Is Domain Highlighting Actually Helpful in Identifying Phishing Web Pages?

Aiping Xiong, **Robert W. Proctor**, **Weining Yang**, and **Ninghui Li**, Purdue University, West Lafayette, Indiana

Objective: To evaluate the effectiveness of domain highlighting in helping users identify whether Web pages are legitimate or spurious.

Background: As a component of the URL, a domain name can be overlooked. Consequently, browsers highlight the domain name to help users identify which Web site they are visiting. Nevertheless, few studies have assessed the effectiveness of domain highlighting, and the only formal study confounded highlighting with instructions to look at the address bar.

Method: We conducted two phishing detection experiments. Experiment 1 was run online: Participants judged the legitimacy of Web pages in two phases. In Phase 1, participants were to judge the legitimacy based on any information on the Web page, whereas in Phase 2, they were to focus on the address bar. Whether the domain was highlighted was also varied. Experiment 2 was conducted similarly but with participants in a laboratory setting, which allowed tracking of fixations.

Results: Participants differentiated the legitimate and fraudulent Web pages better than chance. There was some benefit of attending to the address bar, but domain highlighting did not provide effective protection against phishing attacks. Analysis of eye-gaze fixation measures was in agreement with the task performance, but heat-map results revealed that participants' visual attention was attracted by the highlighted domains.

Conclusion: Failure to detect many fraudulent Web pages even when the domain was highlighted implies that users lacked knowledge of Web page security cues or how to use those cues.

Application: Potential applications include development of phishing prevention training incorporating domain highlighting with other methods to help users identify phishing Web pages.

Keywords: cybersecurity, warnings, decision making, information processing, phishing

Address correspondence to Aiping Xiong, Psychological Sciences Department, Purdue University, 703 Third St., West Lafayette, IN 47906, USA; email: xionga@purdue.edu.

HUMAN FACTORS
Vol. 59, No. 4, June 2017, pp. 640–660
DOI: 10.1177/0018720816684064

INTRODUCTION

Phishing is a social engineering attack, meaning that it psychologically manipulates people into divulging confidential information in the context of cyber security (Orgill, Romney, Bailey, & Orgill, 2004). Phishing is usually implemented as a semantic attack, such as posting false information within emails, directly aiming at exploiting the vulnerabilities of human information processing of meaning during human-computer interaction (Schneier, 2000). Phishing attacks use electronic communication channels (e.g., email messages and social network Web pages) to communicate psychologically manipulated messages intended to persuade potential victims to perform certain actions for the attacker's benefit (Khonji, Iraqi, & Jones, 2013). As the Web site impersonates a reputable organization, victims are tricked into entering personal information and credentials (Parsons, McCormac, Pattinson, Butavicius, & Jerram, 2015), which are then stolen by the attackers. The phishing attacks can result in a series of undesirable consequences, including individual and organizational financial loss, identity theft, erosion of trust of the communication between individuals and companies, and reduction of users' willingness to engage in online transactions (Gopal, Tripathi, & Walter, 2006).

Much research has gone into mitigating phishing attacks. Automated tools for detecting phishing have been developed, including email classification at server and client levels (Cao, Han, & Le, 2008; Fette, Sadeh, & Tomasic, 2007), Web site blacklists (Whittaker, Ryner, & Nazif, 2010), and Web page visual similarity assessment (Fu, Liu, & Deng, 2006). Although several of these automatic solutions stop a large number of phishing attacks, those tools and services do not protect against all phishing attacks because (a) computers have difficulty accurately extracting the meaning of the natural language

messages contained in phishing emails (Stone, 2007) and (b) the attackers' tools continue to evolve (Downs, Holbrook, & Cranor, 2006).

Because the final decision on a Web page's legitimacy is made by the user (Proctor & Chen, 2015), decision-aid tools have been developed to assist people at detecting fraudulent sites when automatic detection fails. Examples are dynamic security skins (Dhamija & Tygar, 2005), browser toolbars such as SpoofStick and Trustbar (Herzberg & Gbara, 2004), and Web browser phishing warnings and Secure Sockets Layer (SSL) warnings (Carpenter, Zhu, & Kolimi, 2014; Felt et al., 2015). However, limited success, ineffectiveness, and usability problems have been found across those decision-aid tools (Sheng et al., 2009; Wu, Miller, & Garfinkel, 2006). To develop effective tools to aid users in combating phishing attacks, studies have been conducted to understand how and why people fall for the attacks (e.g., Downs, Holbrook, & Cranor, 2007; Welk et al., 2015). In general, the results have shown that the user's attention is dominated by deceptive visual cues that reinforce the legitimacy of the Web page. But users rarely pay attention to the security indicators. For the participants who noticed the warnings, poor comprehension and confusion with other security threats were also evident in prior studies (e.g., Bravo-Lillo, Cranor, Downs, & Komanduri, 2011; Dhamija, Tygar, & Hearst, 2006; Sunshine, Egelman, Almuhimedi, Atri, & Cranor, 2009).

Domain Highlighting

A well-crafted phishing site looks visually identical to the impersonated Web site; thus, the phishing Web page intentionally deceives users to fill in their personal information, which they might not do otherwise. Grazioli (2004) examined the responses of tech-savvy Internet consumers to a real site or a deceptive copycat site. Despite a few persons successfully detecting the deception, most participants were unable to do so. Based on the information-processing model of deception detection (Johnson, Grazioli, Jamal, & Berryman, 2001), Grazioli (2004) compared the information-processing behavior between successful and failure deception detection. He found that participants who were successful at detecting the deceptive sites focused on fewer deception cues than participants who were unsuccessful, and the successful participants relied on "assurance" cues (e.g., product warranties) and heavily discounted "trust" cues (e.g., customer's testimonials).

Despite the visual deception of a phishing Web page, the domain name within the uniform resource locator (URL) will always be different from the legitimate one. Therefore, the domain name can serve as a sole telltale sign to detect a phishing Web page (Lin, Greenberg, Trotter, Ma, & Aycock, 2011). Consequently, the mismatch between the real domain name and the impersonated Web page can be expected to serve as a critical measure to detect phishing attacks.

Because the domain name can be easily overlooked by users (Jagatic, Johnson, Jakobsson, & Menczer, 2007), domain highlighting (DH) has been developed to exploit this telltale sign. The owning domain of whatever site a user is currently viewing is highlighted. Specifically, the domain name portion within the URL in the browser's address bar is in black, whereas the rest of the URL is rendered in gray. This method has been implemented by the most popular browsers, including Google Chrome, Microsoft Edge/Internet Explorer, and Mozilla Firefox. Figure 1 illustrates how DH is implemented in the default interface of recent releases of each browser, Firefox on top, Chrome in middle, and Internet Explorer at bottom. The highlighting methods also vary among the browsers; for example, Chrome highlights all content between "//" and the first "/" within the URL, whereas Firefox only highlights the domain name, and Internet Explorer highlights both "https" and the domain name.

The effectiveness of DH is based on three assumptions about users' information processing: (a) Users will naturally attend to the address bar, (b) users will use the domain name to decide the Web site's legitimacy, and (c) users can recognize the domain names. Few studies have evaluated the effectiveness of DH, and the only formal study of which we are aware is that of Lin et al. (2011). They recruited 22 university students to evaluate the legitimacy of 16 Web pages (half legitimate, half fraudulent) of typical phishing targets in two separate rounds. In the first round, the participants evaluated the Web

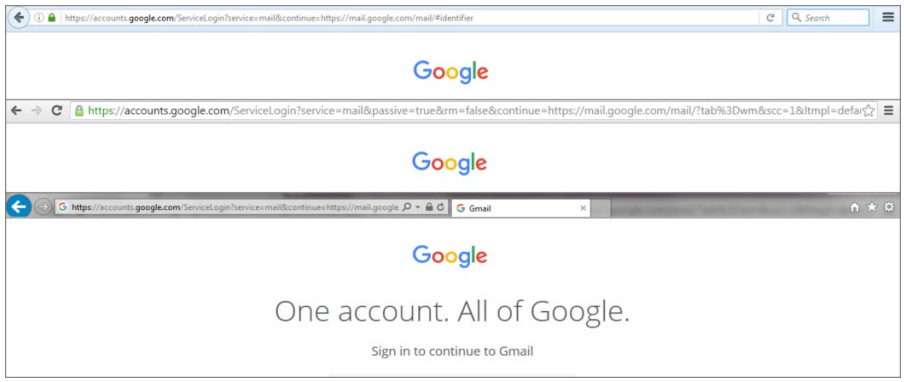

Figure 1. Domain name highlighting in Mozilla Firefox (top), Google Chrome (middle), and Internet Explorer (bottom).

page based on whatever means they chose, whereas in the second round they were instructed to make the decision by looking specifically at the address bar. More fraudulent Web pages were incorrectly rated as safe in the first phase than the second phase, indicating that participants do not always attend to the address bar to determine the site legitimacy. Although Lin et al. argued that DH provides some benefit, the marginal effect they obtained could be a consequence of just directing users to look at the address bar since a control condition without DH was not tested. Also, Lin et al. performed a qualitative analysis of interview results, which showed that most of the participants who attended to the address bar often did not notice the spoofed domain name, indicating a lack of user knowledge.

Eye Tracking

Data about people's visual attention can be collected through eye tracking, the process of monitoring people's eye movements such as the point of gaze and fixation duration (Rayner, 2009). The technology is a straightforward way to observe where users' attention is directed, and it has been deployed in experimental psychology to give insights into the bases of users' behavior (Duchowski, 2007). Nowadays, many eye-tracking techniques are available; one popular form uses video captured by optical devices (i.e., a camera) recording infrared light that illuminates the eyes and provides reference points for the eye tracker. This type of device is mounted remotely from the user and does not require the user to wear any special device, which induces minimal interference with the visual stimulus.

Previously, eye-tracking studies were conducted to identify whether users look at security cues during Web browsing. Whalen and Inkpen (2005) collected eye-tracking data to analyze how users interact with security indicators. No participants gazed at the security indicators in the normal browsing phase; most participants looked at the lock icon when they were "primed" for security, but few made use of its interactive capabilities. Arianezhad, Camp, Kelley, and Stebila (2013) also provided evidence that users do not look at security indicators when they are browsing naturally. When participants were explicitly motivated to pay attention to security information, they still failed to detect phishing Web pages (Neupane, Rahman, Saxena, & Hirshfield, 2015). In another study, an eye-tracking method was designed and implemented with a browser extension to force users to get into the habit of checking the address bar before any inputs in the Web sites (Miyamoto, Iimura, Blanc, Tazaki, & Kadobayashi, 2014). The study results confirmed the effectiveness of forcing participants to check the address bar.

The benefit of DH obtained in Lin et al.'s (2011) study is challenged by the results of eye-tracking experiments, which showed that users' performance in phishing tasks was positively impacted by attention control (Miyamoto et al., 2014; Neupane et al., 2015). Because Lin et al.'s

study had only a small number of participants from a university community and yielded ambiguous findings, we conducted a study (Experiment 1) that followed a similar procedure but with a larger, more representative participant sample and a non-highlighted control condition. The study was conducted online with more than 300 participants, and DH versus domain non-highlighting (DNH) was manipulated to dissociate the effect of directing a user's attention to the address bar from the effect of DH. We also employed an eye tracker in Experiment 2 to obtain eye-movement observations to further test the three assumptions of DH.

GENERAL EXPERIMENTAL DESIGN

The designs of our online study and eye-tracking experiment are in line with the ones implemented previously (Lin et al., 2011). In both experiments, we used the two-phase paradigm. In Phase 1, participants were asked to classify how "safe" each page appeared to be. In Phase 2, participants were instructed to direct their attention to the address bar, which allowed us to evaluate how people used their own knowledge while removing the problems associated with attention. Beyond task performance, the eye-tracking experiment also collected details about how people attend to the address bar during their safety decisions on Web pages for each phase. We also gathered subjective data at the end of the eye-tracking experiment with questions related to the user's awareness and comprehension of DH.

Two sets of 6 different Web pages each, representing 12 different Web sites (see Table 1), were selected and used as stimuli. The Web sites included the most targeted phishing industries, including: e-commerce, banking, social media, and email server (Aaron, Rasmussen, & Routt, 2014). Another two categories were online file-host system and miscellaneous. The Web sites' domain traffic rankings from Alexa.com were approximately equal across the two sets within our geographic region to make the two sets to be encountered as equally as possible during routine Web browsing. Each legitimate Web page was a perfect replica of the original Web site, including the domain name. Fraudulent versions of the Web pages were developed by altering their URLs based on real phishing Web sites listed in PhishTank. The modifications represent similar and complex spoof methods (Lin et al., 2011). For the similar method, fraudulent URLs are visually similar to the legitimate URLs, such as twiller.org looks similar to twitter.com. With the complex method, fraudulent URLs expand the length of the legitimate ones, which makes interpretation of the URL difficult. Table 1 lists the specific Web sites, their rankings, the spoof method, and the modified URLs. Participants were shown snapshots of the Web sites' login Web pages taken from the Firefox browser by excluding all browser chrome except address bar. Each Web page had a version in which the domain was highlighted and another version in which it was not. These were created by enabling or disabling, respectively, the *browser.urlbar.formatting* function under *about:config*.

Participants evaluated the legitimacy of different Web pages (half fraudulent and half legitimate) in two phases for both the online and eye-tracking experiments. Each phase started with the instruction page. For the first phase, the instruction was "Please evaluate the trustworthiness of the Web pages based on any information you can find within the Web page." In the second phase, the instruction was "Please evaluate the trustworthiness of the Web pages by focusing on the address bar," and the address bar area was marked in red square. Trial pages were presented after the instructions. For the online study, the Web page review and safety judgment response were presented in the same page. For the eye-tracking experiment, the safety-judgment response page was separated from the Web page view to obtain a ratio of total fixation time over the whole Web page view time. As in Lin et al.'s (2011) study, the "any information" phase was always conducted prior to the "address-bar focus" phase so that participants would not be predisposed to attend to the address bar by the prior instructions.

This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at Purdue University. Informed consent was obtained from each participant. The experiment data that were stored and analyzed are anonymized.

TABLE 1: Web Page Details of Each Set

Set	Web Site Category	Web Site	Ranking	Spoof Method	Legitimate Web Page URL	Phishing Web Page URL
1	Bank	Chase	22	Complex	https://www.chase.com	http://secureupdate.chaseonline.chase.security.check.bemonline.net/Logon/chase
	Social media	Twitter	8	Similar	https://twitter.com/login/	http://twiller.org/
	Online file host system	Apple	30	Complex	https://secure2.store.apple.com/shop/sign_in?c=……	http://support.apple.com.en-gb.confirm.id.auth.cgi-key.myapple-unlock.user-eu9.url-
	Miscellaneous	Craigslist	9	Similar	https://accounts.craigslist.org/login	http://craigslistersr.us/<anything_here>
	Email	Hotmail	14	Complex	https://login.live.com/login.srf?wa=wsignin……	http://mercymedical.org/wp-content/uploads/hotmail/Sign_in.html
	E-commerce	eBay	7	Complex	https://signin.ebay.com/ws/eBayISAPI.dll?SignIn&UsingSSL=1&pUserID=&co_partnerID=2&siteid=0&ru=……	http://weissoptic.pl/scg/eBa1y-notice/eba1y-accountnot8ce49.html?_ipg=50&_sopbarabule=……
2	Bank	BOA	28	Complex	https://www.bankofamerica.com/sitemap/hub/signin.go	http://www.secure.bankofamerica.com.oho4ueq8b82e6mfdryktxhloawxuphq5p6x.00d8abyf8albs0lcxsnbsazdq.mlauchiroinc.com/sig/login.php?pin=……
	Social media	Facebook	2	Complex	https://www.facebook.com	http://cutt.us/1144QQNN?Secure-public-in-fb
	Online file host system	Dropbox	53	Complex	https://www.dropbox.com	http://makino-sakana.com/dropbox/
	Miscellaneous	LinkedIn	11	Similar	https://www.linkedin.com/uas/login	http://www.linkedin.com.vsbst.com/852d6e7798dab64c468ed354f71b6603/login
	Email	Gmail	43	Complex	https://accounts.google.com/ServiceLogin?service=mail……	http://127maintenance-service.com/q/
	E-commerce	Amazon	3	Complex	https://www.amazon.com/ap/signin?_encoding=……	http://amazstateapp.com/help/custom/login/us/Amazon/ca/tn/529dffc790ab5c0964185bad3c1f54e7/

EXPERIMENT 1: ONLINE STUDY

The online study was designed to dissociate the effect of directing a user's attention to the address bar from that of DH. This was accomplished by having a general user population evaluate Web pages' safety with DH and DNH conditions.

Method

Participants. Three hundred and twenty adults (58% female) were recruited through Amazon Mechanic Turk (MTurk). Their reported ages ranged from 18 to over 50 years, with 76% being between ages 23 and 50. About 38% of the participants reported that they had college credits but no degree, and about 50% had a bachelor's degree or higher. Each participant was compensated with $0.10 cash. Each participant was tested on his or her computing device. Participants were not instructed about phishing attacks or DH. Instead, they were initially told that they were participating in a study on the visual security aids that are typically found in browsers. We limited data collection to participants from the United States because the Web sites we selected are popular within this region. Participants were told to participate in the study only once. However, 26 participants finished the study twice; for those participants, we analyzed only the data from the first time.

Apparatus and procedure. Each participant was asked to evaluate the legitimacy of 6 different Web pages (half fraudulent and half legitimate) in each of two phases. For the first phase, the participants rated the safety of each Web page by clicking buttons on a 5-point scale from 1 (*unsafe*) to 5 (*safe*) based on any information. For the second phase, participants were told to adopt an address-bar focus approach without indication of specifically what to look for in the area, and they then evaluated 6 different Web pages. Order was not counterbalanced because being told to focus on the address bar in the first phase would create a bias to do so with the neutral instructions in the second phase. Participants' demographic information was collected after both phases had been completed. On average, the whole experiment took about 6 minutes.

The Web pages were presented in a randomized order across the participants. Domain highlighted or not was manipulated within each phase. Similar numbers of participants were assigned to four task variants determined by Web pages' legitimacy and DH/DNH. Which subset of sites was assigned to have valid or fraudulent URLs and whether the domains were highlighted in the first or second phase were counterbalanced between subjects.

Results

Following the similar analysis of Lin et al. (2011), for each condition, the ratings of the three relevant Web pages were averaged for each participant, yielding a mean rating. The rating data were then converted into *unsafe* (1 or 2), *unsure* (3), or *safe* (4 or 5) categories based on the rounded value; specifically, the value was rounded as the smaller integer if the decimal digit was smaller than 5; otherwise, it was rounded as the larger integer. The safety judgment results are listed in Table 2.

In Phase 1 with any information instruction, regardless of DH, on average participants rated legitimate Web pages 2.2% incorrectly as unsafe, 15.0% as unsure, and 82.8% correctly as safe. For fraudulent Web pages, participants rated 25.7% correctly as unsafe, 25.3% as unsure, and 49.1% incorrectly as safe. Thus, to a limited extent, legitimate Web pages were discriminated from fraudulent ones. From the table, we can also observe that DH had little effect, with the highlighted and non-highlighted URLs showing similar results for both the legitimate and fraudulent Web pages.

In Phase 2 with address-bar focus instruction, regardless of DH, participants rated 3.5% of the legitimate Web pages as unsafe, 16.3% as unsure, and 80.3% correctly as safe. Thus, the legitimate Web pages were rated as safe when instructed to focus on the address bar compared to when not instructed to do so. The accuracy of the ratings of the fraudulent Web pages was higher in the second phase, with 41.6% rated correctly as unsafe, 26.8% as unsure, and 31.6% as safe. Fraudulent Web pages were correctly judged to be unsafe more often in Phase 2 than in Phase 1, but there was still not much difference between DH and DNH.

Safety decision analysis. Chi-square tests were conducted on the resulting frequencies of

TABLE 2: Percentages of Web Page Safety Decisions, Signal Detection Analysis Measures (*d'* and *c*) Estimates of Each Condition in Experiments 1 and 2

		First Phase (Any Information)				Second Phase (Address-Bar Focus)			
		Domain Highlighting		Domain Non-Highlighting		Domain Highlighting		Domain Non-Highlighting	
Experiment	Measure	Legitimate	Fraudulent	Legitimate	Fraudulent	Legitimate	Fraudulent	Legitimate	Fraudulent
1	Site is unsafe, %	3.1	26.1	1.3	25.2	3.8	46.2	3.1	37.0
	Unsure, %	16.1	26.1	13.8	24.5	17.7	23.4	14.8	30.3
	Site is safe, %	80.7	47.8	84.9	50.3	78.5	30.4	82.1	32.7
	d'	0.30		0.13		0.37		0.36	
	c	0.53		0.60		0.30		0.38	
2	Site is unsafe, %	0.0	18.7	0.0	12.5	0.0	68.8	6.2	50.0
	Unsure, %	25.0	37.5	0.0	18.8	25.0	18.8	6.3	31.2
	Site is safe, %	75.0	43.8	100.0	68.7	75.0	12.4	87.5	18.8
	d'	0.62		0.75		1.69		1.23	
	c	0.62		0.98		0.40		0.50	

safe and unsafe decisions. Analysis confirmed that the legitimate Web pages were classified as safe more often than the fraudulent Web pages, $\chi^2(1) = 250.22$, $p < .001$. There was also a difference between the two instructions, $\chi^2(1) = 17.53$, $p < .001$, and it interacted with Web page's legitimacy decision, $\chi^2(1) = 6.02$, $p = .014$. Further analysis of the two Web page types confirmed that the difference was mainly due to the fraudulent Web pages being classified more accurately with address-bar focus instruction in the second phase, $\chi^2(1) = 23.30$, $p < .001$, whereas the classification of the authentic Web pages did not differ across the two phases, $\chi^2(1) < 1.0$. DH or DNH did not show any effect, $\chi^2(1) = 1.95$, $p = .162$, or interact with Web page's legitimacy decision, $\chi^2(1) < 1.0$.

Signal detection analysis. As the proportion of successful phishing detection ignores the influence from legitimate trials, signal detection theory methods have also been used to assess users' sensitivity (d') and response bias (c) to phishing (e.g., Canfield, Fischhoff, & Davis, 2016). Because each participant in our study received only 3 Web pages for each cell of the experimental design, many of the hit and false alarm rates are 0.0 or 1.0, which makes it inappropriate to calculate and analyze d' and c for each participant. Macmillan and Kaplan (1985) noted that in such cases, "Computing a collapsed d'/c from averaged data can be a reliable, relatively unbiased way to estimate true average d'/c" (p. 196). Consequently, we pooled all participants' data and analyzed them using the rating method specified by Macmillan and Creelman (2004). With this method, empirical receiver operating characteristics (ROCs) with N–1 pairs of hit and false alarm rates (N = number or rating categories; 5 in our case) are generated, from which d'/c is estimated.

We calculated hits (unsafe rating of phishing Web pages) and false alarms (unsafe classification of legitimate Web pages) as follows: First, fraudulent and legitimate Web pages' ratings were ordered from 1 to 5 for the four conditions comprised of two instructions × DH/DNH, with 1 indicating highest certainty of phishing Web page and 5 indicating lowest certainty. Then, for each condition, the total number of responses of each rating was calculated; for each rating, the proportion of trials was computed by summing all responses of that rating and lower. Next, a d' measure and c measure were obtained for each pair of hit and false alarm rates. Finally, for each condition, the four individual d' and c values were averaged to obtain an overall d' and c (see Table 2).

The overall d' values were < 0.5, indicating that participants had limited ability to detect phishing Web pages. But the detectability improvement from Phase 1 (P1) to Phase 2 (P2) was numerically larger ($d'_{P1} = 0.21$, $d'_{P2} = 0.36$) than the change between DNH and DH ($d'_{DNH} = 0.25$, $d'_{DH} = 0.32$), consistent with the percentage response results. Positive c values indicate that participants were biased to identify Web pages as safe. This bias decreased when participants were instructed to focus on the address bar ($c_{P1} = 0.57$, $c_{P2} = 0.34$) but was similar for DNH and DH conditions ($c_{DNH} = 0.42$, $c_{DH} = 0.49$).

Discussion

The results showed a relatively high proportion correct for identifying legitimate Web pages across the two phases, and this aspect of performance was unchanged between DH and DNH. The lack of effect of phase or DH on correct identification of the legitimate pages as safe could be due to a ceiling in ratings being reached already in the first phase. In contrast, the proportion of correctly identified phishing pages was low. With the any information instruction in the first phase, only about 26% of the fraudulent Web pages were identified as unsafe, whereas about 50% were incorrectly rated as safe. When participants were directed to focus on the address bar, identification of phishing Web pages as unsafe improved to nearly 43%, but the effect of DH was not significant. Signal detection analysis of the group data indicated low ability to discriminate the fraudulent Web pages from the legitimate ones, with a bias to classify most Web pages as safe.

In general, our results are similar to those of Lin et al. (2011) in showing a small benefit of directing participants to focus on the address bar. However, the results showed that this benefit was due almost entirely to the address-bar focus instruction and not to DH. And the results were further verified by the signal detection analysis. The take-home messages of Experiment 1 are: (a)

By including the DNH control condition, our results provide evidence that the small improvement at identifying fraudulent Web sites in Lin et al.'s second phase was mainly due to directing participants' attention to the address bar rather than to DH. (b) The increase in classifying phishing sites as unsafe when participants were instructed to focus on the address bar suggests that they do not always attend to the address bar when considering the legitimacy of a Web site. (c) The fact that less than 50% of fraudulent Web pages were accurately judged as such across the two phases demonstrates that drawing people's attention to the address bar is not sufficient to get accurate decisions.

EXPERIMENT 2: EYE-TRACKING STUDY

The results of Experiment 1 imply that users do not know how to use the domain name information even when they look at the address bar and URL. To further understand how users allocate attention during Web page browsing, we conducted an eye-tracking experiment in a laboratory setting to verify whether users noticed DH and whether they were able to use it during Web page safety judgment.

Method

Participants. Thirty-two undergraduate students (10 female) with average age of 19 years enrolled in introductory psychology courses in Purdue University took part for credits toward a course requirement. Participants were told they were engaging in a study on the visual security aids that are typically found in browsers. Four participants had computer science–related majors, five had engineering-related majors, and the other participants were majoring in psychology, marketing, chemistry, others, or undecided.

Apparatus and procedure. All stimuli were presented using a personal computer and displayed on a 22-inch LCD with a 1,920 × 1,200 pixel resolution and a refresh rate of 60 Hz. To measure participants' gaze pattern, we used an EyeLink 1000Plus (SR Research, Ltd., Ontario, Canada) remote desktop mounted eye-tracker system, averaging 0.25° to 0.5° accuracy. Participants sat 80 cm away from the monitor screen using a head support tower mounted to the table. Each participant's head position was fixed on a padded chinrest and a forehead rest to keep the participant's eyes within the range of the eye tracker during the whole experiment. The chinrest height was fixed for each participant, but participants could be comfortably seated by changing their sitting height with the chair lever on the right-hand side. The eye-tracker was placed on the same table in a fixed position between the participant and the monitor, about 50 cm from the participants. Although viewing was binocular, the recording was monocular (a standard procedure in eye-tracking studies).

The procedure of Experiment 2 was similar to that of Experiment 1, except as noted. After signing the consent form, each participant was seated and asked to assess his or her familiarity with the Web sites in a paper list. The experimenter briefly explained any Web site of which the participant was unaware. Then, the participant was asked to place his or her forehead and chin on the rests. The eye-tracking calibration and validation were conducted before the start of the experiment for each participant using a 13-point calibration grid. The accuracy level of the calibration was only accepted when the errors in degrees of visual angle were smaller than 0.5° on average. The participant was required to maintain the same position during the whole experiment.

Two sets of 6 different Web pages were presented in random order in each phase, with one set authentic and the other fraudulent. The two sets' legitimacy was counterbalanced between participants. Whether the domain was highlighted or not was varied between subjects as well. The experiment started with loading the instructions page. The experimenter read aloud the instructions displayed on the monitor and made sure the participant was clear about the tasks that were to be performed in each phase. Trial pages were presented after the participant pressed the space bar, on which the left-hand fingers rested. Each trial started with a rest page for 2 seconds, in which a cross sign was shown at the center of a white blank page. This was followed by a Web page, which was fake or authentic. A response page was presented when the participant finished viewing the Web page by pressing the space bar. The participant then

clicked one of the buttons from 1 (*unsafe*) to 5 (*safe*) using a computer mouse's left button with their right hand. As in Experiment 1, no feedback was given for each Web page's decision response.

Upon completion of the two task phases, each participant answered an exit survey, in which the experimenter asked five questions and wrote down the participant's responses. The first question asked the participant to describe the methods that he or she used to evaluate the safety of the Web page in each phase. After that question was answered, the experimenter described DH to the participant using the Chrome browser and asked whether the participant understood the purpose of this feature and whether he or she had noticed this feature previously. Following the answer, the experimenter explained how DH could help people evaluate a Web page's safety and asked the participant to rate the effectiveness of DH on a 5-point scale (1 = *not effective*, 5 = *very effective*), after which the participant was asked to explain why he or she gave that particular rating. On average, it took about 20 minutes to complete the entire session.

Results and Discussion

Each participant's safety judgment responses and gaze patterns were recorded and extracted from the fixation reports generated by the EyeLink Data Viewer program. First, the safety judgments and signal detection measures were analyzed in the same way as Experiment 1. Second, the collected gaze data were used to compute the number of fixations, mean fixation duration, and visits in the address bar area, referred to as Areas of Interest (AOI). Fixation is defined as a pause made by a participant looking at the AOI to extract meaningful information. The number of fixations identifies the hits on the AOI during the Web page view period. Mean fixation duration was calculated by averaging the durations of all fixations hitting AOI during Web page view period. Visits define the number of glances to the AOI within Web page view. Average total viewing time for each Web page was also calculated for further analysis with regard to the influence of the address-bar focus instructions. Third, a heat map analysis was conducted to explore participants' visual attention allocation across the AOI and the whole Web page. The exit survey data were analyzed in the end.

Safety decision analysis. We compared the safety decisions for the two phases (see Table 2). In Phase 1, regardless of DH, participants rated 12.5% of the legitimate Web pages as unsure and 87.5% correctly as safe. For fraudulent Web pages, participants rated 15.6% correctly as unsafe, 28.1% as unsure, and 56.3% incorrectly as safe. Thus, as in Experiment 1, legitimate Web pages were discriminated from fraudulent ones to some extent. As before, safety decisions between DH and DNH showed similar results for both authentic and fraudulent Web pages.

In Phase 2, the overall decisions for the legitimate Web pages were similar, with 3.1% rated incorrectly as unsafe, 15.6% as unsure, and 81.3% correctly as safe. Thus, the legitimate Web pages were rated as safe to a slightly less extent when participants were instructed to focus on the address bar than when they were not instructed to do so. With address-bar focus instructions, 59.4% fraudulent Web pages were rated correctly as unsafe, 25.0% as unsure, and 15.6% incorrectly as safe. Again, the results were likewise between DH and DNH for both legitimate and fraudulent Web pages (see Table 2).

Chi-square tests were conducted as in Experiment 1 on the frequencies of safe and unsafe decisions. Analysis confirmed that the legitimate Web pages were classified as safe more often than were the fraudulent Web pages, $\chi^2(1) = 30.609$, $p < .001$. There was also a difference between the two phases, $\chi^2(1) = 10.385$, $p = .001$, and the interaction between phase and Web page legitimacy was at the .05 significance level, $\chi^2(1) = 3.833$, $p = .05$. Analysis of each Web page type confirmed that the difference was mainly due to the fraudulent Web pages being classified more accurately in the second phase, $\chi^2(1) = 13.287$, $p < .001$; classification of the authentic Web pages did not differ across the two phases, $\chi^2(1) < 1$. By adding whether the domain was highlighted or not as another factor, no differences were obtained between each condition, $\chi^2(1) < 1.0$.

Signal detection analysis. We estimated *d'* and *c* values in the same way as in Experiment 1. The overall detectability of phishing Web pages indicated modest detection ability (see Table 2). Same as Experiment 1, the detectability of phishing Web pages increased from Phase 1 to

TABLE 3: Gaze Pattern Results for Safety Decisions: Mean Total Fixation Duration (milliseconds), Number of Fixations, Number of Visits, Total Web Page View Time (milliseconds), and Total Fixation Time/Total Web Page View Time in Experiment 2

	First Phase (Any Information)				Second Phase (Address-Bar Focus)			
	Domain Highlighting		Domain Non-Highlighting		Domain Highlighting		Domain Non-Highlighting	
Measure	Legitimate	Fraudulent	Legitimate	Fraudulent	Legitimate	Fraudulent	Legitimate	Fraudulent
Total fixation duration	383	435	707	1,008	1,212	1,495	949	1,412
Number of fixations	1.5	1.5	2.6	3.0	4.2	4.4	3.2	4.2
Number of visits	0.9	0.9	1.0	1.2	1.3	1.1	1.1	1.3
Total Web page view time	8,916	9,796	6,720	7,821	5,148	5,978	4,010	4,711
Total fixation time/total Web page view time, %	5.5	4.8	7.3	10.2	26.0	24.6	26.1	29.7

Phase 2 ($d'_{P1} = 0.67$, $d'_{P2} = 1.49$); the difference between DNH and DH was much less ($d'_{DNH} = 0.97$, $d'_{DH} = 1.14$). As in Experiment 1, participants showed an overall bias to classify Web pages as safe, and this bias was less in Phase 2 than in Phase 1 ($c_{P1} = 0.80$, $c_{P2} = 0.45$). But, the bias was numerically higher with DNH than DH ($c_{DNH} = 0.74$, $c_{DH} = 0.51$).

The safety decision results and signal detection analysis measures were similar to those obtained in the online study, verifying the benefit of instructing participants to focus on the address bar. The results provide further evidence that the improvement at identifying fraudulent Web sites in the second phase was mainly due to directing participants' attention to the address bar rather than to DH. Besides, the benefit of directing users' attention to the address bar was larger in the current experiment than in the online study. Although participants spent 20 minutes in the eye-tracking experiment, the Web pages' view time was in a similar range as the online study. Therefore, we suspect that the better performance and increased sensitivity are mainly due to more engagement in the task of participants within a controlled environment setting as bias in the two experiments was similar.

Eye-movement data analysis. We wanted to understand whether participants naturally looked at the address bar area and how much time they spent on the area. We were also interested in knowing whether there were any differences when instructions were varied and whether the instructions interacted with DH. Gaze pattern measurements results are listed in Table 3. We conducted a mixed analysis of variance (ANOVA) with legitimacy (authentic vs. fraudulent) and instruction (any information vs. address-bar focus) as within-subject factors and domain highlighting (DH vs. DNH) as a between-subjects factor for each measurement.

Total viewing time on each Web page. Total viewing time was calculated by subtracting the Web page presentation time from the rating page presentation time for each participant. ANOVA results showed main effects of instruction, $F(1, 30) = 25.872$, $p < .001$, $\eta_p^2 = .463$, and legitimacy, $F(1, 30) = 16.041$, $p < .001$, $\eta_p^2 = .348$. Longer time was spent for the any information phase than the address-bar focus phase, and the viewing time for fraudulent pages was longer than legitimate pages.

The percentage of the total fixation time relative to the whole Web page view time was also analyzed. The results showed a main effect of instruction, $F(1, 30) = 29.080$, $p < .001$, $\eta_p^2 = .492$: The percentage was higher for "address-bar focus" instructions than the "any information" instructions. Although there was no main effect of legitimacy, $F < 1$, the two-way interaction

between legitimacy and DH was significant, $F(1, 30) = 4.874$, $p = .035$, $\eta_p^2 = .140$. The total fixation time ratio was larger for legitimate Web pages when the domain was highlighted, whereas the ratio was larger for fraudulent Web pages when the domain was not highlighted.

Total fixation time, number of fixations, and number of visits. From Table 3, one can generally see a lower average number of fixations and shorter total fixation time in the any information phase. The number of fixations and the averaged fixation duration were highest when domain names were highlighted in the address-bar focus phase. These patterns seem not to differ between legitimate and fraudulent Web pages.

ANOVA results of total fixation time showed main effects of instruction, $F(1, 30) = 6.895$, $p = .013$, $\eta_p^2 = .187$, and legitimacy, $F(1, 30) = 10.074$, $p = .003$, $\eta_p^2 = .251$. No terms involving DH were significant or approached significance, Fs < 2.057. Number of fixations showed a similar pattern as total fixation time. The main effect of instruction, $F(1, 30) = 6.557$, $p = .016$, $\eta_p^2 = .179$, was significant, but the main effect of legitimacy, $F(1, 30) = 3.297$, $p = .079$, $\eta_p^2 = .099$, only approached significance. When the domain name was highlighted, the number of visits of the authentic Web pages increased, whereas the number of visits of the fraudulent Web pages decreased. ANOVA results showed that the two-way interaction of Legitimacy × DH was at the .05 level, $F(1, 30) = 4.036$, $p = .053$, $\eta_p^2 = .119$.

We also conducted a correlation analysis between the total fixation time and the ratings. For authentic Web pages, there was no significant correlation, $p = .707$, but the ratings of fraudulent Web pages were negatively correlated with the total fixation time, $r_{cor} = -.220$, $p < .001$. This negative correlation was significant for the any information phase, $r_{cor} = -.251$, $p < .001$, but not the address-bar focus phase. Further, this negative correlation was significant for the DNH condition, $r_{cor} = -.327$, $p < .001$, but not the DH condition.

Based on the eye-movement analysis, we further confirmed that participants can differentiate legitimate and fraudulent Web pages. In general, participants fixated less frequently and for less time on the legitimate Web pages than on the fraudulent ones. When directed to look at the address bar area, participants dwelled more, made more fixations, and visited more often. Therefore, the effectiveness of attending to the address bar area was further verified. However, like the safety-judgment results, DH did not show any significant impact on the eye movements' results either, indicating that participants did not attend to the highlighted domain at all or attended to it but did not know what it indicated, which was also evident in the correlation analysis.

Visual attention on Web pages. To understand whether participants attended to the highlighted domain or not, we did a further analysis based on heat maps. Heat maps can illustrate visual attention by manifesting the fixation locations and fixation duration across regions within the stimuli (Duchowski, 2007; Rayner, 2009). The fixation regions are highlighted by shades in the green-red color spectrum, with green indicating the shortest fixation durations and red signifying the longest durations. The heat maps were generated based on fixations within a Web page view period by using a Gaussian function, and the outer 10% of the fixation distributions were trimmed from the raw data (SR Research, 2009). Thus, shaded areas in the heat maps received 90% of total fixations during Web page view.

We listed the results of Chase Bank (Figures 2 and 3) and Hotmail (Figures 4 and 5) as two examples to illustrate the details of the pattern obtained in the heat maps. A close look at the heat maps in each figure suggests that participants' visual attention was primarily clustered at the region of the Web page content in the any information phase (first row of Figures 2–5). In particular, participants tended to spend more time looking at the sign-in and logos (shown as red shades), consistent with previous studies (e.g., Arianezhad et al., 2013). In addition, it is worth noting that participants were attending to the address bar area but to a much less extent in the first phase, regardless of whether the domain name was highlighted or not. Furthermore, comparison between the first rows of legitimate (Figures 2 and 4) and fraudulent (Figures 3 and 5) Web pages showed that participants seemed to pay more attention to the address bar when viewing fraudulent Web pages. When directed to focus on the address bar, participants spread their attention mainly across the whole URL in the address bar to a larger extent than in the first phase (see Figures 2–5 bottom row). Most important, participants mainly paid attention

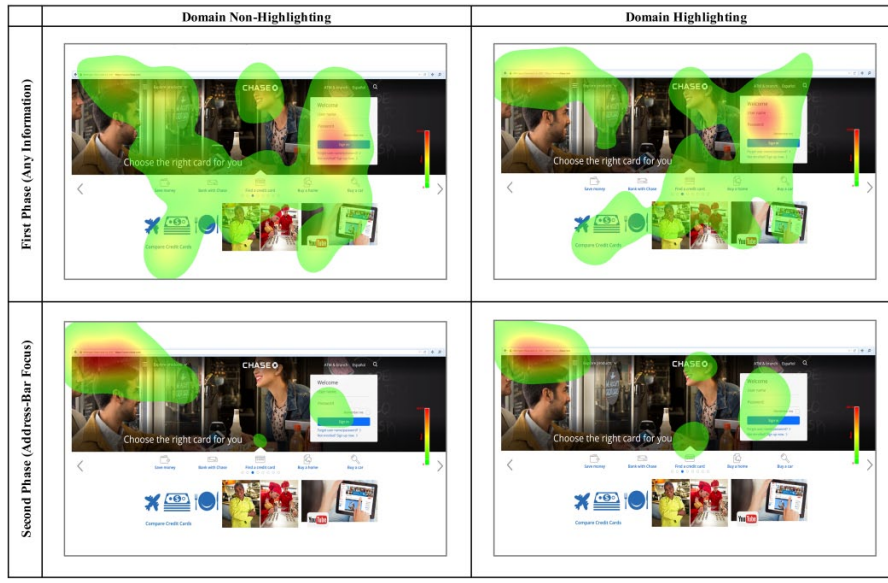

Figure 2. Legitimate Web page heat map, Chase.

Figure 3. Fraudulent Web page heat map, Chase.

to a small region at the very left part of the address bar when they looked at DNH fraudulent Web pages (see Figures 3 and 5 left panel), indicating that they usually relied on security lock or the beginning of a URL for the safety decision. When the domain name was highlighted, more attention was paid to the domain (see Figures 3 and 5 right panel). Thus, participants seemed aware of the phishing, but the high proportion of incorrect safety decisions of fraudulent Web pages suggests participants lacked knowledge of reliable cues to identify Web page's legitimacy.

One heat map for each of the 12 Web pages in each condition was also created for reference, and these are presented in Figures 6 to 13 in the Appendix. Within each figure, the heat maps are listed according to the order in Table 1, and Web pages from each set are paired, starting with Web pages from Set 1. Figures 6 to 9 are results of each condition for legitimate Web pages and Figures 10 to 13

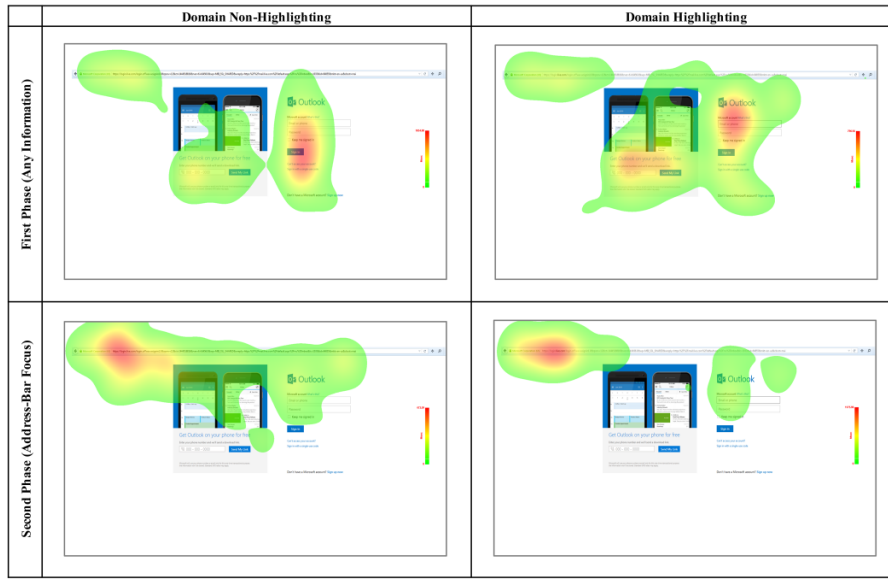

Figure 4. Legitimate Web page heat map, Hotmail.

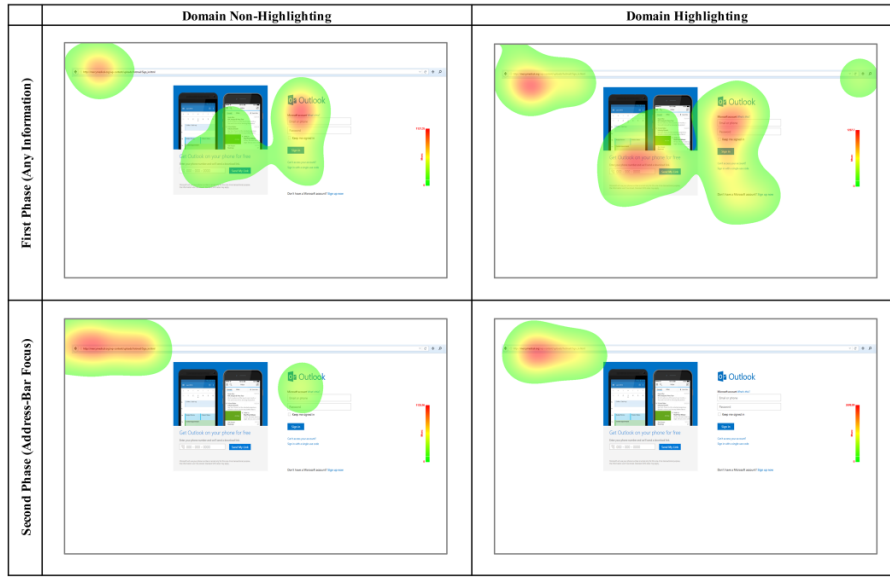

Figure 5. Fraudulent Web page heat map, Hotmail.

are results of each condition for fraudulent Web pages (see the figure caption for details).

The first phase in the current experiment was designed to simulate a natural Web page view. Although participants paid most of their attention to the visual cues, security information was also checked, but to a lesser extent. The security cues that participants used were mainly the green lock in the very beginning of the URL. When participants were motivated to pay more attention to the address bar, they extended their attention across the whole URL but still focused on the very left part. When the domain name was highlighted, participants attended more to the domain name, and their attention was also less distributed. Combined with the judgment results and eye-gaze

TABLE 4: Exit Survey Results Summary

	First Phase	Second Phase
Question 1 (methods used to evaluate the safety of the Web page in each phase)		
Address bar area only	4	21
Web page appearance only	15	0
Mixture of address bar and Web page's appearance	13	11
Question 2 (noticed domain highlighting previously)		
Yes	11	
No	20	
Question 3 (understood the intention of domain highlighting)		
Yes	5	
No	27	
Question 4 (effectiveness of domain highlighting)		
1 (*not effective*)	4	
2	6	
3	11	
4	11	
5 (*very effective*)	0	

pattern analysis, these findings indicate that participants may not understand the intent of the DH and cannot connect it to the security.

Exit survey analysis. Answers and comments were recorded for each of the five exit survey questions. The frequency of answers for Questions 1 to 4 are summarized in Table 4. For the first question, the methods participants used were categorized into three types separately by the experimenter and a rater who was blind to the purpose of the study. The categories were sufficiently distinct that there was no disagreement between the two raters.

Participants mainly relied on Web page appearance cues in Phase 1. For participants who stated that they used the address bar area for the safety judgment, most mentioned the green lock icon, HTTPS, or URL length, and none mentioned the DH feature. For the participants who mentioned URL length, all assumed that the longer the URL, the less secure the Web page, suggesting their lack of knowledge of URLs. Most participants were not aware of the DH feature previous and did not understand its intention. After the experimenter explained how DH could help people evaluate a Web page's safety, 21 participants gave a rating of < 4 for the effectiveness of DH, indicating a perceived ineffectiveness. When those participants were asked as a final question why they thought DH was ineffective, "hard to notice" was the most frequent reply. In general, the exit survey results suggest that DH is not effective at attracting users' attention and the safety usage of the DH is not self-explanatory.

GENERAL DISCUSSION

The primary motivation for our study was to assess the effectiveness of DH at influencing users' safety decisions of Web pages. The eye-tracking results showed that users' attention was directed to focus on the highlighted domain when they were explicitly asked to identify a Web site as safe or unsafe by focusing on the address bar area. However, the safety decision results indicate that users did not necessarily recognize the domain name or did not know that it can be used to decide the Web site's legitimacy. The failure of the latter two assumptions of DH was further verified in the exit survey of Experiment 2, in which 11 of 16 participants in the DH condition said they did not notice this feature previously, and 15 of them did not know what it was used for. Therefore, besides drawing people's attention, the knowledge required to process the security indicator's information in a meaningful way seems more critical for correct safety decisions.

The heat map results also showed that participants who attended to the address bar without explicit instructions mainly focused their attention to the very left part of the address bar area, which is consistent with their subjective report in the exit survey of using the lock or HTTPS. However, this natural behavior pattern may become a vulnerability for security decisions because the Certificate Authorities can be subverted (e.g., Microsoft, 2007) and fraudulent certificates can also be obtained by hackers (e.g., Bright, 2011a, 2011b). Besides, a significant number of phishing attacks have been reported being run on sites over HTTPS, for which SSL certifications have been issued (Netcraft, 2013). Relatedly, Wogalter and Mayhorn (2008) examined how quality seals associated with Web sites influence the credibility beliefs of the Web site and found the fictitious seals were rated as high as or higher than the seals actually used by reputable Web sites. Seals of approval, like the lock in the address bar area, are created by third-party organizations to represent a conformation to some set of standards of the Web site. It indicates some level of security and confidentiality and enhances the user's perceived credibility of the Web site. Wogalter and Mayhorn's results suggest a lack of discrimination or over-trust of the indicators issued by third parties. Therefore, it is dangerous for users to take for granted that the presence of a lock or HTTPS assures safety, which motivates the promotion of domain name in identifying phishing Web pages.

We also obtained a similar pattern of the fraudulent Web pages for which no security indicators were presented in the very left location, suggesting that users may follow their daily reading direction to process URL information. URLs work best when they are short and simple because this makes clear what they are. However, lengthy and complex URLs are used commonly, which confuses users. Thus, some people tend to judge the long and complex URLs as unsafe. One solution to increase the effectiveness of DH is to move the highlighted domain to the very beginning of the URL, which caters to the habit of users to attend to that location. Putting the domain name at the beginning of the URL may also facilitate users' recognition of the domain name because the "/" structure used in the URL is compatible to the structure used in PCs for file management, with which users are familiar. Placing the domain name at the beginning of the URL also puts the highlighted domain name closer to the Extended Validation (EV) certificates or other security indicators. Thus, a better detection rate is expected because there are multiple cues together to help users identify fraudulent Web pages.

Another reason to put emphasis on the domain name is that the placements of security indicators are not consistent across Web browsers. For example, the lock (Web site identification information) is located to the right end of the address bar in Internet Explorer, whereas the lock is placed before the URL in the address bar area for Chrome and Firefox. For Safari, the lock is placed just before the domain name. Thus, for Internet Explorer, users who do not notice the lock can still use the highlighted domain to facilitate the safety decision.

Limitations

In the present study, we presented Web page snapshots to the participants. Although this method kept participants focused on the tasks they were performing, they did not have a chance to interact with the Web pages as they would in the real-life environment. Also, participants were instructed to focus on security during both experiments, whereas security is often a secondary concern for users. Even though participants were more highly motivated than in the real world, the poor detection of phishing Web pages suggests that the results may be worse in naturalistic settings. Finally, the participants in Experiment 2 were young, educated students, and the eye-movement results need to be validated further in the general population of Web users.

Practical Implications

Task performance results of Experiments 1 and 2 suggest that the ineffectiveness of DH is probably due to people's lack of knowledge about how to use the highlighted information, which was confirmed by the heat map results showing that participants' visual attention distribution was impacted by the DH. Therefore, training people how to use the domain name to identify a phishing Web page becomes essential for improving the effectiveness of DH (e.g., Kumaraguru et al., 2009). We suggest the possibility to describe DH explicitly with other phishing prevention techniques, such as a phishing warning. In the hybrid

method, users would be instructed in the warning that they cannot rely on the content of the Web page and to check the highlighted domain name to verify that the Web site is the one it purports to be.

Because people also have difficulty discriminating the credibility of Web sites based on domain names (Wogalter & Mayhorn, 2008), the legitimate domain name or the method how to find the legitimate domain name should also be included in the warning for a comparison. Besides, a highlighted phishing name together with the legitimate domain name in the warning provide an instance of the common domain name phishing methods of which users should be aware. McDougald and Wogalter (2014) evaluated whether the use of relevant highlighting could benefit comprehension with a set of warning-related pictorials and found that comprehension of warning pictorials was higher for a relevant highlighting condition than less relevant highlighting and no highlighting conditions. The use of appropriately placed highlighting could benefit the design of a complex symbol by pointing out pertinent areas to aid in determining its intended conceptual meaning. Therefore, the hybrid method that we suggest is promising for helping users to avoid falling for phishing attacks. It also uses every phishing case as an opportunity to train users to exercise their knowledge. This method should foster habits in users to check each visited Web page's domain name.

CONCLUSION

In the two reported experiments, the task performance results showed some benefit of directing users' attention to the address bar, but DH did not provide effective protection against phishing attacks. Although participants' eye movements only showed statistical differences between phases and Web page legitimacy, the heat map results revealed that participants' fixations were scattered less when the domain name was highlighted than when it was not. Thus, participants indeed fixated more on the highlighted domain, although this apparently had little benefit on detecting that a Web page was fraudulent. Our findings suggest that the failure of DH is probably due to people's lack of knowledge about how to use the highlighted information.

ACKNOWLEDGMENT

This research was supported by a National Security Agency Grant as part of a Science of Security lablet through North Carolina State University. We thank Huangyi Ge for aid in setting up the online experiment, Wanling Zou for helping to conduct the laboratory experiments, and Seoyeon Lim for assisting with data analysis. The eye tracker was funded in part by the Department of Psychological Sciences and the College of Health and Human Sciences at Purdue University.

APPENDIX

Heat Maps for Different Web Page Types in Each Phase

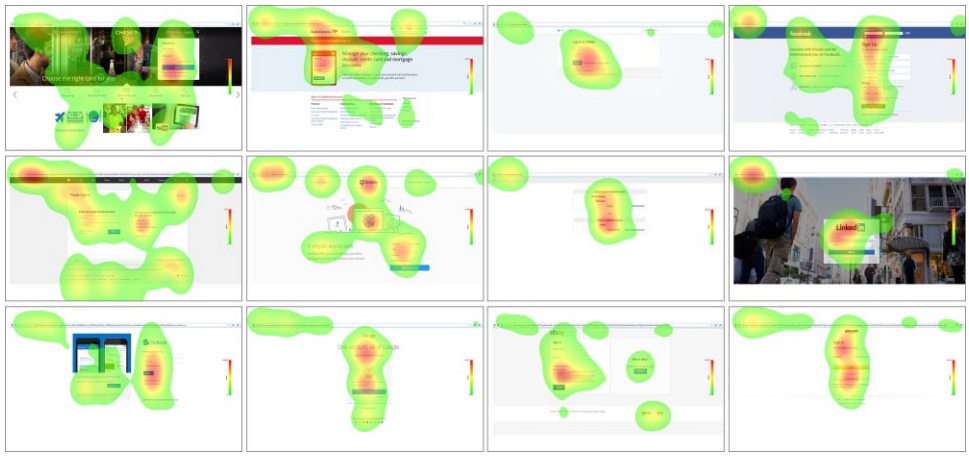

Figure 6. Phase 1 heat maps of all legitimate Web pages with domain non-highlighted.

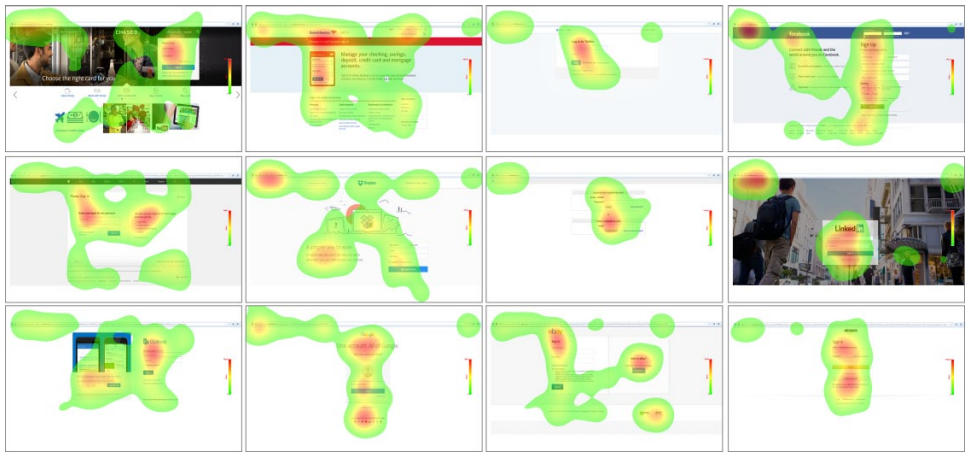

Figure 7. Phase 1 heat maps of all legitimate Web pages with domain highlighted.

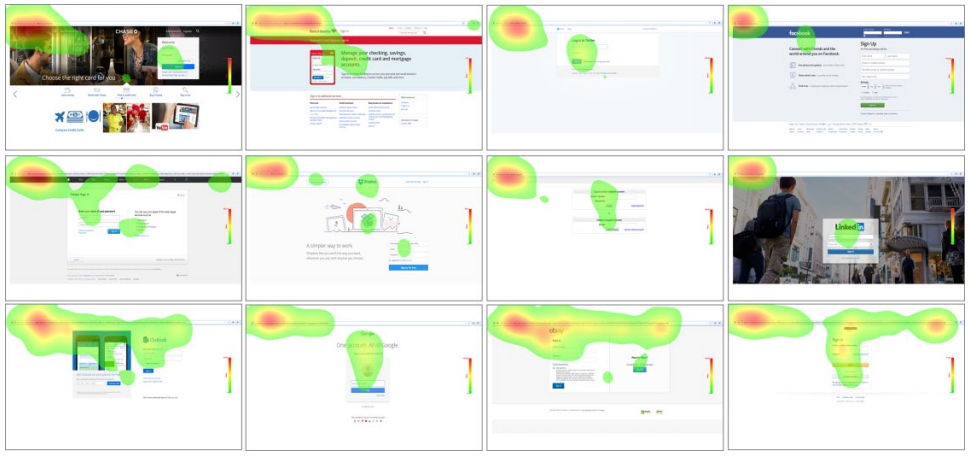

Figure 8. Phase 2 heat maps of all legitimate Web pages with domain non-highlighted.

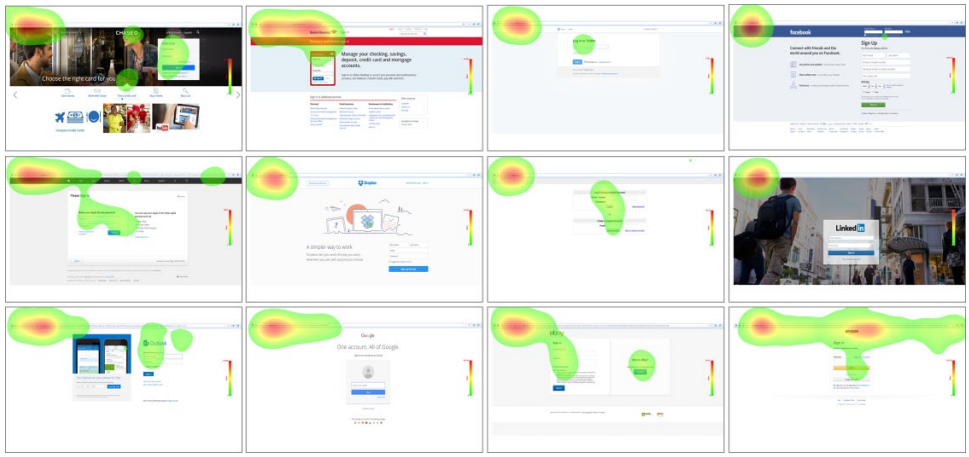

Figure 9. Phase 2 heat maps of all legitimate Web pages with domain highlighted.

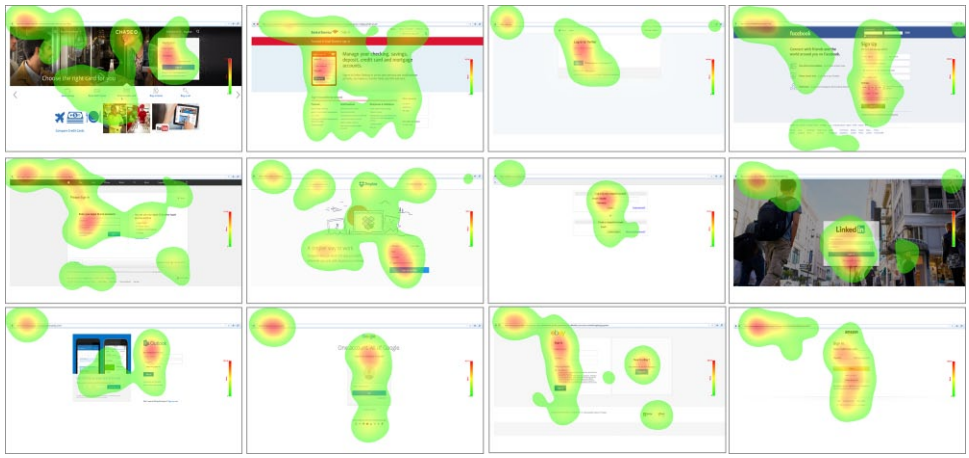

Figure 10. Phase 1 heat maps of all fraudulent Web pages with domain non-highlighted.

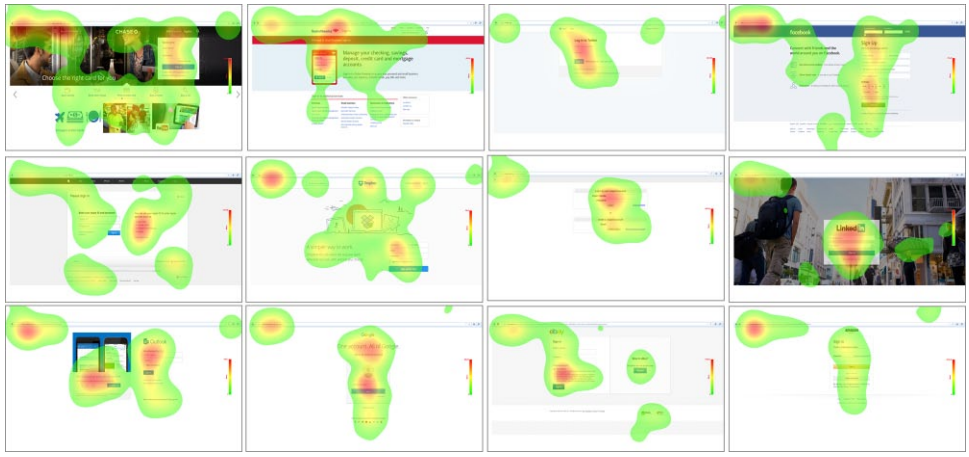

Figure 11. Phase 1 heat maps of all fraudulent Web pages with domain highlighted.

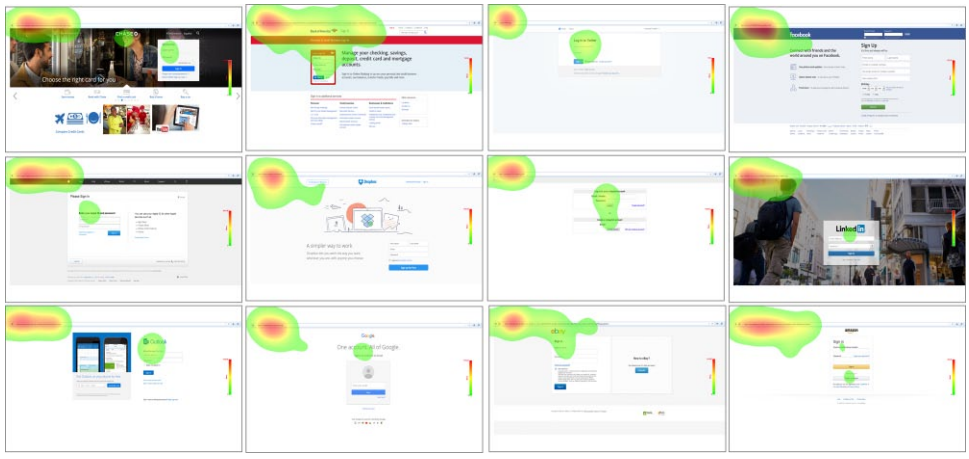

Figure 12. Phase 2 heat maps of all fraudulent Web pages with domain non-highlighted.

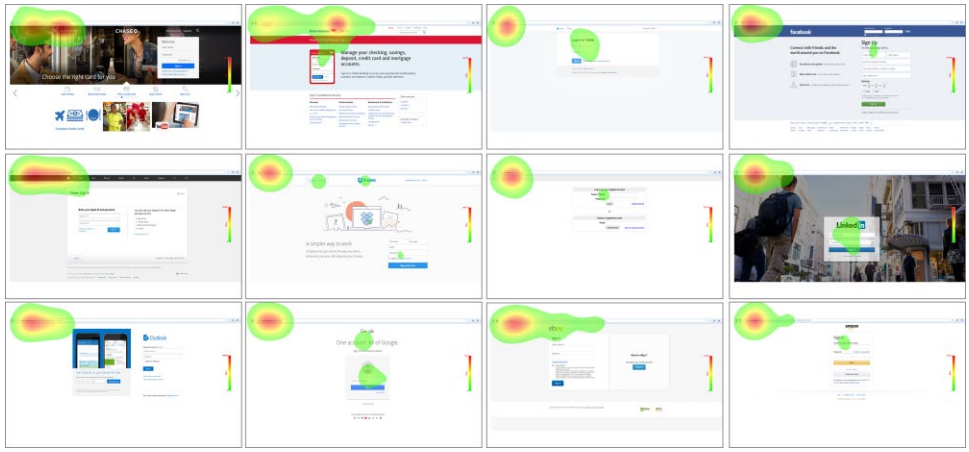

Figure 13. Phase 2 heat maps of all fraudulent Web pages with domain highlighted.

KEY POINTS

- We pursue a two-level measurement of users' information processing of a security indicator (the domain name), examining decisions about whether a Web page is safe and tracking the locations on which users fixate while making their safety decisions.
- Domain highlighting has little benefit on judgments of Web page legitimacy, although there is some benefit of directing a person's attention to the address bar.
- When participants attend to the domain highlighting function, lack of knowledge of the domain name or how to use the domain name limits the effectiveness of the highlighting.

REFERENCES

Aaron, G., Rasmussen, R., & Routt, A. (2014). *Global phishing survey: Trends and domain name use in 2h2013*. Retrieved from http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_2H2013.pdf

Arianezhad, M., Camp, L. J., Kelley, T., & 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* (pp. 105–116). New York, NY: ACM.

Bravo-Lillo, C., Cranor, L. F., Downs, J., & Komanduri, S. (2011). Bridging the gap in computer security warnings: A mental model approach. *IEEE Security & Privacy*, *9*, 18–26.

Bright, P. (2011a). *Another fraudulent certificate raises the same old questions about certificate authorities*. Retrieved from http://arstechnica.com/security/2011/08/earlier-this-year-an-iranian/

Bright, P. (2011b). *Independent Iranian hacker claims responsibility for Comodo hack*. Retrieved from http://arstechnica.com/security/2011/03/independent-iranian-hacker-claims-responsibility-for-comodo-hack/

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, *58*, 1158–1172.

Cao, Y., Han, W., & Le, Y. (2008). Anti-phishing based on automated individual white-list. In *Proceedings of the 4th ACM Workshop on Digital Identity Management* (pp. 51–60). New York, NY: ACM.

Carpenter, S., Zhu, F., & Kolimi, S. (2014). Reducing online identity disclosure using warnings. *Applied Ergonomics*, *45*, 1337–1342.

Dhamija, R., & Tygar, J. D. (2005). The battle against phishing: Dynamic security skins. In *Proceedings of the 2005 Symposium on Usable Privacy and Security* (pp. 77–88). New York, NY: ACM.

Dhamija, R., Tygar, J. D., & Hearst, M. A. (2006). Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 581–590). New York, NY: ACM.

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the Second Symposium on Usable Privacy and Security* (pp. 79–90). New York, NY: ACM.

Downs, J. S., Holbrook, M., & Cranor, L. F. (2007). Behavioral response to phishing risk. In *Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit* (pp. 37–44). New York, NY: ACM.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice*. London: Springer-Verlag.

Felt, A. P., Ainslie, A., Reeder, R. W., Consolvo, S., Thyagaraja, S., Bettes, A., & Grimes, J. (2015). Improving SSL warnings: Comprehension and adherence. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2893–2902). New York, NY: ACM.

Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. In *Proceedings of the 16th International Conference on World Wide Web* (pp. 649–656). New York, NY: ACM.

Fu, A. Y., Liu, W. Y., & Deng, X. (2006). Detecting phishing Web pages with visual similarity assessment based on earth mover's distance (EMD). *IEEE Transactions on Dependable and Secure Computing*, *3*, 301–311.

Gopal, R. D., Tripathi, A. K., & Walter, Z. D. (2006). Economics of first-contact email advertising. *Decision Support Systems*, *42*, 1366–1382.

Grazioli, S. (2004). Where did they go wrong? An analysis of the failure of knowledgeable Internet consumers to detect deception over the Internet. *Group Decision and Negotiation*, *13*, 149–172.

Herzberg, A., & Gbara, A. (2004). *Trustbar: Protecting (even naive) Web users from spoofing and phishing attacks* (Cryptology ePrint Archive Report 2004/155). Retrieved from http://eprint.iacr.org/2004/155.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*, 94–100.

Johnson, P. E., Grazioli, S., Jamal, K., & Berryman, G. (2001). Detecting deception: Adversarial problem solving in a low base rate world. *Cognitive Science*, *25*, 355–392.

Khonji, M., Iraqi, Y., & Jones, A. (2013). Phishing detection: A literature survey. *IEEE Communications Surveys & Tutorials*, *15*, 2091–2121.

Kumaraguru, P., Cranshaw, J., Acquisti, R., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). *A real-word evaluation of anti-phishing training*. Technical report, Carnegie Mellon University.

Lin, E., Greenberg, S., Trotter, E., Ma, D., & Aycock, J. (2011). Does domain highlighting help people identify phishing sites? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2075–2084). New York, NY: ACM.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Mahwah, NJ: Lawrence Erlbaum.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*, 185–199.

McDougald, B. R., & Wogalter, M. S. (2014). Facilitating pictorial comprehension with color highlighting. *Applied Ergonomics*, *45*, 1285–1290.

Microsoft, Inc. (2007). *Microsoft Security Bulletin MS01-017: Erroneous VeriSign-issued digital certificates pose spoofing hazard*. Retrieved from https://support.microsoft.com/en-us/kb/293818.

Miyamoto, D., Iimura, T., Blanc, G., Tazaki, H., & Kadobayashi, Y. (2014). EyeBit: Eye-tracking approach for enforcing phishing prevention habits. In *Proceedings of the 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security (BADGERS)*. Retrieved from http://www.necoma-project.eu/m/filer_public/d2/46/d24651e7-1f35-4846-b8c3-52f3d941a58e/miyamoto-badgers2014.pdf

Netcraft. (2013). *Phishing alerts for SSL certificate authorities*. Retrieved from http://www.netcraft.com/anti-phishing/certificate-authority-phishing-alerts/

Neupane, A., Rahman, M. L., Saxena, N., & Hirshfield, L. (2015). A multi-modal neuro-physiological study of phishing detection and malware warnings. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 479–491). New York, NY: ACM.

Orgill, G. L., Romney, G. W., Bailey, M. G., & Orgill, P. M. (2004). The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems. In *Proceedings of the 5th Conference on Information Technology Education* (pp. 177–181). New York, NY: ACM.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers & Security*, *52*, 194–206.

Proctor, R. W., & Chen, J. (2015). The role of human factors/ergonomics in the science of security: Decision making and action selection in cyberspace. *Human Factors*, *57*, 721–727.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Schneier, B. (2000). Semantic network attacks. *Communications of the ACM*, *43*, 168.

Sheng, S., Wardman, B., Warner, G., Cranor, L. F., Hong, J., & Zhang, C. (2009). An empirical analysis of phishing blacklists. In *Proceedings of the 6th Conference on Email and Anti-Spam, CEAS'09*. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1286&context=hcii

SR Research. (2009). *Eyelink 1000 user's manual*. Mississauga, Canada: Author.

Stone, A. (2007). Natural-language processing for intrusion detection. *Computer*, *40*, 103–105.

Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., & Cranor, L. F. (2009). Crying wolf: An empirical study of SSL warning effectiveness. In *Proceedings of the USENIX Security Symposium* (pp. 399–416). New York, NY: ACM.

Welk, A. K., Hong, K. W., Zielinska, O. A., Tembe, R., Murphy-Hill, E., & Mayhorn, C. B. (2015). Will the "phisher-men" reel you in?: Assessing individual differences in a phishing detection task. *International Journal of Cyber Behavior, Psychology and Learning*, *5*, 1–17.

Whalen, T., & Inkpen, K. M. (2005). Gathering evidence: Use of visual security cues in Web browsers. In *Proceedings of Graphics Interface 2005* (pp. 137–144). Waterloo, Canada: Canadian Human-Computer Communications Society.

Whittaker, C., Ryner, B., & Nazif, M. (2010). Large-scale automatic classification of phishing pages. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2010*. Retrieved from https://www.isoc.org/isoc/conferences/ndss/10/pdf/08.pdf

Wogalter, M. S., & Mayhorn, C. B. (2008). Trusting the Internet: Cues affecting perceived credibility. *International Journal of Technology and Human Interaction*, *4*, 75–93.

Wu, M., Miller, R. C., & Garfinkel, S. L. (2006). Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 601–610). New York, NY: ACM.

Aiping Xiong is a PhD student in cognitive psychology and human factors at Purdue University. She earned her master's degree in industrial engineering at Purdue University in 2014.

Robert W. Proctor is a distinguished professor at the Department of Psychological Sciences of Purdue University. He received his PhD in experimental psychology at the University of Texas at Arlington in 1975.

Weining Yang works at Google, Inc. He received his PhD in computer science from Purdue University in 2016.

Ninghui Li is a professor in the Computer Science Department of Purdue University, and he got his PhD in computer science at New York University in 2000.

Date received: July 11, 2016
Date accepted: November 13, 2016