Edited preprint version.
In M. H. Birnbaum (Ed.), *Psychology Experiments on the Internet* (pp. 89-117). San Diego, CA: Academic Press.

Title: The Web Experiment Method: Advantages, Disadvantages, and Solutions

Ulf-Dietrich Reips
Allgemeine und Entwicklungspsychologie
Universität Zürich
Attenhoferstr. 9
CH-8032 Zürich
Switzerland
E-mail: ureips@genpsy.unizh.ch
Tel.: +41-1-6342930
URLs: http://www.psych.unizh.ch/genpsy/
http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html

Running head: WEB EXPERIMENT METHOD

The Web Experiment Method:
Advantages, Disadvantages, and Solutions

Ulf-Dietrich Reips
Zürich University

The World Wide Web (WWW) provides a new tool for experimental research. The Web experiment method differs in fundamental aspects from traditional laboratory and field experiments; therefore it can be used to validate previous findings. Web experiments offer (1) easy access to a demographically and culturally diverse participant population, including participants from unique and previously inaccessible target populations; (2) bringing the experiment to the participant instead of the opposite; (3) high statistical power by enabling access to large samples; (4) the direct assessment of motivational confounding; and (5) cost savings of lab space, person-hours, equipment, and administration. These and 13 other advantages of Web experiments are reviewed and contrasted with 7 disadvantages, such as (1) multiple submissions, (2) lack of experimental control, (3) self-selection, and (4) drop out. Several techniques and other detailed solutions are described that avoid potential problems or even turn them into useful features of Web experimentation.

INTRODUCTION

Web experimenting has only just begun, and the method is thriving. Since Norma Welch, John Krantz, and others conducted the first Web experiments in 1995 (Krantz, Ballard, & Seher, 1997; Reips, 1996a, 1996b; Welch, 1995; Welch & Krantz, 1996), this method has been used with increasing frequency, and in more and more areas (see American Psychological Society, 1995; Musch & Reips, chap. 3, this volume; Reips & Musch, 1999). There has been a steady increase in the number of Web experiments as well as in the number of researchers using this method. Musch and Reips (chap. 3, this volume) conducted a survey among Web experimenters which confirmed that factors considered important by most of the respondents were, in order of importance, large numbers of participants, high statistical power, time saving, and the chance to reach participants from other countries. Although they were not considered most important, all predictions of cost savings postulated by theoretical analyses of Web experimental methodology (e.g., Reips, 1995c, 1997b) were supported by the survey. Consequently, most survey respondents stated that they would certainly conduct another Web experiment. This shows that advantages outweigh the disadvantages in Web experimenting.

The term ”Web experiment” was coined to underline this method’s categorical distinctiveness from laboratory and field experiments (Reips, 1995a, 1995c, 1998b). However, the underlying logical criteria are the same as those in the other experimental methods. Hence, the definition of ”experiment” used here requires manipulation of the independent variable(s) and repeatability. If an independent variable is manipulated between-subjects, then participants must be randomly assigned to conditions. Likewise, a quasiweb experiment would involve nonrandom assignment of subjects to conditions (see Bredenkamp, 1996; Campbell & Stanley, 1963; Kirk, 1995). Many of the issues discussed in this chapter are also relevant to other forms of Web research, such as surveys and tests.

Traditionally, in the behavioral sciences, there are two forms of experiments: the laboratory experiment and the field experiment. Both forms are distinct from simple observation in being active ways of manipulating variables to gain knowledge. They come with certain theoretical and practical disadvantages, which often limit their power of explanation in principle. Throughout the chapter I will discuss in more detail how and why Web experiments might provide solutions to these issues. The problems are listed next:

- Number of participants or measures too small in many studies, because researchers set the Type I error probability α to a conventional level (and therefore the power of these studies is low; Erdfelder, Faul, & Buchner, 1996)

- Very limited sample populations, which raises the question whether psychological theories and body of research are too narrow in scope (for decades subjects have been mostly young students of local nationality [Reips & Bächtiger, 1999; Schultz, 1972; Smart, 1966]; therefore many results may be specific for young students and/or the culture in which a study was conducted)
- Limited external validity (in laboratory experiments) or limited control (in field studies)
- Less than optimal voluntariness of participation (a twofold issue: (1) although nobody is forced to participate in psychological experiments, many students feel they need to do it to fulfill course requirements, given low attractiveness of possible alternatives; (2) for situational, social, and procedural reasons participants may develop a sense of obligation to stay in the experiment once they have agreed to participate)
- Motivational confounding, in part following from the previous point
- Institutional and organizational limitations (e.g., lab hours are typically limited to daytimes during weekdays; equipment may limit the number of participants that can be tested in a given interval of time)
- Experimenter bias
- Nontransparency of much of the research process
- Frequently, under the traditional methods, limitation of what *can* be done experimentally (partly to avoid the preceding issues becoming too relevant) and what *is* done (due to the ease of conducting some experiments and the hardship of conducting others; Devereux, 1967; for an example of an experiment that was only feasible to be conducted as a Web experiment see Klauer, Musch, & Naumer, 1999).

Consequently, findings from laboratory experiments should be validated using other methods. This is one general application of Web experiments; another one is to allow for experimental research in previously unexplored areas. The new Web experiment category offers the chance of extending the validity and acceptance of experimental research. This chapter is intended to be a comprehensive overview of methodological issues of online experiments.

To understand why conducting Web experiments is an opportunity for our science, we have to look at some of the problems with traditional experiments. For a certain time, in the late 1960s and 1970s, there was a heightened interest in flaws of these methods. Researchers

like Rosenthal (e.g., 1965, 1966) and Cohen (e.g., 1977) relentlessly directed their enlightening spotlights to some of the dark corners in psychology's methodological basement. They discovered experimenter effects such as the Pygmalion effect (Brophy & Good, 1974; Rosenthal & Jacobsen, 1968), volunteer bias (Rosenthal, 1965; Rosenthal & Rosnow, 1969; Rosnow & Rosenthal, 1966), low power and other problems in experimental design (Cohen, 1977; Gigerenzer et al., 1989), demand characteristics (Orne, 1962; Rosenthal & Fode, 1973), and predominance of undergraduate psychology students as experiment participants (Jung, 1969; Schultz, 1972; Smart, 1966). Some issues were addressed, for example, the Hawthorne effect[1] (Roethlisberger, 1977) through use of nonexperimental control groups (Adair, 1973). All other problems listed here remain, however, and because they could not be resolved with the methods at hand or were too complicated to deal with, they have more or less been treated with a certain slippery indifference. Of course, these problems do not affect all areas of psychological research to the same degree. Also, in many studies, many researchers work very hard to minimize many of these effects in their studies. However, there are reasons many researchers have to work so hard to minimize these effects, and some issues are so fundamentally linked to the method that no measures can be taken to reduce their influence. Here Web experiments add some support to the basement.

MOIST BEAMS AND MILDEW STAINS ROTTING IN INDIFFERENCE

(and how Web experiments might be used for structural support)

Generalizability (Demographic)

One of the most obvious reasons for criticism is that most psychological research is done with undergraduate psychology students. As McNemar put it in 1942: ”The existing science of human behavior is largely the science of the behavior of sophomores” (p. 333). Although in hypothetico-deductive research it might not really matter who the participants are (Bredenkamp, 1996; Reips, 1999a), it would matter if results from psychological research are to be generalized to the general population and if the results of experiments differ in different populations (Reips, 1999a). Smart (1966) and Schultz (1972) noted that more than 80% of all psychological studies are conducted with students as participants, while only about 3% of the general population are students. Basically, this picture has not changed (Reips & Bächtiger, 1999).

Birnbaum (chap. 1, this volume) has written: ”Some say that psychological science is based on research with rats, the mentally disturbed, and college students. We study rats because they can be controlled, the disturbed because they need help, and college students because they are available.” If in research using traditional methods the main reason for the restriction in participant demography is the limited availability of participants, then Web experiments should be looked at as a serious alternative. Demographic characteristics of self-selected participants in Web experiments include a much greater diversity than is found in laboratory studies (e.g., Reips, 1996b). Demographics of Internet users is expected to rapidly approach similarity with demographics of the general population (Graphics, Visualization, & Usability Center, 1999). Therefore, a promising application of Web experiments would be the replication of results found in laboratory experiments, aiming at increasing the external validity of the body of research. For instance, in a recent Web experiment (Reips, 1999b) I was able to replicate findings from a study that looked at question context effects in answering behavior depending on whether a survey was administered by mail or by telephone (Schwarz & Hippler, 1994a). Apparently these context effects produced similar differences. It mattered whether survey questions were presented on one Web page (the ”mail survey” situation) or on consecutive Web pages (the ”telephone survey” situation). However, I could not replicate a numerical labeling effect (Schwarz & Hippler, 1994b) in that Web experiment (this is probably due to a different question used as the material, though).

Supporting the practice of using a demographically narrow sample within a study, one can argue that the processes studied in much of psychological research are independent of person type, and studying a uniform population to reduce error variance increases power holding *N* constant. This applies to Web experiments also, and the Web offers a nice feature for aiming at specific person types. One can direct Web participants to different Web studies using *characteristics-dependent redirecting*, which displays different Web pages depending on what was answered on earlier Web pages.

How can we be sure that the results of our studies are not specific to one culture? We need more replications with people from different cultures. The Web experimental method offers paths right onto the desktops of people from all over the world.

Web experiments provide the researcher with easy access to a much wider and geographically diverse participant population. As the term implies, the World Wide Web spans the whole globe and even extends into outer space (there have been live WWW conferences with space shuttle crews). The powerful implication of this development for psychological research comes from particular characteristics of the Internet, namely the nearly complete freedom of space and time. Communication is almost instantaneous between any two points on earth. Basically, this means that in Web experiments many persons from the general population are as accessible to the researcher as students in the local campus cafeteria have always been.

Nevertheless, local accessibility and density of the Web remain unbalanced. The numbers both of Web servers and of Web clients are the highest in industrialized countries. It is estimated that ”Internetization” of U.S. households will approach 58% by 2003 (Internet.com LCC, 1999). Because average income and education of Internet users are higher than the average of the general population (Graphics, Visualization, & Usability Center, 1999), certain strata of the population might be overrepresented in Web research.

On the other hand, the vast size of the Web offers room for tens of thousands of small ”virtual communities” with a wide variety of user profiles (Rheingold, 1993). This allows for targeted recruitment, including participants from very unusual or previously inaccessible populations such as drug dealers (Coomber, 1997), people with particular head injuries (Browndyke, Santa Maria, Pinkston, & Gouvier, 1999), or people suffering from panic attacks (Stones & Perry, 1997).

The advantage of eased access is twofold, as Web experiments allow people to experience psychological research who would never have had the chance to do so due to geographical, cultural, or social barriers. Scientific psychology becomes more accessible.

Web sites with Web experiments, especially online laboratories (e.g., Birnbaum, 1998; Bülthoff, van Veen, Givaty, Braun, & Georg, 1997; Pagani & Lombardi, 1997; PsychExps, 1998; Reips, 1995a; Schubert & Waldzus, 1996), mostly include a rationale of the research at hand and give an explanation of what experimental psychologists are doing. This might well heighten the visibility of psychological research.

Generalizability (Situation) and External Validity

It has long been seen as one of the major disadvantages of laboratory experiments that in this highly controlled situation people might produce results that cannot be transferred to their behavior in the "real world" (Martin, 1996). Chapanis (1970) argues that the external validity of laboratory experiments is generally low (however, for a different view see Anderson & Bushman, 1997). If one wants to generalize findings that have originated in laboratory research one runs into the difficulty that the laboratory situation is mostly very different from what is a natural setting for the participants. This gap might create an "artificial" or "sterile" atmosphere in many laboratories, as participants find themselves being in an often unfamiliar place they cannot personally relate to. It seems quite obvious that it is possible that many people will behave unusually under these circumstances. In Web experiments a large percentage of participants remains in the familiar situation at the computer at home or at work. According to the 10th GVU WWW user survey 79% of respondents *daily* browse the WWW from home, and 57% from work, while 77% *never* browse from school, 72% *never* browse from public terminals, and 60% say they *never* access the WWW from places other than the previous choices (Graphics, Visualization, & Usability Center, 1998).

Besides familiarity with the physical situation and its associated comfort there is another important aspect to Web experimenting: bringing the experiment to the subject instead of the opposite. Web experiments spare participants from scheduling, from transportation, from hassles finding campuses, buildings, rooms within buildings, and the right person to talk to. No walks through long hallways, only a few mouse clicks. Extrapolating current technical development, it will soon be common to connect to the WWW with wireless laptop and palmtop computers. While the dependence on a technical interface may limit external validity in Web experiments, the wide variety of situations that allow for access will increase external validity in Web experiments, compared to laboratory experiments. To make another point for this claim: Participants in Web experiments can freely choose at which time of day (or night) and on which day they wish to participate.

Generalizability (Time)

Laboratories usually are subject to all kinds of institutional regulations, including limited access times. It is probably no overstatement that (with certain exceptions, such as sleep research laboratories) psychology's body of research says a lot about people's behavior during weekday daytime working hours. Web experiments can be accessed all the time, all around the clock. Data from Web experiments lend themselves very easily to making post-hoc comparisons depending on time of day. Furthermore, these data might be less influenced by interactions between participants' biological rhythms and levels of the independent variable(s) used, as it is likely that in Web experiments with self-selected access personally comfortable participation times will be chosen. Because thousands of Web pages can be served at the same time, there is practically no limit to simultaneous use of the materials. Consequently, no scheduling difficulties arise, and overlapping sessions do not produce organizational nightmares as well.

This sounds like complete freedom for participants. But what if an experimenter wants to control for time of day or weekday? Web experiments offer a previously unattainable technical capability of interactively controlling display of experimental materials dependent on preset variables. For example, it is possible to only serve an experiment's materials during a certain hour while during all other times logging in leads to a different experiment. That way one might do a study limited to people surfing the Internet during their local nighttime between 2 and 3 a.m. Additionally, one could cross this *time-dependent redirecting* with a second factor, such as one that is based on the participant's domain. That way one could determine whether persons connecting from U.S. American educational institutions (.edu) differ from those connecting from commercial institutions (.com). In turn, this way one can test whether it makes a difference when and from where someone participates. Other variables readily available for redirecting are the Web page a surfer comes from, the operating system he or she is using, the screen resolution the monitor is set to, left- versus right-handedness (determined by type of mouse arrow movement), or any input the participant might give in a form built into the first page.

Volunteer Bias

It has been shown that complete voluntariness of participation in laboratory experiments may lead to results different from those obtained with less voluntary participation (Rosenthal & Rosnow, 1969; Rosnow & Rosenthal, 1966; Spiel, 1988). Also, there is evidence that

motivation for participation may be a moderating variable. Oakes (1972) compared behavior in group discussions of persons recruited through advertisements in two newspapers with such behavior by "coerced" students and found dramatic differences. In most studies with psychology students, the students participate to fulfill an assignment for a class. As has been noted before, the psychological body of research is mainly based on studies conducted with students (Buchanan & Smith, 1999; Reips & Bächtiger, 1999). Having to fulfill a research assignment in a college course often leaves students with the choice between serving as a subject for extra credit and a not particularly desirable alternative (such as writing a paper). These students are likely to be less-than-eager participants. Consequently, much of our current psychological knowledge has been derived from the behavior of students in "not perfectly voluntary" situations. These people are probably less motivated than people who seek out and choose to participate in Web experiments. Perhaps these Internet participants are more likely to exercise their freedom to quit the experiment than those who serve in the lab. Participants in Web experiments can "leave" with the touch of a button.

Of course, something has to be said about the downside of the lower commitment to participation in Web experiments. Resulting dropouts could adversely affect the conclusion. Techniques for reducing the dropout rate and limiting its impact will be discussed later in this chapter.

Statistical Power

It has been noted that the statistical power of studies in psychological research is notoriously low (Erdfelder et al., 1996). As Erdfelder et al. remind us: "There are only two ways to raise the power if the null hypothesis (H0), the alternative hypothesis (H1), and the test statistics have already been specified: One must increase either the sample size *N* or the Type I error probability α" (p. 2). In laboratory experiments one often is forced to take the unusual step of increasing α, if one wants to draw meaningful conclusions from experiments, as there are theoretical limits to increased precision of the error measurement and more control of unwanted variability, and pragmatic barriers to an increase of *N*. Time, lab space, and finances limit the number of participants one can afford. In Web experiments, however, it is no problem to reach the "ideal" calculated sample size (not too small and not too large) while remaining at a conventional α-level. On the WWW the participant pool is of almost unlimited size.

To support the claim that in Web experiments it is no problem to achieve the optimal calculated sample size Figure 1 shows monthly visits to the *Web's Experimental Psychology*

Lab (Reips, 1995a). Currently, the number of visits is about 3000 a month. The issue of ”hits” and ”visits” is a complicated one. In Figure 1 visits are single downloads of the plain (not the animated) entry pages to the Web’s Experimental Psychology Lab. This gives a rough estimate of true visits by people, as I could determine by four other methods: WebTracker (FxWeb Web Technologies, 1999); JavaScript; figures for certain images; figures for visits from certain IP addresses. The estimate has to be corrected by subtracting robot visits and cases when users switch between language versions, when they reload the pages, when their cache is turned off and they go back, and when they return to the page at the end of an experiment (unless they went to the experiment directly). On the other hand, one has to add those cases when people get the pages from proxies (this number has been increasing in recent years), when several persons connect using the same IP addresses, and all cases for the Zurich version of the lab, as these are not included. On both the positive and the negative side a few other cases are possible. Of course, visits are not ”people who have been in the Web lab exactly one time”. However, with the help of persistent cookies I am able to estimate quite accurately how many people return (about 75% do not) and how often.

Figures similar to the number of visitors to the Web’s Experimental Psychology Lab can be reached for single Web experiments through extensive advertising in newsgroups and on highly frequented Web sites (Jochen Musch, personal communication, April 6, 1998). Being able to potentially reach large numbers of participants is the single most important factor for Web experimenters to conduct Web experiments (Musch & Reips, chap. 3, this volume).

Insert Figure 1 about here

Detectability of Motivational Confounding

An issue strongly related to the voluntariness of (respectively eagerness for) participation is possible confounding of experimental manipulation with motivation and/or task difficulty. In traditional laboratory experiments, levels of a participant’s motivation to engage in the experimental task might be confounded with levels of the independent variable. However, those participants who are in the less motivating condition usually will not indicate so by leaving the situation. They will stay and finish the experiment, thereby possibly contaminating the data. In Web experiments with a between-subjects design a differential dropout rate in experimental conditions would indicate such a motivational confounding. This

information can then be used to address issues such as task difficulty or task attractiveness, for example by introducing control conditions.

With participants being able to terminate the experiment at any time during a Web experiment, *condition-independent* dropout due to motivational issues can become much more problematic than in laboratory experiments. It can be assumed that many participants in Web experiments will not stay in the experimental situation, as their commitment might be lower. On the other hand, their commitment is not based on reasons such as course credit considerations or ”peer pressure” by the experimenter, thereby creating less error variance (Reips, 1996b, 1997b, 1999).

How can dropout behavior be used to detect motivational confounding? Spotting motivational confounding requires a between-subjects design for the Web experiment, that is random distribution of participants to at least two conditions. Generally, selective dropout in one condition might make an experiment worthless (unless dropout is the dependent variable, of course). Participants in a less motivating or boring experimental condition, who in a laboratory experiment would have stayed due to, for example, course credit considerations, might very likely drop out of a Web experiment. In this case the laboratory experiment data would be contaminated by motivational confounding; the Web experiment data would allow for detection of this effect. Put as a general rule: The less free participants are to leave the experimental situation the less likely it is that motivation-related confounding variables will become salient. Consequently, the heightened degree of voluntariness in Web experiments allows for detecting confounds with variables that potentially decrease motivation, such as task difficulty. Between-subjects Web experiments with voluntary participation have a built-in detection device for this type of confounding, which will be discussed in the section on ”Solutions.”

Experimenter Effects and Demand Characteristics

Basic problems that trouble experimental research and tie in with the issue of external validity are demand characteristics (Orne, 1962) and experimenter effects (e.g., Barber & Silver, 1968; Rosenthal & Fode, 1973). Demand characteristics are – often subtle – clues that influence participants in the experimental situation by ”demanding” certain reactions. They include such clues given by the experimenters. Experimenter effects are ”mildew stains” in psychological science, as they might have biased a large portion of experimental results. As long as experimenters are present in experiments, there is the potential that they might give subtle clues or make errors that might systematically bias their data. However, the notion of

experimenter effects also includes biases introduced by the experimenters during execution of the experiment, data transformation, data analysis, and interpretation of the results. ”Every attempt should be made to minimize those demand characteristics that might become confounding variables. … Demand characteristics can be controlled by automating as much of the experiment as possible” (Martin, 1996, p. 72). Although automation might reduce external validity in some cases due to the artificiality of the situation, it will increase it as long as some laboratory-specific demand characteristics are avoided. Web experiments are automated by definition, and experimenter influence is minimized to preparation and data analysis.

Which advantages does the Web experiment method offer regarding experimenter effects? In most Web experiments there is no interaction between experimenters and participants. This avoids a large source of potential experimenter effects. However, experimenters also bias experiments by making errors in the preparation of the experiment, data transformation, data analysis, and interpretation of the results. The more a Web experiment is automated, the less opportunity there is for undetectable experimenter effects (however, because there is no one to ask questions of, instructions for Web experiments must be particularly clear – this problem is addressed later in this chapter). Also, there might be a higher chance for detection of experimenter errors in these tasks, because the material and the records are available on the Internet and are open to review. On this issue another Web experimenter commented:

> I do not believe that detecting subtle errors in experimental design or data analysis is going to be easier for Web experiments than traditional lab-based ones. We have a reaction time study using visual and sound stimuli. Unbeknownst to us, the sound files included about 25 msec of silence at the start of each sound. This was being added to reaction times. This type of error can occur in the lab as easily as on the Web, of course, but in either case it may go undetected for some time. (Kenneth McGraw, personal communication, January 15, 1999).

Incidentally, I had heard about a problem with this particular experiment from one participant, before reading the comment. The participant wrote:

> Try the reaction time experiments. Some of the experiments on reaction time, in my browser, give me an unintended advance warning

> of the coming of the sound cue (on my Netscape screen, I see the system's hourglass just before the sound is presented--clearly, that is a problem in the software---if the sound is delivered before the trial starts, I won't have that cue...If all the sounds are delivered well before the start of the experiment, then I may have too long a wait and give up on doing the experiment. Instead, the sound is being delivered as part of each trial's pre-trial waiting period, so it gives a cue of the timing of the sound onset.

Although this participant erroneously attributed the faulty material issue to the Web experiment software and concluded to critically look at these issues with Web experiments, he noticed that something must be wrong and nicely showed how the open nature of Web research indeed allows for detection and quick correction of errors through feedback by participants and the research community.

Taken together, the mildew stains in psychological laboratory research might well cover some established findings, identifying them as tainted beams, as Web experiments now allow for the validation of traditional research methods. The point is that some of the problems that have plagued the methodology of psychology experiments can be eliminated or at least addressed through the use of Web experiments in lieu of or alongside traditional field or lab research.

OTHER ADVANTAGES OF WEB EXPERIMENTING

Costs

Web experimenting reduces costs, because of savings of lab space, experimenters, equipment, and administration. Web experiments run around the clock and allow for simultaneous access by a large number of participants. Several participants might use the experimental material at the same time without knowing of each other. Consequently, there are no scheduling problems. Institutional regulations (e.g., limited lab hours) lose their confining impact on research. Most Web server programs allow for thousands of simultaneous connections. To deal with large numbers of simultaneous participants in Web experiments is only feasible because no experimenter needs to be present during the experiment. Once a Web experiment is programmed and on the WWW no salaries have to be paid for experimenters.

To conduct Web experiments is financially attractive in even more ways: no rooms for laboratories are needed, and no bureaucracy regarding scheduling, insurance, and so forth. In their simplest form, Web experiments can be run from a desktop computer, which can still function as a personal computer. Although some money for the transmission of the data has to be paid, comparable Web experiments are much more cost-effective than traditional experiments.

The cost argument should be seen within the proper scope. Although one could argue that costs of data coding and data entry from paper to computer are also saved by most Web studies that use scripts to code and store the data, this is no additional advantage in comparison to other computer-driven experiments. Furthermore, one has to bear in mind that some very expensive areas within psychology are not likely to profit from the advantages of Web experimentation, as they rely on specialized equipment (PET scans, MRIs, EEGs, etc.).

General Advantages for the Research Process

Openness, one of the fundamental principles of science (Merton, 1942/1973), can be achieved much better in Web experiments than in laboratory experiments. Traditional experiments may contain features that are not described in the method section that may turn out to be important. Public Web experiments (there are also hidden Web experiments resembling the field experiment type, e.g., Hänze & Meyer, 1997) are openly accessible and can remain indefinitely on the WWW for documentation purposes. Institutions such as publishing houses or online libraries can house these materials and guarantee their accuracy.

Also, there might be development of online agencies (e.g., by funding institutions) that will do searches and collect experimental materials. This allows other researchers to look at the materials and to know in detail the procedure used. Direct hypertext links to articles in online journals such as *Psycoloquy* (1999) may be established and will provide readers with the possibility of experiencing the described experiment from a participant's perspective. Interactive editing of the materials will allow fast setup and modification of Web experiments. The data can also be made accessible online and open for reanalysis. All stages of psychological research will become more public. (For a more elaborate vision of the future research process see Reips, 1997a, 1998a).

The Web experimental method opens the door to research areas that were almost inaccessible for established methods. Examples include studies with children (Reips, 1999c) or with participants from very unusual, specific target populations such as Melungeons (Melungeon Heritage Association, 1999), people with diabetes (newsgroups alt.support.diabetes, alt.support.diabetes.kids, de.sci.medizin.diabetes, misc.health.diabetes), disabled artists (newsgroup alt.support.disabled.artists), mothers of triplets (newsgroup alt.parenting.twins-triplets), people living in long-distance relationships (alt.support.relationships.long-distance), drug dealers (Coomber, 1997), or people with panic attacks (Stones & Perry, 1997). In the past, finding and contacting people who had the desired characteristics would have required an enormous effort, not to mention transportation to the laboratory. By using newsgroups and Web experiments, studies with specific, unusual samples can be conducted almost as easily as those with standard sample populations.

Other Procedural Advantages

Literal Web experiments (i.e., those conducted on the WWW) technically permit one to check for the number of nonparticipants. Because all visits to a Web page are written to a log file the number of visitors to the Web page with a hyperlink to the first page of a Web experiment that do not follow the link approximately equals the number of nonparticipants in that Web experiment (for Web experiments using no-cache tags in the HTLM headers, and corrected for robot visits). Of course, such a comparison is only an assessment of participation rate among those people who visit the Web page or Web pages that offer a hyperlink to the Web experiment, if there are no other ways the experiment is announced (finding Web experiments through search engines can be prohibited by using so-called *robot exclusion tags* in the header of the Web pages containing the experimental material). In laboratory experiments usually one does not know how many people walked past the sign advertising an

experiment. Another way participant's traces on the Web (by analyzing "referer" information) can be used is mentioned later as the *multiple site entry technique*.

With Web experiments it is always possible to validate results by administering the same materials to a local sample in the laboratory. This allows specifically for assessment of volunteer bias, demand characteristics, experimenter effects, and external validity. Another advantage of Web experiments is the technical variance. This issue will be discussed in the "Disadvantages" section, as it might appear to be a disadvantage on first view.

Ethical Issues

In Web experiments, participation is completely voluntary throughout the whole experiment. Participants may drop out of the experiment at any time, and therefore a Web experiment might have much less of a restricting influence on participants than laboratory experiments.

Public display of Web experiments allows for better control of ethical standards. Participants, peers or other members of the scientific community might look at an experiment and communicate any objections by e-mail. As in other research, if an experimenter were to violate ethical guidelines, that person would face the disapprobation of peers and possible sanctions from organizations (e.g., APA) that also restrict membership and publication to people who are ethical in their research. Also, the Internet allows for sanctions by individual participants, an issue that needs further discussion among Web experimenters, as the implications can be grave.

Although current ethical standards in psychological research, such as those published by national psychological associations (e.g. American Psychological Association, 1992; Föderation der Schweizer Psychologinnen und Psychologen, 1997), cover most of the ethical issues related to conduct of Web experiments, the international character of the WWW might create some conflicts. Standards are not the same everywhere, and feelings about ethical issues in conducting research might vary internationally just as those about uncovering faces in public, the death penalty, or the limitation of free speech for certain controversial issues in some countries but not others (e.g., Nazi propaganda is legal in the United States, but not in Germany). For example, ethical aspects of research conducted at British or American institutions is mostly controlled by local review boards. These boards often require a "participant consent form," which is rather unusual at continental European institutions.

DISADVANTAGES AND SOLUTIONS

The list of advantages of Web experiments is long. Nevertheless, Web experimenters should not overlook the beams in their own eyes. In this section, potential problems will be discussed and some solutions offered.

Control Issues ("Cheating")

It seems that for most researchers questions about control of possible "cheating" are the first ones that come to mind when they think about conducting Web experiments, and indeed almost all Web experimenters take some precautionary measures to ensure a minimal degree of control (Musch & Reips, chap. 3, this volume). To guard against multiple submissions, one may simply ask participants not to participate more than once or ask them about their seriousness regarding their answers. Another solution is to provide repeaters with a checkbox to indicate their status. Also, it is common to check the participants' e-mail addresses or computer addresses (IPs). Both pieces of information help to uniquely identify the participants. However, this information is not completely accurate. One cannot easily draw the conclusion that independence of observations is guaranteed. People could also log in twice from different computers (a rare case) or – increasingly more likely – be assigned different IPs through dynamic addressing (Reips, 1999a). Hence, checking for multiple participation by using a more conservative combination of IP address and e-mail address might be a reliable alternative. Although e-mail addresses can be checked for validity, it is difficult to determine whether the same person participates repeatedly using more than one valid e-mail address. Most such cases can be sorted out by only including the first data set from a given IP address. Unfortunately, this conservative procedure will also abandon some valid data, as the same IP may be used by different people working on the same computer. Another case of shared IPs are proxy servers that can be used by a large number of people. Nevertheless, relying only on the first data sets from IPs also decreases the likelihood of biasing effects produced by participants talking to each other. The chance that a whole group of participants enters an experiment with the same hypothesis in mind seems less likely on the Web than in labs on campuses where students may exchange information on experiments and speculate on what the experiments are about (a phenomenon that can also be observed if Web experiments are announced in newsgroups; in this case the experimenter knows about what is being speculated, however).

At a time when occurrence of dynamic addressing was still very rare, I conducted an experiment on learning of consistency of causal mechanisms (Reips, 1995b, 1996a, 1996b,

1997b), which turned out to be the earliest between-subject experiment conducted on the Web (Musch & Reips, chap. 3, this volume). I had to exclude only 4 cases out of 880 in that Web experiment using a strong criterion of IP plus e-mail address plus "similar IP (all but digits after last period) within a 2 day period." Although there may be more cases of multiple participation in Web experiments of shorter duration (the one mentioned took about 45 minutes to complete), we may be safe to assume that "cheating behavior" is rare. However, if one wants to be cautious or wishes to determine true percentage of "individual identities" one might use measures such as the techniques described next.

Proxy servers can make the submissions of different users appear to have been sent from a single person. To counter this as well as other problems, one may use a *password technique* (Reips, 1997b, 1999; see Figure 2, top). Although handing out passwords reduces the number of participants prior to start of the Web experiment and may be used as a way of countering the problem of self-selection, it requires considerable effort. Individual user IDs and passwords have to be created and set in the Web server, and both need to be mailed to preselected participants. A technique that can be applied after conducting a Web experiment is the *subsampling technique* (Reips, 1997b, 1999; see Figure 2, bottom). This procedure consists of randomly drawing a number of participants' e-mail addresses and mailing them a follow-up questionnaire, in which one might repeat some of the questions from the Web experiment to control for the proportion of false answers. Also, one might ask for faxing of a personal ID.

Insert Figure 2 about here

Another procedure that may be used to secure data integrity is to ask for various personal identification items that can be checked independently or are difficult to counterfeit:

- Ask for fax of a personal ID
-
- Ask for bank account number, if a small monetary reward is to be transferred for participation
- Ask for e-mail addresses of two people who will confirm identity
- Ask for URL of personal home page
- Any combination of items above

Additionally, data quality can be controlled for by checking internal consistency and date and time consistency of answers. A feature that is built into newer versions of HTML are so-called Cookie", which allow for checking of repeated visits to a Web page. A Cookie is

> a unique string of letters and numbers that the Web server stores in a file on your hard drive. This method is used by Web designers to track visitors to a Web site so the visitors do not have to enter the same information every time they go to a new page or revisit a site. For example, Web designers use Cookies to keep track of purchases a visitor wants to make while shopping through a Web catalog. Cookies may work through a single visit to a Web site, such as when tracking a shopping trip, or may be set to work through multiple sessions when a visitor returns to the site (Netscape, 1999).

However, using Cookies is frowned upon in the Internet community, as they can be misused to create user profiles for commercial purposes. Using Cookies in Web experiments will not capture those participants who have this feature turned off in their Web browser.

Generally, multiple participation by the same persons is not very likely, as most Web experiments are not that thrilling. It seems unlikely that people would repeat long or boring experiments. As mentioned above, dynamic addressing makes it increasingly difficult to detect individual computers, but looking at data collected at a time of almost no dynamic addressing supports the presumption that multiple participation is rare.

Control Issues (Experimental Control)

Another issue is the desire to have as much control of the experimental situation as possible. Earlier in the chapter it was discussed how this desire contradicts external validity of experiments. Much of the situation cannot be controlled in Web experiments. Consequently, conducting Web experiments requires careful experimental design, including criteria such as factorial between-subjects design with randomized distribution of participants to the experimental conditions (Reips, 1996b, 1997b). Randomization can be realized in Web experiments through the use of so-called CGIs, small computer programs that cooperate with the Web server (Kieley, 1996; Reips, 1996b). Consequent use of these techniques, in combination with a large sample, will almost always detect an effect, which will be highly generalizable, as most variables (apart from the independent variables, of course) are allowed to vary freely. As Martin (1996) states: "As a rule of thumb, the more highly controlled the

experiment, the less generally applicable the results, if you want to generalize the results of your experiment, do not control all the variables." Some experimenters might disagree with this rule, and I am in accord with them for certain quasi-experimental and correlational designs. However, as a rule of thumb the rule is quite logical for between-subjects experiments with random distribution of participants to conditions, as this also means random distribution of all uncontrolled influences to the experimental conditions.

Finally, one technique that might be attractive for researchers who want to know who their participants are is the *participant pool technique* (see Figure 3). Persons who sign up for this pool provide the Web experimenter with their demographic data and can be paid for participation. Additionally, the experimenter keeps a record of who participated in which Web experiments. This technique also allows for drawing stratified samples.

Insert Figure 3 about here

Self-Selection

Self-selection can be considered the most serious problem in online research, as there is no perfect solution as long as participation is voluntary. Just as one might expect selective participation of politically interested persons in a survey on a Web page titled "To participate in a survey on political ignorance, click here" there seem to be pockets of self-selected areas in the WWW between which there is not much surfing traffic. Although we cannot do much about the possibly detrimental nature of self-selection, we can make an attempt to estimate its influence on our results. The technique used here is the *multiple site entry technique* (see Figure 4). Several entry pages or entry hyperlinks on different Web sites lead to the first page of the Web experiment. Later the data of participants coming from the different sources are compared. If they are similar for entry pages with very different content, then one is safe to conclude that self-selection did not play much of a role in determining the results.

Insert Figure 4 about here

Dropout

Comparatively high dropout rates are the downside of the voluntary nature of participation in Web experiments. Birnbaum and Mellers (1989) showed that even with equal

dropout rates in all conditions, if there are dropouts, even a between-subjects experiment can lead to wrong conclusions about the direction of an experimental causal effect. This happens if participants drop out of different conditions for different reasons. Therefore, Birnbaum and Mellers recommend checking dropouts for any correlations of demographic or other background variables with the experimental conditions. This requires assessment of these variables before introduction of conditions. A way of countering dropout in Web experiments is a *warm-up phase* before the point of random distribution of participants to experimental conditions. I used this technique in one of my first Web experiments, where the experimental manipulation was not introduced until about 20 minutes into the Web experiment, resulting in a dropout rate of only 9.7% (dropout was defined as not filling out any of the questionnaires at the end of the Web experiment; Reips, 1996a, 1996b, 1997b). Musch and Reips (chap. 3, this volume) found that the average dropout rate in Web experiments is 34% (median, 35%), with a range from 1 to 87%. The large range suggests that there are other motivational factors that influence dropout, two of which could be verified empirically: monetary rewards for participation and placement of assessment of personal information (PI). Frick, Bächtiger, and Reips (1999) manipulated these two factors in a Web experiment and found the following drop out rates: 5.7% in the condition with PI and financial incentive information (FI) at the beginning of the experiment, 13.2% with PI at end and FI present, 14.9% with PI at start and no FI, and 21.9% in the PI at end / no FI condition. Musch and Reips showed that completion of Web experiments was 86% if some form of reward (individual payments or lottery prizes) was offered, up from 55% without such rewards.

Other factors might also be at work determining continued participation. Some might be technical in nature: In a Web experiment on the reversed hindsight bias, we found that a Javascript version of the Web experiment resulted in a 13.4% larger drop out than an otherwise identical CGI version (Schwarz & Reips, 1999). Several factors have been theoretically explored (Reips, 1996b, 1997b, 1999a) and can be turned into a list of recommendations for Web experimenters:

- Create an attractive Web site by
  - Using a nice looking Web design
  - Having people create links to your site
  - Putting up signs that this could be an interesting site (awards, comments)
  - Not using commercial banners
  - Having multilingual pages
  - Offering various versions (frames, no frames, text).

- Emphasize your site's high trustworthiness by
  - Providing the name of your institution
  - Emphasizing the scientific purpose
  - Ensuring (and keeping) confidentiality
  - Providing contact information.
- Offer a gratification (e.g., the chance of winning a prize).
- Offer feedback (individual or general; for a procedure that can be used to offer individual feedback see Schmidt, chap. 3, this volume).
- Use a Web design that results in systematic shortening of loading times.
- Provide participants with information about their current position in the time structure of the experiment.
- Use the "high entrance barrier" technique.

The *high entrance barrier technique* (Reips, 1999a) is a package of procedures that can be applied to provoke early dropout and ensure continued participation after someone makes the decision to stay. This means bundling of demotivating factors at the very beginning of a Web experiment (i.e., on the general instructions page). Motivating factors should come to work increasingly thereafter, enticing participants to continue with the experiment. Several high entrance barrier techniques are listed here:

- Tell participants participation is serious, and that science needs good data.
- Personalize – ask for e-mail address and/or phone number.
- Tell them you can trace them (via their computer's IP address).
- Be credible: tell them who you are, and what is your institutional affiliation.
- Tell them how long the Web experiment will take.
- Prepare them for any sensitive aspects of your experiment (e.g. "you will be asked about your financial situation").
- Introduce your experimental manipulation after a warm-up phase.
- Tell them what software they will need (and provide them with hyperlinks to get it)
- Perform Java, JavaScript, and plug in tests (but avoid Active-X).
- Make compliance a prerequisite for winning the reward.

Despite all precautions, dropout in self-selected within-subjects designs poses a serious threat to the validity of this type of research. In between-subjects designs, dropout can be turned into a tool for detecting confoundings, as has been described in the "Advantages" section.

Technical Variance

Usually, in a highly controlled experiment, a researcher tries to minimize all error variance, such as that which comes from the technical instruments used. Conducting experiments over the Internet involves a worldwide network of cables, routers (computers at cross points of Internet traffic), satellites, plugs, and so on, so that naturally there is a fluctuation in ”net lag”. This is a very fundamental property of the Internet, as information never goes the same way, but is split up into thousands of ”packets” which try to find their own way to later be reassembled. A second source of technical variance lies in the wide range of computers, monitors, speakers or ear phones, Web browsers, and net connections used by Web surfers.

In between-subjects Web experiments these types of technical variables can be seen as advantageous, because possible undetected sources of systematic error are replaced by random error (Reips, 1996b, 1997b). For example, if a technical problem with a monitor used in a laboratory experiment interferes with whatever is studied, then it will systematically change the results. This will not happen in a Web experiment, because every participant sits at a different monitor. Control is replaced by randomness, and therefore generalizability is increased. Of course, generalizability might be limited by *systematic* changes in the environment, such as uniform changes of Web browser features.

Interaction with Participants

Often in experiments it happens that participants have questions regarding the understanding of the instructions. Then they ask the experimenters. Because this is not feasible in Web experiments, comprehension of material and task cannot be supported by the experimenters. Although this lack of interaction in Web experiments may reduce experimenter effects, lack of comprehension may become a serious threat to the experiment’s validity. A solution to this problem is to use pretests for comprehension difficulties with the materials. Also, providing participants with the opportunity of giving feedback by writing e-mail or using comment text boxes on Web pages might allow for quick detection of misunderstandings and for appropriate revision of instructions.

Comparative Basis

One might argue that the comparative basis (i.e., other similarly conducted studies of the same questions) for research done using the Web experimental method is too small to draw any valid conclusions from Web experiments. However, the number of Web

experiments is increasing rapidly (Musch & Reips, chap. 3, this volume). So far, the evidence for high validity of Web experiments, if compared with laboratory experiments (or, for that matter, validity of laboratory experiments, if one reverses the argument as a consequence of the discussed problems with local laboratory experiments) is overwhelmingly positive (Krantz & Dalal, chap. 2, this volume; Musch & Reips, chap. 3, this volume).

Limits (Epilogue on External Validity)

Not all experiments can be done on the Web as there are some things that cannot be done on the Web. It has not been implemented in Web experiments yet to attach sensors to people, take PET or MRI scans, take EEGs, inject drugs, feed foods to taste, touch others over the Web, or do any of a number of other things. Has it? It might just be a matter of a few weeks….

A disadvantage of Web experiments that cannot be changed, as it is built-in, is that participants will always be at computers. Consequently, computer-independent behaviors cannot yet be measured in Web experiments. However, good psychological theories should be able to predict behavior of people sitting at computers as well!

Summary

In this chapter I discussed methodological advantages and disadvantages of Web experiments, in part comparing the Web experimental method with the traditional laboratory experimental method. Advantages I presented were (1) ease of access to a large number of demographically and culturally diverse participants as well as (2) ease of access to very rare, specific participant populations; (3) a certain justification for generalization of findings in Web experiments to the general population; (4) generalizability of findings to more settings and situations, as there are reasons to believe that external validity in Web experiments is high; (5) avoidance of time constraints; (6) avoidance of organizational problems, such as scheduling difficulties, as thousands of participants may participate simultaneously; (7) completely voluntary participation; (8) ease of acquisition of just the optimal number of participants for achieving high statistical power while being able to draw meaningful conclusions from the experiment; (9) detectability of motivational confounding; (10) reduction of experimenter effects; (11) reduction of demand characteristics; (12) cost savings of lab space, person hours, equipment, administration; (13) greater openness of the research process; (14) ability to assess to the number of nonparticipants; (15) ease of comparing results with results from a locally tested sample; (16) greater external validity through greater

technical variance; (17) ease of access for participants (bringing the experiment to the participant instead of the opposite); (18) public control of ethical standards.

The presentation of disadvantages was accompanied by suggestions of possible solutions. (1) Possible multiple submissions can be avoided or controlled by collecting personal identification items, by checking internal consistency as well as date and time consistency of answers, and by using techniques such as sub-sampling, participant pools, or handing out passwords. Evidence was presented that multiple submissions are rare in Web experiments. (2) Generally, experimental control may be an issue in some experimental designs, but is less of an issue when using a between-subjects design with random distribution of participants to experimental conditions. (3) Self-selection can be controlled by using the multiple site entry technique. (4) Dropout is high in Web experiments, especially, if no financial incentives are given for participation. However, dropout can be turned into a detection device for motivational confounding. Also, dropout can be reduced by implementing a number of measures, which are listed in this chapter. (5) The reduced or absent interaction with participants during a Web experiment creates problems, if instructions are misunderstood. Possible solutions are pretests of the materials and providing the participants with the opportunity for giving feedback. (6) The comparative basis for the Web experiment method is low. This will change. (7) External validity of Web experiments may be limited by their dependence on computers and networks. Also, many studies cannot be done on the Web.

For many areas of experimental psychology, advantages of the Web method outweigh the disadvantages. As more data are collected by this method, new problems will no doubt be discovered and new solutions proposed. Nevertheless, it seems safe to venture that the Web method will soon stand as a new structure side by side with older methods in psychology's research basement, supporting our house and driving out the moisture that threatens laboratory research.

REFERENCES

Adair, J. G. (1973). *The human subject.* Boston: Little, Brown.

Anderson, C., & Bushman, B. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1,* 19-41.

American Psychological Association (1992). *Ethical principles of psychologists and code of conduct* [WWW document]. Available URL: http://www.apa.org/ethics/code.html

American Psychological Society. (1995). *Psychology experiments on the Net* [WWW document]. Available URL: http://psych.hanover.edu/APS/exponnet.html.

Barber, T. X., & Silver, J. J. (1968). Fact, fiction, and the experimenter bias effect. *Psychological Bulletin Monograph Supplement, 70,* 1-29.

Birnbaum, M. H. (1998). *Decision research center* [WWW document]. Available URL: http://psych.fullerton.edu/mbirnbaum/dec.htm

Birnbaum, M. H. (in press). Decision making in the lab and on the Web. In M. H. Birnbaum (Ed.), *Psychology experiments on the Internet.* San Diego, CA: Academic 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.

Bredenkamp, J. (1996). Grundlagen experimenteller Methoden [Foundations of experimental methods]. In E. Erdfelder, R. Mausfeld, T. Meiser, & G. Rudinger (Eds.), *Handbuch Quantitative Methoden* (pp. 37-46). Weinheim, Germany: Psychologie Verlags Union.

Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences.* New York: Holt, Rinehart & Winston.

Browndyke, J. N., Santa Maria, M., Pinkston, J., & Gouvier, W. D. (1999). Online neuropsychology project: Head injury and prevention knowledge among professionals and laypersons. Manuscript submitted for publication.

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.

Bülthoff, H., van Veen, H.-J., Givaty, G., Braun, S., & Georg, P. (1997*). Web-experiments: Psychophysics on the Internet* [WWW document]. Available URL: http://exp.kyb.tuebingen.mpg.de/Web-experiment/index.html

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171-246). Chicago: Rand McNally.

Chapanis, A. (1970). The relevance of laboratory studies to practical situations. In D. P. Schultz (Ed.), *The science of psychology: Critical reflections.* New York: Appleton Century Crofts.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press.

Coomber, R. (1997). Using the Internet for survey research. *Sociological Research Online, 2* [WWW document]. Available URL: http://www.socresonline.org.uk/socresonline/2/2/2.html

Devereux, G. (1967). *From anxiety to method in the behavioral sciences.* The Hague: Mouton.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28,* 1-11.

Föderation der Schweizer Psychologinnen und Psychologen. (1997). *Berufsordnung der Föderation der Schweizer Psychologinnen und Psychologen FSP* [Occupation principles of the Confederation of Swiss Psychologists] [WWW document]. Available URL: http://www.fsp.psy.ch/BerufsordnungD.html

Frick, A., Bächtiger, M. T., & Reips, U.-D. (1999). Financial incentives, personal information and drop-out rate in online studies. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, & A. Werner (Eds./Hrsg.), *Current Internet science – trends, techniques, results. Aktuelle Online Forschung – Trends, Techniken, Ergebnisse.* Zürich: Online Press. [WWW document]. Available URL: http://dgof.de/tband99/

FxWeb Web Technologies. (1999). *Webtracker version 2* [WWW document]. Available URL: http://www.fxWeb.com/tracker/

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, UK: Cambridge University Press.

Graphics, Visualization, & Usability Center. (1998). *GVU's tenth WWW user survey graphs* [WWW document]. Available URL: http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/graphs/graphs.html#general

Graphics, Visualization, & Usability Center. (1999). *GVU's WWW user surveys* [WWW document]. Available URL: http://www.cc.gatech.edu/gvu/user_surveys

Hänze, M., & Meyer, H. A. (1997). Feldexperimente und nicht-reaktive Messung im World Wide Web [Field experiments and nonreactive measurement on the World Wide Web]. In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott, & G. Strube (Eds.), *CAW-97. Beiträge zum Workshop "Cognition & Web"* (pp.141-148). Freiburg, Germany: IIG-Berichte 1/97.

Internet.com LCC. (1999). *One-quarter of US households online* [WWW document]. Available URL: http://www.cyberatlas.com/big_picture/demographics/household.html

Jung, J. (1969). Current practices and problems in the use of college students for psychological research. *Canadian Psychologist, 10*(3), 280-290.

Kieley, J. M. (1996). CGI scripts: Gateways to World-Wide Web power. *Behavior Research Methods, Instruments, & Computers, 28,* 165-169.

Kirk, R. E. (1995). *Experimental Design: Procedures for the behavioral sciences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Klauer, K. C., Musch, J., & Naumer, B. (1999). *On belief bias in syllogistic reasoning.* Manuscript submitted for publication.

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

Krantz, J. H., & Dalal, R. (in press). Validity of Web-based psychological research. In M. H. Birnbaum (Ed.), *Psychology Experiments on the Internet.* San Diego, CA: Academic Press.

Martin, D. W. (1996). *Doing psychology experiments* (4th ed.). Pacific Grove, CA: Brooks/Cole.

McNemar, Q. (1942). Opinion-attitude methodology. *Psychological Bulletin, 43,* 289-374.

Melungeon Heritage Association. (1999). *Melungeon Heritage Association* [WWW document]. Available URL: http://pluto.clinch.edu/appalachia/melungeon/index.html

Merton, R. K. (1973). The normative structure of science. In N. W. Storer (Ed.), *The sociology of science: Theoretical and empirical investigations* (pp. 267-278). Chicago: University of Chicago Press. (Original work published 1942)

Musch, J., & Reips, U.-D. (in press). A brief history of Web experimenting. In M. H. Birnbaum (Ed.), *Psychology experiments on the Internet.* San Diego, CA: Academic Press.

Netscape. (1999). *Netscape communicator glossary* [WWW document]. Available URL: http://home.netscape.com/communicator/glossary.html

Oakes, W. (1972). External validity and the use of real people as subjects. *American Psychologist, 27,* 959-962.

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

Pagani, D., & Lombardi, L. (1997). *Online Psychology Lab Padua* [WWW document]. Avialable URL: http://www.psy.unipd.it/personal/laboratorio/surprise/htmltesi/index.html

Parsons, H. M. (1974). What happened at Hawthorne? *Science, 183,* 922-932.

PsychExps. (1998). [WWW document]. Available URL: http://www.olemiss.edu/PsychExps/

Psycoloquy. (1999). [WWW document]. Available URL: http://www.princeton.edu/~harnad/psyc.html

Reips, U.-D. (1995a). *The Web's Experimental Psychology Lab* [WWW document]. Available URL: http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLab.html

Reips, U.-D. (1995b). *AAAbacus, the first cognitive learning experiment on the WWW* [WWW document]. Available URL: http://www.psych.unizh.ch/genpsy/Ulf/Lab/archiv/aaabacus.html

Reips, U.-D. (1995c). *The Web experiment method* [WWW document]. Available URL: http://www.psych.unizh.ch/genpsy/Ulf/Lab/WWWExpMethod.html

Reips, U.-D. (1996a, April). Experimentieren im World Wide Web [Experimenting in the World Wide Web] [Abstract]. *Experimentelle Psychologie: Tagung experimentell arbeitender Psychologen* [Proceedings of the Experimental Psychologist's Conference], *Germany, 38,* 256-257.

Reips, U.-D. (1996b, October). *Experimenting in the World Wide Web.* Paper presented at the 1996 Society for Computers in Psychology conference, Chicago. [Also as WWW document]. Available URL: http://www.psych.unizh.ch/genpsy/reips/slides/scipchicago96.html

Reips, U.-D. (1997a). Forschen im Jahr 2007: Integration von Web-Experimentieren, Online-Publizieren und Multimedia-Kommunikation [Science in the year 2007: Integration of Web experimenting, online publishing, and multimedia communication]. In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott, & G. Strube (Eds.), *CAW-97.*

Beiträge zum Workshop "Cognition & Web" (pp. 141-148). Freiburg, Germany: IIG-Berichte 1/97.

Reips, U.-D. (1997b). Das psychologische Experimentieren im Internet [Psychological experimenting on the Internet]. In B. Batinic (Ed.), *Internet für Psychologen* (pp. 245-265). Göttingen: Hogrefe.

Reips, U.-D. (1998a). Forschung in der Zukunft [Future science]. In T. Krüger & J. Funke (Eds.), *Psychologie im Internet: Ein Wegweiser für psychologisch interessierte User* (pp. 115-123). Weinheim: Beltz.

Reips, U.-D. (1998b). Web-Experiment [Web experiment]. In F. Dorsch, H. Häcker, & K.-H. Stapf (Eds.), *Psychologisches Wörterbuch* (Vol. 13, pp. 943-944). Bern: Huber.

Reips, U.-D. (1999a). Theorie und Techniken des Web-Experimentierens [Theory and techniques of Web experimenting]. In B. Batinic, A. Werner, L. Gräf, & W. Bandilla (Eds.), *Online Research: Methoden, Anwendungen und Ergebnisse*. Göttingen: Hogrefe.

Reips, U.-D. (1999b). *Online answering biases*. Manuscript submitted for publication.

Reips, U.-D. (1999c). Online-Forschung mit Kindern [Online research with children]. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, & A. Werner (Eds./Hrsg.), *Current Internet science – trends, techniques, results. Aktuelle Online Forschung – Trends, Techniken, Ergebnisse.* Zürich: Online Press. [WWW document]. Available URL: http://dgof.de/tband99/

Reips, U.-D., & Bächtiger, M.-T. (1999). *Are all flies drosophilae? Participant selection bias in psychological research*. Manuscript in preparation.

Reips, U.-D. & Musch, J. (1999). Web experiment list [WWW document]. Available URL: http://www.psych.unizh.ch/genpsy/Ulf/Lab/webexplist.html

Rheingold, H. (1993). *The virtual community: Homesteading at the electronic frontier*. [Also as WWW document, avialable URL: http://www.rheingold.com/vc/book/]. Reading, MA: Addison-Wesley.

Rice, B. (1983). The Hawthorne defect: Persistence of a flawed theory. *Psychology Today, 16,* 70-74.

Roethlisberger, F. J. (1977). *The elusive phenomena: An autobiographical account of my work in the field of organized behavior at the Harvard Business School.* Cambridge, MA: Division of Research, Graduate School of Business Administration. (Distributed by Harvard University Press)

Rosenthal, R. (1965). The volunteer subject. *Human relations, 18,* 389-406.

Rosenthal, R. (1966). *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts.

Rosenthal, R., & Fode, K. L. (1973). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science, 8,* 183-189.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development.* San Francisco, CA: Holt, Rinehart & Winston.

Rosenthal, R., & Rosnow, R. L (1969). The volunteer subject. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research.* New York: Academic Press.

Rosnow, R. L., & Rosenthal, R. (1966). Volunteer subjects and the results of opinion change studies. *Psychological Reports, 19,* 1183-1187.

Schmidt, W. C. (in press). The Server-Side of Psychology Web Experiments. In M. H. Birnbaum (Ed.), *Psychology Experiments on the Internet.* San Diego, CA: Academic Press.

Schubert, T., & Waldzus, S. (1996). *Laboratory of Social Psychology Jena* [WWW document]. Available URL: http://www.uni-jena.de/~ssw/labor.htm

Schultz, D. P. (1972). The human subject in psychological research. In C. L. Sheridan (Ed.), *Readings for experimental psychology* (pp. 263-282). New York: Holt.

Schwarz, N., & Hippler, H. (1994a). Subsequent questions may influence answers to preceding questions in mail surveys. *ZUMA-Arbeitsberichte,* 94/07.

Schwarz, N., & Hippler, H. (1994b). The numeric values of rating scales: A comparison of their impact in mail surveys and telephone interviews. *ZUMA-Arbeitsberichte,* 94/08.

Schwarz, S., & Reips, U.-D. (1999). Drop-out wegen JavaScript: "Das habe ich unmöglich wissen können" – Ein Web-Experiment zum Reversed Hindsight Bias [Drop out caused by JavaScript: "I could not have expected this to happen" – A Web experiment on the reversed hindsight bias]. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, & A. Werner (Eds./Hrsg.), *Current Internet science – trends, techniques, results. Aktuelle Online Forschung – Trends, Techniken, Ergebnisse.* Zürich: Online Press. [WWW document]. Available URL: http://dgof.de/tband99/

Smart, R. (1966). Subject selection bias in psychological research. *Canadian Psychologist, 7a,* 115-121.

Spiel, C. (1988). Experiment versus Quasiexperiment: Eine Untersuchung zur Freiwilligkeit der Teilnahme an wissenschaftlichen Studien [Experiment versus quasi-experiment:

An investigation of voluntariness of participation in scientific studies]. *Zeitschrift für experimentelle und angewandte Psychologie, 25,* 303-316.

Stones, A., & Perry, D. (1997). Survey questionnaire data on panic attacks gathered using the World Wide Web. *Depression and Anxiety, 6,* 86-87.

Welch, N. (1995). *Demonstrations in auditory perception* [WWW document]. Available URLs: http://www.music.mcgill.ca/auditory/Auditory.html and http://gl15.bio.tu-darmstadt.de/auditory/auditory.html

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

Figure 1 Number of monthly visits to the Web's Experimental Psychology Lab, September 1997 to February 1999.

Figure 2 The password technique (top) and the subsampling technique (bottom).

Figure 3 The participant pool technique.

Figure 4 The multiple site entry technique.

---

[1] Although the original "Hawthorne effect" at the Hawthorne plant may well not have been what is meant by the term (Parsons, 1974) "it is certainly possible for an experimental manipulation to cause a change in behavior independent of what the manipulation was" (Martin, 1996).

Figure 1.

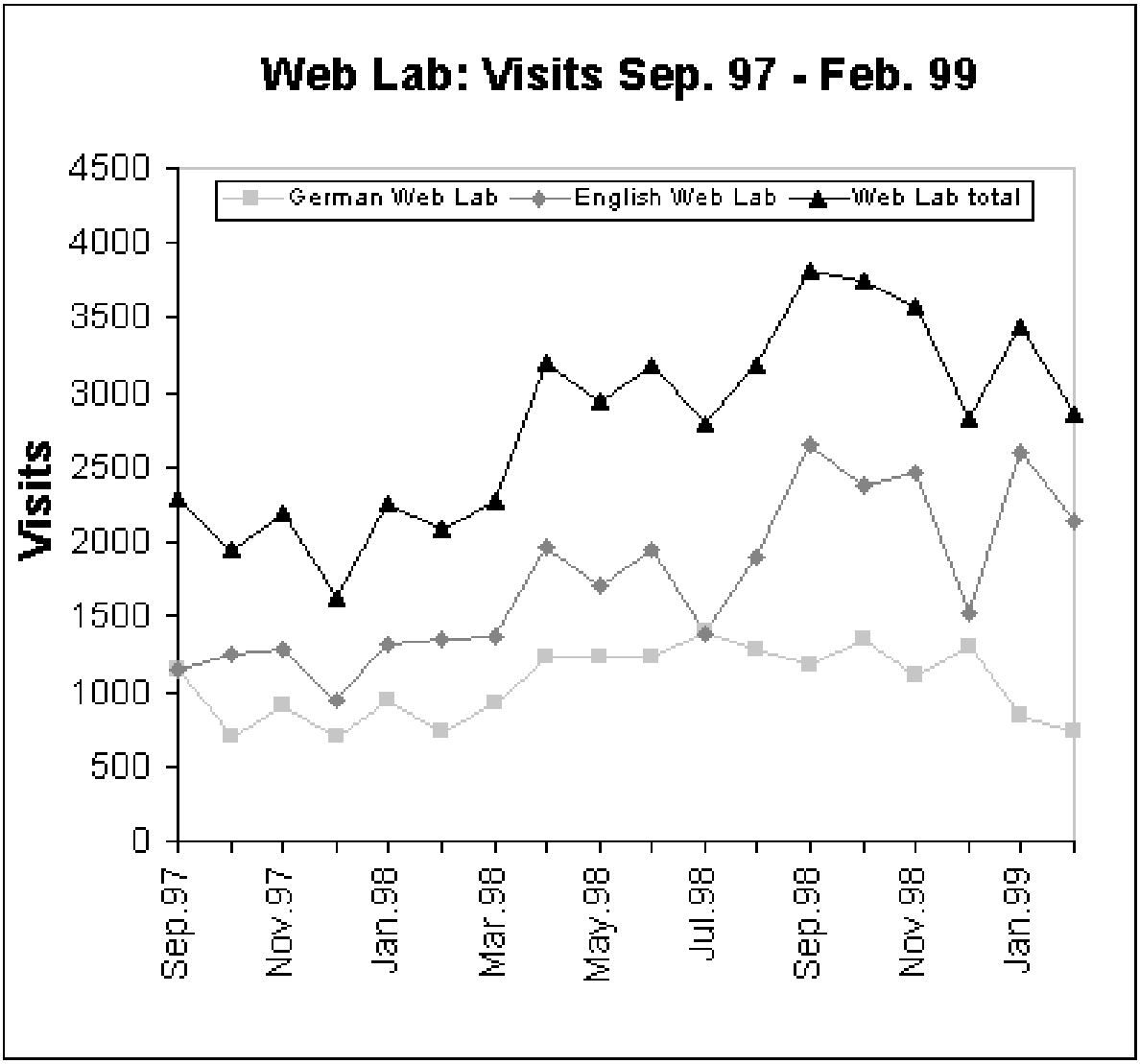

Web Lab: Visits Sep. 97 - Feb. 99
German Web Lab
English Web Lab
Web Lab total
Visits
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Sep.97
Nov.97
Jan.98
Mar.98
May.98
Jul.98
Sep.98
Nov.98
Jan.99

Figure 2.

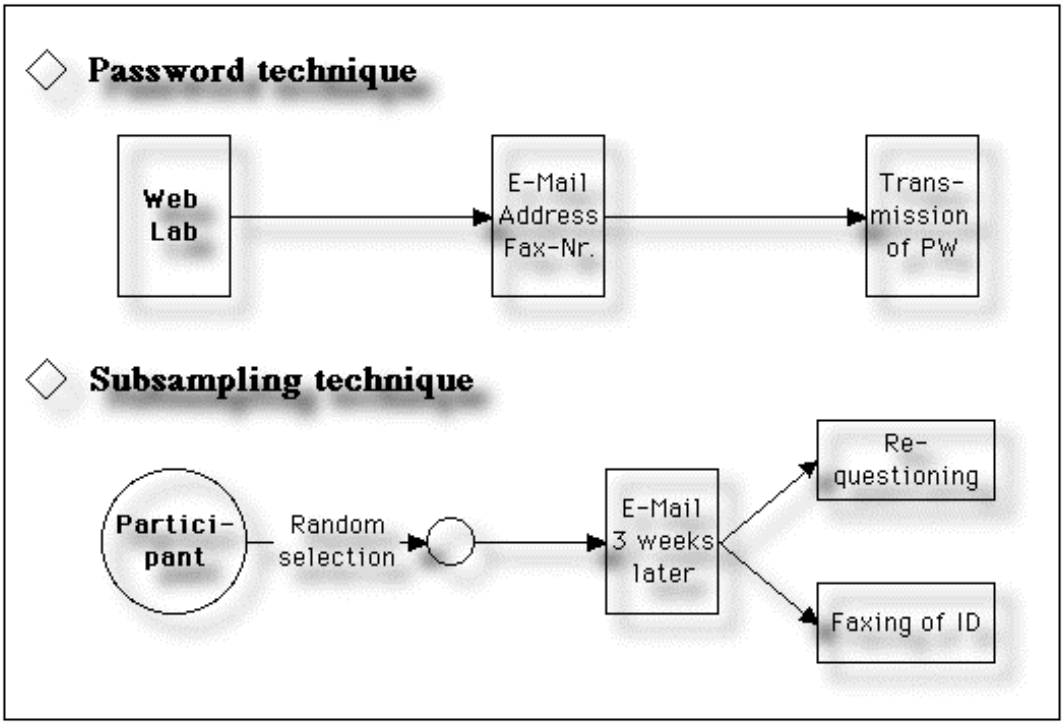

Password technique
Web Lab
E-Mail Address Fax-Nr.
Trans-mission of PW
Subsampling technique
Partici-pant
Random selection
E-Mail 3 weeks later
Re-questioning
Faxing of ID

Figure 3.

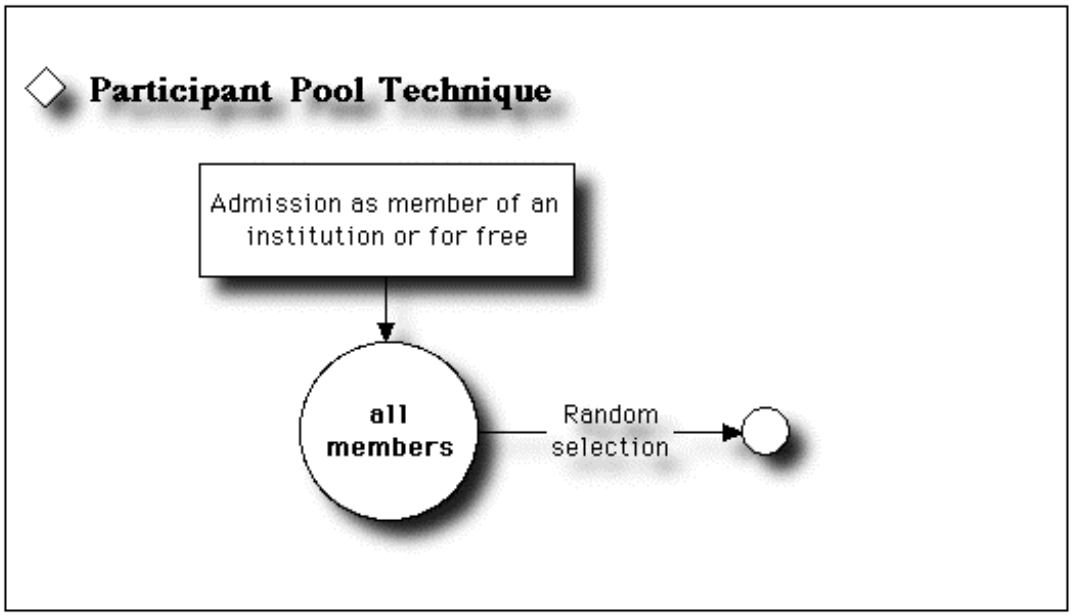
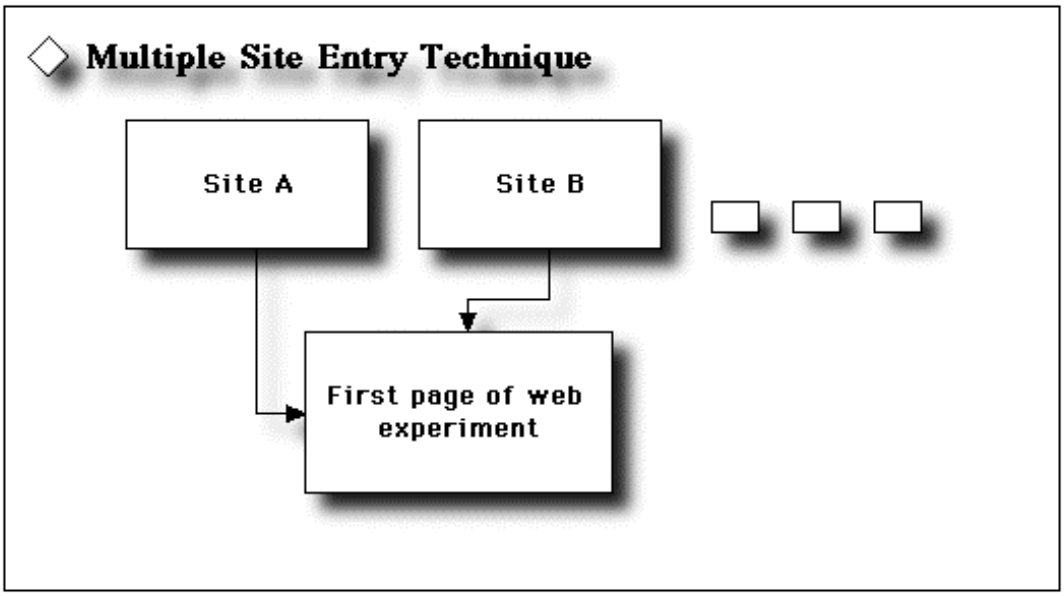

Participant Pool Technique
Admission as member of an institution or for free
all members
Random selection

Figure 4.


Multiple Site Entry Technique
Site A
Site B
First page of web experiment