Methodological and Ethical Issues in Conducting Social Psychology Research via the Internet

MICHAEL H. BIRNBAUM

California State University, Fullerton

When I set out to recruit highly educated people with specialized training in decision making, I anticipated that it would be a difficult project. The reason I wanted to study this group was that I had been obtaining some very startling results in decision-making experiments with undergraduates (Birnbaum & Navarrete, 1998; Birnbaum, Patton, & Lott, 1999). Undergraduates were systematically violating stochastic dominance, a principle that was considered both rational and descriptive, according to cumulative prospect theory and rank dependent expected utility theory, the most widely accepted theories of decision making at the time (Luce & Fishburn, 1991; Quiggin, 1993; Tversky & Kahneman, 1992). These theories were recognized in the 2002 Nobel Prize in Economics shared by

Daniel Kahneman, so these systematic violations required substantial changes in thinking about decision making.

The results were not totally unexpected, for they had been predicted by my configural weight models of decision making (Birnbaum, 1997). Nevertheless, I anticipated the challenge that my results might apply only to people who lack the education required to understand the task. I wanted to see if the results I obtained with undergraduates would hold up with people highly trained in decision making, who do not want to be caught behaving badly with respect to rational principles.

From my previous experiences in research with special populations, I was aware of the difficulties of such research. In previous work, my students and I had printed and mailed materials to targeted participants

construction of data files, it did not allow the possibility of anonymous participation, and it could not work for those without e-mail. It also annoyed people by clogging their e-mail folder with unsolicited (hence suspicious) material.

Since 1995, a superior technique, HyperText Markup Language (HTML) forms, has become available. This method uses the World Wide Web (WWW) rather than the e-mail system. HTML forms allow one to post a Web page of HTML, which automatically codes the data and sends them to a CGI (Common Gateway Interface) script, which saves the data to the server. This method avoids clogging up mailboxes with long surveys, does not require e-mail, and can automatically produce a coded data file ready to analyze.

Anything that can be done with paper and pencil and fixed media (graphics, photographs, sound, or video) can be done in straight HTML. For such research, all one needs are a Web site to host the file, a way to recruit participants to that URL, and a way to save the data.

To get started with the technique, it is possible to use one of my programs, (e.g., SurveyWiz or FactorWiz), which are available from the following URL: http://psych.fullerton. edu/mbirnbaum/programs/

These free programs allow a person to create a Web page that controls a simple survey or factorial study without really knowing HTML (Birnbaum, 2000c). Readers are welcome to use these programs to create practice studies, and even to collect pilot data with nonsensitive content. The default CGI included saves the data to my server, from which you can download your data.

To go beyond the minimal requirements—for example, to learn how to install server software on your lab's computer (Schmidt, Hoffman, & MacDonald, 1997), add dynamic functionality (present content that depends on the participant's behavior), or

measure response times (Birnbaum, 2000b, 2001a; McGraw, Tew, & Williams, 2000b)—see the resources at the following URLs: http:// ati.fullerton.edu and http://psych.fullerton. edu/mbirnbaum/www/links.htm

POTENTIAL ADVANTAGES OF RESEARCH VIA THE WWW

Some of the advantages of the new methods are that one can now recruit participants from the entire world and test them remotely. One can test people without requiring them to reveal their identities or to be tested in the presence of others. Studies can be run without the need for large laboratories, without expensive dedicated equipment, and without limitations of time and place. Online studies run around the clock anywhere that a computer is connected to the WWW. Studies can be run by means of the (now) familiar browser interface, and participants can respond by (now) familiar methods of pointing and clicking or filling in text fields.

Once the programming is perfected, data are automatically coded and saved by a server, sparing the experimenter (and his or her assistants) from much tedious labor. Scientific communication is facilitated by the fact that other scientists can examine and replicate one's experiments precisely. The greatest advantages probably are the convenience and ease with which special samples can be recruited, the low cost and ease of testing large numbers of participants, and the ability to do in weeks what used to take months or years to accomplish.

On the other hand, these new methods have limitations. For example, it is not yet possible to deliver stimuli that can be touched, smelled, or tasted. One cannot deliver drugs or shocks, nor can one yet measure Galvanic Skin Response, PET scans, or heart rates. It is not really possible to control or manipulate the physical presence of other people. One can

(Birnbaum & Hynan, 1986; Birnbaum & Stegner, 1981). Each packet contained a self-addressed and stamped envelope as well as the printed materials and cover letter. We then sent, by mail, reminders with duplicate packets containing additional postage out and back. As packets were returned, we coded and entered data, and then verified the data. All in all, the process was slow, labor-intensive, expensive, and difficult.

I had become aware of the (then) new method for collecting data via the Web (HTML forms), and I decided to try that approach, which I thought might be more efficient than previous methods. I knew that my targeted population (university professors in decision making) could be reached by e-mail and would be interested in the project. I thought that they might be willing to click a link to visit the study and that it would be convenient for them to do online. I was optimistic, but I was unprepared for how successful the method would prove.

Within 2 days of announcing the study, I had more than 150 data records ready to analyze, most from people in my targeted group. A few days later, I was receiving data from graduate students of these professors, then from undergraduates working in the same labs, followed by data from people all over the world at all hours of the day and night. Before the study ended, I had data from more than 1,200 people in 44 nations (Birnbaum, 1999b). Although the results did show that the rate of violation of stochastic dominance varied with gender and education, even the most highly educated participants had substantial rates of violation, and the same conclusions regarding psychological processes were implied by each stratum of the sample. That research led to a series of new studies via the Web, to test new hypotheses and conjectures regarding processes of decision making (Birnbaum, 2000a, 2001b; Birnbaum & Martin, in press).

My success with the method encouraged me to study how others were using the Web in their research, and to explore systematically the applicability of Web research to a variety of research questions (Birnbaum, 1999b, 2000b, 2000c, 2001a, 2002; Birnbaum & Wakcher, 2002). The purpose of this chapter is to review some of the conclusions I have reached regarding methodological and ethical issues in this new approach to psychological research.

The American Psychological Society page of "Psychological Research on the Net," maintained by John Krantz, is a good source for review of Web experiments and surveys (http://psych.hanover.edu/research/exponnet.html). In 1998, this site listed 35 studies; by 1999, the figure had grown to 65; as of December 9, 2002, there were 144 listings, including 43 in social psychology. Although not all Web studies are listed in this site, these figures serve to illustrate that use of the method is still expanding rapidly.

The Internet is a new medium of communication, and as such it may create new types of social relationships, communication styles, and social behaviors. Social psychology may contribute to understanding characteristics and dynamics of Internet use. There are now several reviews of the psychology of the Internet, a topic that will not be treated in this chapter (see Joinson, 2002; McKenna & Bargh, 2000; Wallace, 2001). Instead, this chapter reviews the critical methodological and ethical issues in this new approach to psychological research.

MINIMUM REQUIREMENTS FOR ONLINE EXPERIMENTING

The most elementary type of Internet study was the e-mail survey, in which investigators sent text to a list of recipients and requested that participants fill in their answers and send responses by return email. This type of research saved paper and mailing costs; however, it did not allow easy coding and

offer real or simulated people at the other end of a network, but such manipulations may be quite different from the effects of real personal presence (Wallace, 2001).

When conducting research via the Web, certain issues of sampling, control, and precision must be considered. Are people recruited via the Web more or less "typical" than those recruited by other means? How do we know if their answers are honest and accurate? How do we know what the precise conditions were when the person was being tested? Are there special ethical considerations in testing via the Web?

Some of these issues were addressed in "early" works by Batinic (1997), Birnbaum (1999a, 1999b), Buchanan and Smith (1999), Pettit (1999), Piper (1998), Reips (1997), Schmidt (1997a, 1997b), Schmidt, Hoffman, et al. (1997), Smith and Leigh (1997), Stern and Faber (1997), and Welch and Krantz (1996). Over the last few years, there has been a great expansion in the use of the WWW in data collection, and much has been learned or worked out. Summaries of more recent work are available in a number of recent books and reviews (Birnbaum, 2000b, 2001a; Janetzco, Meyer, & Hildebrand, 2002; Reips, 2001a, 2001b, 2002). This chapter will summarize and contribute to this growing body of work.

EXAMPLES OF RECRUITING AND TESTING PARTICIPANTS VIA THE INTERNET

It is useful to distinguish the use of the WWW to recruit participants from its use to test participants. The cases of four hypothetical researchers will help to illustrate these distinctions. One investigator might recruit participants via the Internet and then send them printed or other test materials by mail, whereas another might recruit participants from the local "subject pool" and test them via the Internet. A third investigator might both recruit and test via the Web, and a fourth might use the WWW to recruit criterion groups in order to investigate individual differences.

In the first example, consider the situation of an experimenter who wants to study the sense of smell among people with a very rare condition. It is not yet possible to deliver odors via the WWW, so this researcher plans to use the WWW strictly for recruitment of people with the rare condition. Organizations of people with rare conditions often maintain Web sites, electronic newsletters, bulletin boards, and other means of communication. Such organizations might be asked to contact people around the world with the targeted, rare condition. If an organization considers research to be relevant and valuable to the concerns of its members, the officers of the organization may offer a great deal of help in recruiting its members. Once recruited, participants might be brought to a lab for testing, visited by teams traveling in the field, or mailed test kits that the participants would administer themselves or take to their local medical labs.

The researcher who plans to recruit people with special rare conditions can be contrasted with the case of a professor who plans to test large numbers of undergraduates by means of questionnaires. Perhaps this experimenter previously collected such data by paper-and-pencil questionnaires administered to participants in the lab. Questionnaires were typed, printed, and presented to participants by paid research assistants, who supervised testing sessions. Data were coded by hand, typed into the computer, and verified.

In this second example, the professor plans to continue testing participants from the university's "subject pool." However, instead of scheduling testing at particular times and places, this researcher would like to allow participation from home, dorm room, or wherever students have Web access. The advantages of the new procedure are largely convenience

to both experimenter and participants, along with cost savings for paper, dedicated space, and assistants for in-lab testing, data coding, and data entry. This hypothetical professor might even conduct an experiment, randomly assigning undergraduates to either lab or Web, to ascertain if these methods make a difference.

In the third example, which describes my initial interest in the method, the experimenter plans to compare data from undergraduates tested in the lab against special populations, who could be recruited and tested via the Web. This experimenter establishes two Web pages, one of which collects data from those tested in the lab. The second Web page contains the same content but receives data from people tested via the WWW. The purpose of this research would be to ascertain whether results observed with college students are also obtained in other groups of people who can be reached via the Web (Birnbaum, 1999b, 2000a, 2001a, 2001b; Krantz & Dalal, 2000).

In the fourth example, a researcher compares criterion groups, in order to calibrate a test of individual differences or to examine if a certain result interacts with individual differences (Buchanan, 2000). Separate Web sites are established that contain the same materials, but people are recruited to these URLs by methods intended to reach members of distinct criterion groups (Schillewaert, Langerak, & Duhamel, 1998). One variation of this type of research is cross-cultural research, in which people who participate from different countries are compared (Pagani & Lombardi, 2000).

Recruitment Method and Sample Characteristics

Although the method of recruitment and the method of testing in either Web or lab are independent issues, most of the early Web studies recruited their participants from the Web and recruited their lab samples from use

the university's "subject pool" of students. Therefore, many of the early studies compared two ways of recruiting and testing participants.

Because the "subject pool" is finite, at some universities there is competition for the limited number of subject-hours available for research. In addition, the lab method usually is more costly and time-consuming *per subject* compared to Web studies, where there is no additional cost once the study is on the WWW. For these reasons, Web studies usually have much larger numbers of participants than do lab studies.

Participation in psychological research is considered an important experience for students of psychology. Web studies allow students to be able to participate in psychological studies even if they attend universities that do little research or if they engage in "distance" learning (taking courses via TV or the Web).

Table 16.1 lists a number of characteristics that have been examined in comparisons of participants recruited from subject pools and via search engines, e-mail announcements, and links on the Web. I will compare "subject pool" against Web recruitment, with arguments why or where one method might have an advantage over the other.

Demographics

College students who elect to take psychology courses are not a random sample of the population. Almost all have graduated high school, and virtually none have college degrees. At many universities, the majority of this pool is female. At my university, more than two thirds are female, most are between 18 and 20 years of age, and none of them has graduated college. Within a given university, despite efforts to encourage diversity, the student body is often homogeneous in age, race, religion, and social class. Although many students work, most lower-division students are supported by their parents, and the

Table 16.1 Characteristics of Web-Recruited Samples and College Student "Subject Pool" Participants

Characteristic	Recruitment Method	
	Subject Pool	Web
Sample sizes	Small	Larger
Cross-cultural studies	Difficult	Easier
Recruitment of rare populations	Difficult	Easier
Sample demographics	Homogeneous	Heterogeneous
Age	18-22	18-80
Nationality/culture	Relatively homogeneous	Heterogeneous
Education	12-14 years	8-20 years
Occupation	Students, part-time work	Various, more full-time
Religion	Often homogeneous	More varied
Race	Depends on university	More varied
Self-selection	Psychology students	Unknown factors; depends on recruitment method
Gender	Mostly female	More equal, depending on recruitment method

positions they hold typically are low-paying, part-time jobs.

Those who choose to take psychology are not a random sample of college students; at my university, they are less likely to be majors in engineering, chemistry, mathematics, or other "hard" subjects than to be undeclared or majoring in "soft" subjects. This is a very specialized group of people.

There are three arguments in favor of sticking with the "subject pool." The first is that this pool is familiar to most psychologists, so most reviewers would not be suspicious of a study for that reason. When one uses a sample that is similar to those used by others, one hopes that different investigators at similar universities will find similar results. Second, it is often assumed that processes studied are characteristic of all people, not just students. Third, the university student pool consists of a homogeneous group. Because the student body *lacks* diversity, one expects that the variability of the data resulting from individual differences on these characteristics will be small. This homogeneity should afford greater power compared with studies with equal-sized samples that are more diverse.

Web Participants Do Not Represent a Population

When one recruits from the Web, the method of recruitment will affect the nature of the sample obtained. Even with methods intended to target certain groups, one cannot control completely the nature of the sample obtained via the Web. For example, other Webmasters may place Web links to a study that attract people who are different from the ones the experimenter wanted to recruit. A study designed to assess alcohol use in the general population, for example, might be affected by links to the study posted in self-help groups for alcoholics. Or links might be posted in an upscale wine tasting club, which might recruit a very different sample. There are techniques for checking what link on the Web led each participant to the site, so one

can study how the people got there (Schmidt, 1997a), but at the end of the day one must concede that the people recruited are not a random sample of any particular population.

I think it would be a mistake to treat samples recruited from the Web as if they represented some stable population of "Web users." First, the populations of those who have *access* to the Web and those who *use* the Web are (both) rapidly changing. Those with access could potentially get on the Web; however, the population of people who use the Web at any given time is a nonrandom sample of those with access. Second, those who receive the invitation to participate are not a random sample of all Web users. Third, those who agree to complete a study are not a random sample of those who see the invitation. Fourth, there usually is no real control of who will receive the invitation, once it has been placed in a public file on the Web.

Researchers who have compared data from the Web against data from students have reported that participants recruited via the Web are on average older, better educated, more likely male (though females may still be in the majority), and more likely employed in full-time jobs. Layered over these average differences, most studies also reported that the variance, or diversity, of the samples is much greater on the Web on all characteristics than one usually finds in the subject pool. Thus, although one finds that the average years of education of a Web sample is greater than the average number for college sophomores, one also finds adults via the Web who have no high school diploma, which one does not find in college samples.

Effect of Diversity on Power and Generality

The theoretical problem of diversity is that it may produce greater error variance and therefore reduce power compared to a study in which participants are homogeneous.

However, in practice, the greater diversity in Web samples may actually be a benefit for three reasons.

First, demographic characteristics rarely have shown large correlations with many dependent variables, so diversity of demographics does not necessarily generate much added error variance. Second, sample sizes in Web studies usually outweigh any added variance resulting from heterogeneity, so it is usually the Web studies that have greater power (Musch & Reips, 2000; Krantz & Dalal, 2000). Third, with very large samples, one can partition the data on various characteristics and conduct meaningful analyses within each stratum of the sample (Birnbaum, 1999b, 2001a). When similar results are found with males and females, with young and old, with rich and poor, with experts and novices, and with participants of many nations, the confidence that the findings will generalize to other groups is increased.

In the case that systematically different results are obtained in different strata, these can be documented and an explanation sought. If results do correlate with measures of individual differences, these can be better studied in samples with greater variance. Web samples are likely better for studying such correlations precisely because they do have greater variation. For example, if a person wanted to study the correlates of education, a "subject pool" sample would not have enough variance on education to make the study worthwhile, whereas a Web-recruited sample would have great variance on this variable.

Cases in which demographic characteristics have been shown to correlate with the dependent variable include surveys intended to predict the proportion who will vote Republican or Democrat in the next election. Here the interest is not in examining correlations but in forecasting the election. Neither surveys of college students nor large convenience samples obtained from volunteers via the WWW would likely yield accurate predictions of the

vote (Dillman & Bowker, 2001). It remains to be seen how well one might do by statistically "correcting" Web data based on a theory of the demographic correlates. For example, if results depend on gender and education, one might try to weight cases so that the Web sample had the same gender and education profile as those who vote. It remains to be seen how well one might do with this approach.

The failure of the famous 1936 *Literary Digest* Poll, which incorrectly predicted that Alf Landon would defeat Franklin D. Roosevelt for president, is a classic example of how a self-selected sample (even with very large sample size) can yield erroneous results (Huff, 1954). Bailey, Foote, and Throckmorton (2000) reviewed this topic in the area of sex surveys. In addition to sampling issues, these authors also discussed the conjecture that people might be less biased when answering a questionnaire via the Internet than they would when responding in person.

Some survey researchers use the Web to collect data from what they believe is a *representative* sample, even if it is not random. These researchers establish a sample with proportions of various characteristics (gender, age, education) that match those in the general population, much like the Nielsen Families for TV ratings, and then return to this group again and again for different research questions. See the URL http://www.nielsenmedia.com/whatratingsmean/

Baron and Siepmann (2000) described a variation of this approach, in which a fixed list (of 600 selected volunteers) is used in study after study. A similar approach has been adopted by several commercial polling organizations that have established fixed samples that they consider to be representative. They sell polling services in their fixed sample for a price.

No method yet devised uses true random sampling. Random digit dialing, which is still popular among survey researchers, has two serious drawbacks: First, some people have more phone numbers than others, so the method is biased to obtain relatively more of such people. Second, not everyone telephoned agrees to participate. I am so annoyed by calls at home, day and night, from salespeople who pretend to be conducting surveys that I no longer accept such calls. Even though the dialing may be random, therefore, the sample is not.

Because no survey method in actual use employs true random sampling, most of these arguments about sampling methods are "armchair" disputes that remain unresolved by empirical evidence. Because experiments in psychology do not employ random sampling of either participants or situations, the basis for generalization from experiments must be based on proper substantive theory, rather than on the statistical theory of random sampling.

EXPERIMENTAL CONTROL, MEASUREMENT, AND OBSERVATION

Table 16.2 lists a number of characteristics of experiments that may differ between lab and Web studies. Some of these distinctions are advantages or disadvantages of one method or another. The issues that seem most important are experimental control of conditions, precision of stimulus presentation, observation of behavior, and the accuracy of measurement.

Two Procedures for Holding an Exam

To understand the issue of control, consider the case of a professor who plans to give an exam intended to measure what students learned in a course. The exam is to be held with closed books and closed notes, and each student is to take the exam in a fixed period of time, without the use of

Table 16.2 Comparisons of Typical Lab and WWW Experiments

Comparison	Research Method	
	Lab	Web
Control of conditions, ease of observation	Good control	Less control; observation not possible
Measurement of behavior	Precise	Sometimes imprecise; pilot tests, lab vs. Web
Control of stimuli	Precise	Imprecise
Dropouts	A serious problem	A worse problem
Experimenter bias	Can be serious	Can be standardized
Motivation	Complete an assignment	Help out; interest; incentives
Multiple submission	Not considered	Not considered a problem

computers, calculators, or help from others. The professor might give the exam in a classroom with a proctor present to make sure that these rules are followed, or the professor might give the exam via the Internet.

Via the Internet, one could ask the students if they used a computer or got help from others, for example, but one cannot be sure what the students actually did (or even who was taking the exam) with the same degree of certainty as is possible in the classroom. This example should make clear the lack of control in studies done via the WWW.

Precision of Manipulations and Measurements

In the lab, one can control the settings on a monitor, control the sound level on speakers or headphones, and control the noise level in the room. Via the Web, each user has a different monitor, different settings, different speakers or earphones, and a different background of noise and distraction.

In addition to better control of conditions, the lab also typically allows more precise measurement of the dependent variables as well as affording actual observation of the participant.

The measurement devices in the lab include apparatus not currently available via the Web (e.g., EEG, fMRI, eye-trackers). In addition, measurement of dependent variables such as response times via the WWW face a variety of problems resulting from the different conditions experienced by different participants. The participant via the WWW, after all, may be watching TV, may have other programs running, may decide to do some other task on the computer in the middle of the session, or may be engaged in conversation with a roommate. Such sources of uncertainty concerning conditions can introduce both random and systematic error components compared to the situation in the lab. Krantz (2001) reviewed additional ways in which the presentation of stimuli via the WWW lacks the control and precision available in the lab. Schmidt (2001) has reviewed the accuracy of animation procedures.

At the user's (i.e., participant's) end, there may be different types of computers, monitors, sound cards (including no sound), systems, and browsers. It is important to check that the Web experiment works with the major systems (e.g., Windows, Mac, Linux) and different browsers for those systems (Netscape Navigator, Internet Explorer, etc.).

Different browsers may render the same page with different appearances, even on the same computer and system. Suppose the appearance (such as, for example, the spacing of a numerically coded series of radio buttons) is important and is displayed differently by different browsers (Dillman & Bowker, 2001). If so, it may be necessary to restrict which browsers participants can use, or at least keep track of data that are sent by different browsers.

The Need for Pilot Work in the Lab

Because the ability to observe the participant is limited in WWW research, one should do pilot testing of each Web experiment in the lab before launching it into cyberspace. As part of the pilot-testing phase, one should observe people as they work through the online materials and interview them afterward to make sure that they understood the task from the instructions in the Web page.

Pilot testing is part of the normal research process in the lab, but it is even more important in Web research because of the difficulty of communicating with participants. Pilot testing is important to ensure that instructions are clear and that the experiment works properly. In lab research, the participant can ask questions and receive clarifications. If many people ask the same question, the lab assistant will become aware of the problem. In Web research, the experimenter needs to anticipate questions or problems that participants may have and to include instructions or other methods for dealing with them.

I teach courses in which students learn how to conduct their research via the Web. By watching participants in the lab working on my students' studies, I noticed that some participants click choices before viewing the information needed to make their decisions. Based on such observations, I advise my students to place response buttons *below* the material to be read rather than above it, so

that the participant at least has to scroll past the relevant material before responding. It is interesting to rethink paper-and-pencil studies, for which the same problem may have existed but may have gone undetected.

A student of mine wanted to determine the percentage of undergraduates who use illegal drugs. I recommended a variation of the random response method (Musch, Broeder, & Klauer, 2001), in order to protect participant anonymity. In the method I suggested, the participant was instructed to privately toss two coins before answering each question. If both coins fell "heads," the participant was to respond "yes" to the question; if both were "tails," the participant was to respond "no." When the coins were mixed (one heads and one tails), the participant was to respond with the truth. This procedure allows the experimenter to calculate the proportions in the population, without knowing for any given participant whether the answer was produced by chance or truth.

For example, one item asked, "Have you used marijuana in the last 24 hours?" From the overall percentages, the experimenter should subtract 25% "yes" from the "yes" percentage (those who got two heads) and multiply the difference by 2 (because only half of the data are based on truth). For example, if the results showed that overall 30% said "yes," then $30\% - 25\% = 5\%$ (among the 50% of those who had mixed coins and who responded truthfully by saying "yes"); so, 5% times $2 = 10\%$, the estimated percentage who actually used marijuana during the last 24 hours, assuming that the method works.

The elegance of this method is that no one (besides the participant) can ever know from the answers whether or not that person did or did not use marijuana, even if the participant's data were by identified by name! However, the method does require that participants follow instructions.

I observed 15 pilot participants who completed my student's questionnaire in the

laboratory. I noticed that only one participant took out any coins during the session. Indeed, that one person asked, "It says here we are supposed to toss coins; does that mean we should toss coins?" Had this study simply been launched into cyberspace without pilot testing, we might never have realized that most people were not following instructions.

I suggested that my student add instructions emphasizing the importance of following the procedure with the coins and that she add two questions to the end of the questionnaire: "Did you actually toss the coins as you were instructed?" and "If not, why not?" About half of the participants responded that they had *not* followed instructions, giving various reasons, including "lazy," "I had nothing to hide," and "I don't mind telling the truth." Still, even among those who *said* they had followed instructions, one can worry that either by confusion or dissimulation, these students still had not followed instructions. And those who say they followed instructions are certainly not a random sample of all participants.

Clearly, in the lab, one could at least verify that each student has coins, tosses the coins for each question, and gives at least the superficial appearance of following instructions. In the lab, it might be possible to obtain blind urine samples against which the results of the questionnaire method could be compared. Questions about the procedure would be easy for the participant to ask and easy for the experimenter to answer. Via the Web, we can instruct the participant and we can ask the participant if he or she followed instructions, but we cannot really know what the conditions were with the same confidence we have when we can observe and interact with the participant.

The Need for Testing of HTML and Programming

One should also conduct tests of Web materials to make sure that all data coding

and recording are functioning properly. When teaching students about Web research, I find that students are eager to upload their materials to the Web before they have really tested them. It is important to conduct a series of tests to make sure that each radio button functions correctly and that each possible response is coded into the correct value in the data file (see Birnbaum, 2001a, Chapters 11, 12, and 21).

My students tell me that their work has been checked, yet when I check it, I detect many errors that my students have not discovered on their own. It is possible to waste the time of many people by putting unchecked work on the Web. In my opinion, such waste of people's time is unethical; it is a kind of vandalism of scientific resources and a breach of trust with the participant. The participants give us their time and effort in the expectation that we will do good research. They do not intend that their efforts will be wasted. Errors will happen because people are human; knowing this, we have a responsibility to check thoroughly to ensure that the materials work properly before launching a study.

Testing in Both Lab and Web

A number of studies have compared data collected via the WWW with data collected in the laboratory. Indeed, once an experiment has been placed online, it is quite easy to collect data in the laboratory, where the experimenters can better control and observe the conditions in which the participants completed their tasks. The generalization from such research is that despite some differences in results, Web and lab research reach much the same conclusions (Batinic, Reips, & Bosnjak, 2002; Birnbaum, 1999b, 2000a, 2000b, 2001a, 2002; Birnbaum & Wakcher, 2002; Krantz, Ballard, & Scher, 1997; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000a, 2000b). Indeed, in cognitive psychology, it is assumed that if one programs the

experiments correctly, Web and lab results should reach the same conclusions (Francis, Neath, & Surprenant, 2000; McGraw et al., 2000a).

However, one should not expect Web and lab data to agree exactly, for a number of reasons. First, the typical Web and lab studies differ from each other in many ways, and each of these variables might make some difference. Some of the differences are as follow.

Web versus lab studies often compare groups of participants who differ in demographic characteristics. If demographic characteristics affect the results, the comparison of data will reflect these differences (e.g., Birnbaum, 1999b, 2000a).

WWW participants may differ in motivation (Birnbaum, 2001a; Reips, 2000). The typical college student usually participates as one option toward fulfilling an assignment in a lower-division psychology course. Students learn about research from their participation. Because at least one person is present, the student may feel some pressure to continue participation, even though the instructions say that quitting is at any time permitted. Participants from the Web, however, often search out the study on their own. They participate out of interest in the subject matter, or out of desire to contribute to scientific progress. Reips (2000) argued that because of the self-selection and ease of withdrawing from an online study, the typical Web participant is more motivated than the typical lab participant.

Analyses show that Web data can be of higher quality than lab data. For this reason, some consider Web studies to have an advantage over lab studies (Baron & Siepmann, 2000; Birnbaum, 1999b; Reips, 2000). On the other hand, one might not be able to generalize from the behavior of those who are internally motivated to people who are less motivated.

When we compare Web and lab studies, there are often a number of other confounded variables of procedure that might cause significantly different results. Lab studies may use different procedures for displaying the stimuli and obtaining responses. Lab research may involve paper-and-pencil tasks, whereas WWW research uses a computer interface. Lab studies usually have at least one person present (the lab assistant), and perhaps many other people (e.g., other participants). Lab research might use dedicated equipment, specialized computer methods, or manipulations, whereas WWW research typically uses the participant's self-selected WWW browser as the interface. Doron Sonsino (personal communication, 2002) is currently working on pure tests with random assignment of participants to conditions to examine the "pure" effect of the set of Web versus lab manipulations.

Dropouts and Between-Subjects Designs

Missing data, produced by participants who quit a study, can ruin an otherwise good experiment. During World War II, the Allies examined bullet holes in every bomber that landed in the United Kingdom after a bombing raid. A probability distribution was constructed, showing the distribution of bullet holes in these aircraft. The decision was made to add armor to those places where there were *fewest* bullet holes.

At first, the decision may seem in error, perhaps misguided by the gambler's fallacy. To understand why the decision was correct, however, think of the missing data. The dropouts in this research are the key to understanding the analysis. Even though dropouts were usually less than 7%, the dropouts are the whole story. The missing data, of course, were the aircraft that were shot down. The research was not intended to determine where bullets hit aircraft, but rather to determine where bullet holes are in planes that *return*. Here, the correct decision was made to put extra armor around the pilot's seat and to

strengthen the tail because few planes with damage there made it home. This correct decision was reached by having a theory of the missing data.

Between-subjects designs are tricky enough to interpret without having to worry about dropouts. For example, Birnbaum (1999a) showed that in a between-subjects design, with random assignment, the number 9 is rated significantly "bigger" than the number 221. It is important to emphasize that it was a between-subjects design, so no subject judged both numbers. It can be misleading to compare data between subjects without a clear theory of the response scale. In this case, I used my knowledge of range-frequency theory (Parducci, 1995) to devise an experiment that would show a silly result. My purpose was to make people worry about all those other between-subjects studies that used the same method to draw other dubious conclusions.

In a between-subjects design, missing data can easily lead to wrong conclusions. Even when the dropout rate is the same in both experimental and control groups, and even when the dependent variable is objective, missing data can cause the observed effect (in a true experiment with random assignment to conditions) to show the opposite conclusion of the truth.

Birnbaum and Mellers (1989) illustrated one such case in which, even if a treatment is harmful, a plausible theory shows how one can obtain equal dropouts and the harmful treatment appears beneficial in the data. All that is needed is that the correlation between dropping out and the dependent variable be mediated by some underlying construct. For example, suppose an SAT review course is harmful to all who take it; for example, suppose all students lose 10 points by taking the review. This treatment will still look beneficial if the course includes giving each student a sample SAT exam. Suppose those who do well on the sample exam go on to take the

SAT and those who do poorly on the practice test drop out. Even with equal dropout rates, the harmful SAT review will look beneficial because those who do complete the SAT will do better in the treatment group than will the control group.

In Web studies, people find it easy to drop out (Reips, 2000, 2001a, 2001b, 2002). In within-subjects designs, the problem of attrition affects only external validity: Can the results be generalized from those who finish to the sort of people who dropped out?

For between-subjects designs, however, attrition affects internal validity. When there are dropouts in between-subjects designs, it is not possible to infer the true direction of the main effect from the observed effect. Think of the bullet holes case: Dropouts were less than 7%, yet the true effect is opposite the observed effect, because to protect people from bullets, places with the *fewest* bullet holes should get the most armor.

Because of the threat to internal validity of missing data, this topic has received some attention in the growing literature of Web experimentation (Birnbaum, 2000b, 2001a; Frick, Bächtiger, & Reips, 2001; Reips, 2002; Reips & Bosnjak, 2001; Tuten, Urban, & Bosnjak, 2002). An idea showing some promise is the use of the "high threshold" method to reduce dropouts (Reips, 2000, 2002). The idea is to introduce manipulations that are likely to cause dropouts early in the experimental sessions, *before* the random assignment to conditions. For example, ask people for their names and addresses first, then present them with a page that loads slowly, then randomly assign those who are left to the experimental conditions.

Experimenter Bias

A potential advantage of Web research is the elimination of the research assistant. Besides the cost of paying assistants, assistants can bias the results. When assistants

understand the purpose of the study, they might do things that bias the results in the direction expected. There are many ways that a person can reinforce behaviors and interfere with objectivity, once that person knows the research hypothesis (Rosenthal, 1976, 1991).

A famous case of experimenter bias is that of Gregor Mendel, the monk who published an obscure article on a genetic model of plant hybrids that became a classic years after it was published. After his paper was rediscovered, statisticians noticed that Mendel's data fit his theory *too* well. I do not think that Mendel intentionally faked his data, but he probably did see reasons why certain batches that deviated from the average were "spoiled" and should not be counted (see http://www. unb.ca/psychology/likely/evolution/mendel. htm). He might also have helped his theory along when he coded his data, counting medium-sized plants as either "tall" or "short," depending on the running count, to help support his theory.

Such little biases as verbal and nonverbal communication, flexible procedures, data coding, or data entry would be a big problem in certain areas of research, such as ESP or the evaluation of benefits of "talk" psychotherapies, where motivation is great to find positive effects, even if the effects are small. The potential advantage of Web experimentation is that the entry and coding of data are done by the participant and computer, and no experimenter is present to possibly bias the subject or the data entry in a way that is not documented in the materials that control the study.

Multiple Submissions

One of the first questions any Web researcher is asked is "How do you know that someone has not sent thousands of copies of the same data to you?" This concern has been discussed in many papers (Birnbaum, 2001a; Reips, 2000; Schmidt, 1997a, 2000), and the consensus of Web experimenters is that multiple submission of data has not been a serious problem (Musch & Reips, 2000).

When sample sizes are small, as they are in lab research, then if a student (perhaps motivated to get another hour of credit) participated in an experiment twice, then the number of degrees of freedom for the error term is not really what the experimenter thinks it is. For example, if there were a dozen students in a lab study, and if one of them participated twice, then there were really only 11 independent sources of error in the study. The consequence is that the statistical tests need to be corrected for the multiple participation by one person. (See Gonzalez & Griffin, Chapter 14, this volume.)

On the WWW, the sample sizes typically are so large that a statistical correction would be minuscule, unless someone participated a very large number of times. Several methods have been proposed to deal with this potential problem.

The first method to avoid multiple submission is to analyze why people might participate more than once and then take steps to remove those motivations. Perhaps the experiment is interesting and people don't know they should participate only once. In that case, one can instruct participants that they should participate only once.

If the experiment is really enjoyable (e.g., a video game), perhaps people will repeat the task for fun. In that case, one could provide an automatic link to a second Web site where those who have finished the experiment proper could visit and continue to play with the materials as much as they like, without sending data to the real experiment.

If a monetary payment is to be given to each participant, there might be a motive to be paid again and again. Instructions might specify that each person can be paid only once. If the experiment offers a chance at a prize, there might be a motive to participate more than once to give oneself more chances. Again,

instructions could specify that if a person participates more than once, only one chance at the prize is given, or even that a person who submits multiple entries will be excluded from any chance at a prize.

A second approach is to allow multiple participation but ask people how many times they have previously completed the study. The experimenter would then analyze the data of those who have not previously participated separately from those who already have. This method also allows one to analyze how experience in the task affects the results.

A third method is to detect and delete multiple submissions. One technique is to examine the Internet Protocol (IP) address of the computer that sent the data and delete records submitted from the same IP. (One can easily sort by IP or use statistical software to construct a frequency distribution of IP.) The IP does not uniquely identify a participant because most of the large Internet service providers now use dynamic IP addresses and assign their clients one of their IPs as they become available. However, if one person sends multiple copies during one session on the computer, the data would show up as records from the same IP. (Of course, when data are collected in a lab, one expects the same IP to show up again and again because that same lab computer is used repeatedly.)

A fourth method that is widely used is to request identifying information from participants. For example, in an experiment that offers payments or cash prizes, participants are willing to supply identifying information (e.g., their names and addresses) that would be used to mail payments or prizes. Other examples of identifying information are the last four digits of a student's nine-digit ID, e-mail address, a portion (e.g., last four digits) of the participant's Social Security Number, or a password assigned to each person in a specified list of participants.

A fifth method is to check for identical data records. If a study has a large number of items, it is very unlikely that two records will have exactly the same responses.

In my experience, multiple submissions are infrequent and usually occur within a very brief period of time. In most cases, the participant has apparently pushed the *Submit* button to send the data, read the thank you message or debriefing, and used the *Back* button on the browser to go back to the study. Then, after visiting the study again, and perhaps completing the questionnaire, changing a response, or adding a comment, the participant pushes the *Submit* button again, which sends another set of data. If the responses change between submissions, the researcher should have a clear policy of what to do in such cases. In my decision-making research, I want each person's last, best, most considered decision. It is not uncommon to see two records that arrive within 2 minutes in which the first copy was incomplete and the second copy is the same except that it has responses for previously omitted items. Therefore, I always take only the last submission and delete any earlier ones from the same person. There might be other studies where one would take only the first set of data and delete any later ones sent after the person has read the debriefing.

If this type of multiple submission is considered a problem, it is possible to discourage it by an HTML or JavaScript routine that causes the *Back* button not to return to the study's page. Cookies, in this case data stored on the participant's computer indicating that the study has already been completed, could also be used for such a purpose. Other methods, using server-side programming, also are available (Schmidt, 2000); these can keep track of a participant and refuse to accept multiple submissions.

I performed a careful analysis of 1,000 successive data records in decision making (Birnbaum, 2001b), where there were chances at cash prizes and one can easily imagine a motive for multiple entries. Instructions stated

that only one entry per person was allowed. I found 5% were blank or incomplete (i.e., they had fewer than 15 of the 20 items completed) and 2% contained repeated e-mail addresses. In only one case did two submissions come from the same e-mail address more than minutes apart. These came from one woman who participated exactly twice, about a month apart, and who interestingly agreed on 19 of her 20 decisions. Less than 1% of remaining data (excluding incomplete records and duplicate e-mail addresses) contained duplicate IP addresses. In those cases, other identifiers indicated that these were from different people who were assigned the same IPs, rather than from the same person submitting twice. Reips (2000) found similar results, in an analysis done in the days of mostly fixed IP addresses.

ETHICAL ISSUES IN WEB AND LAB

The institutional review of research should work in much the same way for an online study as for an in-lab study. Research that places people at more risk than the risks of everyday life should be reviewed to ensure that adequate safeguards are provided to the participants. The purpose of the review is to determine that the potential benefits of the research outweigh any risks to participants, and to ensure that participants will have been clearly warned of any potential dangers and clearly accepted them. A comparison of ethical issues as related to lab and Web is presented in Table 16.3.

Risks of Psychological Experiments

For the most part, psychology experiments are not dangerous. Undoubtedly, it is less dangerous to participate in the typical 1-hour experiment in psychology than to drive 1 hour in traffic, spend 1 hour in a hospital, serve 1 hour in the armed forces, work 1 hour in a mine or factory, serve 1 hour on a jury, or shop 1 hour at the mall. As safe as psychology experiments are, those done via the Internet must be even safer than lab research because they can remove the greatest dangers of psychology experiments in the lab.

The most dangerous aspect of most psychology experiments is the trip to the experiment. Travel, of course, is a danger of everyday life and is not really part of the experiment; however, this danger should not be underestimated. Every year, tens of thousands are killed in the United States, and millions are injured, in traffic accidents. For example, in 2001, the Department of Transportation reported 42,116 fatalities in the United States and more than 3 million injuries (see http://www.dot.gov/affairs/nhtsa5502.htm).

How many people are killed and injured each year in psychology experiments? I am not aware of a single such case in the last 10 years, nor am I aware of a single case in the 10 years preceding the Institutional Review Board (IRB) system of review. I suspect that if this number were large, professionals in the IRB industry would have made us all aware of them.

The second most dangerous aspect of psychology experiments is the fact that labs often are in university buildings. Such buildings are known to be more dangerous than most residences. For example, many buildings of public universities in California were not designed to satisfy building codes because the state of California previously exempted itself from its own building codes in order to save money.

An earthquake occurred at 4:31 a.m. on January 17, 1994, in Northridge, California (see http: //www.eqe.com/publications/northridge/executiv.htm). Several structures of the California State University at Northridge failed, including a fairly new building that collapsed (see http://www.eqe.com/publications/northridge/commerci.htm). Had the earthquake happened during the day, there undoubtedly

Table 16.3   Ethical Considerations of Lab and Web Research

Ethical Issue	Experimental Setting	
	Lab	Web
Risk of death, injury, or illness	Trip to lab; unsafe public buildings; contagious disease	No trip to lab; can be done from typically safer locations
Participant wants to quit	Human presence may induce compliance	No human present; easy to quit
"Stealing" of ideas	Review of submitted papers and grant applications	Greater exposure at earlier date
Deception	Small numbers—concern is damage to people	Large numbers—risk to science and all society
Debriefing	Can (almost) guarantee	Cannot guarantee; free to leave before debriefing
Privacy and confidentiality	Data on paper, presence of other people, burglars, data in proximity to participants	Insecure transmission (e.g., e-mail); hackers, burglars, increased distance

would have been many deaths resulting from these failures (see http://geohazards.cr.usgs.gov/northridge/). Some of those killed might have been participants in psychology experiments.

Certainly, the risks of such dangers as traffic accidents, earthquakes, communicable diseases, and terrorist attacks are not dangers caused by psychology experiments. They are part of everyday life. However, private dwellings usually are safer from these dangers than are public buildings, so by allowing people to serve in experiments from home, Web studies must be considered less risky than lab studies.

Ease of Dropping Out From Online Research

Online experiments should be more acceptable to IRBs than lab experiments for another reason. It has been demonstrated that if an experimenter simply tells a person to do something, most people will "follow orders," even if the instruction is to give a potentially lethal shock to another person (Milgram, 1974). Therefore, even though participants are free to leave a lab study, the presence of other people may cause people to continue who might otherwise prefer to quit. However, Web studies do not usually have other people present, so people tested via the WWW find it easy to drop out at any time, and many do.

Although most psychological research is innocuous and therefore legally exempt from review, most psychological research is reviewed anyway, either by a campus-level IRB or by a departmental IRB. It is interesting that nations that do not conduct prior review of psychology experiments seem to have no more deaths, injuries, property damage, or hurt feelings in their psychology studies than we have in the United States, where there is extensive review. Mueller and Furedy (2001) have questioned if the IRB system in the United States is doing more harm than good. The time and resources spent on the review process seems an expense that is unjustified by its meager benefits. We need a system that quickly recognizes research that is exempt and identifies it as such, in order to save valuable

scientific resources. For more on IRB review, see Kimmel (Chapter 3, this volume).

Ethical Issues
Peculiar to the WWW

In addition to the usual ethical issues concerning the safety of participants, there are ethical considerations in Web research connected with the impact of one's actions on science and other scientists. Everyone knows a researcher who holds a grudge because he or she thinks that another person has "stolen" an idea revealed in a manuscript or grant application during the review process. Very few people see manuscripts submitted to journals or granting agencies, certainly fewer than would have access to a study posted on the Web. Furthermore, a Web study shows what a person is working on well ahead of the time that a manuscript would be under review. There undoubtedly will be cases in which one researcher will accuse another of stealing ideas from his or her Web site without giving proper credit.

There is a tradition of "sharing" on the Web; people put information and resources on the Web as a gift to the world. However, Web pages should be considered published, and one should acknowledge credit to Web sites in the same way that one cites journal articles that have contributed to one's academic work.

Similarly, one should not interfere with another scientist's research. "Hacking" into another person's online research would be illegal as well as improper, but even without hacking, a person can do many things that could adversely affect another's research program. We should proscribe interference of any kind with another's research, even if the person who takes the action claims to act out of good motives.

Deception on the WWW

Deception concerning inducements to participate would be a serious breach of ethics

that not only would annoy participants but would affect others' research as well. For example, if a researcher promised to pay each participant $200 for 4 hours and then failed to pay the promised remuneration, such fraud not only would be illegal but also would give a bad name to all psychological research.

Similarly, promising to keep sensitive data confidential and then revealing such personal information in the newspaper would be the kind of behavior we expect from a reporter in a nation with freedom of the press. Although we expect such behavior from a reporter, and we protect it, we do not expect it from a scientist. Scientists should keep their word, even when they act as reporters.

Other deceptions, such as were once popular in social psychology, pose another tricky ethical issue for Web-based research. If Web researchers were to use deception, it is likely that such deceptions would become the source of Internet "flames," public messages expressing anger and disapproval. The deception thus would become ineffective, but worse, people would soon come to doubt anything said by a psychologist. Such cases could easily give psychological research on the Web a bad reputation. The potential harm of deception to science (and therefore to society as a whole) probably is greater than the potential harm of such deception to participants, who would be more likely annoyed than harmed.

Deception in the lab may be discovered and research might be compromised at a single institution for a limited time among a limited number of people. However, deception on the Web could easily create a long-lasting bad reputation for all of psychology among millions of people. One of the problems of deceptive research is that the deception does not work—especially on the Web, it would be easy to expose and publicize to a vast number of people. If scientists want truthful data from participants, it seems a poor idea to have a dishonest reputation.

Probably the best rule for psychological experiments on the WWW is that false information should not be presented via the Web.

I do not consider it to be deception for an experimenter to describe an experiment in general layman's terms without identifying the theoretical purpose or the independent variables of a study. The rule against false information does not require that a person give complete information. For example, it is not necessary to inform subjects in a between-subjects design that they have received Condition A rather than Condition B.

Privacy and Confidentiality

For most Web-based research, the research is not sensitive, participants are not identified, and participation imposes fewer risks than those of "everyday life." In such cases, IRBs should treat the studies as exempt from review.

In other cases, however, data might be slightly sensitive and partially identifiable. In such cases, identifiers used to detect multiple submissions (e.g., IP addresses) can be cleaned from data files once they have been used to accomplish their purpose.

As is the case with sex surveys conducted via personal interview or questionnaire, it usually is not necessary to keep personal data or to keep identifiers associated with their data. However, there may be such cases, for example, when behavioral data are to be analyzed with medical or educational data for the same people. In such cases, security of the data may require that data be sent or stored in encrypted form. Such data should be stored on a server that is well protected both from burglars who might steal the computer (and its hard drive) and from hackers who might try to steal the electronic information. If necessary, such research might rely on the https protocol, such as that used by online shopping services to send and receive credit card numbers, names, and addresses.

As we know from the break-in of Daniel Ellsberg's psychiatrist's office, politicians and government officials at times seek information to discredit their "rivals" or political "enemies." For this reason, it is useful to take adequate measures to ensure that information is stored in such a way that if the records fell into the wrong hands, they would be of no use to those who stole them. Personal identifiers, if any, should be removed from data files as soon as is practical.

A review of Web researchers (Musch & Reips, 2000) found that "hackers" had not yet been a serious problem in online research. Most of the early studies, however, dealt with tasks that were not sensitive or personal— nothing of interest to hackers. However, imagine the security problems of a Web site that contained records of bribes to public officials, or lists of their sexual partners. Such a site would be of great interest to members of the tabloid press and political opponents. Indeed, the U.S. Congress set a bad example in the Clinton-Lewinsky affair by posting to the Web testimony in a grand jury hearing that had not been cross-examined, even though such information is sealed by law. In most psychology research, with simple precautions, it is probably more likely that information would be abused by an employee (e.g., an assistant) on the project than by a "hacker" or burglar.

The same precautions should be taken to protect sensitive, personal data in both lab and Web. Don't leave your key to the lab around, and don't leave or give out your passwords. Avoid storing names, addresses, or other personal information (complete Social Security Numbers) in identifiable form. It may help to store data on a different server from the one used to host the Web site. Data files should be stored in folders to which only legitimate researchers have password access. The server that stores the data should be in a locked room, protected by strong doors and locks, whose exact location is not public.

In most cases, such precautions are not needed. Browsers automatically warn a person when a Web form is being sent unencrypted, with a message to the effect that "the data are being sent by e-mail and can be viewed by third parties while in transit. Are you sure you want to send them?" Only when a person clicks a second time, assuming the person has not turned this warning off, will the data be sent. A person makes an informed decision to send such data, just as a person accepts lack of privacy when using e-mail. The lack of security of e-mail is a risk most people accept in their daily lives. Participants should be made aware of this risk but not bludgeoned with it.

I don't believe it is necessary to insult or demean participants with excessively lengthy warnings of implausible but imaginable harms in informed consent procedures. Such documents have become so wordy and legalistic that some people agree or refuse without reading. The purpose of such lengthy documents seems to me (and to our participants) to be designed to protect the researcher and the institution rather than the participant and society.

Good Manners on the Web

There are certain rules of etiquette on the WWW, known as "Netiquette." Although these, like everything else on the WWW, are in flux, there are some that you should adopt to avoid getting in trouble with vast numbers of people.

1. Avoid sending unsolicited e-mails to vast numbers of people unless it is reasonable to expect that the vast majority *want* to receive your e-mail. If you send such "spam," you will spend more time responding to the "flames" (angry messages) than you will spend analyzing your data. If you want to recruit in a special population, ask an organization to vouch for you. Ask leaders of the organization to send the message describing your study and why it would be of benefit to the members of the list to participate.

2. Do not send attachments of any kind in any e-mail addressed to people who don't expect an attachment from you. Attachments can carry viruses, and they clog up mailboxes with junk that could have been better posted to the Web. If you promised to provide the results of your study to your participants, you should post your paper on the Web and just send your participants the URL, not the whole paper. If the document is put on the WWW as HTML, the recipient can safely click to view it. If your computer crashed after you opened an attachment from someone, wouldn't you suspect that attachment to be the cause of your misery? If you send attachments to lots of people, odds are that someone will hold you responsible.

3. Do not use any method of recruiting that resembles a "chain letter."

4. Do not send blanket e-mails with readable lists of recipients. How would you feel if you got an e-mail asking you to participate in a study of pedophiles, and you saw your name and address listed among a group of registered sex offenders?

5. If you must send e-mails, keep them short, to the point, and devoid of any fancy formatting, pictures, graphics, or other material that belongs on the Web. Spare your recipient the delays of reading long messages, and give them the choice of visiting your materials.

CONCLUDING COMMENTS

Because of the advantages of Web-based studies, I believe use of such methods will continue to increase exponentially for the next decade. I anticipate a period in which each area of social psychology will evaluate Web methods to decide whether they are suitable to that area's paradigm. In some areas of research, such as social judgment and decision making,

I think that the method will be adopted rapidly and soon taken for granted. As computers and software improve, investigators will find it easier to create, post, and advertise their studies on the WWW. Eventually, investigators will regard Web-based research as they now regard using a computer rather than a calculator for statistics, or using a computer rather than a typewriter to prepare a manuscript.

REFERENCES

Bailey, R. D., Foote, W. E., & Throckmorton, B. (2000). Human sexual behavior: A comparison of college and Internet surveys. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 141-168). San Diego: Academic Press.

Baron, J., & Siepmann, M. (2000). Techniques for creating and using Web questionnaires in research and teaching. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 235-265). San Diego: Academic Press.

Batinic, B. (Ed.). (1997). *Internet für Psychologen.* Göttingen, Germany: Hogrefe.

Batinic, B., Reips, U.-D., & Bosnjak, M. (Eds.). (2002). *Online social sciences.* Seattle: Hogrefe & Huber.

Birnbaum, M. H. (1997). Violations of monotonicity in judgment and decision making. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 73-100). Mahwah, NJ: Lawrence Erlbaum.

Birnbaum, M. H. (1999a). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods, 4*(3), 243-249.

Birnbaum, M. H. (1999b). Testing critical properties of decision making on the Internet. *Psychological Science, 10,* 399-407.

Birnbaum, M. H. (2000a). Decision making in the lab and on the Web. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 3-34). San Diego: Academic Press.

Birnbaum, M. H. (Ed.). (2000b). *Psychological experiments on the Internet.* San Diego: Academic Press.

Birnbaum, M. H. (2000c). SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet. *Behavior Research Methods, Instruments, & Computers, 32,* 339-346.

Birnbaum, M. H. (2001a). *Introduction to behavioral research on the Internet.* Upper Saddle River, NJ: Prentice Hall.

Birnbaum, M. H. (2001b). A Web-based program of research on decision making. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 23-55). Lengerich, Germany: Pabst Science Publishers.

Birnbaum, M. H. (2002). Wahrscheinlichkeitslernen. In D. Janetzko, M. Hildebrand, & H. A. Meyer (Eds.), *Das Experimental-psychologische Praktikum im Labor und WWW* (pp. 141-151). Göttingen, Germany: Hogrefe.

Birnbaum, M. H., & Hynan, L. G. (1986). Judgments of salary bias and test bias from statistical evidence. *Organizational Behavior and Human Decision Processes, 37,* 266-278.

Birnbaum, M. H., & Martin, T. (in press). Generalization across people, procedures, and predictions: Violations of stochastic dominance and coalescing. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on decision research.* New York: Cambridge University Press.

Birnbaum, M. H., & Mellers, B. A. (1989). Mediated models for the analysis of confounded variables and self-selected samples. *Journal of Educational Statistics, 14,* 146-158.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17,* 49-78.

Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Violations of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes, 77,* 44-83.

Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of heredity and environment. *Journal of Experimental Social Psychology, 17,* 159-182.

Birnbaum, M. H., & Wakcher, S. V. (2002). Web-based experiments controlled by JavaScript: An example from probability learning. *Behavior Research Methods, Instruments, & Computers, 34,* 189-199.

Buchanan, T. (2000). Potential of the Internet for personality research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 121-140). San Diego: Academic Press.

Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide Web. *British Journal of Psychology, 90,* 125-144.

Dillman, D. A., & Bowker, D. K. (2001). The Web questionnaire challenge to survey methodologists. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 159-178). Lengerich, Germany: Pabst Science Publishers.

Francis, G., Neath, I., & Surprenant, A. M. (2000). The cognitive psychology online laboratory. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 267-283). San Diego: Academic Press.

Frick, A., Bächtiger, M. T., & Reips, U.-D. (2001). Financial incentives, personal information, and drop-outs in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 209-219). Lengerich: Pabst Science Publishers.

Huff, D. (1954). *How to lie with statistics.* New York: Norton.

Janetzko, D., Meyer, H. A., & Hildebrand, M. (Eds.). (2002). *Das Experimental-psychologische Praktikum im Labor und WWW* [A practical course on psychological experimenting in the laboratory and WWW]. Göttingen, Germany: Hogrefe.

Joinson, A. (2002). *Understanding the psychology of Internet behaviour: Virtual worlds, real lives.* Basingstoke, Hampshire, UK: Palgrave Macmillan.

Krantz, J. H. (2001). Stimulus delivery on the Web: What can be presented when calibration isn't possible? In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet Science* (pp. 113-130). Lengerich, Germany: Pabst Science Publishers.

Krantz, J. H., Ballard, J., & Scher, J. (1997). Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. *Behavior Research Methods, Instruments, & Computers, 29,* 264-269.

Krantz, J. H., & Dalal, R. (2000). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: Academic Press.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first order gambles. *Journal of Risk and Uncertainty, 4,* 29-59.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000a). The integrity of Web-based experiments: Can you trust the data? *Psychological Science, 11,* 502-506.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000b). PsychExps: An on-line psychology laboratory. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 219-233). San Diego: Academic Press.

McKenna, K.Y.A., & Bargh, J. A. (2000). Plan 9 from cyberspace: The implications of the Internet for personality and social psychology. *Personality and Social Psychology Review, 4*(1), 57-75.

Milgram, S. (1974). *Obedience to authority*. New York: Harper & Row.

Mueller, J., & Furedy, J. J. (2001). The IRB review system: How do we know it works? *American Psychological Society Observer, 14*. Retrieved March 23, 2003, from http://www.psychologicalscience.org/observer/0901/irb_reviewing.html

Musch, J., Broeder, A., & Klauer, K. C. (2001). Improving survey research on the World-Wide Web using the randomized response technique. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 179-192). Lengerich, Germany: Pabst Science Publishers.

Musch, J., & Reips, U.-D. (2000). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-87). San Diego: Academic Press.

Pagani, D., & Lombardi, L. (2000). An intercultural examination of facial features communicating surprise. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 169-194). San Diego: Academic Press.

Parducci, A. (1995). *Happiness, pleasure, and judgment*. Mahwah, NJ: Lawrence Erlbaum.

Pettit, F. A. (1999). Exploring the use of the World Wide Web as a psychology data collection tool. *Computers in Human Behavior, 15*, 67-71.

Piper, A. I. (1998). Conducting social science laboratory experiments on the World Wide Web. *Library & Information Science Research, 20*, 5-21.

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model.* Boston: Kluwer.

Reips, U.-D. (1997). Das psychologische Experimentieren im Internet. In B. Batinic (Eds.), *Internet für Psychologen* (pp. 245-265). Göttingen, Germany: Hogrefe.

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-117). San Diego: Academic Press.

Reips, U.-D. (2001a). Merging field and institution: Running a Web laboratory. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 1-22). Lengerich, Germany: Pabst Science Publishers.

Reips, U.-D. (2001b). The Web experimental psychology lab: Five years of data collection on the Internet. *Behavior Research Methods, Instruments, & Computers, 33*, 201-211.

Reips, U.-D. (2002). Standards for Internet experimenting. *Experimental Psychology, 49*(4), 243-256.

Reips, U.-D., & Bosnjak, M. (2001). *Dimensions of Internet science*. Lengerich, Germany: Pabst Science Publishers.

Rosenthal, R. (1976). *Experimenter effects in behavioral research: Enlarged edition.* New York: Irvington.

Rosenthal, R. (1991). Teacher expectancy effects: A brief update 25 years after the Pygmalion experiment. *Journal of Research in Education, 1*, 3-12.

Schillewaert, N., Langerak, F., & Duhamel, T. (1998). Non-probability sampling for WWW surveys: A comparison of methods. *Journal of the Market Research Society, 40*(4), 307-322.

Schmidt, W. C. (1997a). World-Wide Web survey research: Benefits, potential problems, and solutions. *Behavioral Research Methods, Instruments, & Computers, 29*, 274-279.

Schmidt, W. C. (1997b). World-Wide Web survey research made easy with WWW Survey Assistant. *Behavior Research Methods, Instruments, & Computers, 29*, 303-304.

Schmidt, W. C. (2000). The server-side of psychology Web experiments. In M. H. Birnbaum (Ed.), *Psychological Experiments on the Internet* (pp. 285-310). San Diego: Academic Press.

Schmidt, W. C. (2001). Presentation accuracy of Web animation methods. *Behavior Research Methods, Instruments & Computers, 33,* 187-200.

Schmidt, W. C., Hoffman, R., & MacDonald, J. (1997). Operate your own World-Wide Web server. *Behavior Research Methods, Instruments, & Computers, 29,* 189-193.

Smith, M. A., & Leigh, B. (1997). Virtual subjects: Using the Internet as an alternative source of subjects and research environment. *Behavior Research Methods, Instruments, & Computers, 29,* 496-505.

Stern, S. E., & Faber, J. E. (1997). The lost e-mail method: Milgram's lost-letter technique in the age of the Internet. *Behavior Research Methods, Instruments, & Computers, 29,* 260-263.

Tuten, T. L., Urban, D. J., & Bosnjak, M. (2002). Internet surveys and data quality: A review. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 7-26). Seattle: Hogrefe & Huber.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297-323.

Wallace, P. M. (2001). *The psychology of the Internet.* Cambridge, UK: Cambridge University Press.

Welch, N., & Krantz, J. H. (1996). The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: Experience and results. *Behavior Research Methods, Instruments, & Computers, 28,* 192-196.