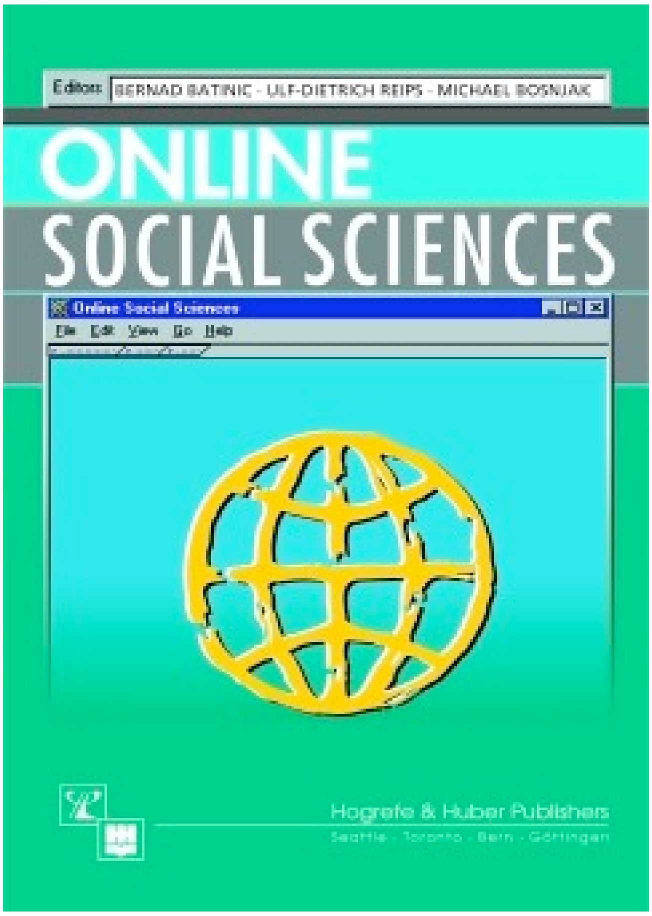
Reprint of

Reips, U.-D. (2002). Theory and techniques of Web experimenting. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online Social Sciences*. Seattle: Hogrefe & Huber.



Order at Amazon

13 Theory and Techniques of Conducting Web Experiments

Ulf-Dietrich Reips

In the Internet, there are three main possibilities for collecting data that allow us to draw conclusions on human behavior. These are:

- Nonreactive data collection,
- Online surveys, and
- Web experiments.

Web experiments, the core theme of this chapter, have an advantage over the other two methods because of their ability in proving causal relationships between variables. This advantage implicitly facilitates deductive, hypothesis testing research, in which the much disputed question of representativeness is not posed – in comparison to other study procedures. In the following sections, I will introduce a demonstration of the potential of Web experiments in comparison to the two aforementioned non-experimental online research methods. The first part of the chapter explains the historical and theoretical background for different types of research methods. The second part of the chapter is devoted to several useful techniques for Web experimenting, such as *warm-up*, *password*, *subsampling*, *participant pool*, and *multiple site entry*.

13.1 Nonexperimental Methods

13.1.1 Nonreactive Data Collection

Nonreactive data collection refers to the use and analysis of existing databases, for example server logfiles or newsgroup contributions. This method is referred to as *nonreactive* because, at the time of the investigation, the producer of the data hardly assumes that the information will be used for a study and, therefore, does not behave differently than normal in *reaction* to this fact. Indeed, at the time when the behavior traces are left even the researchers are rarely aware of this fact. Typical characteristics of nonreactive data collection include:

- Investigation of correlations, but not of causalities is possible.
- Data are not gathered by scientists.
- Modest data collection expenditures required.
- Examination of rare behavioral patterns is facilitated.
- Nonmanipulable events can be studied.
- Greater ecological validity provided, i.e., the data are polled without any outside influences ("naturally" so to speak).

An example for the use of nonreactive data is in the study of communicative behavior among members of a mailing list, with the help of interaction frequencies pertaining to e-mail headers in contributions (see Stegbauer & Rausch, this volume).

13.1.2 Online Surveys

The second, and most commonly used method of online research is the survey (see Batinic & Bosnjak, 1997). The frequent use of surveys on the Internet can be explained by the ease with which it seems surveys can be constructed, conducted, and evaluated. In most cases, the motivational factor behind surveys is determined by the potential of reaching conclusions on a greater number of persons (in this case the sampled population) than were actually interviewed (the sample). Representativeness, which is absolutely necessary for the fulfillment of this aim – can only be attained at great cost outside of the Internet. Already the time spent between the survey and its evaluation decreases the data's significance with regard to the "world's" current situation. Representativeness is most likely if each individual member of the sample is randomly selected from that set of persons for whom the results are expected to make general statements upon.

In 1993, Ross Perot, former American presidential candidate, made the painful discovery of the actual extent to which a nonrepresentative survey can bypass reality. He asked TV viewers – in a sort of fit of "nonrandomized" sampling – to fill in a postcard inserted in a magazine called "TV Guide". Ninety seven percent of the entrants answered the question on whether the government should cut spending positively. In a random sample shortly afterwards, the same question only met with 67% approval (Tanur, 1994). Inferring that almost everybody supported spending cuts would be an instructive example for a conclusion reached many years beforehand: "Direct, intuitive observation, accompanied by questioning, imagination, or creative intervention, is a limited and misleading prescientific technique." (Monte, 1975, p.63).

The supposed ease with which surveys can be conducted is, at second glance, a deceiving mirage and does not lead to the source of eternal knowledge, rather into a grinding quagmire teeming with quality control measures. For example, what must be considered when conducting surveys easily filled the 800 pages book by Lyberg, Biemer, Collins, De Leeuw, Dippo, Schwarz, & Trewin (1997). For online surveys, Gräf, Brenner, and Bandilla all describe in their contributions to the present volume how methodological quality can be guaranteed.

The third principal method of collecting data in the Internet is referred to as the *Web experiment* (Musch & Reips, 2000; Reips, 1995b, 1996b, 1998, 2000a), or Internet-based experiment. The characteristics of the experimental method will be discussed in the following section.

13.2 The Experimental Method

13.2.1 Definition and History

A social scientific experiment is a systematic method for exploring and checking causal influences on behavioral patterns of interest. Its effectiveness lies in the fact that only this method is suitable for examining *causal* explanations of behavior (Bredenkamp, 1996; Kirk, 1995; Martin, 2000; Wormser, 1974), and making statements that transcend the claim of a *correlative* connection.

When they hear the word "experiment", many people first think of physical experiments. In the tradition of natural science, experiments are characterized by the fact that the object of research seems to be governed by little natural variance. A piece of metal will always slide down a slope with the same velocity. Consistent with these findings, the natural sciences (except for biology) concentrated on the central aspects of the development of research methods (i.e., the invention of apparatuses and the limitation of measurement errors). On the other hand, it was clear to both the biological and social sciences that differences between various studied objects actually are defined by the characteristics of the object and therefore remain, even if measurement errors can be excluded. In the natural setting of a playground, one can easily observe that, in contrast to a piece of metal, children can slide down a chute differently. Very early on, the task of recording variations of physiology and behavior in a statistical distribution became a necessary component of social and biological scientific research. But how should one proceed to record the data?

Towards the end of the eighteenth century, agricultural scientists accomplished productive and pioneering achievements in the area of conducting experiments, and at the same time advanced the related field of statistics. For instance, faced with the problem that spreading manure in a field under well controlled conditions (in a physical science sense) would yield excellent results on one plot, yet cause an agricultural disaster in another plot, they came to the conclusion that certain factors which probably influenced the field's harvest should be controlled. Also, the various types of manure (i.e., the variations of the target factor) needed to be compared with each other, and the manure conditions needed to be compared with a no treatment condition to see if manure has an effect at all.

The question of if and at which values the differences found in such a comparison could be considered significant were to be answered by statistics, by then a rapidly evolving discipline.

Only recently have the nonbiological natural sciences seen it fit to employ statistic methods for the analysis of occurring variances. (In view of experimental demonstrations in lessons, generations of pupils have had a hunch that there is this necessity…).

13.2.2 Generalizability of Results According to a Hypothetical-Deductive Understanding of Science

The population of Internet users is growing at an extremely fast pace. Various user analyses show that all population parameters are converging with those of the normal population (IntelliQuest Information Group, Inc. http://www.intelliquest.com/about/release41.htm; Graphics, Visualization, & Usability Center, 1997-2001). In the foreseeable future, in "First World" countries the group of Internet users will be nearly as representative as the TV audience is for the general population. This means that, with the expanded spread of Internet access, Web experiments will make experimental research as accessible to just as many participants from the general population as to the "usual subjects" contacted in university cafeterias. Because so few studies exist with confirmatory evidence on other population groups but college students, the limited range in the test participants' characteristics can, from a general point of view, be regarded as a detriment for the significance of experimental research in the social sciences. Nevertheless, representativeness is of relatively little importance in experimental research, because the (inductive) generalization of results plays a minor role in the (deductive) investigation of hypotheses[1]. Results from

[1] Here, a discussion of "representativeness" in the gir-l mailing list (Archive Mailing-List German Internet Research Mailing List gir-l, 1997), and in particular Jochen Musch's contribution dated 17.10.97, is to be mentioned.

survey research conducted online, on the other hand, depend on where and how in the Internet the participants are contacted[2] (Brenner, this volume).

Deduction is the method whereby a conclusion drawn from a general case is applied to a specific case. In hypothetical-deductive research, a verifiable causal hypothesis (symbolized by a small circle in Figure 13.1) is derived from a theory (large circle in Figure 13.1). The hypothesis is then tested in an experiment. In other words, it is compared to data obtained in an empirical test. If the results of the experiment do not repudiate the hypothesis, the theory is (once again) considered confirmed and useful. *Induction* is the process whereby a conclusion drawn from a specific case is applied to the general. An observation leads to a hypothesis around which, following an inductive train of thought, a theory is built. This need not have anything to do with reality apart from the initial observation. The theory does not undergo a *test* (see Figure 13.1). Typical inductive practice is the application of data obtained from a sample to the population of which the sample was extracted from. The actual research procedure as a whole is constantly on the move back and forth between deduction and induction.

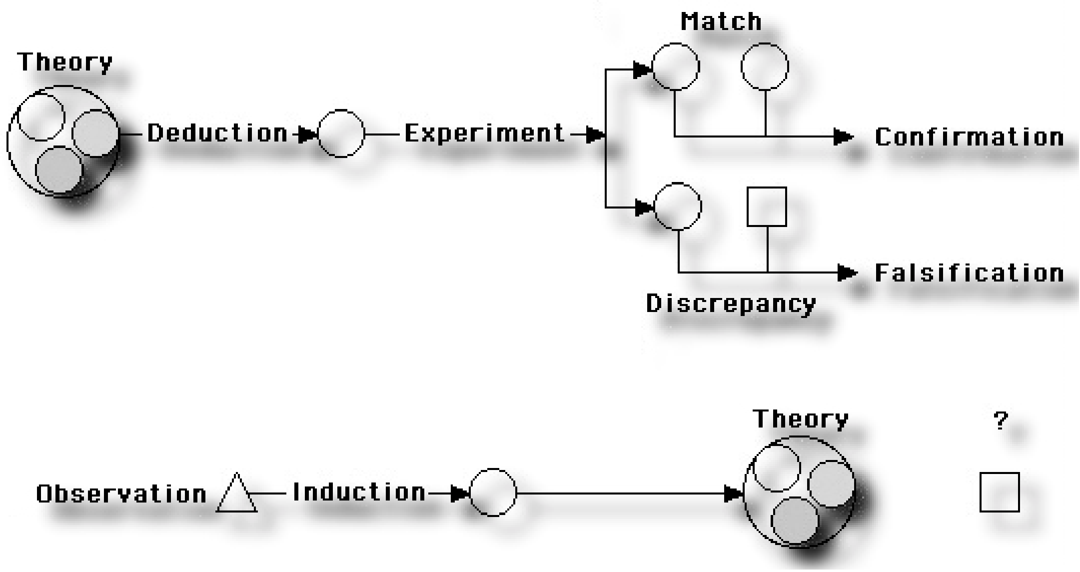


Figure 13.1: Hypothetical-deductive research (top) allows an experimental test of hypotheses, whereas a purely inductive procedure (bottom) does not. Explanation of symbols in text.

[2] We recently conducted a Web experiment involving an IQ test with speed components. Resulting distributions were clearly different for two samples gained by placing links on two distinctive Websites.

13.2.3 Characteristics of Experiments

A defining characteristic of an experiment is the *active creation* of specific situations. In an experiment, a situation is actively and methodically produced in which the event to be investigated – and if possible only this event – can unfold in its entirety. Active creation deems the bothersome wait for a spontaneous occurrence of the event of interest unnecessary. Also, it facilitates the preparation of optimal, controlled circumstances that can in no way interfere with the proper interpretation of the situation. The aim of the created situation could be a decision, for instance, if two theories predict different results for one experiment (*experimentum cruxis*). In the simplest case, it could also be the finding of one answer to the question of how manipulation of a variable affects another variable (*explorative experiment*). In most cases, experiments are conducted on the basis of a hypothesis (i.e., in order to investigate theoretically founded hypotheses).

A second important characteristic of the experiment is its *repeatability*, an issue logically related to its active creation characteristic. Because experiments are repeatable, their results can be studied from an inter-subjective perspective.

The third characteristic of experiments is their *variability*. Because the experimenter controls the circumstances, and because experiments are repeatable, she can introduce completely new variables or alter existing variables in further experiments. By varying the variables, through isolation and combination, the effects of independent variables on the dependent variable(s) can be confirmed and quantified.

The fourth characteristic of experiments is the allocation of participants to different experimental conditions[3]. In a typical experiment the participants are allocated one or more variations (*levels*) of the independent variable(s) to be tested against each other. The participants' behavior is then measured on one or more dependent variables.

13.2.4 Experimental Logic

All scientists – and all of us in daily life – try to find out something about relationships between events. In the social sciences, human behavior is permanently in the focus of interest.

In almost every form of examination, even we behavioral scientists encounter problems in generalizing the results of our research when we wish to measure the influence of clearly defined, physically measurable circumstances and their effects on equally clearly defined behavior. Imagine a reaction time experiment in which the

[3] Whether this characteristic is of importance is debatable. Tests without this characteristic can be referred to as "incomplete experiments".

participants have to depress a button as quickly as possible after hearing a high, medium, or low frequency sound. An example: The head of a company which develops security systems for atomic power plants is interested in your work and asks you, "Excuse me, I hear you conducted a reaction time experiment. Can you tell me what frequency warning sirens should have so that the security personnel will react within a second?" Because you precisely controlled many different variables in the experiment, you answer, "If you can guarantee that your security personnel is male, 30 years old, mildly extravert, and does not suffer from hearing defects, has a slightly below average IQ of 95, and is sitting in a 4 x 4 x 3 meter, soundproof room at a temperature of 21° Celsius, has nothing else to do and can hear the alarm bells (that are always introduced by a prior warning tone) on a Taka 450xm stereo headphone… – maybe then I could answer your question."

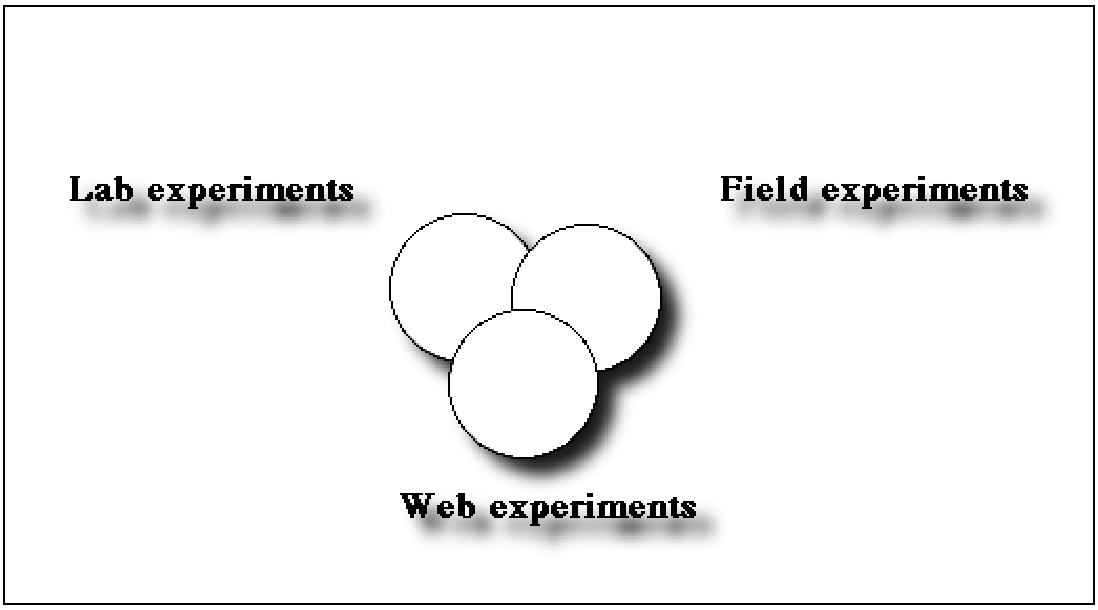


Figure 13.2: "Discovery spotlights" on the *phenomenological space*, as produced by different forms of experiments

In order to avoid research results from being constrained in their significance to an individual combination of situational characteristics, one should avoid *controlling* or *keeping constant* all variables in any given experiment. For, "the more highly controlled the experiment, the less generally applicable the results" (Martin, 2000, p.28). Random variations of the greatest possible amount of variables facilitate generalizability. When participants with whatever traits are randomly allocated to the different conditions of an experiment it cannot be ruled out, but is statistically improbable, that a characteristic manipulated by the experimental conditions systematically influences the differences in the participants' behavior with respect to the dependent variable. A variable that *systematically* varies with the variation of the dependent variable is known as a *confounded* variable.

In contrast to other methods, the characteristics of an experiment (e.g., control groups, randomized allocation of conditions to participants, counterbalancing) allow an examination of causal explanations (for example, assumed reasons for product preferences in shopping behavior).

13.2.5 Types of Experiments

There are three kinds of experiment: laboratory experiments, field experiments, and – more recently – Web experiments (Figure 13.2 symbolizes the three types of experiment as overlapping spotlights on the *phenomenological space*). *Laboratory experiments* are experiments conducted in the traditional sense of the word, for example in a university laboratory. Here, the researchers have the advantage of being able to control many factors precisely. However, the artificial setting can cause the participants to behave in a different manner than usual. For this reason, many researchers tend to go "where life takes place" and conduct experiments in everyday environments. Experiments such as these are known as *field experiments* because of the "naturalness" of such environments. Understandably, anything can cross the researcher's path when the potential for control over the situation is reduced. In consideration, the already classic pioneering study by Muzafer and Carolyn Sherif, who in 1953 conducted an impressive field experiment on group dynamics in a holiday camp, seems all the more remarkable (Sherif & Sherif, 1954). Unfortunately, field experiments are more cost intensive, a reason why they are rarely conducted. Another disadvantage of conducting field experiments is the ethical problem of not informing participants about their involvement in an experiment, because they do not know that they are participating in an experiment. Participants may leave the "action radius", unknowingly creating a methodological problem.

Web experiments (Reips, 1996a, 1996b, 1998, 2001a; see Figure 13.3) are the logical extension of the computer-based laboratory experiment. With the help of a Web browser, the participant can access the laboratory computer – which now is called a *Web server* – from her own computer that may be at a far distant location. In many respects, the experiment is conducted as if the participant were seated in front of a computer in a lab – with the only exception that the experimental information, normally visible on the laboratory computer's monitor, is now transferable to the participant's screen anywhere in the world. Any type of input by the participant, for instance mouse clicks or movements, sound and video signals, text entries or hypertext document requests can be registered by the Web server and answered accordingly. Response times, the name and location of the participant's computer, and the type of browser used are recorded with other data in a logfile. This logfile can be formatted and filtered according to the requirements of the statistical analysis.

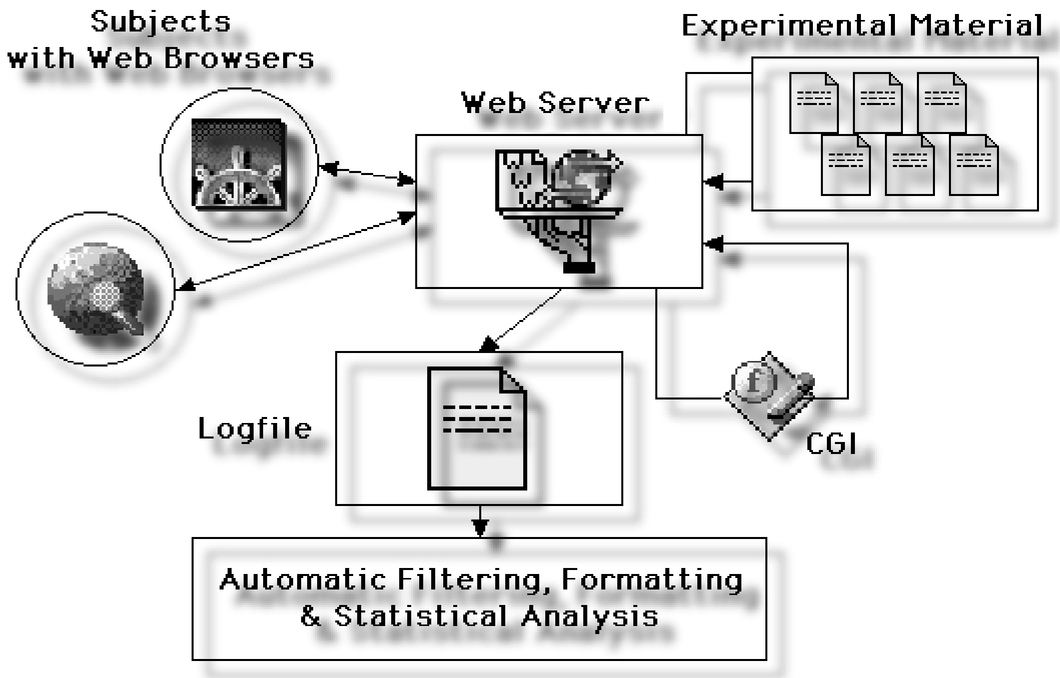


Figure 13.3: Components of a web experimental setup: Web server, web browser, experimental materials to be sent out via Internet, Common Gateway Interface (CGI) scripts, logfile

In contrast to the disadvantages of the lab experimental method, Web experiments offer many advantages:

- Easy access to a large and geographically diverse pool of potential participants, including possible participants from very specific and previously inaccessible target groups.
- The experiment comes to the participant instead of vice versa.
- Ease of acquisition of just the optimal number of participants allows for achieving high statistical power while being able to draw meaningful conclusions from the experiment[4].

[4] Contrary to a widespread belief a study is not generally better the more participants take part in it. At large sample sizes even the tiniest effects become *statistically* "significant", even though they might not be significant in terms of effect size.

- Reduced cost, because neither a laboratory, nor an experimenter assistant are required (More extensive lists of arguments in favor of Web experiments can be found in Reips (1995b, 2000a, 2000b).

First analyses and studies reveal that Web experiments:

- Usually are valid (Krantz, Ballard, & Scher, 1997; Krantz & Dalal, 2000) and sometimes even generate higher quality data than laboratory studies (Birnbaum, 2001; Reips, 2000a),
- facilitate research in previously inaccessible areas (e.g., Bordia, 1996; Coomber, 1997), and
- on the whole, more than even out the scientific and practical disadvantages (Reips, 1997, 2000a).

Figure 13.4: The web experimental psychology lab's logo

The great freedom for participants of being able to begin and discontinue participation in an openly accessible online study at any time has consequences for researchers. On account of this, *self selection* is a problem of great importance. Mostly, questionnaires and experiments on controversial topics are affected. For instance, if one were to post a Web page on the Internet with the title "To take part in an opinion poll on abortion click here" one can rest assured that one will only receive opinions from people generally or specifically interested in the topic of abortions. Before conducting a Web experiment one should consider carefully the likelihood of self-selection and take necessary precautions. These should include factorial design of experiments with variations of the factors, suitable test topics and replications using other samples, which might exist due to other Web connections to an experiment. Also, the multiple site entry technique described later in this chapter can be used to control for biases due to self-selection.

In the summer of 1995, the *Web Experimental Psychology Lab* (Reips, 1995a, 2001a, 2001b) was set up at the University of Tübingen for the purpose of conducting Web experiments (Figure 13.4 shows the Web lab's logo). In 1998, the laboratory moved its physical base to the University of Zürich. At last count (November 2000), the laboratory registered an average of around 4,000 visits per month (for historical numbers see Figure 13.5). Since 1996, a host of further experimental laboratories have been established on the World Wide Web (WWW) (listing in chronological order):

- Interactive CyberLab for Decision-Making Research (http://www.etl.go.jp/~e6930) [April, 1996].
- Laboratory of Social Psychology Jena (http://www.uni-jena.de/~ssw/labor.htm) [June, 1996].
- Experimental Server Trier (http://cogpsy.uni-trier.de:8000/TEServ-e.html) [June, 1997].
- Max-Planck Institute for Biological Cybernetics Tübingen (http://exp.kyb.tuebingen.mpg.de/Web-experiment/index.html) [November, 1997].
- Online Psychology Lab Padua (http://www.psy.unipd.it/personal/laboratorio/surprise/htmltesi/index.html) [May, 1997].
- Decision Research Center (http://psych.fullerton.edu/mbirnbaum/dec.htm) [started online experiments in March, 1998].
- ZUMA Online Research (http://www.or.zuma-mannheim.de/inhalt/onlinelabor/) [May, 1998].
- Psycholinguist Laboratory Scotland (http://surf.to/experiments) [September, 1998].
- PsychExps (http://www.olemiss.edu/PsychExps/) [Fall 1998; invites participation of Web experiments from other researchers].

Further Web experiments can be accessed on the Web experiment list (http://www.genpsy.unizh.ch/Ulf/Lab/webexplist.html), on the classic Web site by the American Psychological Society (http://psych.hanover.edu/APS/exponnet.html), and on William E. Snell Jr.'s Web site to accompany his introductory psychology class on "Psychological Perspectives on Human Behavior" (http://psychology2.semo.edu/websites/web41.htm). Guidelines for conducting experiments in the Web Experimental Psychology Lab can be found on the Web site and in Reips (2000a)[5]. In principle, the Web laboratory is open to cooperative Web experiments.

To complete the picture, the possibility of conducting *undercover* experiments on the WWW (Hänze & Meyer, 1997) should be mentioned. In a series of Web experiments, Hänze and Meyer manipulated images and background colors on the Web pages of a popular icon archive. Hence, for color they were able to replicate the *mere exposure effect* (increasing preferences for often perceived and encountered stimuli) postulated by Kunst-Wilson and Zajonc (1980).

[5] Conditions for the use of the Web lab are listed at http://www.genpsy.unizh.ch/Ulf/Lab/WebLabCond.html

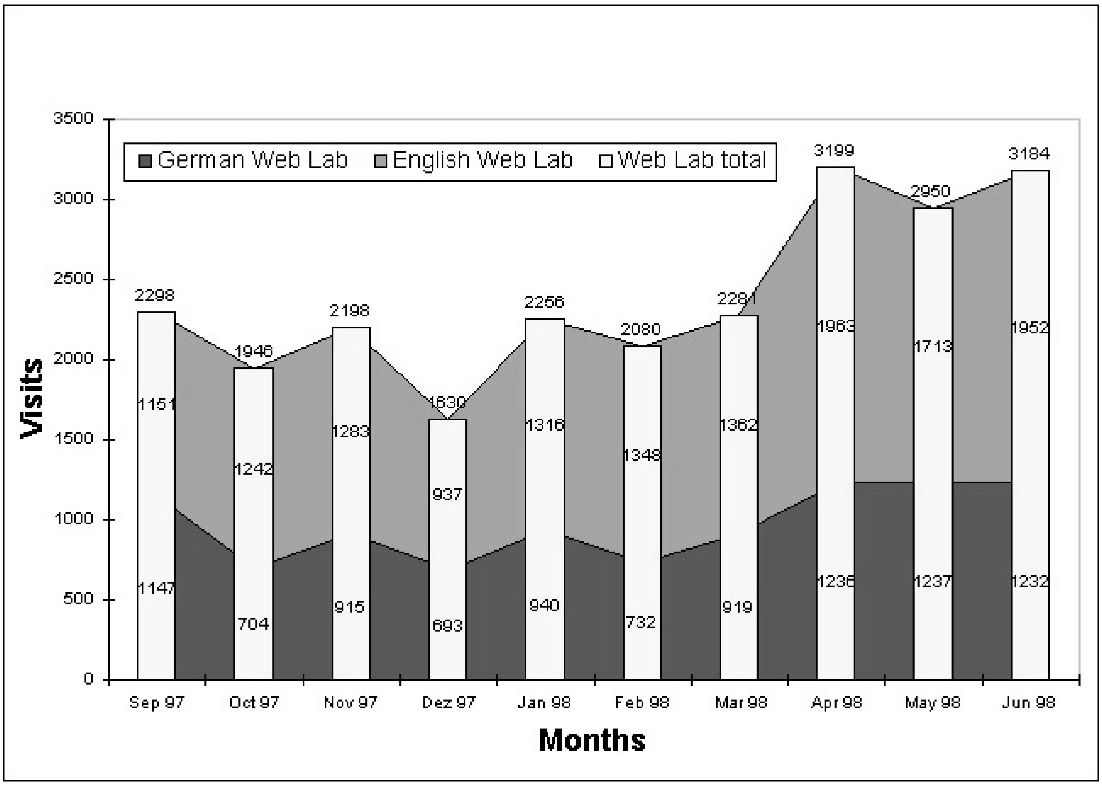


Figure 13.5: Visits to the Tübingen version of the web experimental psychology lab, September 1997 to June 1998

13.3 Motivational Techniques When Conducting Web Experiments

13.3.1 Recruitment and Motivation

Finding and motivating clients and, as in this case, experimental participants to stay on a Web site is a general point of discussion in online business. As will be demonstrated, a few particular characteristics must be regarded when recruiting and encouraging participants to take part in online research projects. To begin with, the Web site should be reasonably attractive to look at. This can be achieved by:

- Incorporating appealing design (e.g., graphics),
- external links to the Web site (registration in search engines),
- comments stating that this part of the Web might be interesting to explore (prize draws, links to press reports, display of awards),

- multilingualism, and
- offering various viewing possibilities (frame/no-frame/text/print versions).

The integrity of the request should be underlined by the corresponding explanation:

- The name(s) of the main institution(s),
- clarification of the scientific value of the project,
- indication of an e-mail contact address, and
- guarantee of data confidentiality (and compliance with this guarantee!).

Participants should be informed of how much time will be required for taking part in the Web experiment. Offering small rewards for taking part and drawing (cash) prizes amongst respondents can be of help. Certainly, an explanation of one's own status in the study as well as receiving information about aim and background of the study both count as further incentives for participating (see Bosnjak & Batinic, in this volume).

13.3.2 Ensuring Incentives for Staying (Dropout Reduction)

Engaging participants for an experiment is one matter, keeping them engaged is another. From a methodological point of view, a low dropout quota is important for a Web experiment. This applies even more to nonexperimental surveys, because a comparison rarely takes place between the different variations of an independent variable. However, whilst the significance of a questionnaire might depend upon the fact that, for example, those who show little interest in the survey will to a greater extent stop answering the questionnaire, selective dropout by participants under one of many different circumstances could be used as a detection device for confounded variables in the Web experiment. This way, potential artifacts could be managed better in the future.

Many Web surfers leave Web sites early because loading times for documents, in particular images, are too long. A proven method of confronting this problem is the *systematic reduction in loading times* the further a respondent advances in the experiment, also called the *High Hurdle Technique*. For instance, most experiments in the Web Experimental Psychology Lab are structured in such a way that the *Web Laboratory page* (http://www.genpsy.unizh.ch/Ulf/Lab/WebExpPsyLab.html) requires the most loading time. The *General Instructions* pages for the various experiments is the second slowest page to load. The *Specific Instructions* pages appear even quicker and the *Experiment Materials* nearly all use text and images that had been used previously, and therefore will be loaded "in no time" from the hard drive's cache. Whoever decides to drop out because of long loading times will be more likely to do this sooner rather than later.

A similar method is advisable when confronting "psychological loading time", (i.e., the "internal barriers" which precede participation). Immediately on the General Instructions page, the participant could be informed of her obligations when deciding to participate, and should be asked to disclose her telephone number or e-mail address. The respondent should also (perhaps with a little test) be notified of the requirement of a certain type of Web browser and should be asked to download the correct software. Furthermore, one should – also for ethical reasons – notify the participant about time needed and other potentially unpleasant aspects of participating in the experiment. The largest hurdle should be straight at the start.

Experienced Web surfers are aware of the fact that they are a "snooping breed". The suggestive power of the term "Web browser" rather implies a sort of drifting or "leafing through" than a planned analysis of contents. What to do with the "snoops"? In order to separate those who are simply peeking from those who are earnestly intent on participating in a Web experiment, the *Warm Up Technique* is a tried method. If, in a Web experiment, the actual experimental manipulation only occurs after a warming up phase, one can deduct that those who ended early did not break off the experiment because of the manipulation.

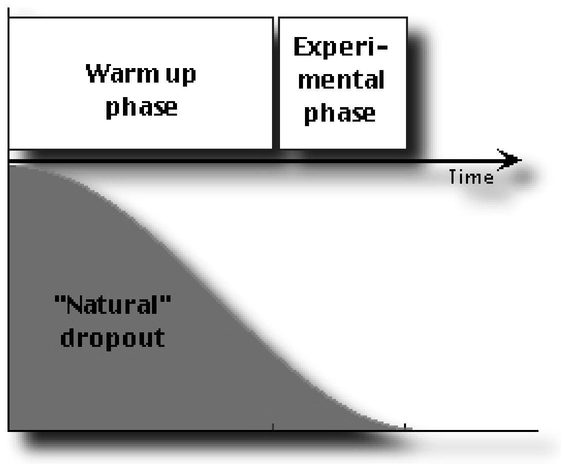


Figure 13.6: The warm up technique separates "snooper" dropout from dropout in reaction to a Web experiment

Revealing to the participant what awaits him will reduce the likelihood of his leaving the experiment early because of an unpleasant surprise. It is therefore advisable to briefly outline the experiment's contents, e.g., "Instructions – Phase 1: A questionnaire with three questions – Phase 2 – two questionnaires with five questions each. Total time: approximately 25 minutes." One could also notify the participant of his progress during the course of the experiment. If a participant knows how long it will take approximately she is more likely to complete the experiment.

Whether announcements, instructions or questionnaires are used – any text should be written in a positive, personal style. The writing most likely will succeed if the participant is imagined as a partner in a cooperative project (who he/she is anyway.)

13.4 Techniques in Practice

Nearly all of the techniques presented in this chapter were put into practice in the first between subjects experiment conducted on the WWW, which was started in 1995 (Musch & Reips, 2000; Reips, 1996a, 1996b). At http://www.genpsylab.unizh.ch/archiv/first/WWWExpE.html, where this Web experiment is documented, one can put oneself in the position of a participant and experience *the influence of causal mechanism information during acquisition of causal knowledge*. Of a total of 880 participants who took part in this Web experiment with a duration of up to 60 minutes between 18 January and 16 September 1996, 255 dropped out of the experiment during the warming-up phase. During this phase, an additional 29 persons fell victim to technical problems caused on the server side. In total, therefore, 596 participants started the actual experiment. During this phase the dropout ratio added up to less than ten percent (58), with the exception of ten technical faults. "Dropout" during the experimental phase referred of course to the fact that the participants did not complete any of the questionnaires. A subset of the 58 persons could have been "insiders" (other scientists interested in finding out about the method) or "questionnaire shy" people.

Of the remaining 528 participants, these included 25 who refused to reveal their telephone number and/or e-mail address and whose data were excluded as a result. Also excluded were four more sets of data produced within two days after prior participation from the same or only marginally different computer addresses despite different e-mail addresses. The probability that these might have been different people is nevertheless large, since it is likely that people tell their colleagues about the Web experiment, who then also participate from the same office building. Yet, to be on the safe side, all eventualities (e.g., deceptive attempts) must be considered and prevention of data contamination should be a goal with highest priority. No event was registered whereby one person signed up under the same e-mail address or telephone number more than once. To conclude, data from 499 experiment participants on the WWW were included in the statistical analysis.

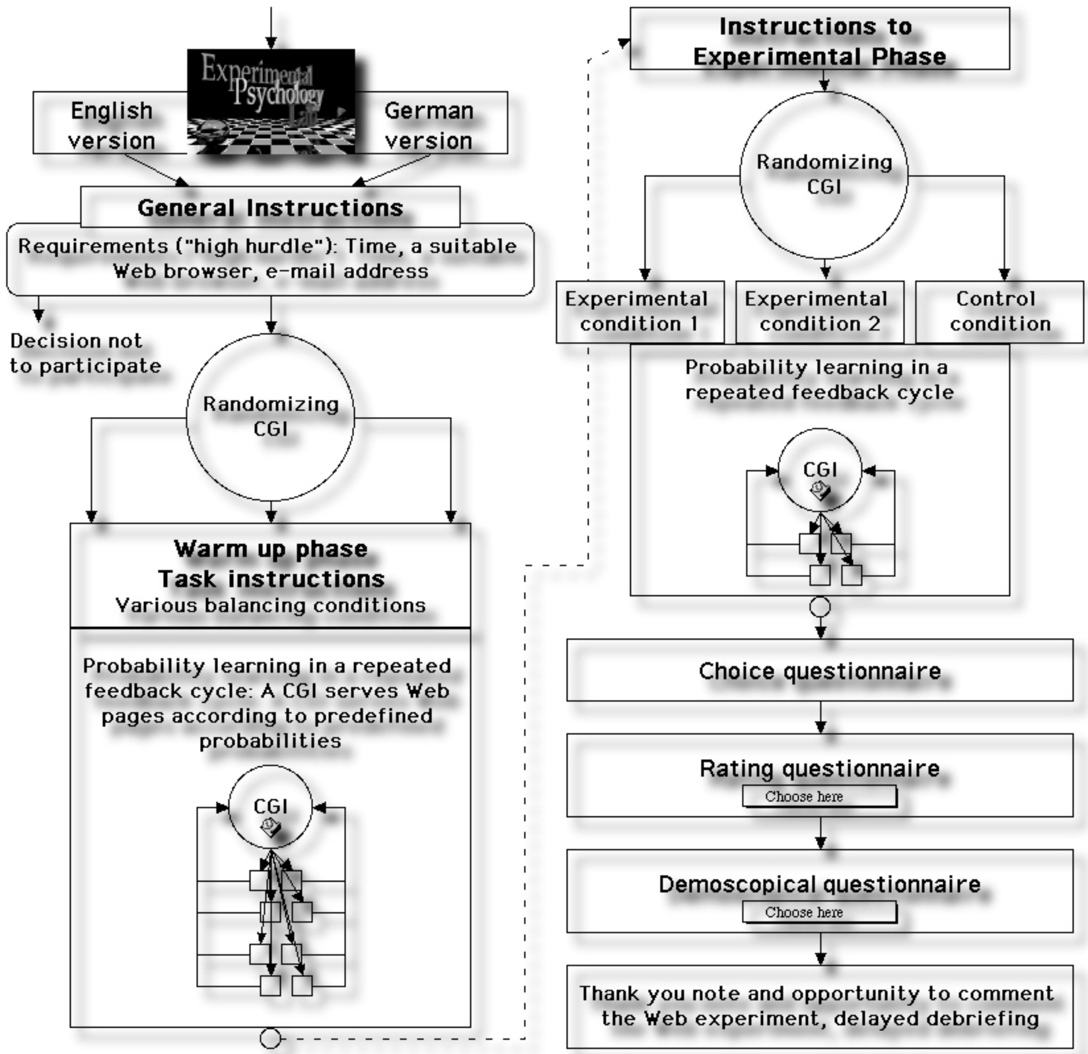


Figure 13.7: Procedure flowchart of the web experiment on the influence of causal mechanism information in the acquisition of causal knowledge (Reips, 1996a, 1996b, 2000b)

13.5 Control Techniques for Conducting Web Experiments

13.5.1 Avoiding Data Garbage

At the end of the previous section, several techniques were presented which help avoid collecting unusable data. One need not only check for double e-mail addresses or IP-addresses, one could also notify participants in advance. One can rest assured that the majority of the WWW population is not as versed in the medium's technology as to recognize the barriers of inspection of multiple participants. In general, one should be able to count on the goodwill of the participants – previous experience has shown that only very rarely will somebody disregard the request to take part only once (Reips, 1996b, 2000a). To a greater extent, this rule applies to experiments in which participation involves a certain amount of expenditure, because of its length and because of the "high hurdles" for the participant described above. It is equally supportive if the Web site is able to indicate that a genuine contractor or an institution of benefit to the public is involved in the research.

But what should be done with the enthusiastic participants, those that desperately want to see the whole thing again? Offer them a channeling possibility, the option of participating again. This could be implemented via a hyperlink or with a question at the beginning. One should generally try to compile verifiable information with the questions. Other questions can serve as indicators of accuracy and coherence: anyone claiming to be 18 years old and a grandmother at the same time should be handled with caution.

```
<SCRIPT language="JavaScript">
function browsertest (){
document.write("Congratulations, your browser has passed the JavaScript test!!!") }
</SCRIPT>
<FONT size=7>JavaScript Test</FONT>
<SCRIPT LANGUAGE="JavaScript">
{ browsertest(); }
</SCRIPT>
```

Figure 13.8: A JavaScript that tests a web browser's JavaScript compatibility

Merely from a technical point of view, data garbage can be avoided by testing if the participant's Web browser supports all necessary features such as scroll-down menus and JavaScript (see Figure 13.8). If not, the participant should be referred to a source where the adequate Web browser can be downloaded. With the help of JavaScript one can also detect the participant's set monitor resolution.

13.5.2 Special Techniques

With the aid of certain tricks one can attain more control. For instance, with the installation of *password* technology, truly interested participants can be traced (see Figure 13.9, top). With this procedure, people interested in participating have to register their e-mail and snail-mail addresses and, if necessary, their fax and telephone numbers. Right away, or as soon as a sufficient amount of registrations has been compiled, the users interested in taking part are sent their login and password, although the distribution medium can be altered. (This variation also facilitates a later comparison of those groups that received their password via e-mail, those that received it in the mail and those that had their passwords faxed). Participants can be contacted reliably in the case of further questions. As a result, any participation in the experiment can be attributed to different people.

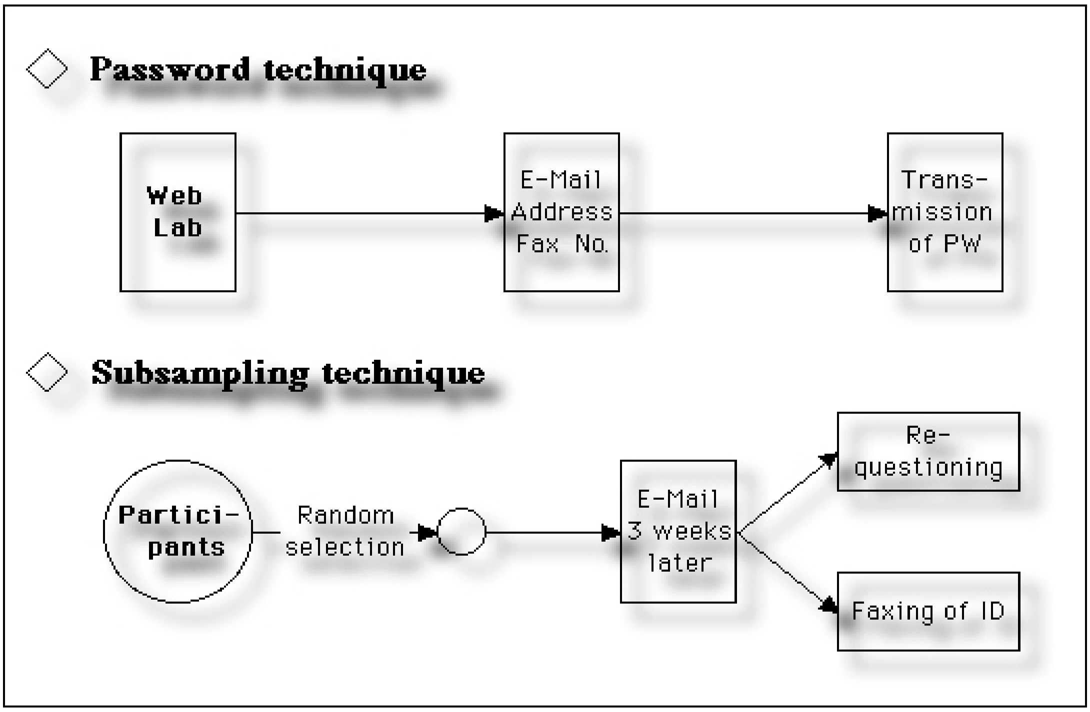


Figure 13.9: Password technique (top) and subsampling technique (bottom)

Another method is a random extraction of a *sub sample* from the total set of all test participants (see Figure 13.9, bottom). The members of this sub sample receive an e-mail that they only need to reply to with an empty e-mail. Firstly, one obtains an estimated value of the proportion of wrong address entries. Secondly, for the participant the expenditure of sending an empty reply is relatively minor, so that one can count on a large response rate. In a further posting the participants are asked to answer certain control questions with reference to the study. One can also request that the people in the sub sample fax back their ID card, and might then contact them repeatedly and rigorously until it is entirely clear who participated only once, and who participated in the experiment on numerous occasions.

Instead of extracting the sub sample after the Web experiment, one can also ensure from the onset that information on the test participants is available. This can be obtained via the *Participant Pool Technique* (see Figure 13.10). Members of a participant pool (also sometimes referred to as an “online panel”, e.g., Maruccia, 1999, and similar to a “closed pool” approach, Voracek, Stieger, & Gindl, 2001) must sign an agreement stating their availability to take part in Web experiments or surveys. They enter their demographic data, which of course is processed together with possible test results and experimental data under a strict guarantee of confidentiality. Accordingly, one can reward participants for their participation in the studies. Participant pools also have the advantage of disclosing other studies a participant has taken part in. Additionally, a scientist can also use the data compiled on participants for “ordering” specifically selected persons with particular characteristics (e.g., a ratio of 50% men and 50% women). The participant pool technique (with formation of the pool on the Internet) is an ideal opportunity for researchers in smaller research facilities who have little access to participants and who, despite their use of the Internet, wish to exercise a modicum of control over whom takes part in their studies.

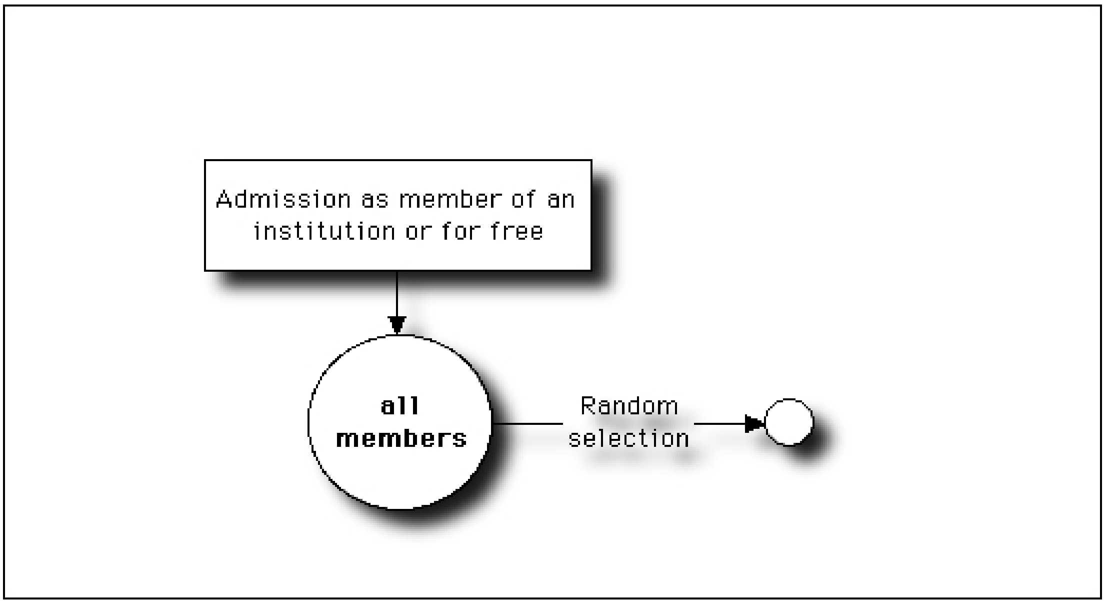


Figure 13.10: The participant pool technique

In the evaluation of data, techniques can also be used to ensure the quality of the data. One example was already mentioned above: only allow and evaluate first participation from one computer address (IP). Also, relative document download times, which are recorded by most Web server programs, can be used as a data filter. In a learning experiment, for example, one would not want to include the data of those participants who took a break during the test. This break can be defined precisely as a certain percentage increase in the average interval time between two learning stages.

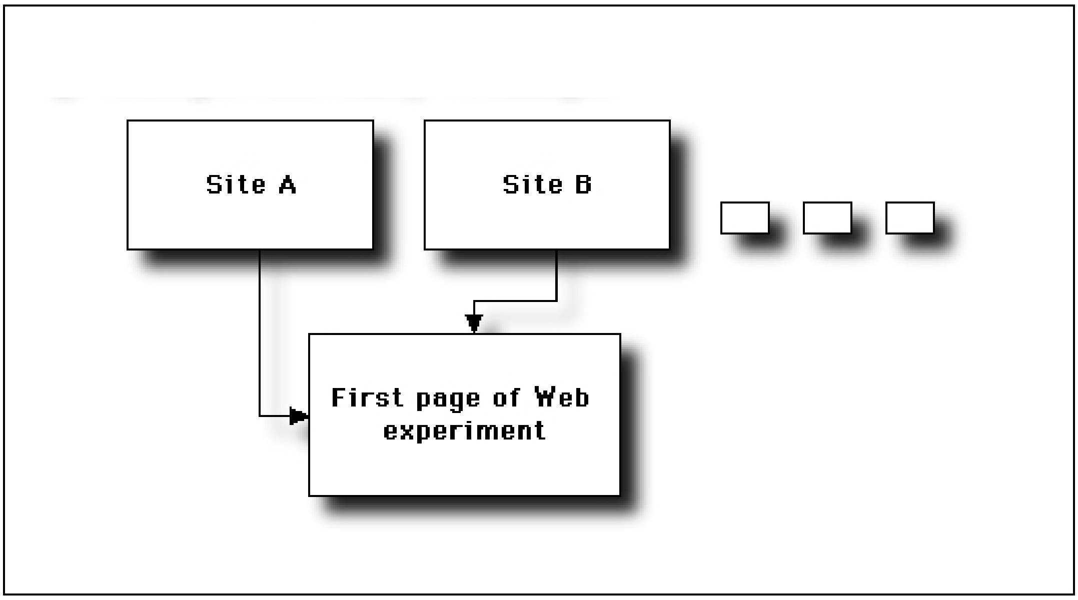


Figure 13.11: The multiple site entry technique

13.6 Epilogue

Web experiments are the method whereby the experimental method is transferred into the virtual world of the WWW. Thanks to the logic of experiments, causal relationships between variables can be examined without being hampered by the quest for representativeness in Internet samples that might be creating a stigma of researchers wandering naïvely into unfamiliar terrain. The WWW is suitable as a space for conducting experiments. The method also holds the promise of working in previously inaccessible areas (for instance with particular groups of people). Should the reader be interested in conducting a Web experiment, then come and visit one of the online laboratories, for example the Web Experimental Psychology Lab, Zürich: http://www.genpsy.unizh.ch/Ulf/Lab/WebExpPsyLab.html.

References

Batinic, B. & Bosnjak, M. (2000). Fragebogenuntersuchungen im Internet [Internet surveys]. In B. Batinic (Ed.), *Internet für Psychologen* (2nd ed.) (pp. 287-317) [Internet for psychologists]. Göttingen: Hogrefe.

Birnbaum, M. H. (2001). Online research and decision making. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 23-55). Lengerich: Pabst Science Publishers.

Bordia, P. (1996). Studying verbal interaction on the Internet: The case of rumor transmission research. *Behavior Research Methods, Instruments, & Computers, 28*, 149-151.

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: Psychologie Verlags Union.

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

Graphics, Visualization, & Usability Center (1997-2000). *GVU's www user surveys*. [Online]. Availble: http://www.cc.gatech.edu/gvu/user_surveys.

Hänze, M. & Meyer, H. A. (1997). Feldexperimente und nicht-reaktive Datenerhebung im World Wide Web [Field experiments and nonreactive data collection on the world wide web]. In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott, & G. Strube (Eds.), *CAW-97: cognition & web* (pp.79-92). Freiburg: IIG-Berichte.

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

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

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

Kunst-Wilson, W. R. & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557-558.

Lyberg, L., Biemer, P., Collins, M., De Leeuw, E., Dippo, C., Schwarz, N., & Trewin, D. (Eds.). (1997). *Survey measurement and process quality.* New York: Wiley.

Martin, D. W. (2000). *Doing psychology experiments* (5th. ed.). Belmont: Wadsworth.

Maruccia, F. (1999). Online Panels – ein Methodentest [Online panels – a test of method]. In U.-D. Reips, B. Batinic, W. Bandilla, M. Bosnjak, L. Gräf, K. Moser, & A. Werner (Eds.), *Current Internet Science – Trends, Techniques, Results. Aktuelle Online Forschung – Trends, Techniken, Ergebnisse*. [Online]. Available: http://dgof.de/tband99/.

Monte, C. F. (1975). *Psychology's scientific endeavor.* New York: Praeger.

Musch, J. & Reips, U.-D. (2000). A brief history of web experimenting. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 61-85). San Diego, CA: Academic Press.

Reips, U.-D. (1995a). *The web's experimental psychology lab.* [Online]. Available: http://www.genpsy.unizh.ch/Ulf/Lab/WebExpPsyLab.html.

Reips, U.-D. (1995b). *The web experiment method.* [Online]. Available: http://www.genpsy.unizh.ch/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, 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. [Online]. Available: http://www.genpsy.unizh.ch/reips/slides/scipschicago96.html.

Reips, U.-D. (1997). 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: Cognition & web* (pp.141-148). Freiburg: IIG-Berichte.

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

Reips, U.-D. (2000a). The web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-114). San Diego, CA: Academic Press.

Reips, U.-D. (2000b). Das psychologische Experimentieren im Internet [Psychological experimenting on the Internet]. In B. Batinic (Ed.), *Internet für Psychologen* (2nd ed.) (pp. 319-343) [Internet for psychologists]. Göttingen, Germany: Hogrefe.

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

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

Repräsentativität. Thread in mailing list (1997). *Archiv Mailingliste German Internet Research List gir-l.* [Online]. Available: http://infosoc.uni-koeln.de/archives/gir-l/msg00253.html ff.

Schmidt, W. C. (1997). World wide web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers, 29,* 274-279.

Sherif, M. & Sherif, C. W. (1954). *Groups in harmony and ten*sion. New York: Harper and Row.

Tanur, J. M. (1994). The trustworthiness of survey research. *Chronicle of Higher Education, 40,* B1-B3.

Voracek, M., Stieger, S., & Gindl, A. (2001). Online replication of results from evolutionary psychology research: Sex differences in sexual jealousy in imagined scenarios of sexual vs. emotional mate infidelity. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 91-112). Lengerich: Pabst Science Publishers.

Wormser, R. (1974). *Experimentelle Psychologie*. München: Reinhardt.