Erschienen in: Online social sciences / Batinic, Bernad; Reips, Ulf-Dietrich et al. (Hrsg.). - Seattle : Hogrefe & Huber, 2002. - S. 69-80. - ISBN 0-88937-257-8

5 Context Effects in Web Surveys

Ulf-Dietrich Reips

5.1 Abstract

The quality of data collection on the Internet depends in part on potentially biasing surface characteristics of Web pages, such as pop-up menus. Also, answering behavior to survey questions is highly sensitive to cognitive contexts evoked by other survey questions. A Web experiment with two experimental phases is used to identify six potentially biasing factors. Manipulations of scale type, reading directionality, cursor entry position, question order, and – in contrast to findings from offline research – type of numerical labeling do not show effects. However, it is shown that it makes a difference whether survey questions are presented together on one Web page or separately on consecutive Web pages. Results are discussed with regard to general as well as online specific implications for the design of survey instruments and multiple response scales.

5.2 Introduction

The World Wide Web's (WWW) increasing popularity led to the beginning of this communication media's use as a data source for online research in various fields (Hewson, Laurent, & Vogel, 1996; Musch & Reips, 2000). Online data collection methods can be classified into three major subtypes: nonreactive data collection, Web experiments, and online surveys (Reips, 1997, this volume). Especially in online surveys the dependent measure often consists in the user's choice on a multiple response scale. Due to the analogy between online questioning and offline questioning it can be assumed that multiple response scales are prone to answering biases that have been discovered in the use of this measure in offline research (e.g., Schwarz & Hippler, 1994a, 1994b). Early investigations of computer-assisted surveying (e.g., Kiesler & Sproull, 1986) support this assumption. In a meta-analysis of 159 correlations from 29 studies on paper-and-pencil versus computerized test administration, Mead and Drasgow (1993) mostly found very high agreement.

Form and demand characteristics of response scales are dependent on the medium they are used in. The WWW offers a number of ways in constructing response scales. These scales may appear similar to those used in paper-and-pencil studies. On the other hand, scales may also look and function quite differently. Consider, for example, scales that take the form of "pop-up menus" on Web pages. At first glance, a pop-up menu looks like a piece of text within a rectangular shape (see Figure 5.1, top). This shape sometimes is in a different shade of gray or in a different color than the background of the Web page. Also, it might appear in a three-dimensional fashion. If the user clicks with her mouse cursor on the pop-up menu, then a scrollable menu appears (see Figure 5.1, bottom). With pop-up menus, before clicking on it, the user does not see the whole scale at once. Consequently, until clicking on it the user doesn't know how many items a scale contains. This is an obvious difference from scales administered in a paper-and-pencil format.

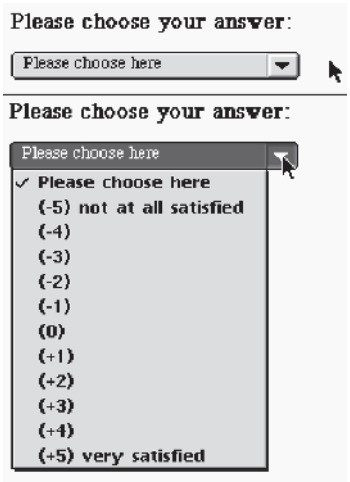

Figure 5.1: A pop-up menu in closed state (top)
and opened (bottom)

One goal of the present study is to find an answer to the question of whether such surface characteristics of WWW typical response scales may lead to biases in answering behavior.

In the literature certain answer-biasing influences are identified as independent of the research method that is used. For example, Schwarz and Hippler (1994a) were able to show that the cognitive context evoked by questions in a survey may influence the respondent's answering behavior in other questions. As a consequence, the present Web experiment aimed at replicating this and other effects including the influence of numeric labeling, question order, and scale polarity in questionnaires that were answered on the WWW.

5.2.1 Possible Biases Studied

The first question to be examined is one of *type of scale*. Are there differences in answering behavior depending on whether a scale is realized as a pop-up menu or as a button scale? Because in most Web browsers a pop-up menu closes as soon as the mouse arrow leaves the opened field, it can be assumed that choices are not equally easy to make. The user has to go back to the top and open the menu again, which may discourage her from choosing "strenuous" answers. Also, for some choices, hand movement is longer than for others. Therefore, answers on pop-up menus may differ systematically from answers on button scales.

A second issue to be looked at is *numerical labeling*. Different numerical labels of the same category on a scale may suggest different meanings even if the verbal labels at the end points are identical. For example, Schwarz & Hippler (1994a) found a more positive evaluation of politicians on eleven point scales in telephone interviews and mail surveys, if the numerical labels ranged from –5 to +5 instead of ranging from 0 to 10. Apparently, this is due to an attribution of a negative quality to numbers that carry a minus sign. Can this result be replicated in a Web experiment?

The third question to be answered involves the *natural reading direction* in western cultures. Does it matter whether the positive end of a scale is on the left or top versus the right or bottom?

Fourth, if the scale is in the form of a pop-up menu, then does it matter whether the *cursor entry and cursor entry label* are at the top of the scale or at the bottom of the scale?

Finally, later items in a survey may influence answering behavior in earlier items. Schwarz and Hippler (1994b) found such influences in mail surveys when compared to telephone interviews. In mail surveys, readers may read through the whole survey before answering the first question. They may go back to earlier questions and change their answers, or they may even choose to make one of the last items the first one they answer. This is not possible in telephone interviews. A comparable situation emerges in Web-based surveys, where survey items may either be placed on a single Web page or one by one on consecutive Web pages. If the hypothesis holds true that some questions will stimulate cognitive contexts that are capable of influencing answers to other questions then results in a single Web page condition should resemble those from a mail survey. The resulting pattern from a survey that is presented on multiple Web pages containing one item each should be similar to data from a telephone interview. In their study, Schwarz and Hippler (1994a, p.4) "asked respondents how much money they would be willing to donate to support the suffering citizens of Russia." The donation question was either immediately preceded or immediately followed by two questions about taxes, thereby evoking a cognitive spending context. As predicted, in the mail survey condition the amount indicated in the donation

question was affected independent of whether the context questions preceded or followed the donation question. In the telephone interview, the answer to the donation question was only influenced by preceding questions.

Besides replication of Schwarz and Hippler's results using single versus multiple Web pages, the present Web experiment explores bidirectionality of context effects. While Schwarz and Hippler only looked at the effect a spending context may have on donations, it seems consequential to realize a design, which allows for evaluation of both directions a context influence could flow. The donation context may have influenced cognitions on spending as well. A second change is made in this study in regard to the focus of the spending question. While Schwarz and Hippler asked for the willingness to donate (i.e., an intended behavior), the present study asks for past donations (i.e., the memory of past behaviors).

5.3 Method

In order to explore whether several types of answering biases apply to online surveys, a Web experiment was conducted. As the name implies, Web experiments are experiments conducted on the WWW. They are relatively new in experimental research. The first Web experiments were conducted in 1995 (Musch & Reips, 2000). A typical Web experiment works as follows (see Figure 5.2, Reips, 1998, 2000b, this volume). At its core is a Web server computer program residing on a computer that is permanently connected to the Internet. Participants retrieve the experimental materials using Web browsers on their computers. The Web server serves the experimental materials, depending on the participants actions and small helper applications. These so-called CGIs (Common Gateway Interface script) can be used to perform a number of tasks, including randomized distribution of participants to experimental conditions (Kieley, 1996; Reips, 1996b, 1998, 2000b). All page requests and form inputs made by participants are written to a logfile, which can then be analyzed. Figure 5.2 shows a schematic description of a Web experiment. Definitions, discussions, and descriptions of Web experiments can be found in Birnbaum (2000), Krantz, Ballard, and Scher (1997), and Reips (1995, 1998, 2000a, 2000b, this volume).

Krantz and Dalal (2000) have shown that most studies involving both WWW and laboratory samples show high agreement between data from both types of experiments. A recent survey showed a similar view is shared by most Web experimenters, who indicate a confidence in the ability of Web experiments to produce valid results (Musch & Reips, 1998).

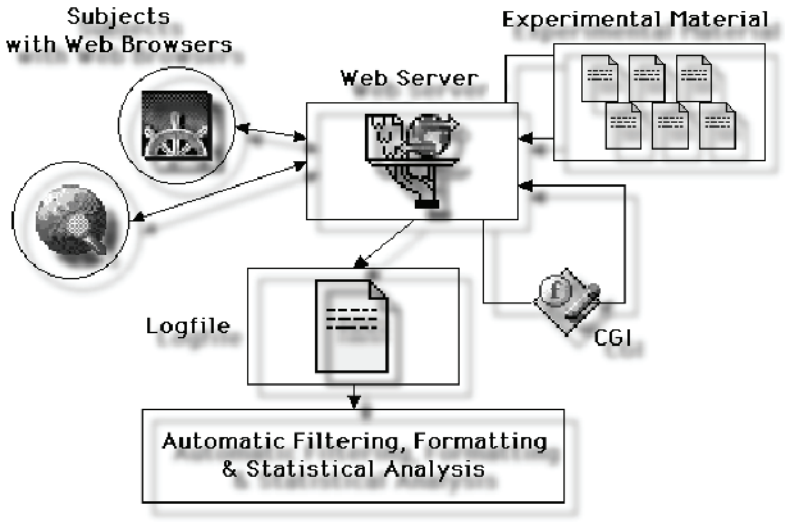

Figure 5.2: Schematic description of a web experiment
(from Reips, 1996b)

The Web experiment was announced with a different entry page URL in the mailing list *gir-l (German Internet Research List)*, in order to check for a possible expert effect. As an incentive for participation and correct submission of their e-mail address, participants received a humorous story sent out by e-mail after completion of a series of experiments.

5.3.1 Participants

334 people started the experiment, as determined by distinct e-mail addresses and distinct computer addresses (IPs). 190 times the German entry page was used; 144 times the English entry page. Of the participants using the German language version, 41 were possible experts, as they used the URL announced in gir-l. Due to technical error (erroneous hyperlinking, leading to recurring pages), 24 data sets had to be discarded for the second experimental phase. 292 participants answered the demographical questionnaire at the end of the experiment (German: 158; English: 134), resulting in a drop out rate of 5.8%.

5.3.2 Procedure

Those people who followed the hyperlink from the Web Experimental Psychology Lab to the experiment were shown a Web page containing general instructions. This page informed about confidentiality of participation, asked for submission of the participant's e-mail address, and contained tests for pop-up menu compatibility and

JavaScript compatibility of the participant's Web browser. The Web experiment was built according to the "high hurdle technique" (Reips, this volume), which aims at reducing drop-out during the course of the Web experiment. After reading the general instructions page, participants clicked the "submit button" as an indication that they agreed to participate in the Web experiment. They then were randomly distributed to one of the 12 conditions of the first experimental phase (see Figure 5.3). These conditions were constructed according to a 2 (scale type) x 2 (numerical labeling) x 2 (reading directionality) design, with an additional dichotomous independent variable (cursor entry position) nested in the pop-up menu scale type.

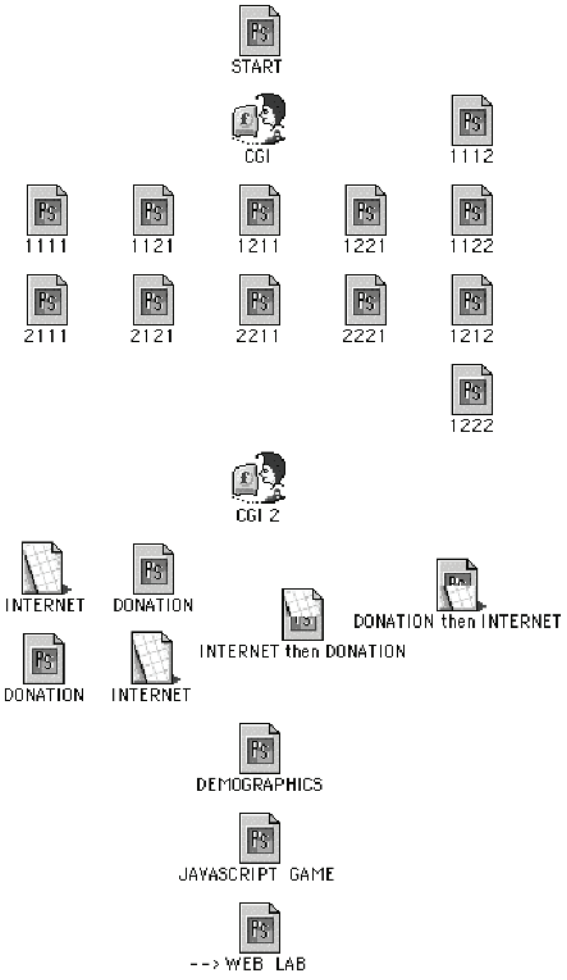

Figure 5.3: Experimental procedure: start, first experimental phase (12 conditions), second experimental phase (4 conditions), demographical questionnaire, JavaScript game. CGIs allow for randomized distribution of participants to conditions.

Each condition was realized on a single Web page containing the question "Overall, how satisfied are you with the quality of your connection to the Internet?" and the answering scale with the verbal end labels "not at all satisfied" and "very satisfied".

Pressing the "submit button" at the bottom of the Web page triggered random distribution via CGI to one of four conditions in the second experimental phase. These conditions resulted from administering presentation of the two spending context and donation context questions on one or two Web pages in both possible question orders. The questions were: "How many thousandth parts of your income would you pay for a good and reliable Internet connection?" and "On average, how many thousandth parts of your income did you give to the charities during the last two years?". Asking for fractions of income was a way to ensure an answer format that is largely independent of currency, socioeconomic class, and culture.

Following the second experimental phase participants were asked to answer three additional questions about their age, their sex, and their expert status regarding scientific online research. At the end participants had the opportunity to play a game on which no data were recorded. All other input made by the participants was written to a logfile residing on the Web server. The data in this logfile provided the basis for statistical analyses.

The Web experiment can be looked at on the WWW at the URL http://www.psych.unizh.ch/genpsy/Ulf/Lab/archive/89452Estart.html. (Offering the possibility of participation in the discussed experiment to you, the reader, is a general advantage of Web experiments, which may lead to an increased understanding of "Procedure" sections like this and make our science's research much more transparent. For a discussion see Reips, 1997, 2000a.)

5.4 Results

To check whether experts for online research show a different performance than non-experts two *t* tests were computed. An alpha level of .05 was used for all statistical tests. There were no significant differences, neither on the satisfaction question in the first experimental phase, $t(323) = .56$, $p = .58$, nor on the difference in income allocation for a donation and an Internet connection, $t(243) = .30$, $p = .77$. Thus, it was concluded that meta-knowledge about online research had no biasing influence on answering behavior in this Web experiment. Consequently, the data of both groups were pooled for the following analyses.

5.4.1 Experimental Phase 1

A 2 (language) x 2 (scale type) x 2 (numerical labeling) x 2 (reading directionality) ANOVA revealed a significant main effect for language $F(1, 309) = 6.68$, $MSE = 5.42$, as well as a significant interaction between language and reading directionality, $F(1, 309) = 6.93$, which carried over to a three-way interaction with numerical labeling and the four-way interaction, Fs(1, 309) = 6.25 and 3.92, respectively, $p < .05$. Also, there was no significant effect for cursor entry position, $t(164) = .59$, $p = .55$. Figure 5.4 shows the mean results acquired during this experimental phase.

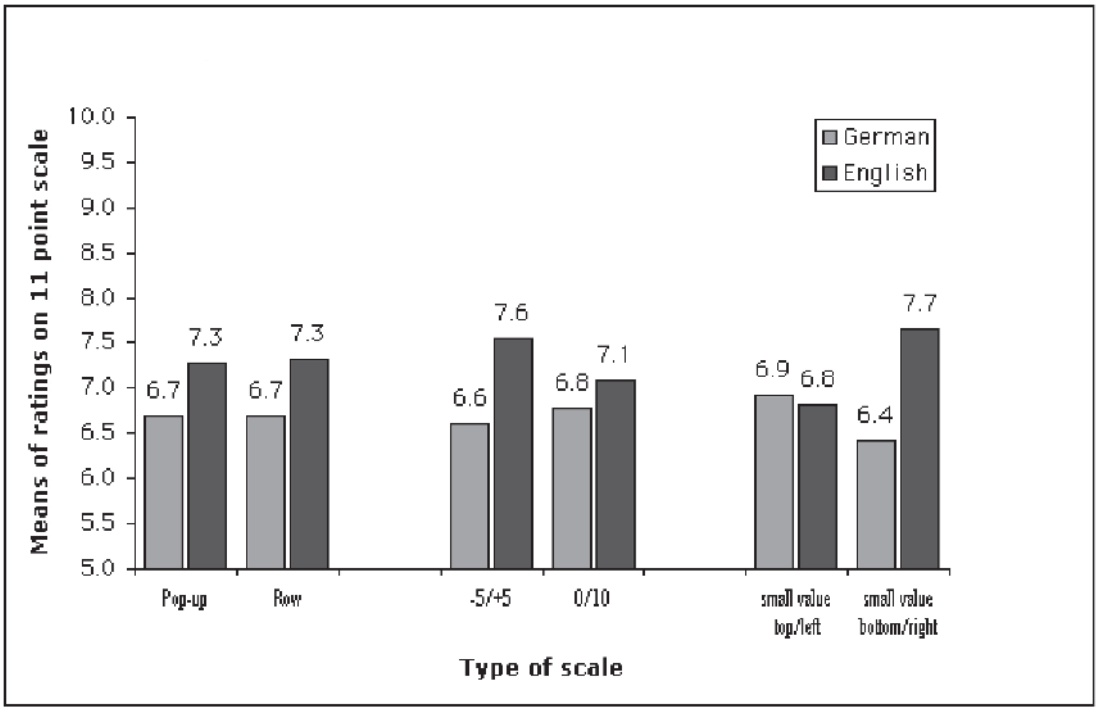

Figure 5.4: Mean ratings during first experimental phase for satisfaction with Internet connection

5.4.2 Experimental Phase 2

A 2 (question separation) x 2 (question order) ANOVA revealed a significant main effect for question separation on the difference in income allocation for a donation and an Internet connection, $F(1, 229) = 8.62$, $p = .004$. There was no significant main effect for question order, and no significant interaction, Fs(1, 229) = .01 and .81, respectively. For the means see Figure 5.5.

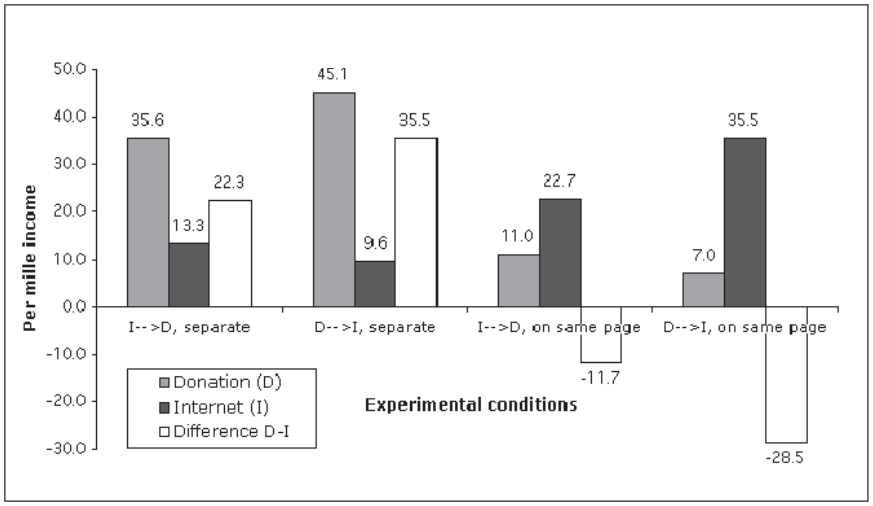

Figure 5.5: Mean reported donations, intended expenses for Internet connection, and differences between both figures, in thousandth of income

5.4.3 Demographical Data

60.1% (161) of those participants who answered the question for sex indicated that they are male, 39.9% (107) checked the mark for "female". Distribution of stated age is shown in Figure 5.6. Out of 267 participants who answered the "insider" question 34.1% (91) stated that they considered themselves to be an "insider".

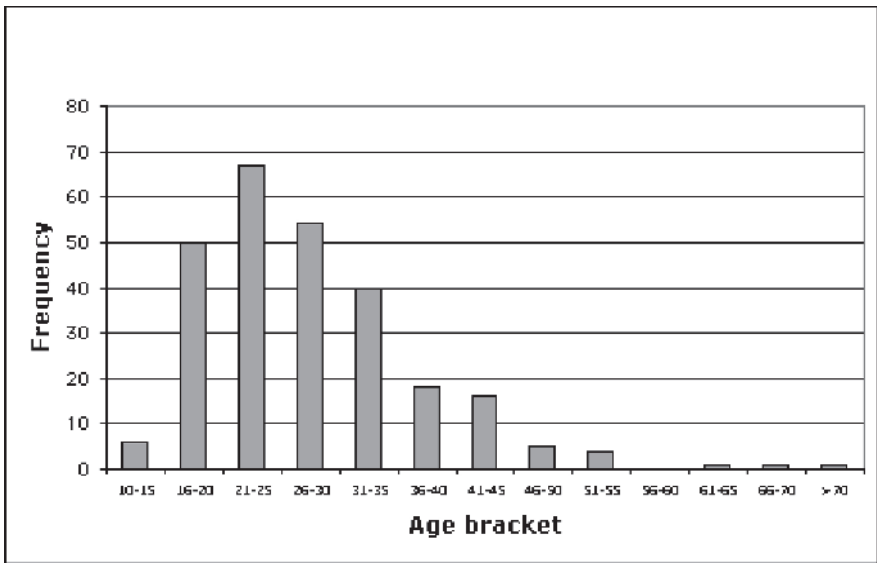

Figure 5.6: Distribution of stated age

5.5 Discussion

Pending replication of the present results it can be cautiously concluded that:

- Whether multiple questions are asked on the same or on different Web pages can lead to different answers. This seems to be due to varying cognitive contexts evoked by the questions themselves.
- This context effect is not only present for intended actions (e.g., how much money to donate, see Schwarz & Hippler, 1994a), but extends to statements about past behaviors (e.g., how much money one has donated). These results support the assumption that such statements about past behaviors might be based on an estimate of how one would behave in the present.
- Surface characteristics of Web questionnaires like pop-up menus versus button scales and numerical labeling don't seem to have an influence on answering behavior of Web participants.
- The difference found in the first experimental phase between language versions of the experiment might be due to real speed differences in Internet connection. The average speed of Internet connections is higher in English language societies than in German language societies. Other possible explanatory candidates for this difference are subtle semantic variations in meanings of terms, or cultural differences, such as a potential tendency towards more pronounced statements in the U.S.A. For evaluations of people, such as evaluations of politicians in Schwarz and Hippler (1994b), one might be able to replicate their findings on the effect of numerical labeling, since such evaluations are less factual.

Going further than Schwarz and Hippler (1994a), the design of the present study allows for the cautious proposal of a hypothetical "total context model." In other words, the total amount of expenditure behavior, no matter whether intended or remembered, is a salient cognitive barrier, which cannot easily be penetrated by context effects. This total amount is then distributed according to a function determined by the relative strength of cognitive contexts involved, and the possibility to show and/or change answering behavior.

A general conclusion that can be drawn is that WWW specific surface characteristics of surveys such as pop-up menus do not seem to make a difference. It remains open to further clarification whether the quality of online data collection is generally prone to similar effects as offline data collection, or whether there are medium specific biases. The primary practical WWW specific implication for the design of survey instruments is that one should not assume that it does not matter whether survey questions are asked together on a single Web page or separated with a Web page for each question.

References

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

Hewson, C. M., Laurent, D., & Vogel, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the Internet. *Behavior Research Methods, Instruments, & Computers, 28*, 186-191.

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

Kiesler, S. & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*, 402-413.

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.

Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449-458.

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. (1995). *The web experiment method*. [Online]. Available: http://www.genpsy.unizh.ch/Ulf/Lab/WWWExpMethod.html.

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.psych.unizh.ch/genpsy/reips/slides/scipchicago96.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*. Freiburg: IIG-Berichte.

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

Schwarz, N. & Hippler, H. (1994a). Subsequent questions may influence answers to preceding questions in mail surveys. *ZUMA-Arbeitsberichte*, 7.

Schwarz, N. & Hippler, H. (1994b). The numeric values of rating scales: A comparison of their impact in mail surveys and telephone interviews. *ZUMA-Arbeitsberichte*, 8.