CHAPTER

Working with human subjects

15

15.1 INTRODUCTION

Research into human-computer interaction (HCI) almost invariably involves the participation of human subjects. Whether you are running a focus group, leading a collaborative design process, running a controlled study, or conducting an ethnographic investigation, you need to engage people in your work.

Although this may sound simple, it isn't. As anyone who has done so can tell you, working with human subjects involves many challenges. Finding the right subjects is often difficult and time consuming, especially for evaluation of systems designed for specific populations or situations.

The real fun can begin when the subjects are ready to begin participating in your study. Research ethics require that participants must be treated fairly and with respect. This means that they must be provided with information about the nature of the study, which they can use to make a meaningful decision as to whether or not they really want to be involved. This notion of *informed consent* is a critical component of modern research on human subjects.

Although some of the details may differ, the general challenges involved in finding and informing research subjects apply to any form of research involving human participants, regardless of the type of person involved. The additional challenges that online research presents in each of these areas are described in Section 15.3 and in Chapter 16. Although different research communities may have a preference for one or the other, this chapter uses the terms subject and participant interchangeably.

15.2 IDENTIFYING POTENTIAL PARTICIPANTS

You've just built a novel two-handed interface for an architectural modeling tool. Your design allows users to use one device in their dominant hand to draw lines while a second device can be used to pan and zoom, allowing easier and more fluid construction of lines and boundaries. Having implemented a prototype supporting these capabilities, you'd like to run some usability tests to see how well your ideas work in practice. This leaves you with a problem: who should participate in your study? There are plenty of potential users with two hands, but having the physical ability to manipulate your tool is just a start. People without the appropriate training and experience will be unable to tell you if your tool succeeds in its primary goal—supporting the work

Research Methods in Human-Computer Interaction. http://dx.doi.org/10.1016/B978-0-12-805390-4.00015-7

of an architect. Narrowing your pool of potential participants to architects would be your next logical step, but even this limitation may not be fine-grained enough. Are you willing to accept architecture students? This might help if there is a school of architecture nearby, but students may lack real-world experience. This might lead you to insist upon professional architects, who may be hard to find. HCI researchers are familiar with these and related challenges in finding appropriate study participants.

In the early days, in the late 1970s and early 1980s, many of the participants in HCI research were workers in corporate computing environments. This population of relatively early users was professionally motivated to participate in studies aimed at improving the systems that they used. As computer use spread more broadly into society and academic groups became active centers of HCI research, students became available (often just walking down the hall) and easily motivated (via cash or food) pools of participants. Countless studies involving computer science or psychology undergraduates have been published over the years.

So, what's wrong with recruiting undergraduate students—or other easily found subjects—in HCI research? Often, nothing. If you are interested in evaluating interfaces intended for use by undergraduate students, this approach is perfect. However, tests that draw on a homogeneous, nonrepresentative group of participants may be open to criticism: results may not apply to users from a different demographic group. Even if a specific menu arrangement in a word-processing program works well for (predominantly young, male) computer science students, it may not work well for retired women. In a case like this, the mismatch may simply limit the extent to which you can claim that your study answers the problem.

The number of participants is another crucial factor. Different forms of research require different numbers of participants. Studies with too few participants may not yield generalizable results, while studies with too many participants are unnecessarily expensive and time consuming.

15.2.1 WHICH SUBJECTS?

In selecting participants, you should strive to find people with *personal attributes* and *goals* appropriate for your study. By personal attributes, we mean demographic, educational, vocational, and avocational details. Some studies may simply need computer users, while others need participants of a certain gender, age range, education level, professional background, or any combination of these characteristics.

Each individual's goals, background, and motivations may play a role in determining how appropriate they are for your study. Insufficiently interested subjects may be unlikely to contribute constructively, no matter how well they match your other criteria. Even with the right physical attributes, an architect who is strongly opposed to the use of computers for modeling would probably not make a good subject for studying the architectural tool described above. On the other hand, some studies might benefit from the perspective of less-motivated participants, who might be more critical and less forgiving of shortcomings than enthusiasts. The participation of these less-motivated users can be particularly useful when studying tools that may

be used by a broad range of users in nonvoluntary circumstances, such as mandatory workplace timesheet reports. Unmotivated users can also be useful for studies aimed at understanding the factors that might influence reluctance to adopt new technology.

Expertise is always an important consideration: study participants should have expertise that is comparable to that of the expected users. We usually define expertise in terms of two largely separable dimensions: computer expertise and domain expertise—knowledge of the problems, systems, goals, and tools used in a specific line of work. If you are testing a tool that is built for highly trained professionals who rarely use complex computer applications, you'll be looking for users who may be computer novices, even though they have significant domain expertise. In other cases, you might be looking for sophisticated computer users who are using a new type of software for an unfamiliar task: computer experts but domain novices. Any differences in expertise between your target population and the participants in your study may lead to results that are hard to interpret.

Interfaces that are intended for use by a broad audience present relatively little difficulty in terms of user characteristics. General-purpose desktop computing tools and interfaces on widely used communications devices are likely to be used by many motivated users, so study participants do not need to meet many specific criteria and can often (but not always) be similar to each other.

The need for appropriate participants becomes more apparent with tools that are designed for specific populations. Children and adults have vastly different cognitive and physical abilities, which directly influence their ability to act as useful study participants. Similarly, cultural differences between users may play a significant role in task performance for communication systems. Whenever possible, studies of tools designed for specific ages, genders, social backgrounds, and physical or cognitive abilities should involve participants who fit the appropriate category. Asking college students to evaluate a tool designed for elderly users would almost certainly be inappropriate. Ethnographic studies (Chapter 9) of specific users and situations are also sensitive to the appropriateness of the participants. If study participants are not the intended users of a system, you can only make limited claims about the utility of the system for the intended population.

Systems designed for domain experts can be particularly challenging in this regard. As the construction of tools for highly specialized tasks requires a detailed understanding of domain-specific work practices, there is a natural tendency to use techniques such as participatory design to involve users in system design. This inclusion may lead to valuable insights, but domain experts who were involved in the design of a tool may have biases in favor of the resulting design, making them inappropriate candidates for subsequent usability tests or other summative evaluations.

Differences between users can also be an important part of study design. Investigations of potential gender differences in organizing certain forms of information would require both male and female participants. Similarly, an experiment exploring the role of user motivation in understanding the effectiveness of a given interface design may need participants who are highly motivated, as well as those who are not at all motivated.

Additional care is necessary when study designs require multiple groups that differ in some dimension. Ideally, the groups would differ in the relevant attribute but be comparable in all others. Any other differences would be possible confounding variables—factors that could be responsible for observed differences. In the study of gender differences in information management, the male and female groups should be comparable in terms of education, age, income, professional experience, and as many other factors as possible. If the women were significantly younger than the men, it might be hard to determine whether any performance differences were due to age or gender: further experimentation may be necessary.

Although these issues may be most important for controlled experiments, the identification of an appropriately general group of participants is always a challenge. Appropriate recruiting methods can help, but there are no guarantees. Despite your best efforts to find a representative population, you always face the possibility that your group of participants is insufficiently representative in a way that was unanticipated. As this bias is always possible, it is best to explicitly state what steps you have taken to account for potentially confounding variables and to be cautious when making claims about your results.

15.2.2 HOW MANY SUBJECTS?

Determining the number of participants to involve in a research study is a trade-off between the information gained in the study and the cost of conducting it. Studies with a very large number of participants—say, tens of thousands—probably involve many people of different ages, educational backgrounds, and computer experience. Any outcome that you see consistently from this population may therefore not be something that can be explained away by the specific characteristics of the individual participants: it is likely to be a "real" effect. Huge studies like this are particularly helpful for controlled experiments in search of statistically significant results. Even subtle differences can be statistically significant if the populations are sufficiently large.

Unfortunately, large studies are difficult and expensive to run, involving substantial costs for recruiting, enrolling, conducting the study, and managing data. If the participants are not at your workplace, there may be travel involved, and many studies pay people for their time. If your study allows you to involve many people at once—perhaps 20 people in a roomful of computers—you may be able to achieve some efficiencies in terms of the time involved. However, research that involves one-on-one interactions between a researcher and a participant may have costs that grow linearly with the number of participants.

At the other extreme, a study with one individual has very real limitations. This study would be relatively inexpensive, but also very limited. Because this study would not have a range of users with different characteristics, any results would run the risk of telling you more about the participant than they did about the research question at hand. If you're conducting an ethnographic study (Chapter 9) with one person, you may learn a great deal about how that person performs certain types of

work, but you have no idea about how representative the person's habits are: you may get unlucky and find someone who is completely unlike colleagues in the field. As studies with few participants rarely, if ever, produce statistically significant results, the conclusions that you can draw from these small studies are extremely limited.

Controlled experiments or empirical studies require a sample group of participants large enough to produce statistically significant results. The research design (the number of independent variables, within or between subjects) will play a role as well. Experiments involving larger numbers of independent variables and between-subjects (as opposed to within-subjects) experiments can require more participants (see Chapter 3). Limitations on resources can often lead researchers to substitute the feasible experiment—the design that requires fewer participants—for the experiment they'd prefer to be doing. In some cases, statistical techniques can be used to determine the minimum number of subjects necessary for a result of a given significance (Chapter 3). Usually, you want at least 15–20 participants: smaller studies may miss potentially interesting results.

The inclusion of more participants gives you more statistical power. As each participant comes with costs in terms of time, energy, and money, there are always good arguments in favor of limiting the size of the study. However, larger populations—ranging from several dozen to several hundred participants—offer the possibility of stronger statistical significance or the identification of subtle effects that would not be significant in smaller populations.

Statisticians have developed a range of techniques for determining the number of participants necessary for establishing statistically significant effects with differing degrees of confidence: Cook and Campbell (1979) is a classic text in this area. These techniques can help you understand how many participants you need *before* your study starts, thus minimizing the chances for painful problems further down the line.

By contrast, case studies and ethnographic studies (Chapters 7 and 9) can often be conducted with a small number of users. If your goal is to gather requirements from domain experts, in-depth discussions with two or three motivated individuals may provide a wealth of data. The length of the session also plays a role here: ethnographic observations generally take more time per participant—and therefore place more demands upon the participants—than controlled experiments.

Usability studies can also be successfully conducted with a small set of participants. These studies may use a combination of expert reviewers equipped with guidelines and heuristics, followed by user-based testing, to identify potential usability problems with proposed interface designs (Chapter 10). Although early work in this area was interpreted to mean that studies involving as few as five participants might be sufficient for finding 2/3 of usability problems (Nielsen and Molich, 1990), this claim has been the subject of significant debate, with more recent work suggesting that significantly more participants might be necessary for effective coverage (Hwang and Salvendy, 2010; Schmettow, 2012). User skills and background can play an important role in determining the number of evaluators needed: as evaluators with experience both in usability and in the problem domain can be more effective, fewer numbers of so-called "double experts" may be needed (Nielsen, 1992). Of course, these highly skilled participants can be incredibly hard to find and enroll.

The nature of the participants required for your study often plays a role in this decision. Studies that involve systems for general use by a broad range of users should be able to attract a suitably large pool of participants, even if hundreds of people are needed. On the other hand, research aimed at studying very specific populations may need to rely on substantially smaller pools of participants: there simply aren't tens of thousands of potential participants for the study of a tool for space shuttle astronauts. Studies of domain experts often face challenges in this regard.

Finding a suitably large participant pool can be particularly challenging for research involving people with disabilities (see Chapter 16 for more information). In addition to being an often-overlooked segment of society, people with disabilities often face significant challenges in transportation, making trips to research labs difficult. Studies with these users are often smaller, tending towards observational case studies with two or three users (Steriadis and Constantinou, 2003), rather than controlled experiments, see Chapter 16 for more details.

The time required for each participant is another important factor. Studies that require a single session of limited length (perhaps a few hours) can enroll larger numbers of participants than ethnographic observations that may involve several days or controlled experiments that require multiple sessions conducted over a period of weeks. As the time required from each participant—both in terms of direct involvement and the elapsed interval from start to finish—increases, it becomes more difficult to recruit and retain people who are willing to commit to that level of involvement.

How many participants should your study have? You should start by using your design as a guide. Ethnographies and case studies can be successfully completed with as few as two or three people. Numbers vary wildly for controlled experiments: although studies with as few as 12 users are not uncommon in HCI, results with 20 or more users are more convincing. From that base, you might expand to involve as many subjects as you can reasonably afford to include. You should then add a few more for pilot tests, replacements for participants who drop out, and a margin for error. Investigation of related work in the research literature can help in this regard: basing your population on a population used in similar prior work can be a good strategy. If there is no clearly related work, you might be able to use a smaller population, and perhaps an experimental methodology isn't the most appropriate method to start with (see Chapter 1 for more information).

15.2.3 RECRUITING PARTICIPANTS

Once you have determined *who* your participants are and *how many* you need, you must find them and convince them to participate.

If you work for a large corporation that frequently performs user studies, you may be able to draw upon the expertise of a dedicated group that maintains rosters of people interested in user studies and generates participant pools for research. Those who don't have such resources available (i.e., most of the professionals who conduct HCI studies) generally must do their own legwork.

The characteristics of your desired participants play an important role in determining how you will go about finding them. If you have relatively few constraints,

recruiting is relatively simple. Advertisements and flyers on your college, university, or corporate bulletin boards (both physical and electronic) can entice users. However, this must be done carefully: if you wish to get participants with a wide range of ages and education by recruiting on a university campus, you should be careful to explicitly recruit faculty and staff, as well as students. Notices in local newspapers and on community-oriented websites can be useful for recruiting an even broader group of participants.

More specific requirements are likely to require more focused recruiting efforts. Increased specificity in advertisements is a starting point: you might specifically indicate that you are looking for female college students. Community groups, professional organizations, and similar groups can be helpful for finding people with other, more specific characteristics. Many of these groups will be willing to pass along messages to members, particularly if the research may be of interest to them. If you can find a group of people that meet your specific needs, it may help to go to them. If you can give a short presentation at a meeting and make yourself available for questions later, you may encourage otherwise reluctant people to participate. Email lists and online groups can be helpful in this regard as well, but these tools should be used carefully: sending out messages that don't comply with the policies of the posted group or lists is inappropriate. Sending unsolicited email messages directly to individuals is almost certainly a bad idea. Although an email message that comes from a trusted mailing list might be well received, the same message sent directly by an individual might be seen as annoying junk email.

Focused ethnography and long-term case studies require fewer subjects, but the effort involved in enrolling each participant may be greater. These projects may require building cooperative arrangements with companies, schools, other organizations, and individuals in order to identify appropriate subjects. Many academic researchers address these challenges by bringing in outside organizations as collaborators. In addition to creating a formal agreement, collaboration can also provide funds that support the efforts of the cooperating organizations.

Incentives can often motivate people to participate. Many undergraduates have been lured into research sessions by promises of cash or pizza. If you can pay your subjects for their time, do so. Gifts can be more appropriate for some participants—particularly children. If you don't have enough funds to pay all participants, you can offer to enter them in a raffle for a desirable prize. Compensation can also be a motivator that can elicit desired behavior: in one study on interruption, researchers asked participants to both complete a memory task and respond to interrupting signals. In order to entice participants to complete both tasks, extra payment was given to the subjects with the best performance (Gluck et al., 2007). Incentives for organizations that assist in recruiting can also be useful. In addition to the research collaborations described above, you might pay groups as consultants (see the Menu Task Performance Studies with Blind Users sidebar for an example).

Although financial and other incentives are routinely used to encourage participation in research studies, it is certainly appropriate to consider the potential impact that prospects of financial gain might have on participant behavior. Researchers have

long known that participants may examine "demand characteristics "—trying to provide responses that they think will please the researcher (Orne, 1962). Although careful researchers will always be on the lookout for opportunities to reduce potential sources of bias, when working with incentives it might be particularly important to stress to participants that they will be rewarded regardless of the answers they give.

MENU TASK PERFORMANCE STUDIES WITH BLIND USERS

Task performance with hierarchical menus has been the subject of many studies over the years, leading to a general consensus that menus with many choices at each of a few levels (broad, shallow trees) lead to faster task completion than menus with a few choices at each of many levels (narrow, deep structures), see Chapter 1. As these studies have generally been conducted with sighted users, who could rely upon a visual scan to quickly identify items in a long list, we were interested if these results would hold for blind users who rely upon the serial presentation of items by screen readers. To address this question, we designed a study based on an early experiment that looked at breadth versus depth in web-based choices from an encyclopedia (Larson and Czerwinski, 1998).

Experimental studies involving blind people can be particularly challenging to run. As blind people often face challenges in transportation, asking users to come to our university offices would be inconvenient. We also knew that we wanted a particular population: experienced users of a particular screen-reader package, who did not have any residual vision.

For this research project, we collaborated with the National Federation of the Blind (NFB), who helped identify potential participants and provided us with access to space in their offices, where we were able to run the study. NFB was paid as a consultant on the project and study participants were compensated as well. Due to the specific nature of the participants, compensation was significantly higher than is customary for similar studies. With NFB's help, we were able to recruit a sufficient number of qualified participants, and we found that, like sighted individuals, our blind participants fared better with broad, shallow menu structures (Hochheiser and Lazar, 2010).

Compensation should be commensurate with the amount of time requested and the type of participants involved. Busy professionals may command a higher fee than students or children. For longer ethnographic or case studies, particularly with domain experts, direct payment for study participation is unlikely to account for the value of their time. In these cases, finding ways to pay experts as consultants may be the best approach. For formative studies aimed at capturing requirements for systems to be used by domain experts, the ability to use the software being developed in their daily work might be a powerful enticement.

Special populations may require creative incentives and accommodations. If you are working with children, you might give them small toys as gifts for participating (cash compensation for accompanying parents is probably always welcome). Elderly people or others without easy access to transportation may be interested in participating but may be unable to make the trip to your lab or office. You might consider trying to conduct your study in participants' homes, community centers, or other locations that would be easy for interested participants to travel to.

Some studies may have additional requirements that require screening of interested participants to determine whether or not they meet important criteria. For example, tools designed for novices should probably not be evaluated by people who work professionally with similar interfaces. Initial questions and interviews with potential subjects can be important tools for ensuring that an individual is appropriate for your study. Specific questions about education, age, experience, and other important attributes can be asked to verify that there is indeed a good match. If you take this approach, you might also consider asking whether they are willing to be contacted in the future for subsequent studies. People who agree to future contact can form the basis for a homegrown database of study participants. Maintaining such a database may involve a fair amount of work, but it can be potentially very useful if you plan to run many studies.

Your database of potential subjects can be an important safety net in the event of difficulties along the way. You may start out with 15 (or 20, 30, or 60) participants with confirmed appointments, only to find that several cancel at the last minute or simply fail to show up. Other problems associated with participant characteristics may force you to dig deeper for a wider range of ages, skills, or backgrounds. If the participants in your study of a general-purpose tool for managing personal photos are all men between 35 and 40 years old (or women over 60), you might have a hard time arguing that your results are indeed generalizable. It's easy to argue that better planning and participant screening might help with this problem, but such details are often not obvious from the beginning. If you're faced with this dilemma, your best option might be to dig deeper into your list, inviting more participants to form a larger (and hopefully more representative) study.

Experiments that involve multiple experimental conditions may require reanomizing participants into roughly equal-sized groups. If you are comparing performance across user attributes—such as age or gender—your groups must differ in the relevant attributes, while remaining as comparable as possible for other characteristics. If your potential pool of participants is large, you need to select participants in a manner that minimizes any potential bias in selection: selecting the first names from a list that is sorted by gender may get you a group of subjects that is entirely male or female. See Chapter 4 for more discussion of these and related issues in population sampling.

15.3 CARE AND HANDLING OF RESEARCH PARTICIPANTS

Studies with human participants put researchers in a privileged position. As “scientific experts”, researchers have expertise, experience, and contextual knowledge that

make them well equipped to understand the reasons for conducting the experiment and the potential costs and benefits involved in participation in a study. Potential participants may lack some or all of this relevant background.

Research studies should be designed to protect participants. Informed consent—the notion that research participants should be provided with the information needed to make a meaningful decision as to whether or not they will participate—is the cornerstone of this protection. Academic and industrial organizations that conduct research with human subjects generally rely on institutional review boards (IRBs) to review proposed research for any possible risks and to guarantee that appropriate procedures for informed consent are being followed.

15.3.1 RISKS AND CONCERNS OF RESEARCH PARTICIPANTS

Participation in a research study involves multiple agreements between the participant and the researcher. The participant agrees to perform certain tasks as needed by the experiment and the experimenter frequently agrees to provide some incentive or compensation to the participant. Perhaps more importantly, experimenters agree to conduct responsible research that protects participants' rights, health, privacy, and safety.

Risks to participants are often most pronounced in medical research, where investigation of new drugs, devices, and procedures can lead to health risks, particularly when things don't work as intended (or hoped). However, physical harm is not necessarily the only relevant concern. Famous psychology experiments have shown how research that places people in uncomfortable situations can cause significant emotional distress (see the Milgram's Experiment and Stanford Prison Experiment sidebars). Although some HCI experiments might raise these concerns, most of the studies in our field are low risk. Some studies may lead to fatigue (from mouse movements) or eye strain, but these risks are minor. Regardless of the level of risk involved, researchers must treat human participants with dignity and respect.

MILGRAM'S EXPERIMENT

Perhaps the most famous example of deception in psychology research, Stanley Milgram's obedience experiment illustrates one possible extreme associated with research on human subjects.

In this study, subjects were told that they were participating in a study of the effect of punishment on learning. They were asked to administer tests to another subject—a “learner”—who would have to identify a word that had previously been associated with a stimulus word. Subjects were told that they had to

administer an electric shock to the learner if incorrect answers were given and that the voltage of the shock should be increased after each incorrect answer. Shocks were described as being "extremely painful," but incapable of causing permanent damage (Milgram, 1963).

This description was an elaborate deception aimed at concealing the true goal of the experiment: a study of the limits of obedience. As the "learner" was in fact a colleague of the experimenter's, no actual shocks were administered. However, the subject did receive a mild shock to provide evidence of the authenticity of the equipment and the learner acted as if shocks had been applied. The experimenter participated actively in the deception, urging subjects to continue with the experiment even when they expressed reluctance.

The results of the study were intriguing: of 40 participants, all continued giving shocks until after the point where the "learner" kicked on the wall and stopped responding to the test questions. Most (26 out of 40) of the participants administered the maximum level of shock—two steps beyond "Danger: Severe Shock." Participation caused discomfort including nervous laughter, embarrassment, and seizures for several subjects.

This experiment would not have worked without deception: had the subjects known that they were not actually administering potentially painful shocks, they presumably would have been even less reluctant to participate. The deception created a scenario in which obedience had a real cost, in terms of the distress associated with inflicting harm on a fellow human being.

Milgram's experiment would not be considered appropriate human subjects research in most current research environments. The extreme nature of the psychological distress involved in these experiments and the strong reactions experienced by some of the participants raise serious questions as to whether such research can ever be conducted responsibly (Milgram, 1963).

Virtual environments provide interesting possibilities for subsequent investigations of similar phenomena without raising the ethical concerns associated with Milgram's experiment as originally executed. In a "virtual reprise" of those experiments, subjects were asked to administer shocks to a female virtual human in an immersive environment. The use of a computer-generated character eliminated the need for deceit, thus removing some of the possible ethical objections. Although participants knew that they were interacting with a computer-generated avatar, they responded to the situation as if they were working with a real person, particularly if they could see the avatar (as opposed to communicating via a text chat interface) (Slater et al., 2006).

THE STANFORD PRISON EXPERIMENT

Many interesting and important questions about human behavior in difficult situations can only be examined by conducting studies that expose participants to the risk of significant psychological distress. As interesting as these questions may be, the risks are substantial enough to make this research effectively off limits.

The Stanford prison experiment, conducted by Philip Zimbardo and his colleagues during the summer of 1971, provides an example of both the risks and insight potentially associated with research that exposes participants to significant emotional distress. In order to examine the social forces associated with prisons, the researchers divided a group of Stanford undergraduates (all males) into "guards" and "prisoners." Prisoners were arrested at their homes, blindfolded, placed in uniforms, and incarcerated in a makeshift prison constructed in the basement of Stanford's psychology building. Guards were not given training—they were simply told to do what was necessary to maintain order.

The researchers and participants were all surprised by their responses. Both guards and prisoners completely fell into their roles. Guards humiliated prisoners, using tactics such as awaking prisoners throughout the night for "counts" and placing people in solitary confinement to establish their authority and prevent rebellion. Prisoners temporarily lost their personal identity, thinking of themselves only by their prisoner number. They were passive, depressed, and helpless. One prisoner suffered significant stress, including crying and rage. Both the guards and the researchers responded like real prison staff, believing that he was faking. Dr. Zimbardo—the professor in charge of the experiment—found himself acting like a prison warden, bristling at concerns for the well-being of the prisoners—who were, after all, innocent bystanders. Originally planned for 2 weeks, the study was terminated after six days, out of concern for the participants (Haney et al., 1973; Zimbardo, 2008b).

The observation that seemingly ordinary people would quickly assume the role of sadistic prison guards raises serious questions about the role of context in determining human behavior. Although we would all like to think that we would not behave abusively in such contexts, the Stanford Prison Experiment raises the concern that environment and expectations can play a huge role in encouraging seemingly inhuman behavior. This lesson continues to have significant relevance, through explorations including Kyle Alvarez' 2015 film *The Stanford Prison Experiment* and Philip Zimbardo's outspoken commentary on the behavior of guards at the Abu Ghraib prison in Iraq (Zimbardo, 2008a,b).

The Stanford prison experiment also provides a cautionary tale regarding the evolution of research ethics. Despite the known potential for harm, this

study was approved by Stanford's Human Subjects Review Board, participants signed an informed consent form, and a 1973 review from the American Psychological Association determined that the study had been consistent with existing ethical guidelines (Zimbardo, 2008b). Changing views on responsible research—influenced at least in part by this—have led to a much more conservative view of appropriate research. Philip Zimbardo publicly apologized for his role in the study (Zimbardo, 2008b) and the establishment of beneficence—maximizing of benefits while minimizing harm (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979)—argued for research that would strive to avoid the harms seen in the prison experiment. It is hard to imagine a study with this degree of potential harm being approved by any modern IRB.

Specific definitions of the responsibilities of researchers grew out of concerns about inappropriate medical procedures conducted during the mid-20th century (see the Informed Consent: Origins and Controversies sidebar). In 1979, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research published the Belmont Report (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). This document established three principles for the treatment of research participants: respect for persons, beneficence, and justice. Respect for persons involves allowing individuals to make independent and autonomous decisions regarding their participation in research. Researchers must allow participants to make judgments and must provide the information necessary for making those judgments. Special consideration must be given in cases of illness or disability that may limit an individual's ability to make independent decisions. Beneficence refers to the need to minimize possible harm while maximizing possible benefits. Justice requires that neither the burdens of participating in research nor the benefits of the research should be limited to certain populations, particularly when some groups of people may be easily manipulated (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979). These principles form the basis for informed consent.

INFORMED CONSENT: ORIGINS AND CONTROVERSIES

Famous (or infamous) medical research experiments conducted during the mid-20th century led to the development of modern concepts of informed consent and appropriate treatment of research participants. Nazi Germany's use of concentration camp prisoners in often brutal and barbaric medical experiments led to the Nuremberg code, which established some of the principles behind informed consent.

(Continued)

INFORMED CONSENT: ORIGINS AND CONTROVERSIES—CONT'D

The US Public Health Service Syphilis Study at Tuskegee involved hundreds of black men with syphilis over 40 years. Although they were told that they were being treated, no treatment was in fact given, and efforts were actively made to prevent participants from getting treatment (Centers for Disease Control and Prevention, 2007). Several other studies in the US involving administration of drugs or treatment without consent were conducted in the US after the end of World War II (Pellegrino, 1997). More recently, drug trials conducted by Western companies in countries such as India have raised concerns about the nature of informed consent across such cultural and financial divides (Sharma, 2005).

The costs associated with these studies are not limited to the substantial harm inflicted upon the subjects. These unethical experiments reflect poorly on science and scientists in general, harming public trust and increasing reluctance to participate. One study of both white and black residents of Detroit found that black residents were more likely to have heard of the Tuskegee experiments. They were also more likely to be distrustful of researchers and less likely to participate in research (Jones, 1993; Shavers et al., 2000).

15.3.2 PROTECTING PRIVACY

Participants should also be assured that their privacy will be protected. Researchers should obtain consent for the collection and storage of personal information; limit the information collected to that which is necessary; identify the uses that will be made of any information; limit the use, disclosure, and retention of the information; securely protect any information; disclose policies and procedures; provide a means for addressing concerns regarding compliance with information practices; and be accountable for those practices. Patrick provides a set of more than 30 questions suitable for addressing adequacy of practices in each of these areas (Patrick, 2007b).

The use of photography and video or audio recording presents special challenges regarding the privacy of participants. Photos, videos, and audio recordings can be very useful tools for illustrating the use of an interface, but they can also unambiguously identify individuals as having participated in a research project. There are several steps that you should take in any project before you start the shutters snapping or cameras rolling. You should clearly tell participants what you are recording and why. If you are going to consider using images of participants in any publications or reports, participants should be fully informed of this possibility. These practices should be mentioned in your informed consent forms (Section 15.3.4) and discussed with participants. If you are video recording, you might consider recording a portion of the discussion, taking care to include footage of the participants explicitly agreeing to be video recorded. You should plan your photos or videos carefully: if you are really interested in what is going on with the interface, take pictures and video of

the inputs and display—not the faces of the participants. You might be able to shoot over the users' shoulders to get a fuller view without identifying your participants. Similarly, audio recordings captured for potential distribution should minimize use of the participant's voices—record the voices of the research staff if necessary.

Although data minimization may limit risks to participant privacy, the associated loss of detail may not be acceptable in some circumstances. Studies of the clinical use of electronic medical records have used audio, video, screen capture, and related techniques to collect rich records of the technical and interpersonal aspects of the use of these tools in practice (Asan and Montague, 2013; Weibel et al., 2015). Given the particularly sensitive nature of medical data, such records should be made carefully, protected through encryption and appropriate security, and only used by authorized research staff members.

More generally, data storage and backup choices should also consider participant privacy. Research data storage should preserve both privacy and availability of data, particularly given increased mandates for data sharing. Storage on local hard drives or on digital media stored in locked file cabinets can often be best for protecting privacy, with somewhat reduced availability. Although cloud-based services may offer easier data sharing, privacy protections may be weaker. However, cloud providers are increasingly offering services with higher levels of security and access control. If you have any concerns, check with your IRB or research office.

Appropriate choices in dissemination—particularly in publishing—can also help protect participant's privacy. If you must show people in action, you might consider using image-manipulation techniques, such as blurring or black bars over the eyes to hide the identity of the participants. Pictures or videos of the research staff might be more appropriate for distribution. Finally, you might provide an alternative for participants who are concerned about their privacy: you might not need video or audio recordings of every individual in your study. Case studies and interviews with participants might take care to ensure that descriptions, quotes, and other details are presented in a manner that removes all materials that might jeopardize participants' well-being, health, and livelihood.

15.3.3 INSTITUTIONAL REVIEW BOARDS

Universities, hospitals, corporations, and other organizations that conduct research often have standing committees that review and approve projects involving human subjects. These IRBs examine proposed studies for appropriate practices, procedures, goals, and disclosures. By conducting this review prior to the start of research with human subjects, IRBs protect all of the groups and individuals that may be affected by the research. Participants are protected by examination of proposed research for any elements that may be manipulative, coercive, or otherwise abusive. Proposals that contain any such elements should not be approved by IRBs. Researchers and institutions benefit from the knowledge that the proposed research has been reviewed for issues that may cause embarrassment or legal liability. Although this review is certainly not foolproof, it generally works well in practice.

IRB review and approval for proposed research generally begins when a researcher submits materials relating to proposed research. A description of the proposed research, draft informed consent forms, instructions to be provided to users, questionnaires, and materials to be used during the course of the research are some of the items that might be required. Upon receipt of these materials, the IRB will review them for completeness and content. The board may approve the research, request additional information, require revision of materials, or take other steps as appropriate.

As research cannot begin until the IRB approval is complete, it is generally best to start this process early. Some research funding agencies will not release any funds until appropriate IRB approvals have been obtained. As each IRB has its own rules, it is important that researchers understand and follow the appropriate procedures for their institution. Many IRBs have websites that describe policies and provide relevant forms. It is a good idea to familiarize yourself with this material. Although some boards consider applications on a rolling basis, others have scheduled meetings, with published submission deadlines for consideration at each meeting. Attention to detail is particularly important for boards that meet on a set schedule: if your IRB meets bi-monthly, minor omissions in a proposed package may lead to a 2-month delay in acquiring the necessary approval.

Some IRBs—particularly those at large research institutions with affiliated medical schools—may spend much of their time focusing on drug or treatment studies. If your IRB falls into this category, board members may not be aware of the techniques used in HCI research (and described in this book). You may have to spend some time and effort explaining ethnography, research based on online data sources, or other techniques that they are not familiar with. If you run into this sort of challenge, you should stress the widespread application of these techniques, and the existing body of research from groups such as the Association of Internet Researchers (http://www.aoir.org) or the Ethnographic Praxis in Industry Conference (EPIC, https://www.epicpeople.org/). It's best to approach such discussions from a collegial, not confrontational, perspective.

IRB policies for US government and government-funded institutions (including almost all universities and colleges in the United States) are dictated by the Federal Policy for Protection of Human Research Subjects, otherwise known as the *Common Rule* (45 CFR, part 46) (Department of Health and Human Services, 2009). The Common Rule describes requirements for institutional review, including categories of studies that can be given expedited review or exempt from review altogether. Studies that both involve "minimal risk," and fall into one of nine specified categories can be candidates for expedited review. Although most of these categories are specific to biomedical research, categories involving collection of data from recordings made for research purposes and research involving group characteristics or behavior are directly of interest to HCI studies. Similarly, studies involving educational strategies; test surveys, interviews, or observations of public behavior; and studies of existing data or research data can be considered exempt from full IRB review (Department of Health and Human Services, 2009). Generally, institutions will have specific sets of forms for studies that require full IRB review, expedited review, or exemptions, with expedited review often quicker than full review. Studies that qualify as exempt must

still be reviewed by IRB staff, who will then provide documentation indicating that the study is, in fact, exempt.

Although the paperwork required by some IRBs may feel like a nuisance, you should consider your IRB as an ally. By insisting upon procedures, IRBs protect researchers and institutions from problems associated with research that goes wrong. IRBs can also provide helpful feedback in situations that may raise questions. Some projects may blur the lines between participating in the research and acting as a collaborative partner. For example, projects involving participatory design may involve ethnographic observation of users in the workplace. Is informed consent necessary in this case? Although the conservative approach of requiring informed consent is unlikely to be inappropriate, discussing this question with a member of your IRB might provide insight into your institution's policies regarding such research. Many IRBs require researchers to take training courses before conducting any studies involving human subjects research. These courses may not seem exciting, but they can provide valuable information that might prove helpful when you are preparing informed consent materials.

Organizations that infrequently engage in human subjects research may not have an established IRB. This may be particularly true for small companies that run occasional user studies. If you find yourself in such a situation, it may be helpful to discuss matters with appropriate professionals in your organization, including community relations staff and legal counsel. IRBs from nearby research institutions may be willing to provide feedback as well. The use of informed consent forms and proper procedures is always appropriate, even in the absence of a formal review from an IRB.

Researchers and regulators share a common interest in optimizing IRB review requirements and procedures. In the US, a 2011 Advanced Notice of Proposed Rule-Making (ANPRM) suggested significant changes to IRB procedures, including the creation of an "IRB of record" provision that would allow one review for all participants in a multi-institution study, as opposed to separate review for all institutions, as is the current process. Other changes proposed in this ANPRM include changes to methods for determining risk, new categories of studies designed to minimize review requirements for low-risk studies, and revised rules for determining when informed consent is not required, among others (Cohen and Lynch, 2014). Further discussion of these issues included a "request for information" regarding the use of single IRB review for multi-institution studies funded by the National Institutes of Health (National Institutes of Health, 2014). A Sep. 2015 notice in the US Federal Register revived the federal review proposals with a Notice of Proposed Rule-Making (NPRM) in response to the 2011 ANPRM. The proposed rules retain many of the changes that might facilitate HCI research, including the single IRB proposal and the possibility of defining "certain types of social and behavioral research conducted with competent adults" as exempt from IRB review (Federal Register, 2015). There is widespread agreement that some reform of these rules is needed. However, the complexity of the regulations and the 4-year gap between the ANPRM and the NPRM suggest that acceptance of these proposals and implementation of any changes may be a gradual

process (Schrag, 2015; Cohen and Lynch, 2014). A Jun. 2016 report issued by the US National Academies, raised multiple concerns about the proposed rule-making, suggesting that it be withdrawn and that a national commission on human subjects research be established (National Academies of Sciences, 2016).

15.3.4 INFORMED CONSENT

The notion of informed consent has two parts. "Informed" means that study participants must understand the reason for conducting the study, the procedures that are involved, potential risks, and how they can get more information about the study. Without this information, participants do not have the information necessary to make a truly meaningful decision as to whether or not they wish to participate. If potential participants are not told that the use of a specific virtual-reality environment can occasionally cause nausea, particularly sensitive individuals may agree to participate without being aware that they might be subjecting themselves to an unpleasant experience. For these reasons, researchers should strive to clearly provide information that is relevant and necessary for appropriate decision-making. Truly informing potential participants means that the information must be provided in a manner that is comprehensible. The reason for the study, the procedures being used, and other details should be provided in a manner that is clear, accessible, and free from professional jargon.

The second, equally important notion is "consent": participation in research studies should be entirely voluntary and free from any implied or implicit coercion. Potential participants should not be given any reason to believe that a decision not to participate will lead to repercussions or retaliation, whether in the form of punishment by employers; withholding of medication or the use of a system; or disapproval from the researcher. Researchers in academic settings should be very careful about giving students credit for coursework in exchange for their participation in studies: if an alternative means of earning the credit are not provided, some students may feel that their grades will suffer if they decline to participate. In such circumstances, participation would be coerced, not consensual.

In most cases, researchers provide participants with an informed consent document that contains several sections (Office for Human Research Protections, 1998):

- *Institution and Researcher Identification:* Who is responsible for the research? Specifically, which individuals and institutions are conducting the study?
- *Contact Information:* Who should participants contact if they have questions or concerns? This section should contain names and contact information for the researchers in charge of the study, as well as for representatives of the IRB or other appropriate body.
- *Title and Purpose*: Why is the study being done?
- *Description of Procedures:* What will be asked of participants? For HCI studies, this probably involves using one or more interface variants, discussing goals and needs, commenting on design proposals, and other related tasks.

- *Duration:* How long will each participant be involved in the study? This should tell the user how much time will be involved. If there are multiple sessions, the number of sessions, the length of each session, and the elapsed interval required should all be specified.
- *Risks:* What risks might be involved in participation? Medical trials may involve the risks of unknown drug side-effects, but the risks are generally less severe in HCI studies. Fatigue, boredom, and perhaps slight discomfort due to repetitive motion are possible risks for studies involving desktop computers. Virtual-reality systems may involve some risk of nausea or disorientation. Studies involving mobile devices, computers in cars, or other interfaces in nontraditional settings may involve additional health or safety risks. Evaluation of the potential distractions caused by computing devices in cars should probably not be conducted in cars driving on public roads! Other interfaces involving social interactions may pose emotional risks, if tasks or content may prove upsetting to participants (see Milgram's Experiment sidebar). The privacy risks of photography and video or audio recording are discussed above; projects involving online conferencing or ongoing use of online chat systems may present similar concerns. Experimenters should, of course, design studies to minimize all risks. Any remaining risks should be described in detail in informed consent forms and then discussed honestly and thoroughly with study participants.
- *Benefits:* What are the benefits of participation? Some researchers may provide participants with ongoing access to software that is being evaluated. In other cases, financial or material compensation is the main benefit.
- *Alternatives to Participation:* What other options are available? For most HCI studies common alternatives include simply not participating, opting out of the study at any point in time, and continuing to use the software that was being used before the study.
- *Confidentiality:* Participants' privacy should be respected. This section of the form generally includes comments indicating that personally identifying information will not be used or published in any way. Confidentiality is a particularly important issue for HCI research involving observation of user behavior such as search or information use activity. Web search, email organization, and other activities may reveal sensitive personal information that could compromise confidentiality. Proper protection of participant privacy involves limiting the use, disclosure, and retention of data; taking appropriate measures to protect data, including encryption and secure storage; openly describing policies and practices; providing avenues for challenging compliance with data protection procedures; and providing for training and related measures to ensure accountability (Patrick, 2007b).
- *Costs/Additional Expenses:* Are there any financial expenses or other costs associated with participation? Although such costs may not be inappropriate, they may discourage some users from participating. If you are going to ask participants to make costly trips to travel to your location, to purchase software

for their computer, or to spend significant amounts of time entering data into diaries, you need to make sure that they are aware of these costs.

- *Participant's Rights:* This section should make three important points:
  - Participation is voluntary.
  - Participants can choose to stop participating at any time, without penalty. The informed consent form should describe what will be done with data for participants who withdraw.
  - Participants have the right to be informed of any new information that will affect their participation in the study.
- *Supplemental Information:* Where should participants go for further information? This section should list resources that can be used for additional information, including (but not limited to) descriptions of the research program and institutional policies and procedures for research involving human subjects.
- *Signature:* Participants should sign a copy of the consent form. The signature should be accompanied by a statement indicating that the participant:
  - has volunteered to participate;
  - has been informed about the tasks and procedures;
  - has had a chance to ask questions and had questions answered;
  - is aware that he/she can withdraw at any time;
  - consented prior to participation in the study (Shneiderman et al., 2016).

The researcher should provide a copy of the consent form to each participant for reference, while retaining the signed copies as documentation of the consent.

Construction of an informed consent document can be a useful step in ensuring that your research meets accepted ethical standards. If you have accounted for the risks, benefits, alternatives, and confidentiality measures associated with your project, the relevant sections of the document should be relatively straightforward to be put together. Similarly, difficulty in construction of these sections may indicate the need to rethink proposed practices in procedures.

Writing clear, concise informed documents is not trivial. One study of informed consent forms for medical research studies found that users preferred simpler statements written at a seventh-grade level (as opposed to at a college graduate level) but the simpler statements did not lead to greater comprehension (Davis et al., 1998). Pilot testing of the consent forms, either as part of a pilot test for an experiment or via reviews by potential participants or collaborators can help identify confusing language or areas that may need clarification. A sample informed consent form is given in Figure 15.1.

Informed consent requires affirmative agreement from an individual who is capable of understanding the implications of agreeing to participation in the research. Research involving participants who are not able to interpret informed consent forms may require additional measures. Requirements and procedures regarding assent for children's participation in research are described in Section 15.3.6.3, while issues relating to participants with disabilities are discussed in Chapter 16.

Local or national legislation may place additional constraints on the content of an informed consent document. In the United States, federal regulations prohibit

INFORMED CONSENT FORM

Evaluating Menu Selection Task Performance

PRINCIPAL INVESTIGATOR: A. Researcher
Department of Computer and Information Sciences
Research University Phone: 555-555-5555
Email: researcher@research.edu

Purpose of the Study: The goal of this study is to understand how computer interfaces might be customized to best suit the needs of users. Participants will be asked to use a menu interface to find items in various multilevel hierarchy designs. Task completion times and subjective responses will be used to determine which (if any) design is most suitable for these users.

Procedures: Participation in this study will involve two phases. In the first phase, you will be asked to use a web browser to make selections from a menu of choices, in order to locate a specified entry. You will be given the opportunity to try a sample task, and then you will be asked to complete multiple tasks with different menu structures. This study should take about one hour to complete.

After you have completed the experimental tasks, we may ask you some questions about the various interfaces. These questions will be designed to help us understand which (if any) of the interfaces you preferred, and why. We may also ask some general questions about your habits and practices with respect to computer use.

Risks/Discomfort: You may become fatigued during the course of your participation in the study. You will be given several opportunities to rest, and additional breaks are also possible. There are no other risks associated with participation in the study. Should completion of either the task or the interview become distressing to you, it will be terminated immediately.

Benefits: It is hoped that the results of this study will be useful for the development of guidelines for the design of user interfaces that will help people use computers more effectively.

Alternatives to Participation: Participation in this study is voluntary. You are free to withdraw or discontinue participation at any time.

Cost and Compensation: Participation in this study will involve no cost to you. You will be paid for your participation.

FIGURE 15.1

Informed consent form.

language in informed consent forms that would waive legal rights or absolve researchers of legal responsibility. The use of informed consent forms—even those that are approved by IRBs (see Section 15.3.3) should not be seen as a green light to move forward with research that may otherwise raise questions regarding respect for the rights and concerns of participants.

Evolution of the research landscape often spurs innovation in informed consent practices. Complexity is often a challenge for studies that may involve some risk to participant privacy, as long consent forms laden with technical information may be difficult to understand. Concerns about the complexity of detailed consent form

for research projects involving genetic studies of biological samples have led to the development of simplified consent forms (Beskow et al., 2010) and studies of comprehensible phrasing for relevant content (Beskow et al., 2015). HCI researchers tackling deceptive studies have struggled with similar challenges (see Section 15.3.6.1). The growth of ubiquitous computing research, with many studies of contextualized computer and device use or interaction in use in homes, schools, and other public places has led some to suggest that consent should be more nuanced and contextualized than a simple blanket agreement to the terms of a research study (Luger and Rodden, 2013a,b).

15.3.5 RESPECTING PARTICIPANTS

The Belmont report describes respect for persons in terms of ensuring that participants are able to make independent and informed decisions about their involvement in research studies. Although this is perhaps most directly applied through the informed consent process described above, truly respecting participants requires consideration of their needs, concerns, and values throughout all aspects of designing the study, conducting the research, and publishing the results.

15.3.5.1 Study design

The experiments conducted by Milgram and Zimbardo arguably arose at least in part because their interest in their research question overshadowed concerns that they may have had about the impact that the research would have on the participants. Although these studies would clearly be considered unethical by current standards, controversies regarding the impact of research on participants still rage. Examination of these debates can shed some light on the challenges raised by some HCI research projects.

A newsworthy study published in July 2014 by researchers from Cornell University and Facebook (Kramer et al., 2014) provides a textbook example of how these concerns might arise in modern HCI research. In order to understand the impact of "emotional contagion", these researchers worked with Facebook to manipulate the presentation of items on users' news feeds. Over the course of 1 week in Jan. 2012, researchers adjusted news feeds, decreasing either the amount of positive or negative emotional content, as determined by the inclusion of words previously shown to be correlated with measures of well-being. Examining news feeds for a large (almost 700,000) group of Facebook users, they found that reductions in positive posts appearing on a users' feed were associated with reductions in positive content in that individuals postings, with comparable effects for reductions in negative content (Kramer et al., 2014).

The Facebook paper raised a firestorm of research ethics controversy immediately upon completion. An "Editorial Expression of Concern" (Verma, 2014) published alongside the paper noted the major concern: the research was conducted without explicit informed consent from participants. The study was consistent with Facebook's data use policies, which describe creation of an account as implicit agreement to

participate in research. Furthermore, Cornell University's IRB determined that the study was not under their jurisdiction, as it was conducted by Facebook. As a result, participants were not informed of their participation, which the editor considered to be "not fully consistent with informed consent" (Verma, 2014). Reports of user concern quickly spread throughout the new media (Goel, 2014), as users complained that they may have been manipulated without their knowledge or consent. Subsequent soul-searching in the academic literature (Fiske and Hauser, 2014; Puschmann and Bozdag, 2014; Ross, 2014) examined the implications for evolving research ethics in the age of social media.

Many issues raised by the Facebook study are thorny questions that are not easily resolved. Should Facebook have informed users and obtained consent? Would that have biased results, as users might have been more sensitized to positive or negative content in posts? Would a design involving consent with some amount of deception (see 15.3.6.1) have been more appropriate?

The participation of Facebook as a corporate sponsor of research complicates matters further. As acknowledged in the paper, and in the Editor's expression of concern, Facebook is a private company and therefore not subject to the requirements of the Common Rule (Verma, 2014). This raises the interesting question of corporate ethics and conflicts of interest—specifically, what are the obligations of corporations that conduct human subjects research that are not subject to external regulation? How, if at all, did the Facebook study differ from the widely used techniques of showing different web site designs to different sets of users to determine which is preferred (so-called "A/B testing"? (Merritt et al., 2010)).

Discussions of corporate research behavior are likely to continue and evolve for the foreseeable future. Not long after the publication of the Facebook paper, online dating site OkCupid published a blog post describing the many ways that they have experimented with manipulations of content, in the hopes of understanding how participants respond to postings describing potential dates (Rudder, 2014). In contrast with Facebook's effort, these experiments did not involve academic researchers as partners, and were not published in a scientific journal. Do these differences change our perceptions of the ethical implications of the work, or our interpretations of the results?

These questions do not have simple answers, but they do illustrate concerns that most researchers would be well advised to consider carefully. Corporations such as Facebook and OkCupid may be able to weather the publicity associated with these potentially controversial research studies, but many academic researchers—specifically, those working with public funds in public in university settings—might want to think twice before conducting studies that might lead participants to feel as if they had not been treated appropriately.

Perhaps an application of the golden rule to study design might be appropriate. Before conducting a study, you might ask how you might feel if asked to participate, or, as in the Facebook study, you later found that your actions might have been part of the study without your knowledge. If you decide that you might not be comfortable, others might have the same reaction, and you might consider revising your study design.

15.3.5.2 Practical issues

Participants are crucial to our studies—without them, HCI research would be all but impossible. We should make every effort to treat participants in a manner that reflects this importance. Compensation for time and effort is certainly helpful, but researchers should also take concrete steps to make participation convenient and enjoyable. Comfortable surroundings may put participants at ease. Ample opportunities for rest or bathroom breaks should be provided, particularly for studies that involve longer research sessions. Flexibility in scheduling and location can be particularly important for some users: enrolling professionals in your study may require that you travel to their workplace or allow for sessions outside of traditional working hours. If your study is fun and convenient, participants may be more likely to help your recruiting efforts by urging friends and colleagues to join in.

When working with human participants in any form of HCI research, you must pay careful attention to your role as a researcher. Participants may be impressed or intimidated by your presence, your use of language, your technical skills, the context of the experiment, or any of a variety of related factors. This is particularly true for observations and contextual inquiry, where you will spend a great deal of time in close contact with one or more participants. Although you should make every reasonable effort to help participants feel as at ease as possible, you should also be aware that your presence may have an impact on observed performance. In some cases, participants may exhibit the "demand characteristics" described above, trying to behave in the manner that they think you are looking for.

Others have claimed that the mere act of participating in an experiment will influence user behavior, in the so-called "Hawthorne effect" (Macefield, 2007). Although this effect has been the subject of significant debate among scientists (McCambridge et al., 2014; Levitt and List, 2011), some suggested responses are appropriate. Researchers should never give feedback regarding user performance during the course of a study and experiments involving the comparison of multiple interfaces should be controlled and "blind"—participants should not know if one of the alternatives is favored by the researchers (Macefield, 2007).

More generally, these concerns about the influence of researchers on experimental results point towards a need to be modest about the results of our research. All experiments have flaws and no single study establishes incontrovertible facts on its own. When reporting results and drawing conclusions, we should avoid overstatement, admit the flaws in our research, and point the way for future work that will bring greater understanding.

15.3.6 ADDITIONAL CONCERNS

15.3.6.1 Potentially deceptive research

Does respect for persons always require complete disclosure regarding research goals and design? Although it might seem as if withholding key details from research participants might be somewhat less than fully honest, complete transparency might not be appropriate in some cases, particularly if knowledge of the goals of the study

might influence participant behavior. When this happens, researchers might resort to a bit of misdirection. Deceptive studies ask participants to perform tasks that are described as relating to a particular goal, when the researcher is actually interested in addressing a different question unrelated to the goal presented to the user. Although concealing the true nature of the study does present some concerns regarding the validity of informed consent, this practice is often necessary, particularly in situations where full disclosure might compromise the realism of the study.

A study involving security and usability provides an example of the use of deception in HCI research (Schechter et al., 2007). This study had two goals: to determine the influence of security feedback and to see if participants using their own data would behave more or less securely than those who were role-playing using someone else's data. As the researchers were concerned that study participants would not behave naturally if they were told that usability was being studied, they were told that the purpose of the study was to "help make online banking better" (Schechter et al., 2007). Participants were asked to perform online banking tasks. Some participants were "role-playing"—they were asked to pretend that they were a specific individual with specific goals in mind; others used their own bank accounts. In addition to finding that security indicators were not particularly helpful, this study found that people using their own bank accounts behaved more securely than those who were role-playing (Schechter et al., 2007).

Schechter et al. (2007) used deceit as a means of setting up conditions that maximized the realism of the experiment. By presenting users with real online banking tasks, they focused the experiment on how actual users might behave when using online banking on their own. If participants had been told that the experiment was examining their behavior regarding security and privacy, they might have paid extra attention to their behavior in these areas. This use of deception may be useful and valid, but it does have its limits. These limits arise from the established psychological concept of demand characteristics (Orne, 1962), which states that participants in a research study may act in a manner that attempts to validate the hypotheses being tested. In this study, participants may have taken the goal of improving online banking to heart, perhaps acting more insecurely than they otherwise might have (Patrick, 2007a).

A notable phishing study provides another example of the complexities of conducting research without full prior disclosure of goals and participant consent. Researchers at Indiana University harvested email addresses from publicly available sources and then conducted a phishing attack that encouraged students to log in to a university server that would verify (but not store) their authentication credentials. Arguing that no real harm would come to participants, and that disclosure and consent would sensitize participants to the goals of the project, and therefore invalidate results, the developers of this study worked closely with the appropriate IRB to carefully design a study protocol that would not require explicit consent. This process required extensive review of relevant regulations and legislation, leading to a novel study design that allowed the research to proceed without compromising on ethical concerns or participant privacy or security (Finn and Jakobsson, 2007a,b; Jagatic et al., 2007; Jakobsson et al., 2008).

Deception in HCI research should be used carefully and sparingly. As deception pushes the limits of the concept of informed consent, researchers should be careful to frame deceptions clearly, justify their use, and minimize any risks—particularly regarding discomfort and distress—that may be involved (see the Milgram's Experiment sidebar for a famous example of deceptive research). Participants in studies involving deception are usually thoroughly debriefed at the end of their participation. Debriefing has been shown to help deceived participants eliminate negative effects and even to have experiences that were more positive than those of participants who have not been deceived (Smith and Richardson, 1983).

15.3.6.2 Longitudinal studies

Many HCI studies necessarily involve designs that ask more of participants than a single visit to a usability lab for a relatively short (generally less than 2 hours) session. Ethnography, observations, case studies, and other in-depth qualitative studies often require repeated interactions with individual participants over weeks or months. Learnability studies might require multiple lab sessions in order to measure retention, while studies of technology in use might involve data collection over an extended time period, potentially including regular interactions with researchers (Azar, 2000; Harrison et al., 2014; Srinivasan et al., 2014). Long-term "field studies" of technologies such as mobile devices are particularly useful for developing understanding of emerging usage patterns that might be overlooked in a brief lab session (Kjeldskov and Skov, 2014).

As important as these studies may be, these longitudinal or "multiwave" studies are also challenging and time consuming. Recruiting, scheduling, and enrolling participants is often hard enough for relatively simple usability studies. When this challenge is extended to include the need for multiple visits or reports and potentially consistent use of a device, over a long period of time, the challenge is even more difficult. Researchers undertaking this challenge should expect a range of difficulties not seen in simpler studies: participants will either actively drop out or passively decline to respond to contacts, mobile devices will break or be lost (Harrison et al., 2014), research team members (particularly students) will move on, etc.

Longitudinal studies should be designed to account for the likelihood of these and other complications. Protocols and requirements for scheduled interactions with participants should be structured to decrease dropouts and nonresponses. Automatic or low-effort data collection through instrumented software (see Chapter 12) or online surveys can take the place of in-person or real-time telephone conversations whenever possible. Any measures that reduce demands placed on participants have the potential to increase your retention rates.

Appropriate incentives might encourage users to stick with a study—for example, users might be allowed to keep the mobile device under study if they complete all phases, or compensation might be "back-loaded", providing the bulk of the financial benefit at the end of the study. Conservative designs will also plan for attrition in both participants and equipment. Enrolling extra participants and purchasing spare devices will increase the likelihood of successfully completing the study, even in the face of attrition and device failure.

Although longitudinal studies may provide insight that is otherwise unavailable (Kjeldskov and Skov, 2014), the costs can be significant. Researchers considering such complex endeavors might ask themselves whether a subset of their goals might be met by simpler lab studies (Kjeldskov et al., 2004), leaving more complex designs for those challenges that simply cannot be addressed in any other way.

15.3.6.3 Working with children

Over the past 20 years, HCI research focusing on children has grown substantially, with regular conferences such as Interaction Design and Children showcasing the work of many research groups, and an exploding number of apps and products engaging children in both recreational and educational activities. As all parents know, working with children can be deeply rewarding and intensely frustrating (often at the same time!). Careful consideration of the differences between child and adult participants will help increase the reward while minimizing the difficulties.

Perhaps most important is the need to ensure that children understand what it means to participate in a research study in general, and in your studies in particular. Although this understanding is, of course, vital for all research participants, it is of particular concern for children who may not be able to understand many of the abstract concepts surrounding research, risk, and consent. When children participate in research studies, parents or legal guardians are generally asked to consent to the participation. When possible, children may also be asked to "assent"—to agree to participate—even if they are not capable of giving informed consent (Society for Research in Child Development, 1991; United States Department of Health and Human Services, 1993). This assent is generally in addition to—not instead of—parental consent.

Similar questions relate to the conduct of the study itself. For studies involving younger children, inviting parents to be in the room while the study is ongoing might be the best way to inspire confidence in the safety and security of the proceedings. When extending such an invitation, you might find that appropriate instruction to the parents is necessary to avoid interference that might bias results. For example, you might suggest that parents not interfere unless the child is clearly upset or distressed. The tables may turn completely in studies involving older children, who might prefer that parents not be present. In these cases, you should be prepared to respectfully discuss potential biases that might be associated with parental observations.

Other seemingly small details might become important when dealing with children in research studies. As children are often very (some parents might say excessively) concerned that their viewpoints are considered and their participation is valued, you might take extra effort to let them know that they are a vital part of a complex process, and that their participation is needed. These concerns are particularly challenging for studies involving children in design activities (Read et al., 2014). As many children are reward-driven, age-appropriate incentives (books are great) can be a great motivator for sitting still and completing tasks. Of course, given limited attention spans, experimental sessions should be designed with the age of the child in mind: 3-hour sessions for 4-year-olds might be doomed from the start.

Given the concern over the safety and security of children, these differences in consent and study design may lead to greater scrutiny of proposed studies by IRBs. Discussing your projects with your local IRB and modeling your materials on previous studies where possible might help ensure appropriate safeguards for underage participants and facilitate the approval process.

An understanding of childhood cognitive development can be invaluable for building a early understanding of what children of various ages can be expected to know and how they might view the world. Juan-Pablo Hourcade's book on Child-Computer Interaction (Hourcade, 2015) provides an invaluable overview. When possible, you might also consider including a child development expert on your research team.

15.3.6.4 Populations with specific concerns

Children are not the only research participants who may need special care and handling. Many HCI research projects involve needs assessments, tool development, and evaluation for projects that either specifically focus on certain groups of users, or aim for universal usability across broad ranges of ability, expertise, and technologies (Lazar, 2007). These populations might include older individuals, members of specific ethnic groups, patients with specific health concerns, families, and many other groups. Working with these "nontraditional" research participants can be both rewarding and challenging, requiring both flexibility and creativity.

Recruitment is often the first struggle, particularly for academic researchers who have traditionally relied on readily available pools of undergraduates as study participants. Finding participants for these more specific studies might require interacting with community organizations (as suggested for working with participants with disabilities, Chapter 16), reaching out via online and print resources, encouraging word of mouth and referrals from friends, and numerous other creative approaches. Experience indicates that there is no "silver bullet": many studies rely on a combination of approaches to meet enrollment goals. Perhaps the only consistent observation is that recruitment will often take longer and cost more than expected, making realistic plans necessary for success.

Finding appropriate subjects from these groups is only a part of the challenge. Scheduling and transportation can be difficult for families, elders, and others, just as they are for individuals with disabilities (Chapter 16). Flexibility is key—anything that you might be able to do to accommodate diverse schedules and living arrangements will be helpful. Consider your experimental design and data collection requirements—lab-based studies that work for some subjects might be inappropriate for studies with diverse user populations. Designs that simplify data collection—preferably simple enough to run on a laptop—will enable the enrollment of participants in participant homes, community centers, and libraries, potentially removing or reducing logistical barriers that might discourage some participants.

Studies involving these user populations should also give careful consideration to specific needs and limitations that might impact participants' ability to meaningfully

participate in research studies. Some participants might struggle with informed consent forms and research project descriptions that might be straightforward for undergraduates. Patients considering participation of studies of medically-related systems might have privacy concerns. Families might be limited in their ability to participate in long sessions. Careful planning and review of proposed protocols, pilot participants, and planning of supplies and materials are particularly vital for studies involving these participants.

15.3.7 INTERNATIONAL CONCERNS

We apologize to readers from outside the United States who might have found themselves frustrated that the discussions of informed consent and IRBs earlier in this chapter are overly focused on the conduct of research in the USA, with little attention paid to matters in other countries. We plead guilty—this chapter is indeed heavily influenced by our experiences in conducting research in the US.

Although the above discussion may seem somewhat parochial, protecting participants in research studies is a global concern. A 2015 listing of human research standards lists dozens of countries with relevant standards (Office for Human Research Protection, 2017). Many international organizations, such as the World Health Organization (World Health Organization, 2015), have their own policies, protocols, and terminologies. For example, outside of the US researchers generally work with "Research Ethics Committees" instead of "Institutional Review Boards."

Despite these differences, overall perspectives are generally fairly well aligned. The European Commission's policies on protecting human participants cites precedents from the European Charter of Fundamental rights, including the "right to the integrity of the person," "respect for private and family life," "protection of personal data," and "Freedom of the Arts and Sciences" (European Commission, 2013). The differences between these rights and the three pillars of US policy as outlined in the Belmont Report—beneficence, respect for persons, and justice—lead to subtle differences in emphasis, but little that would be in substantive disagreement with American policy. For example, the European Commission's policy discussion explicitly covers data protection measures (European Commission, 2013) that would be familiar to many researchers in the US, even though those matters are not discussed in the Common Code.

Given these differences, researchers conducting human studies experiments should always be careful to ensure that they are appropriately versed in the local understanding of human subjects' protections and the related regulatory requirements. This preparation may be particularly important for those conducting research in a culture with which they are unfamiliar, as misunderstandings may lead to difficulties.

Cultural sensitivity is a particular concern for research projects conducted by foreigners working in developing countries. Work in the area known as Information and Communication Technology for Development (ICT4D) often involves the participation of researchers from relatively affluent locales in projects in developing countries, often with participants who do not share their levels of

affluence or education. Some commentators have raised the question of whether or not informed consent is possible in the face of these disparities (Sharma, 2005), while others have developed materials that might use field training of local field workers to help promote research ethics in challenging situations (Merritt et al., 2010). HCI efforts in such circumstances should carefully consider how questions of imbalances in education and financial resources might bias research. Working with local partners is generally a necessity, as is compliance with local regulations. Although requirements will vary across contexts, project review by both the researchers' "home" institution and an appropriate board at the site of the study might be necessary.

15.4 HUMAN SUBJECTS RESEARCH AND THE PUBLIC TRUST

Human Subjects research can be alternatively rewarding and infuriating. The distances between the excitement of a novel insight or a statistically significant result and the frustration of dealing with participants who miss appointments or IRBs who misinterpret studies are all often very short indeed. Although all research endeavors face their share of difficulties, the bureaucratic issues in dealing with IRBs and related paperwork are particularly problematic, as the ever-present temptation to cut corners presents a tantalizing way out. We have all heard excuses like "I don't really need an IRB review for this study" or "we can reuse an existing IRB approval." We've also experienced the difficulties of securing IRB approval, with projects delayed weeks or even months in a seemingly inscrutable bureaucratic process. The temptation to end run these processes may be strong, but it should be avoided.

Although the practical costs of such approaches—including potential difficulties in publication and risk of losing grant funding—are significant, the real problem in short-cutting human-subject protection lies in the abuse of the public trust. Certainly, many years have passed since the Tuskegee experiments, Milgram's experiments, and Zimbardo's prison in a Stanford psychology building, and the overwhelming majority of scientific studies are conducted carefully, ethically, and appropriately. However, ethical questions in research conducted are far from fully settled. The Facebook (Kramer et al., 2014) and OKCupid controversies (Rudder, 2014) illustrate the difficulties that researchers might face if they fail to consider ethical questions before they tackle novel problems. Moving beyond individual researchers, professional norms can also be the source of great controversy. A 2015 study commissioned by the American Psychological Association found that changes to the APA's ethics policies enacted in the early 2000s may have been conducted with undue deference to the goals of the US Central Intelligence Agency and Department of Defense, and may have provided a veneer of approval to torture practices (Hoffman et al., 2015). The ensuing controversy led to substantial upheaval at the APA (Bohannon, 2015).

Researchers involved in these studies—or, for that matter, in prior studies we now think of as abusive or inappropriate—were not necessarily acting in bad faith or out of lack of concern. This is exactly why we need careful, independent review to ensure respect for persons, beneficence, and justice (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979), and to justify the trust that society requires when we ask individuals to participate in our studies and governments and foundations to financially support our work.

15.5 SUMMARY

Working with human subjects is one of the most challenging and informative aspects of HCI research. Finding appropriate participants; informing them of their rights; protecting their privacy; and answering their questions can be time consuming and often tedious, but the results are more than worth the effort. Even when study participants criticize our designs or fail to confirm our cherished experimental hypotheses, they provide invaluable insight that provides a rigorous foundation for our work.

Whatever type of study you are running, it is never too early to plan for recruiting, informed consent documentation, and other aspects of human participation. Proper planning will keep your study from becoming one of the many that have been delayed by unforeseen circumstances including difficulty in finding participants, or delays in IRB approval.

Recruiting entails finding the right number of the right kinds of participants. For usability studies, ethnographic observations of users, interviews, focus groups, and other approaches aimed at gathering requirements or evaluating design proposals, this may mean understanding the audience of users and identifying a sample of participants that is broad enough to reflect the needs and behavior of potential users. Designers and professional developers conducting research of this sort might work with collaborators, marketing teams, professional organizations, or others with appropriate understanding and context to identify both the range of viewpoints that would be needed and possible sources of the appropriate individuals.

Empirical studies require consideration of both the diversity of potential participants and any confounding factors that might contribute to performance differences. Characteristics of desirable participants might both be informed by and influence experimental hypotheses. Students and researchers conducting these studies should be careful to plan their data analysis and recruiting together, to ensure that the participants will be selected to increase the power of the statistical analysis.

Appropriate respect for participants is a cornerstone of all research involving human subjects. These issues are particularly relevant for studies that involve deception. Even when not required by institutional policy to do so, designers and developers would be advised to use formal informed consent forms to help participants make informed decisions. Students and researchers should take the time—again, as early in the process as possible—to understand the regulations in force in their institution, and to make sure that their approvals are in order before starting any project.

Human subjects research in HCI can be an unpredictable and often unsettling process. Unforeseen problems, including misinterpreted tasks and goals, systems failures, and missed appointments, are routine: it's rare that a study (of any sort) goes off completely without a hitch. These matters can complicate data collection and interpretation: if a user chooses an interpretation of a written task that differs from your intent and then completes the task correctly, how do you interpret the result—is it correct or not? What should be done with results from a user who decides to withdraw from a study after completing only a portion of the tasks? As hard and fast rules for handling situations like these are few and far between, you may have to handle each issue on a case-by-case basis. The specific decisions that you make may be less important than how they are enforced: consistent application of policies and procedures will ensure your ability to make meaningful comparisons.

All participants in HCI research studies should be well treated and approached with an open mind. Participating in HCI studies should be fun and engaging whenever possible: by making our studies positive experiences, we encourage people both to participate and to provide useful feedback. As researchers, we should "expect the unexpected": software will crash, devices won't work, and (perhaps most distressingly) users will hate our beloved inventions. High-quality HCI research takes these setbacks in its stride, all the while striving to observe carefully while maintaining respect for the people who give a bit of their time to help our studies along. By watching and listening carefully, we can learn from what users do and how they do it. That, after all, is the point of conducting user studies.

DISCUSSION QUESTIONS

1. University researchers occasionally ask students in a class to participate in research studies. However, this practice may involve elements of coercion, as students may be concerned that refusal to participate may negatively impact their grade. Is voluntary informed consent possible in such a situation? What steps might be taken to reconcile the researcher's need for subjects with the students' right to decline to participate?

2. The virtual reprise of Milgram's experiment (see Section 15.2.1) asked participants to inflict harm upon a computer-generated avatar. This approach eliminates some of the potential ethical concerns associated with the original experiment, but may raise additional questions. As user behavior was similar to what was observed in the original experiments, it is possible that participants in the "virtual" versions would experience similar patterns of nervousness and distress. Do you consider this sort of research to be appropriate? What might be done to protect participants in this sort of experiment?

3. As part of a larger study of how various aspects of interaction in online worlds impact the offline lives of participants, you are interested in observing participants both online and offline. As you know, participants in online games

such as these may not represent a broad cross-section of society. The race and gender of online characters may not reflect those of the real individuals involved and some may choose to hide their "real" identity. Given these challenges, how might you go about finding a group of participants that would be interesting to work with? How might these challenges affect the conclusions that you might be able to draw from your observations and your ability to generalize from those conclusions?

RESEARCH DESIGN EXERCISES

1. You are designing a study to evaluate the effectiveness of a new text entry method for messaging on cell phones. Due to the popularity of messaging among college students, you decide that the undergraduate student body at your school would be an appropriate pool of potential participants. What would you want to know about the habits of these students regarding text messaging? You might be interested in comparing the performance of computer science students against students from other fields. Are there any other attributes of the students that might make for interesting comparisons? Given the male-female imbalance in computer science, what problems might this comparison involve?
2. Your research design for the study of text entry on cell phones involves asking users to perform a set of tasks in a laboratory. As they will not be using their own phones, there is little, if any, privacy risk. What other risks might this study pose, and how would you inform users about them?
3. Find the website or other information about your IRB. Examine the policies and procedures specific to your institution, and write a draft informed consent form for the study described in Exercise 1.
4. Studies of how users respond to events that interrupt their work (Gluck et al., 2007) present a challenge in design. If participants are told that the study is investigating reactions to interruptions, they may be more sensitive to those events than they would otherwise be. A deceptive study, in which the subjects were provided with an alternative description of the goals of the study, might be one way to get around this problem. How might you describe a deceptive study for examining reactions to interruptions? How would you describe this study in an informed consent form? What would you discuss in the debriefing sessions?

REFERENCES

Asan, O., Montague, E., 2013. Technology-mediated information sharing between patients and clinicians in primary care encounters. Behaviour & Information Technology 33 (3), 259–270.

Azar, B., 2000. Online experiments: ethically fair or foul? Monitor on Psychology 31 (4).
Beskow, L.M., Dombeck, C.B., Thompson, C.P., Watson-Ormond, J.K., Weinfurt, K.P., 2015. Informed consent for biobanking: consensus-based guidelines for adequate comprehension. Genetics in Medicine 17 (3), 226–233.
Beskow, L.M., Friedman, J.Y., Hardy, N.C., Lin, L., Weinfurt, K.P., 2010. Developing a simplified consent form for biobanking. PLoS One 5 (10), e13302.
Bohannon, J., 2015. APA Overhauling Policies and Leadership After Torture Report. Retrieved from http://news.sciencemag.org/brain-behavior/2015/07/apa-overhauling-policies-and-leadership-after-torture-report (30.08.15).
Centers for Disease Control and Prevention, 2007. U.S. Public Health Service Syphilis Study at Tuskegee. Retrieved from http://www.cdc.gov/nchstp/od/tuskegee/ (31.05.07).
Cohen, I.G., Lynch, H.F. (Eds.), 2014. Human Subjects Research Regulation: Perspectives on the Future. MIT Press, Cambridge, MA.
Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin Company, Boston.
Davis, T.C., Holcombe, R.F., Berkel, H.J., Pramanik, S., Divers, S.G., 1998. Informed consent for clinical trials: a comparative study of standard versus simplified forms. Journal of the National Cancer Institute 90 (9), 668–674.
Department of Health and Services, 2009. Federal Policy for Protection of Human Research Subjects. Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.
European Commission, 2013. Ethics for Researchers: Facilitating Research Excellence in FP7. Retrieved March 20, 2017 from https://ec.europa.eu/research/participants/data/ref/fp7/89888/ethics-for-researchers_en.pdf.
Federal Register, 2015. Federal Policy for the Protection of Human Subjects, United States Government. Retrieved March 20, 2017 from https://www.federalregister.gov/documents/2015/09/08/2015-21756/federal-policy-for-the-protection-of-human-subjects.
Finn, P., Jakobsson, M., 2007a. Designing and Conducting Phishing Experiments. IEEE Technology and Society Magazine (Special Issue on Usability and Security). pp. 1–21.
Finn, P., Jakobsson, M., 2007b. Designing ethical phishing experiments. IEEE Technology and Society Magazine 26 (1), 46–58.
Fiske, S.T., Hauser, R.M., 2014. Protecting human research participants in the age of big data. Proceedings of the National Academy of Sciences 111 (38), 13675–13676.
Gluck, J., Bunt, A., McGrenere, J., 2007. Matching attentional draw with utility in interruption. In: Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems, ACM, San Jose, CA, pp. 41–50.
Goel, V., 2014. Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry. The New York Times. Retrieved March 20, 2017 from https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0.
Haney, C., Banks, C., Zimbardo, P., 1973. Interpersonal dynamics in a simulated prison. International Journal of Criminology and Penology 1, 69–97.
Harrison, D., Marshall, P., Berthouze, N., Bird, J., 2014. Tracking physical activity: problems related to running longitudinal studies with commercial devices. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication. ACM, Seattle, Washington, pp. 699–702.
Hochheiser, H., Lazar, J., 2010. Revisiting breadth vs. depth in menu structures for blind users of screen readers. Interacting with Computers 22 (5), 389–398.

Hoffman, D.H., Carter, D., Viglucci Lopez, C.R., Guo, A.X., Latfif, S.Y., Craig, D., 2015. Report to the Special Committee of the Board of the American Psychological Association: Independent Review Relating to APA Ethics Guidelines, National Security Investigations, and Torture. Retrieved March 20, 2017 from http://www.apa.org/independent-review/revised-report.pdf.

Hourcade, J.P., 2015. Child-Computer Interaction. CreateSpace Independent Publishing Platform ISBN 978-1514397251. Available from http://homepage.divms.uiowa.edu/~hourcade/book/child-computer-interaction-first-edition.pdf (accessed March 19, 2017).

Hwang, W., Salvendy, G., 2010. Number of people required for usability evaluation: the 10+/−2 rule. Communications of the ACM 53 (5), 130–133.

Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menczer, F., 2007. Social phishing. Communications of the ACM 50 (10), 94–100.

Jakobsson, M., Finn, P., Johnson, N., 2008. Why and how to perform fraud experiments. IEEE Security and Privacy 6 (2), 66–68.

Jones, J.H., 1993. Bad Blood: The Tuskegee Syphilis Experiment. Free Press, New York.

Kjeldskov, J., Skov, M., Als, B., Høegh, R., 2004. Is it worth the hassle? Exploring the added value of evaluating the usability of context-aware mobile systems in the field. In: Brewster, S., Dunlop, M. (Eds.), Mobile Human-Computer Interaction—MobileHCI 2004, vol. 3160. Springer, Berlin Heidelberg, pp. 61–73.

Kjeldskov, J., Skov, M.B., 2014. Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations. In: Proceedings of the 16th International Conference on Human-Computer Interaction With Mobile Devices' Services. ACM, Toronto, ON, Canada, pp. 43–52.

Kramer, A.D.I., Guillory, J.E., Hancock, J.T., 2014. Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences 111 (24), 8788–8790.

Larson, K., Czerwinski, M., 1998. Web page design: implications of memory, structure and scent for information retrieval. In: SIGCHI Conference on Human Factors in Computing Systems Los Angeles. ACM Press, CA, United States.

Lazar, J. (Ed.), 2007. Universal Usability: Designing Computer Interfaces for Diverse User Populations. John Wiley and Sons, Chichester, UK.

Levitt, S.D., List, J.A., 2011. Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. American Economic Journal: Applied Economics 3 (1), 224–238.

Luger, E., Rodden, T., 2013a. An informed view on consent for UbiComp. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, Zurich, Switzerland, pp. 529–538.

Luger, E., Rodden, T., 2013b. Terms of agreement: rethinking consent for pervasive computing. Interacting with Computers 25 (3), 229–241.

Macefield, R., 2007. Usability studies and the Hawthorne effect. Journal of Usability Studies 2 (3).

McCambridge, J., Witton, J., Elbourne, D.R., 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. Journal of Clinical Epidemiology 67 (3), 267–277.

Merritt, M.W., Labrique, A.B., Katz, J., Rashid, M., West Jr., K.P., Pettit, J., 2010. A field training guide for human subjects research ethics. PLoS Medicine 7 (10), e1000349.

Milgram, S., 1963. Behavioral study of obedience. Journal of Abnormal and Social Psychology 67, 371–378.

National Academies of Sciences and Medicine, 2016. Optimizing the Nation's Investment in Academic Research: A New Regulatory Framework for the 21st Century. Retrieved March 20, 2017 from https://www.nap.edu/catalog/21824/optimizing-the-nations-investment-in-academic-research-a-new-regulatory.

National Institutes of Health, 2014. Request for Comments on the Draft NIH Policy on the Use of a Single Institutional Review Board for Multi-Site Research. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-026.html (29.08.15).

Nielsen, J., 1992. Finding usability problems through heuristic evaluation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Monterey, CA, USA, pp. 373–380.

Nielsen, J., Molich, R., 1990. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Seattle, Washington, USA, pp. 249–256.

Office for Human Research Protection, 2017. International Compilation of Human Research Standards, 2017 Edition. Retrieved March 20, 2017 from https://www.hhs.gov/ohrp/sites/default/files/international-compilation-of-human-research-standards-2017.pdf.

Office for Human Research Protections, 1998. Informed Consent Checklist. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/guidance/checklists/index.html (01.01.17).

Orne, M.T., 1962. On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. The American Psychologist 17 (11), 776–783.

Patrick, A., 2007a. Commentary on Research on New Security Indicators. Retrieved from http://www.andrewpatrick.ca/essays/commentary-on-research-on-new-security-indicators/ (22.06.07).

Patrick, A., 2007b. Privacy Practices for HCI Research. Retrieved from http://www.andrewpatrick.ca/essays/privacy-practices-for-hci-research/ (04.06.07).

Pellegrino, E.D., 1997. The Nazi doctors and Nuremberg: some moral lessons revisited. Annals of Internal Medicine 127 (4), 307–308.

Puschmann, C., Bozdag, E., 2014. Staking out the unclear ethical terrain of online social experiments. Internet Policy Review 3 (4).

Read, J.C., Fitton, D., Horton, M., 2014. Giving ideas an equal chance: inclusion and representation in participatory design with children. In: Proceedings of the 2014 Conference on Interaction Design and Children. ACM, Aarhus, Denmark, pp. 105–114.

Ross, M., 2014. Do research ethics need updating for the digital age? Monitor on Psychology 45.

Rudder, C., 2014. We Experiment On Human Beings!. Retrieved from http://blog.okcupid.com/index.php/we-experiment-on-human-beings/ (27.09.15).

Schechter, S., Dhamija, R., Ozment, A., Fischer, A., 2007. The Emperor's New Security Indicators: an evaluation of website authentication and the effect of role playing on usability studies. In: IEEE Symposium on Security and Privacy. IEEE, Oakland, CA.

Schmettow, M., 2012. Sample size in usability studies. Communications of the ACM 55 (4), 64–70.

Schrag, Z.M., 2015. Institutional Review Blog. Retrieved from http://www.institutionalreviewblog.com/ (20.09.15).

Sharma, K., 2005. Can clinical trials ever be truly ethical? The Hindu. Retrieved April 10, 2017 from http://www.thehindu.com/2005/12/06/stories/2005120603081000.htm.

Shavers, V.L., Lynch, C.F.L., Burmeister, L.F., 2000. Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. Journal of the National Medical Association 92 (12), 563–572.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., Diakopoulos, N., 2016. Designing the user interface, sixth ed. Pearson Addison Wesley, Boston.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N., Sanchez-Vives, M.V., 2006. A virtual reprise of the Stanley Milgram obedience experiments. PLoS One 1 (1).

Smith, S.S., Richardson, D., 1983. Amelioration of deception and harm in psychological research: the important role of debriefing. Journal of Personality and Social Psychology 44 (5), 1075–1082.

Society for Research in Child Development, 1991. Ethical Standards for Research with Children. Retrieved from http://www.srcd.org/about-us/ethical-standards-research (08.04.17).

Srinivasan, V., Moghaddam, S., Mukherji, A., Rachuri, K.K., Xu, C., Tapia, E.M., 2014. MobileMiner: mining your frequent patterns on your phone. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. ACM, Seattle, Washington, pp. 389–400.

Steriadis, C.E., Constantinou, P., 2003. Designing human-computer interfaces for quadriplegic people. ACM Transactions on Computer-Human Interaction 10 (2), 87–118.

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html (08.04.17).

United States Department of Health and Human Services, 1993. Institutional Review Board Guidebook: Chapter VI: Special Classes of Subjects. Retrieved from https://archive.hhs.gov/ohrp/irb/irb_chapter6.htm (20.03.17).

Verma, I.M., 2014. Editorial expression of concern: psychological and cognitive sciences. Proceedings of the National Academy of Sciences 111 (24), 10779.

Weibel, N., Rick, S., Emmenegger, C., Ashfaq, S., Calvitti, A., Agha, Z., 2015. LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office. Personal and Ubiquitous Computing 19 (2), 317–334.

World Health Organization, 2015. Ethical standards and procedures for research with human beings. Retrieved from http://www.who.int/ethics/research/en/ (27.09.15).

Zimbardo, P., 2008a. The Lucifer Effect: Understanding How Good People Turn Evil. Random House.

Zimbardo, P., 2008b. Stanford Prison Experiment. Retrieved from http://www.prisonexp.org/ (08.04.17).