correct answer? "Thank you for that helpful hint, Dr. Quattrone. I think I'll guess 71° F rather than 61° F. Will you be offering such helpful hints on the final exam?" Notice that critics of the study accepted the empirical finding but argued that there might be a less interesting *explanation* at work than judgmental anchoring. How do we know that this research finding occurred because of imperfect reasoning rather than the rules of friendly conversation? Quattrone et al. (1984) surgically dissected these two explanations by replicating the study in a separate group of Stanford students who received anchors that could not have possibly been construed as helpful. In this version of the study, the high anchor was a *ridiculously* high 558° F. The patently ridiculous high and low anchors yielded an anchoring effect that was almost identical in size to the original effect. That's clever. That's also dissection.

Dissection can also be a solution to the problems of **scenario studies**. Scenarios studies are studies that manipulate a variable by asking some people what they'd do in one particular situation and asking another group of people what they'd do in a *variation* on that situation. A major drawback of scenario studies is that people often have no idea what they'd really do in a particular situation unless they have been in that particular situation before. We hope none of you have ever *really* had to decide whether to kill one stranger to save five others. Another drawback of scenario studies is that people sometimes know *exactly* what they would do but are reluctant to admit it—especially if the behavior in question is highly sensitive. Imagine a variation on the trolley problem designed to assess racism. Would you sacrifice a White male teenager to save the lives of five White male adults? What about sacrificing a Black male teenager to save the lives of five White male adults? There is no shortage of data showing that many Americans treat Black and White people very differently. But anyone who wishes to use the trolley problem to study racism would face the problem of **socially desirable responding**. This is the finding that many people often *report* that they would do whatever is most socially appropriate. But in real life, many people engage in less than wonderful behavior.

There are two ways to apply the scalpel of dissection to scenario studies. One of the ways is twofold. First, make sure you only manipulate your independent variable (e.g., the ethnicity of a target) on a between-subjects basis. In this way, your manipulation will become a lot less transparent. Second, do anything you can to *disguise* the manipulation or to distract your participants from your true research hypothesis. For example, sticking with the trolley problem, suppose you said nothing about ethnicity, that you emphasized the age of the to-be-pushed person (e.g., push a *teenager* to save five *adults*), and that you subtly manipulated ethnicity by making the to-be-pushed target in question appear dark-skinned or White (in a pretty realistic drawing). You might also make sure that one of the five strangers in need of saving was dark-skinned. Collectively, these precautions would probably reduce some of the most obvious concerns about social desirability. But in our view, race is such a hot-button issue that you'd still have some major worries.

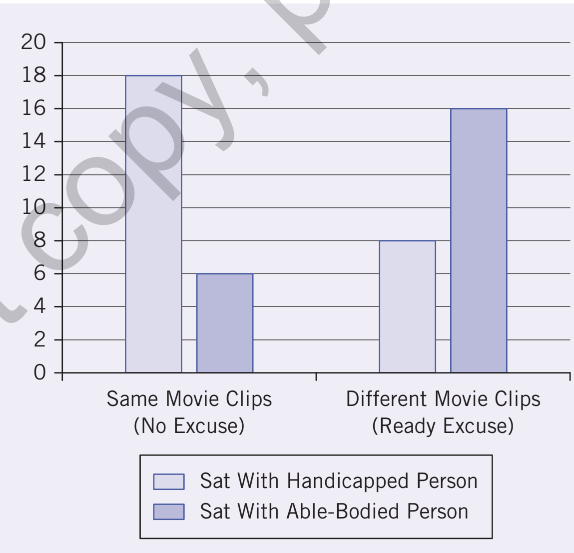
A second way to dissect social desirability concerns and variables such as altruism or racism is to look at real rather than hypothetical behavior. Consider a study by Whitehouse and colleagues. Recall that Swann and colleagues (2014) showed that Spaniards who said their identities were fused with those of their fellow Spaniards were more likely than un-fused Spaniards to say they would throw themselves in front of a trolley to save five Spaniards. Talk is cheap, you might say. And you *might* be right. But if you think all trolley studies have us headed down the wrong track, consider a follow-up study that included real rather than hypothetical sacrifices. In a study of Libyan rebels fighting against the Gaddafi regime, Whitehouse, McQuinn, Buhrmester, and Swann (2014) found that front-line fighters in a very real war reported extremely high levels of *identity-fusion* with their fellow battalion members. In contrast, those who chose to serve in less risky support roles reported more modest levels of identity fusion. People who are truly identity-fused are often willing to back up their beliefs by putting themselves directly in harm's way. One way to dissect a scenario study, then, is to move from a scenario to a real situation.

As a final example of dissection, consider a clever lab study by Snyder, Kleck, Strenta, and Mentzer (1979). Snyder et al. wanted to study *discrimination* (in their case, avoidance) against people with a physical handicap. In their studies, the handicapped person people could potentially avoid was always a person with a leg brace who walked with crutches. In the language we're using here, they wanted to *dissect* people's true wishes from social desirability biases. What the researchers did *not* do was to ask people "Hey, would you usually avoid sitting next to a stranger with a physical disability?" Instead, in two experiments, they pretended to be studying movie preferences. For this reason, participants learned, they would be watching some comedy scenes from the silent movie era and offering their personal ratings of the movie clips. Participants always had a choice of watching the movie clips in one small room or another. In fact, participants could plainly see both of the two small rooms, and they could even see that exactly one fellow participant was already seated in front of each viewing screen in each room. Each of the fake fellow participants was dutifully filling out some paperwork, apparently in anticipation of seeing the film clips.

That was the mundane part. Here's the clever part. Participants were always told that the project focused on two different kinds of film clips. In fact, half the participants were told that they should choose which genre of silent films clips they would like to watch: either "slapstick" or "sad clowns." Recall that there was a different **confederate** (fake participant) in each room. And notice that, because participants' explicit instructions were to choose a *movie genre* (not an able-bodied versus a physically disabled viewing partner), they had a good excuse if anyone were to accuse them of avoiding the handicapped partner ("But this is a movie study, and I just wanted to see those hilarious silent clowns."). To avoid any possible confusion, by the way, there was a clear label with the film clip topics on the main table in each room.

But for other half of the participants in these studies, no such socially acceptable excuse was available. These participants learned (falsely, of course) that the researchers were having some technical difficulties—which meant that only one genre of film clip would be available that day. *These* participants had to choose between sitting next to a handicapped person to watch the film clips and sitting next to a person who was not handicapped—to watch exactly the *same* film clips. Figure 4.5 summarizes the total results across two variations on this study. As shown in the left hand

Figure 4.5 Seating preferences revealed that people only avoided a physically handicapped stranger when there was a good ("face-saving") excuse to do so. Otherwise they behaved as if they preferred the physically handicapped stranger. Results aggregated from Study 1 and Study 2 of Snyder et al. (1979).



Source: Adapted form Synder (1979).

portion of Figure 4.5, in the *same movie* condition, participants were three times as likely to sit with the handicapped person as to sit with the able-bodied person. These data alone might have suggested to some that people have a strong preference for hanging with handicapped strangers. But the right-hand half of Figure 4.5 tells a very different story. When participants had a good excuse for avoiding the presumably handicapped fellow participant, they readily did so. In fact, they were exactly twice as likely to sit with the able-bodied confederate as to sit with the physically handicapped confederate.

Of course, the exact nature of dissection will vary from topic to topic. But the logic is always the same. Come up with a variation on the original independent or dependent variables of interest (e.g., age, anchoring, altruism, attitude change, aversion) that will yield an effect according to one causal account but won't do so (or will yield the opposite effect) according to a competing causal account. Then collect your data and hope to find out which account is correct.

Extension

If neither integration nor dissection is your cup of tea, you don't have to go around thirsty. There are at least two other ways to generate psychological research hypotheses. A third way to do so is called extension, and we're happy to say that if you are familiar with the OOPS! heuristic, you're already pretty familiar with extension. The logic of extension is to test the limits of an established theory or hypothesis by, well, *extending* it—by considering the four aspects of the OOPS! heuristic.

You might thus ask whether a specific hypothesis would still hold true if you selected a different operationalization of a variable (or variables) in a study you wish to extend. The specific way in which researchers operationalize "aggression," for example, has a huge effect on whether they observe gender differences in aggression. Most people think boys and men are simply more aggressive than girls and women. After all, men murder other people, especially strangers, *much* more often than women do (Kellermann & Mercy, 1992). Likewise only about 6 percent of U.S. bank robbers are female. In both the United Kingdom and the United States in 2003, men arrested for robbery of any kind outnumbered women about eight to one (Brookman, Mullins, Bennett, & Wright, 2007). But men are only slightly more likely than women are to kill people they know very well. In fact, when it comes to killing those we love with knives or blades, wives in the United States stab their husbands to death about 50 percent more often than husbands do the reverse (Kellermann & Mercy, 1992). Further, in contrast to most studies of physical violence, many studies suggest that when it comes to *verbal* aggression (e.g., calling a person a hurtful name), women and girls are about as aggressive as men and boys. Finally, when it comes to *relational* aggression (e.g., ostracizing a person, doing something to harm person's relationship with someone else), girls and women appear to be at least as aggressive as, if not more aggressive than, boys and men (Ostrov, Kamper, Hart, Godleski, & Blakely-McClure, 2014). Don't tell anyone we said that, by the way, or we won't be your friend.

You might also extend a past finding by tinkering with occasions (i.e., with time). For example, people sometimes need to update measures to reflect generational changes in how people speak or in what is considered socially acceptable. Long ago, Katz and Braly (1933) published a classic paper on stereotypes in which they asked college students which traits they felt described ten social groups. The social group *labels* Katz and Braly used in their study included "Jews," "English," "Irish," and "Negroes." Further, the list did *not* include "Hispanics" or "Latinos" (who were a much smaller percentage of the U.S. population then than now). Since 1933, though, most Americans have stopped giving much thought to whether a "White" person is "Irish" versus "English." Further, the dated term "Negro" has been replaced by more modern terms such as "Black" or "African-American." Of course, there are much more substantive ways in which one might apply the issue of occasions to extend a study. Are college students today more narcissistic than students of 40 years ago? Are parents today less likely to give their kids common first names than were parents many decades ago? Would people today be as obedient to authority as were the people Stanley Milgram studied in the early 1960s? The answer to all of these questions, by the way, appears to be yes.

Perhaps the most common way in which people are likely to extend a study involves identifying and studying populations that have received little or no past research attention. A lot of work in ethnic studies, gender studies, and cross-cultural psychology has to do with how different populations think, feel, and behave. Of course, people in different populations are also treated differently by others. Research on prejudice, stereotyping and discrimination examines when and why this is true. Sometimes being a member of one population or another can serve as the independent variable in a study. For example, research on the "second shift" shows that in many U.S. families, women who work full-time are often expected by their husbands to begin a second shift of work (e.g., cooking and cleaning) as soon as they get home (Hochschild & Machung, 1989). Wives only rarely hold such unrealistic expectations of their husbands. Along somewhat similar lines, research on depressed entitlement explicitly makes gender the independent variable. Men and women are asked to perform some kind of work, and after doing so, they are asked to evaluate their own work—or to indicate how much money they should be paid for their own work. Studies consistently show that women evaluate their work less favorably and pay themselves less money for the same work than men do (Major, McFarlin, & Gagnon, 1984; Pelham & Hetts, 2001). Both work on the second shift and work on dressed entitlement suggest that men and women may unknowingly internalize traditional ideas about the value of "male" versus "female" work. If you were to apply the concept of population to *extend* this work, you might assess whether the second shift phenomenon is stronger than usual in populations with more traditional gender norms (e.g., among fundamentalist Christians, in Sierra

Leone rather than Canada) or weaker than usual in populations where women enjoy relative equality with men (e.g., among female PhDs, in Sweden rather than Canada).

One can apply the concept of *extension* to a new population at many different levels to generate new research ideas. The first level is to see if a simple observation made in one population is equally true in another. Are older Americans more religious than younger Americans? In our view, such descriptive work is important, but it is even better if you can come up with a research hypothesis that will allow you to test a specific theory. Suppose you believe that older Americans are more religious than younger Americans *because* they know that death is just around the corner. If so, it would be a good start to show that older people are more religious than younger people. But you'd surely want to go further. You might conduct an experiment to see if young people report more religious attitudes than usual if they've recently been thinking about their own deaths. In research on *terror management theory* (the idea that people work very hard to minimize their fears of death), this is exactly what happens (Norenzayan & Hansen, 2006). Research also shows that if you examine differences in religiosity levels in 150 nations across the globe, infant and adult mortality rates are very strong predictors of how religious people are in a given country. This is true even after statistically controlling for the fact that poorer nations tend to be more religious than rich nations (Pelham, 2018b).

A second level of extension involving populations is to see if an *association* (rather than a mean score or percentage) known to exist in one population varies in a different population. For example, among Christians in the United States, women tend to be somewhat more religious than men. That's an *association* between gender and religiosity. Women also report praying and attending religious services more often than men do. But among Muslims and Orthodox Jews, this association is different, with men generally reporting that they are more religious and saying that they engage in more religious behavior (Pew Research Center, 2016). Likewise, older people in the United States are often stereotyped as narrow minded and muddle headed relative to young people. In contrast, in many other parts of the earth, older people are respected for their wisdom and their ability to solve problems that vex younger people.

In some cases, it is highly informative that associations observed in one population occur pretty consistently (rather than varying dramatically) in other populations. Schmitt (2005) conducted a 48-nation study of gender differences in people's attitudes about sexual promiscuity. There were large cultural differences in whether people felt casual sex was OK. For example, people in Finland were much more comfortable with the idea of casual sex than people in Taiwan. However, the effect of gender was much larger than the effect of nation, and in fact there was no nation Schmitt studied where the average man didn't report being substantially more comfortable with casual sex than the average woman was. Schmitt (2005, p. 265) concluded that "culture has an important influence on sociosexuality, but biological sex is the larger and stronger predictor of human mating strategies across the nations of the ISDP [i.e., across all 48 nations studied].” Not everyone has agreed with this conclusion, but few would argue that Schmitt’s global study of nations from Argentina to Zimbabwe wasn’t ambitious. It was certainly *extensive*.

As we hope you recall, the fourth part of the OOPS! heuristic is situations. In our view, researchers probably do not pay enough attention to the specific situations in which research occurs. The core social psychological principle known as the “*power of the situation*” (aka the “pervasiveness of social influence”; Smith & Mackie, 2007) dictates that situations have a huge impact on how people behave. Everyone knows that children are better behaved than usual when their parents are watching—and that their parents are better behaved than usual when state troopers are watching. But the power of the situation goes well beyond these obvious examples. Consider research on *self-stereotpying*. By the time you are a young adult, you’d think you’d have a pretty good idea how feminine you are. But research on self-stereotyping shows that we often change our beliefs about ourselves from one audience to the next. For example, Sinclair, Pappas, and Lun (2009) had female college students engage in an interaction with a man who appeared to share their birthday (to make the women like the man). When women believed that this man held traditional gender attitudes (believing that women “should be cherished and protected by men”), they reported possessing more feminine personality traits than usual (e.g., *sensitive*, *sweet*). You can imagine, then, that if you were measuring people’s self-evaluations, their stereotypes of others, or even how helpful they were to a needy stranger, the exact situation in which you conducted your study (including the age, gender, ethnicity, and style of dress of the experimenter) could have a big impact on what you observed.

In light of findings such as these, Hardin (personal communication, January 5, 2018) has argued that a big mistake that even the most seasoned researchers often make in empirical research papers is to fail to provide essential details about the exact situation in which data collection took place. Here’s how he put it:

> A common problem in empirical papers is an extreme dearth of details regarding the specific situations that participants experienced. From beginning to end, the nature of the all-important situation in which a study took place is usually a mystery. How were participants recruited and approached? What exactly was said to them? What expectations and instructions were participants given at the beginning of the study? How were transitions across tasks handled? Who did participants think the experimenters were? Did they appear to be seasoned experts or clerks dutifully tabulating data points? Did experimenters conform to gender norms for dress and assertiveness? Were they serious, or were they lighthearted? With what groups did the experimenters seem to be affiliated? Did participants really need to focus and pay attention, or was it actually better for them not to give a damn? Researchers should pay careful attention to these

aspects of data collection and make thoughtful choices about them. And when these choices were likely to have been important to their findings, researchers should have reported the details of these choices in their Methods sections.

I think the primary problem with replication in psychology (sometimes Fletcher didn't get the same results Jefferson did) is that we don't know the effective conditions that produce a given effect. Methods sections should be longer; introduction and discussion sections should be shorter.

Application

A final way to generate a novel research idea is arguably a lot easier than the first three, and some would say it's the most important. This fourth way is application. Researchers who conduct **applied research** begin with a basic research finding (e.g., people evolved to prefer sweet and fatty foods, people are not good at multitasking, people often stereotype women as fragile and in need of protection). They then ask themselves whether there is a real-world problem (e.g., obesity, motor vehicle accidents, high rates of unnecessary cesarean sections) that can be better understood through the lens of the basic research finding in question. For example, a great deal of research in cognitive psychology does indeed show that people are not nearly as good at multitasking as they think they are. If you'd like to see a great hands-on demonstration of this fact, just Google "thinker divided attention task" or go directly to https://cat.xula.edu/thinker/perception/attention/divided. *Please* take that two-minute test before you read any further. Done? We're pretty confident that you did *not* get 6 out of 6 on the six-item spelling test and *also* correctly guess the exact shape traced by the little red dot that was moving just to the left of the spelling test. Even if you are a very good speller, focusing enough of your visual attention on the spelling test to get a perfect 6/6 uses up too much of your visual attention to allow you to track the exact shape made by the red dot. Likewise, if you were able to track the dot well enough to know for sure that it drew the state of Texas, you almost certainly missed at least a few of the six spelling words.

But people often *think* they can multitask when they really can't. Why else would people *ever* text and drive? If you look at a text for just 2.0 seconds when driving 60 mph on the freeway, this means you drove more than half the length of a football field while paying no attention to the road! Applied psychologists often begin with basic research findings such as this one (people have little idea how much they suck at multitasking)—and then *apply* them to a real-world problem (such as the dangers of texting or talking on a cell phone while driving). For example, Strayer and Johnston (2001) showed that talking on a cell phone greatly increases the likelihood that drivers will fail to see a traffic sign or get into an accident while driving a realistic computer-simulated car (a high-tech virtual car that simulates almost every aspect of real driving). This was true, by the way, even when participants talked hands-free while trying to drive. In contrast, listening to the radio—or even listening to a book on tape—had no such effects. Presumably this is because these activities do not distract people nearly as much as a two-way conversation does. Such effects have been replicated many times. For example, Stavrinos et al. (2013) showed that to compensate for distracted driving, people often slow down when they would not otherwise have needed to do so and fail to change lanes when they otherwise should have. Such behavior contributes to traffic jams and creates unsafe conditions for others drivers, especially in heavy traffic. If you are wondering if effects like these hold up in a real-world setting, perhaps you should become an applied psychologist yourself. Exactly what questions would you like to ask real drivers (confidentially, of course) about their day-to-day "distracted driving"? How would you assess how safe or how experienced drivers are? How would you measure accident rates? Would you need to control for hours spent behind the wheel? Would it be important to control for age and gender? What about local traffic density?

The list of topics studied by applied psychologists doesn't stop with driving. Almost no real-world behavior is safe from the clutches of applied psychologists who are constantly plotting ways to make people happier, healthier, safer, or more successful. Consider research on exercise. Applied research shows that regular aerobic exercise can prevent premature aging and even increase *brain plasticity* (your brain's ability to adapt to changes by learning and rewiring itself; Molteni, Zheng, Ying, Gómez-Pinilla, & Twiss, 2004; Tomporowski, 2003). So if you want to increase your chances of staying mentally keen your whole life, put away that Scrabble game and put on your sneakers. Likewise, if you want to live longer, applied psychologists would be the first to tell you to "just keep swimming." Biking, running, walking, or playing pick-up soccer will also do the trick. Lifting weights will help you get stronger, by the way, but it does little to increase your cardiovascular fitness. Even if you didn't know about the connection between exercise and brain plasticity, you surely knew that aerobic exercise is good for you. The problem for most people is *adherence* or *commitment* (sticking to an exercise program).

Applied psychologists study that, too. Abby King and colleagues (2007) showed that simply giving middle-aged walkers an encouraging phone call about once per month roughly doubled how much they walked over the course of a year. An encouraging reminder from a computer was almost as effective as a live call from a real person at increasing walking mileage. Having a live person who agrees to work out with you on a regular basis can be even more important. But there are subtler ways in which others can help. Carron, Hausenblas, and Mack (1996) conducted a systematic review of many different exercise intervention studies. They found that having family members who support your intention to keep exercising, which we'd like to dub "*fitness social support*," is strongly associated with sticking to an exercise plan.

In the past few decades, there has been tremendous growth in applied psychological research. Thousands of studies in health psychology, sports psychology, industrial/organizational psychology, political psychology, and clinical psychology qualify as good examples. Likewise, some researchers who study close relationships have gone beyond merely predicting who's attracted to whom or who's most likely to get a divorce, to offering couples practical tips for enjoying happier and healthier relationships (e.g., see Gottman & Silver, 2015). One of the fastest growing areas of applied research examines people's attitudes about climate change (see Maibach & van der Linden, 2016). Most of the applied research in most of these areas began with a researcher who translated a basic research finding into an applied study that examined the implications of this basic finding for an important social problem.

Is the IDEA System Exhaustive?

When we have run the IDEA system by experts in research methods, some of them have pointed out that there are other important ways to generate research hypotheses. Recall that Bill McGuire suggested a long list of specific routes to generating research hypotheses. However, one of our expert reviewers, D. S. Boninger (personal communication, January 2018), suggested an additional approach to generating research ideas that even McGuire did not mention.

It is this: *Be a careful observer.* Pay attention to human behavior, from the routine ways in which we harm others to the delightful ways in which we help and support others. Have you noticed that there are things people do in a group that they would never do alone? How much of our communication is verbal and how much is nonverbal? Do people become more religious when they have children? If so, why? Why does anyone ever bother to vote (when any one vote almost never matters)? Paradoxes and opportunities to test new ideas about human behavior are everywhere. You just have to be on the lookout for them.

STEP 2: OPERATIONALIZATION (DESIGN)

So you think you have a good research idea? Or maybe you know painfully well that you don't, but you expect to have one pretty soon now that you know about the IDEA system. Once you do have a research idea, you'll have to translate the idea into a testable hypothesis. We've already discussed operationalization a few times in this text. Operationalization makes the otherwise *un*observable observable. No one can perform statistical tests on an idea; real research requires data. In this section of the chapter, we offer some practical advice about how to come up with good operational definitions of psychological constructs. Some of this advice may seem a little obvious, but in our experience, even seasoned researchers sometimes miss the obvious when it comes to operationalization.

Read the Literature

Our first piece of advice about operationalization is probably as old as the wheel. It is this: *Don't reinvent the wheel.* Because psychology has been around for about 140 years, a *lot* of people have already spent a *lot* of time coming up with tried and true operational definitions of many, many psychological constructs. So don't spend weeks coming up with your own operational definition of "self-esteem" only to learn that a guy named Morris Rosenberg already did all of that work back in the 1960s—and published a book called *Society and the Adolescent Self-Image* in 1965. That book had been cited by other researchers about 33,000 times when we last checked. Because we both study the self-concept, we happen to know that many people who cite that book do so because it contains Rosenberg's popular self-esteem measure. Rosenberg's face-valid 10-item self-esteem measure (e.g., "I feel that I have a number of good qualities," "On the whole, I am satisfied with myself.") is by far the most widely used self-esteem measure ever developed. This doesn't absolutely guarantee that you should use it, by the way. Tafarodi and Swann (1995) argued that the Rosenberg self-esteem measure doesn't allow us to discriminate easily between "self-liking" and "self-competence." In keeping with the principle of dissection, they thus developed a self-esteem measure that neatly separates these two aspects of self-esteem. But unless the point of your research has to do with separating self-liking and self-competence, you'd probably get less pushback from reviewers is you used Rosenberg's measure.

We'd like to be clear that psychological science is not a popularity contest. We are *not* advocating the use of Rosenberg's self-esteem scale because it's popular. We're advocating it (or any other well-validated measure of anything) because careful work has already been done on it. In fact, so much work had been done on Rosenberg's self-esteem scale by the mid-1990s that Tafarodi and Swann (1995) were able to publish their highly cited paper on the two components of self-esteem—at least in part—because they had a good *criticism* of the famous Rosenberg self-esteem measure.

Not reinventing the wheel includes finding out exactly how wheels work—and how they came to exist in the first place. If you read some key empirical papers on the research topic of your choice, you should learn some valuable lessons about what has been done on a topic, what questions remains unanswered, and what techniques experts suggest for answering those unanswered questions. For example, in the Stavrinos et al. (2013) paper on the perils of texting and driving, the authors suggest a fruitful area for future research. Stavrinos et al. (2013, p. 68) noted that "while driving simulators provide much needed experimental control to test hypotheses with regard to traffic flow theory, it is difficult to truly ascertain the degree to which simulated driving performance maps on to real world driving behavior." This is a clear call for survey studies in which the driving outcomes are real rather than computer simulated. Virtually every empirical research paper ever published contains a discussion section, and in most of these discussion sections you can find thoughtful suggestions for future research.

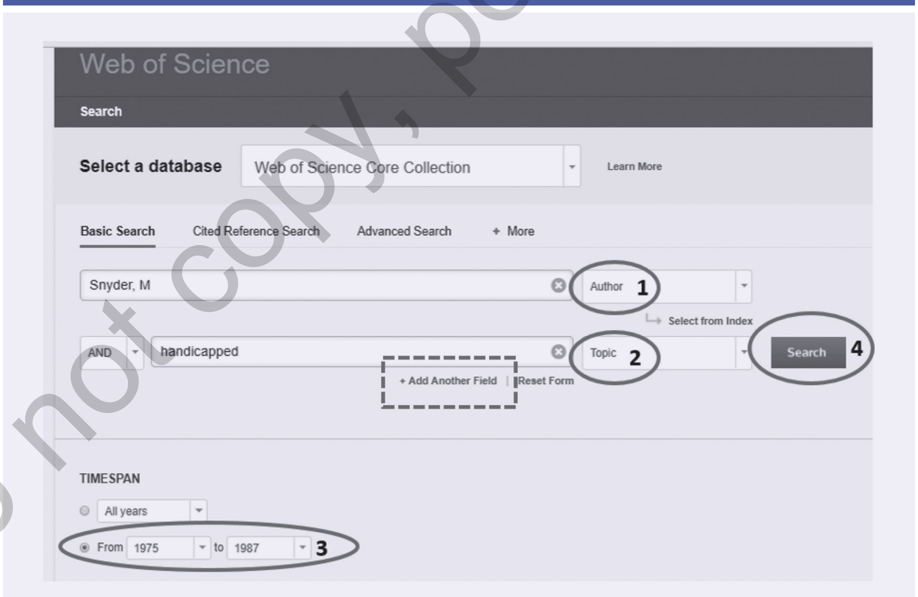
Many people's first effort to learn about a new topic is to Google it. This isn't *always* a big mistake. For example, when the first author Googled "self-esteem measure" in January of 2018, the first few hits proved to be exactly what a beginner should have hoped for. They included a reference to a well-validated "collective self-esteem measure," a newer but increasingly popular "single-item self-esteem measure," and a Wikipedia page and pdf file that each included the details of the popular Rosenberg (1965) self-esteem measure. But as tempting as it is to rely on Google because it is so familiar, there are some big problems with many people's Google habits. In an article in *Education Week*, Wineburg and McGrew (2016) report the distressing results of a series of studies showing that Google searches can get people into a lot of trouble. In one study, Wineburg and McGrew gave 25 Stanford students 10 minutes to compare an entry they Googled from the (a) *American Academy of Pediatrics* with an entry they Googled from (b) the *American College of Pediatricians.* Students could do anything they wanted (including Googling what other sites said about these two sites) to find out what they thought about the entries at the two sites. Just over half the students concluded that the entry from the American College of Pediatricians was "more reliable." Maybe this isn't too surprising. Both groups *sound* reliable enough. I mean, both sites have names that include words like "American" and "Pediatrician"–plus (as Dory would put it) "something about the tentacles." The problem is that the site that the Stanford Googlers found a bit more reliable is a splinter group who broke away from the American Academy of Pediatrics in 2002. The splinter group believes it is wrong for same-sex couples to be able to adopt children. The Southern Poverty Law Center considers them a hate group because they equate homosexuality with pedophilia. If students had been given more time to compare the two sites, they would presumably have discovered the difference between the two sites. But the unfortunate fact is that as much as your first author loves Google, it is an informal rather than a formal way to find things out.

Your first author is also a big fan of Wikipedia, by the way. Whereas most scientists consider Wikipedia the kind of thing that only a middle schooler would use as a scholarly reference, your first author can get lost in Wikipedia entries—learning about topics as diverse as LeBron James and the King James Bible. In our view, Wikipedia does often provide concise topical research summaries for beginners. If you wish to use Wikipedia, though, the key to following good scientific practice would be tracking down and carefully reading the *original research papers* cited in a Wikipedia article—and then deciding for yourself what you think they mean. Our metaphor for using Google and Wikipedia is that doing so is like asking a helpful stranger for directions. Just because the stranger isn't a cartographer doesn't mean you shouldn't trust her to point you in the right direction. But precisely because the stranger isn't a cartographer, you shouldn't rely on her when it is essential that you get something right.

If we may put in a word for a more respectable research tool, we have found the "Web of Science" (a scientific search tool to which most university libraries have a subscription) to be *extremely* useful. This search tool contains citations, abstracts, and often pdfs of complete papers for millions of scientific publications. The Web of Science search tool is also pretty intuitive. In the screen capture you see in Figure 4.6, the first author tracked down a pdf of the classic Snyder et al. (1979) paper on discrimination against people with handicaps—after recalling only that Mel Snyder was an author and knowing only roughly when the paper was published. The numbered ovals in Figure 4.6 offer you some hints about where to click to get Web of Science to perform a typical search. The dashed rectangle, by the way, shows that if you wanted to add an additional search field that was not included in the default options, you'd just click "Add Another Field" to create another search field of your choice. For example, if you wanted to search only in a particular journal, you could make "Publication Name" a new search field and enter any publication name (or any portion thereof), which would further limit your search.

This tool is both very powerful and very flexible. For example, you could use the tool to find anything done by anyone on the topic of "altruism" in the years 2014–2018. To do so, you'd merely reset the form to the default, change "Author" to "Topic," and specify 2014 and 2018 as the low and high ends of the time span. In mid-January of

Figure 4.6 A screen capture from the Web of Science search tool. I entered "Snyder, M" as the "Author," entered "handicapped" as the "Topic," and limited the "TIME SPAN" to 1975–1987. I then got pretty lucky. The paper I desired was the only hit that showed up after I clicked the "Search" button.



Source: Web of Science (2018). Basic Search Tool [Screenshot]. Retrieved from http://apps.webofknowledge.com

2018, doing so would have generated 3,063 results. Needless to say, you'd want to refine that search a bit if you were beginning a new research project. To do this, you could peruse the options on the left hand side of the *results* page and then limit the results to a particular year, to a particular type of paper (e.g., empirical article versus book chapter), or to any of several other subcategories.

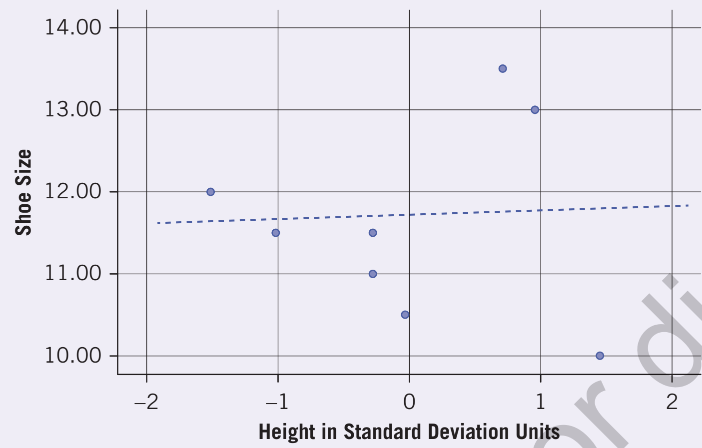
Consult an Expert

If you are lucky enough to be taking a class on research methods from a professor who shares some of your own research interests or if you are flexible enough to focus on areas in which your professor is an expert, you should also be able to take advantage of your professor's knowledge. For example, if you were to ask the second author of this text to name five "must-read" papers on "implicit social cognition" or "unconscious bias," he'd quickly rattle off a few important papers. Furthermore, he'd be able to tell you in exactly what journals each of the papers had appeared and in exactly what year. If you posed the same question to the first author, he'd recommend a somewhat different set of papers, because your authors often have spirited debates about the validity of implicit measurement techniques. But almost any of the papers either of us would recommend should help you decide for yourself what *you* think about trying to measure things such as implicit self-esteem, implicit racial biases, or implicit gender stereotypes. Likewise, if your professor is an expert in childhood social development, behavioral neuroscience, or psycholinguistics, this person would be in a much better position than either of us would be to offer you cutting-edge advice about what to read on those important topics. If you are wondering if it would be an even better idea to track down an expert at a different university who is an expert on the *exact* research topic of your choice, the unfortunate truth is that most experts are busy enough that they barely have time to answer questions for their own students. Emails from strangers often take a back seat to more pressing concerns.

Maximize Statistical Power

Statistical power refers to a study's ability to detect a true effect. Many inexperienced researchers believe that if your hypothesis is correct, your survey or experiment will surely confirm it. However, there are some big problems with this assumption. Two of the biggest problems are that you might fail to observe statistically significant support even for a true hypothesis (a) because your sample size is too small or (b) because you ran into some bad luck when sampling participants. Consider the data on height and shoe size from the eight people listed in Figure 4.7, all of whom are friends or relatives of the first author (so they would best be considered a *highly* convenient *convenience sample*). In these data, the **correlation** between height and shoe size is virtually zero.

Figure 4.7 Correlation between height and shoe size in a very small convenience sample



(You may recall from an introductory psychology course that a true correlation of zero between two variables means that they are completely unrelated.)

This is true in this very small sample mainly because (a) the shortest person sampled happens to have very big feet (especially for his height) while (b) the tallest person sampled happens to have small feet (especially for his height). We hope you already knew that in the U.S population, there is a real and substantial positive correlation between height and shoe size. But the combination of a very small sample and a bit of bad luck meant that this study did not uncover the truth about height and shoe size.

This means that if you wish to maximize your chances of observing a real effect, you should do whatever you can to get the *biggest sample size possible*. As most researchers could tell you, there are analyses you can conduct to calculate statistical power. But such analyses, as useful as they are, are beyond the scope of this introductory text. So our advice for beginners is to take a couple of simple steps to increase statistical power. The first step is to try to gain access to a large sample.

A second step toward maximizing statistical power is to create dependent variables (and sometimes independent variables) that are **continuous** rather than **categorical**. Statistically speaking, a continuous variable can take on many values, from very low to very high. Height, weight, age, IQ, and yearly income are all good examples of continuous variables. In contrast, whether a person is retired, whether a person voted for Donald Trump, and the type of car a person owns (e.g., sedan, truck, or SUV) are all categorical variables—because they can take on only discrete values that are qualitative rather than quantitative. Of course, many variables are *inherently* categorical or continuous (pregnancy is categorical; the number of episodes of *Rick and Morty* one has watched is continuous). But researchers often have a choice about whether they create categorical or continuous variables in their own research. We will go into great detail later in this text about how to create scales with good measurement properties. For now, we suggest you adopt continuous measures whenever possible. Continuous measures are more sensitive than categorical measures, and this usually increases statistical power. Thus, rather than asking participants if they are hungry or not or if they like the music of Beyoncé or not, it would be wiser to ask people one of the two questions you see in Table 4.1.

Because there are seven possible answer options to each of these questions (rather than merely two), these questions would almost certainly offer researchers a lot more statistical power than categorical versions of the same questions would. In addition, as you will learn later, the two model questions in Table 4.1 also make use of **EGWA** scales (empirically grounded, well-anchored scales), which have some highly desirable measurement properties that few other scales have.

Keep It Simple

Some ambitious new researchers think they have to make their studies very complicated. But very often the cleverest studies are the simplest, and this is particularly true for experiments—especially new experiments. Imagine you're studying the effects of taking over-the-counter pain relievers on how painful people say it is to relive their "worst high school memory." This might sound whacky, but Eisenberger, Lieberman, and Williams (2003) used modern brain imaging technology (fMRI) to show that when a person gets socially rejected while playing a computer game, the areas of the brain that get activated by the social pain are exactly the same brain areas that get activated when a person experiences *physical* pain. Following up on this finding, DeWall

Table 4.1 Two Examples of Continuous Dependent Measures

1. How hungry are you feeling right now? (Circle one number.)						
0	1	2	3	4	5	6
not at all hungry		slightly hungry		quite a bit hungry		extremely hungry
2. How much do you like the music of Beyoncé? (Circle one number.)						
−3	−2	−1	0	1	2	3
dislike extremely	dislike quite a bit	dislike slightly	neither like nor dislike	like slightly	like quite a bit	like extremely

and colleagues (2010) showed that taking a standard dose of acetaminophen twice a day for three weeks reduced people's feelings of social rejection in an experimental game. Physiologically, social pain is very much like physical pain. Apparently for that reason, at least one product that reduces physical pain (acetaminophen) also seems to reduce social pain.

Imagine that you wanted to extend this finding in your study of reliving a painful memory. You might figure that the best way to do the new study would be to use four or five different dosages of the drug. After all, isn't a "continuous" independent variable better than a dichotomous one? Not in this case. In this case, you'd almost certainly have to conduct your experiment on a pretty small convenience sample. If you did so, you'd be better off just including the two extreme levels of your manipulation. This might put 20 people into each of your two conditions rather than six to nine people in each of your five conditions. So our advice would be to give some participants a 0 mg placebo and others a 500 mg "standard dose" of acetaminophen. (If you're unfamiliar with placebos, don't worry; we'll officially define them later in this chapter.) Unless you care deeply about the subtle differences between low to medium doses of a drug (or low to medium levels of humor, anger, or test anxiety), you will usually maximize the power and validity of an initial experiment by keeping it simple and treating your two experimental groups as differently as you can practically and ethically manage.

A supplemental advantage of keeping your research design simple is that this usually means any necessary statistical analyses of your findings will also be much simpler than they would be with a more complex design. The difference between being able to conduct a simple *t*-test and a more complex analysis of variance (*ANOVA)* is usually the number of conditions you have in an experiment. There is one pretty clear exception to this rule, by the way. If both your independent variable and your dependent variable are continuous, you can almost always calculate a simple correlation coefficient (Pearson's *r*). However, even the most sophisticated readers often prefer to see a pair of means rather than a correlation coefficient—because means facilitate natural understanding. If you told your readers that the men in your experiment ate 5.1 ounces of chocolate whereas the women ate only 2.0 ounces of chocolate, everyone would know exactly what you meant. But if you told most readers that the point-biserial correlation between gender and chocolate consumption was $r = -.35$ (with men coded 1 and women coded 2), they'd probably ask you to see the means broken down by gender.

Make Sure You Have a Design

Before you begin data collection, it's very important to confirm that your data will be suitable for data analysis after you are done. In short, you need a research design—a framework that lets you make a clear empirical statement. There are many, many specific research designs (many of which will be covered in great detail later in this text). Instead of trying to review them all here, we'll cover a couple of general points about design and then define a couple of popular designs that often prove to be good choices for a first research project.

Assuming that you're familiar with basic descriptive statistics such as means and percentages, we're going to emphasize the fact that almost all theory-driven research is about *associations*. So whether you want to know if people discriminate against physically disabled others (relative to non-disabled others), whether men eat more meat than women, or whether taking acetaminophen reduces social pain, you are asking a question about *covariation*. To assess covariation you always need an *independent variable* (a variable you have identified as a hypothetical cause) and a *dependent variable* (a variable you've identified as a hypothetical consequence). In a typical experiment, this would mean (a) two or more groups who receive different levels of your independent variable and (b) continuous scores on one or more dependent variables. An experimental design that fits this bill is known as a **two-groups design**, and if your dependent measure is continuous, you'll usually be able to conduct a very simple *t*-test to analyze your data, as noted above. If you truly need to create more than two experimental conditions, you'll have to conduct a somewhat more complex ANOVA on your continuous dependent measure(s), as also noted above. The same thing would often hold true if you were conducting a passive observational study in which your independent variable had more than two categories (say, Democrats, Republicans, and Independents).

One pretty common mistake beginning researchers make is to fail to specify both an independent variable and a dependent variable. An even more common mistake beginning *experimenters* make is to confound one manipulation (one independent variable) with another. Having a good experimental manipulation means that people in an experimental group and a control group are treated differently in *one and only one way*. This is why studies of whether drugs work almost always use a **placebo**—a fake pill or drug that participants in the control condition believe to be the real thing. Doing so allows researchers to separate the chemical effects of a drug from the psychological effects (from the belief that one is getting a drug). In many experiments, it is important to keep *experimenters* blind to which conditions participants are experiencing. If experimenters know that some people are tasting Oreos, for example, whereas others are tasting a generic equivalent, the experimenters may give off subtle cues suggesting that one cookie is more delicious than the other. A study in which neither the participants nor the experimenters know which treatment people receive is said to use a **double blind procedure**. When it would be impossible for experimenters to ignore cues that might give away a participant's condition, for example, it might be wise to record crucial parts of the experimental instructions—or to give written instructions that create the manipulation.

If you are *not* conducting an experiment, by the way, it is still important to identify your independent and dependent variables. Likewise it will be important to create (or select) both independent and dependent variables that consist of just one thing. For example, if you wish to see if self-esteem (your independent variable) influences self-reports of romantic satisfaction, be sure that your dependent variable only measures satisfaction. "I am very satisfied with my relationship with my romantic partner" sounds good to us. But a question such as "I am satisfied with my partner because I know she truly loves me" is *not* a good idea because it is a mixture of both (a) satisfaction and (b) feelings of being loved. Even if you end up with a very simple design, however, you will almost certainly need some help from a seasoned expert when it comes to analyzing and interpreting your findings. But the key point is that if you don't have at least two groups to compare or if you identify a dependent variable but not an independent variable in your research, you're not going to have much to say about any kind of covariation. This would make John Stuart Mill very unhappy.

Consider Attention Checks

Speaking of unhappiness, one of the worst things that can ever happen to a researcher is to spend many hours carefully designing a study and learning, once you peek at your data, that your participants were not taking the study seriously. This is particularly worrisome, for example, if you collect data online using a resource such as Amazon's Mechanical Turk. Some people who take part in online studies, simply click answers at random because they want to get to that big fat 75-cent paycheck without having to do much reading. This means that in most studies, it can't hurt to include a couple of attention checks to be sure your participants are really thinking about the questions they're answering. **Attention checks** are unusual questions that have nothing to do with your research hypothesis. Their sole purpose is to see if participants are really reading or listening to your instructions, manipulations, and dependent measures. For example, near the end of a survey you might insert the following: "Now that you are done with the survey, we would like to evaluate how carefully you have been reading these instructions. If you just read this sentence, please enter the number 517 in the little box in the lower right hand corner of this page."

Some researchers prefer to embed attention checks here and there throughout their entire surveys or experimental materials. If your dependent measure is two pages long, you might include a question such as "I am paying careful attention to all of the questions in this survey" in the middle of each page of questions. If you do so, be sure to offer participants an EGWA scale anchored by something like "not at all" and "very much." Further, if you include more than one attention check, be sure that your two or more questions have very different answers. In addition to the "paying attention" question (whose obvious answer is "very much"), you might include a question whose appropriate answer is "not at all." For example, questions such as "I am very good friends with Michelle Obama" will quickly let you know who is really paying attention. Incidentally, it's very important that you make any decisions about whom to exclude from your study *without* knowing whether their data on the real questions confirmed or disconfirmed your hypothesis. Otherwise, you'll be setting yourself up for **experimenter bias** (a problem that occurs in research when a researcher's wishes or expectations about a study influence what he or she observes).

Collect Some Pilot Data

A final tip about operationalizing the variables in your first research study is that, if time and energy allow, it's often a great idea to conduct a **pilot study** (a practice study that is often a simplified version of the real thing) before conducting your full-blown study. Pilot studies often yield preliminary information about important issues such as whether your instructions are clear and whether there is plenty of variation on your dependent variable in the sample you hope to study. For example, if you were to draft a survey on frequency of texting while driving and accident history, you might begin by giving a first draft of the survey to everyone who is enrolled in your research methods class (or your section of a large research methods class). Your classmates could help you out in two different ways. First, they could answer the questions honestly (and confidentially). Second, after completing the survey, your fellow students could comment on any questions they found confusing or worrisome. Thus, someone might report that "I noticed your survey asked if we 'have *ever* been involved as a driver in a motor vehicle crash.' Wouldn't it be better to ask people how many crashes they have been in (as a driver) in the past 12 months? Offering a 12-month window would equalize drivers who've been driving for many years and those who have only been driving a year or two. Asking people exactly how many crashes they've been in should also be better than asking people the yes/no version of the same question."

You may notice that, just as we promised, we have belabored the first two steps of the six-step process of conducting research. That's because, in our view, these two steps are the most important. It's also because they're the two steps that come first. You may be happy to learn that the advice we have to offer about Steps 4 through 6 is much more succinct. Furthermore, the advice we'll offer about some of these steps is that things are different at different places. Although there's a lot of consensus about the definition of an experiment, for example, there is more variation than you might imagine in the exact process by which researchers get permission from an ethics committee (an IRB) to conduct a study. The important third stage of permission, then, is not exactly the same everywhere.

STEP 3: PERMISSION

Overview of Permission

You surely recall that psychologists who wish to conduct and publish empirical research can usually do so only after they have gotten approval (i.e., permission) to do so from an internal review board, better known as an IRB. Further, you might also recall that an implication of the risk-benefit rule is that one should not use deception in research unless there is no other good way to test an important research hypothesis. People don't usually like being deceived. Finally, you probably recall that if researchers *do* deceive their participants in any way, the researchers must carefully debrief their participants, not only to explain the exact nature of the deception and the reasons for it, but also to make sure participants leave a study feeling at least as good about themselves as they did when they arrived.

But you probably did *not* recall that part of what IRBs decide is whether a study is ethical according to local "community standards." In our experience, letting an IRB know that you are doing exactly what other scientists have already done with the blessing of their own local IRBs is no guarantee that they will consider your study ethical. For example, some IRBs look upon minor deceptions as if they are a mildly entertaining necessity when it comes to conducting experiments. Other IRBs may view minor deceptions as a violation of a sacred code, and they may come down pretty hard on researchers who wish to use even pretty small deceptions (e.g., failing to tell participants up front that one of the two teas they will taste in today's "consumer taste test" will have been branded with a Japanese-sounding label that happens to resemble their own first name). Would it really scar Jennifer for life to taste one tea labeled "Jenioki" and one tea named "Elioki"? In our experience, specific IRBs at different colleges and universities vary widely in their view of what is ethical. A great deal of this variation has to do with the naturally existing variation across the United States in local "community standards." But we should also note that even if one were to hold community standards constant, the particular people who happen to serve on IRBs often possess different personal attitudes and values that influence their judgment. If the IRB at your school is much too worried about deception but OK with administering acetaminophen, this is something you may need to take into account as you design your study—and as you write up the IRB application that you may have to submit yourself to your local IRB.

Exemptions and Expedited Reviews

We said "*may* have to submit yourself" for a couple of reasons. First, some instructors like to handle IRB applications by themselves. Others may offer student researchers a great deal of advice and assistance when it comes to completing an IRB application. Second, some categories of research are eligible for **exemption** from IRB approval. Qualifying for exemption means that a study does not have to be reviewed by an IRB at all. One of the most common exemption categories is marketing research, such as new product evaluation and taste testing foods and non-alcoholic beverages. Another common exemption category is research on the "efficacy of different educational techniques." Thus if you wish to conduct an experiment to find out which of two teaching methods is superior, you will not usually need to bother your local IRB.

Because waiting for IRB approval is never fun, you might be happy to learn that many studies in psychology also qualify for "**expedited review**." Because many psychological studies pose what has been dubbed "**minimal risk**" of physical or psychological harm to participants (no risk above and beyond what participants encounter in routine daily activities), the U.S. **Office for Human Research Protections (OHRP)** specifies a wide range of research activities that are eligible for expedited review. An expedited review means that rather than waiting for an entire IRB committee to convene and debate a proposal, a researcher whose proposal qualifies for expedited review can have the proposal reviewed (usually pretty quickly) by just one experienced member of the IRB committee. But just because a category of studies is eligible for expedited review does not mean all studies in that category truly qualify. If you were to go to the OHRP web site (https://www.hhs.gov/ohrp/) and search for "expedited review categories," you'd quickly see that the fact that a study is included in the OHRP expedited review categories list "merely means that the activity is *eligible* for review through the expedited review procedure" [emphasis added]. Sometimes the chair of an IRB will determine that a study proposed for expedited review must be reviewed by the full IRB.

All this being said, where psychological research goes, the list of studies that potentially qualify for expedited review is very long. One of the many categories for expedited review includes "Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies." So a great deal of psychological research could qualify for expedited review. We know of no OHRP policy dictating that deception studies never qualify for expedited review. But in our experience, even a small deception often means that IRB chairs will often decide that a study must undergo a full review. You might be surprised to learn that a lot of medically invasive procedures (e.g., blood draws) qualify for expedited review—assuming, of course, that they are carried out by trained medical professionals. It's also worth noting that there are differences between nations in the specific ethical guidelines that researchers must follow. For example, the ethical rules for exposing children to any kind of physical or psychological risk in Canada are a little stricter than those in the United States (Millum, 2012). Which system is better depends on whether you are more worried about the difficulty of getting permission to study an important problem in kids or more worried that kids could be harmed a lot more easily than adults by being exposed to sensitive manipulations.

A Sample IRB Application

If you have to get IRB approval to conduct a study for a research methods class, you'll need to familiarize yourself well with the exact application used by your college or university.

In our experience, careful attention to detail is one of the keys to getting a timely IRB decision. Being polite to the members of the committee never hurts either. After all, most of them are doing this sometimes taxing committee work for free—because they want to make sure no one gets hurt by taking part in research. If you would like to see an example of a completed IRB application that was approved with no requests for changes (with a small amount of personal information redacted), you can check out a proposal your first author had approved at Montgomery College in the fall of 2017. At the end of this application, we also included the Informed Consent Form that was a part of the application process. Just go to http//SAGEWEB.com to see what a real completed IRB application looks like—at least for one small college in Maryland.

STEP 4: EXECUTION (DATA COLLECTION)

We don't have nearly as much to say about execution as we did about the first three steps of conducting a research project. If we could offer just three pieces of advice, however, they would be this. First, give yourself more time—and perhaps more opportunities—than you think you'll need to complete data collection. Things that seem easy on paper often prove to be a little more difficult than people expect. For example, suppose Dr. Snorkwerth gives you permission to collect data on the 20 students in her human development class. Recognize that only 11 students may actually show up on the only Friday afternoon on which Dr. Snorkwerth was able to squeeze you in. It's better to hope to collect data from 150 people knowing 100 people is what you really need than to plan overconfidently for 100 people and end up with 60. Second, remember that the only cues that potential participants are likely to have about your status will be the way you dress and present yourself during the study. Think carefully about exactly how you'll introduce the study to others and dress professionally during data collection. If people can't see that you take your study seriously, why would you expect them to do so? Third, make careful notes about your experience. If a group of participants seems particularly sleepy or enthusiastic or if it is 8 a.m. rather than 2:30 p.m. when you collect data, these facts themselves—along with your attention checks, of course—may prove to be important data points. And it wouldn't hurt to revisit the list of other research situations that Hardin (2018) suggested can be so important.

STEP 5: CALCULATION (DATA ANALYSIS)

It may seem like we're getting a little ahead of the game to worry about statistical analysis in the very early stages of conducting research. But it's important to think about how you plan to analyze your data *before* you finalize the design of your study. As we already noted, you'll only be able to conduct a *t*-test (more specifically, an independent samples *t*-test) on your data if you have a categorical independent variable with exactly two levels *and* if you have a dependent measure that is continuous. But it's actually a little trickier than that. To know for *sure* if such a *t*-test is really appropriate, you'll also need to take a close look at your raw data once they come in. A *t*-test is only statistically appropriate if scores on your continuous dependent variable really do prove to be continuous. They won't be, for example, if 93 percent of your participants circled either "6" or "7" on your 7-point scale. Further, an independent samples *t*-test is appropriate only if the scores on your dependent variable turn out to be more or less **normally distributed**. That is, the distribution of your scores must resemble the familiar *bell-shaped curve* that, luckily enough, applies to most things that are carefully measured on a continuous scale. If you are unlucky enough that your data are *not* normally distributed, all is not lost. But you'll need some expert advice from your instructor about how to correct this statistical problem. Don't worry. She was probably a statistics minor in graduate school, and she's got your back.

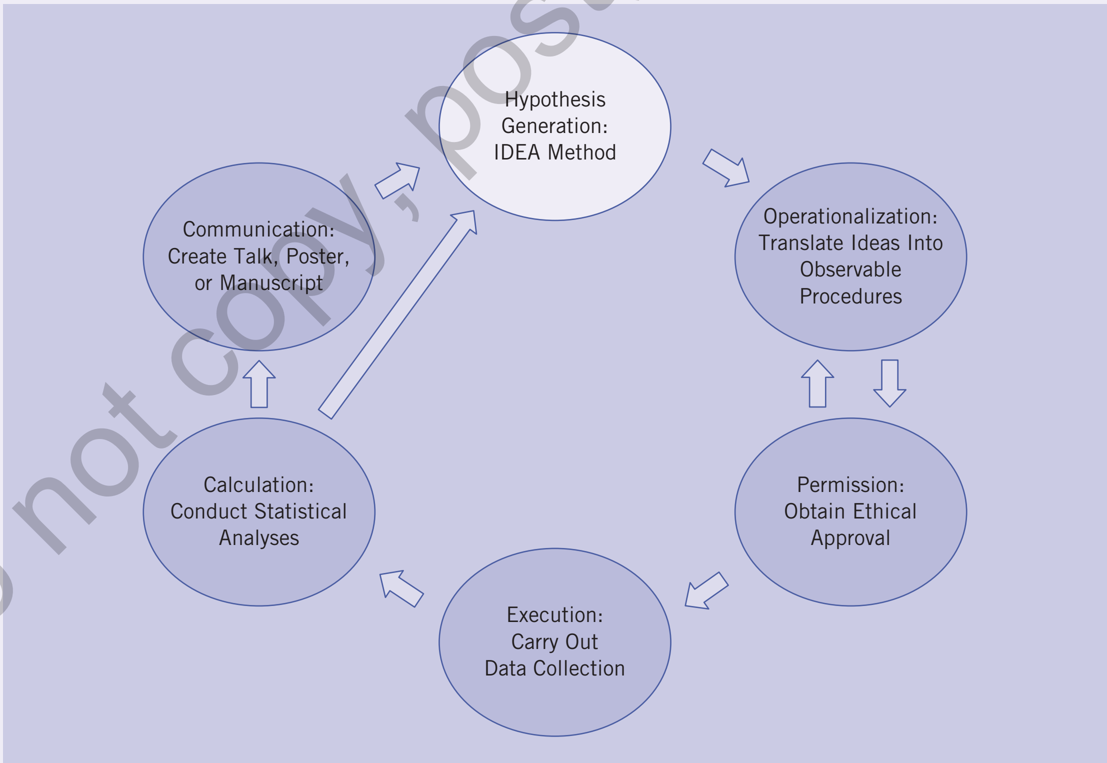
STEP 6: COMMUNICATION

So you think you're done with your research project because you have analyzed your data? Actually, you surely knew painfully well that you are *not* done with a research project just because you have analyzed your data. The final step in any research project is communicating your findings to the scientific world. This might mean creating a conference poster or giving a brief research talk. More often, especially in a course in psychological research methods, this means writing an APA-style empirical research report to summarize what you did, what you found, and what you think it means. By the time you reach that point in your research project, we hope you will have had a chance to read the chapter in this text on "telling the world about it." We won't belabor any of the technical or conceptual points in that chapter here. The only thing we will add here is that there is no better example in psychology of what we call the *Sanborn-Edison equation* than writing empirical research papers. This "equation" is usually attributed to Thomas Edison, but the public speaker Kate Sanborn seems to have beaten Edison to the punch (see https://quoteinvestigator.com/). The version of the equation that eventually became a popular *quotation* is "Genius is 1 percent inspiration and 99 percent perspiration." By this, both Sanborn and Edison meant that most people think ingenious work is the product of natural genius whereas it is actually the product of working your butt off. If you're willing to read and digest our chapter on how to communicate about psychological research and if your instructor is able to give you some advice and feedback on top of this advice, there is no reason you can't do a great job of letting the world know what you found in your research. Good luck! And remember that an ingenious research paper is only about 4 percent equations and 1 percent quotations; the other 95 percent is up to you.

SUMMARY

This chapter introduces new researchers to the basic steps in conducting psychological research. These six steps include: (1) generation (coming up with a hypothesis), (2) operationalization (converting your abstract idea into a concrete study), (3) permission (getting IRB approval whenever necessary), (4) execution (carrying out the actual study), (5) calculation (data analysis), and (6) communication (creating a poster, research talk, or research paper that communicates your work to other researchers). Because the first two steps of this research process are the most daunting and mysterious we pay the most attention in this chapter to spelling them out in great detail. Specifically, we introduce the IDEA model as a heuristic for generating a psychological research idea. IDEA stands for integration, dissection, extension, and application. Likewise, we break down Step 2 (operationalization) into seven sub-steps, which include (1) reading the literature, (2) consulting an expert, (3) maximizing statistical power, (4) keeping it simple, (5) making sure you have a real design, (6) considering attention checks, and (7) collecting pilot data. We also provide a few pragmatic details about exactly how to get IRB approval for your study (permission) as well as offering a few tips about the last few steps of the research process (e.g., giving yourself plenty of time for execution). This entire process is summarized in Figure 4.8.

Figure 4.8 The six-step psychological research process summarized in this chapter



Because we hope you are now familiar with all the steps in this model, we will not belabor them all. However, we would like to draw your attention to two aspects of the research process we did *not* point out explicitly in this chapter. First, notice that the arrows between operationalization (design) and permission run in both directions. This part of the model is bidirectional because sometimes an IRB identifies ethical concerns a researcher had not appreciated that require design changes. This, in turn, leads the researcher to redesign accordingly and resubmit the project for IRB approval. Second, notice that although the model begins with hypothesis generation (the white oval), either of the last two steps of the process (data analysis or communication with other researchers) may lead a researcher to conduct a new study whose main inspiration was the outcome of the study that began the cycle. Thus, the cycle begins again.

STUDY QUESTIONS

1. Consider the four quotations from hypothetical research articles listed below. Next to each quotation indicate whether the study to which the quotation refers was most likely a product of I. Integration, D. Dissection. E. Extension, or A. Application.

   ___ I. "Whereas self-enhancement theories predict that people should prefer the flattering evaluator, self-verification theories predicts that people should prefer the unflattering but accurate evaluator."

   ___ II. "Were patients who scored high in conscientiousness more likely to report taking their blood pressure medication on a daily basis? Did this reduce their susceptibility to a heart attack?"

   ___III. "Whereas this meta-analysis clearly shows that Lieberman's popular likability manipulation reduces conflict during negotiations, the existing manipulation appears to confound likability *per se* and cues of past cooperation. This experiment was designed to see which cue . . ."

   ___IV. "Although laboratory studies of the mere ownership effect reveal a clear and robust effect among college students in the lab, we wanted to see if the effect would be robust in a hunter gatherer culture where ideas of personal ownership are more fluid and malleable than in most Western cultures."

2. Chapter 2 discussed inductive versus deductive techniques for generating research hypotheses. Next to each inductive or deductive approach listed below, indicate where this technique best fits into the IDEA system. Place a letter or two (I, D, E, or A) next to each example.

   ___ I. *Case studies* that suggests that a well-known effect does not apply to a specific population

   ___ II. *Paradoxical incidents* that yield the opposite outcome of what one would expect from a particular theory

   ___ III. Bowlby's "*reasoning by analogy*," which led him to the test the idea that the same basic infant attachment processes observed in other primates would also apply to human infants

___ IV. "*Accounting for conflicting results*" by arguing that when people evaluate others they make an important distinction between how "trustworthy" someone is and how "likable" the person is

___ V. Using the *hypothetico-deductive method* to predict that, if human beings are highly egocentric, then people who live in states with colder climate should be more skeptical of "global warming."

3. In exactly what way does knowing about the OOPS! heuristic help a person generate a novel study that relies on the *extension* component of the IDEA system? Cherise tested the hypothesis that "we are attracted to others who resemble those we already like, even if the resemblance is arbitrary." She did so by showing that college students at a small college in Florida liked a person more when they thought the person had the same first name as their oldest sibling (as opposed to a well-matched control name). How might you *extend* this study using each component the OOPS! heuristic?

4. You meet a researcher at a conference, and she tells you that she mainly studies risk-taking behavior as it relates to "increasing safe driving, safe sex, and saving for one's retirement." What is your best guess about which aspect of the IDEA system she personally uses most routinely to generate new research ideas?

5. Name at least two advantages of using the Web of Science (rather than just Googling a topic) to see what research has already been done in an area. Use the words "precision" and "exhaustive" in your answer.

6. Researchers whose studies take a long time to complete often worry about participant boredom or fatigue. How might attention checks help with this problem? How might you design a study to see if attention checks become more useful when a study takes longer to complete?

7. What is the difference between an exemption from IRB review and an expedited review?

8. Kara received an expedited review of her experiment on "priming trust in close friendships" at Kansas State University. When Kara's collaborator at the University of Virginia asked her local IRB to review exactly the same proposal, they decide to send it out for a complete review rather than expediting it. What can explain this difference?

NOTES

1. We thank the gracious and ingenious David Boninger, who gave us a detailed suggestion for writing exactly this kind of chapter on how to get started in research.

2. The APA-style manual says one should use the label "Notes" even when there is only *one* note. We think this violates the rules of grammar and common decency, and so we added this bogus note.