Proceedings of the Ninth International Symposium on
Human Aspects of Information Security & Assurance (HAISA 2015)

Tracking Risky Behavior On The Web: Distinguishing Between What Users 'Say' And 'Do'

T. Kelley and B. I. Bertenthal

Developmental Cognitive Neuroscience Laboratory
Department of Psychology and Brain Sciences, Indiana University, Bloomington
Indiana, United States of America
e-mail: kelleyt@indiana.edu

Abstract

Modern browsers are designed to inform users as to whether or not it is secure to login to a website, but most users are not aware of this information and even those that are sometimes ignore it. The goal of this research is to assess users' knowledge of security warnings communicated via browser indicators (e.g., https, lock icon in the status bar), and the likelihood that their online decision making adheres to this knowledge. A large sample of participants was recruited from Amazon's Mechanical Turk and their knowledge of cybersecurity was assessed with an online survey. These participants were also instructed to visit a series of secure and insecure websites, and decide as quickly and as accurately as possible whether or not it was safe to login. The results revealed that knowledge of cybersecurity was not necessarily a good predictor of decisions regarding whether or not to sign-in to a website. Moreover, these decisions were modulated by attention to security indicators, familiarity of the website, and psychosocial stress induced by bonus payments determined by response times and accuracy. We suggest that even individuals with security knowledge are unable to draw the necessary conclusions about digital risks while browsing the web. Users are being educated through daily use to ignore recommended security indicators and we surmise that the lack of conformity in website conventions contributes to this behavior.

Keywords

Information security, browser login, security expertise, Mechanical Turk, experiment

1. Introduction

Users on the Internet are regularly confronted with complex security decisions that can affect their privacy. They must decide whether it is safe to enter their username, password, credit card details, and other personal information on websites with very different interfaces and only a few visual clues on whether it is safe to do so. These security indicators include the protocol used, the domain name, the SSL/TLS certificate, and visual elements in the browser window. Very few users understand the technical details of these various indicators.

Not surprisingly, users often get it wrong, either ignoring security indicators completely or misunderstanding them. Many popular websites' are designed in such a way that these indicators are displayed in a suboptimal way, further complicating users' decision making process (Stebila 2010). Moreover, these websites can appear

confusing, because they include no or only partial encryption, but users will treat them as secure even without security indicators if they have been previously visited (Hazim *et al.* 2014). This confusion is due to the manner in which security information is typically deployed, i.e., as communication between technical experts (Garg and Camp 2012).

While several studies have evaluated whether users correctly use security indicators, there has been very little work investigating whether their knowledge of these indicators will predict their behavior (Schechter *et al.* 2007). One reason for this predicament is that it is challenging to design behavioral studies that will realistically simulate the conditions that a user would experience on the Internet (Arianezhad *et al.* 2013).

One real-world condition that is particularly difficult to replicate in an experimental environment is the experience of risk. Many studies ask participants to assume the role of someone else to avoid exposing participants to real risks (Schechter *et al.* 2007, Sunshine *et al.* 2009). Other studies use priming—alerting participants to the fact the study is interested in behaviour related to security—to induce secure-like behaviour (Whalen and Inkpen 2005). It is unlikely, however, that participants playing roles behave as securely as they would when they are personally at risk.

A different strategy is to use monetary incentives and penalties as a method for creating risky decisions. We utilize participants' assumed goal of maximizing payment to put pressure on the participant to act as quickly as possible by offering participants a bonus payment that decreases as the total elapsed time increases.

2. Methodology

By introducing a performance bonus based on both speed and accuracy in completing the task (Figure 1), we sought to increase the motivation and risk taking behaviour of participants (Petzold *et al.* 2010). Our primary question was whether users would ignore or simply miss security indicators when pressed for time. In order to address this question, we wanted a relatively large sample with a broad distribution of knowledge concerning security indicators.

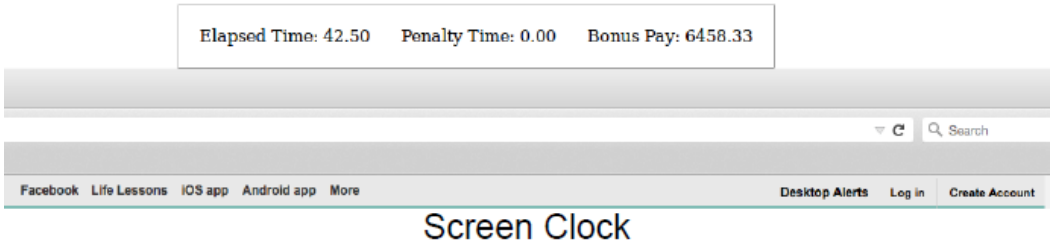

Figure 1: Screenshot of top of experimental task instructions. Note the presence of the sample clock resting on top of the simulated browser chrome

2.1. Participants

The sample consisted of 173 participants ranging in age from 18- to 76-years-old *(M* = 32.6, *SD* = 9.58*)* recruited from Amazon's Mechanical Turk (AMT). Studies have shown that AMT provides more diverse study populations and robust findings in numerous psychological paradigms (Buhrmester *et al.* 2011, Crump *et al.* 2013). There were 100 males and 73 females, primarily Caucasian. Most participants listed Firefox *(N* = 84*)* or Google Chrome *(N* = 81*)* as their primary browser.

2.2. Stimuli

Each trial simulated websites appearing on a Firefox browser. In order to standardize all websites, logins always appeared on the second page of the website. All websites were manipulated in a graphical editing program and presented to participants in a popup window with disabled user interface chrome to minimize confusion between the proxy websites' chrome and their actual browser chrome. This also prevented participants from manipulating the experiment by reloading pages or navigating back and forward outside of our simulated website user interface.

2.3. Procedure

Participants were instructed to decide whether or not to login to a series of websites depending on whether or not they were judged to be secure. The goal was to visit all the websites as quickly as possible, and the pay for completing this task was contingent on how quickly it was completed. If a participant clicked to login to a secure website, the screen advanced to the next one. If a participant did not click to login to a secure website and instead pressed the back button, a penalty screen was displayed for 20 sec and that time was added to their cumulative time. If a participant pressed the back button and the website was insecure, the screen advanced to the next website. If, however, a participant clicked to login to an insecure website, the penalty screen was displayed for 10 sec and that time was added to their cumulative time.

An online survey assessing participants' knowledge concerning security indicators was administered after the experimental task so as not to bias participants' performance. There were three categories of questions: 1) Demographic information (e.g., age, gender, education level), 2) Applied security knowledge (e.g., security indicators, password behaviour), and 3) Technical security knowledge (e.g., DDoS, Phishing, Firewalls).

2.4. Design

This study addressed two questions: 1) Do web security indicators affect participants' behaviour when discerning the safety of encrypted vs. unencrypted websites, and 2) Do web security indicators affect participants' ability to discern between spoofed vs. not spoofed websites. The first question was tested by manipulating whether the security indicators included http or https (https/http

manipulation). The second question was tested by manipulating whether or not the website was spoofed with an incorrect domain name (no-spoof/spoof manipulation). There are four different levels of encryption information displayed by web security indicators:

1. Extended Validation (EV) – green lock and https – full encryption; Extended vetting by certificate authority
2. Full Encryption (FE) – grey lock and https – full encryption; domain validation only
3. Partial Encryption (PE) – triangle with exclamation mark; some (unknown) elements of website encrypted
4. No Encryption (NE) – globe; no encryption of the displayed page

For the spoof manipulation we included all four levels for both spoof and no-spoof websites, but this was not possible for the https/http manipulation because unencrypted websites (http) only display a globe (NE), whereas the encrypted websites (https) display the three other security symbols listed above (1-3). Thus the https/http and no-spoof/spoof manipulations were analysed separately in this study.

Each participant was presented with 16 trials, 8 corresponding to each security manipulation condition (https/http vs. no spoof/spoof). Four trials corresponded to secure websites (https/no spoof) and 4 corresponded to insecure websites (http/spoof). For the https/http manipulation, each secure website included 1 of the 3 valid levels of encryption information (EV, FE, or PE), whereas each insecure website included only the NE indicator. For the spoof/no spoof manipulation, the 4 secure and 4 insecure trials each corresponded to one of the 4 encryption information levels. The secure and insecure websites were counterbalanced between participants and the presentation order of the websites was randomized.

2.5. Metrics and Data Reduction

Applied security knowledge was computed from the number of correct and incorrect security indicators identified in the survey $(\#\, correct\ indicators + 1)/(\#\, incorrect\ indicators + 1)$ resulting in an indicator score ranging from [0.2, 4.0], with a log-normal distribution $\ln N(M = 0.14, SD = 0.58)$.

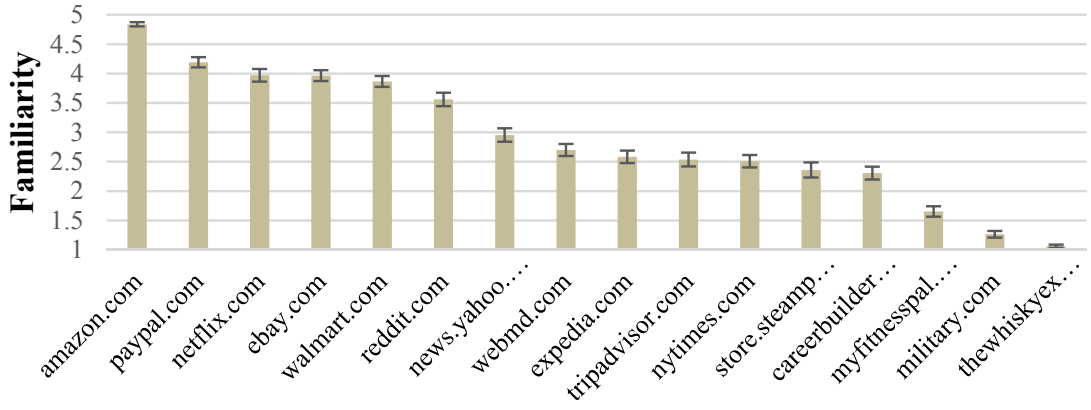

Figure 2: Mean familiarity for each website used in our study

Technical security knowledge was scored from 1 to 5 depending on the number of survey questions answered correctly (0%-20%=1, 21%-40%=2…..). Participants scoring greater than 60% (N = 50) were identified as "High Technical Security Knowledge" (Hi-Knowledge) and participants scoring 60% or less (N = 123) were identified as "Low Technical Security Knowledge" (Lo-Knowledge).

Familiarity of the websites was rated on a 5-point Likert scale. The mean rating was 2.90, and it ranged from a low of 1.00 to a high of 5.00 (Figure 2).

3. Results

The primary question concerned how frequently participants would login to insecure websites. Overall, they were more accurate responding to encrypted than to unencrypted websites (*Mdiff* = 0.55, *95% HDI* = 0.45, 0.69) and to non-spoofed than to spoofed websites (*Mdiff* = 0.46, *95% HDI* = 0.37, 0.57). Critically, the results revealed a strong response bias to login regardless of available security indicators (Figure 3). Participants' lack of sensitivity to the available stimuli was reflected in the relatively low d' in both the https/http manipulation (M = 0.41, SD = 0.66) and the no spoof/spoof manipulation (M = 0.34, SD = 0.67).

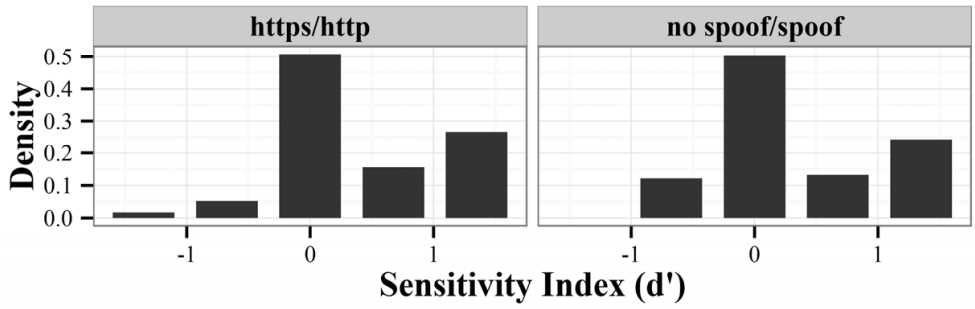

Figure 3: Participants' response bias towards login

Participants performance on the http/https websites was analysed by assessing the percent of accurate logins as a function of technical security knowledge (lo vs. hi) and security manipulation (http vs. https). A 2 x 2 repeated measures analysis of variance (ANOVA) revealed significant main effects for both security manipulation ($F(1,1368) = 382.5$, $p < 0.001$ and technical security knowledge , $F(1,1368) = 3.88$, $p < 0.05$. As can be seen in Figure there was also a significant interaction between manipulation and technical security knowledge $F(1,1368) = 6.94$, $p < 0.01$, because participants with high knowledge were more accurate than those with low knowledge in the https condition (*Mdiff* = 0.12, *95% CI* = 0.02, 0.21, $p < 0.01$), but technical security knowledge had no effect in the http condition (*Mdiff* = -0.02, *95% CI* = -0.11, 0.08, $p > 0.96$).

In a separate analysis of encryption information in the https condition, encryption was found to have a main effect $F(1,676) = 19.56$, $p < 0.001$, but was not involved in any interactions. As can be seen in Figure 4, participants were more accurate in the lock than the no lock encryption condition, The presence of encryption information (FE and EV) led participants to be more accurate (*Mdiff* = 0.12, *95% CI* = 0.06, 0.18, $p < 0.001$), as seen in Figure .

Participants' performance on the no spoof/spoof websites was analysed similarly, but included level of encryption information as a third independent variable. An ANOVA revealed a main effect for manipulation (no spoof vs. spoof), $F(1,1352) = 342.42$, $p < 0.001$ and technical security knowledge (Hi vs. Lo) $F(1,1352) = 41.28$, $p < 0.001$, but not encryption information, $F(3,1352) = 0.79$, $p > 0.37$. Encryption information, however, did interact with security manipulation $F(1,1352) = 36.78$, $p < 0.00$, and there was also an interaction between security manipulation and technical security knowledge, $F(1,1352) = 12.89$, $p < 0.001$.

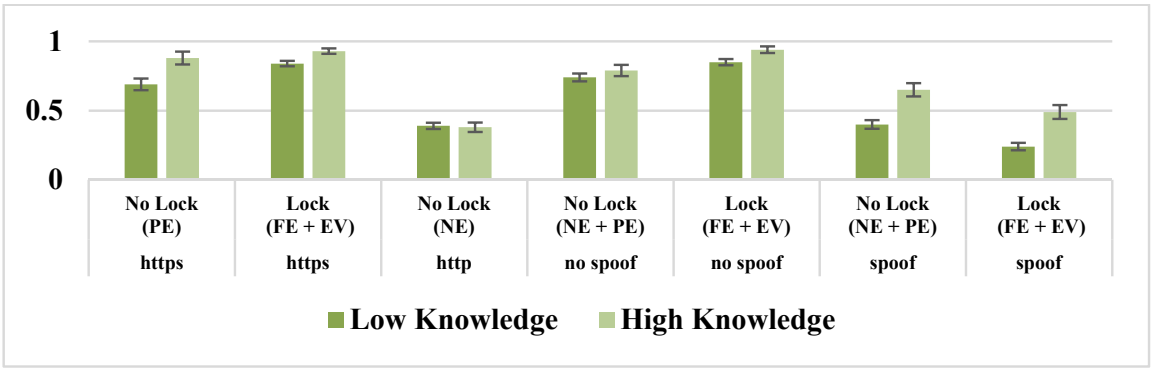

Figure 4: Differences in accuracy by manipulation (https/http and no spoof/spoof), and technical security knowledge

As can be seen in Figure 4, the two-way interaction is accounted for by participants performing more accurately in the no lock than lock conditions at the spoof websites which is opposite their performance at the no spoof websites. The presence of a lock (FE + EV) led to greater accuracy than the lack of a lock (PE + NE) (*Mdiff* = 0.12, *95% CI* = 0.035, 0.20, $p < 0.05$), but that the presence of a lock in the no spoof condition negatively impacted participants' accuracy (*Mdiff* = -0.16, *95% CI* = -0.24,

0.08, $p < 0.001$*)*. The second interaction shows that those with high technical security knowledge were, in general, more accurate than those with low knowledge *(Mdiff* = 0.16, *95% CI* = 0.11, 0.21, $p < 0.001$*)*, but the difference in accuracy due to knowledge did not occur in the no spoof condition *(Mdiff* = 0.07, *95% CI* = -0.02, 0.16, $p > 0.18$*)*, rather, it was found in the spoof condition *(Mdiff* = 0.25, *95% CI* = 0.16, 0.34, $p < 0.001$*)*.

3.1. Effects of Indicator Scores and Website Familiarity

In order to assess whether knowledge of web browser security indicators or website familiarity interacted with participants decisions to login to secure and insecure websites, we added two covariates (indicator score and website familiarity) to the previous analyses.

For the https/http manipulation, the only covariate that had a main effect was indicator score $F(1,1368) = 4.42$, $p < 0.05$, with a higher indicator score correlated with higher accuracy $r_\tau(398) = 0.05, p < 0.05$. There was also a two-way interaction between familiarity and manipulation $F(1,1368) = 16.52$, $p < 0.001$. Familiarity increased accuracy in the https manipulation, but decreased it in the http manipulation (Figure 6). These patterns were modulated by technical security knowledge, leading to a four-way interaction between security knowledge, https/http manipulation, and both indicator score and familiarity, $F(1,1368) = 4.22$, $p < 0.05$. Knowledge of security indicators increased accuracy in the http condition, while familiarity decreased it. Accuracy was unaffected, or reduced, by indicator score in the https condition depending on the presence of encryption information (Lock (FE + EV) vs. No Lock (PE)). Familiarity increased accuracy in the https condition, particularly for lo-knowledge participants, with encryption information present (FE + EV) (Figures 5 and 6).

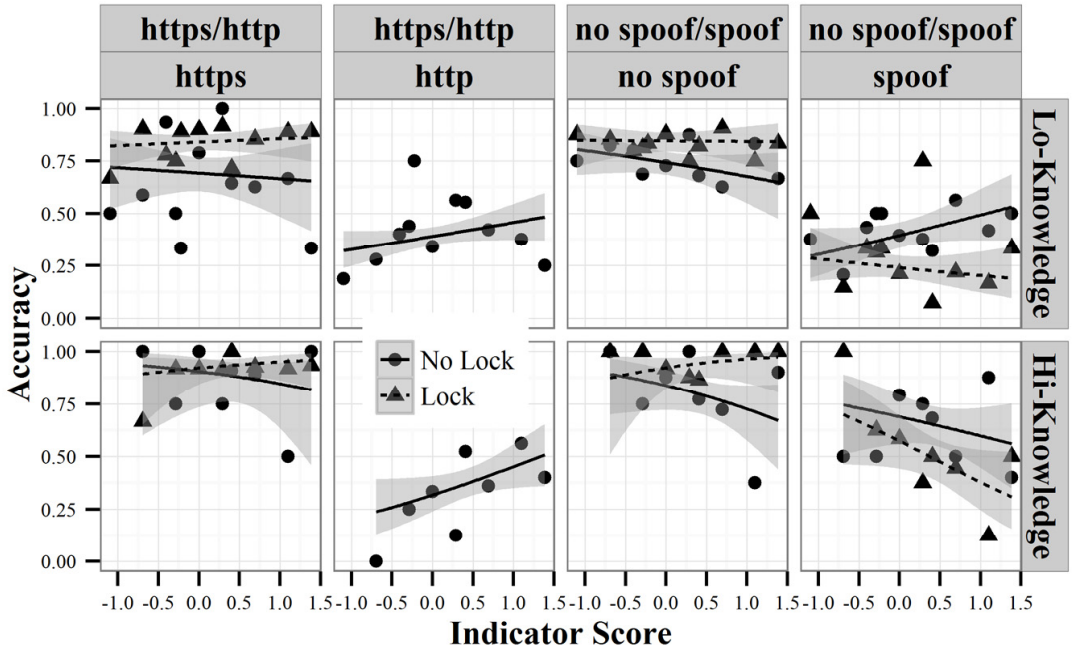

Figure 5: Relationship between participants' use of security indicators and their proportion of correct responses

For the no-spoof/spoof manipulation, there was a main effect of familiarity, $F(1,1352) = 11.61$, $p < 0.01$. Familiarity also interacted with manipulation, $F(1,1352) = 6.42$, $p < 0.05$, as well as with both manipulation and technical security knowledge $F(1,1352) = 7.24$, $p < 0.01$, and there was a four-way interaction with manipulation, technical security knowledge, and encryption information $F(1,1352) = 4.64$, $p < 0.05$.

As observed with the https/http manipulation, familiarity drives logins, but unlike the https/http manipulation, it generally increased accuracy in the spoof condition for the more knowledgeable group. Participants' with high technical security knowledge are better able to take advantage of their familiarity $r_\tau(398) = 0.11, p < 0.05$, but this was not true for lo-knowledge participants, especially for the spoof websites. We hypothesize that hi-knowledge participants are more likely to detect spoof websites (i.e., wrong domain names) as their familiarity increases, whereas this detection process does not apply to lo-knowledge participants.

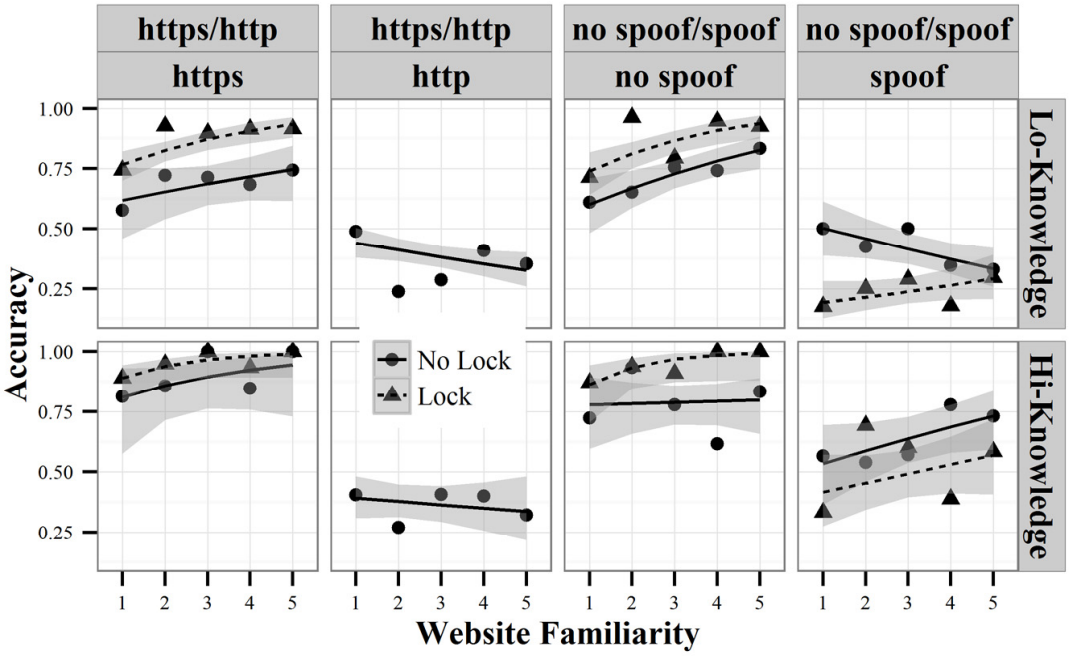

Figure 6: Relationship between participants' proportion of correct responses and their familiarity with the given website.

Participants' indicator scores interacted with no-spoof/spoof manipulation and encryption information $F(1,1352) = 6.91$, $p < 0.01$. In the no-spoof condition, knowledge of indicators improves participants' accuracy, but, in the spoof condition, rather than improving accuracy, attention to indicators, specifically when encryption information was present, reduces participants' accuracy (Figure 5).

4. Discussion & Conclusions

Although these results suggest that security knowledge is related to a decrease in risky behavior, it would be a gross exaggeration to suggest that security knowledge is sufficient to ensure secure and safe behavior on the web. Limiting our analysis to just those participants who scored correctly on at least 80% of the technical security questions (n = 32), we find that they scored correctly on 88% of the secure logins, but just 59% of the insecure logins across both studies.

These results clearly reveal that there is no simple relationship between security knowledge and the likelihood of logging into insecure websites. Although this result might have been predicted for non-experts, we expected that experts would show a lower likelihood of logging in to insecure websites. Given that this study was designed to increase both risk-taking and stress by motivating participants to respond as quickly as possible in order to maximize their pay-off, it is possible that either factor or both inflated the number of errors that were shown by experts. Familiarity of the websites may have also contributed to participants being less likely to check security indicators because they were more likely to revert to habitual behaviour of

logging in to familiar websites. In theory, this should have made all participants more vulnerable to the no spoof/spoof manipulation, but the performance of experts, in particular, was more complex than expected.

Experts are better than non-experts at detecting spoofed websites, but no better at detecting sites without encryption information. One possible explanation for this phenomenon is that experts primarily use the domain name highlighting feature available in modern browsers when identifying insecure websites, while non-experts do not. Assuming that the spoofed URLs are not particularly clever, then modest familiarity of an authentic URL—but not user interface—should expose a fraudulent website if one is aware of domain highlighting. This hypothesis would help explain why experts are good at identifying fraudulent websites, but no better than non-experts when it comes to logging into websites with no encryption. Experts' choices rely on more than familiarity. The presence of security indicators appears to diminish their accuracy when detecting spoofed websites. This suggests that experts find that security cues obscure the presence of an inauthentic URL, leading to a reduction in accuracy when dealing with spoofed websites with encryption information present.

These results clearly suggest that education alone will not be sufficient to change risky behaviors on the web. Just like in our study, a typical Internet user will often be asked to make security decisions against best-practice recommendations on security indicators. In essence, users are being educated through daily use to ignore recommended security indicators. These indicators are also used in an inconsistent fashion, where it is often necessary to have some familiarity with the website to know whether a partial or no encryption indicator is tantamount to commerce on an insecure site. Some of this confusion could be reduced if website designers conformed to the same set of conventions regarding security indicators. This would at the very least give users a better chance to identify insecure and spoof websites where their credentials and financial information can be hijacked.

5. Acknowledgements

Research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. The authors would also like to acknowledge the following people for their assistance: L. Jean Camp, Prashanth Rajivan, Rachel Huss, and Tom Denning.

6. References

Arianezhad, M., Camp, L.J., Kelley, T., and Stebila, D., 2013. Comparative eye tracking of experts and novices in web single sign-on. *In*: *Proceedings of the third ACM conference on*

Data and application security and privacy - CODASPY '13. New York, New York, USA: ACM Press, 105.

Buhrmester, M., Kwang, T., and Gosling, S.D., 2011. Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, 6 (1), 3–5.

Crump, M.J.C., McDonnell, J. V, and Gureckis, T.M., 2013. Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PloS one*, 8 (3), e57410.

Garg, V. and Camp, J., 2012. End User Perception of Online Risk under Uncertainty. *In*: *2012 45th Hawaii International Conference on System Sciences*. IEEE, 3278–3287.

Hazim, A., Felt, A.P., Reeder, R.W., and Consolvo, S., 2014. Your Reputation Precedes You: History, Reputation, and the Chrome Malware Warning. *In*: *Symposium on Usable Privacy and Security (SOUPS)*.

Kruschke, J.K., 2013. Bayesian estimation supersedes the t test. *Journal of experimental psychology. General*, 142 (2), 573–603.

Petzold, A., Plessow, F., Goschke, T., and Kirschbaum, C., 2010. Stress reduces use of negative feedback in a feedback-based learning task. *Behavioral neuroscience*, 124 (2), 248–255.

Schechter, S.E., Dhamija, R., Ozment, A., and Fischer, I., 2007. The emperor's new security indicators an evaluation of website authentication and the effect of role playing on usability studies. *In*: *Proceedings - IEEE Symposium on Security and Privacy*. Oakland/Berkley, CA, USA: IEEE, 51–65.

Stebila, D., 2010. Reinforcing bad behaviour. *In*: *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction - OZCHI '10*. New York, New York, USA: ACM Press, 248.

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

Tukey, J.W., 1949. Comparing individual means in the analysis of variance. *Biometrics*, 5 (2), 99–114.

Whalen, T. and Inkpen, K.M., 2005. Gathering evidence: use of visual security cues in web browsers. *In*: *Proceedings of Graphics Interface 2005*. 137–144.