On the Challenges in Usable Security Lab Studies: Lessons Learned from Replicating a Study on SSL Warnings

Andreas Sotirakopoulos
University of British Columbia
Vancouver, BC, Canada
andreass@ece.ubc.ca

Kirstie Hawkey
Dalhousie University
Halifax, NS, Canada
hawkey@cs.dal.ca

Konstantin Beznosov
University of British Columbia
Vancouver, BC, Canada
beznosov@ece.ubc.ca

ABSTRACT

We replicated and extended a 2008 study conducted at CMU that investigated the effectiveness of SSL warnings. We adjusted the experimental design to mitigate some of the limitations of that prior study; adjustments include allowing participants to use their web browser of choice and recruiting a more representative user sample. However, during our study we observed a strong disparity between our participants actions during the laboratory tasks and their self-reported "would be" actions during similar tasks in everyday computer practices. Our participants attributed this disparity to the laboratory environment and the security it offered. In this paper we discuss our results and how the introduced changes to the initial study design may have affected them. Also, we discuss the challenges of observing natural behavior in a study environment, as well as the challenges of replicating previous studies given the rapid changes in web technology. We also propose alternatives to traditional laboratory study methodologies that can be considered by the usable security research community when investigating research questions involving sensitive data where trust may influence behavior.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Human Factors, Security, Experimentation

Keywords

Usable security, SSL warnings, experimental design, study environment bias, study replication

1. INTRODUCTION

The need for secure communication over the Internet has led to the development of the Secure Socket Layer (SSL) and later the Transport Layer Security (TLS) protocols. In the typical implementation of both protocols, the server is authenticated by the client using its public key. For this purpose, a certificate that binds the server's public key with its identity is used. Web browsers check the certificate that the server supplies during the initialization of the secure communication. If the certificate is invalid or does not appear to be genuine, a warning message is raised. Warnings are often raised due to a mismatch between the host name of the web server and the one in the certificate, the expiration of the certificate, or an unknown/untrusted certificate issuer. An invalid certificate might be the result of a man-in-the-middle-attack or a DNS spoofing attack. However, in most cases the warnings are raised on legitimate sites that have a mis-configured, expired, or self-signed certificate. Browsers display the warnings to the users in order to alert them to the potential danger of falling victim to an attack.

Multiple studies have indicated that displaying warnings to the user should be done with caution and in a clear, easy to understand way, avoiding false positives [17, 21]. Otherwise, users will learn to disregard them [22, 7]. Because many legitimate sites that employ a secure protocol connection have errors with their certificates, it is important to learn how these errors have affected users' perceptions and reactions to SSL warnings.

The purpose of our study was to investigate the effectiveness of SSL warnings by validating and extending a study by Sunshine et al. [18], which was conducted at CMU in 2008 (referred to here as the "CMU study"). By introducing new parameters and addressing limitations of the CMU study, we aimed to understand not only how users react to SSL warnings but also the reasons behind their reactions. Furthermore, we wanted to investigate how factors, such as ecological validity and population sampling, play out in studies that seek to investigate human behavior and perception in regards to computer security and security practices. We also believe that validating findings of previous studies as well as examining users' interactions from a slightly different angle add value to the overall reliability of the findings, and this will enable the research community to better understand the phenomena at hand.

As was done in the CMU study, we conducted a laboratory experiment with 100 participants assigned to one of 5 conditions; in our case this assignment was made based on the browser they used, while in the CMU study, the assign-

Symposium on Usable Privacy and Security (SOUPS) 2011, July 20–22, 2011, Pittsburgh, PA USA

ment was random. Participants performed four information seeking tasks; during two of these tasks, we arranged for SSL warnings to be displayed. As in the CMU study, we removed from the web browser those trusted certificate authorities who were responsible for issuing the certificates for Hotmail as well as all the major banks. This resulted in self-signed certificate warnings to be raised during some of the tasks. We limited our study to self signed certificate warnings, as the warning generation method was reliable and easy to implement.

The tasks during which the warnings were raised required participants to use their personal and sensitive information (e.g., their name and address during the Hotmail account creation task and their actual bank account information during the banking task). As such, we expected them to consider the possible implications of their actions during the study. We recorded the participants' reactions to the certificate warnings and inquired about their perceptions of and reactions to such warnings during an exit questionnaire.

One key difference between our experimental design and that of the CMU study was that our participants constituted a more representative user sample. Instead of recruiting participants from a narrow population of university students or employees, we recruited a sample that was more diverse in terms of age, level of education, and occupation. We made this change in order to investigate whether a more representative sample increases the reliability of results. A second key difference was that we assigned users to their usual web browser in order to reduce the surprise effect that may occur when users are assigned to interact with an unfamiliar browser. A third key difference between the two studies was the recruitment method (described in detail in Section 3.3), which allowed us to observe that a portion of our potential participants chose to not participate in our study due to security concerns upon learning that they would have to use their personal data during the study.

Our data reveals interesting findings. Firstly, there were observable differences in participants' behavior between the two studies. In the CMU study, participants' reaction to the warnings differed depending on the condition (browser/warning) that they were assigned to. This was not the case in our study, where we observed no statistically significant differences between conditions. We attribute this mostly to the fact that when the CMU study began, Firefox 3 (FF3), which was used as the browser in one of the conditions in both studies, had just come out (June 17, 2008). The participants in the CMU study were unfamiliar with FF3's new interface for certificate warnings, making it difficult for them to perform the necessary steps of adding an exception, "trusting" the site, and continuing past the warning.

Secondly, we did not observe any statistically significant differences between how student and non student participants performed. This surprising finding may be of particular interest to the usable security research community due to the common concern that studies that recruit participants solely from student populations, rather than from a more difficult to recruit broader population, may not be generalizable to the broader population. Our study failed to confirm that the use of a student population is an issue.

Finally, we investigated further the factors that contribute to the perception of warnings. In contrast to the CMU study, where the redesigned warning completely altered the native warning including its layout, in our study we only changed the wording and intensified the coloring of the browser's native warning, while maintaining its general layout. While in the CMU study the redesigned warning proved very effective at preventing users from proceeding, in our study there were no statistically significant differences between our custom warnings and the browsers' native ones. Our inability to confirm the CMU findings suggests that wording and coloring are not the factors that contribute the most to the perception of warnings. It remains to be seen whether changing just the general layout of the warning and its novelty would capture users' attention and elicit a more cautious approach to the web site that raised it.

Some of our most significant findings, however, have to do with the reasons that participants gave for their reactions to the warnings that they encountered during the study tasks. These findings are applicable to the experimental design of similar usable security studies, where user behavior is under investigation. As described in more detail in Sections 4 and 5, one third of our participants claimed that their reaction would be different if they were not in a study environment and did not have the reassurance from the study environment (e.g., ethics board approval, the university as a reputable organization) that their information would be safe and secure. Although this potential for bias has been supported by anecdotal evidence in the usable security community, to the best of our knowledge, our study is the first to provide evidence, in the form of self-reported data, about the impact of the environmental bias that a laboratory environment introduces.

In summary, the major contributions of our work are as follows:

1. Validation of the findings of the previous study as well as examining participants' interactions from a slightly different angle (i.e., adding ecological validity by assigning participants to browser of preference).

2. Evidence of behavioral changes of participants towards SSL warning mechanisms employed by browsers as time passes (i.e., differences in reactions of participants towards the FF3 SSL warnings between the two studies).

3. Evidence of a strong bias of the laboratory environment for usable security studies that require participants to use their own data in the study environment, which may cause them to act differently than if they were using the same data at home (i.e., participants' reasons for their reaction to the warnings).

The rest of the paper is organized as follows. In Section 2, we present background and related work. In Section 3, we discuss in detail our methodology and its differences with the CMU approach. In Section 4, we present our results, followed by discussion in Section 5. We conclude in Section 6.

This study was approved as a minimal risk study by the University of British Columbia's (UBC) Behavioural Research Ethics Board (BREB).

2. BACKGROUND AND RELATED WORK

Extensive research has been conducted on the perception and reaction of users to SSL security indicators and warnings. In 2005, Whalen et al. [20] conducted a study using an eye-tracker to investigate whether participants were paying

attention to web browser security indicators. Participants were not observed to pay any attention to the security indicators, such as the lock icon; it was only after they were primed to pay attention to security that the majority of participants took notice of the lock icon. Schechter et al. [16] found that even when security indicators have been removed from information sensitive web sites, participants will provide sensitive information to those sites.

Technologies have been developed to help users defend against malicious web sites, and most modern web browsers employ them. Support for extended validation (EV) certificates and active SSL warnings are two of those. Using an EV certificate, major browsers, such as Internet Explorer 7 (IE7) and FF3, display the certificate owner's name in the address bar and also paint the address bar green. These indicators, however, are still susceptible to phishing attacks using picture-in-picture attacks as shown in a study by Jackson et al. [11]. In that study, many participants were misled by spoofed browser chrome that labeled fake web sites as legitimate. The findings about the inability of EV certificate related security indicators to provide an adequate defensive mechanism are backed up by a study conducted by Sobey et al. [19] in 2008. They used an eye-tracker to observe whether participants paid attention to the EV certificate security indicators. None of the 28 participants clicked on the EV certificate indicator of FF3's address bar, and its existence did not affect their decisions in using online shopping web sites in the study. Active warnings and indicators related to browser security, which force the user to take action in order to overcome the warnings, have been recently introduced in all major browsers. Although they are harder for the user to disregard than passive indicators, research has found that they are often ignored by users [5] and that users eventually get used to them and disregard them [7]. Furthermore, studies like the one done by Biddle et al. [2] have shown that improvements could be made in the warnings (wording and layout) so as to convey the intended message to the user in a clearer, more understood way.

Much of the research on the effectiveness of web browser security warnings has taken place in laboratory studies, which raises concerns that a bias from the settings may exist in these studies [20]. Furthermore, participants tend to be a skewed population of mostly university students [18, 19, 20], a group characterized as WEIRD (Western, Educated, Industrialized, Rich, and Democratic) by Henrich et al. [10]. This raises concerns about the generalizability of results based on such a population. Because usable security studies often seek to obtain insight into the average users' security concerns and behaviour, both realism and a representative sample are important.

3. METHODOLOGY

We next describe the design challenges we faced and the steps we took to address them. We then describe our study design, the recruitment process, the study tasks, and the exit questionnaire. Throughout, we highlight and discuss the differences between our and CMU study protocols. While designing our study, we considered several alternatives to our final approach, e.g., employing deception more extensively, having the study taking place outside of the lab, utilizing software and hardware to record participants reaction more accurately, etc. Although we do agree that such alternatives may have led to improvements in the external validity of the study, we decided to keep the basic structure of the study as close as possible to that employed in the CMU study that we were aiming to replicate.

Figure 1: CMU custom warning

3.1 Methodological Challenges

There were several broad challenges that we had to consider as we designed our study. The first broad challenge was that we did not want our participants to be primed for security so did not want to reveal our study purpose. We therefore took care to obfuscate the purpose both during the recruitment phase (as detailed in section 3.3) and during our interactions with participants in the study session (as detailed in section 3.4).

A second broad challenge was to mitigate some of the limitations of the prior CMU experimental design, both those acknowledged by the CMU researchers and those that we felt should be addressed. The first limitation we identified was that participants in the CMU study were drawn almost exclusively from the CMU student body. A second limitation was that their participants were randomly assigned to the browsers investigated, which might have caused them to alter their normal behavior and become more cautious about SSL warnings if the warning interface was unfamiliar to them. A third limitation was that in the CMU study, the custom warnings designed for IE7 (Figure 1) were radically different in colors, wording, and layout from the native IE7 warnings (Figure 2). We believe that this might also have contributed to participants being surprised, thereby eliciting a more cautious reaction to the warnings. It should be noted here that our study design, although it controlled for, and reduced, the surprise a warning may cause to a user familiar with a particular browser, it may have introduced a new surprise element for the our redesigned warnings. Namely, a participant that was familiar with the warnings of a particular browser might be alerted by our redesigned warnings. Our study did not have enough participants to identify which of the two designs (CMU's or ours) had a bigger surprise effect for our participants.

In an effort to mitigate these limitations, we recruited participants from the broader Vancouver population, instead of limiting ourselves to UBC students. We assigned participants to our conditions according to the browser they nor-

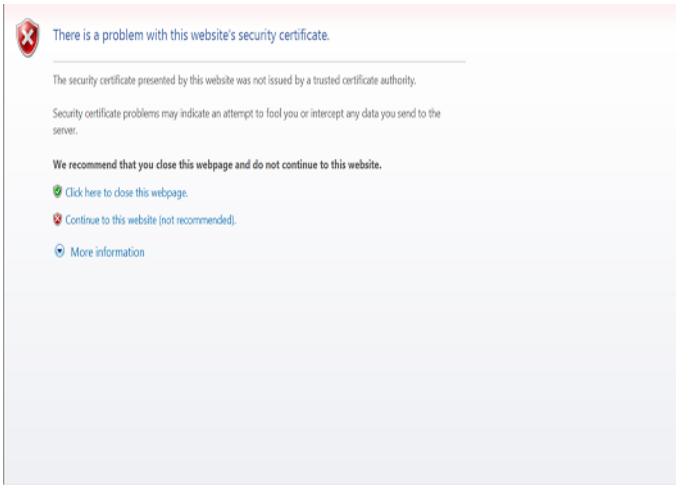

Figure 2: Internet Explorer's native warning

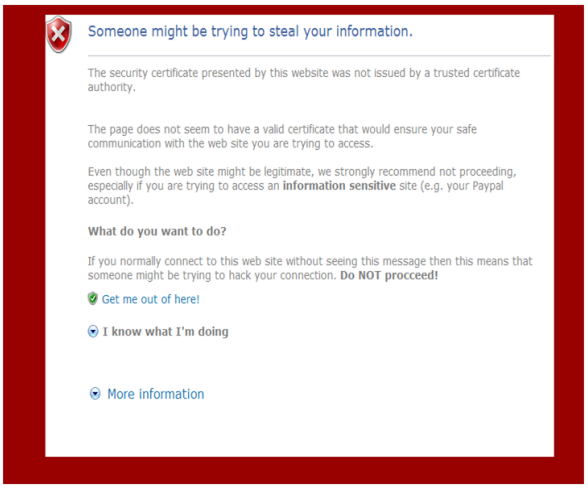

Figure 3: Our custom warning for IE7

mally used, rather than assigning them to random browsers. We also redesigned the custom SSL warnings that were presented to the users, keeping the layout similar to the native warning in an effort to limit the novelty effect that a previously unseen browser interface and warning would have (Figure 3).

3.2 Study Design

Our study was a between subjects experiment with an initial four conditions based on the warning presented (native warning, custom warning) and browser normally used by the participant (Firefox 3.5, Internet Explorer 7). We did not replicate the CMU study's Internet Explorer 6 (IE6) condition, because we could not recruit a sufficient number of participants who used IE6 in their everyday life. Finally, we did not replicate the custom multi-page warning condition present in the CMU study. Instead we substituted it with a Firefox 3 (FF3) condition, in which the custom warning also retained the same layout as FF3's native warning, but with changed colors and wording. We did this because we wanted to investigate if there are differences in security practices between IE and FF users. By using the same colors and wording, but maintaining each browser's layout, we could directly compare reactions to our custom warnings between users of these browsers.

During the study we realized that because we had changed two variables from the CMU study conditions (i.e., breadth of population, use of regular web browser), we would not be able to draw conclusions about the cause of any differences that we might observe during analysis of our data. We therefore added a fifth condition to our study in which we also recruited from a broad population, but we asked participants to use IE7 regardless of which browser they normally used. This condition enabled us to determine whether any differences in the reactions to the warnings were due to the population sample used or the choice of web browser.

To recap, the five conditions were:

1. **[FF Normal Native]** Firefox 3.5 users; the FF3 browser presented its native SSL warning
2. **[FF Normal Custom]** Firefox 3.5 users; the FF3 browser presented our custom SSL warning
3. **[IE Normal Native]** Internet Explorer 7 users; the IE7 browser presented its native SSL warning
4. **[IE Normal Custom]** Internet Explorer 7 users; the IE7 browser presented our custom SSL warning
5. **[IE Random Native]** participants used Internet Explorer 7 regardless of their usual browser; the IE7 browser presented its native SSL Warning

3.3 Recruitment

As one of the main goals for our study was to have a more representative population sample, we decided to seek participants outside the UBC community so that our sample would have diversity in terms of age, gender, occupation, and educational background. Having a broad population should lead to results that are more representative of the average computer user's behavior. It is the authors' belief that in studies like our own, where human behavior is the primary factor under investigation, the population sample is of utmost importance; in order to have statistically valid and generalizable data, we needed to have a good representation of the average user.

We adopted a three step approach for our recruitment. First we advertised the study using flyers posted around the UBC campus and the Vancouver community centers, and using advertisements on Craigslist. In our advertisements, we mentioned that during the study participants would have the task of seeking information from various sources like Google, online banking sites, and online shopping sites. We did not reveal at this point that they would be asked to use their personal information to do so. When potential participants contacted us via email or telephone, we set a date and time for a session via email. After we had arranged a session, we sent a final email with the consent form attached, as required by the UBC BREB, and details about the location of the study. It was only then that we revealed that they would have to retrieve information from various on-line sources, including their bank's online system. For that purpose, they had to have an account with one of the major Canadian banks and needed to remember their bank credentials. We not only wanted to ensure that participants had an online banking account, but that they would have their banking

card number with them (or at least remember it) as this is used on most Canadian banking sites for logging in.

Our reasons for purposely set a date and time for the study prior to revealing that real account information would be used by participants during the study was that we hypothesized that if participants were concerned with the privacy of their information, they would explicitly state that if they canceled the session. Although we had some participants that did not come to their session and who did not provide any reasons for missing it, our hypothesis was confirmed in most cases.

The CMU study's recruitment process advertised the study as a "usability of information sources study." Participants were limited to those that were customers of one particular bank, the one selected by the researchers for use during the study. A screening process was administered that required participants to have used search engines and have performed an online purchase in the last year. In contrast, we advertised our study as one that seeks to investigate the challenges users face when retrieving information online and that we sought to identify the "difficulties people are facing when trying to complete every day tasks online (e.g., search on google.com for information, online banking, online shopping)". We omitted the screening survey done in the CMU study because we were aiming for a broad population, both in terms of occupation and age. We felt that older participants might be relatively unfamiliar with tasks like online shopping so did not include this as a recruitment criteria.

Due to the difficulty of recruiting non-student participants to take part in a study located on the UBC campus, we eventually had to raise the honorarium offered to participants from $10 to $20 CND, in order to meet our recruitment goals.

3.4 Study Protocol and Tasks

We now describe the study protocol and the four tasks that we asked participants to perform.

When a participant arrived, the experimenter gave him a copy of the consent form and asked him to sign it if he agreed to participate in the study. The experimenter then gave a detailed overview of the study session, without revealing the real purpose of the study. This overview was in the form of a script that the researcher had memorized so as to ensure that all participants would receive the same instructions. The four tasks were then simultaneously presented to the participant, and he was asked to review them and ask any questions that he might have. Each task was to find a piece of information and we included a primary source and a secondary source that would enable the participant to retrieve that information. This was a feature of the CMU study, that aimed to mitigate the task focus effect [14] and response bias that have been observed in similar studies [16]. The task focus effect occurs when participants are overly focused on completing the given steps of the experimental task, while the response bias is the altering of participants behavior so as to "please" the experimenter or to comply with what they perceive the study expectations or desired outcome to be.

The first task asked participants to retrieve the surface area of Greece using Google.com as a primary source and Ask.com as the secondary one. The second task asked participants to retrieve the last two digits of their account balance using either their bank's online banking system as the primary source or its telephone banking system as the secondary source. The third task asked participants to locate the price of the hardcover edition of the book Freaknomics using either Amazon.com as the primary source or Barnes and Noble as the secondary source. The fourth task asked participants to create a new email account in order to register with tripadvisor.com, using either Hotmail.com as the primary site or Yahoo.com as the secondary site. The first and third of the tasks were dummy tasks that were there only to obfuscate the real purpose of the study and to reinforce the participants' belief that this study was not about warnings. We counterbalanced the order of the bank and email tasks (i.e., the warning tasks) so that we could control for any order effects in the warnings presented.

The CMU study did not include an email task. Their fourth task asked participants to use the CMU online library catalog or alternatively the library phone number to retrieve the call number of a book. As we wanted to recruit participants from outside UBC, we could not design a similar task. We opted for the email task as described above. Similar to CMU's library task, it is a task with a relatively low risk of personal information exposure for the participant. This is supported by our participants' responses during the exit questionnaire (Appendix B); several claimed that they did not regard this task as an information sensitive one as they did not have to use their own personal information, if they did not want to, to sign up for an email account.

The experimenter did not assist the participants during our study (although many asked for help while performing the tasks, including the dummy tasks), but did not deny that he was part of the research team. In later discussions with of one of the researchers involved in the CMU study (S. Egelman, personal communication, March 31, 2010), we learned that in the CMU study, the researcher who administered the study pretended that he had no connection with the research other than getting paid to sit in the room with the participants and just read the script to them. This subtlety was not reported in the paper describing the CMU study [18]. Although we do not have any data on how this mild role-playing on the part of the researcher might have affected the CMU participants' behavior, it is conceivable that participants might have felt that their information was under greater risk of leakage as no one responsible for the actual study was in the room.

3.5 Exit Questionnaire

After the completion of the four tasks, participants were directed to an online questionnaire on SurveyMonkey (Appendix B). Similar to the CMU study, the questionnaire asked 45 questions in six categories. The first set of questions investigated participants' understanding of and reaction to the bank warning in the study. The second set asked the same questions about the Hotmail warning. The third set asked questions to gauge their general understanding of SSL certificates and invalid certificate warnings. The fourth set gauged participants' prior exposure to identity theft and other cyber threats. The fifth set, asked them about their technical experience, including their experience with computer security. Finally, the sixth set asked general demographic questions like age, gender, and education level. We duplicated most of the questions from the CMU study, but we did add some questions in order to further investigate our research interests (e.g., if the participant would perform

differently when using his own PC instead of the PC in the lab environment).

After the questionnaire was completed, we debriefed participants and revealed the purpose of our study. We also explained to them the utility of SSL and the SSL warnings and advised them about the safe way to consider and react to the certificate warnings in the future.

4. RESULTS

4.1 Participant Characteristics

A total of 217 individuals responded to our recruitment notices. After a time slot was assigned to them, and we sent them the second email revealing that they would need to use their bank credentials, 23 individuals dropped out. Twenty of these individuals explicitly stated that they were not interested in participating because they felt that their personal information would be at risk, while the other 3 dropped out without explaining their reasons. In the end, one hundred participants took part in our study. The remaining 94 were either excluded due to not having an online banking account, did not show up for their appointment, or the appropriate condition already had 20 participants. Participants that revealed that they did not have an online banking account at the time of the laboratory session they were informed that they are not eligible for participating into the study, were given the honorarium and their data (if any were collected) were discarded.

Our participants were almost equally divided by gender (56 female, 44 male) and had a broader age distribution than in the CMU study (Figure 4). We did, however, find it difficult to recruit working non-students participants to take part in our study. Although diverse, our participants were still younger than the general population, according to Statistics Canada. Despite this, we believe that we have achieved an adequately distributed population in terms of age groups to meet our study goal of a representative population as we managed to have participants from a broader age spectrum than the previous study. This diversity allowed us to draw conclusions as to whether recruiting from the general population yielded different results than the previous study. We hypothesized that it would, because students as a social group, fit a certain internet profile which might not be more representative of the general population. Out of our 100 participants 40 were students. As presented in Section 4.3, there was no significant correlation between a participant being a student and their reaction to the warnings.

Our participants were technically sophisticated, scoring a 2.14 on a 0 to 4 Likert scale (0=I often ask others for help with the computer, 4=Others often ask me for help with the computer), when they self-evaluated (in the exit questionnaire) their technical skills based on how often they ask for help from others or others ask help from them with computers. This is similar to the participants in the CMU study, who had a mean score of 1.90 on the same scale.

Of the 100 participants, 80 were assigned to conditions according to the browser they mostly used and the other 20 were assigned to IE7. Of those, 9 were Firefox users, 2 used Safari, and the rest used Internet Explorer in their every day life.

4.2 Effect of Browser/Warning on Behavior

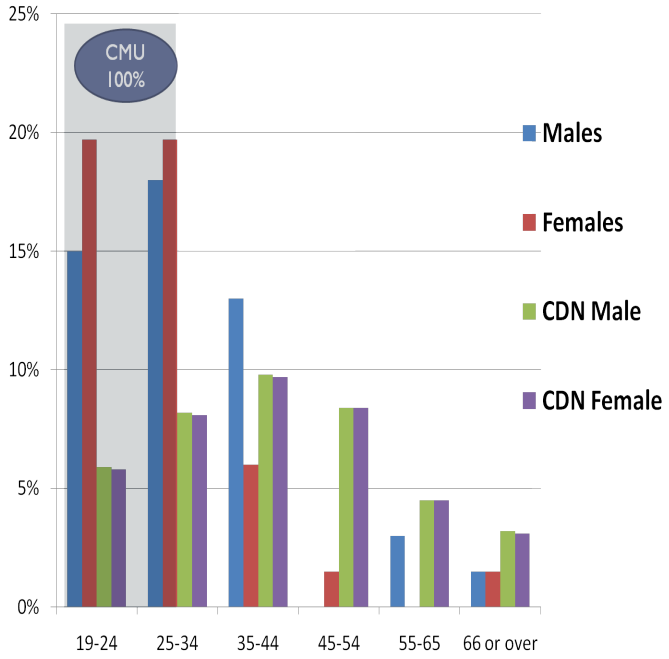

Figure 4: Participants age/gender distribution in our study (CDN: Statistics for Internet use in Canada)

We first examined how the different browsers and warnings affected participants' behavior. The optimal way in which any warning should work is to discourage the user from providing sensitive information to malicious or suspicious sites, while allowing them to continue if the warning is raised on a legitimate web site due to a mis-configured certificate. Table 1 presents the data from both our study and the CMU study for the banking task.

We used the Fisher's exact test to analyze our results in terms of differences in behaviors between the conditions (i.e., browser/warning used) in our study. No statistically significant differences were observed between the various conditions in our study. FF3's native warning was not significantly more effective than IE7's (p=0.358). Furthermore, no differences were observed in the effectiveness of our redesigned warnings when compared to the browsers' native ones (IE7 - IE7 Custom, p = 0.634; FF3 - FF3 Custom, p=0.358). Finally, we examined whether the effect of randomly assigning participants to browsers yielded different results than if they were allowed to use their usual browser and observed no significant difference between IE7 Normal and IE7 Random (p=0.5).

In Table 2, we present our results for the 5 conditions for the Hotmail task. The table shows the number of participants who chose to proceed to sign up with a new Hotmail account despite the SSL warning. Similarly, to the bank task, we observed no statistically significant differences between conditions. FF3's native warning was not significantly more effective than IE7's (p=0.358). Furthermore, no differences were observed in the effectiveness of our redesigned warnings when compared to the browsers' native ones (IE7 - IE7 Custom, p=0.5; FF3 - FF3 Custom, p=0.5). Finally, we examined whether the effect of randomly assigning par-

	FF3	FF3 custom	IE7		IE7 custom
CMU	11 (55%)	N/A	18 (90%)		9 (45%)
UBC	16 (80)%	17 (85%)	N: 14 (70%)	R: 15 (75%)	14 (70%)

Table 1: Comparison of results between the two studies for participants who chose to ignore the SSL warning at the bank sign-in web site. (N: Participants using their normal browser, R: participants are assigned to browsers randomly)

	FF3	FF3 custom	IE7		IE7 custom
UBC	14 (70)%	16 (80%)	N: 16 (80%)	R: 17 (85%)	16 (80%)

Table 2: Results for participants who chose to ignore the SSL warning at the hotmail sign up web site. (N: Participants using their normal browser, R: participants are assigned to browsers randomly)

ticipants to browsers yielded different results if they were allowed to use their normal browser and found no significant difference between IE7 Normal and IE7 Random (p=0.5).

All participants were able to complete all the tasks using either the primary or the secondary method given to them to retrieve the information.

4.3 Effect of Participant Characteristics

We did not observe any statistical difference in participants' reactions based on their gender, level of education (University undergraduate or higher versus Professional degree or lower), or their student status. We found this result to be surprising as we had hypothesized that there would be differences. We should note that the majority of our participants were highly educated with 41% of them having a college education and 21% having a graduate level education. Our population is therefore not necessarily representative of the general internet user population; and many of our non-student participants may be similar to current students. Still, the fact that no significant differences were observed does not confirm concerns that the results of studies that rely solely on students as their population sample are simply not generalizable.

4.4 Effect of Task Context

We hypothesized that upon seeing the warning at the bank site, our participants would stop and choose the alternative contact method (i.e., use phone banking) rather than risk sharing their information with the site. We also hypothesized that when the warning at Hotmail was displayed, they would continue as there was no real danger to their personal information (we did not instruct them to use their actual information). Our hypotheses were not confirmed. Almost all (96/100) participants treated SSL warnings at both sites (bank and Hotmail) similarly (i.e., ignoring or heeding them in both cases). Only 2 participants stopped at the bank site but proceeded at Hotmail; an equal number of them proceeded at the bank site because they felt that it was trustworthy, whereas they did not continue with the Hotmail task, later explaining that they had heard of security incidents at Hotmail. It is worth noting that 3 participants reported that they did not see a warning during the Hotmail task. The warning was actually displayed to them, but they did not notice it, rather they disregarded it and clicked through. This is evidence of how habituation and task priority can affect the user's perception while performing everyday tasks on the Internet.

Hypothesized Action for FF3 Warning	
I would proceed ignoring the warning	14%
I would proceed if the site was not information sensitive	28%
I would leave the web site	43%
Other, please explain	15%

Table 3: Self reporting of participants' intended action when presented with a screenshot of the Firefox 3.0 self signed certificate warning while trying to reach a hypothetical web site

Reason for My Reaction to the Bank Warning	
It is a study	33%
Calling the bank is time consuming	15%
I wanted to complete the task	13%
I am used to the warning	10%
I trust my bank's web site	25%
Other	4%

Table 4: Participants' responses, in the online survey, when asked why they ignored the warning at their bank's web site

4.5 Participants' Intended Reactions to a Warning vs Actual Reactions

As was done in the CMU study, we presented participants with a screen shot of a Firefox 3.0 self-signed certificate warning during the exit survey. We asked what they would do if they saw this warning prior to entering a web site (e.g., www.example.com) and provided a multiple choice question with four choices; "I would leave the web site", "I would proceed if the web site was not information sensitive", "I would proceed to the web site", "Other, please explain". As shown in Table 3, the majority of participants claimed that they would leave the web site or would proceed if the site was not information sensitive. However, this is not what the majority of the participants actually did during the study for either the information sensitive task of logging into their bank account task 1 or during the Hotmail sign up task 2. We will elaborate on this point further in the discussion.

4.6 Participants' Reasoning About their Reactions to the Warnings

Although the CMU authors speculated that the study environment might have affected their results, they did not report any data in support of this. We were able to measure through participants' self-reports interesting findings about the reasons behind their actions upon seeing the SSL warning at the bank log-in web page (Table 4). We asked them, through an open ended question in the exit questionnaire, to explain why they chose to ignore or heed the warning. One third of those who ignored it explicitly claimed in their response that they ignored the warning because they felt safe in the study environment (e.g., building on campus, consent form handed prior to the study). Another 13% claimed that they did so because they wanted to complete the task. These findings raise concerns about the utility of study designs like our own, which are common in the usable security field, as over 40% of our participants reported that the study environment influenced their actions in a manner that may be inconsistent with their usual behaviors.

5. DISCUSSION AND LIMITATIONS

The overall aim of our experiment was to investigate computer security behavior in the context of SSL warnings. Interesting inferences about the bias introduced by laboratory environment, the perception of warnings (and how this changes over time) by the users as well as the elements of warnings that draw attention and affect user bahaviors can be drawn from the fact that we did not observe any statistically significant differences between the impact of different warnings, as implemented by the two browsers we investigated and our own designs. We should note, however, that our limited sample size with only 20 participants in each condition may have hindered our ability to detect a difference. Failure to reject the null hypothesis does not alone provide sufficient statistical evidence to suggest that the null hypothesis is true. Further studies will be required to verify our findings.

Additionally, results of our data analysis raised some interesting topics, the discussion of which we feel will benefit the research community. From the analysis, as presented in Section 4.6, we believe that there is a significant impact of the laboratory environment, which introduces uncertainty in the results, even if one considers cognitive dissonance as a reason for the disparity between actual actions and self-reported would-be behaviors of participants. This affected not only the results gathered and their quality, but also the profile of participants that took part in our study. This impact, we believe, is due to systematic limitations of the experimental method; therefore, the key points we present may be applicable to other studies that have similar experimental design and overall goals.

5.1 Impact of Browser Used

As discussed in Section 4.2, no statistically significant differences were observed across the different browser conditions. This finding does not confirm the results of the CMU study, in which participants assigned to the FF3 condition tended to behave differently from the ones assigned to FF2 or IE7, and similarly to those assigned to the single page custom warning condition. In our study, all conditions (i.e., browser/warnings), including the custom warnings, did not yield a statistically different response from our participants. We interpret this difference in findings as evidence of habituation. In the CMU study, the design of the custom warning was similar to the FF3's warning. The design of SSL certificate warnings in FF3 introduced a radically new way of interaction with the user, which was completely different from the pop up windows that FF2 and IE6 had been employing. We believe that participants in these two conditions encountered a situation that was completely new to them. Even IE7, where the warning covers the whole page, requires only one click instead of three clicks in FF3. While making it more difficult to overcome a warning might work for a certain period of time, over time web users may have become accustomed to it and learned the necessary steps they need to take in order to go past the warning. Results of our study also demonstrate how difficult it is to replicate a study on web behavior as the Web is a rapidly changing environment. Not only does the environment change, but so do the users' perceptions and practices within the environment.

5.2 Impact of Age and Occupation

We had hypothesized that age and occupation would affect how participants responded to SSL warnings. In particular, we expected that students fitting a certain internet profile might behave differently than the general population. Our hypothesis was supported by prior research on privacy that indicate gender, age, education, and experience online to be associated with privacy concerns. Dommeyer and Gross, in a 2003 study aiming at consumer knowledge of privacy-related laws and practices, reported that young people are more likely to use privacy protection mechanisms [6]. However, contrary to our hypothesis, our results did not confirm that education, age, or the student status of the participant affected the way that they reacted to warnings. Previous research in phishing by Dhamija et al. in 2006, which had both students and university staff as participants, has also reported that demographic differences (e.g., sex, age, and education) did not affect participant security behavior [5]. Clearly, further studies are required to tease out when these participants characteristics have an impact on end user's privacy and security concerns and behaviors.

5.3 Impact of Intense Colors and Clear Wording

As discussed in Section 3.1, for our custom warnings, we purposely redesigned IE7's and FF3's native warnings, maintaining a similar layout but making the colors more intense (red) and the language clearer. Previous research by Geffen et al. indicates that straightforward exposition in privacy notices can develop trust [8]. However, the impact of these changes was not the one we expected. It appears that the participants' conviction of the safety of their information was such that they disregarded our custom warnings with ease. Although the warning informed the participants that someone might be trying to steal their information and they were clearly instructed not to go to the web site, they disregarded the warning and continued. This was in contrast to what was observed in the CMU study where the layout of the custom IE7 warning was altered radically and the warning was more effective in deterring participants from ignoring it.

Our findings support the notion that users do not necessarily believe or even read warnings and messages they come

across while performing a task; rather, they use their past experience and knowledge (in this case, their knowledge of the warning's layout and how to overcome it) to complete their tasks. Warnings that present users with a new layout frequently have been shown to address the problem of habituation better than the traditional ones as shown in the study by Brustoloni and Villamarin-Salomon [3]

5.4 Differences in Self-Reported and Observed Actions

During our exit questionnaire, we presented participants with a screenshot of a Firefox 3.0 SSL warning raised due to a self signed certificate. We asked what they would do if they saw this warning prior to entering a web site (e.g., www.example.com). As described in Section 4.5, the majority of participants claimed that they would leave the web site or would proceed if it was not information sensitive. This is very different from participants' actual behaviors that we observed when they were presented with an SSL warning during the banking and Hotmail tasks. In those cases, the majority of participants ignored the warnings and proceeded, even when facing custom warnings that had intense colors and strong wording. The difference in self-reported and observed actions has been reported in other usable security studies

It should be noted that this question was asked in the late stages of the exit questionnaire. Therefore, it is quite possible that participants understood the actual purpose of the study (i.e., their reaction to SSL warnings) by the time they answered the question. In that case, as an result of cognitive dissonance, they may have been biased towards providing the "correct", or at least a logically justified, response (i.e., security of the laboratory environment) rather than accurately report their usual behavior.

It would be interesting to study further the reasons behind such differences. We are aware of the problems of self-reporting in terms of reliability, as shown in studies like [9] where participants claimed to take particular actions regarding their privacy, when in fact they actually took different ones during the tasks. However, based on our collected data, we believe that the usable security research community also has to take into account the bias the laboratory environment brings into the experiment by offering a safe environment in which the participants interact. In order to conclude which method produces more reliable results for the context under investigation, we would like to design an experiment that would ask participants in a survey for qualitative responses on what their reaction would be if presented with a security feature. At another time (either before or after), their actions could be observed in a natural setting as they are presented with warnings during their normal tasks. Although we are aware of the challenges that such an experimental design would involve (e.g., ethics approval, lack of controlled environment), it is conceivable that the experiment could be conducted outside a lab environment with minimal or no awareness on behalf of the participants. This way the laboratory bias would be eliminated, and the reliability of the survey responses would be clear.

5.5 Difficulty in Recruiting a Representative Population Sample

One limitation of our study was our inability to recruit a representative population pool. The participant pool is of utmost importance for any study that requires a representative sample. Without such a sample, it is hard to generalize results or draw conclusions, especially when what is investigated is the behavior or views of individuals. In our study, we recruited a broad population in terms of age, education, and occupation; however, in reality, our participant population is still skewed. Participants were not randomly recruited, rather they volunteered; and it became clear during the recruitment process that many security concerned people opted not to participate.

Due to our recruiting method, we were able to establish communication with potential participants before making it clear that they would be required to use sensitive, personal information (i.e., log into their actual online banking site). This allowed us to observe the fact that the most security aware or cautious individuals decided not to take part in the study, either on their own or because of advice given by someone in their environment (e.g., spouse). This refusal occurred even though we stressed that no information would be recorded during the study. We also sent the consent form, as an attachment to the same email, which should have provided an additional sense of safety and security.

This raises a concern about the statistical validity of the recruited sample when sensitive information must be used by a participant and he has prior knowledge of that. The problem is that a considerable percentage of the potential participants will not take part in the study out of fear of leakage of their information. The systematic error introduced here is due to the fact that the users who take part in studies similar to our own fit a certain behavioral profile and so conclusions can be drawn only for this behavioral profile. This potentially affects the reported severity of the problem under investigation as we are not recruiting as participants those users who do the "right thing" (i.e., keep their information safe and private). As a result the generalizability of the conclusions drawn on the results gathered is degraded.

In addition, as previously discussed, we decided during the study to increase the compensation for participants from $10 to $20 CND, similarly to the CMU researchers. We did that to encourage participation, as it we found it difficult to recruit an adequate number of participants at the beginning of the study. One concern we had with recruitment and our effort to have a representative sample was that $10 might be too little to entice participation by individuals of certain financial demographics. However, research by Russell et al. [15] regarding participation in medical studies indicates that participants with higher education and financial security are more inclined to volunteer for a study with little or no financial incentive. In fact, we observed that same effect. When we increased our honorarium, although the flow of participants increased, we started having issues with participants that clearly registered purely for the money. Certain individuals, knowing that is common practice to give the honorarium to the participant if they show up, regardless of the outcome of the session took advantage of this practice. During the screening process, their responses made them appear to be eligible to participate and it was only during the study session they would reveal that they were not eligible to take part (e.g., did not have online bank accounts). In this regard, it is clear that increasing the financial incentive might lead to an even more skewed population sample or to participants that have the goal of receiving the money with no regard to the other aspects of volunteering for a study

(e.g., being interested in giving their time for the sake of research in general). This is another interesting challenge with laboratory studies that seek to investigate participants' behavior. Researchers need participants to act naturally and not alter their behavior in order to meet the researchers' perceived needs, so as to be eligible to receive compensation.

5.6 Impact of Study Environment

Another limitation of our study was the impact of the study environment, despite our efforts to mitigate the known challenges of conducting a laboratory study. As discussed in our results section, we asked participants to explain why they chose to ignore or heed the SSL warnings with which they were presented. One third explicitly stated that it was the study itself (i.e., being conducted by UBC researchers and having approval from BREB) that made them trust the procedure and the experimental setup and ignore the warning in order to enter their personal information into the site. Furthermore, some of those participants claimed that they would do otherwise if this was not a laboratory experiment and they had seen the warning at a public PC or while connected through a public network, or even on their own computer.

Another 13% of participants claimed that they ignored the warnings because they wanted to complete the task. If considered conservatively, these participants can be interpreted as a task focus effect (i.e., the participant is continuing to the web site because she feels that she should do as asked) [14].

These two types of responses make us question the very utility of laboratory study designs in usable security, when *user practices and behavior* are studied. Although this is not, by any means, new knowledge in the field of behavioral studies; it is the first time to the best of our knowledge, that empirical data have been collected in a usable security study, suggesting inadequacy of the laboratory study as a tool for this purpose. Even if measures are taken to mitigate this issue, in our opinion, it is very hard, if not impossible, to make sure that the measures have been successful.

We argue that using laboratory studies as an experimental methodology contradicts popular computer security advice. We essentially ask our participants to perform in public and to potentially unsafe environments (as perceived by some "privacy fundamentalists" [1]) those actions that they have been told/trained to perform in private and safe environments. Consequently, we have a systematic error introduced by the study design and the potential participant behavioral profile. Namely, (1) we either recruit participants who are by nature prone to unsafe behavior (as discussed in section 5.5, we had difficulty recruiting security sensitive participants) or (2) we recruit participants that put quite justified good faith in the study environment due to the perceived reputation of the research institution. In either case, the conclusions drawn from the results collected under such circumstances can hardly be considered reliable and representative of *actual* user practices. It is the authors' belief that although laboratory user studies are invaluable in usability research, they may be less suitable for the purpose of research in usable security when the investigation includes the need for participants to use personal data in the study environment. The systematic error introduced by the study environment seems to be a hard to correct or avoid.

It can be argued that biases like the one coming from the laboratory environment affect all conditions equally, so the error introduced, although present, can be disregarded. We agree with this argument for those studies when the goal is a tightly controlled evaluation of the differences between conditions or an evaluation of alternative approaches. In our case, and in many other similar studies in usable security, our aim was twofold. On the one hand, we sought to conduct a comparative evaluation that investigated how well the different warning implementations fared in deterring participants from making insecure decisions. On the other hand, we also sought to investigate our participants' reaction to a SSL warnings in general. The latter was the reason why we chose not to assign participants to conditions randomly, rather to let them choose the browser they felt most comfortable with (i.e., the one they used in their everyday life). By doing so we wanted to increase ecological validity, which in turn would allow us to observe participants' (more) natural behavior towards SSL warnings. This is where the bias, although present in all conditions, became problematic. It did not, directly, affect our ability to compare warning implementations. It rather impaired our ability to assess participants' behavior and practices in a reliable manner. Given the nature of the study, this bias affected the core of the experiment and affected, ultimately, our ability to make reliable judgments of warning implementations. For example, participants that did not perceive the situation as dangerous would have a cavalier approach in any condition towards the corresponding warning and the potential risks the warning was meant to warn against, thus canceling the warnings' very purpose of alerting the user against a risk. It might be that when the matter under investigation is tightly related to participants' behavior, perception, and practices, then the laboratory bias, although present in all conditions, is one that generates an error that cannot be ignored.

Although our results indicate that there is a strong bias introduced by the laboratory environment, we believe that lab studies are still useful to the usable security research community. The findings can be compared with prior research that was conducted under similar conditions and they may indicate differences over time. In our case, we found that over time users have apparently familiarized themselves with the new warning layouts and the warnings have become less effective, as shown by the difference between the effectiveness of FF3 warnings as found in our study and the earlier CMU study when the layout was novel for participants.

Measures to increase ecological validity (e.g., adding tasks to conceal the real purpose of the study) are often inadequate to mitigate the biases introduced by the study environment. It might be possible to successfully conceal the real purpose of an experiment by adding more irrelevant tasks and devising elaborate scenarios to create a greater sense of realism. However, adding too many tasks could render the study infeasible due to time constraints, and scenarios do not seem to be successful in creating a realistic environment [16].

Some studies have received ethical approval to study the participants' behaviors without their prior knowledge. However such studies have proved to be equally problematic, as happened in the case of [12] where, although the research investigating phishing had received ethical approval to use members of Indiana University community as participants without their prior consent, there was a backlash as some of the participants felt embarrassed and outraged, even threatening legal action against the researchers. This example is

not given to argue against field studies. On the contrary, it is the authors' belief that field observations experiments can be more useful in examining the security behavior of users, since the setting is not artificial and the participants may be unaware of the researcher's manipulations of the environment and measurement of their actions. However, it must be noted that a field study, especially with a privacy or security focus has to consider more than just logistical issues. It also has to take into account the possible resentment on the part of the participants, even when no actual legal concerns are present, as it happened in the Indiana University case.

We suggest that alternatives to a regular laboratory study may have to be sought for those cases where we seek to investigate human behavior in computer security, particularly when we require participants use their private data. In an effort to maintain a sense of realism without compromising the ethical aspect of a study, Jakobsson et al. in [13] designed an experiment that took into consideration both realism and ethical concerns, demonstrating safe techniques that they developed to examine phishing attacks on an auction site (rOnl). Taking advantage of rOnl's features, they were able to commence phishing attacks against its users without actually compromising the identities of the users that fell victims of their scheme. Moreover, researchers could devise experiments in lieu of laboratory studies that use methods to log users' behavior while they unknowingly perform everyday actions. Then, after asking for the users' consent, they could analyze these logs (e.g., contacting the IT department of a big corporation and installing a proxy that generates SSL warnings on particular sites and logging the reaction). Such post-hoc designs have been used by the HCI research community in investigating web use [4]. Although we are aware of the challenges in terms of research ethics and acquiring the cooperation of organizations and individuals, it is our belief that the benefits in terms of the reliability of results will be substantial. We propose that additional attention and effort should be dedicated towards such experimental designs to support the move away from laboratory studies in future usable security research that investigates user behaviors and requires a high degree of ecological validity.

6. CONCLUSION

We presented the experimental design that we used to investigate the effectiveness of SSL warnings and to validate the findings of the prior CMU study. Our findings do not confirm the findings of that study, rather they suggest that a completely new contextual environment is in effect. Participants may be now aware of the new warning designs and have already learned how to bypass them in order to achieve their primary task. We presented evidence that suggests that after a certain period of time, the actual warning design does not seem to influence how participants, accustomed to the interface of the browser, perceive and react to it. Further studies are required to validate whether our null results for the impact of participant demographics actually confirm the null hypothesis.

Furthermore, our analysis also raised concerns about the limitations of laboratory studies for usable security research on human behaviors when ecological validity is important. This, we believe, is the most important contribution of our study. It is our belief that the aforementioned limitations of the particular experimental method are applicable to other studies in the field of usable security. The observed reluctance of security concerned people to take part in our study raises concerns about the ability of such studies to accurately and reliably draw conclusions about security practices and user behavior of the general population. Finally, we suggested alternative study methodologies that might be free of the systematic errors and could yield more reliable results. Such methodologies might complement current experimental designs in order to mitigate the limitations inherent in any design approach.

Acknowledgements

We thank study participants for their time, and members of the Laboratory for Education and Research in Secure Systems Engineering (LERSSE) who provided valuable feedback on the earlier drafts of this paper. Cormac Herley provided feedback in May 2010 on the design of the project. Comments from the participants of SOUPS Usable Security Experiment Reports (USER) Workshop 2010 and ISSNet annual workshop helped us to improve this research. Comments from the anonymous reviewers as well as from shepherd Stuart Schechter were instrumental in improving the paper. This research has been partially supported by the Canadian NSERC ISSNet Internetworked Systems Security Network Program.

7. REFERENCES

[1] M. S. Ackerman, L. F. Cranor, and J. Reagle. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce*, EC '99, pages 1–8, New York, NY, USA, 1999. ACM.

[2] R. Biddle, P. C. van Oorschot, A. S. Patrick, J. Sobey, and T. Whalen. Browser interfaces and extended validation ssl certificates: an empirical study. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, CCSW '09, pages 19–30, New York, NY, USA, 2009. ACM.

[3] J. C. Brustoloni and R. Villamarín-Salomón. Improving security decisions with polymorphic and audited dialogs. In *SOUPS '07: Proceedings of the 3rd symposium on Usable privacy and security*, pages 76–85, New York, NY, USA, 2007. ACM.

[4] A. Cockburn and B. McKenzie. What do web users do? an empirical analysis of web use. *Int. J. Human-Computer Studies*, 54:903–922, 2001.

[5] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 581–590, Montréal, Québec, Canada, 2006. ACM.

[6] C. J. Dommeyer and B. L. Gross. What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies. *Journal of Interactive Marketing*, 17(2):34–51, 2003.

[7] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In *CHI '08: Proc. of the SIGCHI conf. on Human factors in Computing*

Systems, pages 1065–1074, New York, NY, USA, 2008. ACM.

[8] D. Gefen, E. Karahanna, and D. W. Straub. Trust and tam in online shopping: An integrated model. *MIS Quarterly*, 27(1):pp. 51–90, 2003.

[9] J. Gideon, L. Cranor, S. Egelman, and A. Acquisti. Power strips, prophylactics, and privacy, oh my! pages 133–144. ACM Press New York, NY, USA, 2006.

[10] J. Henrich, S. Heine, and A. Norenzayan. Most people are not weird. *Nature*, (466):29, 2010.

[11] T. Jackson. How our spam filter works. Technical report, Google, 2007.

[12] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Commun. ACM*, 50(10):94–100, 2007.

[13] M. Jakobsson and J. Ratkiewicz. Designing ethical phishing experiments: a study of (rot13) ronl query features. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 513–522, New York, NY, USA, 2006. ACM.

[14] A. Patrick. Commentary on research on new security indicators - essay. http://www.andrewpatrick.ca/essays/commentary-on-research-on-new-security-indicators, 2007.

[15] M. L. Russell, M. G. Donna, and E. Burgess. Paying research subjects: participants' perspectives. *Journal of Medical Ethics*, 26(2):126–130, 2000.

[16] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperor's new security indicators. In *Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 51–65, Washington, DC, USA, 2007. IEEE Computer Society.

[17] D. W. Stewart and I. M. Martin. Intended and unintended consequences of warning messages: A review and synthesis of empirical research. *Journal of Public Policy and Marketing*, 13(1):1–19, 1994.

[18] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying Wolf: An empirical study of SSL warning effectiveness. In *Proceedings of 18th USENIX Security Symposium*, pages 399–432, 2009.

[19] S. J. B. R. van Oorschot P. C. Patrick A. S. Exploring user reactions to new browser cues for extended validation certificates. In *Proceedings of the 13th European Symposium on Research in Computer Security*, pages 411–427, 2008.

[20] T. Whalen and K. M. Inkpen. Gathering evidence: use of visual security cues in web browsers. In *Graphics Interface*, pages 137–144. Canadian Human-Computer Communications Society, 2005.

[21] M. Wogalter. Purpose and scope of warnings. In *Handbook of Warnings*, pages 3–9. Lawrence Erlbaum Associates, 2006.

[22] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems(CHI '06)*, pages 601–610, New York, NY, USA, 2006. ACM.

APPENDIX A

Consent Form and Recruitment Flyer

THE UNIVERSITY OF BRITISH COLUMBIA

Consent Form

Information Usability Study

Principal Investigator:

The principal investigator of this research is Dr. Konstantin Beznosov from the Electrical and Computer Engineering Dept. of UBC. You can contact him at beznosov@ece.ubc.ca or 604 822 9181

Co-Investigators:

Andreas Sotirakopoulos, Masters Student
Levi Stoddard, Undergraduate Student
Kirstie Hawkey, Post-Doctoral Fellow

All co-investigators are from the Electrical and Computer Engineering Dept. of UBC. You can contact them at 604 827 3410

Purpose:

The purpose of the study is to learn more about the behavior of users while searching for information online. In particular, we are interested in identifying challenges users might face.

Study Procedures:

You will be asked to complete a series of tasks such as retrieving information from well known web sites (e.g. Google.com, Amazon.com) as well as from your bank's web site. In addition, after you complete the tasks, you will be asked to complete a short online survey regarding your reasoning during the tasks. The whole session will be approximately one hour long. You will be observed during the session by one researcher, and your voice will be digitally recorded. No login information will be retained in the study.

Confidentiality:

The identities of all participants will remain anonymous and will be kept confidential. Identifiable data and audiotapes will be stored securely in a locked cabinet or in a password protected computer account. All data from individual participants will be coded and their anonymity will be protected in any reports, research papers, and presentations that result from this work.

Remuneration/Compensation:

We are very grateful for your participation. You will receive $20 as compensation for participating in this project.

Contact for information about the study:

If you have any questions or require further information about the project you may contact Dr. Konstantin Beznosov at 604 822 9181 or Andreas Sotirakopoulos at 778 322 3907.

Contact for concerns about the rights of research subjects:

If you have any concerns about your treatment or right as a research subject, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604 822 8598 or if long distance e-mail rsil@ors.ubc.ca.

Consent:

Your participation in this study is entirely voluntary and you may refuse to participate or withdraw from the study at any time.

Your signature below indicates that you have received a copy of this consent form for your own records and indicates that you consent to participate in this study.

Participant Signature Date

Printed Name of the Participant Signing Above

Researcher Signature Date

Printed Name of the Researcher Signing Above

THE UNIVERSITY OF BRITISH COLUMBIA

Be part of a UBC research project!

UBC researchers are conducting a study about challenges users face when retrieving information online.

All participants will receive **$20**.

We require volunteers to participate in a single **45 minute** session.

During this session, you will be asked to complete a series of tasks that will help us draw valuable conclusions on the **difficulties people** are facing when trying to complete everyday tasks online (e.g., search google.com for information, online banking, online shopping).

If you would like to participate in this study, please contact Andreas Sotirakopoulos at **andreass@ece.ubc.ca (**or call 778-322-3907).

Help us to help the community make the Internet a better place for all.

Appendix B

Online Survey

Usability of Information Sources Study

1. Bank Warning Message

The following questions are related to the warning you saw at your bank's web site.

1. Before this study, have you ever seen the warning you saw at your bank's web site?

- Yes
- No
- I'm not sure

2. Did you read the full text of the warning at your bank's web site? Why/Why not?

3. When the warning at your bank's site was displayed to you what was your first reaction?

4. What did you believe the warning in the bank web site meant?

- That the site was infected with a virus
- That the site was going to install spyware on my pc
- That someone might have forged the security certificate and could be evasdroping your communication
- None of the above

Other (please specify)

5. After seeing the warning message at the bank web site, did you beleive there may be some risk involved with accesssing the website?

- Yes
- No
- Not Sure

Usability of Information Sources Study

6. If you chose Yes or Not Sure in question 5 please rate the amount of risk you feel you were warned against

	Very low risk	Low risk	Neutral	High risk	Very high risk
Risk					

7. What action, if any, did the warning at the bank web site want you to take?

- To not continue to the web site
- To be careful while continuing to the web site
- To continue to the web site
- I did not feel it wanted me to take any action

Other (please specify)

8. Please explain why you chose to either heed or ignore the warning at your bank's web site.

2. Yahoo.com Warning Message

The following questions are related to the warning you saw at the Yahoo web site.

1. Before this study, had you ever seen the warning you saw at the Yahoo.com e-mail sign up web site?

- Yes
- No
- I'm not sure

2. Did you read the full text of the warning at the Yahoo.com e-mail sign up web site? Why/Why not?

3. When the warning at the Yahoo.com e-mail sign up web site was displayed to you what was your first reaction?

Usability of Information Sources Study

4. What did you believe the warning in the Yahoo.com e-mail sign up web site meant?

- That the site was infected with a virus
- That the site was going to install spyware to your pc
- That someone might have forged the security certificate and could evasdroping your communication
- Nothing of the above

Other (please specify)

5. After seeing the warning message at the Yahoo.com e-mail sign up web site, did you beleive there may be some risk involved with accesssing the website?

- Yes
- No

6. Rate the amount of risk you feel you were warned against

	Very low risk	Low risk	Neutral	High risk	Very high risk
Risk					

7. Did you believe the warning at the Yahoo.com e-mail sign up web site?

- Yes
- No

Other (please specify)

8. Why did or did not believe the warning at the Yahoo.com e-mail sign up web site?

9. Please explain why you chose to either heed or ignore the warning at the Yahoo.com e-mail sign up web site.

3. Security Decision Factors

Usability of Information Sources Study

1. How much did the following factors influence your decision to heed or ignore the warnings?

	No influence at all:0	1	2	3	4	5	Strongly influence: 6
The text of the warning							
The colors of the warning							
The choices that the warning presented							
The destination URL							
The look and feel of the destination website							
Other factors (please describe below)							

2. If there were any other factors, please describe them here.

3. Which factor had the most influence on your decision?

- The text of the warning
- The colors of the warning
- The choices that the warning presented
- The destination URL
- The look and feel of the destination website

Other factors (please specify)

4. Technical Experience

1. Rate yourself on this scale:

	I often ask others for help with the computer				Others often ask me for help with the computer
Computer help					

2. Do you have a degree in an IT-related field (e.g., computer science, electrical and computer engineering, etc.)?

- Yes
- No

Usability of Information Sources Study

3. Have you attended a computer security conference in the past year?

- ◯ Yes
- ◯ No

4. Have you ever taken or taught a course on computer security?

- ◯ Yes
- ◯ No

5. How often do you inform yourself regarding computer security news and updates (e.g., reading online articles, computer magazines, etc.)?

- ◯ Very often (2 times a week or more)
- ◯ Often (Weekly)
- ◯ Somewhat often (Monthly)
- ◯ I do not ever read about computer security developments

Other (please specify)

6. What is a man in the middle attack?

- ◯ A virus installed on my pc
- ◯ A kind of Denial of Service attack
- ◯ Someone is eavesdroping on my communication
- ◯ None of the above
- ◯ I do not know

Other (please specify)

7. What is a security certificate?

8. What is a self-signed certificate?

5. Online Security Questions

Usability of Information Sources Study

1. Have you ever had any online account information stolen?

- ◯ Yes
- ◯ No

2. Have you ever found fraudulent transactions on a bank statement?

- ◯ Yes
- ◯ No

3. Have you ever had your social insurance number stolen?

- ◯ Yes
- ◯ No

4. Have you ever been notified that your personal information has been stolen or compromised?

- ◯ Yes
- ◯ No

6. Warning Message

Usability of Information Sources Study

1. Imagine you are trying to visit a web site and see the warning message shown above. What does it mean?

2. What would you do if your web browser displayed this message?

- ◯ I would proceed, ignoring the warning
- ◯ I would proceed if the web site was not information sensitive
- ◯ I would leave the web site

Other (please specify)

3. Please justify your choice in question 2

7. Warning Message cont.

1. Imagine you are trying to visit a web site and see the warning message shown above. What does it mean?

Usability of Information Sources Study

2. What would you do if your web browser displayed this message?

- ◯ I would proceed ignoring the warning
- ◯ I would proceed if the web site was not information sensitive
- ◯ I would leave the web site

Other (please specify)

3. Please justify your choice in question 2

8. Demographics

1. What is your age?

- ◯ 18 or under
- ◯ 19-29
- ◯ 30-39
- ◯ 40-49
- ◯ 50 or over

2. What is your gender?

- ◯ Male
- ◯ Female

3. What is your highest level of education?

- ◯ Some high school
- ◯ High school diploma
- ◯ College degree
- ◯ Graduate Degree
- ◯ Professional degree (including trade school)

Other (please specify)

Usability of Information Sources Study

4. If you have additional comments please write them below.