RECRUITING PARTICIPANTS

CHAPTER HIGHLIGHTS

• Subject pools: Characteristics, software, practical issues
• Recruiting off campus
• Service learning: Benefits; conflicts and multiple relationships
• 12 tips for recruiting in the field (Dustin’s Dozen)
• Recruiting online: Social media; adverts and snowballing; ethical concerns
• Types of sampling
• Response issues
• Incentives: Practical considerations
• Amazon Mechanical Turk and online paid panels

OVERVIEW

Far in advance of recruiting participants, researchers confront the challenge of who these people will be and how they will be reached, given that this information is required for institutional review board (IRB) approval. Thinking back to the definition in the Code of Federal regulations (45 CFR 46; see Chapter 4), doing research with human subjects requires IRB review. In this definition, the human subjects are living, and the researcher is interacting or intervening with the individual or obtaining identifiable private information about that person. True, research may involve archival information that does not include identifiable information, but the focus here will be on subjects with whom one plans to interact.
This chapter will emphasize the practical aspects of recruiting participants and cover the most common sources of participants, starting with the “closest link,” university subject pools, and ending with the “farthest link,” online crowdsourcing platforms like Amazon MTurk® and paid participant panels. Issues in recruiting off campus will be discussed, as will the use of vulnerable populations and the challenges in securing permissions (see also Chapter 4). The use of incentives will be covered, as will the timing of your request over the course of the semester. Concepts in sampling will be introduced and related to internal validity.

**WHO PARTICIPATES IN RESEARCH: AN OVERVIEW**

When we think about scientific findings regarding human nature, we need to ask ourselves about the participants (people) in that research. In particular, we need to think about how representative they are of the population in question across the demographic categories of interest (e.g., age, gender, education, race, and income). *Representativeness (sampling)* is the degree to which your sample reflects the population as a whole. Much of the information we have about human nature comes from a very limited slice of humanity. This is a problem for science. You may not be able to solve that problem in your research, but at the very least, you need to be aware of the “who” of your research and the “how” of your sampling. Both who participates and how you obtain those individuals may limit your results. Such issues are part of sampling bias, as discussed in this chapter.

In research, especially in social science (and psychology in particular), there is a heavy reliance on subject pools for research participants. A *subject pool* is formed of individuals who have agreed to serve as participants for research. In a study titled “The Weirdest People in the World?” Joseph Henrich, Steven Heine, and Ara Norenzayan (2010) talked about the predominance of participants who come from societies that are “Western, Educated, Industrialized, Rich, and Democratic” (hence WEIRD) in research. This paper should be required reading for anyone conducting research with human subjects, particularly studies that rely on subject pools. Henrich et al. argued that it is “Western, and more specifically American, undergraduates who form the bulk of the database in the experimental branches of psychology, cognitive science, and economics, as well as allied fields” (p. 61).

There is increasing recognition that unless we increase the diversity of our samples, we have a skewed picture of human nature. The title of Robert Guthrie’s 1976 book *Even the Rat Was White* points to this lack of representativeness in psychology and social science more broadly. Echoing a theme we saw when talking about the value in adding qualitative to our heavily quantitative approaches (see Chapter 6), Paul Rozin (2009) talked about the kinds of research we should be publishing. In such research, there is an emphasis on questions that reflect greater diversity, not only of participants, the focus of this chapter, but also of approaches. Are we making any progress on this front? Although there seems
to be an increase in the number of non-U.S. authors publishing in psychology journals and an increase in multiauthor collaborations (Piocuda, Smyers, Knysh, Harris, & Rai, 2015), Jorge Piocuda et al. did not report whether there was internationalization of the samples in that research. Research from Jeffrey Arnett (2008) showed that for 2007, “in 67% of American studies published in JPSP, the samples consisted of undergraduate psychology students” (p. 604). JPSP, the Journal of Personality and Social Psychology, is one of the premier journals in the discipline.

Try This Now

What is the potential impact of using WEIRD people, and American undergraduates in particular, on our understanding of human nature? Are there kinds of studies where you think using WEIRD people would be less likely to limit the results?

THE SUBJECT POOL: THE WORKHORSE OF SOCIAL SCIENCE RESEARCH

A good deal of research is done using American undergraduates, and most of these people come from subject pools in psychology. There are two general types of subject pools: unpaid and paid. **Unpaid subject pools** are those typically linked to course participation requirements, where participating in research (or an alternative, see Chapter 4) is a requirement of the course. Unpaid pools may also provide a source of extra credit in a course. It is important to reinforce the fact that participation in such pools is voluntary; we cannot require people to participate in research. **Paid subject pools** are typically available at large research universities where faculty research grants fund the pools. Participants may be students, staff, faculty, and even individuals unrelated to the institution; these volunteers are paid to participate in research projects. Such pools may be offered within psychology departments, medical schools, or business schools, as examples.

THE DRAWBACKS TO SUBJECT POOLS: CONCERNS ABOUT INTERNAL VALIDITY

As indicated in the research reported by Arnett (2008) and by Henrich et al. (2010), there is a lack of diversity in undergraduate subject pools. Such pools are thus unrepresentative of the public at large. There are other ways in which these pools are unrepresentative. Some of these differences have to do with when people are likely to participate and in what form (in person vs. online).
Across the subject pool, participants differ in ways that may affect the internal validity of the research project. For example, there are differences in extraversion between those who choose to participate in person (higher levels of extraversion) compared to those who select to participate online. Furthermore, those who participate earlier in the semester are likely to be more conscientious and to be women than those who participate later in the semester (Witt, Donnellan, & Orlando, 2011). Using the NEO-PI-3®, a widely used personality inventory (see Chapter 5), research looking at the timing of participation in the semester showed that higher levels of extraversion and openness were associated with participating later in the semester (Aviv, Zelinski, Rallo, & Larson, 2002). As A. L. Aviv et al. (2002) argued, the differences in personality associated with participating at different times of the semester create confounds and undermine the internal validity of the research unless steps are taken to address these problems.

**Try This Now**

Given these differences in personality characteristics and when people decide to participate, what might happen if we run an experiment with sequential conditions, such that all of Condition 1 is run in September and all of Condition 2 is run in November?

As Aviv et al. (2002) discussed, this kind of sequential approach may lead to unrepresentativeness, given that personality may be confounded with time of the semester. Another type of research approach that is problematic is a study that is conducted over a short period of time (e.g., 1–2 weeks). In both of those scenarios (sequential and short time frame), the people who volunteer at an early point in time likely will differ from those who volunteer later in the semester. And these differences may interact with the variable(s) in the study.

How would you change the way the subject pool is managed to eliminate this problem? The solution offered by Aviv et al. (2002) is to eliminate choice. By randomly assigning participants to experiments, these potentially damaging differences in the nonrandom distribution of personality characteristics are eliminated. Aviv et al. stated that there are psychology departments in the United States that assign participants in this manner. It is unclear how common this approach is.

**Try This Now**

How does your institution recruit and assign participants for experiments?
LIMITS ON THE NUMBER OF PARTICIPANTS AVAILABLE FROM UNPAID SUBJECT POOLS

Participants are valuable because they are limited in number, typically being generated through course requirements. For that reason, most unpaid subject pools set restrictions on the number of participants any given researcher may acquire through the pool. There is typically a priority system of who has access and in what order. The size of the institution and hence the number of students enrolled has a direct bearing on the number of participants available for research. Research is more difficult to conduct at small institutions because of the limited number of participants available.

REVISIT AND RESPOND

- List two potential threats to the internal validity of a study when you use a subject pool and how researchers have tried to combat these threats.

KEEPING TRACK OF PARTICIPANTS: ONLINE PARTICIPANT MANAGEMENT SYSTEMS

Increasingly institutions are turning to software not only to create studies (e.g., SurveyMonkey® and Qualtrics®) but also to manage recordkeeping between researchers and participants. One of the most widely used cloud-based participant management systems is Sona Systems® (https://www.sona-systems.com/default.aspx), which describes itself as the “global leader in university research software” that “serves the top 50 universities in the US” according to its website.

This kind of software brings researchers and participants together. Researchers and participants create accounts; researchers can post their studies online, and participants can sign up online for these posted studies. Many features help to streamline the research process. Researchers can list criteria for participation (e.g., athlete status, class year, or gender), send reminders, automatically award credit, reserve rooms for running their studies, document no-shows or excused absences, as well as several other conveniences. Participants have a way to document their participation rather than relying on signed informed consent documents, which are easy to misplace.
The software is relatively straightforward, and this approach to managing participants is widely accepted. As Sona Systems states on its website, “The days of bulletin boards and sign up sheets are over.” As you might expect, there is a fee for such participant management software. For Sona, the price depends on the volume of usage, measured by the number of participant sessions managed in the system.

**PRACTICAL ISSUES IN COMMUNICATING ABOUT RECRUITING**

Whether you are recruiting from the subject pool or elsewhere (more on that shortly), some aspects to the study affect its success, in terms of both attracting participants and internal validity.

**Study Labels**

Earlier in the book (see Chapter 3 on demand characteristics and cover stories; see Chapter 5 on scale labels) we discussed the ways in which research is a communication process. What participants see or hear about the research influences their perceptions about what is going to take place. One of these influences is the title of the study.

Consider the following two titles for the same project:

- **Perceptions of Women in Sports**
- **Perceptions of Femininity and Women’s Weightlifting**

The research in question (mentioned in Chapters 6 and 7) is assessing how judgments of femininity and masculinity are affected by exposure to women holding barbells of different weights [5 lbs (80 oz) or 25 lbs (400 oz)], but including femininity in the title would likely bias participants. When you advertise your study (whatever the platform), you want to describe the nature of the study in general terms without providing information that discloses the hypotheses. This same caution is in order at the beginning and end of your informed consent where you describe the study (see Appendix B at the end of the book for a sample informed consent document).

**Study Length**

You recall that one of the elements of informed consent (see Chapter 4) is a statement of how long the research could be expected to take. Normally researchers determine this length by pilot testing or having a friend or colleague complete the study (e.g., fill out the questionnaires) and then use that estimate on the informed consent and in advertising the study. Typically students in a research pool who are participating for course credit
Recruiting Participants

are required to fulfill a given number of hours of participation (e.g., five), although some institutions state the requirement in terms of a specific number of studies. Students are astute; if they need to participate in a given number of hours, they often select studies that take more time (up to a point) to reduce the absolute number of studies they need for their course requirement. If your study takes 30 minutes, or 45 minutes, or 60 minutes, it may have more “takers” than if it were only a 15-minute study. The length of the advertised study also plays a role in crowdsourcing platforms like Amazon Mechanical Turk (discussed later in the chapter). If potential workers think that the ratio of pay/length of task is unfavorable (e.g., $0.25 for a 30-minute survey), you may have few participants (called “Workers”).

Time of the Semester

The time of the semester you recruit participants can affect who signs up. Women and those who are conscientious are more likely to sign up earlier in the semester (Witt et al., 2011); those who are extraverts and exhibit openness to experience are more likely to sign up later in the semester (Aviv et al., 2002). Beyond these differences, which may affect the internal validity of your study, there is the basic problem of running out of participants in the pool. If you wait until after Thanksgiving or Spring Break to collect your data from the pool, there are likely to be few subject hours left. In addition, you have little time to analyze the data and complete your research paper before the semester ends.

Time of Day and Day of Week

If you are using a subject pool where students have to appear in person to take a study, it makes sense to offer multiple data collection sessions scheduled at different times of the day and throughout the week. Most in-person sessions would be held later in the afternoon (after classes end) when rooms to run research are available. It is important to consider that some participant groups (e.g., athletes) are restricted in their availability. Offering evening as well as afternoon sessions thus makes sense. In addition, think about the days of the week that should produce the greatest level of participation. One recommendation is that you offer sessions Monday–Thursday, with less emphasis on Friday. Some researchers recommend Tuesday–Thursday because students returning to campus after a weekend may not remember a Monday obligation.

Attrition or Experimental Mortality

Attrition, or dropping out (also called experimental mortality, see Chapters 3 and 8), is likely to be a problem in longitudinal research or some within subjects designs where you are worried about carryover effects and want to schedule sessions with time in between. An additional possibility is the case of pre–post research (see Chapter 8 for each of these research designs). In the case of longitudinal research, you are collecting data from the
same people, sometimes for years. In the within subjects approach, where carryover effects are a concern, you may want the effect of one treatment level to wear off before assessing the effect of another treatment level (e.g., the effect of different doses of caffeine on memory). In the pre–post approach, you may want to test whether an intervention has an effect extended in time. For that reason, you may schedule a break between data collection sessions and request that people return for a follow-up session on a different day. You cannot require participants to return, nor can you withhold research credit if they don’t.

**Try This Now**

What strategies might you employ to encourage people to return for follow-up sessions?

The most effective way to persuade people to return is to explain the importance of their participation and the effect that dropping out would have on the research. Remembering that the research process is one of communication, establishing a connection to your participants by treating them with respect, and thanking them for participating helps to create a positive climate for the research process. This positive climate is easier to establish in person than online, but the same principles of respect and civility hold for online studies.

**REVISIT AND RESPOND**

- Imagine that you are doing a study on the effect of wearing glasses on perceptions of professionalism in the workplace. Come up with a title that reflects the recommendations for labeling your study (to avoid demand characteristics). When would you schedule your study (day of week, time of day) for maximum representativeness?

**RESEARCH ON SENSITIVE TOPICS AND THE ROLE OF THE IRB**

Earlier in the book (see Chapter 5) we discussed the order of questions in a survey. Don Dillman et al.’s (Dillman, Smyth, & Christian, 2014) advice was to place sensitive questions as far into the survey as possible so that some level of trust between the participants and the researcher has a chance to be established. A level of trust could come from clear communication, for example, organizing the material by topic and labeling sections numerically. But what about the situation when the research topic itself is sensitive, such as sexual assault or addiction?
Try This Now

In your opinion, does the researcher have an obligation to inform participants that the topic may be sensitive? Could warning have an effect on who participates (i.e., the representativeness of the sample)?

Beyond establishing a level of trust, the issue of the sensitivity of the topic is an ethical one. Researchers have an ethical obligation to alert participants to material that may be upsetting, and most researchers endorse alerting participants to that possibility. Since sensitive research may be relevant to researchers (concerns for safety of interviewees who may be violent), members of ethics committees (who read the proposal), transcribers of information (who read interviews with sensitive material), and even readers of publications that include that material.

RECRUITING OFF CAMPUS

Not every research question can be evaluated using a sample of undergraduate college students, nor do we want to rely on such pools to develop our understanding of human nature (Henrich et al., 2010). One principle to remember is that the nature of the research question influences the selection of the participants; target recruitment for locations where you are likely to find people with the characteristics you seek. If you were interested in organic food, for example, you might go to local farmers’ markets (understanding that you need permission to recruit at private venues).

When you move off campus, the difficulty of recruiting participants typically increases substantially. The following sections address how to recruit successfully in the field.

USING YOUR PERSONAL CONNECTIONS

One of the best ways to recruit people off campus is to use your personal connections. These connections can be direct, for example, your membership in an off-campus group (e.g., soccer club or religious organization) or indirect (e.g., knowing someone who is a member of a group, or whose parents or relatives or neighbors belong to a group or

---

1There are concerns in the classroom about teaching topics that may include sensitive material; alerting students to this material is referred to as including a trigger warning (Flaherty, 2014).
organization). Gaining access to workplace environments for research can also be facilitated through connections, using one's relatives as the primary example. Your institution's career services office may be another good source of contacts. Given the interconnectedness possible through social media, reaching out to friends and acquaintances is easy.

In Ann Devlin et al. (2013), one sample in this study of responses to multicultural art in a therapist's office was a community group. The data collection took place in the large activity space of a community health clinic. The sample was obtained through identifying connections to the community. The director of the health clinic, a spouse of a colleague at the author's institution, not only helped advertise the study in the community but also attended every data collection session to assist with translations (the study was offered in English and Spanish). The sessions were run at lunchtime for convenience; pizza and beverages were provided; and an incentive (a $10 voucher to use at a local grocery store) was offered.

**USING YOUR INSTITUTION'S CONNECTIONS**

In addition to the personal connections, your institution or place of work may have formal connections to other institutions or facilities. Your school may be part of an athletic conference or a consortium of schools that shares resources; through work you may belong to a larger network of similar job types or categories. Your institution may belong to organizations as well (e.g., National Association of Independent Colleges and Universities or National High School Coaches Association). Doing an Internet search using the term “organizations in higher education” will produce a list of possibilities to investigate; the same approach could be tried with any job category. Another resource from institutions is the alumni network.

**BUREAUCRACY**

As we discussed in the section on ethnography (see Chapter 6), gaining access is one of the challenges to conducting research in the community. Remember that IRBs (see Chapter 4) will expect verification of permission to gain access to a particular sample.

**VULNERABLE POPULATIONS IN THE COMMUNITY**

As we saw in Chapter 4, vulnerable populations trigger *Full IRB Review*, the highest level. The vulnerable population most often selected for student research is children, defined as those who have not reached the age of majority. Many layers of permissions are required for vulnerable populations (see Chapter 4 for coverage of permission with children).
When doing research with populations where the state has control, such as inmates in correctional facilities or patients in state psychiatric facilities, the permission, if granted at all, may take as long as a year to secure.

**PHYSICAL SECURITY ISSUES IN CONDUCTING RESEARCH OFF CAMPUS**

Whether doing research off campus with vulnerable populations or not, security issues for both the participant and the researcher are relevant. Often discussion about security issues focuses on the data itself, especially online data, and how it should be safeguarded (see, for example, Quinn, 2015), but here we will focus on aspects of physical security.

For both the participant and the researcher, one major question is where the research will take place and who else will be there. To enhance security, recommendations are usually to conduct the research in a public place (e.g., a café or a library) unless doing so would put the participant at risk (see Chapter 4). The presence of others may provide security for the participant (strength in numbers) but also for the researcher. If the research deals with sexual assault, addiction, or another topic considered sensitive where the participant might become agitated and/or someone connected to the participant might confront the researcher, having others at the location of the research is advisable. If research is being conducted in a school and having someone else in the room would limit disclosure by the participant or make the participant uncomfortable, a recommendation is to conduct the research in a quiet area but with the door open or ajar. As a researcher, you don’t want to be in a potentially compromising situation, which might lead to accusations of misconduct.


These recommendations include assessing risk in the field (e.g., determining whether public transportation is available); investigating housing and living environments ahead of time (e.g., understanding the layout of dense housing complexes); and telling the interviewee that others know your whereabouts and your schedule. Assessment of risk is particularly important if you are in areas where there is crime, it is the evening, or you are in someone’s home.

**REVISIT AND RESPOND**

- What kinds of connections do you have that would enable you to conduct research with off-campus populations? What steps would you take to safeguard your physical security and the security of your participants? Beyond the steps that are outlined in this chapter, what other ideas do you have about securing your physical safety?
SERVICE LEARNING COURSES AND RECRUITING PARTICIPANTS: OPPORTUNITIES AND COMPLICATIONS

Service learning courses are those in which students integrate academic material with internships or volunteer work in the community (and typically receive course credit in the process). This approach is a form of experiential learning where students put into action what they have learned (or are learning) in the classroom.

**Try This Now**

Before reading further, what are the advantages you see to using service learning experiences to collect data? What are the potential drawbacks? [Think back to Zimbardo’s (Haney, Banks, & Zimbardo, 1973) double role as researcher and prison superintendent in Chapter 4.]

Conducting research in the setting where you are currently doing an internship or otherwise participating in service learning raises potential ethical conflicts and may be considered exploitative (see the next section on conflicts of interest and multiple relationships). Because doing service learning and conducting research at the same time may create conflicts about motivations and goals, it makes more sense to do research in a service learning setting *after* you have concluded your course work there. If the setting is large enough, another possibility is to do research in another part of the setting where you are not interning (e.g., a different classroom, unit, department, or branch).

After you conclude your internship, you will likely continue to have access to the setting, but you will no longer be a “*double agent*” (Yanos & Ziedonis, 2006). The idea in being a double agent is that you may not be acting solely in the best interest of your client (here, the people you serve in the service learning environment), if you are also interested in them as research participants. In other words, if you are both an intern and a researcher, you have a potential conflict of interest.

CONFLICTS OF INTEREST AND MULTIPLE RELATIONSHIPS

Participants are not available in unlimited numbers; most subject pools have restrictions to the number of people you can obtain through the pool. Given those pressures, researchers...
often turn to sources where they have connections, such as clubs, athletic teams, religious groups, and career-related List Serves. When we approach people to whom we have a connection, several ethical issues emerge, especially if there are power differences in the people asking and the people responding. For example, given that one interest area for student research is athletics, you can imagine someone who is a team captain or fourth year student recruiting younger team members as participants.

**Try This Now**

What conflicts of interest might there be when fourth year students or team captains recruit players from their teams to be research participants?

The American Psychological Association (APA) Code of Ethics has a section (2010a, 3.05) dealing with avoiding multiple relationships. The gist of this guideline is that if you are in a professional role with a person (e.g., therapist or teacher) you should not enter into another kind of role with that person (e.g., as a researcher collaborator) if the possibility exists that the new relationship might impair your judgment or effectiveness in your professional work as a psychologist. Although we are not assuming the researcher is a psychologist, we are talking about the idea that having multiple relationships with potential participants should raise warning flags. Lynne Roberts and Peter Allen (2015) discussed this issue in the context of research in education and suggest that when one’s students are possible participants, such students should be considered a vulnerable population. Imagine the situation where an assistant lacrosse coach, enrolled in an MA program, wants to use members of his or her team as participants in a study of personality, athlete goals and strivings (as reflected in personal narratives), and career objectives. How can the researcher avoid the possibility of coercion in this situation?

Some researchers in this kind of situation have someone else approach the team members; in addition, the guarantee is made that the data will be collected anonymously and that participants’ identities will not be known to the researcher. Nevertheless, if there are few players, it is possible that some identifying information may be revealed in the narratives the players produce. Yet another possibility, and one that makes more sense in terms of reducing the possibility of coercion, is to use another team at the institution, or from another institution altogether. It is likely that the research questions are not about a specific team (e.g., lacrosse) but about athletes more generally. For that reason, a substitute team, either at one’s institution or at another institution (making use of one’s contacts) is a reasonable approach and one that avoids the conflict of interest.
What are the benefits of conducting research at service learning sites? What are the potential drawbacks? What does it mean to be a “double agent” in the context of research? If you want to collect data from a group where you are a member, how can you avoid the possibility of coercion in requesting their participation?

DUSTIN’S DOZEN: TIPS FOR COLLECTING DATA IN THE FIELD

(Quoted or in some sections closely adapted from Devlin, 2006, pp. 141–144.)

We have talked about places one might find participants, such as service learning courses or clubs or other institutions. Next we will talk about recommendations for collecting data in public places such as train stations or plazas. Sarah Ibrahim and Souraya Sidani (2014) categorized strategies into those that are proactive (direct) and reactive (indirect). Proactive approaches include face-to-face encounters at such sites as community centers, health centers, churches, centers for seniors, farmers’ markets, and street fairs. I would also add libraries, community parks, and parades (before and after the event). Reactive (indirect) approaches include working through community leaders, snowballing, word of mouth, and media.

The many useful suggestions from a former graduate student, Dustin Wielt, have been compiled into a list of do’s and don’ts for collecting data in the field. Dustin’s MA project was a study of the attributions people made about murderers as a function of the murderer’s criminal history, psychiatric history, and social status. Over the two months he collected his data (a sample of more than 300 participants), his major collection point was the local Amtrak train station, but he also collected data in a laundromat. Here are Dustin’s recommendations:

1. **Dress decently.** As Dustin said, don’t dress like a “bum.” If you are a male student, wear a shirt tucked in and a belt. Don’t wear a baseball cap facing backward. Basically, if you are negotiating with a business owner about whether you can collect data at his or her establishment, dress appropriately. And dress appropriately when you collect the data, of course!

2. **The inverse correlation.** There seems to be an inverse correlation between the likelihood of gaining permission to collect data at a facility and the size of the bureaucracy. For example, Dustin asked permission of the local Amtrak police and the Amtrak counter personnel for permission to distribute questionnaires at
the local train station. They granted permission, with the caveat that if anyone complained, he would be asked to leave. Then they “shook” on it. This agreement worked very well. But when Dustin pursued gaining permission from the much larger regional train station by calling the number of an Amtrak official there, the response was “no.” The higher up you are required to go to get permission (and the less likely you are to do it in person), the less likely it is that the permission will be granted. This same experience was replicated in seeking permission from the local owner of a laundromat (“yes”) versus from the manager of a chain of laundromats (“no”).

3. Avoid “No Solicitation” locations. As a matter of policy, establishments that have “No solicitation” signs posted are less likely to agree to your research.

4. Carefully select your “first ask” of the day. As your first “ask” of the day, try to identify someone who seems likely to cooperate with you. As an example, Dustin found that women who were reading books were likely to agree to participate. If they said “yes,” the people sitting around them became curious and were more likely to agree. If the first person you ask who is sitting in a group says “no,” it is less likely that others will then have the courage to override this social norm and say “yes.” Dustin estimated that on average, approximately 60% of those he first asked said “yes” immediately, and if they asked a question about the research, another 10% said “yes.”

Although there may be difficulties, the train station turned out to be a good location for data collection. In Dustin’s sample, he had an age range of 18–84, and the sample was almost normally distributed. There was also racial and socioeconomic diversity, as well as a fairly good gender split (53% women, 47% men). There may be other ways in which this sample is not representative (people who take trains may be different in certain respects from those who do not), but it is a reasonable place to start.

Although he did not implement it, another strategy Dustin considered was the use of a confederate (see Chapter 4) who could be approached first for a given day’s data collection session to say “yes,” thereby establishing a climate of cooperation. In the two months that Dustin collected data at the train station, he had only two to three people who were asked twice to fill out the questionnaire, so the number of repeats was quite low, and a confederate could be effective in this situation with a high turnover of people.

5. Personalize your “pitch.” Ask the individual for participation in YOUR project, as opposed to simply asking for participation. It is harder to turn down someone when you are turning down the individual as opposed to a representative of a large organization.
6. **Be mobile.** Rather than putting yourself at a desk or in a stationary position, which strangers may be reluctant to approach, move around the setting. If you are collecting data at a train station, a laundromat, or some other facility, put your questionnaires on a clipboard and approach people for their help. This approach works much more effectively than waiting for people to approach you.

7. **Approach people with a respect for their personal space.** Stand about 5–6 feet away when you first ask people to participate. If they agree to participate, you can move closer to hand them the clipboard. Dustin recommended having about 10 clipboards available (in a backpack or tote) to hand out and then “reloading” them with fresh questionnaires so that materials are always available. He provided pens (which he retrieved) for all participants. He placed the debriefing form, face down, at the end of the questionnaire so that it was immediately available to participants when they finished the questionnaire. Make sure that the informed consent information in the introduction to the study is printed in large enough type so that people can easily read it.

8. **Be ready to assist those with low reading levels.** At the laundromat, Dustin found several people for whom English was a second language or for whom reading at a sixth-grade level was a problem. Although he asked people if they had any questions before they started filling out the questionnaires, on occasion, people asked him for definitions of words (one example is the word “generous”). At this point, he offered to read the questionnaire to this particular individual, and she agreed. Questionnaires can be designed for a minimum reading level, and certain word processing programs have reading-level indicators (refer to explanation in Chapter 5 under “Length and Difficulty of Measures”). You can also pilot your materials by asking several people to read your questions and comment on the appropriateness of the reading level for your target population. Although questionnaires can be designed for a given minimum reading level, it is still the case that the researcher may need to assist participants. When this happens, the researcher then will need to make a decision about the validity of the data (because people’s responses may change in a socially desirable direction) and whether these data should be included in the analyses.

9. **Limit the length of your survey.** Dustin estimated that a 10–15-minute questionnaire is the maximum length that will work, especially in places like train stations where people are waiting but may not arrive too much in advance of their trains.
He also said that the day before Thanksgiving was his most successful data collection session, where he collected more than 100 questionnaires as people began their holiday travel. Trains were running 10–15 minutes late all day, so that even those who had planned to arrive just a few minutes before the train’s scheduled departure had “time on their hands.”

10. Be sensitive to the day of the week and the time of day. Dustin found that weekends were better than weekdays for cooperation as weekend travelers were more likely to be traveling for pleasure as opposed to business. Business people traveling on Monday mornings were the least cooperative.

At the laundromat, the best time to approach people was the moment after they had put their clothes in the dryer (rather than when they were folding clothes, for example).

11. Safeguard your personal information. Dustin recommended listing the department’s phone number rather than your home or dorm number as a contact number on the debriefing sheet. This is a matter of safety. He was once called at 3:30 a.m. by someone who was drunk and disoriented. This person wanted more information about the study but was clearly in need of other kind of assistance. As standard procedure, some IRBs now require that researchers give their e-mail contact, not their phone contact.

12. Be ready for challenges. Challenges to research may come in many different guises. Some people may want to engage you in a lengthy conversation, and you may need to extricate yourself and move on. Others may criticize your research because they know something about questionnaire construction or the topic you are studying. The best advice is to remain polite. Reiterate that they are under no obligation to fill out the questionnaire.

IDENTIFYING INFORMATION

In addition to these very good suggestions from Dustin, an additional recommendation is to wear a hangtag with your identifying information, in particular your institutional affiliation. If you are collecting data in the town where your institution is located, this information will be readily recognizable (e.g., your college seal and name) and will help establish your credibility.
REVISIT AND RESPOND

- Which of Dustin’s recommendations seem to focus on permissions and which on approaches to potential participants? Of Dustin’s recommendations, list the five you think are most important and justify your choices. Add a recommendation you think he overlooked.

OTHER SOURCES OF PARTICIPANTS: THE ONLINE APPROACH

Given the limitations of subject pools and the difficulties associated with recruiting in the community, researchers increasingly seek other sources of participants. These sources include social media, online crowdsourcing platforms like Amazon Mechanical Turk, and paid panels (although such panels are somewhat expensive). An online crowdsourcing platform provides a way to obtain feedback from people (paid or unpaid) about an idea or topic. A paid panel is a group of individuals with particular characteristics who are paid to provide feedback about a product or topic.

Use of Social Media for Recruiting

Social media provide not only the opportunity for investigating their contents (i.e., studying behavior online) but also tools for recruiting. Among outlets, Facebook® has received the most attention as a recruiting tool. In their recent paper, Michal Kosinski, Sandra Matz, Samuel Gosling, Vesselin Popov, and David Stillwell (2015) outlined the strengths of the tool as well as some of the drawbacks to consider.

Among the problems they cited are the tendency for participants to rush through the survey, the lack of control over the circumstances in which participants complete the survey, the psychological distance between the researcher and the participant because they are not face to face, and the potential difficulties with communication if there are cultural or linguistic differences. Despite these difficulties, Kosinski et al. (2015) presented evidence of the high quality of data obtained using Facebook.

In using Facebook, Kosinski et al. (2015) discussed the merit of different kinds of incentives and discouraged the use of financial incentives. Instead, they said that providing feedback about performance or scores would encourage greater attention to the task and more honest responding in the case of self-report measures. This idea seems especially well suited for the kinds of research on personality they study (e.g., providing feedback about how participants’ personality profiles compare with averages). It is unclear how easy it would be to adapt this recommendation for all kinds of studies, but it is worth
considering. The basic idea is that people are likely to participate because they will learn something specific about themselves. There are also economic advantages to using this kind of informational feedback (i.e., it’s inexpensive).

**ONLINE USE OF ADVERTS (ADVERTISEMENTS) VERSUS SNOWBALL SAMPLES**

Much of the research discussed in Kosinski et al. (2015) focuses on the use of placing advertisements on Facebook (adverts; an informal British term for advertising) to attract participants, but of course, this costs money. Their recommendation is to use adverts for studies focused on difficult-to-reach populations. As an example of difficult-to-reach populations, Kosinski et al. mentioned people who are stigmatized in what is called the “offline world” (p. 545). Research using these social media platforms has its own vocabulary; people choosing to do research on these platforms need to understand the vocabulary as well as the particulars of the technology.

Students and other researchers increasingly use Facebook as a way to post studies, relying primarily on what is known as snowball sampling (as introduced in Chapter 2; additional kinds of sampling are discussed later in the chapter). In snowball sampling, those first asked to participate (usually one’s friends) in turn ask THEIR friends, and this process is repeated in a sequence. Kosinski et al. (2015) discussed the disadvantages of snowball sampling in their article (especially concerning representativeness) but concluded that the difficulties can be overcome. Their primary example was a project called “myPersonality” that originally used snowball sampling but ultimately, as a result of its popularity, “went viral,” and in four years had more than 6 million participants (p. 544). Importantly, this is not random sampling, where each person in the population has an equal chance of being selected. Your friends are probably like you and therefore create an inherent bias in the sample. But Kosinski et al. went on to argue that most samples are not representative anyway (from undergraduate psychology students to Amazon MTurkers) and that Facebook provides a reasonable “high quality” alternative (p. 545) that would minimize the disadvantages of snowball sampling, given the (a) potential size of the sample and (b) diversity of the potential population of Facebook respondents.

Addressing the use of Web-based surveys more broadly, Samuel Gosling, Simine Vazire, Sanjay Srivastava, and Oliver John (2004) dealt with the issue of the validity in such surveys. They reported good news. The samples are diverse, at least to the extent of traditional samples; the samples are not filled with highly maladjusted people; the format of the interaction is not an issue; and the results are not biased by less serious participants.
(e.g., rushing through the surveys) or by the anonymity the Internet provides (e.g., being dishonest in responding). By and large the findings are consistent with the findings from more traditional approaches.

**ETHICAL ISSUES IN ONLINE ENVIRONMENTS: THE FACEBOOK EMOTIONAL CONTAGION STUDY**

An important point Kosinski et al. (2015) made deals with ethical issues. In their view, there are few if any substantive guidelines for conducting research in online environments such as Facebook. They argued that IRBs (and federal agencies and the APA) need to begin considering what these ethical challenges might be.

One issue being discussed is whether using information from Facebook (data mining, that is, examining large datasets to discover patterns) qualifies as research with human subjects. Think back to our definition of human subjects at the beginning of this chapter (and see Chapter 4). Lauren Solberg (2010) argued that such information does not qualify as research. In her view acquiring data this way would qualify as using human subjects only if we were talking about their private information. In the event that the Department of Health and Human Services (HHS) Office for Human Research Protections (OHRP) and local IRBs determine that such use of Facebook does constitute research with human subjects, Solberg recommended that such research be considered exempt or at the most expedited (see Chapter 4). Solberg (2010) is among several researchers worried about what has been called the “mission creep” (p. 335) of IRBs (see, for example, Schneider, 2015), and she hopes that OHRP will help set guidelines to limit what she considers the overreaching of IRBs.

This issue of the degree of privacy of your data posted on social media will continue to receive attention. As an example, the research of Adam Kramer, Jamie Guillory, and Jeffrey Hancock (2014) was recently in the news. Their study involved the manipulation of the News Feed function on Facebook and retrieval of the subsequent postings. This study was a collaboration of researchers at Facebook and Cornell University in which the emotional content of the News Feed function (which is managed by an algorithm set and controlled by Facebook) was modified without users’ consent and without IRB review. **Emotional contagion** is the convergence of emotions across individuals. The research showed that emotional contagion might occur in the absence of interpersonal contact. The research question was whether the emotional content expressed via the News Feed would lead people to post content of their own that was emotionally consistent with that posted on the News Feed. The research was not considered to be under the purview of the Cornell IRB because the study was being conducted by Facebook, a company that is not bound by the Code of Federal Regulations regarding IRBs.
There has been considerable discussion about the ethics of this Facebook research (e.g., Meyer, 2014). When establishing a Facebook account, any user agrees to Facebook’s data use policy, which allows Facebook to use the information for research, among other purposes. If there had been IRB review, what might the outcome have been?

**Try This Now**

Given what you know about the work of IRBs from Chapter 4 and your own use of Facebook, (a) do you think an IRB review of the Kramer et al. (2014) emotional contagion study was needed? (b) If so, what steps for approval would have been required?

In all likelihood, the outcome would have been a waiver of the requirement for informed consent. Why? In this situation there was no evidence that users’ participation on Facebook in this situation involved more risk than they would encounter in everyday life. As this study of emotional contagion shows, the Internet is a brave new world in terms of the privacy of information. Fortunately, researchers are beginning to provide resources to help manage the privacy issues involved with Internet studies. Robert Kraut et al. (2004) provided a very nice flow chart (see Figure 1, p. 110), which may help researchers determine whether informed consent should be obtained and documented in a given Internet study. The authors also gave advice about how to manage debriefing in such studies and discussed the kinds of research unsuited to the Internet.

**REVISIT AND RESPOND**

- What are some of the major challenges that Kosinski et al. (2015) identified in doing research on Facebook? Why did they discourage using financial incentives? What did they use instead? What is an advert? A snowball sample? What does the research of Gosling et al. (2004) show about the validity of Web-based surveys? What is the “emotional contagion” study, and what ethical issues did it raise?

**SAMPLING**

The next section will cover different approaches to sampling. The “who” of your participants and the “how” of their recruitment directly impact the generalizability of your results. We want to be able to say that our sample mirrors, or is representative of, the population of interest (e.g., the student body).
Probability and Nonprobability Samples

In our coverage of Internet research, we talked about snowball sampling (see Chapter 2), a procedure for procuring participants that depends on people taking the survey and then typically forwarding the link to others. Who are these people in the snowball sample? When we ask who the participants in the study were, we are concerned with aspects of generalizability. To whom could we extend the findings? The answer to that question depends on the nature of the topic being studied, the people in the sample, and how they were selected.

Two general classifications are sometimes distinguished: nonprobability sampling and probability sampling. Snowball sampling is a form of nonprobability sampling; we can't make a precise estimate of the representativeness of those in the sample. Probability sampling is the ability to make such precise statements about the representativeness of your sample. As we talked about earlier in the chapter, there is arguably a limited degree of representativeness in the samples used in social science because so many of them rely on undergraduate subject pools. These pools offer nonprobability samples.

Nonprobability Samples: Snowball Samples and Convenience Samples

In the earlier section dealing with social media such as Facebook, we talked about snowball sampling. You encourage people who have taken your survey to send it to other people they know. These friends in turn ask others, and so on, gathering momentum (the size of the sample grows) like a snowball rolling down the hill. This type of sampling is also called chain sampling or referral sampling. Convenience samples (as introduced in Chapter 2) are samples that are easily obtainable (e.g., people studying in the library whom you approach about participating in research or the participant pool). Like snowball samples, convenience samples are nonprobability samples. Because convenience samples use those readily available, they often fail to mirror the population of interest. In the case of college students, for example, students enrolled in a particular major, such as psychology, who are part of the participant pool, may not mirror the student population as a whole in class year, gender, race, income, or a host of other ways.

REVISIT AND RESPOND

- Using the concept of representativeness, explain why an undergraduate subject pool is a nonprobability sample.
Types of Probability Sampling

In probability sampling, you can make precise estimates of representativeness. There are a variety of types of probability sampling, which will be discussed next. A core characteristic of probability sampling is the concept of representativeness.

RANDOM SAMPLING  The essence of random sampling is that every member of the population has an equal chance of being selected. If you randomly sampled students at your institution to acquire your sample, this would mean having access to every name in the population and using a random approach (e.g., picking names at random from the student database) to select the participants. This is an ideal and has been approximated through surveys that use landlines and random digit dialing to reach target participants.

A question you (and others) may ask is what happens to sampling approaches as the use of landlines declines? In research dealing with health, Stephen Blumberg and Julian Luke (2014) stated that a potential for bias exists when surveys rely on landline-based surveys because age groups vary in their reliance on wireless versus landlines. For example, in the 23–29-year-old group, 65.7% live in wireless-only households, in data from the National Health Interview Survey Early Release Program summarized by Blumberg and Luke (http://www.cdc.gov/nchs/nhis.htm). Every six months, the National Center for Health Statistics releases data on the use of wireless technology by households in the United States. This source is an important database for those trying to reach survey respondents via landlines or wireless technology and puts the issue of possible bias into perspective.

STRATIFIED RANDOM SAMPLING  One potential drawback to random sampling is that your sample might not include members who have a specific demographic characteristic of interest (e.g., a particular income level or a particular race). In stratified random sampling, you are typically making an adjustment to the results that would emerge in straightforward random sampling. If you want your sample to reflect specific subgroups within the population (representative of a particular demographic characteristic such as class year), and you want to make sure your final sample captures these characteristics, then you create these subgroups, called strata, and sample randomly within those strata. In our example of class year, there would be four strata, one for each of the class years (first year, sophomore, juniors, seniors). To perform stratified random sampling, you would need access to the demographic characteristics of interest, which might be available from an institutional researcher or registrar (in the case of an educational institution), or from city, state, or national databases, depending on the research question.

PROPORTIONATE SAMPLING  When you move to proportionate sampling or stratification, you take the additional step of making sure that the sample you draw from each
strata is proportionate to the percentage of people in that subgroup in the population as a whole. In our example of class year, the class sizes may not be equal for a variety of reasons (e.g., how many were admitted in a given year and how many transferred or dropped out). Using proportionate sampling means that you are less likely to over- or underrepresent the targeted characteristics of those in your strata. You would need access to the same kind of demographic statistics for proportionate sampling as for stratified random sampling.

**SYSTEMATIC SAMPLING** In systematic sampling, where the full population can be identified (like all students who attend a college), every $k$th element (e.g., tenth student name) after a random start is selected. Researchers sometimes use this approach with a student directory as a way to approximate a random sample. The $k$th element is known as the sampling interval and can be determined by dividing the population size by the desired sample size. Thus, if we have an institution of 2,000 students and we want a sample of 100 students, we would select every twentieth student after a random start; if we want a sample of 200 students, we would use every tenth student.

**CLUSTER SAMPLING** In cluster sampling, the sample is divided into clusters, typically based on naturally occurring groups (e.g., classrooms, hospital floors, or neighborhoods), and then a random sample of the clusters is selected. The cluster is the sampling unit, but each individual in a particular cluster is included in the analyses. As an example, you might do a study with classrooms in a school district where you don’t have the resources to include every classroom. To obtain some assessment about the performance of children in classrooms across the school district, you would pick a random sample of these classrooms.

**REVISIT AND RESPOND**

In general, explain why using probability sampling is preferable to nonprobability sampling. What makes something a random sample? How do stratified random and proportionate sampling differ? In what way is systematic sampling like random sampling? For cluster sampling, what does it mean to say that a cluster is the sampling unit?

**NONRESPONSE BIAS AND THREATS TO INTERNAL VALIDITY**

In an ideal world, your project would use random sampling and everyone selected would respond and complete the measures in your study. Life is seldom like that (never, in fact). Most of the time, a portion of your sample (more than 50% in many
studies where the researcher has no connection to the population) will simply not respond. When you divide the number of people who responded by the number of people invited, this gives you the percentage responding. In reporting response rates, even a simple formula like that (responded/invited) is typically adjusted to account for such things as surveys that were undeliverable. It is important for researchers to be clear about how their response rate is calculated. Otherwise, we have an example of a questionable ethical practice (here, reporting your response rate as higher than it was).

Try This Now

Have you ever received a survey in the mail or been contacted by telephone and not responded? As you think about your research, what level of responding would be adequate, in your view?

NONRESPONSE AND NONRESPONSE BIAS

Before proceeding, we need to distinguish between nonresponse and nonresponse bias. Nonresponse rate is simply a numerical statement of the number of people who did not respond. Nonresponse bias occurs when those who do not respond are different from those who do respond on one or more characteristics that are important to your study.

Does nonresponding automatically create nonresponse bias? No, but there are situations when nonresponding does create bias. Imagine you are looking at the relationship between where someone was raised (rural vs. urban) and the number of walks he or she had taken in the arboretum on campus and other measures of connectedness to nature. When you examine your data, you see that those who did not respond lived farther away from the arboretum than did those who responded to your survey (you could ascertain this through their addresses in the student directory). In this instance, that differential nonresponse does create nonresponse bias because the distance of the residence hall from the arboretum might affect their use of the arboretum.

When you are using a student population, the demographic makeup of the population is probably available to you through the college website (e.g., geographic distribution, gender, and percentage of students on sports teams) or through data available from the institutional researcher. By checking these data, you can see to what extent your sample mirrors the larger population, at least on those characteristics that are known. Whether nonresponse bias is present depends on the centrality of those characteristics to your hypotheses. For those wishing to learn more about estimating nonresponse bias in mail
surveys, a frequently cited article by Scott Armstrong and Terry Overton (1977) is recommended. Of the methods they discussed, comparing your results to known values of the population (such as archival information about the student body) is the easiest approach.

Many of the recommendations in Dillman et al.’s (2014) book deal with achieving the highest return rate on your survey to avoid (a) nonresponse and (b) nonresponse bias. This book is a core resource for survey researchers. Given that nonresponse bias is not that easy to estimate (Wagner, 2010), taking steps to avoid it is important.

**REVISIT AND RESPOND**

- Why is nonresponse not necessarily nonresponse bias? For basic demographic characteristics of a college sample, where could you find information about the population?

**RESPONSE RATES AND REPORTING THEM**

What do we know about rates of responding to surveys? One clear message is that return rates are dropping (Baruch, 1999; Wagner, 2010). Pew Research Center data ([http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/](http://www.people-press.org/2012/05/15/assessing-the-representativeness-of-public-opinion-surveys/)) show that whereas a response rate of 36% was found to their telephone surveys in 1997, that rate had dropped to 9% by 2012.

Yehuda Baruch (1999) distinguished between questionnaires that are returned and questionnaires that are returned and usable. He argued that it is the returned and usable return rate that should be reported. Otherwise, researchers are inflating their success.

**Mode of Delivery**

Another consideration is that response rates differ by mode of delivery, with response rates to online surveys typically lower than those to mail surveys (Tuten, 2010). Nevertheless, the response rates for online surveys may change over time in a given study. In one study (Kongsved, Basnov, Holm-Christensen, & Hjollund, 2007), the superiority of paper-and-pencil surveys was shown in terms of the initial response rate (in contrast to an Internet distribution). After a reminder, however, the completeness with which the questionnaire was filled out was significantly higher for the Internet than for the paper-and-pencil version. Furthermore, the authors argued that as the Internet becomes more familiar to people as a way to respond to questionnaires, it likely will be a more effective mode of delivery than the paper-and-pencil approach.
INCENTIVES: PRACTICAL ISSUES

Given the need to encourage people to respond to surveys or otherwise take part in research, you might ask whether incentives are effective in increasing response rates. The answer is “yes.” Chapter 4 covered the ethical aspects of offering incentives, in particular, whether they could be considered coercive. Here the focus is the practical aspects of incentives.

A point made in Chapter 4 is that there are no guidelines at the federal level that prohibit offering incentives. Often the determination of approving incentives is made at the local level in terms of (a) whether they are permitted and, if so, (b) their type and value.

Perspectives on Financial Incentives

There is evidence that small cash incentives work well (Dillman et al., 2014), but incentives may be prohibited by some departments. In a frequently cited meta-analysis on the effectiveness of incentives, Allan Church (1993) reported that initially sending rewards (either monetary, 19.1% return rate above controls; or nonmonetary, 7.9% return rate above controls) significantly increased response rates more than did promising rewards when the survey was returned (called promised incentives). The nonmonetary rewards were highly variable, including small tokens (e.g., golf balls), entry into a lottery, or in one study, a turkey! Church recommended against using rewards delayed until survey return (promised incentives) and questioned whether nonmonetary rewards are worth the return on investment, given the greater success with upfront monetary rewards. A 2015 meta-analysis (Mercer, Caporaso, Cantor, & Townsend, 2015) again showed the advantage of prepaid incentives in contrast to promised incentives in increasing response rates, particularly for mail surveys.

In the literature on incentives, one theme that has emerged is that lotteries, although often used in online research, may not be particularly effective incentives. In a chapter from a book covering online behavioral research methods (Gosling & Johnson, 2010), Anja Göritz (2010) reflects on the use of incentives. In online research, the most widely used approach is lotteries (also called sweepstakes or prize draws) perhaps because they are easier to implement than prepaid incentives given the Web modality. At the same time, Göritz stated, “Regarding the type of incentive to be used, lotteries are usually cheap but may be ineffective” (p. 228). She recommended that those with a small research budget “keep the lottery payout to a minimum” (p. 228) presumably because of its limited effectiveness.
REVISIT AND RESPOND

- In terms of responding to surveys, what do we know about the relative effectiveness of (a) incentives versus no incentives? (b) Monetary versus gift incentives? (c) Prepaid versus promised incentives?

AMAZON MECHANICAL TURK (MTURK): THE WORLD AWAITS

At the beginning of the chapter, we talked about the use of institutional subject pools, where there may be limits to the number of participants available to any given researcher. Now we will move to the opposite end of the spectrum, online sources of paid participants, limited only by your ability to pay. In particular, we will focus on Amazon Mechanical Turk (introduced in Chapter 2), which provides a market of paid (by you) online workers who are interested in participating in research. **Workers**, as they are known, sign up for a particular **Human Intelligence Task (HIT)**, which you, the **Requester**, post on the Amazon Mechanical Turk website (https://www.mturk.com/mturk/welcome) (Figure 9.1 shows the home page if you click on this link. Please note that Amazon often uses uppercase on the site to refer to Workers and the Requester, but we will use lowercase hereafter.)

**FIGURE 9.1** Screenshot of Amazon Mechanical Turk Home Page

Source: © Amazon.
Amazon Mechanical Turk (AMT or MTurk) was started in 2005 (Bartneck, Duenser, Moltchanova, & Zawieska, 2015). There are other pools of online workers, such as Incentive, oDesk, or CloudCrowd (Chandler, Mueller, & Paolacci, 2014), but Amazon Mechanical Turk may be the most well known.

As an introduction to what is known about Amazon MTurk, a paper by Gabriele Paolacci and Jesse Chandler (2014) is helpful. They described MTurk as an online labor market, with most participants coming from the United States and India. Summarizing what is known, Paolacci and Chandler stated that workers tend to be “younger (about 30 years old), overeducated, underemployed, less religious, and more liberal than the general population. . . . Within the United States, Asians are overrepresented and Blacks and Hispanics are underrepresented relative to the population as a whole” (p. 185). MTurker workers tend to be more introverted than are college student samples. Furthermore, there is some evidence that there are differences in performance on cognitive tasks, with workers learning less quickly and likely to have more difficulty with complex tasks than is true of college student samples (Paolacci & Chandler, 2014).

As Paolacci and Chandler (2014) pointed out, MTurk is not representative of the general population, but then neither is the undergraduate participant pool. Another issue they raised is that workers want to please researchers; for that reason, social desirability is an issue. One of their recommendations is to use between subjects designs (if appropriate; see Chapter 7) to avoid carryover effects (Paolacci & Chandler, 2014; see Chapter 8). Whether workers understand the task appears to make a difference in the quality of responses (Paolacci & Chandler), so it is a good idea to be very clear in what you are asking in your “HIT.” Two other steps you can take to improve the quality of data are (1) thanking workers and (2) clearly explaining the meaning of the task to be completed (Paolacci & Chandler, 2014).

Amazon MTurk (AMT) is a tool undergoing rapid change (Litman, Robinson, & Rosenzweig, 2015). Compensation is now viewed as the primary motivator, in both India and the United States. The workforce of MTurkers is estimated to be on the order of half a million individuals worldwide with an increasingly larger percentage of workers coming from India (Litman et al., 2015). In the work of Leib Litman et al., financial compensation emerged as the primary motivation for completing HITs (more than killing time, having fun, gaining self-knowledge, and doing interesting tasks). Participants from India were more sensitive to the financial compensation issues, whereas U.S. participants produced high-quality data independent of compensation. For participants from India, Litman et al. argued that quality assurance steps are still necessary to assure the level of data is high.

With regard to platforms like MTurk, the rules of the game are changing rapidly. In other words, with regard to the nuts and bolts of use, what is written now likely will be out of date by the time this book is published. For that reason, we concentrate here more
on the conceptual issues and general approach rather than on the particulars of setting up a HIT. Nevertheless, to improve understanding of the material that follows, it might be helpful to know a few basics:

- AMT offers internal templates within which to create simple HITs. If you want to conduct more complicated research using a between subjects design and random assignment to condition, your best approach is to create your survey in Qualtrics or SurveyMonkey and embed your survey link in the HIT (see Chapter 5 for coverage of online survey software).
- Most of the workers come from the United States and India. There are more complicated ways to acquire participants with particular characteristics, but in the simplest approach, you can set qualifications within your HIT that specify country of location and the percentage of HIT approvals workers must have.
- Workers are evaluated in terms of whether they finish the HIT and whether you approve their work. You can therefore specify that workers for your study have a particular level of HIT approval rating, such as 95%.
- What you pay workers is a moving target, but you are unlikely to get people to accept your HIT for a study lasting more than 10 minutes if you offer less than $0.50 (and that amount is steadily creeping upward). In research about the characteristics of MTurkers, workers now report that they are attracted to MTurk to earn money.
- If workers do not like your HIT (e.g., think it took longer than you said it would), they may post negative comments about you on MTurk list serves or forums. If you are thin-skinned, this form of data collection may not be for you.
- You will have to provide your social security number to be a requester or a worker (this presumably has to do with issues related to income earned and/or money paid).
- Amazon charges a commission fee.

Online sources of participants are a particularly attractive option for small institutions with limited participant pools, where it would not be uncommon for data collection to spread over two semesters if not into the next year to acquire the needed number of participants for a given study. Dan Johnson and Lauren Borden (2012) agreed with this perspective and saw the use of Amazon Mechanical Turk as a way to overcome the limited participant pools characteristic of small institutions. Their article provides a nice overview to the nuts and bolts of using MTurk. Michael Buhrmester, Tracy Kwang, and Samuel Gosling (2011) also encouraged the use of this tool because it provides a somewhat inexpensive way to access a more diverse pool than in traditional approaches (i.e., subject pools), and the quality of the data is as high (e.g., regarding psychometric standards) as is true of other approaches.
QUESTIONS OF VALIDITY IN USING AMAZON MTURK

A core question for researchers who contemplate using MTurk is the extent to which participants solicited from MTurk are similar to other samples you might use, such as undergraduate subject pools. Chandler et al. (2014) addressed one aspect of this question. Among the concerns are that MTurkers might be non-naïve (that is, they might be knowledgeable about the research task) in a variety of ways. For example, workers might participate in multiple related studies and thus learn something about the tasks and the hypotheses. Forums and list servers exist for MTurkers to share their views of research, particular research projects, and even particular requesters (researchers). This is another way workers might learn about the goals of research.

On the plus side, Chandler et al. (2014) pointed out that the MTurk participant pool is more diverse than you would find in academic participant pools and that the quality of data is as good, sometimes better, than data collected from traditional approaches. The quality of these data has been demonstrated in a variety of research domains (social, clinical, personality, and cognitive).

Chandler et al. (2014) provided several specific ways to decrease duplicate workers (see their article for the programming specifics). Because the sophistication of workers is growing, Chandler et al. recommended avoiding what they call commonly used experimental procedures (e.g., the trolley problem, which is used in evaluating moral dilemmas). The trolley problem (there are variations) is a hypothetical ethical dilemma problem involving the participant, who is standing by a lever in the train yard. Pulling the lever can divert the runaway trolley from its current course (which will kill five people if the participant does nothing); if the participant pulls the lever, the trolley will be diverted to a secondary track, where one person tied to that track will then be killed. The participant must decide what to do.

Another recommendation is to use attention check questions (questions that assess whether you are reading carefully; a commonly used question is “I am having a heart attack right now”; Chandler et al., 2014; Peer, Vosgerau, & Acquisti, 2014).

REVISIT AND RESPOND

- What is Amazon Mechanical Turk? What do we know about the validity of data collected on AMT and its comparability to data collected through more traditional approaches? Most workers come from which two countries? What is the primary motivation to participate in MTurk studies? Name a step you can take to assure that you will get workers to complete your task (“HIT”) with care.

Attention check questions:
questions inserted in a survey to check whether the respondent is actually reading the items (e.g., “I am having a heart attack right now”).

Copyright ©2018 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
ONLINE PAID PANELS

Online panels consist of individuals who have typically been asked to sign up to participate in research. They may be customers, patients, or people who possess desirable target demographics (e.g., high income or luxury car ownership). These individuals have known demographic characteristics and can be recruited for studies that request those characteristics. There are many companies (e.g., Research Now and Toluna) that conduct research with such panels, which are often used by market research companies. Greenbook (http://www.greenbook.org), a guide for companies that want market research, has a list of such panels on its website.

SurveyMonkey and Qualtrics also offer paid panels for conducting research. Qualtrics Panel Services uses the services of other companies, including Research Now, Toluna, GMI, Clearvoice, and SSI (Survey Sampling International), to provide panels to interested customers (in other words, they outsource this). SurveyMonkey has its own panel service called Survey Monkey Audience. Its website lists examples from $1 to $3/response, but it provides the following criteria for determining the price: number of responses, targeting criteria, and survey length. Respondents come from the people who themselves take surveys on SurveyMonkey (https://www.surveymonkey.com/mp/audience/our-survey-respondents/).

Using these panels costs money, significantly more than the cost of collecting data on Amazon Mechanical Turk. If your research requires a specific consumer segment (e.g., people interested in installing a high-end sound system in their luxury vehicle), you might expect to pay more than $50/respondent. Some advice for using such panels comes from Ron Sellars, of GreyMatter Research and Consulting, in a blog posting dealing with online panel quality (http://www.greenbookblog.org/2011/05/31/how-do-you-assure-online-panel-quality/). If you need participants with targeted demographic characteristics, you might investigate what such panels have to offer, but be prepared to spend money.

REVISIT AND RESPOND

- What are online panels? What kinds of firms use them? Why are such panels expensive? (Think about the cost of finding people with particular characteristics.)

Summary

With the information from this chapter, you have a very good idea of the samples used in research—how representative they are and where we find them. You also know how to look beyond the subject pool to recruit participants in the community or online, and the
specific obstacles that may stand in your way, especially concerning vulnerable populations. Definitions of sampling approaches have been discussed, and you know the difference between nonprobability sampling, for example, through snowballing, and probability sampling, including random sampling, stratified random sampling, proportionate sampling, systematic sampling, and cluster sampling. There is a good deal of practical advice in this chapter, from labeling your study to Dustin’s Dozen for research in the field. You have the information to make the best use of incentives (what works and what doesn’t). Amazon Mechanical Turk has been introduced as an online labor market, and you have a beginning knowledge of how it works and why it is an important source for small departments.

In the event that you did not consider them earlier, here are the REVISIT and RESPOND prompts that appeared in this chapter.

- List two potential threats to the internal validity of a study when you use a subject pool and how researchers have tried to combat these threats.
- Imagine that you are doing a study on the effect of wearing glasses on perceptions of professionalism in the workplace. Come up with a title that reflects the recommendations for labeling your study (to avoid demand characteristics). When would you schedule your study (day of week, time of day) for maximum representativeness?
- What kinds of connections do you have that would enable you to conduct research with off-campus populations? What steps would you take to safeguard your physical security and the security of your participants? Beyond the steps that are outlined in this chapter, what other ideas do you have about securing your physical safety?
- What are the benefits of conducting research at service learning sites? What are the potential drawbacks? What does it mean to be a “double agent” in the context of research? If you want to collect data from a group where you are a member, how can you avoid the possibility of coercion in requesting their participation?
- Which of Dustin’s recommendations seem to focus on permissions and which on approaches to potential participants? Of Dustin’s recommendations, list the five you think are most important and justify your choices. Add a recommendation you think he overlooked.
- What are some of the major challenges that Kosinski et al. (2015) identified in doing research on Facebook? Why did they discourage using financial incentives? What did they use instead? What is an advert? A snowball sample? What does the research of Gosling et al. (2004) show about the validity of Web-based surveys? What is the “emotional contagion” study, and what ethical issues did it raise?
- Using the concept of representativeness, explain why an undergraduate subject pool is a nonprobability sample.
- In general, explain why using probability sampling is preferable to nonprobability sampling. What makes something a random sample? How do stratified random and proportionate sampling differ? In what way is systematic sampling like random sampling? For cluster sampling, what does it mean to say that a cluster is the sampling unit?
- Why is nonresponse not necessarily nonresponse bias? For basic demographic characteristics of a college sample, where could you find information about the population?
- In terms of responding to surveys, what do we know about the relative effectiveness of (a) incentives versus no incentives? (b) Monetary versus gift incentives? (c) Prepaid versus promised incentives?
- What is Amazon Mechanical Turk? What do we know about the validity of data collected on AMT and its comparability to data collected through more traditional approaches? Most workers come from which two countries? What is the primary motivation to participate in MTurk studies? Name a step you can take to assure that you will get workers to complete your task (“HIT”) with care.
- What are online panels? What kinds of firms use them? Why are such panels expensive? (Think about the cost of finding people with particular characteristics.)