Self-Identified Experts Lost on the Interwebs

Timothy Kelley
Indiana University Bloomington
107 S. Indiana Ave.
Bloomington, IN, 47405
kelleyt@indiana.edu

Suzanne Lien
Santa Clara University
500 El Camino Real
Santa Clara, CA 95053
slien@scu.edu

L. Jean Camp
Indiana University Bloomington
107 S. Indiana Ave.
Bloomington, IN, 47405
ljcamp@indiana.edu

Douglas Stebila
Queensland University of Technology
Brisbane, Queensland, Australia
stebila@qut.edu.au

ABSTRACT

Security cues found in web browsers are meant to alert users to potential online threats, yet many studies demonstrate that security indicators are largely ineffective in this regard. Those studies have depended upon self-reporting of subjects' use or aggregate experimentation that correlate responses to sites with and without indicators.

We report on a laboratory experiment using eye-tracking to follow the behavior of self-identified computer experts as they share information across popular social media websites. The use of eye-tracking equipment allows us to explore possible behavioral differences in the way experts perceive web browser security cues, as opposed to non-experts.

Unfortunately, due to the use of self-identified experts, technological issues with the setup, and demographic anomalies, our results are inconclusive. We describe our initial experimental design, lessons learned in our experimentation, and provide a set of steps for others to follow in implementing experiments using unfamiliar technologies, eye-tracking specifically, subjects with different experience with the laboratory tasks, as well as individuals with varying security expertise. We also discuss recruitment and how our design will address the inherent uncertainties in recruitment, as opposed to design for an ideal population. Some of these modifications are generalizable, together they will allow us to run a larger 2x2 study, rather than a study of only experts using two different single sign-on systems.

1. INTRODUCTION

Web browsers use security indicators — such as the presence of a lock icon and additional security certificate information in the browser chrome and the use of the protocol "https" in the location bar — to notify users of possible security properties or risks of a given website. These are security 'cues'.

Security cues found in web browsers are meant to alert users to potential online threats, yet there are many studies that demonstrate that security indicators are largely ineffective in communicating security information to users. However, as Schechter *et al.* demonstrated, users are often unaware of these cues [13]. Other studies have shown that, even when users are aware, many people are unsure about how to interpret them, or ignore them. Technical expertise, however, has been shown to be a mitigating factor in a user's vulnerability to online fraud.

There is evidence that a user's technical ability is a key factor in mitigating the risk posed by malicious websites and email [14, 2, 20, 9]. Yet the methods by which computer security experts differentiate between malicious and benign websites and email are not well understood. The goal of our experimental work was to compare the use of experts and nonexperts in their use of cues. Previous eye-tracking studies have examined single population groups to identify a group's attention to security cues, making note of technical expertise, but have not compared the differences between novice and expert users. While these previous studies have noted which subjects had computer experience, those subjects were not tested in separate population groups to compare their behavior. Larger, survey-based studies show differences in cognitive tasks, such as card sorting or accurately describing communication channels [3, 18, 7], but are unable to identify the behavioral components of those differences.

Our study focuses on single sign-on authentication protocols. These types of protocols allow a user to use their account from one site—called the *identity provider*—to login to another site — called the *relying party.* Given the proliferation of websites with which users interact, single sign-on protocols allow users to reduce the number of username/password combinations they have to remember (or, given that poor password practices are relatively common, this means users may be reusing the same username/password combination at fewer sites). This comes at the cost of trusting authentication to a single server. Moreover, login interaction becomes more complicated, as three parties — the user, the relying party, and the identity provider — are now involved in each login, and the user needs to ensure they are interacting with the right party at the right time.

LASER 2012, July 18 - 19, Arlington, VA, USA

We draw on past work on usable security as well as work in cognitive-neuroscience to design an experiment to identify the behavioral components of online decision making processes, especially in the context of user authentication on the web. Identifying these behaviors will be useful in identifying the nature of experts' tacit knowledge when it comes to making security decisions. It will also allow us to obtain some insight on the usability of single sign-on authentication.

Our study is designed to identify what sources of information and features experts use when evaluating websites. It is also designed to be mobile, allowing us to collect data from a wide range of experts. However, instead of our planned results, we instead have developed a series of guidelines on how to create an experiment when addressing technical errors, and demographic issues. The errors of researcher optimism and resulting errors were resolved in running and analyzing the experiments; and some of the lessons learned may be generalizable to other domains beyond eye tracking.

The primary recommendations we have are discussed more in depth in Section 6. However, the first, and most obvious, is ensure that one understands the technical limitations of the equipment and testing environment. When examining experts and novices, one should have a well developed method for objectively differentiating between abilities. When dealing with security, how important is it to minimize task completion bias—where subjects are more interested in completing a task than paying attention the security. To evaluate the attention paid to security, task completion bias should be minimized. In order to mimic activities of daily computing, task completion bias should be present and significant. Finally, one must be prepared to address extreme differences in the population from which participants can be recruited and the actual participant who show up.

We begin by describing the initial motivation for the experiment, in Section 2. We describe our initial assumptions and methodology in Section 3. We then integrate both the results and our critique of our implementation. Essentially, our laboratory experience and training with eye-tracking was not up to the standard needed for data compilation. A basic recommendation of this section can be found in the design of TCP: start slow. In Section 5 we delve into the problems we experienced in more depth. The Section 6 we enumerate six findings that are generalizable for the study of security cues, with a focus on our particular method. Using these changes in study design, the Indiana team was able to suggest changes to the protocol to the QUT coauthors at the Queensland University of Technology. The collaborators were able to run a successful—recently submitted—experiment comparing experts and novices using the revisions suggested in this paper.

2. BACKGROUND

Web browser security cues.

Several studies have sought to understand how users employ security cues in web browsers. Whalen and Inkpen used eye-tracking equipment to examine where users looked when browsing different websites and found that many users could identify the location of security indicators, but failed to interact with them [19]. Sobey *et al.* reported similar results when they examined users' responses to extended validation (EV) SSL certificates as provided by web browser security cues [15]. A study by Sunshine *et al.* showed that users will ignore SSL certificate warnings presented to them by web browsers [17], confirming many of the results found in Schechter *et al.*'s work.

Expertise and security decisions.

Previous studies have shown that in some scenarios people studying in technical fields are less likely to be decieved by online threats, and are more likely to understand security cues. An early study by Friedman *et al.*, examined three population groups. Their study found that, on a whole, users from a high-technology neighborhood in California were able to better describe and identify the correct representation of an encrypted communication channel, than communities that were less technical [7].

Similarly, a study by Jagatic *et al.* showed that students majoring in technical fields were unlikely to fall for standard phishing attacks and were roughly half as likely to fall for spear-phish attacks as students majoring in non-technical fields [9]. Sobey *et al.* found that expert users were able to correctly identify and interpret web browser security indicators [15]. However, Sunshine *et al.*'s study demonstrated that experts were more security-aware only in specific situations [17].

Single sign-on.

Single sign-on protocols allow a user to use their account from one identity provider to authenticate to multiple relying parties. Several standardized and proprietary single sign-on protocols exist. Within organizations, single sign-on has become quite common, and the venerable Kerberos protocol underlies modern enterprise single sign-on systems such as Microsoft Active Directory.

Our focus is on the single sign-on on the public Internet. The OpenID protocol [11] is an open standard that has seen some adoption; one notable feature of OpenID is that it allows anyone to set up an identity provider, rather than requiring a small number of centralized identity providers. The OAuth protocol [1] is a delegated authorization protocol which provides services closely related to single sign-on. There has been significant adoption of OAuth; in 2008, the social networking site Facebook introduced a single sign-on featured called Facebook Connect which employs OAuth, and in 2009 the microblogging site Twitter began requiring the use of OAuth for all interactions with its API.

Single sign-on systems have been studied to some extent in the literature. Pashalidis and Mitchell [12] gave a taxonomy of single sign-on systems. Recently, Sun *et al.* [16] did a usability study in which they explored users' understand and opinions on single sign-on systems.

3. METHODOLOGY

Our study was designed to expand on previous security usability research by providing an explicit study of computer experts to ascertain to what extent they utilize web browser security cues. Compared with previous usability studies on web browser security cues, our research is notable for explicitly comparing the behaviors of novices and experts, for working in the context of single sign-on authentication as opposed to direct logins, and for using eye-tracking equipment to gain additional insight into user behavior.

We conducted a two-part experiment with subjects who were recruited from the graduate program (Masters and PhD) at Indiana University's School of Informatics and Computing. Subjects performed a sequence of eye-tracking tasks and then completed a survey. Subjects were paid $1 for each online task completed and $8 dollars for the completion of the survey, to a maximum of $15. They could withdraw at any point and were paid for what they completed.

We assumed that subjects recruited from this graduate program would fit the description of Jagatic *et al.* of having technical expertise. Jagatic *et al.*'s study, as well as Friedmann's study demonstrated that general technical expertise was a good indicator of security awareness [9, 7]. This awareness, however, seems to be limited by task familiarity, as we find in our results.

The eye-tracking tasks asked subjects to complete a set of information sharing tasks using either Facebook Connect or OpenID and Twitter's single sign-on architecture. The survey task was designed to collect demographic information, as well as subjects' risk and benefit perceptions of sharing information online. As part of the survey task we also queried subjects on their reasons for completing or not completing parts of the eye-tracking task.

Our setup allowed us to run two experiments simultaniously within the same lab. However, there were, as we discovered later, differences in lighting, which affected post-processing.

3.1 Eye-Tracking Task

Subjects completed an eye-tracking task asking them to share information via Facebook Connect or OpenID and Twitter. They were randomly assigned to either the Facebook Connect tasks or the OpenID and Twitter tasks. The list of tasks appears in Table 3.1. Moreover, the order of tasks within the list was also randomly shuffled for each user. Subjects viewed webpages using Firefox 4.0 on Dell laptops with 12in monitors running Windows Vista.

We used an eye-tracking device developed by Thomas Busey at Indiana University's Department of Psychology and Brain Science and analyzed the eye-tracking data using the open source software, Expert Eyes [4]. This eye-tracking device consists of two cameras mounted on a pair of glasses which the participant wears, with one camera recording the subject's eye movements and the other camera recording the screen. Via post-processing, the video recordings are correlated and a video is produced showing the subject's gaze overlaid on the screen recording. We attempted to reduce head movement by using a chin rest. See Figures **??** and **??** for examples of the captured eye recording and corresponding video recording with gaze overlaid.

We chose this particular eye-tracking hardware/software combination due to its portability. Our protocol was meant to collect data from a large number of subjects with a wide demographic background, and in many cases subjects are unable to travel to a university to act as subjects in an experiment. For this particular experiment, we only had IRB approval to conduct research with experts recruited from Indiana University. Since then, our protocol has been expanded to allow us to collect data from experts and novices in many different settings, including security conferences and the local farmer's market. This is meant to give us a better external validity to our study.

Subjects used pregenerated logon information for all accounts to minimize exposure of subjects' identities. Our OpenID identity provider was not configured to use SSL, thus all OpenID login information was passed unsecured through HTTP; Facebook and Twitter login did employ HTTPS on the single sign-on authentication page. After a subject finished the online tasks, we cleared the web browser's cache and deleted all cookies. We informed subjects that they need not complete the sharing task, but if they did not complete the task they needed to explain why they did not do so in the following survey task in order to be paid for the task.

Before the subjects began the information sharing tasks, we calibrated the eye-tracking equipment using a series of dots on which they focused (Figure 1). We repeated the calibration after the eye-tracking task to ensure that the equipment did not slip during the task, and to correct for slippage, if it did occur.

Figure 1: Sample Calibration Point

3.2 Survey Task

After subjects had finished the eye-tracking task we asked them to complete several surveys.

The first survey was a task completion checklist, which included questions to obtain qualitative information on why subjects may not have completed a task. For each task, we asked subjects to inform us if they were able to complete an online task. For those tasks which the subjects did not complete, we asked them to give us a reason for why they could not complete the task. This allowed us to identify difficulties in our online tasks, as well as possible security reasons subjects declined to finish a given task.

The second survey collected subjects' demographic information. The information collected included gender, age, household income, and education level, as well as how often subjects used the internet and their browser and operating systems preferences.

For the third survey, subjects were randomly assigned to complete one of two possible surveys: either a survey asking subjects to report their perceptions of the risks of sharing information online, or a survey on subjects' perceptions of the benefits of sharing information online. For this study, we used the online task portion, the demographic information, and a rudimentary analysis of subjects' risk/benefit perceptions. A more thorough investigation of the risk/benefit

Facebook	**OpenID and Twitter**
Share a story on CNN using Facebook Connect (F1)	Post a comment on LiveJournal using OpenID (O1)
Comment on a CNN story using Facebook Connect (F2)	Post a comment on a blog at Blogspot using OpenID (O2)
Post a comment on LiveJournal using Facebook Connect(F3)	Post a review on SourceForge using OpenID (O3)
Rate a movie on RottenTomatoes and post it on Facebook (F4)	Share a story on CNN using Twitter (O4)
Share an item from Amazon using Facebook Connect (F5)	Comment on a CNN story using Twitter (O5)
Login to Yahoo! Mail using Facebook Connect (F6)	Share an item from Amazon on Twitter (O6)
Import friends from Yahoo! Mail into Facebook (F7)	Rate a movie on RottenTomatoes and post a story about it on Twitter (O7)

Table 1: List of information sharing tasks for eye-tracking task.

Age	Male	Female
19-25	9	1
26-30	2	1
31-35	1	0

Table 2: Age ranges given by subjects.

Household Income	Male	Female
$x < 10000$	3	0
$10000 < x < 20000$	5	0
$20000 < x < 30000$	2	0
$50000 < x < 75000$	1	1
Don't Know	1	1

Table 3: Household Income reported by subjects.

portion is part of another, ongoing study.

4. RESULTS AND TECHNICAL ISSUES

We conducted the experiment on 14 subjects (12 male, 2 female), all of whom completed the entire experiment. Most subjects were between the ages of 19 and 25 (Table 2). Household income was highly variable with income ranging from less than $10,000 to at most $75,000, with most subjects making between $10,000 and $20,000 (Table 3). All subjects were graduate students at the School of Informatics and Computing at Indiana University, but most subjects had completed at least one year of a master's degree (Table 4).

All subjects used the internet on a daily basis and most were familiar with multiple web browsers and operating systems (Tables 5 and 6). Chrome was the most popular exclusive browser, with three subjects reporting only using Chrome. Windows was the most popular exclusive operating system with two subjects exclusively using Windows. No one exclusively used a Linux OS.

Our initial recruitment of subjects focused on graduate students, and given the broad experience demonstrated using various web browsers and operating systems, we feel they represent good examples of technical expertise, not necessarily computer security expertise. In Jagatic *et. al*, found that technical expertise was capable of mitigating the threat of phishing attacks, however, we found that this expertise is highly contextual [9].

Education Level	Male	Female
College Graduate	3	0
Master's Degree	6	2
Ph.D.	3	0

Table 4: Education levels reported by subjects.

Web Browser	Male	Female
Firefox	5	2
IE	5	1
Safari	3	0
Opera	1	0
Chrome	10	1

Table 5: Number of subjects using a given web browser.

OS	Male	Female
Windows	9	1
Macinotsh	7	1
Linux	5	2

Table 6: Number of subjects using a given OS.

4.1 Eye-Tracking Issues

Luckily, we only have one non-human related technical issue. On the first subject we ran, one of the wires in the camera was loose, leading to interesting footage of that subject's eye. However, after that subject was done, we fixed the eye-tracking equipment and the problem was resolved.

4.1.1 Problems with Small Screens

One of the critical limitations we ran into was the use of a laptops with overly small monitors. While subjects were able to complete the tasks, the cameras were not of a high enough resolution to capture more accurate information from the screen. This resulted in a significant amount of guessing of the location of a subject's gaze, which was crucial to our experiment.

Another problem with the small screen size was it pulled subjects' gaze overly downward. This led to significant parts of the data becoming inaccessible because we could not create an eyefit model when the pupil was obscured (Figure 2. This sort of data error occurred in 7 of the remaining 13 subjects, and —removing our protocol aware subject— 6 of the 12 subjects were afflicted. In some of the cases, we could still extract some information, but it made the eye fit model much more complicated.

4.1.2 Focus and Lighting Issues

We had focus issues in three cases, two of these cases were in data that had not been affected by the previously mentioned difficulties. We either positioned the eye camera too far from the subject's eye, or focused on the wrong part

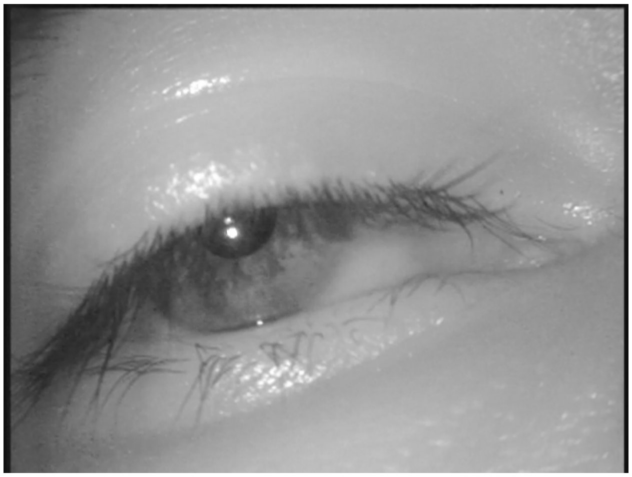

Figure 2: No, this person is not asleep. They're trying to complete the online tasks. The pupil is obscured making calibration of the eye gaze difficult.

of the subject's eye. In these cases, we also noticed that our second setup was not as well lit as the first computer setup. This led to grainy resolution, which confused Expert Eyes when trying to identify pupil locations.

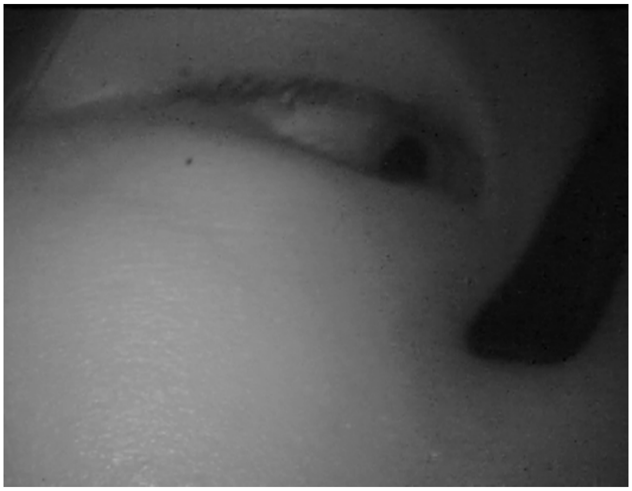

Figure 3: Poor lighting and focus make identifying pupil locations difficult.

The Expert Eyes software uses dark-pupil illumination to detect the location of the pupil against the rest of the eye [4]. Thus, when there is not enough contrast between the pupil and the background (Figure 3), it is difficult to produce a good eye-fit model.

Altogether, the various technical issues resulted in only two eyetracking subjects' data being completely analyzed. One of these knew the protocol leading to an effective sample size of one. This required us to rely on informal observations rather than accurate data. These informal observations, however, led to important realizations about our survey methodology and demographic issues.

4.1.3 Online Task Completion

Task	Completed Online	Completed on Survey
F1	8	1
F2	9	0
F3	9	0
F4	9	0
F5	8	1
F6	9	0
F7	9	0

Table 7: Number and method of subjects completing each Facebook task.

Task	Complete Online	Completed on Survey
O1	3	1
O2	2	2
O3	2	2
O4	4	0
O5	4	0
O6	4	0
O7	4	0

Table 8: Number and method of subjects completing OpenID and Twitter tasks.

Most subjects were able to complete the Facebook tasks, at least after instruction. There were two subjects that had difficulty due to technical issues when trying to complete the Facebook tasks. See Table 7 for the exact numerical results. The reasons given for not being able to complete the Facebook tasks were:

1. "The password is always wrong" (F1)
2. "There was some problem with the facebook password. It kept asking password everytime while trying to share even after logging in" (F5)

On the otherhand, subjects seemed to have great difficulty in completing the OpenID and Twitter tasks. In particular, the OpenID tasks proved troublesome to subjects. Task O1 only had 60% online completion, while tasks O2 and O3 were only completed by 40% of the subjects. The Twitter tasks, however, were completed by all subjects assigned the OpenID and Twitter tasks.

The online completion of OpenID tasks is slightly less disappointing as our knowledgable subject was selected for the OpenID and Twitter tasks. That subject was able to correctly identify the lack of security in our OpenID identity provider, and listed that as the reason for not completing those tasks. Table 8 shows the results when we remove the knowledgeable subject from the data. However, removing the knowledgeable subject meant that we were left with only four subjects being selected for the OpenID and Twitter tasks — less than half of the subjects chosen for the Facebook tasks.

Aside from the security reasons given by our knowledgeable subject, subjects' reasons for not completing the online task were:

1. "I thought the jobs were done. I did not realize the task is not complete." (O1-O3)
2. "OpenID didn't work." (O2)

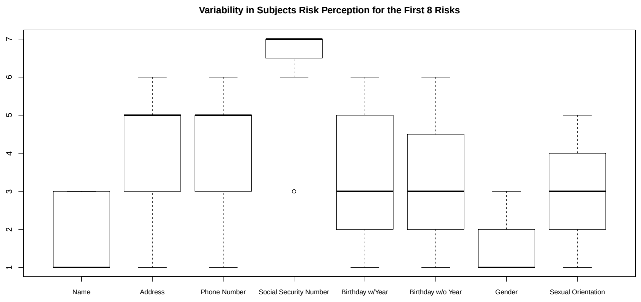

Figure 4: Subjects' risk perception of the first eight categories given. 1 is least amount of risk; 7 is greatest.

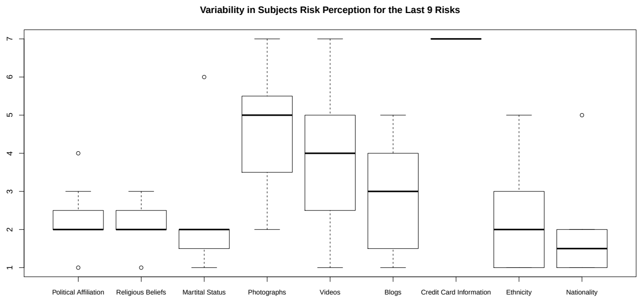

Figure 5: Subjects' risk perception of the last nine categories given. 1 is least amount of risk; 7 is greatest.

3. "The account did not have credentials to review a project." (O3)

4.2 Demographic and Human Factors

In addition to the disappointing male-female ratios in our study, most subjects were either from India or China. Facebook is not the primary social networking site in either of those countries. This meant that in most cases, we had to instruct subjects on how to navigate Facebook, OpenID, or Twitter in order for them to be able to complete the tasks.

Our instruction led to an unnatural use environment that, given our informal observations, distracted subjects. For example, in four subjects we noticed that the subjects were following our fingers as we showed them which buttons to click on to share stories and comment on CNN. Normal browsing habits do not typically involve following a human's finger as they tell one where to click.

4.2.1 *Risk/Benefit Perceptions*

We performed a rudimentary analysis of the subjects' risk/benefit perceptions to acertain any consensus in how subjects percieve sharing information online. Subjects were asked rate a collection of 17 different information categories on either its perceived risk or benefit when sharing that information online. Each category could take a value from 1–least beneficial/risky–to 7–most beneficial/risky.

In terms of risk, there was strong agreement that sharing social security numbers (Figure 4) and credit card information (Figure 5)was risky, while sharing marital status and nationality were viewed as low risk (Figure 5). In terms of benefits, however, there was little consensus. It appears that there is a fairly strong agreement that sharing one's gender online is of moderately high benefit (Figure 6).

5. DISCUSSION

This choice of technology for this study was a key part of the design. We chose the eye-tracking equipment because of its portability. This will allow us to gather data from a wider variety of experts and novices in different locations, leading to greater external validity. However, the choice to gather data from a diverse population also exposed limitations in our methodology.

Our inexperience with the technology was a hinderance to data collection, but it does highlight the necessity in understanding the equipment. More importantly though, it demonstrates the need to control the experimental environment to ensure that the results are valid. The failure to control our laboratory environment due to unforseen circumstances is a critical learning experience for adapting the equipment to mobile data collection.

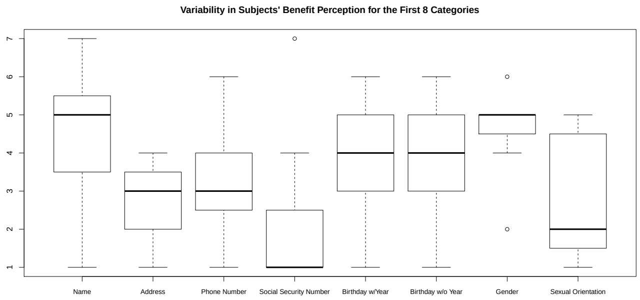

Figure 6: Subjects' benefit perception of the first eight categories given. 1 is least amount of benefit; 7 is greatest.

For example, our use of two testing areas allowed us to collect data more quickly, but it illuminated the difficulties in controlling lighting. The lighting became an issue because our subjects were unfamiliar with the tasks and required our instruction to complete the tasks. Our presence behind subjects, while instructing them, obscured the light in one of the two testing locations. If our subjects were familiar with how to complete the tasks, we may not have discovered the problems until we began mobile data collection.

The desire to collect data from a diverse population also underscored the necessity for objective and culturally aware metrics. Our experiment limited users to Firefox 4.0 and Windows Vista, but, as our results show, only half of the subjects used Firefox as a browser. Moreover, most of our subjects were unfamiliar with Facebook and OpenID.

Ideally, we would like only subjects familiar with the technologies being tested, as suggested in Section 6. However, it may be more informative to identify subject reported familiarity with the technologies being tested. This would lead to more complicated data analysis, but might also help identify differences in security expertise and task expertise.

Another consideration is incentive. We paid subjects $1 for each online task completed. Completion of a task—as

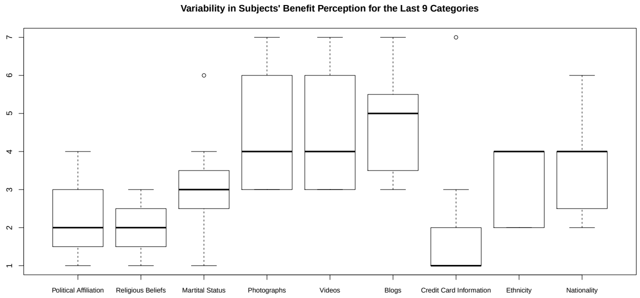

Figure 7: Subjects' benefit perception of the last nine categories given. 1 is least amount of benefit; 7 is greatest.

we instructed subjects—was either doing the task online, or explaining to us why they did not complete the task online. This may have induced subjects to complete the tasks, rather than pay attention to security cues. We mention in Section 6 that the option used in Queensland University of Technology was to pay subjects the full amount regardless of how many tasks they completed.

In work by Busey *et al.*, however, they chose not to pay subjects at all [5]. This allowed them to recruit subjects that were self-motivated to participate, rather than inticing subjects. This would have the advantage of limiting motivational biases in task completion.

Our study used generated identies for subjects, rather than having subjects use their own credentials. This may have lead subjects to pay less attention to the security cues as their was no possible damage to their accounts. We suggest in Section 6 that users should use their own accounts if they want, but provide them with generated accounts if they either do not want to use their own, or if they do not have accounts for the technologies being studied.

Another aspect of experts and novices that emerged in thinking about the data was speed of information acquisition. Security usability eye-tracking studies generally look at whether subjects look at security cues provided by the browser, and for how long. As we discussed further ways to analyze the data, it was mentioned that experts may not observe cues for as long as novices, and that it may be difficult to find clusters of viewing patterns because experts can pick out information much quicker than novices [10].

This phenomena has been demonstrated in many different fields from igo—an oriental strategy game—to finger print identification [8, 21]. This suggests that further research into experts and novices in computer security may not be able to find any significant differences between gaze locations, or even gaze durations, but that the dynamics of the gaze patterns may be a more accurate indicator of the differences between the groups [21]. While we did not incorporate a method for analyzing the gaze dynamics in our Queensland University of Technology study, we are working with Tom Busey to develop a methodology to use for our existing data and data we acquire in the future.

This type of behavior also suggests further avenues of research in usable security. For example, if experts are taking in security information from security cues in browser chrome, do they suffer from change blindness making them more susceptible (or just as bad as novices) when the security cues are spoofed [6]?

6. CHANGES TO STUDY DESIGN

In addition to ensuring that the technical issues observed in Section 4 are resolved, we make several recommendations on how the study design should be changed to obtain valuable results.

1. *Ensure a sufficiently diverse study population.* Our study participants were mostly from either China or India, where Facebook was not the primary online social network, and thus participants were unfamiliar with using Facebook. Recruitment for this type of study should focus on users that regularly employ the social networking tools used in the study.

   In any study, prepare for unexpected distributions of participants, and attempt to over-recruit.

2. *Where possible, have participants use their own login credentials.* In our study, participants were provided with pre-generated usernames and passwords so that they did not have to use their own login credentials. This is beneficial from a privacy perspective as subjects do not risk revealing personal information during the study. However, previous research, such as that by Schechter et al. [13], has shown that participants using their own login credentials tend to be more security-aware than participants using provided login credentials. This requires greater sensitivity in data handling to protect participant privacy and obtain ethics approval, but may place participants in a more realistic setting.

3. *Classify subjects as experts or novices based on responses to skill-testing questions rather than self-reported opinions.* Subjects were classified as experts or novices based on their self-reported expertise. Individuals may have different opinions on what qualifies as expertise or may be biased about their own expertise.

   Research in security usability would benefit from a standardized instrument assessing technical expertise.

4. *Include non-single-sign-on tasks to obtain baseline information on use of security cues.* Our study was designed to assess how subjects use security cues during single sign-on, but our study protocol did not include any tasks that involved authentication without single sign-on. As a result, we are unable to compare whether participants use security cues in single sign-on more or less than in direct authentication (non-single-sign-on) settings.

5. *Allow users to choose which web browser to use.* Individuals use different web browsers, with no web browser currently having more than 30% market share. As a result, requiring subjects to use one web browser reduces the ecological validity of the study. Allowing subjects to choose which web browser to use would place them in a more natural computing environment, albeit at the cost of greater complexity in data analysis.

6. *Pay subjects regardless of number of tasks completed.* We originally chose to pay subjects on a per-task basis

to induce subjects to complete the study instead of collecting their money and walking away. However, this can introduce a task focus, where subjects aim to complete the required tasks with little heed to other factors. This is particularly problematic in the context of security research, as task-focused-users may ignore security cues with their focus on task completion.

7. CONCLUSIONS

The goal of the original experiment was to present motivating background work and an innovative way to explore these. Indeed the second experiment did, hopefully, reach these goals. Serendipitously the experiment we developed had been expected to be implemented in two cultural contexts; however, the initial context remains unexplored. However, during our learning process we developed a set of heuristics to guide the remaining experimentation. We enumerate those above. In addition we recommend a slow start, with one or two participants for unpublished experimentation to begin. Our recommendation that there be a uniform instrument for assessing technical expertise is problematic. Those standard evaluation instruments become archaic almost more quickly than these can be accepted in the academy. Due to the rate of change in technology, a different approach to standardized instruments is needed. This is exacerbated that such an instrument would ideally integrate computer science, psychology, security, and sociology.

8. ACKNOWLEDGEMENTS

We would like to acknowledge the great help Brandi Emerick and Tom Busey were in teaching us how to use the eyetracking equipment and working with us to deal with data analysis problems. The errors in this study were entirely the responsibility of the lead Indiana faculty; the patience was from Tom and Bradi; the lessons learned to QUT; and the painful but fruitful task of learning, the entire IU team.

9. REFERENCES

[1] The OAuth 1.0 Protocol, Apr. 2010.

[2] C. L. Anderson and R. Agarwal. Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioral Intentions. *MIS QUARTERLY*, 34(3):613–643, Sept. 2010.

[3] F. Asgharpour, D. Liu, and L. J. Camp. Mental models of computer security risks. *Workshop on the Economics of Information Security*, 2007.

[4] J. S. Babcock and J. B. Pelz. Building a lightweight eyetracking headgear. In *Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004*, pages 109–114, New York, New York, USA, 2004. ACM Press.

[5] T. a. Busey and J. R. Vanderkolk. Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision research*, 45(4):431–48, Feb. 2005.

[6] P. J. Durlach. Change blindness and its implications for complex monitoring and control systems design and operator training. *HUMAN-COMPUTER INTERACTION*, 19(4):423–451, 2004.

[7] B. Friedman, D. Hurley, D. C. Howe, E. Felten, and H. Nissenbaum. Users' conceptions of web security. In *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*, page 746, New York, New York, USA, 2002. ACM Press.

[8] K. Itoh, H. Kitamura, Y. Fujii, and T. Nakada. Neural substrates for visual pattern recognition learning in Igo. *Brain research*, 1227:162–73, Aug. 2008.

[9] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50(10):94–100, Oct. 2007.

[10] J. Kruschke. Personal conversation with John Kruschke of Indiana University's Department of Psychology and Brain Sciences, 2011.

[11] OpenID Foundation. Specifications, 2010.

[12] A. Pashalidis and C. J. Mitchell. A Taxonomy of Single Sign-On Systems. In R. Safavi-Naini and J. Seberry, editors, *Proc. 8th Australasian Conference on Information Security and Privacy (ACISP) 2003*, volume 2727 of *LNCS*. Springer, 2003.

[13] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The Emperor's New Security Indicators. *Security and Privacy, IEEE Symposium on*, 0:51–65, 2007.

[14] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs. Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 373–382. ACM, 2010.

[15] J. Sobey, R. Biddle, P. van Oorschot, and A. S. Patrick. Exploring User Reactions to New Browser Cues for Extended Validation Certificates. In S. Jajodia and J. Lopez, editors, *Proc. 13th European Symposium on Research in Computer Security (ESORICS) 2008*, volume 5283 of *LNCS*, pages 411–427. Springer, 2008.

[16] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov. What makes users refuse web single sign-on?: an empirical investigation of {OpenID}. In L. F. Cranor, editor, *Proc. 7th Symposium on Usable Privacy and Security (SOUPS) 2011*, pages 4:1—-4:20. ACM, 2011.

[17] J. Sunshine, S. Egelman, H. Almuhimedi, N. Atri, and L. F. Cranor. Crying Wolf: An Empirical Study of {SSL} Warning Effectiveness. In *Proc. 18th {USENIX} Security Symposium*, 2009.

[18] R. Wash. Folk Models of Home Computer Security. In *Symposium on Usable Privacy and Security (SOUPS)*. ACM Press, 2010.

[19] T. Whalen. Gathering evidence: use of visual security cues in web browsers. In *Proceedings of Graphics Interface 2005*, pages 137–144, 2005.

[20] R. T. Wright and K. Marett. The Influence of Experiential and Dispositional Factors in Phishing: An Empirical Investigation of the Deceived. *Journal of Management Information Systems*, 27(1):273–303, July 2010.

[21] C. Yu, T. Busey, and J. Vanderkolk. Discovering Correspondences between Fingerprints Based on the Temporal Dynamics of Eye Movements from Experts. In H. Sako, K. Franke, and S. Saitoh, editors,

Computational Forensics, volume 6540 of *Lecture Notes in Computer Science*, pages 160–172. Springer Berlin / Heidelberg, 2011.