

HUMAN RESEARCH AND DATA COLLECTION VIA THE INTERNET

Michael H. Birnbaum

*Department of Psychology, California State University, Fullerton,
California 92834–6846; email: mbirnbaum@fullerton.edu*

Key Words anonymity, browser-based research, CGI, computers in psychology, computer programming, experimental dropouts, experimenter bias, HTML, HTML forms, Internet research, Java, JavaScript, methodology, missing data, on-line research, Perl, survey, Web experiment, WWW

■ **Abstract** Advantages and disadvantages of Web and lab research are reviewed. Via the World Wide Web, one can efficiently recruit large, heterogeneous samples quickly, recruit specialized samples (people with rare characteristics), and standardize procedures, making studies easy to replicate. Alternative programming techniques (procedures for data collection) are compared, including client-side as opposed to server-side programming. Web studies have methodological problems; for example, higher rates of drop out and of repeated participation. Web studies must be thoroughly analyzed and tested before launching on-line. Many studies compared data obtained in Web versus lab. These two methods usually reach the same conclusions; however, there are significant differences between college students tested in the lab and people recruited and tested via the Internet. Reasons that Web researchers are enthusiastic about the potential of the new methods are discussed.

CONTENTS

INTRODUCTION	804
Computer Developments that Set the Stage	804
TECHNIQUES	806
Getting Started: HTML Forms	806
Client-Side Programming	807
Server-Side Programming	809
Run Your Own Server	811
Stimulus Delivery	811
METHODOLOGY	813
Multiple Submissions	813
Dropouts in Between-Subjects Research	816
Recruiting Participants Via the Web	818
Sampling Bias and Stratified Analysis	820

Response Bias	821
Experimenter Bias	822
PILOT TESTING IN THE LAB	822
WEB VERSUS LAB	824
SUMMARY AND CONCLUSIONS	827

INTRODUCTION

In the last decade, a new protocol for sending information on the World Wide Web (WWW), hypertext transfer protocol (HTTP), created a new way to conduct psychological research. This new technique allows researchers to collect data from participants all over the world 24 hours a day and seven days per week. Surveys and experiments can be delivered quickly to anyone connected to the Web and data can be saved automatically in electronic form, reducing costs in lab space, dedicated equipment, paper, mailing costs, and labor. Once an experiment or survey is properly programmed, data can be stored in a form ready for analysis, saving costs of data coding and entry that used to be an expensive and time-consuming part of the research process.

Computer Developments that Set the Stage

By early 1995, a number of changes were under way that created the conditions for a new approach to psychological research:

- Computer hardware and software improved; it became less expensive to own computers and easier to operate them.
- The number of individuals who owned or had access to personal computers, email, and the WWW was increasing exponentially.
- More and more valuable content was being added to the WWW.
- Speed of connections to the Internet were improving and browsers, the programs that display WWW content, were now able to handle more complex media and programming languages of Java and JavaScript (initially called “LiveScript”).
- Hypertext markup language (HTML) 2 supported the technique of forms.

The new standard of HTML 2, introduced in late 1994, supported forms, which allowed a person viewing a Web page to easily send data back to a designated server, which could process, code, filter, and save data electronically. This technique made it possible for a person, even without an email account, to be able to participate in a survey from an Internet-connected computer (e.g., in a public library or campus computer facility) even if that computer was not configured to send email.

Within a few months, psychologists began using this technique to collect data in surveys and experiments. A number of these “pioneers” contributed chapters

to Birnbaum's (2000b) edited book, which included a chapter by Musch & Reips (2000) that summarized the (short) history of psychological experiments (not surveys) on the WWW [see also Musch (2000)].

Musch & Reips (2000) noted that the first psychological experiments (with manipulated variables) conducted via the Web were those of Welch & Krantz (1996) and Krantz et al. (1997). Krantz (1998) maintains a Web site, "Psychological Research on the Net," (<http://psych.hanover.edu/research/exponnet.html>) that lists experiments currently running on the Internet. The site had 35 links to on-line studies on June 17, 1998. By May 11, 1999, there were 65 links, and on May 10, 2003, there were 150 links, including 45 to projects in social psychology and 30 to studies of cognitive psychology.

Several of these links lead to multiple experiments. For example, Jonathan Baron (<http://www.psych.upenn.edu/~baron/qs.html>) has been running about 50 studies per year from his site. The PsychExps site at Ole Miss (<http://psychexps.olemiss.edu/>) listed 33 lab and research studies on May 10, 2003, most of which were in cognitive psychology. Similarly, the student research project site at my Decision Research Center of Fullerton (<http://psych.fullerton.edu/mbirnbaum/decisions/thanks.htm>) listed 21 studies on that same date. Ulf-Dietrich Reips's Web Experimental Psychology Lab (<http://www.psychologie.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html>), which began operations in 1995, had links to 16 active studies in English and 12 in German plus about 80 in its archive on this same date in 2003. The Web Experiment List (<http://genpsylab-wexlist.unizh.ch/>) listed 166 studies in 2003. Although not all studies are listed in these sites (and some items are duplicates), the numbers give an indication of the rapid expansion of this method of doing research. A few hours visiting the links would convey to the reader a general impression of the kinds of studies being done this way.

Although any given work may have more than one focus, I divide my review in three distinct topic areas, according to my judgment of a work's main thrust. These three areas are (a) Techniques that describe or analyze "how-to" issues. For example, how can we randomly assign participants to between-subjects conditions? (b) Web methodology, which deals with internal and external validity of Internet research. For example, what are the threats to internal and external validity if people drop out of Web-based experiments, and what are the considerations of using particular techniques to reduce such dropouts? (c) Web versus lab comparisons, which includes findings specific to the Internet and comparisons of Web and lab studies of the same phenomena. For example, do experiments in decision making that are conducted via the Web yield the same conclusions as those done in the lab? In what ways do results obtained in Web and lab differ?

A fourth major area of research involves social psychology of the Internet, which deals with the Internet as a new social situation or communication medium. For example, do people exhibit different social behaviors on-line and in person? This topic is the focus of a separate review (Bargh & McKenna 2004) and will not be covered here except in relation to methodological issues in Web research.

TECHNIQUES

Many books and articles describe techniques that can be useful for conducting research via the WWW. Items that are primarily commercial and directed to the general audience (e.g., books on HTML, JavaScript, Java, Perl, etc.) are not summarized here, but those that are specifically directed to psychological researchers are discussed.

Getting Started: HTML Forms

Probably the easiest way to get started with the techniques of Web-based research is to make a survey or experiment using one of the free programs (e.g., Birnbaum 1998) to create the Web page for a simple survey or factorial experiment. These programs are available from the following URL: <http://psych.fullerton.edu/mbirnbaum/programs/>.

SurveyWiz and FactorWiz are Web pages that make Web pages for collecting data via the WWW. Within each Web page is a list of instructions for use (see also Birnbaum 2000c, 2001a). Creating a Web form with SurveyWiz is as simple as typing the questions and pushing buttons for the type of input device desired. The program supports text boxes, which are boxes into which a participant can respond by typing a number or short answer. SurveyWiz and FactorWizRB also support radio buttons, which allow the user to click along a rating scale.

The most fundamental technique of Web-based research is HTML, the formatting language used to compose Web pages. A Web page (a document) can contain formatted text, links to other information on the Web, pictures, graphics, animations, and other media such as sounds or video. Many software products are available to create Web pages, including free ones such as Birnbaum's (2000c) FactorWiz, WEXTOR by Reips & Neuhaus (2002), Schmidt's (1997b) WWW Survey Assistant, and White & Hammer's (2000) Quiz-o-matic (which makes self-scoring quizzes). Despite the availability of free and commercial software to make Web pages, a Web researcher needs to have a basic understanding of the "tags" (commands) of HTML. There are many free tutorials available on the WWW for learning HTML, as well as many commercial books on the subject.

The only introductory text in English on Web research is Birnbaum's (2001a), intended for behavioral researchers who are new to Web research. It includes a brief introduction to HTML, including the technique of forms (Chapters 2–5), spreadsheet and statistical software useful in Web research (Chapters 6, 7, and 12–15), an introduction to JavaScript (Chapters 17–19), and a brief introduction to advanced topics such as Java, Authorware, and other methods (Chapter 20). Several chapters cover psychological content with examples of how to program the experiments, explanations of the examples, and instructions on how to analyze the data. A compact disk (CD) accompanies the book with many programming examples, experiments that are ready to run, and data for those experiments. There are also chapters on methodology, recruiting, and ethics of Web studies. Materials

for using this book with both undergraduate and graduate students are available from the following URL: <http://psych.fullerton.edu/mbirnbaum/psych466/>.

Books by Janetzko (1999) and Janetzko et al. (2002) also include much useful information on how to execute psychology experiments via the WWW. The latter is an edited book with useful treatments of methodology and statistics, as well as many interesting examples of experiments written by authors of the chapters, ready to run from the accompanying CD. Unfortunately, these works are available only in German at this time.

It is worth mentioning that psychologists in German-speaking nations took to Web research very quickly. The first edition of *Internet für Psychologen* (Batinic 2000) appeared in 1997, and the *German On-line Research Society* began its annual meetings that year (Batinic et al. 1999). Within two years, the society invited papers in English as well as German (e.g., Birnbaum 1999c), and recently, much of their work has become available in English (Batinic et al. 2002, Reips 2001b, Reips & Bosnjak 2001).

Client-Side Programming

The “server” is the computer that stores Web pages and delivers (“serves”) them to people who send requests to view them. The “browsers,” programs that request and display Web pages, or the computers that run programs to contact the server, are known as the clients. Client-side programs run on the participant’s computer rather than on the server. They are often implemented by a “plug-in,” an extra software component that helps Web browsers run such programs. Modern Web browsers, such as Internet Explorer, Netscape Navigator, and Mozilla (among others), come with JavaScript and Java installed, which are the two most popular languages for client-side programming. These languages are “free,” in that neither programmer nor participant needs to buy anything.

Authorware programs, which also run on the client computer, are executed by means of the Authorware Player, a plug-in that can be downloaded free; however, the Authorware program (used to create experiments, demonstrations, or other content) is expensive.

JavaScript is a programming language that has many uses in Web-based research (Birnbaum 2001a). JavaScript programs can be included as source code in a Web page, and the program loads and runs on the client (i.e., the participant’s machine). By running on the client computer, the program does not burden the server or network with computation and communication delays, which might slow down interactions within an experiment. Including source code in the Web page makes it easy for other researchers to study the programs used and to build upon them. Commercial programmers do not want people to be able to copy and reuse their work, but academic scientists thrive when knowledge is available for inspection, criticism, and modification.

A potential problem with running a program on the client’s computer is that one relies on the participant to have the proper software installed and running.

If a person did not have a compatible browser with the proper plug-ins, then client-side programs would not work for that user. Indeed, many users turned off JavaScript and Java when these programs first became available, fearing security gaps. Today, such fears have mostly been allayed by good experiences with the programs and concerns have been outweighed by the value they provide. Although one now expects to find JavaScript and Java on most users' machines, it is still the case that most users would need to make a special effort to download and install certain plug-ins such as the Authorware player, even though such plug-ins are free.

JavaScript is the focus of several works demonstrating its usefulness in psychological research. Baron & Siepmann (2000) show how it can be used to control questionnaire studies, with materials randomized for each participant. Birnbaum & Wakcher (2002) illustrate how to use JavaScript to control a probability learning study (Birnbaum 2002). They also provide a link to a brief tutorial with references to useful resources, at the following URL: <http://psych.fullerton.edu/mbirnbaum/brmic/>. Lange (1999) shows how JavaScript can be used to accomplish some interesting and useful effects in her on-line museum of perception, and Schmidt (2001) reviews its use in delivering animations on the Web.

Java is a very powerful programming language (which should not be confused with JavaScript). Java can be used to make both stand-alone programs and applets (small programs) that can be precompiled and sent as byte codes along with a Web page, in much the same way that images are sent and incorporated in a page's display. The browser then uses its Java engine to execute the programs on the client computer. For discussions of this approach, with many excellent applications to cognitive psychology experiments, see Francis et al. (2000) and Stevenson et al. (1999). McClelland's (2000) on-line book, *Seeing Statistics*, has a number of Java applets that allow one to "see" and manipulate graphical displays to get a better understanding of fundamental statistical ideas (available at <http://www.seeingstatistics.com/>).

Instruction on Java by McClelland is available from the Advanced Training Institutes in Social Psychology (ATISP) site: <http://ati.fullerton.edu/>.

An advantage of Java is the power provided by a fully object-oriented language, with superior control of graphics. In computer programming, an "object" can be conceived as a block of code that is abstract enough that it can be specialized in new ways or generalized for new purposes that were not part of its original design. An object can combine the roles of functions and data structures. A key to programming in the language is to learn how to create objects that will have the greatest generality, can be used as building blocks, and can be reused in new ways. These concepts can be difficult to master, but once a person has learned the language, it opens up many possibilities for controlling graphics and measuring events (such as the position of the mouse) that represent behaviors of the participant. The language of Java is still in flux, so programmers have had to revise their programs to preserve compatibility as new versions of the language were introduced (Francis et al. 2000).

Authorware is a commercial program that can be used to do many of the same tasks as can be accomplished by Java. A possible advantage of Authorware is that it is not as difficult to learn as a programming language. One can create experiments (or other interactive media, such as tutorials) by means of a graphical user interface on which icons representing loops, decisions, and other high-level operations can be placed on a flow-line. Authorware can be used to create cognitive psychology experiments, with accurate control of stimulus timing and measurement of response times (McGraw et al. 2000a,b; Williams et al. 1999). The PsychExps site at Ole Miss is a powerful, virtual laboratory in cognitive psychology, in which investigators from many institutions use the site to run new experiments and lab demonstrations of classic experiments. Instructional material on the Authorware approach is available from the ATISP Web site and from the PsychExps Web site: <http://psychexps.olemiss.edu/>.

The goal of client-side programming (in JavaScript, Java, Authorware, etc.) is that the program should run equally well on any client's system and browser. However, this ideal has yet to be fully achieved, and one must test any new program with the major systems (Mac, Windows, etc.), major browsers (Internet Explorer, Netscape Navigator, Mozilla, etc.), and various versions of those browsers. Because languages such as Java and JavaScript are still young and developing, a programmer needs to be flexible and energetic to maintain code that will stay compatible with new forms of these languages.

Server-Side Programming

By using a server-side common gateway interface (CGI) program, one can at least guarantee that all users will be able to run the experiment. A server-side program runs on the server, so it does not require the participant to have any special hardware or software (beyond the basic ability to read HTML Web pages).

Schwarz & Reips (2001) compared the use of JavaScript against CGI in a Web experiment and found more and more attrition (more dropouts) in the JavaScript condition compared to the CGI condition, as users worked through the experiment. When their research was conducted, many Web users lacked compatible JavaScript or had it turned off in their browsers. In addition, Internet Explorer's version of JavaScript was "buggy" and not completely consistent with Navigator's. Schwarz & Reips did not find any significant effects of this variable, however, on their theoretical conclusions in a study of the hindsight bias.

Today, so many Web sites depend on JavaScript that it would be frustrating to spend much time on the Web without it. I suspect that if the Schwarz & Reips study were repeated today, there would be less of a difference in attrition between these two ways of controlling an experiment, assuming both were equally well programmed.

Perl is a "free" and open language that can be downloaded and used without charge. It is probably the most widely used language for server-side, CGI programming. Many Web sites are devoted to this language. Schwartz (1998) presents a

brief tutorial on Perl. Instructional material by Schmidt (2000) on server-side programming, with links to useful resources, can be retrieved from the Web site from the ATISP site.

Many different ways are often available to accomplish the same task in computer programming, including programming for the Web. But the pros and cons of different methods should be considered in selecting the “best” way to handle a given situation. Take, for example, the problem of random assignment to between-subjects conditions. This can be accomplished by a CGI script running on the server, which means that it will work for all users but may slow down the experiment as the participant waits for a response from the server. When participants wait for a response from the server, they may think that the experiment “froze up” and quit the experiment. Alternately, a client-side program, such as a JavaScript routine, can handle random assignment to conditions; in this case it frees the server, but it may not work for all users. Finally, random assignment can be handled by very simple HTML, which ensures it works and works quickly for the greatest number of participants.

A simple solution for random assignment via HTML is to create a page with a list of months and ask the participant to click the month of his or her birth. Each month would be linked to one of the conditions of the study. For example, January might be assigned to Condition 1, February to Condition 2, March to Condition 1, etc. During the course of the study, the association of birth months to conditions can be counterbalanced to allow a check for any horoscope effects and to equalize the probability of assignment to conditions. Birnbaum (1999a) used this approach.

Another variation of this HTML technique would be to ask the user to click on the last digit of some form of ID, such as driver’s license number, Social Security number, etc., with different digits linked to different conditions. The assignment of digits to conditions can be counterbalanced (in a Latin Square, for example) over time.

When choosing such a random device, it is important to be clear that the outcomes can still be random even when outcomes may not be equally probable. For example, there are more odd birthdays (date within the month) than even ones, so if people were assigned to two conditions by odd or even birthdays, one expects slightly more than half of participants to end up in the odd condition. Counterbalancing the assignment equalizes the numbers in the two groups as well as making possible a test of whether being born on an odd- or even-numbered birthday makes a difference for the behavior under examination.

A simple JavaScript program can also handle random assignment to conditions; for example, see Birnbaum (2001a, Chapter 17). Birnbaum’s example script can be easily modified for any number of conditions. In the script, a random number is selected using the pseudo random number generator of JavaScript. In case the user does not have JavaScript installed, the Web page reverts to the HTML birthday method. Other client-side methods to handle random assignment include Java (Francis et al. 2000) and Authorware (McGraw et al. 2000b), among others.

For descriptions of how assignment can be accomplished by means of server-side CGI scripts, see Morrow & McKee (1998) and Schmidt (2000). A CGI script in Perl to handle random assignment to conditions, written by William Schmidt, is available from http://psych.fullerton.edu/mbirnbaum/programs/PERL_script2.htm.

Another task that can be accomplished in many ways on the Web is animation. For a study of different ways of creating animations on the Web, see the article by Schmidt (2001), which reviews the pros and cons of animated graphics interchange formats (gifs), movies, Flash animations, Java, and JavaScript.

Although certain tasks can be done well on either the server or the participant's computer, other tasks can or should only be done on the server. For example, the routine to save the data must be run on the server. A simple script for sorting and saving data by William Schmidt is available from the above-listed ATISP site. Other cases where server-side programming should be used (rather than client-side programming) include password control of access to files and scoring of on-line tests when there may be an incentive to cheat. Any task where security is an issue should only be run on the server. Schmidt (2000) describes many other examples of server-side programming.

Run Your Own Server

Schmidt et al. (1997) describe the advantages of operating your own server (see also Reips 1997). At the ATISP site (<http://ati.fullerton.edu>), Schmidt explains how to download and install the free Apache Web server on a PC, which includes the free distribution of Perl. All that is required is a desktop computer attached to the Internet and the free software. The system works best with a fixed Internet protocol (IP) address, but Schmidt notes there are even ways to work around that limitation. With a Mac running operating system (OS) X, Apache and Perl come preinstalled and can be easily turned on, as described by Reips, in the following URL: http://ati.fullerton.edu/ureips/all_you_need/index.html.

Many campus computing centers restrict what a professor or student can do on the campus network (there are legitimate security concerns of allowing students to add CGIs to a university server, for example). In response to concerns for security, some campus centers may arbitrarily put excessive restrictions on what a professor can do. To overcome such restrictions and to provide maximal control, many Web experimenters now run their own Web servers, and some have even moved their research off campus to be hosted by commercial Internet providers, where they enjoy greater freedom.

Stimulus Delivery

As computer hardware and software improve, it becomes easier and easier to deliver visual and auditory stimuli via the WWW. But there are limitations on what can be delivered and the precision of the stimulus delivered (Krantz 2001). A potential problem is that different users have different computers, systems,

browsers, monitors, and speakers, so the actual stimulus received may differ quite a bit from what the experimenter intended. The stimuli received may vary from person to person. Fortunately, perception obeys constancy properties that help make many such variations irrelevant to the results (Krantz 2001); people quickly adapt to fixed background conditions (such as brightness setting on a monitor), so that such variables, which change from person to person, have little effect. Nevertheless, such lack of control must be taken into consideration.

Even something as simple as how a browser and system displays different colors cannot be taken for granted. There were supposed to be 216 so-called Web-safe colors that would be displayed the same by all browsers (Birnbaum 2001a), but even these are not produced exactly the same by Internet Explorer and Netscape Navigator. For more information on color see the following sites:

<http://seurat.art.udel.edu/Site/Cookbook.html>

<http://hotwired.lycos.com/webmonkey/00/37/index2a.html?tw=design>

<http://www.visibone.com/>

<http://www.websitetips.com/color/>

Questions have also been raised involving the accurate control and measurement of temporal intervals on the participant's computer. Despite problems with stimulus control and response measurement, a number of classic perceptual and cognitive experiments have been successfully implemented on the Web. These Web studies appear to work quite well, despite imperfections introduced by the Web medium (Eichstaedt 2001, 2002; Gordon & Rosenblum 2001; Hecht et al. 1999; Horswill & Coster 2001; Krantz 2001; Lange 1999; Laugwitz 2001; Ruppertsberg et al. 2001; Schmidt 2001; Stevenson et al. 1999; Van Veen et al. 1998).

When the participant has to wait for a large visual or auditory file to download before it can be presented, he or she might get bored and quit the experiment. In recent years, attention has been devoted to speeding up transmission on the WWW by compressing pictures, graphics, and sounds, and finding ways to display a stimulus before the file has completely downloaded. One familiar method for sending graphics is to send a "progressive" file that presents a grainy but complete version of the final picture quickly and then improves the image as the details are received and interpreted (Krantz 2001).

The mp3 standard for auditory files has made it possible to send relatively small files of music or speech that sound quite good. Additional information with links to on-line resources are in Birnbaum (2001a, Chapter 14) and Krantz (2001).

By means of streaming media, it is possible to send video movies with sound in (almost) real time, once a buffer is filled. The idea of streaming media is that the first part of a file can begin to play while the rest of the file is still being downloaded. For information on QuickTime and free streaming servers, see <http://www.apple.com/quicktime/>.

If one has a good connection to the Web and the proper plug-ins, one can even watch "live" television via the Web. For example, with the Windows Media Player,

one can view the educational, Annenberg/CPB channel from the following URL: <http://www.learner.org/channel/channel.html>.

With the Flash Player and RealPlayer plug-ins, one can view the German television channel, *Deutsche Welle*, and receive “live” broadcasts from the following URL: <http://www.dw-world.de>.

At the time of this writing, these techniques are of good enough quality for psychological research only with computers that have fast connections to the Internet and which have the proper plug-ins installed (e.g., QuickTime, RealPlayer, Windows Media Player, etc.). However, it is likely that by the time this chapter is published, most people with Internet access will have the appropriate connection and software needed to display experiments with streaming video, making such research practical to conduct via the Web.

METHODOLOGY

Web methods help solve some old problems of lab research with undergraduates, but they create new problems that require special solutions. A number of works discuss advantages and possible drawbacks of Web-based research (Bandilla 2002; Birnbaum 1999b, 2001a, 2003; Dillman 2000; Dillman & Bowker 2001; Hewson et al. 1996; Petit 1999; Piper 1998; Reips 1997, 2000, 2001a; 2002b; Schiano 1997; Schmidt 1997a; Smith & Leigh 1997).

Perhaps the three most important advantages of Web research over lab research with the undergraduate “subject pool” are as follows: On the Web one can achieve large samples, making statistical tests very powerful and model fitting very clean. With clean data the “signal” of systematic deviations can be easily distinguished from “noise.” Second, Web studies permit generalization from college students to a wider variety of participants. Third, one can recruit specialized types of participants via the WWW that would be quite rare to find among students. At the same time, studies conducted via the WWW have several potential problems and disadvantages. These methodological pros and cons are reviewed in the next sections, including methods to address some of the potential problems introduced by Web research.

Multiple Submissions

In lab research it has rarely been considered a problem that a participant might serve twice in an experiment and thus reduce the effective degrees of freedom. However, in Web research, the possibility of multiple submissions has received considerable attention (Birnbaum 2000b, 2001a,b, 2004; Reips 2000, 2001a, 2002b; Schmidt 1997a, 2000). Table 1 summarizes methods intended to prevent multiple submissions or to detect them so that they can be removed from the data.

Each method is based on a particular theory of why a person might want to submit multiple sets of data. The first idea is that people may think that if it is good

TABLE 1 Avoiding or detecting multiple submissions

Method	Tactic
Instructions	Tell people to participate only once
Remove incentives	Rewards not available for those who participate more than once
Replace incentive	Provide alternative site for repeated play
Use identifiers	Common gateway interface (CGI) script allows only one submission; option: replace previous data or refusal to accept new
Use identifiers	Filter data to remove repeats
Use Internet protocol (IP), email address	Check for repeated IP addresses
Passwords	Allow participation by password only
Cookies	Check cookie for previous participation
CGI scripts	CGI checks for referring page and other features
Log file analysis	Can detect patterns of requests
Subsample follow up	Contact participants to verify ID
Check for identical data records	Filter identical or nearly identical records

to participate, it would be better to participate many times. So the first method is simply to ask people not to participate more than once.

The next methods are to remove or replace incentives to participate more than once. For example, if each participant is given a payment or a chance at a prize, there is an incentive to collect multiple payments or improve one's chances to win the prize. Instructions could then state that each person can be paid only once or receive no more than one chance at the prize. If the study is interesting (e.g., a video game or intellectual challenge), one could provide a separate site where those who want to continue to play with the materials can do so without adding to the regular data file.

Identifiers, such as student numbers, phone numbers, email addresses, mailing address, demographics, or names, can be used to identify each record of data. It is an easy matter to sort by such identifiers to detect multiple submissions from the same person.

In some research, it might be intended that each person should participate repeatedly to see how behavior changes with practice. Each person might be asked how many times he or she had already done the task. Date and time stamps on each record can be used to identify the temporal spacing in which submissions were received.

The Internet protocol (IP) address can be used to spot multiple submissions. The IP is a number such as 137.151.68.63, which represents a computer on the campus of California State University, Fullerton. Although IP addresses do not

uniquely identify a person, when two submissions arrive in a short period from the same IP, they likely represent multiple submissions from the same person. A conservative procedure is to remove records coming from the same or similar IP addresses even if they are separated in time (Reips 2000). A similar IP is an address with all but the number after the last dot the same; for example, 137.151.68.63 and 137.151.68.15 are two different computers on my home campus. Reips found, even in the days of mostly fixed IP addresses, that this conservative procedure of deleting any submissions from similar IPs would have almost no impact on the data set.

Of course, when data are collected in labs, one should expect the same IP address to come up again and again as different participants are tested on the same machine. Another complication is that IP addresses are now mostly “dynamic,” which means that two different users from an Internet service provider might receive the same IP on different days, because the Internet service provider (e.g., AOL) will assign IP addresses to users as they come and go. This also means that the same user might come to a study from two different IP addresses on two occasions. Therefore, it is useful to obtain additional identifiers, such as the last four digits of a Social Security number or driver’s license number, or an email address.

Instead of checking data for multiple submissions and removing them, it is possible to program the server to refuse to accept multiple submissions from the same ID (IP or other identifier). A CGI script could be used to check the identifiers and refuse to save subsequent submissions by the same ID, if this were deemed necessary.

One can use passwords to identify each participant and use a CGI script to monitor admission to the study, or one can place passwords in the data file to identify each record of data. Cookies (data stored on the participant’s computer) can also be used to keep track of how many times a person at that computer has participated. However, this procedure would not prevent a person from participating from different computers or from erasing the cookie, which some people do as a general protection of their privacy.

CGI scripts can be used to check for the referring page that sent the data. The referring page is the WWW document that sent data to the CGI. One can find not only what uniform resource locator (URL) (i.e., Web page) called the CGI (sent data), but one can also detect what URL referred the participant to that Web page.

It is good policy to check the referring page (that sent the data), as a general security measure. If that page is not located on the home server (the server that housed the study), it might mean that a hacker is sending data from some other site. One can also glean valuable information from an analysis of the server’s log file, which keeps track of requests and submissions of files and data. See also Breinholt & Krueger (1999) and Reips (2001a, 2002b). Reips has written software that can be used to analyze log files in order to determine where in a study the participant might have dropped out, for example. He also advises that Web experimenters keep log files for possible reanalysis by other investigators.

Reips (2000, 2001b) suggests following up a subsample of participants to check their identities, should this issue become a concern. For example, in a study in

which each participant has a chance at a prize and provides an email address to be contacted in case they win, one might check a subsample of those addresses to verify the identities of those people.

One can also search for identical data records. If a study is long enough, it is unlikely that two records would be the same unless the same person submitted data twice. Should one find a number of identical submissions in a short time interval, they are likely multiple copies.

Although the issue of multiple submissions is one of the first questions that Web experimenters are asked, Web researchers are of the consensus that this issue has not been a real problem (Birnbaum 2000b, 2001b; Musch & Reips 2000). In my experience, I find that multiple submissions are rare and easy to detect. They typically occur when a participant scrolls to the bottom of the questionnaire, clicks the submit button, reads the “thank you” page, and uses the “back” button on the browser to return to the study. The person may then add a comment or change a response, and click the submit button to navigate forward again. Such cases are easy to spot and filter from the data, because they come in sequentially with the same IP, the same email address, and for the most part, the same data in a short period.

In a careful analysis of 1000 data records, I found only one case in which the same person sent data on two different occasions. One woman participated exactly twice, months apart, and interestingly, agreed on 19 of her 20 decisions (Birnbaum 2001b). In the days of mostly fixed IPs, Reips (1997) analyzed 880 records and found only four submissions from the same or similar IP addresses. Reips’s procedure is conservative since it might eliminate two different students who both participated from the same campus.

A researcher should have a clear prior policy on whether to retain the new data or the old, in case data have changed between submissions. In my decision-making research, I always take the last submission, since I want the person’s last, best, most considered decision. However, in some areas of research (perhaps where debriefing might alter a person’s response), one might have a policy to take only the first set of data and delete any subsequent ones that might have been affected by debriefing. Reips’s LogAnalyzer program (<http://genpsylab-logcrunsh.unizh.ch/>) allows one to choose either first or last submissions for its analysis.

Dropouts in Between-Subjects Research

A serious threat to internal validity of a between-subjects experiment occurs when there are dropouts, people who begin a study and quit before completing it. Even when there are equal dropout rates in experimental and control groups, dropouts can cause the observed results to show the opposite of the true effects (Birnbaum 2003, Birnbaum & Mellers 1989). For example, a harmful treatment might look helpful.

Consider a workshop intended to help students prepare for the SAT. Suppose participants are randomly assigned to a control group or treatment group who

receive the program. Suppose this workshop actually lowers scores on the SAT, but also includes a practice test with advice not to take the real SAT exam if you did poorly on the practice exam. Such a program increases the correlation between performance on the test and likelihood of taking the test (as opposed to dropping out). Such an intervention could easily look helpful (because those who dropped out were those who would have achieved low scores), even though the treatment itself lowers the average SAT. Birnbaum & Mellers (1989) presented numerical examples illustrating how in a true experiment, with equal dropout in both experimental and treatment groups, the harmful treatment can produce a higher mean level of performance among those who do not drop out. Thus, dropouts can make a true experiment give false results.

Web-based research has been found to have larger dropout rates than lab studies. In the lab, other people are present, so a person would have to explain that he or she would like to quit the study and leave early. Web participants are free of any such possible social pressure or embarrassment. They simply click a button to quit a study and do something else (Birnbaum 2003). Knapp & Heidingsfelder (2001) and Frick et al. (2001) discuss the issue and describe procedures intended to reduce dropouts. See also O'Neil & Penrod (2001).

The "high-hurdle" and "warm-up" techniques of Reips (2000, 2002b) are intended to cause those who would drop out of a study to do so before the random assignment to conditions. In the high-hurdle technique, one asks for personal information early and includes a page that loads slowly, in order to encourage impatient or resistant people to drop out early, leaving cooperative participants behind. By asking for identifying information early, it may help those quit early who would be loath to provide the information later, and it may create more social pressure among those who do provide the information to stick it out to the end of the study. Frick et al. (2001) presented an empirical application of this approach. If the method works as theorized, then the method exchanges a threat to external validity (How do results with self-selected cooperative people generalize to others?) for the threat to internal validity (Do different people drop out of the treatment and control groups?).

Bosnjak (2001) and Bosnjak & Batinic (2002) consider the related issues of why people may not wish to "drop in" to participate in surveys or to answer all items. Welker (2001) notes that one of the reasons that surveys sent via regular mail have higher response rates than surveys sent by email is that not everyone reads his or her email. In a study with a fixed population of people with email accounts, a random sample of 900 was selected to receive invitations to a survey by email. It was found that the overall response rate was only 14%; however, for this sample of 900, 330 never read their email during the two months of the survey. When these nonreaders were taken into account, the response rate was figured to be 22%, which is higher, but still rather low.

Self-selection of participants means that results of any survey may not be accurate as a description of the population to which inference may be desired. By repeated reminders and combinations of electronic and postal mailings, it is

possible to increase response rates (Dillman 2000). Tourangeau (2004) documents a trend in survey research: The refusal rates and nonresponse rates to surveys of all types has doubled in the last decade; that review includes discussion of surveys by mail, email, phone, and Internet.

Recruiting Participants Via the Web

The Web can be used not only to test participants, but also to recruit them. Indeed, four types of studies should be distinguished, composed of the four combinations of recruitment method (Web or other) and testing method (Web or lab). By recruiting from the Web, it is possible to easily obtain large sample sizes; one can obtain samples that are heterogeneous with respect to age, education, income, social class, and nationality; and it is possible to reach people with special or rare characteristics. One might recruit people with rare characteristics via the Web and then test them in the lab. Or one might recruit people from the local “subject” pool and test them in the lab. Similarly, one might recruit and test in the Web or one might recruit from the “subject pool” and test via the WWW. Examples of these four types of studies are discussed in Birnbaum (2003).

Birnbaum (2001a, Chapter 21) and Reips (2000, 2001a,b) review a number of different techniques for recruiting participants via the WWW. One method is passive: Let people find the study via search engines or links. One can register an experiment with search engines that still accept “suggest a site” and wait for participants to find the study on the Web.

A study of sexual behavior (suggested to *Yahoo!*) was initiated in 1995 at Fullerton, which recruited entirely by this passive method. In the next four years, the study was completed by more than 10,000 participants who completed a questionnaire of more than 400 items (Bailey et al. 2000). During the course of that study, several other Web sites placed links to the study, which illustrates the point that even though one may choose a method of recruitment, that method does not completely control recruitment. It seems likely that it makes a difference if a study is linked in a site for incest victims as opposed to one for swingers. The people who place the links are being “helpful,” but they may influence how people come to find the study. Information concerning the link that sent a person to the site can be recorded, and data can be analyzed separately according to what link referred each person to the study, but the lack of complete control over recruiting should be clear from this example.

Another method of recruitment is to announce a study via email. However, it is bad manners to send unsolicited email messages to people who don’t want to receive them (Birnbaum 2003, Hewson et al. 1996), so it is recommended that one should *not* send any request that might be seen as “spam” (electronic “junk” mail) to all members of a mailing list. Such spam may generate “flames” (angry messages) to the entire list, which could injure one’s reputation and be counterproductive. Instead, one should seek cooperation of a relevant organization with an Internet presence. If the organization vouches that the researcher is a serious scientist and

that the research would be of interest or benefit to its members, the organization can provide good help and save the researcher from going down in “flames.”

Pam Drake, a former graduate assistant of mine, wanted to recruit elderly people interested in genealogy. She contacted a relevant on-line genealogy organization, which agreed to back her research by sending a request to participate to its members. Within a week, she had more than 4000 data records from her targeted group, and she had received many friendly messages of encouragement and support. Other ways that an organization could help would be by listing a link to your study in its Web site, posting a message to an electronic bulletin board, or including information about the study in its newsletter.

One can also recruit by requesting a link in sites that list on-line psychology studies, such as the sites maintained by Krantz (1998) and by Reips (1995), mentioned above. There are also sites that list free things on the Web. If your experiment yields an estimate of one’s IQ, for example (Wilhelm & McKnight 2002), it could be described as a free IQ test. Many “self-discovery” sites would be happy to list a link to ability or personality tests that provide feedback.

Banners, or advertising links placed in Web sites, might also be used. These are usually commercial announcements, so one runs the risk of looking like “free enterprise,” which usually means there’s a catch (i.e., not free). Some scientists have explored this method and estimated that less than 1% of those who view the banner will click the link. Because this method has expense associated with it, it may not be cost-effective; nevertheless, there may be some situations in which one might want to test the sort of people who would click such links placed at certain special sites (Buchanan 2000).

For example, in personality testing, a person might want to study visitors to body-piercing sites, to see if their personalities differ systematically from those who visit “traditional family values” sites. The multiple-site entry technique (Reips 2000, 2002b) uses different methods to recruit different types of people to sites that contain the same basic content, to see if the data collected at the different sites are systematically correlated with the recruitment method. As a check on the recruitment method, one can ask the participants to indicate the group with which they more closely identify.

Many personality taxonomies have been developed based on paper-and-pencil analyses of data from college students. Because college students are homogeneous on several variables (age, education, etc.) it is reasonable that correlates of such variables would not emerge in factor-analytic studies of personality. Because Web studies reach more heterogeneous participants, recruitment via the Web has advantages over traditional studies of undergraduate personality (Buchanan 2000, 2001; Buchanan & Smith 1999a,b; Hertel et al. 2002; Mueller et al. 2000; Pasveer & El-lard 1998). McKenna & Bargh (2000), Schillewaert et al. (1998), and Reips (2000, 2002b) discuss the multiple-site entry method, which has good potential for establishing criterion groups that can be used to develop and validate personality tests.

The Internet is a convenient medium for cross-cultural or international research as well. Here the criterion groups are nationalities or cultures. Once an experiment

is running on the Web, participants from anywhere in the world can access it. An investigator can form collaborations with scientists in other nations who can recruit to the same site from different strata of their populations. It is then easy to compare data from participants recruited similarly in different nations. Studies by Pagani & Lombardi (2000) and Pohl et al. (2002) illustrate this technique. It would be quite useful in an area where one wishes to compare data across cultures to investigate evolutionary theories of human behavior (such as jealousy, as in Voracek et al. 2001).

One method of regulating the selection of participants tested via the Web is to recruit an on-line panel (Baron & Siepmann 2000, Göritz et al. 2002). Such a panel consists of a group of people who have been preselected on some basis (e.g., for representative distributions of gender, age, education level, etc.) to be members of a panel that will be used repeatedly, much like the families used for Nielson television ratings.

Sampling Bias and Stratified Analysis

It would be a mistake to treat data recruited from the Web as if they represented a sample of some stable population of “Web users.” The list of people with access to the Internet is expanding every day, and no method yet devised has been shown to reach those users in a random fashion. By using different techniques, many different substrata of Web users might be reached, but there is no guarantee that any particular method of recruitment would yield a sample representative of some particular population.

One approach to this problem in experimental research is to analyze the research question separately within each substratum of a heterogeneous sample. With large numbers of participants in Web studies, it becomes possible to separate the data by age, education, gender, and other demographic variables and to analyze the data separately in each slice. If the same conclusions are reached in each subsample, one begins to believe that the results have generality to other groups besides college sophomores (Birnbaum 1999b).

When the purpose of the research is to estimate a population parameter from a Web sample, however, then neither college sophomores nor self-selected Web participants should be considered as random samples of any particular population. Those who take lower division psychology courses are not a random sample of college students, let alone of some wider population. Those who enroll in psychology are more likely to be female and less likely to be engineers than the campus average. Those who are recruited via the Web are on average older than college students, have greater mean education, and greater variance in age and education than do college students.

As an example of research intended to estimate a population parameter, one might want to forecast the results of an election by a survey of potential voters. For this purpose, neither college undergraduates, nor self-selected samples (such as are easy to recruit via the Web) are appropriate. In this case, participants are said

to be self-selected because they decide whether to participate and whether to drop in or drop out. For more discussion of problems of self-selection, see Birnbaum (1999b, 2000b, 2001a, 2003), Bandilla (2002), Brenner (2002), Bosnjak & Batinic (2002), Corkrey & Parkinson (2002), Dillman (2000), Dillman & Bowker (2001), Musch et al. (2001), Reips (2000, 2002c), and Stanton (1998).

For an example of a survey of a sample randomly selected for invitation from a particular membership list of Internet users, see Lukawetz (2002). Lukawetz found that those who used the Internet less frequently were less likely to respond to the survey and more likely to respond late. Thus, even when a population can be defined, there is evidence that self-selected volunteers are not a random sample of that group.

The problem with self-selection is that the results may not generalize from the self-selected sample to some larger population. For example, a survey of political opinions recruited from a Web site for gun owners might yield a different conclusion with respect to people's opinions of political candidates than a survey recruited from a Web site devoted to environmental issues. Even if the same survey materials were used, neither group would be likely to forecast the outcome of the next election.

The most famous case of misuse of data from a self-selected sample is the *Literary Digest's* erroneous prediction that Alf Landon would defeat Franklin D. Roosevelt for the presidency in 1936. Even though the sample size was very large, the readers of this magazine who self-selected to mail in the survey were much more likely to favor Republican candidates than was the average voter (Huff 1954).

Response Bias

The input devices used to accept responses can influence how a participant would respond. In fact, some input devices should be avoided completely. For example, the check box should not be used because it allows only two responses: yes or no (Birnbaum 2001a). If a box is unchecked, does that mean that the participant responded "no" or does it mean the person did not respond to the item? One should allow at least three responses (yes, no, and no response) and sometimes even four (yes, no, no answer, and refuse to answer) for a "yes-no" question.

The pull-down selection list is a very complicated answering device. The user must click on the device to see a list of alternative answers. The user must then scroll through the choices to select an answer. As in a "yes-no" item, there should be no preselected answer, or people who skip the item will be recorded as having made that answer. The preselected choice should be coded as a nonresponse and given a label such as "choose from this list." The distance and direction of each possible response from the preselected choice may influence the results.

For example, suppose we want to determine in what category a person's monthly income falls. Suppose the default choice is listed as \$0, and many levels are listed between \$0 and \$200 per month. It is likely that this procedure would yield lower estimates of monthly income than if the default were more than \$1 million

per month, if there were many levels between \$1 million and \$10,000 per month, and where one must scroll down past many options to see lower ranges of income.

The size of a text box for a numerical answer may also influence the response. For example, consider the question, "How many traffic accidents have you had?" A size = 1 input text box (which suggests to the participant that a 1-digit number is expected) would likely produce smaller numbers than a size = 3 box (which signals that a 3-digit number might be expected). Birnbaum (2001a, Chapter 5) described an experiment on the value of the St. Petersburg gamble in which significantly different averages were obtained by different lists of values in a pull-down list and these results were in turn different from the mean obtained with a large text box. See Dillman & Bowker (2001) and Reips (2002a,d) for further discussion of how changes in arrangements or display as well as other surface features might alter the results one obtains.

Experimenter Bias

When an experimenter or her assistant knows the research hypothesis, receives responses from the participant, and then enters the data for analysis, small biases can enter the data file in many ways. The experimenter might give additional undocumented instructions, might provide subtle reinforcing cues, might repeat a question as if expecting a different answer, and might miscode or mistype the data in the expected direction. An error in the expected direction might not be noticed, but an error in the unpredicted direction might be caught. A possible advantage of Web studies is that these many sources of experimenter biases can be standardized and documented so that other scientists can repeat the experiment exactly. By eliminating the lab assistant, some of the sources of bias are eliminated.

The potential drawback is that if the participant has a question about the instructions, there is no lab assistant to explain or clarify the task. This means that every aspect of an experiment, including the wording of instructions, must be carefully thought out before launching an experiment on the Internet. Furthermore, one must take into consideration the variety of people around the world who may have very different interpretations of the instructions.

PILOT TESTING IN THE LAB

As noted earlier, it is possible to have several thousand people complete a Web study in a few days, so before launching a study, it must be thoroughly checked. It would be a kind of vandalism of scientific resources to waste people's time completing a flawed research instrument. Such checking should involve at least three main parts.

First, decide if the data analysis will answer the question addressed by the research. Those who design a study without planning the analysis usually fail to devise an experiment that answers the original question (Birnbaum 2001a, 2003).

Second, check the programming of the study, including coding of the data. I have found surveys on-line in which the same code value was used for missing data as for one of the legitimate answers. In one case, if a person were from India or if he or she did not respond, the coded value was the same, 99. Fortunately, this problem was caught before much data had been collected.

I find many problems that my students fail to notice because they do not check their work carefully. I have seen students make mistakes in code such that when the participant answers the second question, the answer to the first is replaced. Click a radio button to Question 3, and the one for Question 2 goes out. Part of the checking process should include sending test data to make absolutely sure that each button and each response field is functioning properly and that the experimenter knows where the data go in the file. One should also view the experiment using several types of computers, monitors, browsers, and versions of those browsers to preview the different experiences that participants will have with these variations.

This second part of pilot testing should include testing some naive participants in the lab, and interviewing them to make sure that they understood the instructions. Data from this lab test should be analyzed to make sure that the method of analysis works and is error free. It is often when analyzing data that students notice that they have made a coding error in their programming. Based on the pilot study, any improvements should be made and then rechecked before the study is put on-line.

As a third part of the pretest, participants in the lab test should also be closely observed, as some interesting things may be discovered. One of my students placed a response device above the material on which the response was to be made. Watching people, I saw that some participants answered before they scrolled down to see the material to which they were supposed to be responding.

In another student project, participants were instructed to take out some coins and toss them, in a variation of the random response technique (Musch et al. 2001). For example, consider the question of whether a person used an illegal drug in the past 24 hours. The random response technique, if it works, allows the researcher to estimate the proportion of people who used the drug, without knowing for any individual whether he or she used the drug. Participants were instructed to toss two coins, and if they got two heads, they were to respond "yes" and if they got two tails, they were to respond "no;" however, if they got a heads and tails or tails and heads, they were to respond with the truth. Thus, a "yes" or "no" does not indicate for any person if it came from the coins or from the truth. But the researcher can subtract off 25% heads and 25% tails and multiply by two to get the correct answer. For example, if 30% of the overall sample said "yes," it means that 10% of the truth-tellers said "yes." This is a very clever method, if it works as planned.

Out of 15 participants pretested in the lab, however, only one student actually took out any coins, and that one asked first if she should actually do what the instructions said to do. Had this study been placed on the Web, the experimenter never would have known what the participants were doing (or in this case, not doing). The solution in this case was to add additional instructions, and to add questions asking the participants if they did, indeed, use coins or not, and if not,

why not. Even after the instructions emphasized the importance of following the instructions, about half of undergraduates tested conceded that they did not, in fact, use the coins, giving reasons such as “lazy” or “I had nothing to hide” as explanations for not following the instructions.

Lists of other tests to perform and standards to check before launching a study on the Web are given by Birnbaum (2001a, Chapters 12 and 21) and Gräf (2002). Reips (2002b,d) also provides lists of standards and “dos and don’ts” that one should consider before launching one’s study. These lists will likely continue to expand as more is learned about on-line research.

WEB VERSUS LAB

Once an experiment is running on the Web, it is a simple matter to replicate the experiment in the lab. A number of studies compared data obtained in the lab against data obtained via the Web, to see if the two ways of conducting research reach the same conclusions (Birnbaum 1999b, 2001a; Horswill & Coster 2001; Krantz & Dalal 2000; Krantz et al. 1997; Musch & Klauer 2002; Reips 2002b,c; Van Veen et al. 1998; Voracek et al. 2001).

Krantz & Dalal (2000) reviewed nine studies comparing Web and lab versions and concluded that results from the two methods yielded surprising agreement. Subsequent studies have generally conformed to this result. However, that conclusion should not be taken to mean that there would be no significant difference between subjects recruited from the Web and college students, nor does it mean that no significant effect of different types of monitors and browsers should be expected.

Those working with cognitive psychology experiments take it for granted that if they program their experiments properly, the same basic results should be observed (Stevenson et al. 1999, Williams et al. 1999). Indeed, such Web experiments consistently replicate classic results, even when the independent and dependent variables involve manipulation or measurement of brief time intervals, which are thought to be less precisely controlled or measured via the Web (Eichstaedt 2002, Francis et al. 2000, McGraw et al. 2000b).

However, the thesis that the same conclusions might be reached in lab and Web research does not mean that Web and lab studies will not yield significantly different results. Birnbaum (1999b), for example, set out to recruit a special population of participants—people with advanced education in decision-making. He found that people holding doctorates who have read a scientific work on decision theory are significantly less likely to violate stochastic dominance than undergraduates who have not studied decision-making. With undergraduates, 68% said they would prefer to play a gamble with a 0.85 probability to win \$96, a 0.05 probability to win \$90, and a 0.10 probability to win \$12 rather than a gamble with a 0.90 probability to win \$96, 0.05 to win \$14, and 0.05 to win \$12, even though the second gamble dominates the first. The 95 people with doctorates and who had read on

decision-making had 50% violations on this choice, which is still very high but significantly lower than the rate observed with undergraduates.

Significant correlations were also found with gender: Female undergraduates violated stochastic dominance more often than male undergraduates. Indeed, the undergraduate sample in Birnbaum's (1999b) "lab" study was 73% female, whereas the participants recruited from the Web had only 56% females. However, the conclusions with respect to models of decision-making were no different in these groups, even though the rates of violation were significantly different in demographic categories.

Birnbaum (1999b, 2000a, 2001b; Birnbaum & Martin 2003) conducted a series of Web experiments with thousands of participants, indicating that cumulative prospect theory (Tversky & Kahneman 1992) is not descriptive of risky decision-making. Five new paradoxes have been devised that yield contradictions to this theory. Cumulative prospect theory implies that people's judgments should satisfy stochastic dominance, coalescing, lower cumulative independence, upper cumulative independence, and upper tail independence. However, data from thousands of participants tested in the lab and via the Web show that these properties are systematically and significantly violated. Web experiments have now tested a variety of techniques for displaying the risky gambles, and the results continue to hold up with thousands of participants. In addition, violations of restricted branch independence and of distribution independence are opposite the pattern predicted by cumulative prospect theory's weighting function.

This series of experiments would have taken decades to accomplish had the studies been run by the techniques used in lab research. This advantage in efficiency is the probably the greatest advantage of Web research over previous methods of research and will be the primary motivation of those who switch to this new method. In these tests of decision-making properties, tests with undergraduates in classrooms and in labs also yield the same conclusions as those from the Web.

Web studies, by having larger samples, usually have greater power than lab studies, despite greater noise due to technical variations (hardware, systems, and browsers) and demographic diversity in Web samples. Data quality can be defined by variable error, constant error, reliability, or validity. Comparisons of power and of certain measures of quality have found cases where Web data are higher in quality by one or another of these definitions than are comparable lab data, though not always (Birnbaum 1999b, 2000a; Reips 2002c; Tuten et al. 2002).

Many Web researchers are convinced that data obtained via the Web can be "better" than those obtained from students (Baron & Siepmann 2000, Musch & Reips 2000, Reips 2000), despite the obvious advantage that the lab offers for control. Based on what has been done, however, I don't think it appropriate to argue yet that Web data are better than lab data, because the relevant comparisons have not been completed and the Web-versus-lab variable is not really a single independent variable. For example, those recruited via the WWW have been older, better educated, and perhaps more motivated than the usual undergraduate sample tested in the lab.

Part of the difference may occur because the average participant in early Web studies had more education than did the typical undergraduate in the lab. Perhaps more highly educated participants are also higher in intelligence, or better at reading and following directions. Birnbaum (1999b), who intended to recruit highly educated participants, found that only 2% of females who have earned a doctorate violated transparent dominance (a rational property of decision making) while tested via the WWW. In contrast, 14% of female undergraduates tested in the lab violated the same property. Is this difference due to being tested in the lab versus on-line, or is it due to the difference in education or its unmeasured correlates? For comparison, 12% of females tested via the Web with some college (but no degree) violated the same property. So, if we consider conformity to transparent dominance as an index of data quality, the Web data would indeed be judged higher in quality, but most of the apparent difference might be explainable by the confounded effects of education or other correlates of the method of recruitment.

Part of the belief that Web data are better than lab data may also be because many participants in early Web studies participated as volunteers out of interest, whereas students often participate as one way to complete an assignment (Reips 2000). Many Web experimenters have been pleasantly surprised (as I was) to receive many messages from Web participants expressing interest and encouragement in the research. It makes one think that Web participants are more motivated than college students and perhaps take the tasks more seriously. Interestingly, the confidence that Web experimenters have in their data runs contrary to the impression held by some that on the Web people are more likely to misrepresent themselves.

Bargh & McKenna (2004) reviewed research on the social psychology of the Internet. Social psychologists theorize that the Internet creates new social situations. There is a great deal of theorizing backed by a little data that people may behave differently, perhaps taking on a new identity via the Web, where they might feel anonymous (Joinson 1999, Sassenberg & Kreutz 2002), compared to how they would respond in person. It has also been argued to the contrary, that people might be more honest with a computer than with a human interviewer on sensitive matters, such as details of one's sex life (Bailey et al. 2000). This new and growing literature has already been reviewed in several books (Döring 2002a, Gackebach 1998, Joinson 2002, Suler 1996, Wallace 2001). If people do behave differently, it might mean that research conducted via the WWW would reach systematically different conclusions from those obtained in the lab.

Kraut et al. (1998) suggested that spending a lot of time on the Internet might interfere with one's social life with real people, and might lead to depression. Brenner (2002) provides a skeptical review of the topic of "Internet addiction," a term referring to pathologically excessive use of the Internet to the detriment of one's real life. Döring (2002b) reviews studies of cyber romance, noting that these relationships are quite common and real. Janetzko (2002) reviews the topic of dialog and interview "bots," computer programs that emulate the behavior of humans. Perhaps one day these bots will be perfected and we will have an Internet sequel to the tale of Pygmalion. Perhaps people will fall in love with a computer

program. For more on the social psychology of the Internet, see the review by Bargh & McKenna (2004).

SUMMARY AND CONCLUSIONS

Even if people do not love the computers themselves, it is clear that many psychologists have indeed become enamored of the possibilities provided by these new methods for conducting psychological research via the WWW.

In the short time since Web experimentation has become possible, a great deal has been learned about the techniques, methodology, and results of Web studies. Programs and instructional materials are now available to make it relatively easy for a new investigator in this field to be able to implement simple studies to run via the WWW. Investigators who have begun research using these techniques have for the most part judged the method to be successful and plan to continue research of this type. Although some methodological problems have emerged in Web-based research because of the lack of control in Web studies, many investigators consider the advantages in experimental power, low cost, and convenience of testing via the Web to outweigh its disadvantages. Experience with studies done via the Web and in the lab indicates that if Web studies are properly designed, one can replicate lab results in many fields of psychology. Significant differences between results of lab studies with undergraduates and Web studies of people recruited via the Web have been reported, but these may be due to the differences in people tested rather than the method.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation Grants SBR-94,10572, SES 99-86,436, and BCS-01,29453. I thank Ulf-Dietrich Reips for many helpful comments and suggestions on a previous draft of the manuscript.

The Annual Review of Psychology is online at <http://psych.annualreviews.org>

LITERATURE CITED

- Bailey RD, Foote WE, Throckmorton B. 2000. Human sexual behavior: a comparison of college and Internet surveys. See Birnbaum 2000b, pp. 141–68
- Bandilla W. 2002. Web surveys—an appropriate mode of data collection for the social sciences? See Batinic et al. 2002, pp. 1–6
- Bargh JA, McKenna KYA. 2004. The Internet and social life. *Annu. Rev. Psychol.* 55:573–90
- Baron J, Siepman M. 2000. Techniques for creating and using web questionnaires in research and teaching. See Birnbaum 2000b, pp. 235–65
- Batinic B, ed. 2000. *Internet für Psychologen*. Göttingen, Ger.: Hogrefe
- Batinic B, Reips U-D, Bosnjak M, eds. 2002. *Online Social Sciences*. Göttingen, Ger.: Hogrefe & Huber
- Batinic B, Werner A, Gräf L, Bandilla W,

- eds. 1999. *Online Research: Methoden, Anwendungen, Ergebnisse*. Göttingen, Ger.: Hogrefe
- Birnbaum M. 1998. *SurveyWiz and FactorWiz*. <http://psych.fullerton.edu/mbirnbaum/programs/>
- Birnbaum MH. 1999a. How to show that $9 > 221$: collect judgments in a between-subjects design. *Psychol. Methods* 4:243–49
- Birnbaum MH. 1999b. Testing critical properties of decision making on the Internet. *Psychol. Sci.* 10:399–407
- Birnbaum MH. 1999c. *WWW Research and Decision Making*. Keynote address to German Online Res. Soc. Meet., Nürnberg, Ger.
- Birnbaum MH. 2000a. Decision making in the lab and on the Web. See Birnbaum 2000b, pp. 3–34
- Birnbaum MH, ed. 2000b. *Psychological Experiments on the Internet*. San Diego: Academic
- Birnbaum MH. 2000c. SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet. *Behav. Res. Methods Instrum. Comput.* 32:339–46
- Birnbaum MH. 2001a. *Introduction to Behavioral Research on the Internet*. Upper Saddle River, NJ: Prentice Hall
- Birnbaum MH. 2001b. A Web-based program of research on decision making. See Reips & Bosnjak 2001, pp. 23–55
- Birnbaum MH. 2002. Wahrscheinlichkeitslernen. In *Das Experimentale-psychologische Praktikum im Labor und WWW*, ed. D Janetzko, M Hildebrand, HA Meyer, pp. 141–51. Göttingen, Ger.: Hogrefe
- Birnbaum MH. 2004. Methodological and ethical issues in conducting social psychology research via the Internet. In *Handbook of Methods in Social Psychology*, ed. C Sansone, CC Morf, AT Panter, pp. 359–82. Thousand Oaks, CA: Sage
- Birnbaum MH, Martin T. 2003. Generalization across people, procedures, and predictions: violations of stochastic dominance and coalescing. In *Emerging Perspectives on Decision Research*, ed. SL Schneider, J Shanteau, pp. 84–107. New York: Cambridge Univ. Press
- Birnbaum MH, Mellers BA. 1989. Mediated models for the analysis of confounded variables and self-selected samples. *J. Educ. Stat.* 14:146–58
- Birnbaum MH, Wakcher SV. 2002. Web-based experiments controlled by JavaScript: an example from probability learning. *Behav. Res. Methods Instrum. Comput.* 34:189–99
- Bosnjak M. 2001. Participation in non-restricted Web surveys: A typology and explanatory model for item non-response. See Reips & Bosnjak 2001, pp. 193–207
- Bosnjak M, Batinic B. 2002. Understanding the willingness to participate in online surveys—the case of e-mail questionnaires. See Batinic et al. 2002, pp. 81–92
- Breinholt G, Krueger H. 1999. A tutorial on creating logfiles for event-driven applications. *Behav. Res. Methods Instrum. Comput.* 31:410–15
- Brenner V. 2002. Generalizability issues in Internet-based survey research: implications for the Internet addiction controversy. See Batinic et al. 2002, pp. 93–113
- Buchanan T. 2000. Potential of the Internet for personality research. See Birnbaum 2000b, pp. 121–40
- Buchanan T. 2001. Online personality assessment. See Reips & Bosnjak 2001, pp. 57–74
- Buchanan T, Smith JL. 1999a. Research on the Internet: validation of a World Wide Web mediated personality scale. *Behav. Res. Methods Instrum. Comput.* 31:565–71
- Buchanan T, Smith JL. 1999b. Using the Internet for psychological research: personality testing on the World-Wide Web. *Br. J. Psychol.* 90:125–44
- Corkrey R, Parkinson L. 2002. A comparison of four computer-based telephone interviewing methods: getting answers to sensitive questions. *Behav. Res. Methods Instrum. Comput.* 34:354–63
- Dillman DA. 2000. *Mail and Internet Surveys: The Tailored Design Method*. New York: Wiley

- Dillman DA, Bowker DK. 2001. The Web questionnaire challenge to survey methodologists. See Reips & Bosnjak 2001, pp. 159–78
- Döring N. 2002a. *Sozialpsychologie des Internet*. Göttingen, Ger.: Hogrefe
- Döring N. 2002b. Studying online love and cyberromance. See Batinic et al. 2002, pp. 333–56
- Eichstaedt J. 2001. An inaccurate timing filter for reaction time measurement by JAVA applets implementing Internet-based experiments. *Behav. Res. Methods Instrum. Comput.* 33:179–86
- Eichstaedt J. 2002. Measuring differences in preactivation on the Internet: the content category superiority effect. *Exp. Psychol.* 49:283–91
- Francis G, Neath I, Surprenant AM. 2000. The cognitive psychology online laboratory. See Birnbaum 2000b, pp. 267–83
- Frick A, Bächtiger MT, Reips U-D. 2001. Financial incentives, personal information, and drop-out in online studies. See Reips & Bosnjak 2001, pp. 209–19
- Gackebach J, ed. 1998. *Psychology and the Internet: Intrapersonal, Interpersonal, and Transpersonal Implications*. San Diego: Academic
- Gordon MS, Rosenblum LD. 2001. Audiovisual speech Web-lab: an Internet teaching and research laboratory. *Behav. Res. Methods Instrum. Comput.* 33:267–69
- Görizt A, Reinhold N, Batinic B. 2002. Online panels. See Batinic et al. 2002, pp. 27–47
- Gräf L. 2002. Assessing Internet questionnaires: the online pretest lab. See Batinic et al. 2002, pp. 49–68
- Hecht H, Oesker M, Kaiser A, Civelek H, Stecker T. 1999. A perception experiment with time-critical graphics animation on the World Wide Web. *Behav. Res. Methods Instrum. Comput.* 31:439–45
- Hertel G, Naumann S, Konradt U, Batinic B. 2002. Personality assessment via Internet: comparing Online and paper-and-pencil questionnaires. See Batinic et al. 2002, pp. 115–33
- Hewson CM, Laurent D, Vogel CM. 1996. Proper methodologies for psychological and sociological studies conducted via the Internet. *Behav. Res. Methods Instrum. Comput.* 28:186–91
- Horswill MS, Coster ME. 2001. User-controlled photographic animations, photograph-based questions, and questionnaires: three Internet-based instruments for measuring drivers' risk-taking behavior. *Behav. Res. Methods Instrum. Comput.* 33:46–58
- Huff D. 1954. *How to Lie with Statistics*. New York: Norton
- Janetzko D. 1999. *Statistische Anwendungen im Internet*. Munich: Addison-Wesley
- Janetzko D. 2002. Artificial dialogues—dialogue and interview bots for the World Wide Web. See Batinic et al. 2002, pp. 357–64
- Janetzko D, Meyer HA, Hildebrand M, eds. 2002. *Das Experimentale-psychologische Praktikum im Labor und WWW [A practical course on psychological experimenting in the laboratory and WWW]*. Göttingen, Ger.: Hogrefe
- Joinson AN. 1999. Social desirability, anonymity, and Internet-based questionnaires. *Behav. Res. Methods Instrum. Comput.* 31:433–38
- Joinson AN. 2002. *Understanding the Psychology of Internet Behaviour: Virtual Worlds, Real Lives*. Basingstoke, Hampshire, UK: Palgrave Macmillian
- Knapp F, Heidingsfelder M. 2001. Drop-out analysis: effects of the survey design. See Reips & Bosnjak 2001, pp. 221–30
- Krantz JH. 1998. *Psychological Research on the Net*. <http://psych.hanover.edu/research/exponnet.html>
- Krantz JH. 2001. Stimulus delivery on the Web: What can be presented when calibration isn't possible? See Reips & Bosnjak 2001, pp. 113–30
- Krantz JH, Ballard J, Scher J. 1997. Comparing the results of laboratory and world-wide web samples on the determinants of female attractiveness. *Behav.*

- Res. Methods Instrum. Comput.* 29:264–69
- Krantz JH, Dalal R. 2000. Validity of Web-based psychological research. See Birnbaum 2000b, pp. 35–60
- Kraut R, Patterson M, Lundmark V, Kiesler S, Mukopadhyay T, Scherlis W. 1998. Internet paradox: a social technology that reduces social involvement and psychological well-being? *Am. Psychol.* 53:1017–31
- Lange M. 1999. Museum of perception and cognition website: using JavaScript to increase interactivity in web-based presentations. *Behav. Res. Methods Instrum. Comput.* 31:34–45
- Laugwitz B. 2001. A Web-experiment on colour harmony principles applied to computer user interface design. See Reips & Bosnjak 2001, pp. 131–45
- Lukawetz G. 2002. Empirically quantifying unit-nonresponse-errors in online surveys and suggestions for computational correction methods. See Batinic et al. 2002, pp. 403–15
- McClelland G. 2000. *Seeing Statistics*. Florence, KY: Duxbury Press
- McGraw KO, Tew MD, Williams JE. 2000a. The integrity of Web-based experiments: Can you trust the data? *Psychol. Sci.* 11:502–6
- McGraw KO, Tew MD, Williams JE. 2000b. PsychExps: an on-line psychology laboratory. See Birnbaum 2000b, pp. 219–33
- McKenna KYA, Bargh JA. 2000. Plan 9 from cyberspace: the implications of the Internet for personality and social psychology. *Personal. Soc. Psychol. Rev.* 4:57–75
- Morrow RH, McKee AJ. 1998. CGI scripts: a strategy for between-subjects experimental group assignment on the World-Wide Web. *Behav. Res. Methods Instrum. Comput.* 30:306–8
- Mueller JH, Jacobsen DM, Schwarzer R. 2000. What are computing experiences good for: a case study in on-line research. See Birnbaum 2000b, pp. 195–216
- Musch J. 2000. Die Geschichte des Netzes: Ein historischer Abriss. See Batinic 2000, pp. 15–37
- Musch J, Broeder A, Klauer KC. 2001. Improving survey research on the World-Wide Web using the randomized response technique. See Reips & Bosnjak 2001, pp. 179–92
- Musch J, Klauer KC. 2002. Psychological experimenting on the World Wide Web: investigating context effects in syllogistic reasoning. See Batinic et al. 2002, pp. 181–212
- Musch J, Reips U-D. 2000. A brief history of Web experimenting. See Birnbaum 2000b, pp. 61–87
- O'Neil KM, Penrod SD. 2001. Methodological variables in Web-based research that may affect results: sample type, monetary incentives, and personal information. *Behav. Res. Methods Instrum. Comput.* 33:226–33
- Pagani D, Lombardi L. 2000. An intercultural examination of facial features communicating surprise. See Birnbaum 2000b, pp. 169–94
- Pasveer KA, Ellard JH. 1998. The making of a personality inventory: help from the WWW. *Behav. Res. Methods Instrum. Comput.* 30:309–13
- Piper AI. 1998. Conducting social science laboratory experiments on the World Wide Web. *Libr. Inf. Sci. Res.* 20:5–21
- Pohl RF, Bender M, Lachmann G. 2002. Hind-sight bias around the world. *Exp. Psychol.* 49:270–82
- Reips U-D. 1995. *Web Experimental Psychology Lab*. <http://www.psychologie.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html>
- Reips U-D. 1997. Das psychologische Experimentieren im Internet. In *Internet für Psychologen*, ed. B Batinic, pp. 245–65. Göttingen, Ger.: Hogrefe
- Reips U-D. 2000. The Web experiment method: advantages, disadvantages, and solutions. See Birnbaum 2000b, pp. 89–117
- Reips U-D. 2001a. Merging field and institution: running a Web laboratory. See Reips & Bosnjak 2001, pp. 1–22
- Reips U-D. 2001b. The Web experimental psychology lab: five years of data collection on the Internet. *Behav. Res. Methods Instrum. Comput.* 33:201–11
- Reips U-D. 2002a. Context effects in Web surveys. See Batinic et al. 2002, pp. 69–79

- Reips U-D. 2002b. Standards for Internet experimenting. *Exp. Psychol.* 49:243–56
- Reips U-D. 2002c. Theory and techniques of conducting Web experiments. See Batinic et al. 2002, pp. 219–49
- Reips U-D. 2002d. Internet-based psychological experimenting: five dos and five don'ts. *Soc. Sci. Comput. Rev.* 20:241–49
- Reips U-D, Bosnjak M, eds. 2001. *Dimensions of Internet Science*. Lengerich, Ger.: Pabst Sci.
- Reips U-D, Neuhaus C. 2002. WEXTOR: a Web-based tool for generating and visualizing experimental designs and procedures. *Behav. Res. Methods Instrum. Comput.* 34:234–40
- Ruppertsberg AI, Givaty G, Van Veen HAHC, Bülthoff H. 2001. Games as research tools for visual perception over the Internet. See Reips & Bosnjak 2001, pp. 147–58
- Sassenberg K, Kreutz S. 2002. Online research and anonymity. See Batinic et al. 2002, pp. 213–27
- Schiano DJ. 1997. Convergent methodologies in cyber-psychology: a case study. *Behav. Res. Methods Instrum. Comput.* 29:270–73
- Schillewaert N, Langerak F, Duhamel T. 1998. Non-probability sampling for WWW surveys: a comparison of methods. *J. Mark. Res. Soc.* 40:307–22
- Schmidt WC. 1997a. World-Wide Web survey research: benefits, potential problems, and solutions. *Behav. Res. Methods Instrum. Comput.* 29:274–79
- Schmidt WC. 1997b. World-Wide Web survey research made easy with WWW Survey Assistant. *Behav. Res. Methods Instrum. Comput.* 29:303–4
- Schmidt WC. 2000. The server-side of psychology web experiments. See Birnbaum 2000b, pp. 285–310
- Schmidt WC. 2001. Presentation accuracy of Web animation methods. *Behav. Res. Methods Instrum. Comput.* 33:187–200
- Schmidt WC, Hoffman R, MacDonald J. 1997. Operate your own World-Wide Web server. *Behav. Res. Methods Instrum. Comput.* 29:189–93
- Schwartz A. 1998. Tutorial: PERL, a psychologically efficient reformatting language. *Behav. Res. Methods Instrum. Comput.* 30:605–9
- Schwarz S, Reips U-D. 2001. CGI versus JavaScript: a Web experiment on the reversed hindsight bias. See Reips & Bosnjak 2001, pp. 75–90
- Smith MA, Leigh B. 1997. Virtual subjects: using the Internet as an alternative source of subjects and research environment. *Behav. Res. Methods Instrum. Comput.* 29:496–505
- Stanton JM. 1998. An empirical assessment of data collection using the Internet. *Pers. Psychol.* 51:709–25
- Stevenson AK, Francis G, Kim H. 1999. Java experiments for introductory cognitive psychology courses. *Behav. Res. Methods Instrum. Comput.* 31:99–106
- Suler J. 1996/2003. *The Psychology of Cyberspace*. <http://www.rider.edu/~suler/psyber/psycyber.html>
- Tourangeau R. 2004. Survey research and societal change. *Annu. Rev. Psychol.* 55:775–801
- Tuten TL, Urban DJ, Bosnjak M. 2002. Internet surveys and data quality: a review. See Batinic et al. 2002, pp. 7–26
- Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5:297–323
- Van Veen HAHC, Bülthoff HH, Givaty G. 1998. Psychophysics tests on the Internet: an evaluation. *Perception* S27:179
- Voracek M, Stieger S, Gindl A. 2001. Online replication of evolutionary psychological evidence: sex differences in sexual jealousy in imagined scenarios of mate's sexual versus emotional infidelity. See Reips & Bosnjak 2001, pp. 91–112
- Wallace PM. 2001. *The Psychology of the Internet*. Cambridge, UK: Cambridge Univ. Press
- Welch N, Krantz JH. 1996. The World-Wide Web as a medium for psychoacoustical demonstrations and experiments: experience and results. *Behav. Res. Methods Instrum. Comput.* 28:192–96

- Welker M. 2001. E-mail surveys: non-response figures reflected. See Reips & Bosnjak 2001, pp. 231–38
- White RJ, Hammer CA. 2000. Quiz-o-matic: a free Web-based tool for construction of self-scoring on-line quizzes. *Behav. Res. Methods Instrum. Comput.* 32:250–53
- Wilhelm O, McKnight PE. 2002. Ability and achievement testing on the World Wide Web. See Batinic et al. 2002, pp. 151–80
- Williams JE, McGraw KO, Tew MD. 1999. Undergraduate labs and computers: the case for PsychExps. *Behav. Res. Methods Instrum. Comput.* 31:287–91