

Theorie und Techniken des Web-Experimentierens

Ulf-Dietrich Reips

Es gibt drei prinzipielle Möglichkeiten, Daten im Internet zu sammeln, die Schlüsse über menschliches Verhalten erlauben. Es sind:

- nichtreaktive Datenerhebung,
- Befragungen und
- Web-Experimente.

Web-Experimente, denen dieses Kapitel gewidmet ist, bieten gegenüber den beiden anderen Methoden den Vorteil, daß man mit ihnen kausale Zusammenhänge überprüfen kann. Dieser Vorteil erlaubt implizit ein deduktives, hypothesenprüfendes Forschen, bei dem sich die bezüglich anderer Datenerhebungsverfahren vieldiskutierte Frage der Repräsentativität der Ergebnisse nicht stellt.

Um besser verdeutlichen zu können, was Web-Experimente im Vergleich zu den beiden nichtexperimentellen Online-Forschungsmethoden leisten können, sollen diese hier zunächst kurz dargestellt werden.

Nichtexperimentelle Methoden

Nichtreaktive Datenerhebung

Die *nichtreaktive Datenerhebung* bedeutet die Nutzung bestehender Datenbestände, wie zum Beispiel Server-Log-Dateien oder Newsgroup-Beiträge. *Nichtreaktiv* ist diese Methode, weil die Produzenten der Daten im Moment ihrer Produktion nicht oder kaum vermuten, daß diese Daten einmal für eine Untersuchung genutzt werden – ja, meist wissen es zu diesem Zeitpunkt selbst die Untersucher noch nicht – und deshalb nicht in *Reaktion* auf diese Tatsache anders handeln als sonst. Typische Kennzeichen der nichtreaktiven Datenerhebung:

- Erlaubt das Überprüfen von korrelativen, aber nicht von kausalen Zusammenhängen.
- Die Daten werden oft nicht von Wissenschaftlern gesammelt.
- geringer Aufwand bei der Datenerhebung
- Erlaubt die Untersuchung selten auftretenden Verhaltens.
- Nicht herstellbare Ereignisse können studiert werden.

- Hohe ökologische Validität, d.h. die Daten sind ohne Beeinflussung von außen (sozusagen ganz “natürlich”) entstanden.

Ein Beispiel für die Nutzung nichtreaktiver Daten ist die Untersuchung von Kommunikationsverhalten anhand der mittels der Mailingheader erhobenen Interaktionshäufigkeiten in einer Mailing-Liste (Stegbauer & Rausch, in diesem Band).

Befragung

Die zweite und bei weitem am häufigsten genutzte Methode der Online-Forschung ist die *Befragung* (vgl. z.B. Batinic & Bosnjak 1997). Die Häufigkeit von Umfragen im Netz erklärt sich daraus, daß es recht einfach erscheint, Umfragematerialien zu erstellen und eine Umfrage dann durchzuführen und auszuwerten. Meist steht hinter Umfragen der Wunsch, aus den Antworten Aussagen über eine größere Menge von Personen (in diesem Fall dann die Grundgesamtheit) zu machen als nur über die Personen, die befragt wurden (die Stichprobe). Repräsentativität, die für die vollständige Erfüllung dieses Wunsches eine absolute Notwendigkeit ist, ist allerdings schon bei Umfragen außerhalb des Internet nur mit großem Aufwand näherungsweise zu erreichen. Bereits die Zeit, die zwischen Befragung und Auswertung verstreicht, verringert die Aussagekraft der Daten in bezug auf den aktuellen Zustand “der Welt”. Repräsentativität ist dann am wahrscheinlichsten, wenn jede einzelne Person der Stichprobe per Zufall aus der Menge der Personen ausgewählt wird, auf die man die Untersuchungsergebnisse verallgemeinern möchte.

Wie sehr eine nicht repräsentative Umfrage an der Realität vorbeigehen kann, mußte beispielsweise Ross Perot, der ehemalige amerikanische Präsidentschaftskandidat erfahren. Im Jahr 1993 bat er – sozusagen in einem Anfall nichtrandomisierter “Stichprobenziehung” – Fernsehzuschauer darum, eine in der Zeitschrift “TV Guide” eingeklebte Umfrage-Postkarte einzuschicken. 97% der Einsender beantworteten die auf dieser Postkarte gestellte Frage, ob die Regierung größere Sparmaßnahmen einleiten solle, positiv. Die gleiche Frage bekam in einer zufällig gezogenen Stichprobe nur 67% Zustimmung (Tanur 1994). Ein lehrreiches Beispiel für ein bereits einige Jahre vorher gezogenes Fazit: “Direct, intuitive observation, accompanied by questioning, imagination, or creative intervention, is a limited and misleading prescientific technique.” (Monte 1975, S. 63).

Die scheinbare Leichtigkeit, mit der sich eine Umfrage durchführen läßt, erweist sich bei genauerem Hinsehen also als eine trügerische Fata Morgana, die nicht der Quell schnell sprudelnden Wissens, sondern eine wüste Ödnis staubtrockener Plackerei mit Qualitätssicherungsmaßnahmen ist. Was man bei der Durchführung von Umfragen berücksichtigen sollte, ist beispielsweise in dem 800-seitigen Buch von Lyberg et al. (1997) nachzulesen. Gräf (1997) und die Beiträge von Bandilla und Hauptmanns in diesem Band zeigen, wie die methodische Qualität speziell bei Online-Umfragen sichergestellt werden kann.

Die dritte prinzipielle Methode der Datenerhebung im Internet ist das *Web-Experiment*, also das Experiment im WWW. Wodurch ein Experiment gekennzeichnet ist, wird im folgenden Abschnitt erläutert.

Die experimentelle Methode

Begriff und Geschichte

Ein sozialwissenschaftliches Experiment ist eine systematische Methode, ursächliche Einflüsse bestimmter Umstände auf ein interessierendes Verhalten zu entdecken und zu überprüfen. Die besondere Stärke dieser Methode liegt darin, daß nur mit ihr *kausale* Erklärungen von Verhalten überprüft werden können (Wormser 1974), also Aussagen, die über die Behauptung eines *korrelativen* (ungerichteten) Zusammenhangs hinausgehen.

Viele Leute denken bei dem Wort "Experiment" zunächst an physikalische Experimente. In der naturwissenschaftlichen Tradition sind Experimente dadurch gekennzeichnet, daß der Forschungsgegenstand wenig natürlicher Varianz unterworfen zu sein scheint. Ein Metallstück rutscht überall mit der gleichen Geschwindigkeit eine Schräge hinab. Folgerichtig stand in den nichtbiologischen Wissenschaften im Zentrum der Weiterentwicklung von Forschungsmethoden stets die Erfindung von Apparaten und die Eindämmung von Meßfehlern. In den biologischen Wissenschaften und den Sozialwissenschaften war es hingegen von Anfang an offensichtlich, daß Unterschiede zwischen verschiedenen Untersuchungsobjekten Eigenschaften der Objekte selbst sind und also auch unter Ausschluß aller Meßfehler bestehen bleiben. In dem natürlichen Setting "Spielplatz" ist leicht zu beobachten, daß Kinder im Gegensatz zu Metallstücken eine Rutsche äußerst verschieden hinabgleiten können. Die Variation von Physiologie und Verhalten in Form statistischer Verteilungen zu erfassen, wurde so sehr früh ein notwendiger Bestandteil der Forschung in den sozialen und biologischen Wissenschaften. Aber wie sollte man vorgehen, um sie zu erfassen?

Ertragreiche Pionierarbeit auf dem Gebiet des Experimentierens und der damit verbundenen Entwicklung der Statistik leisteten seit dem ausgehenden 18. Jahrhundert die Agrarwissenschaftler. Sie standen vor dem Problem, daß etwa das Ausbringen eines Düngers auf einem Feld auch unter im physikalisch-naturwissenschaftlichen Sinne hochkontrollierten Bedingungen sehr gute Ergebnisse liefern konnte, die gleiche Behandlung auf einem anderen Feld aber eher ein landwirtschaftliches Desaster reflektierte. Es zeigte sich bald, daß erstens bestimmte Faktoren, die wahrscheinlich einen Einfluß auf den Ertrag eines Feldes haben, kontrolliert werden müssen und zweitens die verschiedenen Düngerformen (also die Varianten des interessierenden Faktors) miteinander und – drittens – mit einer Nichtbehandlung verglichen werden müssen. Die Frage, ob und ab welchem Wert ein in einem solchen Vergleich gefundener Unterschied als bedeutsam erachtet werden sollte, hatte die sich als Folge dieser Erkenntnisse rapide entwickelnde Statistik zu beantworten.

Erst in jüngerer Zeit sehen sich auch die nichtbiologischen Naturwissenschaften gezwungen, statistische Methoden zu benutzen, mit denen sich die auftretenden Varianzen analysieren lassen. (Generationen von Schülern hatten angesichts experimenteller Unterrichtsdemonstrationen eine eindringliche Ahnung, daß diese Notwendigkeit besteht).

Repräsentativität im hypothetico-deduktiven Wissenschaftsverständnis

Die Population der Internetbenutzer wächst extrem schnell. Verschiedene Nutzeranalysen zeigen, daß alle Populationsparameter mit denen der allgemeinen Bevölkerung konvergieren (IntelliQuest Information Group, Inc. <<http://www.intelliquest.com/about/release41.htm>>; Graphics, Visualization, & Usability Center 1997). In absehbarer Zeit wird die Gruppe der Internet-Benutzer wohl so repräsentativ für die Allgemeinpopulation sein wie derzeit die Gruppe der Fernsehkonsumenten. Dies bedeutet, daß Web-Experimente mit der weiteren Verbreitung von Internet-Anschlüssen sehr viele Personen der Allgemeinpopulation der Experimentalforschung genauso zugänglich machen wie die "üblichen Subjekte" aus Universitäts-Cafeterien. Insgesamt betrachtet ist die geringe Streubreite in den Merkmalen der Versuchspersonen zwar ein Manko für die Aussagekraft experimenteller sozialwissenschaftlicher Forschung, weil es zu wenige verbindende Untersuchungen gibt, die bestätigende Ergebnisse mit anderen Bevölkerungsgruppen gefunden haben. Trotzdem ist Repräsentativität für die experimentelle Forschung relativ unerheblich, denn die (induktivistische) Generalisierbarkeit von Ergebnissen spielt beim (deduktivistischen) Überprüfen von Hypothesen keine Rolle.¹

Deduktion ist der Schluß vom allgemeinen auf den spezielleren Fall. Beim hypothetico-deduktiven Forschen wird zunächst aus einer Theorie (in Abbildung 1 symbolisiert durch einen großen Kreis) eine überprüfbare Kausalhypothese (kleiner Kreis) abgeleitet. Dann wird diese Hypothese anhand eines Experiments geprüft, d.h. den in einem empirischen Test gefundenen Daten gegenübergestellt. Wenn die Ergebnisse des Experiments nicht gegen die Hypothese sprechen, dann gilt die Theorie (ein weiteres Mal) als gestützt. *Induktion* ist der Schluß vom Speziellen auf das Allgemeine. Eine Beobachtung führt zu einer Hypothese, um die dann beim rein induktivistischen Vorgehen eine Theorie gebaut wird, die mit der Realität außerhalb der Ausgangsbeobachtung nichts zu tun haben muß. Es kommt nicht zu einem *Test* der Theorie (vgl. Abbildung 1). Auch die Verallgemeinerung der aus einer Stichprobe gewonnenen Eindrücke auf die Population, aus der die Stichprobe gezogen wurde, ist ein typisches induktives Vorgehen. Der tatsächliche Forschungsprozeß als Ganzer bewegt sich stets im Wechselspiel zwischen Deduktion und Induktion.

¹ Hier sei auf einen Thread mit dem Thema "Repräsentativität" in der gir-I-Mailing-Liste (Archiv Mailing-Liste German Internet Research List gir-I, 1997) verwiesen, und dort speziell auf den Beitrag von Jochen Musch vom 17.10.1997.

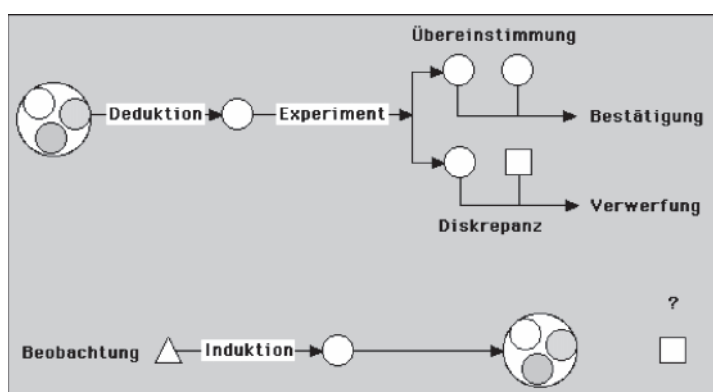


Abbildung 1: Hypothetico-deduktives Forschen (oben) erlaubt den experimentellen Test von Hypothesen, rein induktives Vorgehen (unten) nicht. Erläuterungen der Symbole im Text.

Kennzeichen des Experiments

Ein definierendes Kennzeichen des Experiments ist die *Herstellung* besonderer Umstände. Im Experiment wird aktiv und planmäßig eine Situation geschaffen, in der sich der Vorgang, der untersucht wird – und möglichst nur dieser – optimal entfalten kann. Die Herstellung erspart lästiges Warten auf spontanes Stattfinden des interessierenden Vorgangs, und es erlaubt die Vorbereitung günstiger, kontrollierter Verhältnisse, in denen fast nichts die klare Interpretation des Sachverhalts stören kann. Das Ziel der so hergestellten Situation kann eine Entscheidung sein, etwa, wenn aufgrund zweier Theorien unterschiedliche Ergebnisse eines Experiments vorhergesagt werden (*experimentum cruxis*). Im einfachsten Fall kann es aber auch die Beantwortung der Frage sein, wie sich die Manipulation einer Variablen auf eine andere Variable auswirkt (*exploratives Experiment*). Meist werden Experimente jedoch hypothesengeleitet eingesetzt, d.h. zur Überprüfung theoretisch fundierter Hypothesen (s.o.).

Ein zweites wichtiges Kennzeichen des Experiments ist seine *Wiederholbarkeit*, die logisch aus der planmäßigen Herstellung folgt. Weil Experimente wiederholbar sind, sind ihre Ergebnisse intersubjektiv überprüfbar.

Drittes Charakteristikum des Experiments ist seine *Variierbarkeit*. Weil eine Experimentator die Bedingungen beherrscht und Experimente wiederholbar sind, kann er in weiteren Experimenten unabhängige Variablen verändern oder ganz neu einführen. Durch Variationen per Isolation und Kombination der Variablen lassen sich die Wirkungen der unabhängigen Variablen auf die abhängige(n) Variable(n) bestimmen und gegebenenfalls auch quantifizieren.

Viertes Kennzeichen eines Experiments ist die Zuteilung der Versuchsteilnehmer auf verschiedene experimentelle Bedingungen.² In einem typischen Experiment wer-

² Ob dieses Kennzeichen wesentlich ist, ist umstritten. Versuche ohne dieses Kennzeichen können als *unvollständige Experimente* bezeichnet werden.

den die Versuchspersonen zufällig einer oder mehreren gegeneinander zu testenden Ausprägungen der unabhängigen Variablen zugewiesen. Das Verhalten der Versuchspersonen wird dann bezüglich einer oder mehrerer abhängiger Variablen gemessen.

Die experimentelle Logik

Alle Wissenschaftstreibenden – und wir alle im Alltag – versuchen, etwas über Beziehungen zwischen Ereignissen herauszufinden. In den Sozialwissenschaften ist dabei immer menschliches Verhalten eines der interessierenden Ereignisse.

Bei jeder Untersuchungsform haben wir Verhaltenswissenschaftler selbst dann Probleme, die Ergebnisse unserer Forschung zu verallgemeinern, wenn wir den Einfluß eindeutig definierter, physikalisch meßbarer Ereignisse auf ein ebenso klares Verhalten messen wollen. Stellen Sie sich vor, Sie machen ein Reaktionszeitexperiment, bei dem die Versuchsteilnehmer möglichst schnell einen Knopf drücken sollen, nachdem sie einen Ton niedriger, mittlerer oder hoher Frequenz hören. Die Chefin einer Firma, die Sicherheitssysteme für Kernkraftwerke entwickelt, interessiert sich für Ihre Arbeit und fragt Sie: “Sagen Sie, wie ich höre, haben Sie ein akustisches Reaktionszeitexperiment gemacht. Können Sie mir sagen, welche Frequenz Warntöne haben sollten, damit das Wachpersonal bei einem Alarm innerhalb einer halben Sekunde reagiert?” Da Sie in Ihrem Experiment viele Variablen genau kontrolliert haben, antworten Sie: “Wenn Sie garantieren können, daß das Wachpersonal männlich, 30 Jahre alt, mäßig extravertiert und frei von Hörschädigungen ist, einen leicht unterdurchschnittlichen IQ von 95 hat, in einem auf 21° Celsius temperierten 4 x 4 x 3 m großen störungsgeräuschfreien Raum sitzt, sonst nichts zu tun hat und die Alarmtöne, die jedesmal eine Sekunde vorher durch ein Warnsignal angekündigt sein müssen, mit einem Taka 450xm Stereo-Kopfhörer hört ... – dann kann ich Ihnen vielleicht eine Antwort geben.”

Um zu vermeiden, daß Forschungsergebnisse in ihrer Aussagekraft auf solch eine einzige Kombination von Umständen beschränkt bleiben, sollte man nicht alle Variablen in einem Experiment *kontrollieren* beziehungsweise *konstant halten*. Denn je kontrollierter ein Experiment ist, umso weniger generalisierbar sind die Ergebnisse (Martin 1996, S. 7). Besser für die Verallgemeinerbarkeit ist es, möglichst viele Umstände zufällig variieren zu lassen. Wenn Versuchspersonen mit was immer für Eigenschaften zufällig den verschiedenen Versuchsbedingungen eines Experiments zugewiesen werden, dann kann es zwar nicht ausgeschlossen werden, ist aber statistisch unwahrscheinlich, daß eine andere als die durch die Versuchsbedingungen manipulierte Eigenschaft systematisch die Unterschiede im Verhalten der Versuchspersonen bezüglich der abhängigen Variablen beeinflusst. Eine Variable, die *systematisch* mit den Ausprägungen der unabhängigen Variable variiert, nennt man eine *konfundierte* Variable.

Die Eigenschaften des Experiments (z.B. Kontrollgruppen, randomisierte Bedingungs- und Teilnehmerzuordnung, Balancierung) erlauben es im Gegensatz zu allen anderen Methoden, kausale Erklärungen (beispielsweise angenommene Gründe für Produktpräferenzen im Kaufverhalten) zu überprüfen.

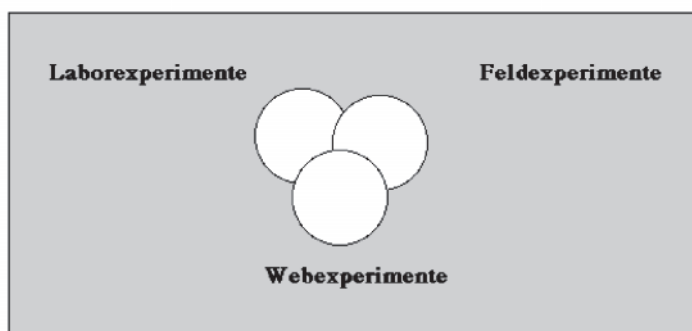


Abbildung 2: Formen des Experiments

Spielarten des Experiments

Es gibt drei Formen des Experiments: Laborexperimente, Feldexperimente und – seit kurzem – Web-Experimente (vgl. Abbildung 2). Laborexperimente sind Experimente, die ganz traditionell zum Beispiel in einem Universitätslabor durchgeführt werden. Dort haben die Forscher die besten Möglichkeiten, alle Faktoren genau zu kontrollieren. Die dadurch manchmal entstehende, zur Sterilität neigende, Atmosphäre kann aber dazu führen, daß die Versuchsteilnehmer sich in dieser für sie fremden Umgebung ganz anders verhalten, als sie es sonst tun würden. Deshalb gehen manche Forscher dorthin, “wo das Leben stattfindet” und versuchen, in relativ alltäglichen Umgebungen Experimente zu realisieren. Wegen der “Natürlichkeit” der Umgebung heißen solche Versuche *Feldexperimente*. Verständlicherweise kann einem wegen der fehlenden Kontrollmöglichkeiten des Labors dabei alles mögliche in die Quere kommen, besonders wenn man ein paar Kindergruppen durch die Gegend scheucht. Um so bewundernswerter ist die schon klassische Pionierarbeit von Muzafer und Carolyn Sherif, die bereits 1953 in einem Ferienlager ein eindrucksvolles Feldexperiment über Gruppenprozesse durchführten (Sherif & Sherif 1954). Leider sind Feldexperimente meist sehr kostenintensiv, weshalb sie selten durchgeführt werden. Ebenfalls ein Nachteil von Feldexperimenten ist das ethische Problem, daß man die Teilnehmer in der Regel hinterher nicht über das Experiment aufklären kann, weil sie gar nicht wissen, daß sie an einem Experiment teilnehmen und sich deshalb vom “Ort der Handlung” entfernen.

Web-Experimente (Reips 1996a, 1996b, 1997b; vgl. Abbildung 3) sind die konsequente Erweiterung von Laborexperimenten mit Schreibtischcomputern: eine Versuchsperson nimmt zur Teilnahme von ihrem Computer mit Hilfe eines Web-Browsers Kontakt auf zum Laborcomputer, auf dem ein Web-Server läuft. Das Experiment spielt sich in vieler Hinsicht genauso ab, als würde die Versuchsperson vor einem Computer im Labor sitzen – nur daß das Experimentalmaterial, das sonst auf dem Bildschirm des Laborcomputers gezeigt würde, statt dessen auf den Bildschirm der Versuchsperson ir-

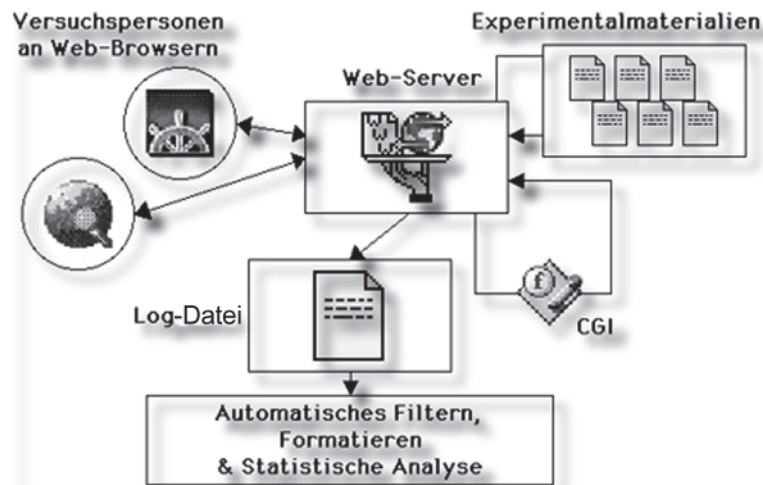


Abbildung 3: Komponenten des Web-Experimentierens: Web-Server, Web-Browser, über das Internet verschicktes Experimentalmaterial, CGI-Programme, Log-Datei

gendwo auf der Welt transferiert wird. Jeglicher Input der Versuchsperson, etwa Mausklicks und -bewegungen, Ton- und Videosignale, Texteingabe oder Dokumentabfragen, kann vom Web-Server aufgezeichnet und auf vorbestimmte Weise beantwortet werden. Antwortzeiten, Name und Standort des Computers der Versuchsperson und Art des verwendeten Web-Browsers werden zusammen mit weiteren Angaben in eine Log-Datei geschrieben. Diese Log-Datei kann dann gefiltert und so umformatiert werden, daß sie einer statistischen Analyse unterzogen werden kann.

Web-Experimente bieten – in Abhebung von den Nachteilen der laborexperimentellen Methode – einige Vorteile:

- Leichte Erreichbarkeit eines zahlenmäßig und geographisch fast nicht begrenzten Versuchspersonenpools, in dem auch Versuchspersonen aus sehr spezifischen und bisher nicht zugänglichen Zielgruppen sind.
- Das Experiment kommt räumlich zur Versuchsperson anstatt umgekehrt.
- Die wegen einer optimal großen Versuchspersonenzahl hohe statistische Power erlaubt das Beibehalten eines konventionellen α -Niveaus.
- Reduzierte Kosten, da weder Versuchsräume noch anwesende Experimentatoren gebraucht werden.³

Erste Analysen und vergleichende Studien zeigen, daß Web-Experimente

- sehr valide sein können (Krantz, Ballard & Scher 1996),
- Forschung in bisher unzugänglichen Bereichen erlauben (Bordia 1996) und
- insgesamt die wissenschaftstheoretischen und praktischen Vorteile die Nachteile mehr als ausgleichen (Reips 1997a).

³ Eine längere Liste von Argumenten zu Web-Experimenten findet sich bei Reips (1996a).



Abbildung 4: Das Web-Labor für Experimentelle Psychologie

Die große Freiheit der Versuchspersonen, ein Web-Experiment jederzeit beenden zu können, hat ihren Preis für die Forschenden. Ein Problem, das dadurch in stärkerem Maße eine Rolle spielt, ist *Selbstselektion*. Hiervon sind beispielsweise oft Umfragen, aber auch Experimente zu brisanten Themen betroffen. Stellt man beispielsweise eine Web-Seite ins Netz, auf der steht "Um an einer Umfrage über § 218 teilzunehmen, bitte hier klicken", dann kann man sich ziemlich sicher sein, nachher nur die Meinung generell bis speziell an der Schwangerschaftsabbruch-Thematik Interessierter erhalten zu haben. Vor der Durchführung eines Web-Experiments sollte man die Möglichkeit der Selbstselektion genau erwägen und geeignete Vorkehrungen treffen. Dazu gehören faktorielle Versuchspläne mit Variationen auf den Faktoren, geeignete Untersuchungsgegenstände und Replikationen mit anderen Stichproben, die auch zum Beispiel durch einen zweiten Web-Zugang zum Experiment entstehen können.

Zum Zweck der Durchführung von Web-Experimenten wurde im Sommer 1995 an der Universität Tübingen das *Web-Labor für Experimentelle Psychologie* (Reips 1995) gegründet, das im April 1998 physisch an die Universität Zürich umgezogen ist (Abbildung 4 zeigt das Logo des Web-Labors). Es kann mittlerweile eine Zugriffszahl von ca. 2.000-3.000 Besuchen im Monat vorweisen (vgl. Abbildung 5). Seither sind eine Reihe weiterer Experimentallabors im WWW eingerichtet worden:

- *Sozialpsychologische Werkstatt Jena*
(<http://www.uni-jena.de/~ssw/zugang.htm>)
- *Trierer Experimentalserver*
(<http://cogpsy.uni-trier.de:8000/TEServ.html>)
- *Max-Planck-Institut für Biologische Kybernetik Tübingen*
(http://exp.kyb.tuebingen.mpg.de/web-experiment/index_de.html)
- *Online Psychology Lab Padua*
(<http://www.psy.unipd.it/personal/laboratorio/surprise/htmltesi/index.html>)

Listen weiterer Web-Experimente gibt es zum Beispiel bei der American Psychological Society (<http://psych.hanover.edu/APS/exponnet.html>) und auf der Seite des Psychologie-Einführungskurses "Psychological Perspectives on Human Behavior" von Snell (<http://psychology2.semo.edu/websites/web41.htm>).

Richtlinien für die Durchführung eines Experiments im Web-Labor finden sich bei Reips (1996b); das Web-Labor steht prinzipiell für die kooperative Durchführung von

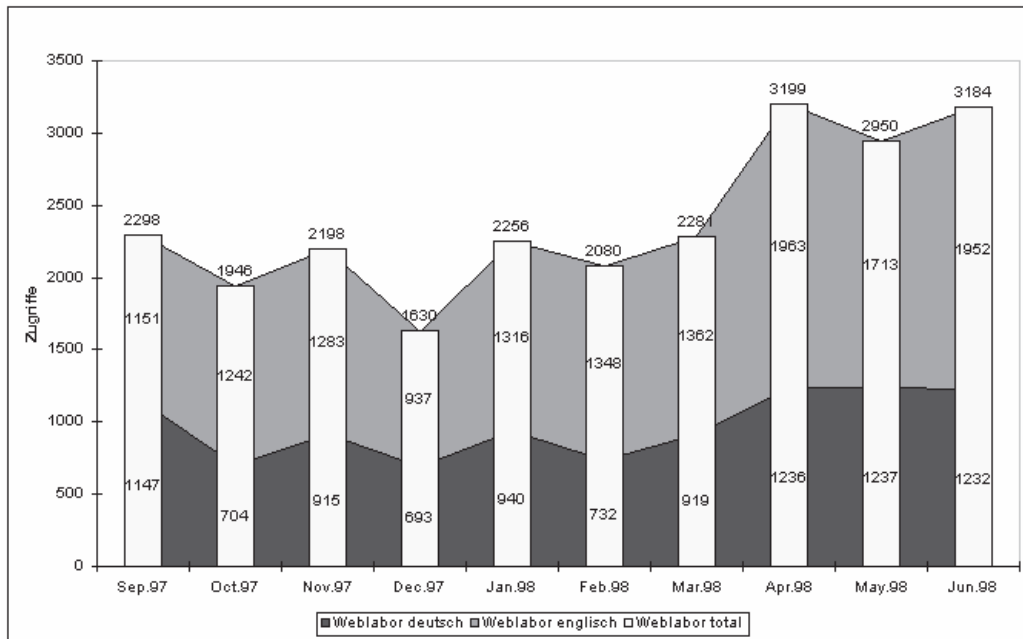


Abbildung 5: Besuche in der Tübinger Version des Web-Labors für Experimentelle Psychologie, September 1997 bis Juni 1998

Web-Experimenten zur Verfügung. Der Vollständigkeit halber sei hier auch auf die Möglichkeit *verdeckter* Experimente im WWW hingewiesen (Hänze & Meyer 1997). Hänze und Meyer manipulierten in experimenteller Absicht bestimmte Abbildungen und Hintergrundfarben auf den Web-Seiten eines etablierten Icon-Archivs. Sie konnten so den von Kunst-Wilson und Zajonc (1980) beschriebenen *mere exposure effect* (zunehmende Präferenz für wahrgenommene Reize, denen man oft begegnet) für Hintergrundfarbenstimuli replizieren.

Motivationale Techniken des Web-Experimentierens

Anwerbung und Motivation

Kunden oder – in diesem Fall – Versuchsteilnehmer anzuwerben und sie dazu zu motivieren, eine Weile auf der Website zu verweilen, ist ein allgemeines Thema im Online-Geschäft. Wie sich zeigen wird, gilt es bei der Anwerbung und Motivation von Versuchspersonen für Online-Forschung, ein paar Besonderheiten zu berücksichtigen.

Zunächst sollte die Website generell attraktiv sein. Erreicht werden kann dies durch:

- ein ansprechendes Design (z.B. Graphiken),
- Verweise von außen auf die Website (Anmeldung bei Suchmaschinen),
- Hinweise darauf, daß dies ein interessanter Fleck im Web sein könnte (verliebene Preise, Verweise auf Presseberichte),

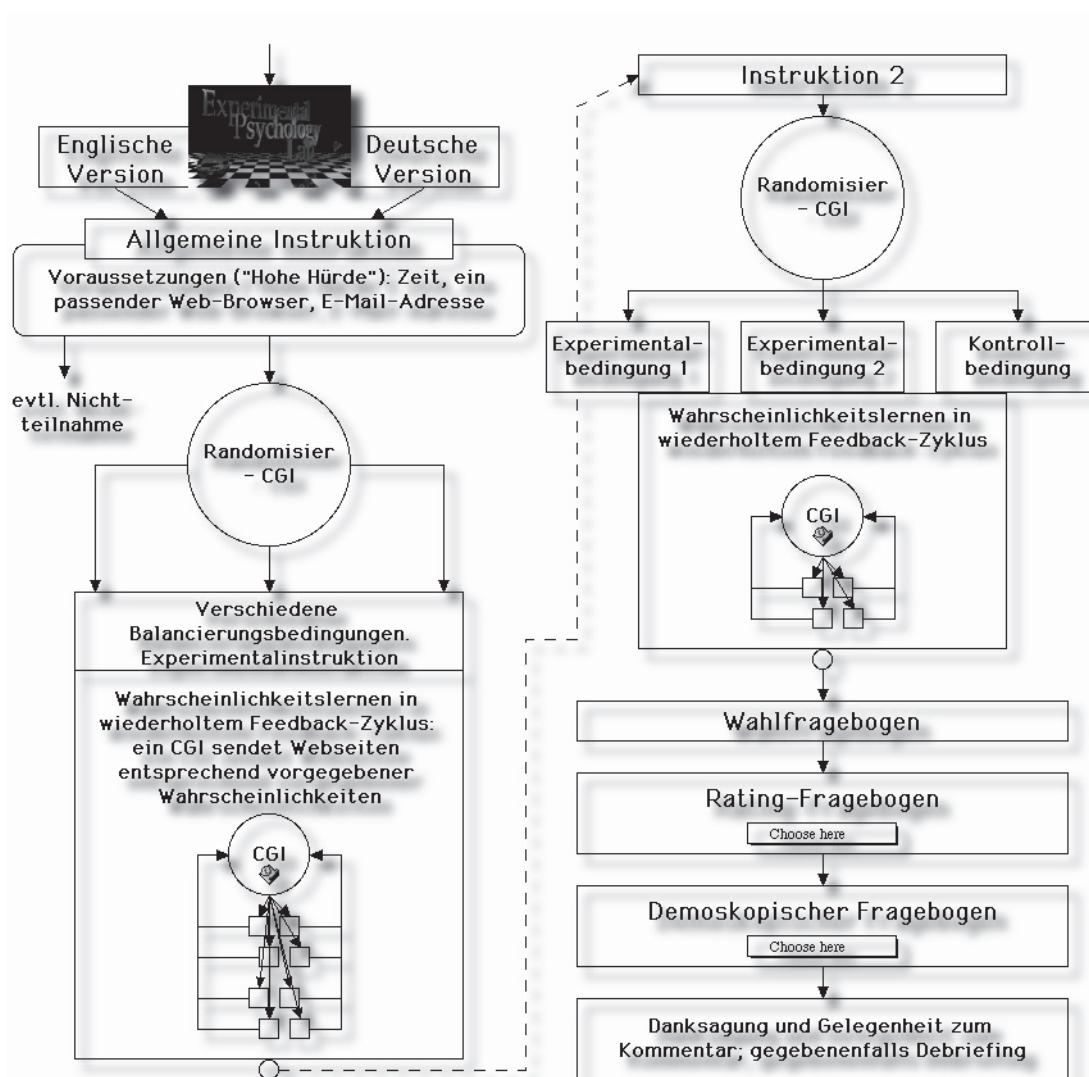


Abbildung 6: Ablauf eines Web-Experiments am Beispiel einer Untersuchung zur Bedeutung kausaler Mechanismen beim Kausallernen (Reips, 1996b)

- Mehrsprachigkeit (vor allem Englisch) und
- das Anbieten verschiedener Möglichkeiten des Betrachtens (Frame/No-Frame/Text/Print-Version).

Die Seriosität des Anliegens sollte durch entsprechende Hinweise unterstrichen werden:

- Angabe der tragenden Institution(en),
- Verweis auf die Wissenschaftlichkeit des Unterfangens,
- Angabe einer E-Mail-Kontaktadresse und
- Zusicherung der Vertraulichkeit (und Einhaltung derselben!).

Die Versuchsteilnehmer sollten erfahren, wieviel Zeit die Teilnahme an den Web-Experimenten jeweils in Anspruch nimmt. Es hilft, eine kleine Belohnung für die Teilnahme in Aussicht zu stellen und (Geld-)Preise unter den Versuchsteilnehmern zu verlosen. Selbstverständlich ist auch eine zu erwartende Rückmeldung über den Sinn und den eigenen Platz in der Untersuchung ein Anreiz teilzunehmen (vgl. Bosnjak & Batinic in diesem Band).

Sicherung der Verweilmotivation (Abbruchreduktion)

Versuchsteilnehmer zu gewinnen ist eine Sache, sie zu halten eine andere. Aus methodischen Gründen ist es wichtig, daß die Abbruchquote bei einem Web-Experiment niedrig ist. Dies gilt noch mehr bei Befragungen, da dort ja meist kein Vergleich zwischen mehreren Ausprägungen einer unabhängigen Variablen stattfindet. Während allerdings die Aussagekraft einer Umfrage damit kippt, daß zum Beispiel diejenigen Leute vermehrt aussteigen, die das Thema der Umfrage nicht interessiert, kann das selektive Aussteigen von Versuchspersonen in einer von mehreren Bedingungen in einem Web-Experiment als Hinweis auf eine Konfundierung sogar genutzt und ein potentieller Artefakt damit kontrolliert werden.

Viele Web-Surfer verlassen Sites vorzeitig, weil die Ladezeiten der Dokumente, besonders der Graphikdateien, zu lang sind. Eine gute Methode, diesem Problem zu begegnen, ist die *systematische Verringerung der Ladezeiten*, je weiter jemand in die Site vorstößt. Beispielsweise sind die meisten Experimente im *Web-Labor für Experimentelle Psychologie* so strukturiert, daß die Web-Labor-Seite (<http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLabD.html>) die größte Ladezeit beansprucht und die *generelle Instruktion* der Experimente am zweitlangsamsten aufgebaut wird. Die *experimentenspezifische Instruktion* ist dann noch schneller und die *Experimentalmaterialien* schließlich bestehen meistens fast nur noch aus Text und den Graphiken, die bereits auf den vorherigen Web-Seiten verwendet wurden und deshalb “in no time” aus dem Festplatten-Cache geladen werden. Wer also wegen zu langer Ladezeit abbricht, wird dies eher früher als später tun.

Ein ähnliches Vorgehen empfiehlt sich bei den “psychologischen Ladezeiten”, den “inneren Hürden” vor einer Teilnahme. Beispielsweise sollte die Versuchsperson auf der generellen Instruktionseite auf die Verbindlichkeit bei Entschluß zur Teilnahme hingewiesen und um die Angabe ihrer E-Mail-Adresse oder ihrer Telefonnummer gebeten werden. Sie sollte (unter Umständen mittels eines kleinen Tests) gegebenenfalls mit der Notwendigkeit eines bestimmten Browser-Typus konfrontiert und zum Runterladen desselben veranlaßt werden. Außerdem sollte man die Versuchsperson – auch aus ethischen Gründen – schonungslos über die Zeitdauer und andere möglicherweise unangenehme Aspekte des Experiments aufklären. Die höchste Hürde sollte, wie gehabt, am Anfang stehen.

Routinierte Web-Surfer wissen, daß sie ein “schnupperndes Völkchen” sind. Nicht umsonst suggeriert der Begriff “Web-Browser” eher zielloses “Durchblättern” von Zeitschriften als geplantes Studieren. Was tun mit der “Schnuppererei”? Um in einem

Web-Experiment diejenigen, die nur mal kurz reinschauen wollen, bevor sie weitersurfen, von denjenigen zu trennen, die sich ernsthaft festgebissen haben, hat sich die *Aufwärm-Technik* bewährt. Wenn die eigentliche experimentelle Manipulation in einem Web-Experiment erst nach einer Aufwärmphase stattfindet, dann kann eindeutig festgestellt werden, daß die Abbrecher nicht wegen dieser Manipulation ausgestiegen sind.

Ankündigungen dessen, was die Versuchsperson zu erwarten hat, verringern die Wahrscheinlichkeit, daß sie wegen einer unangenehmen Überraschung ihre Teilnahme terminiert. Es empfiehlt sich also, eine grobe Beschreibung des experimentellen Ablaufs zu geben, zum Beispiel “Instruktion – Phase 1 – ein Fragebogen mit drei Fragen – Phase 2 – zwei Fragebögen mit je fünf Fragen. Gesamtdauer ca. 25 Minuten.” Zusätzlich kann man der Versuchsperson während des Web-Experiments Rückmeldungen über ihre zeitliche Position in diesem Ablauf geben. Denn wer weiß, wie lange es ungefähr noch dauert, bleibt eher dabei.

Ob Ankündigungen, Instruktionen oder Fragebögen – jeglicher Text sollte in positiver, persönlicher Sprache verfaßt werden. Tip: am ehesten gelingt das, wenn die Versuchsperson beim Schreiben als Partnerin in einer gemeinsamen Unternehmung imaginiert wird, die sie ja auch ist.

Die Techniken in der Praxis

Fast alle der in diesem Kapitel geschilderten Techniken wurden in dem ersten im WWW durchgeführten Lernexperiment eingesetzt (Reips 1996c). Unter <http://134.2.133.25/WWWExpOrgD.html>, wo dieses Web-Experiment als dokumentarisches Zeugnis weiterläuft, kann man sich in die Sicht einer Versuchsperson hineinversetzen. Von insgesamt 880 Versuchspersonen, die zwischen dem 18.1. und 16.9.1996 an diesem ca. 30- bis 60-minütigen Web-Experiment teilnahmen, brachen 255 das Experiment in der Aufwärmphase ab. 29 weitere wurden in dieser Phase Opfer eines technischen Problems auf der Server-Seite. Insgesamt begannen also 596 Versuchspersonen das eigentliche Experiment. In dieser Phase betrug die Abbruchquote weniger als zehn Prozent (58), wenn man von zehn technischen Ausfällen absieht. “Abbruch” bedeutete in der Experimentalphase natürlich einfach, daß die Versuchsperson keinen der Fragebögen ausfüllte. Ein großer Teil der 58 Personen könnten also “Insider” (andere Wissenschaftler, die sich über die Methode informieren wollten) oder Fragebogen-Scheue sein.

Unter den verbleibenden 528 Versuchsteilnehmern befanden sich 25, die weder eine E-Mail-Adresse noch eine Telefonnummer angegeben hatten und deren Daten deshalb ausgeschlossen wurden. Desgleichen wurde mit vier Datensätzen verfahren, die innerhalb eines Zeitraums von zwei Tagen nach einer ersten Teilnahme zwar mit einer anderen E-Mail-Angabe, aber doch von der gleichen oder annähernd gleichen Computeradresse aus entstanden. Es ist zwar gut möglich, daß dies jeweils verschiedene Personen waren, aber es sollte allen Eventualitäten (z.B. Täuschungsversuchen) vorgebeugt werden. Es kam nicht ein einziges Mal vor, daß jemand unter Angabe der

gleichen E-Mail-Adresse oder Telefonnummer mehrfach teilnahm. Letztlich gingen die Daten von 499 Versuchspersonen aus dem WWW in die Auswertung ein.

Kontrolltechniken des Web-Experimentierens

Contra Datenschrott

Bereits am Ende des vorangehenden Abschnitts wurden einige Techniken geschildert, die verhindern helfen, daß erhobene Daten unbrauchbar sind. Die Überprüfung von E-Mail-Adressen und IP-Adressen auf Doppelungen und zu große Ähnlichkeit kann man nicht nur durchführen, man kann dies von vornherein auch ankündigen. Dabei kann man getrost damit rechnen, daß ein Großteil des WWW-Publikums ein zu geringes technisches Verständnis des Mediums hat, als daß es die Grenzen der Überprüfbarkeit von Mehrfachteilnahmen erkennen könnte. Generell hat man auch den Goodwill der Teilnehmer auf seiner Seite – nach bisheriger Erfahrung wird es nur selten vorkommen, daß jemand sich über die Aufforderung, nur einmal teilzunehmen, hinwegsetzt. Dies gilt in besonderem Maße, wenn die Teilnahme an einem Web-Experiment wegen seiner Länge und der oben beschriebenen “hohen Hürden” für die Versuchsperson ei-

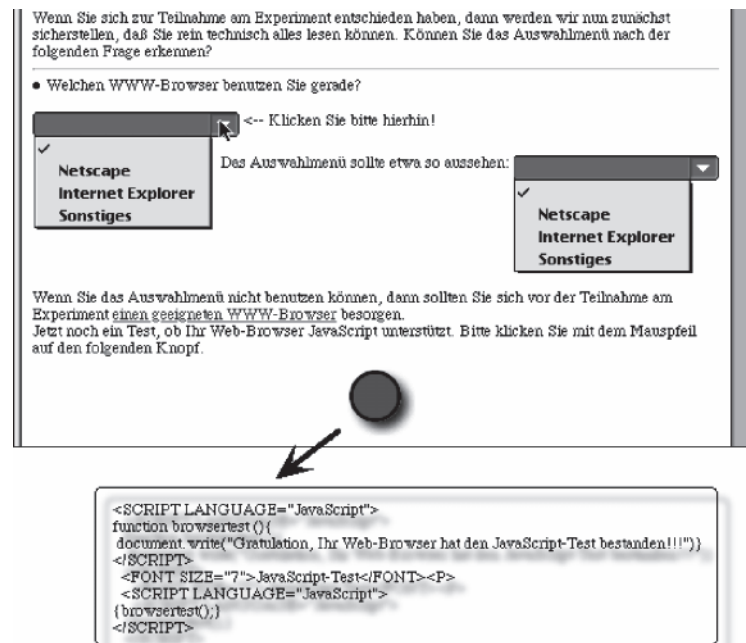


Abbildung 7: Aufklappmenü und JavaScript-Test. Unten ein JavaScript, mit dessen Hilfe sich die JavaScript-Tauglichkeit eines Web-Browsers bestimmen läßt.

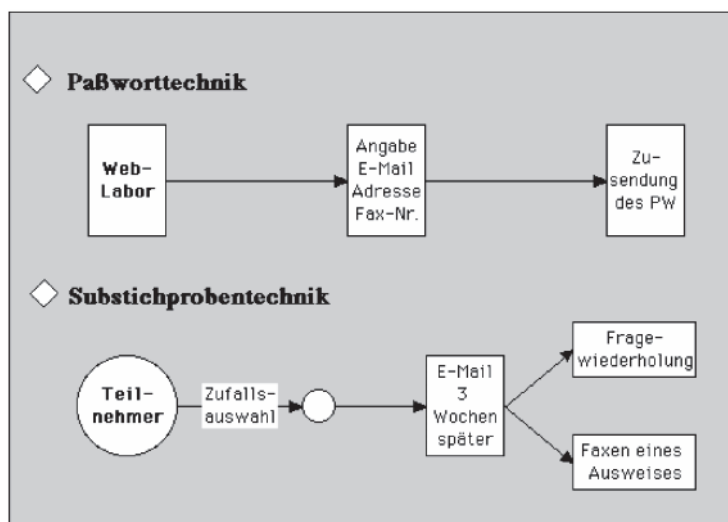


Abbildung 8: Passworttechnik und Substichprobenverfahren

nen nicht unerheblichen Aufwand bedeutet. Ebenso unterstützend ist es, wenn die Web-Site erkennen läßt, daß eine seriöse und/oder gemeinnützige Institution hinter dieser Forschung steht.

Was aber tun mit den Begeisterten, denen, die sich das Ganze unbedingt nochmal anschauen wollen? Man gibt ihnen eine Kanalisierungsmöglichkeit, die Chance, quasi außer der Reihe ein zweites Mal teilzunehmen. Dies kann beispielsweise über ein Link geschehen, oder über eine Frage am Anfang. Generell sollte man versuchen, mit Fragen überprüfbare Informationen zu sammeln. Andere Fragen können als Konsistenzdiagnostika dienen: daß jemand gleichzeitig angibt, 18 Jahre alt und Großmutter zu sein, ist ein Grund zur Vorsicht.

Auf der rein technischen Ebene läßt sich Datenschnitt dadurch vermeiden, daß man testet, ob der Web-Browser einer Versuchsperson alle benötigten Features, wie zum Beispiel Aufklappmenüs und JavaScript, unterstützt (vgl. Abbildung 7). Wenn nicht, dann sollten die Versuchspersonen auf eine Quelle für einen adäquaten Web-Browser verwiesen werden. Mit Hilfe von JavaScript läßt sich auch feststellen, welche Monitorauflösung eine Versuchsperson eingestellt hat.

Spezialtechniken

Mit Hilfe besonderer Kniffe läßt sich eine zusätzliche Kontrolle erreichen. Beispielsweise kann man ernsthaft an einer Teilnahme interessierte Versuchspersonen mit Hilfe der *Passworttechnik* ermitteln (vgl. Abbildung 8 oben). Bei diesem Verfahren melden sich die an einer Versuchsteilnahme interessierten Personen auf einer Web-Site unter Angabe ihrer E-Mail- und Snail-Mail-Adressen und gegebenenfalls ihrer Fax- und Telefonnummern an. Sobald eine genügend große Anzahl von Anmeldungen vorliegt, bekommen die Interessenten ihr Login und Paßwort zugesandt, wobei das Medium der

Zusendung variiert werden kann. (Diese Variation erlaubt einen späteren Vergleich zwischen den Gruppen, die ihr Paßwort per E-Mail bekommen haben, denen, die es per Brief erhielten und denen, denen es zugefaxt wurde). Auf diese Weise kann jede Teilnahme am Experiment einer anderen Person zugeschrieben werden. Außerdem kann man die Versuchspersonen bei eventuell auftauchenden Fragen zuverlässig nochmal kontaktieren.

Eine weitere Technik ist das zufällige Ziehen einer *Substichprobe* aus der Menge aller Versuchsteilnehmer (s. Abbildung 8 unten). Die Personen in dieser Substichprobe erhalten eine E-Mail zugeschickt, die sie zunächst nur mit einer wortlosen Reply beantworten sollen. Erstens erhält man so einen Schätzwert für den Anteil falscher E-Mail-Adreßangaben. Zweitens ist der Aufwand, einen wortlosen Reply abzuschicken, sehr gering für die Versuchsperson, so daß man eine hohe Rückmelderate erzielen sollte. In einer weiteren E-Mail werden die Versuchspersonen dann gebeten, bestimmte Kontrollfragen bezüglich der Untersuchung zu beantworten. Man kann die Personen der Substichprobe auch beispielsweise um das Zufaxen eines Ausweises bitten und rigoros-penetrant so lange immer wieder kontaktieren, bis eindeutig geklärt ist, wer nur einmal und wer mehrfach am Web-Experiment teilgenommen hat.

Anstatt nach der Durchführung eines Web-Experiments mit der Ziehung einer Substichprobe zu arbeiten, kann man auch von vornherein dafür sorgen, daß man Informationen über die Versuchspersonen hat. Dies erreicht man mittels der *Versuchsperson-Pool-Technik* (vgl. Abbildung 9). Mitglieder eines Versuchspersonen-Pools müssen sich bei Eintritt in den Pool dazu bereit erklären, prinzipiell zur Teilnahme an Web-Experimenten oder Befragungen zur Verfügung zu stehen. Sie geben ihre demographischen Daten an, die selbstverständlich unter Wahrung der Vertraulichkeit zusammen mit eventuellen Testergebnissen und den Experimentaldaten verwaltet werden müssen. In diesem Rahmen ist es auch möglich, die Versuchspersonen für jede Teilnahme an einer Untersuchung zu entlohnen. Versuchspersonen-Pools haben den Vorteil, daß bekannt ist, an welchen anderen Studien eine Versuchsperson teilgenommen hat. Außerdem kann eine Wissenschaftlerin die über die Versuchspersonen abgespeicherten Daten dazu nutzen, gezielt Versuchspersonen mit bestimmten Merkmalen fast wie in einem Katalog zu "bestellen" – im einfachsten Fall zum Beispiel je 50% Frauen und Männer. Die Versuchspersonen-Pool-Technik unter Bildung eines Versuchspersonen-Pools im Internet ist eine gute Chance für Forschende an kleineren Forschungseinrichtungen, die wenig Zugang zu Versuchspersonen haben und gerne trotz Nutzung des Internet eine gewisse Kontrolle darüber haben möchten, wer an ihren Untersuchungen teilnimmt.

Bei der Auswertung der Daten kann man ebenfalls Techniken anwenden, die die Datenqualität sichern. Ein Beispiel wurde schon weiter oben genannt: jeweils nur die erste Teilnahme von einer Computeradresse (IP) aus zuzulassen beziehungsweise auszuwerten. Auch die von den meisten Web-Server-Programmen aufgezeichneten relativen Abrufzeiten der Dokumente kann man als Datenfilter nutzen. In einem Lernexperiment wird man zum Beispiel die Daten aller Versuchspersonen nicht berücksichtigen wollen, die irgendwann während des Versuchs eine Pause gemacht ha-

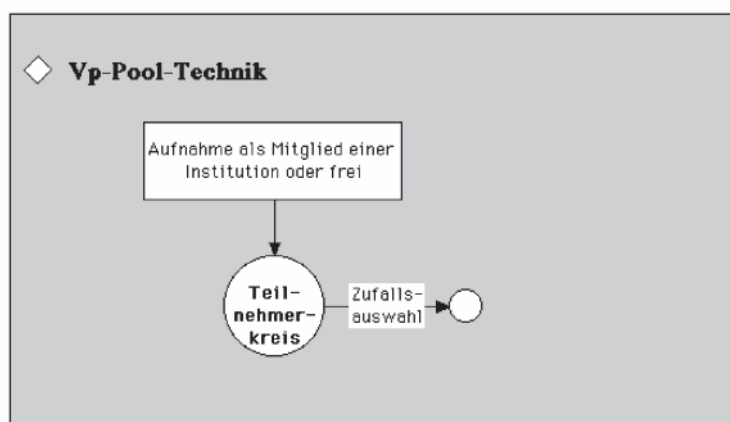


Abbildung 9: Nutzung eines Versuchspersonen-Pools

ben, was sich als ein Anstieg um einen bestimmten Prozentsatz der durchschnittlichen Intervalldauer zwischen zwei Lerndurchgängen dieser Versuchsperson genau definieren läßt.

Epilog

Web-Experimente sind die Übertragung der experimentellen Methode in die virtuelle Welt des World Wide Web. Dank der Logik des Experiments lassen sich mit Web-Experimenten Annahmen über kausale Zusammenhänge überprüfen, ohne daß dabei die Frage nach der Repräsentativität die Forschenden von vornherein als naiven Stoßtrupp auf dem Weg in unbekannte Welten erscheinen läßt. Das World Wide Web eignet sich als Raum für die Durchführung von Experimenten und bietet die Chance, auch in bisher unzugänglichen Bereichen (etwa mit ganz speziellen Gruppen von Personen) mit dieser Methode zu arbeiten. Sollten Sie Interesse daran haben, selbst ein Web-Experiment durchzuführen, dann besuchen Sie doch einmal eines der Online-Labors, zum Beispiel das Web-Labor für Experimentelle Psychologie: <http://www.uni-tuebingen.de/uni/sii/Ulf/Lab/WebExpPsyLabD.html> (in Tübingen) oder <http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebExpPsyLabD.html> (in Zürich).

Literatur

- Batinic, B. & Bosnjak, M. (1997). Fragebogenuntersuchungen im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 221-243). Göttingen: Hogrefe.
- Bordia, P. (1996). Studying verbal interaction on the Internet: The case of rumor transmission research. *Behavioral Research Methods, Instruments, & Computers*, 28, 149-151.

- Gräf, L. (1997). Pretest von WWW-Umfragen. In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott & G. Strube (Hrsg.), *CAW-97: Beiträge zum Workshop "Cognition & Web"* (S. 51-62). Freiburg: IIG-Berichte.
- Graphics, Visualization, & Usability Center (1997). *GVU's WWW user surveys*. URL: <http://www.cc.gatech.edu/gvu/user_surveys>.
- Hänze, M. & Meyer, H. A. (1997). Feldexperimente und nicht-reaktive Datenerhebung im World Wide Web. In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott & G. Strube (Hrsg.), *CAW-97: Beiträge zum Workshop "Cognition & Web"* (S. 79-92). Freiburg: IIG-Berichte.
- Krantz, J. H., Ballard, J. & Scher, J. (1996). Comparing the results of laboratory and world-wide web samples on the determinants of female attractiveness. *Behavioral Research Methods, Instruments, & Computers*, 29, 264-269.
- 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. (Hrsg.)(1997). *Survey measurement and process quality*. New York: Wiley.
- Martin, D. W. (1996). *Doing psychology experiments* (4th. ed.). Pacific Grove: Brooks / Cole.
- Monte, C. F. (1975). *Psychology's scientific endeavor*. New York: Praeger.
- Reips, U. -D. (1995). *Das Web-Labor für Experimentelle Psychologie*. URL: < <http://www.uni-tuebingen.de/uni/sii/Ulf/Lab/WebExpPsyLabD.html>> (seit April 1998 auch <<http://www.genpsyserv.unizh.ch>>).
- Reips, U. -D. (1996a). *Methodisches zu Web-Experimenten*. URL: < <http://www.uni-tuebingen.de/sii/Ulf/Lab/WebExpMethode.html>>.
- Reips, U. -D. (1996b). *Experimenting in the World Wide Web*. Paper presented at the 1996 Society for Computers in Psychology Conference. Chicago, IL.
- Reips, U. -D. (1997a). Forschen im Jahr 2007: Integration von Web-Experimentieren, Online-Publizieren und Multimedia-Kommunikation In D. Janetzko, B. Batinic, D. Schoder, M. Mattingley-Scott & G. Strube (Hrsg.), *CAW-97: Beiträge zum Workshop "Cognition & Web"* (S. 141-148). Freiburg: IIG-Berichte.
- Reips, U. -D. (1997b). Das psychologische Experimentieren im Internet. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 245-265). Göttingen: Hogrefe.
- Reips, U. -D. (1997c). Tips zum Erstellen von WWW-Dokumenten. In B. Batinic (Hrsg.), *Internet für Psychologen* (S. 499-517). Göttingen: Hogrefe.
- Reips, U. -D. (1997d). *Richtlinien für die Benutzung des Web-Labors*. URL: < <http://www.psych.unizh.ch/genpsy/Ulf/Lab/WebLabBed.html>>.
- Repräsentativität*. Mailingliste-Thread (1997). Archiv Mailingliste German Internet Research List gir-l. URL: <<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. *Behavioral Research Methods, Instruments, & Computers*, 29, 274-279.
- Sherif, M., Sherif, C. W. (1954). *Groups in harmony and tension*. New York: Harper and Row.

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

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