

# Information search in the laboratory and on the Web: With or without an experimenter

MICHAEL SCHULTE-MECKLENBECK and OSWALD HUBER  
*University of Fribourg, Fribourg, Switzerland*

The focus of this study is the effect of the location (laboratory vs. Web) of experiments on active information search in decision-making tasks. In two experiments, participants were confronted with two different search method versions (list vs. keyword) for acquiring information about a task from a database. The amount and type of information gathered and the time required for task completion were measured. In Experiment 1, significantly more information was searched for in the laboratory than on the Web when the list version was employed, whereas there was no difference between locations in the keyword version. In Experiment 2, the participants were assigned randomly to the Web or the laboratory condition. The results of Experiment 1 were replicated. Whereas location (and the presence or absence of an experimenter) had an effect on the absolute amount of information gathered in both experiments, the relative distribution and type of information items did not differ.

In recent years, conducting experiments and other studies on the Web has become increasingly popular. Web-based studies are marked by several characteristics: (1) the procedure, including instructions and participants' reactions, is implemented on the computer via the Web; (2) participants may come to the respective website and end the procedure whenever they choose to without prejudice; (3) it is easy to access large samples; (4) motivational confounding can be detected; and (5) the experimenter has access to the number of nonparticipants. (For a detailed discussion of the advantages and disadvantages of Web-based studies, see Reips, 2000, 2002b).

Most Web-based studies use participants' reactions on a response scale (e.g., a yes/no or an interval scale) as a dependent variable. Studies validating laboratory-based experiments on the Web generally reveal that behavior in the two conditions differs only slightly, if at all (Krantz, Ballard, & Scher, 1997; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Mehta & Sivada, 1995; Musch & Klauer, 2002; Pohl, Bender, & Lachmann, 2002).

However, there are areas of psychological research in which dependent variables other than reactions measured on response scales are relevant. One example is the search for information by the participant. In experiments involving this type of behavior, the participant is asked to work on a task (for example, to solve a problem or make a decision). In order to complete the task, the participant

has to actively search for (subjectively) relevant information. The type of information that is searched for, and the sequence in which the search takes place, are the central dependent variables of the behavioral analysis of the process. To our knowledge, no experimental comparisons of search behavior in a decision task between the Web and the laboratory are available. In this study, we investigate such a comparison with the example of risky decision-making tasks and ask the following: (1) Is active search for information in a risky decision-making task identical on the Web and in the laboratory? (2) Are there differences between information acquisition methods (i.e., how a person gets information) used on the Web and in the laboratory?

In the first section of this article, we discuss the advantages, disadvantages, and validity of Web studies. We then present the main results of recent research on risky decision making in the laboratory and on the Web. In the second and third sections, we describe our experiments and their results. Finally, we discuss these results in the light of the theoretical outline.

## **Advantages and Disadvantages of Web Studies**

Conducting experiments on the Web has gained much interest in recent years (Batinic, Reips, & Bosnjak, 2002; Birnbaum, 2000; Reips & Bosnjak, 2001). As is the case with every new method, questions of its advantages and disadvantages have arisen. Reips (2000, 2002b) suggests that, among other things, online research facilitates access to samples by reaching large populations from different countries and groups, enables the delivery of the experiment to the participant, produces high statistical power due to large sample size, and offers the direct assessment of motivational factors through analysis of dropout behavior in certain conditions.

---

We thank Tom Buchanan, Rich Carlson, an anonymous reviewer, and Jonathan Vaughan for their comments on previous drafts of the manuscript. Correspondence concerning this article should be addressed to M. Schulte-Mecklenbeck, Department of Psychology, University of Fribourg, Rue de Faucigny 2, 1700 Fribourg, Switzerland (e-mail: michael.schulte@unifr.ch).

However, it is also noted that with multiple submissions, it is possible for participants to do an experiment several times; there is a lack of experimental control (e.g., the experimenter has no influence on location and experimental setup); self-selection may be at work; and technical variance could occur, such as different access speeds and use of various operating systems or browsers.

**Validity of Web studies.** As we previously mentioned, conducting research on the Web has many distinct advantages that make it attractive to the researcher. Thus, it is necessary to find out whether or not the disadvantages actually are harmful and, if so, under what circumstances. In order to investigate the influence of measurement on the dependent variable, a study is run on the Web and compared with a study run in the laboratory. If there is no difference in the results, the Web version is considered to be valid.

Krantz and Dalal (2000) divided their review of the validity of Web studies into three types of research designs: surveys, correlational studies, and experimental studies. In a survey comparing traditional mail with e-mail delivery, Mehta and Sivada (1995) found that response rates were equivalent and mean responses comparable. In a correlational design study, Buchanan and Smith (1999) put a psychological test (the Self-Monitoring Scale; Gangestad & Snyder, 1985) on the Web. The authors ran a sample and compared the results with published results of traditional tests. Comparable internal consistency of the Web sample with the pen-and-paper test and a factorial analysis with excellent goodness-of-fit measures were the results. Finally, for experimental designs, two studies of ratings of attractiveness should be mentioned. Krantz et al. (1997) used line drawings as stimuli, whereas Epstein, Klinkenberg, Wiley, and McKinley (2001) developed the method further by using ratings of attractiveness of photographs. Both studies yielded comparable results between the laboratory and the Web groups.

### Risky Decision-Making Studies

In a risky decision-making task, a decision maker has to choose one of at least two alternatives. At least one of the alternatives has undesirable consequences, and the occurrence of these consequences is uncertain. For example, a patient is asked to decide whether he wants to take a drug that may have dangerous side effects or to undergo surgery. Undesirable consequences of the latter are presented as complications during surgery and the possibility that the surgery may not be successful. Other well-researched types of risky decisions are gambles, such as roulette, lotteries, and bets.

**Quasi-realistic risky tasks versus gambles.** Risky decisions have been investigated experimentally in decision theory for over 50 years. An impressive number of empirical studies has been conducted, and important theories have been developed (Lopes, 1995). According to the experimental results of classical decision theory, the central factors determining decision behavior are the subjective values of the outcomes and their subjective

probabilities. In generalizing the results of classical decision theory, one must consider that most of these experiments use simple gambles. These are presented, for example, in the forms of roulette wheels or urns with balls of different colors.

In recent years, a number of experiments with quasi-realistic risky tasks have been conducted and have shown that decision behavior in gambling situations cannot be generalized to nongambling decision-making tasks (Huber, Beutter, Montoya, & Huber, 2001; Huber & Macho, 2001; Huber, Wider, & Huber, 1997; Ranyard, Hinkley, & Williamson, 2001; Williamson, Ranyard, & Cuthbert, 2000). These studies used quasi-realistic choice tasks in which the decision maker is confronted with the description of a realistic situation or scenario with at least two alternatives. Previously conducted research has revealed that behavioral differences between choices in quasi-realistic risky tasks and choices among gambles include the following:

1. **Low interest in probabilities:** If the experimental setting allows the decision maker to choose which information he or she wants to examine, most decision makers are usually *not* interested in probability information (Huber et al., 1997).

2. **Use of *risk-defusing operators* (RDOs):** If decision makers realize that an otherwise positive alternative may lead to a negative outcome, they search for an RDO. This operator should reduce the risk involved and is used instead of probability information, for example (Huber et al., 2001).

An RDO is an action intended by the decision maker to be performed in addition to a specific alternative and is expected to decrease risk. RDOs are common in everyday risky decision situations (e.g., taking out insurance, getting a vaccination, making a backup copy of one's computer files, or wearing protective gear in order to avoid contact with a toxic substance). In most quasi-realistic tasks, the majority of decision makers actively search for RDOs. Finding an RDO has a distinct effect on choices. If a decision maker finds an RDO or an RDO is available, he or she chooses the risky alternative in question much more frequently than if no RDO is found.

To obtain insight into the process of information gathering, the method of *active information search* (AIS; Huber et al., 1997) was developed. The participant is first given a short description of the decision situation. Then, he or she can ask questions in order to obtain more information from the experimenter. The questions are recorded, and responses are given from a list of standardized answers. The participant may ask any and as many questions as he or she likes. Pilot studies are used to optimize the short description and to standardize the answers.

In addition to this basic version, a second AIS method has been studied: In the *list version*, the participant selects from a list of questions. The list version evokes more questions than does the basic version, but the distribution of question types—for example, questions

about the probability of the negative outcome, questions about the availability of risk-defusing actions—is equal (Huber et al., 2001).

**Laboratory- versus Web-based experiments.** There are only a few studies that have compared laboratory and Web experiments using decision-making tasks. Birnbaum (1999) and Anderhub, Müller, and Schmidt (2001) found small differences between laboratory and Web conditions using gambles as tasks. These differences were due to larger variances in the Web sample. Both studies attributed these results to differences between the two sample populations. Shavit, Sonsino, and Benzion (2001) used lotteries and controlled the population sample by using students in both the Web and the laboratory conditions. The participants were more risk seeking on the Web. The general choice behavior was nevertheless comparable.

Experiment 1 was intended to compare the use of two AIS-type tasks in the laboratory and on the Web. A computerized version of the basic and list versions is provided, thereby guaranteeing comparable experimental conditions at both locations.

## EXPERIMENT 1

### Method

**Participants.** The participants in the Web condition were recruited via e-mails to four German newsgroups (de.sci.sociologie, de.alt.umfragen, de.sci.psychologie, and de.sci.misc); the main topics of these groups are psychology, sociology, and science. From December 17, 2001 through February 17, 2002, 120 users accessed the experiment's Web page. The participants had to go through a registration process in which information regarding their e-mail address, sex, and age was requested. Immediately after registration, an e-mail with a link to the entrance page, a user name, and a password (which could be used only once) was sent to each participant. Forty-four participants (36.44%) had to be excluded due to incomplete data (see Dropouts in the Results section for details). The remaining 76 participants (32 female, 44 male) had a mean age of 28.73 years ( $SD = 4.21$  years).

The participants in the laboratory condition were recruited via the student e-mail list of the University of Fribourg. Every student attending the university is registered on this list, and its use is frequent. Psychology students were excluded from participation because decision making and online research are regular psychology lecture topics at the University. Twenty-nine participants (18 female, 11 male) with a mean age of 24.56 years ( $SD = 3.51$  years) replied to the initial message. Several appointment times were proposed by the experimenter, and the participants were asked to choose a convenient one. Every participant completed the experiment.

Twenty 6.75 coupons to an online bookstore were raffled among all of the participants as an incentive for participation.

Reips's (2000) multiple-site entry technique was used to obtain insight into the origin of the participants of the Web sample. The idea was to generate several entry pages and hyperlinks leading to the starting page. After the experimenter checked the log files of the Web server, an analysis of the participants from different locations was possible. The present experiment was announced in two different locations: newsgroups and a Google ad (see <http://adwords.google.com> for a description). Participant turnout totaled 120 for the newsgroups and, interestingly, 0 for the Google ad. It is surmised that since only a small number of page views was financed, the resulting presentation period was short.<sup>1</sup>

**AIS method on the Web: List and keyword versions.** In order to examine the research questions proposed, the AIS method of presenting information to the participants was used. This method does not affect information acquisition behavior in terms of distribution of question types between the list and keyword versions. Furthermore, it enables the experimenter to detect in which items of information the participant is interested.

Two versions (list and keyword) were differentiated. The list version is a computerized AIS version that lends itself to being implemented on the computer and, therefore, also on the Web. As an alternative method, we introduced the keyword version (see basic AIS version), in which the participant entered an individually selected keyword and was then confronted with a list of questions containing that keyword. The two versions are described below.

**List version.** The participants were presented with a list of questions (in hyperlink form) and selected one at a time by clicking. The questions were formulated according to category: *probability* (questions concerning the probability of the occurrence of an event), *new alternative* (questions concerning options not presented by the experimenter), *control* (questions dealing with the decision maker's control over the external event or over negative consequences), *worst-case plan* (questions demanding information about what can be done in case a negative event occurs), *situation* (questions asking for general background information not linked to a specific alternative) and *consequences* of a certain alternative. Control and worst-case questions are instances of questions for RDOs. The participants could click on as many questions as they wished.

The Appendix can be consulted for a list of sample questions and answers. A list of all the questions and the prepared answers for all tasks used can be obtained from the authors (in German) upon request.

**Keyword version.** In the keyword version, the participant entered a keyword into a database (similar to using a search engine) and decided which information from the resulting list was interesting. The participants searched through the same questions used in the list version, except that now they accomplished this by entering a keyword. A list of corresponding questions in hyperlink form was presented; when a participant clicked on a question of interest, the answer was shown.

**Decision-making tasks.** A warm-up task and two main tasks were used in the experiment. As the warm-up, the *post office task* was used. The two main tasks were the *ticks task* and the *mechanics task*. The basic structure of all three tasks was the same. (See the Appendix for an English translation of the texts of the three tasks.)

**Design.** In this study, three independent variables were examined: location (laboratory vs. Web), version (list vs. keyword), and task (*ticks* vs. *mechanics*). The independent variables of location and version were varied between participants, whereas task was varied within participants. The procedure was exactly the same for both the laboratory and the Web conditions. Both groups were instructed via the same Web pages and interacted with the database in exactly the same manner. The laboratory group participated in the experiment in a room at the Department of Psychology of the University of Fribourg, where a computer was provided which had access to the database over the intranet of the university. An experimenter was present during the entire experimental session.

The Web group accessed the experiment via the Web through a link provided in a postregistration e-mail.

**Apparatus.** A Pentium I (133 MHz, 32Mb RAM) with a LAMP (Linux Apache MySQL PHP) environment was installed. Linux Mandrake 6.2 was used as an operating system to host the experiment for both the laboratory and the Web versions. MySQL 3.23.41 served as database, and PHP 4.0.7 generated the necessary Web pages. An Apache Web Server 1.3.20 connected the components to the Web. Note that LAMP provides a server-sided solution that needs only a browser on the client's (i.e., the participant's) side. No further installation of software is necessary; therefore, compatibil-

**Table 1**  
**Number of Questions and Words for the Decision-Making Tasks**

Task	Questions	Words
Post	39	390
Ticks	89	932
Mechanics	106	1,228

ity problems are rare. The use of a database to record the participant's clicks is an elegant method of data collection, because easy data access is assured. The participants were randomly assigned to the versions through a PHP query after registration.

The list version was generated via a list of all the questions used in the experiment. (See Table 1 for details on the number of questions and number of words relevant for the text search in the keyword version for each task.) In the keyword version, a full text search through the database of possible questions was used for information gathering.

All actions taken during the experiment were recorded and written to the database with a *timestamp* (current time of an event that is recorded by a computer) for each individual step.

## Results

**Dropouts.** Analysis of dropouts is an important tool in online research. In the current experiment, the laboratory and Web conditions yielded different dropout behaviors. A dropout in this experiment was defined as someone who had registered (by submitting an e-mail address and receiving a password) and who did not start the experiment, had started but not finished all the tasks of the experiment, or just lurked through the experiment without clicking on any question. Forty-four (36.44%) participants dropped out of the Web sample; significantly more dropped out of the keyword version (30) in comparison with the list version [14;  $\chi^2(1, N = 44) = 5.82, p = .016$ ]. By contrast, the laboratory sample had no dropouts at all.

**Clicks.** In this study, a click was defined as the clicking on a hyperlink in order to obtain information about a question of interest. This measure was used for both the list and the keyword versions. Clicks were analyzed using an ANOVA with location (laboratory vs. Web) and version (list vs. keyword) as between-subjects factors. Task (*ticks* vs. *mechanics*) was varied as a within-subjects factor. Task yielded no significant difference—the *ticks* and *mechanics* tasks resulted in comparable amounts of information. Location was significant [ $F(1,74) = 14.66, p < .001$ ], indicating that more items were clicked on in the laboratory condition ( $M = 12.03, SD = .87$ ) than in the Web condition ( $M = 7.91, SD = .64$ ; see Figure 1). Version was also significant [ $F(1,74) = 4.62, p = .035$ ]. More items were clicked on in the list ( $M = 11.13, SD = .78$ ) than in the keyword condition ( $M = 8.81, SD = .74$ ).

In addition, a significant location  $\times$  version interaction was found for the laboratory condition [ $F(1,74) = 6.88, p = .011$ ]. A post hoc analysis of the laboratory condition revealed a significant difference between versions [ $t(52) = 3.01, p = .004$ ]. More items were accessed in the list than in the keyword version. No effect was found for the Web condition.

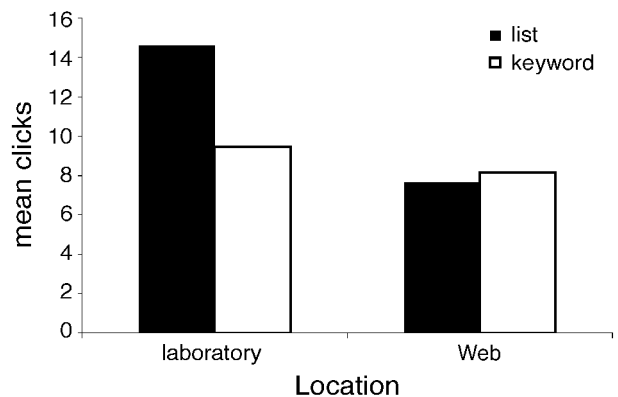
In the keyword version, no differences between laboratory and Web were found. Methodological problems may account for this. For the keyword version, the number of dropouts was larger on the Web than in the laboratory. It is important that a significant difference in the number of keywords entered was found between the two locations (laboratory:  $M = 7.50, SD = 5.68$ ; Web:  $M = 4.41, SD = 3.70$ ;  $t(94) = 3.03, p = .003$ ). Although the participants in the laboratory sample entered nearly twice as many keywords to gain information, there was no difference in the number of clicks (see Figure 1).

**Categories.** On the basis of Huber et al. (2001), six categories of information were formed for this analysis: *probability, new alternative, control, worst-case plan, situation, and consequences* (see Table 2 for descriptive statistics). An ANOVA with location (laboratory vs. Web), version (list vs. keyword), and category as between-subjects factors was conducted. Clicks served as a dependent variable. A significant main effect was found for categories [ $F(1,900) = 98.83, p < .001$ ].

A post hoc Scheffé test revealed that the participants used the categories *situation* and *consequences* more often than the other categories (see Table 2). These results are in accordance with the findings of Huber et al. (1997; Huber et al., 2001), confirming that probability information is of less interest than situational and consequential information.

A second analysis of categories was performed in order to obtain a measure of chosen categories unaffected by the number of information items examined. The dependent variable was the proportion of information searched for in each category by the participants.  $\chi^2$  tests on the difference between locations were performed. No significant difference could be found in any of the categories.

**Time.** Each individual click in the experiment was automatically attached to a timestamp. This procedure makes it possible to calculate completion time for the tasks. Time was analyzed by an ANOVA with location (laboratory vs. Web) and version (click vs. search) as between-



**Figure 1. Experiment 1: Mean number of clicks as a function of location and version.**

**Table 2**  
**Descriptive Statistics for Clicks in**  
**Location  $\times$  Version  $\times$  Category in Experiment 1**

Version	Category	Location			
		Laboratory		Web	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
List	Probability	1.61	1.31	0.57	0.87
	New alternative	0.68	0.90	0.24	0.43
	Control	1.46	1.67	0.76	0.83
	Worst-case plan	0.86	0.85	0.57	0.65
	Situation	5.96	2.73	3.22	2.24
Keyword	Consequences	4.50	2.63	2.22	1.80
	Probability	0.93	1.27	0.77	0.98
	New alternative	0.26	0.53	0.03	0.18
	Control	0.56	1.05	0.52	1.05
	Worst-case plan	0.56	0.64	0.40	0.69
	Situation	4.00	3.34	3.26	3.89
	Consequences	3.15	2.63	3.26	3.27

subjects factors. Location was significant [ $F(1,153) = 37.03, p < .001$ ], indicating that more time was used to complete the task in the laboratory ( $M = 367.3, SD = 22.7$  sec) than on the Web ( $M = 192.9, SD = 17.4$  sec). No effect was found for version.

## Discussion

These results suggest that the location where the research was conducted had an effect on the number of items considered relevant. The participants in the Web condition did not use as much information as did those in the laboratory. This finding contradicts previous comparison studies that showed no difference between the laboratory and the Web (Buchanan & Smith, 1999; Epstein et al., 2001; Krantz et al., 1997; Krantz & Dalal, 2000; Mehta & Sivada, 1995). However, this effect was limited to the list version. The keyword version showed a ceiling effect on number of information items looked at, with certain hints of methodological problems in information acquisition.

The differences observed in Experiment 1 can be interpreted as either an influence of measurement between Web and laboratory conditions or merely a difference between Web and laboratory participant samples.

In Experiment 1, the influences of these two factors cannot be separated. This problem is inherent to most validity studies using Web methodology.

The fact that in a study a difference between Web and laboratory conditions was revealed does not prove that there is a reliable difference due to administration conditions if the samples are different. A second experiment was conducted to overcome this confound.

## EXPERIMENT 2

The goal of Experiment 2 was to solve the problem of different samples. This was done by assigning participants from a single population randomly to the Web and laboratory conditions (see Shavit et al., 2001). The ex-

perimental design and the decision-making tasks from Experiment 1 were used again.

## Method

**Participants.** The participants were recruited via an announcement at the University of Fribourg. Eighty students volunteered and were randomly allocated to either the laboratory or the Web condition.

An e-mail was sent to 40 students requesting their participation in the Web condition. Of these, 32 students (19 female, 13 male) with a mean age of 23.75 years ( $SD = 4.64$ ) responded and completed the experiment. This represents a dropout rate of 20% for the Web condition (list version, 3; keyword version, 5; n.s.).

In the laboratory condition, several appointment times were proposed to the 40 students (26 female, 14 male). Every participant completed the experiment. The participants in this sample had a mean age of 24.80 years ( $SD = 4.54$  years).

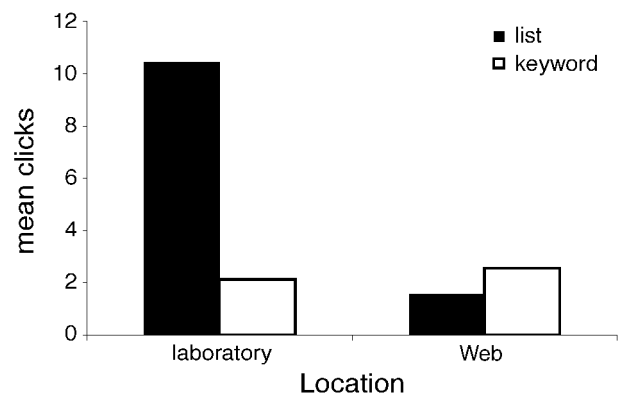
Each participant received course credit for participation.

**Design, Apparatus, and Procedure.** The method, tasks, design, and apparatus used in Experiment 1 were used in this experiment as well.

## Results

**Clicks.** An ANOVA with location (laboratory vs. Web) and version (list vs. keyword) as between-subjects factors was conducted. Task (*ticks* vs. *mechanics*) was used as a within-subjects factor. Task yielded no significant difference; the *ticks* task and the *mechanics* task resulted in comparable interest in information. Location was significant [ $F(1,252) = 198.06, p < .001$ ], indicating that more items were clicked on in the laboratory condition ( $M = 6.27, SD = .21$ ) than in the Web condition ( $M = 2.06, SD = .22$ ; see Figure 2). Version was also significant [ $F(1,252) = 149.62, p < .001$ ]. More items were clicked on in the list ( $M = 5.99, SD = .10$ ) than in the keyword ( $M = 2.34, SD = .23$ ) condition. In addition, a significant interaction between location and version [ $F(1,252) = 239.99, p < .001$ ] was found. A post hoc analysis of the laboratory condition revealed a significant difference between versions [ $t(199) = 19.81, p < .001$ ]. More items were accessed in the list than in the keyword version. No effect was found for the Web condition.

An analysis of entered keywords in the two locations demonstrated a larger number of keywords entered in the



**Figure 2. Experiment 2: Mean number of clicks as a function of location and version.**

**Table 3**  
**Descriptive Statistics for Clicks in**  
**Location × Version × Category in Experiment 2**

Version	Category	Location			
		Laboratory		Web	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
List	Probability	1.24	1.26	0.41	0.61
	New alternative	0.42	0.50	0.26	0.45
	Control	0.84	0.85	0.68	0.81
	Worst-case plan	0.15	1.68	0.59	0.66
	Situation	3.87	2.63	3.15	2.23
Keyword	Consequences	3.51	1.65	2.18	1.71
	Probability	0.85	1.18	0.80	1.04
	New alternative	0.21	0.48	0.04	0.20
	Control	0.45	0.97	0.57	1.14
	Worst-case plan	0.48	0.62	0.45	0.73
	Situation	4.94	4.29	2.57	2.91
	Consequences	3.03	2.48	3.10	3.05

laboratory condition ( $M = 5.68$ ,  $SD = 3.85$ ) in comparison with the Web condition ( $M = 3.27$ ,  $SD = 2.88$ ). This significant difference [ $t(118) = 1.98$ ,  $p = .05$ ] emerged although there was no difference in the number of clicks.

**Categories.** An ANOVA with location (laboratory vs. Web), version (list vs. keyword), and categories (*probability vs. new alternative vs. control vs. worst-case plan vs. situation vs. consequences*) as between-subjects factors was also conducted. A significant main effect was found for categories [ $F(1,852) = 109.91$ ,  $p < .001$ ].

A post hoc Scheffé test resulted in significant, more frequent use of the *situation* and *consequences* categories in comparison with the other categories (see Table 3). These results are in accordance with the findings of Huber et al. (1997; Huber et al., 2001), demonstrating that probability information is of less interest than situational and consequential information.

A second analysis of categories was performed in order to obtain a measure of chosen categories unaffected by the number of information items looked at. The dependent variable was the proportion of information searched for in each category by the participants. Note that this analysis works with a proportion of hits in a certain category and not with the absolute number of hits. This assumption enables the calculation despite the small number of clicks in three of the four conditions (see Figure 2).  $\chi^2$  tests on the difference between locations were performed. In no category was there a significant difference.

**Time.** Time was analyzed by an ANOVA with location (laboratory vs. Web) and version (click vs. search) as between-subjects factors. Location was significant [ $F(1,67) = 27.99$ ,  $p < .001$ ], indicating that more time was used to complete the tasks in the laboratory ( $M = 292.25$ ,  $SD = 20.41$  sec) than on the Web ( $M = 144.23$ ,  $SD = 19.13$  sec). No effect was found for version.

## Discussion

Because the online population is steadily growing and consequently becoming more and more diverse, it is dif-

ferent from the student population. Not only do age, sex, and educational distributions differ, but, for the most part, economic well-being also varies. The argument that measuring different populations (laboratory and Web) leads to differences in results is a common one.

The purpose of Experiment 2 was to overcome the confounding factors between location and population sample that were present in Experiment 1. It was shown that the control of population does not change the general result of differences between locations. Therefore, the interpretation that the laboratory and Web conditions have a central effect on information acquisition appears to be appropriate.

## GENERAL DISCUSSION

The purpose of this study was to investigate differences in information acquisition behavior with data collected in the laboratory and on the Web. The main results of our experiments can be summarized as follows:

1. Differences emerged in the amount of information gathered between the Web and the laboratory conditions. These cannot be explained by sample differences. Using the same sample (students), the results of Experiment 1 were replicated in Experiment 2.

2. The participants in the list version of the AIS method searched for more information than did those in the keyword version (in the lab). On the Web, such differences were not found.

3. Different dropout behavior is found on the Web. Fewer participants dropped out of the experiment in the list version than in the keyword version. The lab condition showed no dropouts at all.

4. If we look at the distribution of question categories as a dependent variable, we see no differences (i.e., no interaction) between locations and versions. This result is also valid if we use the percentage of participants who asked at least one question in the respective category.

### Why Do Participants Behave Differently in the Laboratory and on the Web?

The presented literature on comparisons between laboratory and Web suggests that there should be no difference between the two research locations. There is a crucial difference between those studies and ours—namely, the response mode of the participants. In a pen-and-paper or Web-based questionnaire, participants can estimate the duration of the task and the number of answers requested. In an information search task, participants are expected to search as long as necessary for *them* to complete the task. No hint of what constitutes a good amount of search depth is given; participants decide on their own, and their decisions are different when the experiment is run on the Web.

In combination with this, a second critical factor is the presence of an experimenter in the laboratory. There are at least two possible explanations, which are not mutu-

ally exclusive: (1) The experimenter is an authority figure for the participant and, in his or her presence, the participant is more inclined to make an effort; and (2) the presence of the experimenter signals for the participant that the experiment is important; therefore, the participant is motivated to search for more information. Approximately twice the number of entered keywords were measured in the laboratory in comparison with the Web.

For future studies, a laboratory condition with an experimenter leaving the participant on his or her own during the experiment could be of interest.

### **Why Do People With the List Version Search for More Information Than Do Those With the Keyword Version (in the Laboratory)?**

In the list version, a given list of questions is scanned by the participants, and questions of interest are chosen. In the keyword version, the task is as follows: Think of a keyword, write it into the appropriate field, scan through the results, and choose an information item of interest. Cognitive load and, therefore, effort made in completing a task are obviously higher for the keyword version. This is recognized as one reason for differences.

In addition, methodological problems seemed to arise in the keyword version. Using a common search engine on the Web enables one to find results for nearly every possible keyword. This is because the number of stored pages reaches the billions on these search engines. In our experiment, the search space was limited to the number of questions in the database (an average of 98). The participants had to learn which keywords yielded results and which did not. Search behavior was therefore influenced by the number of successful keywords.

A hint of problems in the keyword version is also given by the number of dropouts, which was much larger than in the list version.

An improvement of the keyword version seems important. This could be accomplished by enlarging the database of questions and using smarter search devices than a full text search.

### **Why Does the Dropout Behavior Differ Between Locations and Versions?**

Dropout is a feature of Web experimenting that should be controlled and reduced (Reips, 2002b). Nevertheless, a positive aspect of dropout is the insight into possible confounding of experimental conditions. In our experiment, we minimized the dropout rate by asking for personal information and giving information about incentives at the beginning of the experiment (see Frick, Bächtiger, & Reips, 2001, for experimental manipulation of these two factors). Nevertheless, the two locations—laboratory (no dropouts) and Web (36% dropout rate in Experiment 1, 20% in Experiment 2)—resulted in different dropout behaviors. Two possible explanations, which could be confounded, are suggested. The participants had problems in the keyword version be-

cause information search was not always immediately successful (see Reips, 2002a, for a discussion of the detection of motivational confounding). The presence of an experimenter in the laboratory could be a second reason: The participants were socially “trapped” and, therefore, forced to finish the experiment. It does not seem appropriate to leave an experiment when another person (e.g., the experimenter) is present.

### **For Which Dependent Variable is the Difference Between Laboratory and Web Relevant?**

If one is interested in the *absolute* amount of information gathered, there are certain differences between the two locations. Nevertheless, it is not clear from these two experiments which of the two locations is closer to a *real value*.

Often, however, the researcher is interested not in the absolute amount of information, but in the relative amount of searching for different categories and changes in the relative frequency of categories in different experimental conditions. As our results show, the distribution of information search over categories is not affected by the laboratory–Web variation.

We investigated the difference between laboratory and Web experiments using a decision-making task as an example. However, our results are also relevant to other areas of psychological research in which information search behavior is of interest. Examples are the areas of complex problem solving, mental models, knowledge representation, and the evaluation of Web page content.

A scenario for the last point could be the following: The aim is to communicate information about a risky situation (e.g., breast cancer) on a Web page. An AIS experiment is run in order to locate types of information central to the participants. Finally, an analysis of these categories and a comparison with *objective* important information should reveal deficiencies in information gathering. With emphasis on this missing information, an improvement of information presentation should be possible.

To recapitulate, our experiments reveal that there is no simple answer to the question of whether or not a Web study is acceptable for information search tasks. Even if we find distinct differences in search behavior between the laboratory and the Web, the answer is highly dependent on the behavioral aspects in which the researcher is interested. A promising point is that the type of information demanded in a Web study seems to be comparable with that demanded in the laboratory study. Similar information is requested whether or not the experimenter is present.

### **REFERENCES**

- ANDERHUB, V., MÜLLER, R., & SCHMIDT, C. (2001). Design and evaluation of an economic experiment via the Internet. *Journal of Economic Behavior & Organization*, *46*, 227-247.
- BATINIC, B., REIPS, U.-D., & BOSNJAK, M. (Eds.) (2002). *Online social sciences*. Seattle: Hogrefe & Huber.
- BIRNBAUM, M. (1999). Testing critical properties of decision making on the Internet. *Psychological Science*, *10*, 399-407.

- BIRNBAUM, M. (2000). *Psychological experiments on the Internet*. San Diego: Academic Press.
- BUCHANAN, T., & SMITH, J. L. (1999). Using the Internet for psychological research: Personality testing on the World-Wide Web. *British Journal of Psychology*, **90**, 125-144.
- EPSTEIN, J., KLINKENBERG, W., WILEY, D., & MCKINLEY, L. (2001). Insuring sample equivalence across Internet and paper-and-pencil assessments. *Computers in Human Behavior*, **17**, 339-346.
- FRICK, A., BÄCHTIGER, M. T., & REIPS, U.-D. (2001). Financial incentives, personal information, and drop out in online studies. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 209-219). Lengerich, Germany: Pabst Science.
- GANGESTAD, S. W., & SNYDER, M. (1985). To carve nature at its joints: On the existence of discrete classes in personality. *Psychological Review*, **92**, 317-340.
- HUBER, O., BEUTTER, C., MONTOYA, J., & HUBER, O. W. (2001). Risk defusing behaviour: Towards an understanding of risky decision making. *European Journal of Cognitive Psychology*, **13**, 409-426.
- HUBER, O., & MACHO, S. (2001). Probabilistic set-up and the search for probability information in quasi-naturalistic decision tasks. *Risk Decision & Policy*, **6**, 1-16.
- HUBER, O., WIDER, R., & HUBER, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, **95**, 15-29.
- 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. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 36-60). San Diego: Academic Press.
- LOPES, L. L. (1995). Algebra and process in the modeling of risky choice. In J. Busemeyer, R. Hastie, & D. L. Medin (Eds.), *Decision making from a cognitive perspective* (pp. 177-220). San Diego: Academic Press.
- MCGRAW, K. O., TEW, M. D., & WILLIAMS, J. E. (2000). The integrity of Web-delivered experiments: Can we trust the data? *Psychological Science*, **11**, 502-506.
- MEHTA, R., & SIVADA, E. (1995). Comparing response rates and response content in mail vs. electronic mail surveys. *Journal of the Market Research Society*, **17**, 429-440.
- MUSCH, J., & KLAUER, K. C. (2002). Psychological experimenting on the World Wide Web: Investigating content effects in syllogistic reasoning. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences*. Göttingen: Hogrefe.
- POHL, R. F., BENDER, M., & LACHMANN, G. (2002). Hindsight bias around the world. *Experimental Psychology*, **49**, 270-282.
- RANYARD, R., HINKLEY, L., & WILLIAMSON, J. (2001). Risk management in consumers' credit decision making. *Zeitschrift für Sozialpsychologie*, **32**, 152-161.
- REIPS, U.-D. (2000). The web experiment method: Advantages, disadvantages, and solutions. In M. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-114). San Diego: Academic Press.
- REIPS, U.-D. (2002a). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review*, **20**, 241-249.
- REIPS, U.-D. (2002b). Standards for Internet-based experimenting. *Experimental Psychology*, **49**, 243-256.
- REIPS, U.-D., & BOSNJAK, M. (2001). *Dimensions of Internet science*. Lengerich, Germany: Pabst Science.
- SHAVIT, T., SONSINO, D., & BENZION, U. (2001). A comparative study of lotteries-evaluation in class and on the Web. *Journal of Economic Psychology*, **22**, 483-491.
- WILLIAMSON, J., RANYARD, R., & CUTHBERT, L. (2000). Risk management in naturalistic insurance decisions: Evidence from a simulation study. *Risk Decision & Policy*, **5**, 19-38.

#### NOTE

1. In 2001, when the experiment was conducted, Google ads were pay-per-view; Google (<http://adwords.google.com>, retrieved February 5, 2003) has since changed to a pay-per-click policy, which should increase the effectiveness of invested money.

---

## APPENDIX

### Examples of Tasks

#### Post Office

Imagine that you are the head of the post office in the village H. The post office has very cramped conditions and has been faced with the following problem for several years: In November and December, it has to handle many parcels. If the number of parcels to be handled is too large, conditions may become unbearable. The village administration offers you the opportunity to rent the local meeting hall.

#### Ticks

Imagine that you are the director of a pediatric allergy center. It is uncertain whether the center can stay in the rented location, because the owners may sell the building. If this sale take place, the center will have to move out and the director will have to find another location that meets the requirements for treating allergies. You have just now been offered a big house in the forest. The problem is that the forester's house stands in a wood, which is contaminated with a specific kind of ticks.

#### Machine

Imagine that you are the owner of a small factory that produces a kind of weatherproof credit card for ski lifts. The order books are so full that an additional high-capacity machine must be bought to meet the great demand. You can choose between two machines. They differ in their technology and in how prone they are to interruptions caused by breakdowns.



## APPENDIX (Continued)

**Table A1**  
**Sample Questions and Answers for the *Mechanics* Task**

Category	Question	Answer
Probability	What's the probability of a new contract if I buy the machine?	The probability of a new contract is 50%.
New alternative	Can I get expert statements on the two machines?	No experts are available at the moment.
Control	Do I have influence over future orders of my factory?	Future orders depend on the success of current assignments.
Worst-case plan	Can I get insurance against financial loss if one machine breaks down?	Yes, there is insurance against financial loss, but premiums are high.
Situation	What does the factory produce?	The factory produces mechanical parts for watches, electric motors, and electronic microcomponents.
Consequences	What are the consequences if I agree to buy a new machine?	Your employees will be able to work with the latest available technology.

(Manuscript received November 19, 2002;  
revision accepted for publication February 26, 2003.)