# Teaching Johnny Not to Fall for Phish

Ponnurangam Kumaraguru[1], Steve Sheng[2], Alessandro Acquisti[3],
Lorrie Faith Cranor[1,2], Jason Hong[4]
[1]Institute for Software Research, [2]Engineering and Public Policy,
[3]Heinz School of Public Policy, [4]Human Computer Interaction Institute
Carnegie Mellon University
ponguru@cs.cmu.edu

## ABSTRACT
Phishing attacks exploit users' inability to distinguish legitimate websites from fake ones. Different strategies can be employed to solve the problem of phishing: detecting phishing scams, providing better user interfaces, and training users. While a great deal of effort has been devoted to the first two solutions, little research has been done in the area of educating and training users. In this paper we present the results of a user study we conducted to test the effectiveness of existing online training materials. We find that these training materials are surprisingly effective when users actually read them. We then analyze the training materials using principles from learning sciences, and provide some suggestions on how to improve them based on those principles.

## Categories and Subject Descriptors
C.2.0 [Computer - Communication Networks]: General—Security and protection; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Graphical user interfaces; K.4.2 [Computers and Society]: Social issues—Abuse and crime involving computers.

## General Terms
Design, Experimentation, Security, Human Factors.

## Keywords
Training, anti-phishing, phishing, embedded training, email, usable privacy and security.

## 1. INTRODUCTION
Phishing attacks exploit users' inability to distinguish legitimate company websites from fake ones. Phishers send out spoofed emails that look as if they were sent by trusted companies. These emails lead to spoofed websites that are similar or virtually identical to legitimate websites, and lure people into disclosing sensitive information. Phishers use that information for criminal purposes, such as identity theft [25], [28].

People are vulnerable to phishing attacks because spoofed websites look very similar to legitimate websites. In fact, Dhamija et al. showed that people have trouble identifying phishing sites even in tests in which they have been alerted about the possibility of such attacks [10]. Furthermore, when phishers personalize their emails, they can further increase the likelihood that the attack will be successful [20], [24].

Researchers have been working on technical approaches to countering phishing attacks, such as toolbars, email filters, and verified sender addresses [16]. However, these approaches are not foolproof. In a recent study of 10 anti-phishing toolbars, only one toolbar was able to correctly identify over 85% of phishing websites, and that one toolbar also incorrectly identified about one-third of legitimate websites as fraudulent [9]. Furthermore, while automated phishing detection is improving, phishers are adapting their attack techniques to improve their chances of success and avoid detection. Finally, contextual information known to the recipient may be needed to determine whether some email messages are legitimate. Thus, we argue that automated detection systems should be used as the first line of defense against phishing attacks, but since these systems are unlikely to perform flawlessly, they should be complemented with training to help people make better recognize fraudulent email and websites.

However, end-users are not necessarily receptive to computer security education. Computer security is a secondary task that often interferes with people's primary tasks (for example, communicating with others, gaming, business needs, etc.) [2]. In addition, most computer users have little knowledge of computer security, and little desire to spend time learning about it [41]. Thus, training users about computer security can be difficult, and some experts have concluded that anti-phishing education will not work.[1]

In this paper we show that despite the difficulties described above anti-phishing education can be effective. We present the results of a user study in which users spent 15 minutes reading web-based anti-phishing educational materials and then demonstrated significant improvements in their ability to recognize fraudulent websites. However, we also found some areas where these materials could be improved. Using principles from learning sciences, we analyzed the training materials and derived some suggestions for content and presentation of these materials

The remainder of this paper is organized as follows. We present related work in Section 2. In Section 3 we present the design and the results of the user study we conducted to evaluate the effectiveness of the existing online training materials. In Section 4 we present our analyses of training materials using learning science principles. In Section 5 we present the lessons learned from the user study with some general suggestions to improve the training materials. We conclude in Section 6 by discussing conclusions and future research work.

[1] At the fall 2006 Anti-Phishing Working Group (APWG) general meeting some of the speakers asserted that users could not be educated (http://www.antiphishing.org/events/2006_fallGeneralMeeting.html).

## 2. RELATED WORK

The volume of phishing attacks is increasing. According to the Anti-Phishing Working Group (APWG), the number of unique phishing websites reported in August 2006 was 10,091, compared to 7,197 in December 2005 [5]. Gartner estimates the total financial loss in 2006 due to phishing to be $2.8 billion [30]. Not only do victims lose their money and identities, but they also undergo significant emotional stress [25]. Many solutions to this problem have been proposed. We classified them into three categories: (1) preventing and detecting phishing scams; (2) providing better user interfaces; and (3) training the users.

### 2.1 Preventing and Detecting Phishing Scams

One way to combat phishing scams is to prevent spoofed emails and web pages from reaching the end user. This can be achieved in a number of ways: (1) implementing filters to detect and delete emails automatically at the server [19], [38]; (2) finding and shutting down suspicious websites that have domain names similar to trusted brands; (3) installing toolbars to detect phishing websites (described in more detail in the next subsection); and (4) using domain keys and Sender Policy Framework (SPF) to verify the DNS domain of the email server and to reject forged addresses in the SMTP mail from address respectively [12], [37].

Given current Internet technology and regulatory status, phishing attacks cannot be prevented completely. For example, filters are clearly not 100% effective, since phishing emails still routinely reach the inbox of many users. In addition, false positives are a serious concern for email filters. With respect to finding and shutting down suspicious websites, due to cross-border jurisdiction, it is difficult to remove websites hosted in different countries: according to the Anti-Phishing Working Group (APWG), phishing sites stay online on average for 4.5 days [5]. For the domain keys solution to be successful, the adoption rate among organizations needs to be high. In short, techniques for preventing and detecting phishing scams are not foolproof. Consequently, we believe that users also have to be trained in both identifying and detecting these phishing emails. Our ultimate goal is to develop training materials to educate users to identify and detect these phishing emails. The work in this paper presents a study helping us understand how effective current training materials are, and takes us a step closer towards that goal.

### 2.2 Providing Better User Interfaces

Certain solutions provide visual indicators to help users identify potential phishing scams. For example, some anti-phishing toolbars show different colors (such as red, yellow, or green) to indicate the degree of danger of a website, while some provide an estimate of its "spoof rate." Some of the toolbars available are Account Guard [1], EarthLink [14], Google Toolbar [21], IE7 toolbar [31], Netcraft [34], SpoofGuard [40], SpoofStick [39], and Zillabar [47].

Toolbars can be effective because they present potentially relevant aspects of the underlying system model to users (i.e. hidden state such as the age of the website). Having a clearer model of the current state of things can help clarify misconceptions about what the system is doing and help users make better decisions. However, there are three weaknesses in this approach: first, it requires people to install special software (though newer versions of web browsers ship with the software already built in). Second, studies have shown that users often do not understand or act on the cues provided by toolbars [32], [44]. Third, a recent study shows that some anti-phishing toolbars are not very accurate, and even the best toolbars may miss nearly 20% of phishing websites [9]. Other tools, such as PassPet and WebWallet, try to engage users in making an active operation (either by pressing a button or by approving the action) before giving out sensitive information [42], [43], [46]. However, even these solutions ultimately rely on the users' ability to make the right decision.

Ye et al. [45] and Dhamija and Tygar [10] have developed "trusted paths" for the Mozilla web browser that are designed to assist users in verifying that their browser has made a secure connection to a trusted site. Herzberg and Gbara have developed TrustBar, a browser add-on that uses logos and warnings to help users distinguish trusted and untrusted websites [22].

In all the above systems, users are still involved in the decision-making process. These tools can only aid users in making a decision; they do not make the decision for users. Studies have shown that users often disregard the information presented. This may be due to a lack of awareness of the consequences of their behavior. This suggests a need to raise users' awareness about phishing and to train users on how to avoid falling for these attacks.

### 2.3 Training the Users

A few approaches have focused on educating and training users about phishing. The most basic approach is to provide online information regarding phishing. This has been done by government organizations [18], non-profit organizations [3] and business organizations [15]. Another approach allows users to take tests on phishing websites and emails. For example, Mail Frontier [29] has set up a website containing screenshots of potential phishing emails. Users are scored based on how well they can identify which emails are legitimate and which are not. Robila et al have also tried training students in a class room setting, demonstrating that some simple tests plus class discussion helped students be more aware of and be better at recognizing phishing attacks [36].

Researchers have also tried a *contextual training* approach in which users are sent phishing emails to probe their vulnerability. At the end of the study, users are typically given additional materials informing them about phishing attacks in general. This approach has been used at Indiana University in studies conducted on students about contextual attacks making use of personal information (also known as spear-phishing) [24], at West Point [20], [23] and at a New York State Office [35].

In a related paper, we have also developed and evaluated an email-based approach to train people to avoid phishing attacks [26]. We called this approach *embedded training*, in that it trains people during their regular use of email. As in previous studies, we sent our subjects phishing emails, and then presented an intervention warning people who had fallen for our messages. Our study was conducted in a laboratory and interventions were presented immediately when users clicked on a phishing link in the email, rather than at the end of the study. Our goal was to evaluate how effective various intervention designs were and how well people could transfer knowledge from one situation to another. We created several designs based on learning sciences (for example, contiguity and personalization principle [8]), and found that our interventions were more effective than standard security notices.

The work presented in this current paper addresses the questions "Can users be trained to identify phishing websites?" and "What is the effectiveness of existing online training materials?" Our participants were asked to identify whether a set of websites were legitimate or spoofed. In contrast, in the earlier study cited above [26], we addressed the questions "Are the security notices that organizations send out effective?" and "How effective is contextual training by sending phishing emails to users and providing the training materials when they fall for the phishing attacks?"

# 3. USER STUDY

The goal of our study is to determine the effectiveness of available web-based anti-phishing training materials. In this section we present the study design, participant details, and results.

## 3.1 Study Design

We based the design of our user study on Dhamija et al.'s study of phishing websites [11]. Users were given the following scenario: "You have received an email message that asks you to click on one of its links. Imagine that you have clicked on the link to see if it is a legitimate website or a spoofed website." We then presented users with twenty websites and asked them to state whether a website was legitimate or phishing, as well as confidence of their judgments (from a scale of 1-5 where 1 stands for not confident at all, and 5 for very confident).

We used 20 websites for the study: ten of them are phishing sites collected from APWG database; the legitimate websites are from popular financial institutions, online merchants, and a few other random websites. We divided up the twenty websites into two groups (A, B). In our test, users were asked to view a group first (pre test), followed by a fifteen minute break to complete a task prescribed by the conditions below, after which they viewed the second group of websites (post test). We randomized the order of pretest and post test, so that half the users used Group A in the pre test, and half used group B in the pre test. The list of websites used is shown in Table 2 .

We used two experimental conditions: control and training:

- **Control condition**: In this condition, participants were asked to complete a task between pre and post test that was irrelevant to the goals of the study. They were asked to play simple computer games such as solitaire and minesweeper.
- **Training material condition**: Instead of performing an irrelevant task at the break, participants in this condition were asked to read what we judged to be the best educational material on phishing currently available. The rest of the setup was identical to that used for the control condition.

Each group (A and B) includes 5 phishing sites and 5 real sites. We hosted these phishing websites on the local computer by modifying the host DNS file, so our participants were not at risk. Our study design was a between-subjects design. To record and capture the screen of the interviews we used Camtasia Studio [7].

## 3.2 Training Materials

The training materials were selected based on content and popularity from a list of 24 online anti-phishing training materials. Our final selections were eBay's tutorial on spoofed emails, Microsoft's Security tutorial on Phishing and Phishing E-card from the U.S. Federal Trade Commission [15], [17], [31]. We also included a URL tutorial from MySecureCyberSpace, which is a portal for educating people about security risks and countermeasures on the Internet [33]. In Table 1, we present information about the format of the instruction, length of the instructions in number of words, length of the instructions in number of printed pages, number of graphic examples, and what concepts they try to teach about phishing. All the training materials that we used for the study had some form of link to other resources for people to further read about phishing and security in general.

Almost all the training materials started with some basic definition of phishing. An example definition is "Claiming to be sent by well-known companies, these emails ask consumers to reply with personal information, such as their credit card number, social security number or account password." All the materials presented a variation of this definition. Almost all the materials initially also provided definitions of "spoof emails" and then connected them to phishing emails.

These training materials also highlighted some characteristics of phishing emails and provided suggestions for how to avoid falling for such scams. Table 1 presents the characteristics of the emails and the suggestions. Almost all the materials mention some version of "organizations do not request personal information through emails." Finally, these materials also presented information about what to do after falling for phishing emails. These suggestions included: reporting or forwarding the phishing email to spoof@ebay.com, report to FTC, etc.

## 3.3 Participants

We recruited 14 participants for each condition, for a total of 28 subjects. To recruit participants, we posted flyers around campus, posted recruitment messages on university bulletin boards and on craigslist.com.

We screened participants with respect to their knowledge of computers in general, aiming to recruit only participants who could be considered "non-experts". We recruited users who answered "no" to two or more of the following screening questions: 1) whether they had ever changed preferences or settings in their web browser, 2) whether they had ever created a web page, and 3) whether they had ever helped someone fix a computer problem. These questions have served as good filters to recruit non-experts in other studies [13], [26].

Our subjects had the following demographics:

□ Gender: 39% percent of the users were male, and 61% percent of the users were male.

□ Age: 85% of the users were between the ages of 18-34, 7% were between 35-44 years old, 4% were between 45-64 years old, and 4% declined to answer.

□ Education Level: 14% of the users had high school or less education, 39% of the users were college undergraduates, 18% were college graduates, and 29% were post graduate students.

## 3.4 Results

In this section, we present the result of our study. We find that subjects in the training condition demonstrated significant improvements in their ability to recognize fraudulent websites.

### *3.4.1 Effectiveness of Training*

We use two metrics to measure the effectiveness of training: the number of false positives and the number of false negatives. A

false positive is when a legitimate site is mistakenly judged as a phishing site. A false negative is when a phishing site is incorrectly judged to be a legitimate site.

**Table 1: Information about the training materials. N/A is "not applicable". Presenting the signals that the instruction is asking to look for to identify phishing emails and suggestions to avoid falling for phishing attacks.**

| Organization | Content format | Length in words | # of printed pages | # of graphic examples | Cues to look for in the email | Suggestions |
|---|---|---|---|---|---|---|
| Microsoft | Webpage | 737 | 3 | 2 | - Urgency / action status in the email<br>- Greetings in the email<br>- Requesting personal information through email | - Mouse over the link to check whether it is taking to the website that it is claiming |
| eBay | Webpage | 1276 | 5 | 8 | - Sender email address<br>- Greetings in the email<br>- Urgency / action status in the email<br>- Links in the email<br>- Requesting personal information through email<br>- legitimate eBay address versus fake eBay address | - Open a new browser to type in the URL<br>- Never click on the link in the email<br>- How to identify legitimate eBay Address. |
| FTC Phishing E-card | Video | N/A | N/A | N/A | - Requesting personal information through email | - Do not give personal information through emails |
| URL tutorial | Webpage | 236 | 1 | 0 | N / A | N / A |

False negatives are usually worse than false positives in phishing, because the consequence of mistaking a legitimate site to be phishing is a matter of inconvenience, whereas the consequence of mistaking a phishing site to be real can lead to identity theft.

In our analysis, the false positive and false negative rates are calculated as:

$$\text{False Positive Rate} = \frac{\text{number of false positives}}{\text{number of legitimate sites}}$$

$$\text{False Negative Rate} = \frac{\text{number of false negatives}}{\text{number of phishing sites}}$$

We found that for the training group, there is a significant reduction in the false negative rate after the training - from 0.40 to 0.11 (paired t-test: $\mu1=0.40$, $\mu2=0.11$, $p = 0.01$, $DF = 13$). There is no statistical significant change in the false negative rate for the control group (paired t-test: $\mu1=0.47$, $\mu2=0.43$, $p=0.29$, $DF=13$).

We also tabulated the training group's performance by website. We show in Table 2, for each website, the percentage correct rate before training and after training. We find that users made improvements in 11 of the twenty sites, did not change in four sites, and performed worse in 5 of them.

While the false positive rate remained virtually unchanged for the control group, it increased from 0.31 to 0.41 in the training group. However, this increase is not statistically significant. (paired t-test: $\mu1=0.31$, $\mu2=0.41$,$p=0.12$, $DF = 13$). We explain the reason for increase in false positives in detail in Section 3.4.3.

### *3.4.2 User Strategies for determing website legitimacy*

Our analysis shows that users used a variety of strategies to determine website legitimacy and these strategies vary in their effectiveness. This is in alignment with other research. Previous user studies [11], [13] have discussed user's strategies for determining website legitimacy. Dhamija et al. categorized user strategies into five categories: relying on web contents only; relying on content and domain name only; relying on content with address and https; relying on all the above plus padlock icon; and relying on all of the above plus certificates. Downs et al. discuss cues that users are sensitive to when judging the legitimacy of a site. Such cues include secure site lock icons, broken images on the webpage, unexpected or strange URLs, and the indication of an https connection.

**Figure 1: False negative rates for the test groups. N (control) = 14, N(training) = 14**

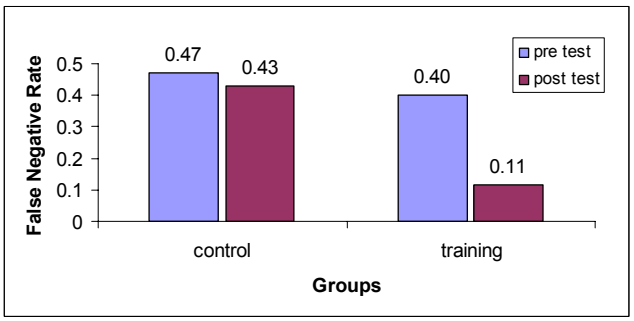

**Table 2: Percentage of correct predictions for the training group before and after the training**

| Website | Real / Spoof | Description | Pre Training % correct (avg conf) | Post Training %correct (avg conf) | Change |
|---|---|---|---|---|---|
| Paypal | Spoof | Fake URL bar displaying the real paypal URL. not requesting too much information | 14 (4) | 71 (4.4) | +57% |
| PNC Bank | Spoof | Bank account update. Pop up window overlay the real PNC Bank, security lock, requesting credit card number. | 57 (3.7) | 100 (4.1) | +43% |
| Citicards | Spoof | Citicard account update. Lock on the page, requesting a lot of information | 42 (4.3) | 85 (4.5) | +43% |
| Royal Bank of Canada | Spoof | Sign in online banking page. Layered information request, URL: www.etsongfc.com/cgi-bins/rbaccess/rbunxcgi/RBC.htm | 42 (3.3) | 85 (4.8) | + 43% |
| HSBC | Spoof | Internet banking login page, layered information request. URL: IP address | 50 (4.0) | 85 (4.8) | + 35% |
| Chase Student Loans | Real | Primitive designed Page. | 28 (4.5) | 50 (4.3) | +22% |
| Paypal | Real | Paypal login page | 85 (4.5) | 100 (4.5) | + 15% |
| Barclays | Spoof | Faked Barclays login page, layered information request. URL: IP address | 85 (4.1) | 100 (4.4) | +15% |
| AOL | Spoof | AOL account update, animation at the beginning that mimics AOL, requesting a lot of information, URL: myaol.com/verifybilling/ | 85 (4.0) | 100 (4.7) | +15% |
| Halifax Bank | Spoof | Halifax bank log in page, security lock, layered information request. When user click on the link, fake prompt that now leaves secure site, URL: www.halifax-cnline.co.uk/_mem_bin/index.php | 85 (4.6) | 100 (4.4) | +15% |
| eBay | Real | eBay register page, requesting lots of information | 28 (5.0) | 42 (4.6) | +14% |
| Etrade | Real | Etrade home page, | 100 (4.1) | 100 (4.2) | 0% |
| eBay | Spoof | Faked eBay login page. URL: IP address | 85 (4.8) | 85 (4.8) | 0% |
| Wellsfargo bank | Spoof | Faked Wellsfargo home page, layered information request, URL online.wellsfargo.wfosec.net/update | 71 (4.0) | 71 (3.8) | 0% |
| Desjardins | Real | Account log in page. Unfamiliar foreign bank. | 57 (3.0) | 57 (3.5) | 0% |
| MBNA business | Real | domain name usecfo.com has nothing to do with MBNA. | 42 (4.3) | 28 (3.5) | -14% |
| Bank of America | Real | Bank of America home page | 83 (4.2) | 57 (3.7) | -26% |
| Chase online | Real | Online banking Log in page, URL: chaseonline.chase.com | 100 (4.5) | 71 (2.8) | -29% |
| Citibank | Real | Citibank login Page | 71 (4.0) | 42 (4.0) | -29% |
| US Bank | Real | Online banking login page. URL: www4.usbank.com | 100 (4.2) | 57 (4.2) | -43% |

During the interview phase of our study, we asked users to think aloud the reasons for their decisions. We recorded these reasons and clustered them around seven categories: design and content, URL, information requested, consistency, search engine, prior knowledge, and security indicators. Table 3 explains these strategies in detail and shows the percentage of times they were used by the subjects for the control group. To ensure free from bias from study administrators, the test administrator only prompted users to speak about their decisions if they did not do so (which usually only happened at the beginning of a study). At no point of time in the study did the test administrators provide hints or ask them to look at certain cues.

### *3.4.3 What users are learning, and what they are not learning*

We compared the strategies that our participants used before and after the training (see Table 4). Our results show that the participants in the training group relied on the design and content of a website, as well as their prior knowledge, less often after the break. Furthermore, they examined the URLs of the webpage and the amount of information requested more often. Both of these

**Table 3: User Strategies with the percentage of usage for the control group**

| Strategies | Examples | Percentage of time used (Control Group) |
|---|---|---|
| Design & Content | - The design of the websites are poor/ professional.<br>- The links (images) are functioning / broken.<br>- There are up to date contact information, copyright statement, privacy and security statements.<br>- There are security locks in the content, verisign symbol, TRUSTe logo | 42% |
| URL | - The URL has number in them.<br>- The address looks suspicious. | 31% |
| Information requested | - Amount of information requested is too much / alright.<br>- The website is / not requesting sensitive information.<br>- It is alright / weird for website to request my information here. | 19% |
| Consistency | - The URLs of different pages are in the same site.<br>- All the links on one page are pointing to the same site.<br>- Logos and colors of different pages match. | 16% |
| Search engine | - using search engine to double check the legitimacy of the site. | 16% |
| Prior knowledge | - I have an account with the company, I know this company.<br>- I have seen the website / know the company.<br>- I have / know someone who is a victim of this site. | 6% |
| Security indicator | - The URL has https in them.<br>- There is secure browser pad lock. | 3% |

results are positive, in that our participants learned to avoid poor strategies and started to adopt good strategies. Finally, we did not observe any significant changes in the control group.

The training materials taught that phishing sites often request sensitive user information (such as credit card PIN number and social security numbers), whereas legitimate companies do not. After training, our participants paid more attention to what information the website is requesting. This leads us to conclude that users are learning this technique.

**Table 4: Percentage change in strategies that users used**

| Strategies | Training (change) | Control (change) |
|---|---|---|
| Design & content | -15% | -1% |
| Prior knowledge | -11% | -5% |
| URL | +23% | +2% |
| Information requested | +13% | -3% |

As for URLs, the Microsoft and eBay training materials teach (1) what is their correct URL of their respective sites, and (2) what are some examples URLs that phishers use to trick people. However, the training materials do not teach about in general how to identify long URLs, specifically the sub-domain ones.

**Table 5: Reasons of failure of users for post training**

| Website | Pre Training % correct (avg conf) | Post Training %correct (avg conf) | Change | Reasons for failure |
|---|---|---|---|---|
| MBNA business (real) | 42 (4.3) | 28 (3.5) | -14% | Domain name usecfo.com has nothing to do with MBNA. |
| Bank of America (real) | 83 (4.2) | 57 (3.7) | -26% | Weird URL |
| Chase online (real) | 100 (4.5) | 71 (2.8) | -29% | Weird URL chaseonline.chase.com, expecting chase.com |
| Citibank (real) | 71 (4.0) | 42 (4.0) | -29% | weird URL web-us.da.citibank.com |
| US Bank (real) | 100 (4.2) | 57 (4.2) | -43% | Weird URL www4.usbank.com, expecting www.usbank.com |

For identifying IP addressed based scams (which has IP address in the URL instead of text), subjects in the training group seems to

perform quite well, as only one user failed to recognize them (and failed twice on it). His rationale was that "both of the two sites do not ask for much information." In contrast, in the control group, our participants failed to identify seven IP-address-based phishing sites.

Phishing sites use deceptive URLs which are hard to detect. In Dhamija's study, 92% of the users fall for the deceptive domain of www.bankofthevvest.com (two "v", instead of "w"). Surprisingly in our study, none of our users in the training group fall for the deceptive domain attack after training. These included halifax-cnline.com (change of "o" to "c" in halifax-online.com), which our training participants noticed the typo immediately.

However our participants had a hard time interpreting longer URLs, especially URLs using sub-domains. For example, many of our participants in the training condition labeled wellsfargo.com.wfcnet.net as legitimate sites because the word wellsfargo.com appear in the name, chaseonline.chase.com and web-da.citibank.com as phishing sites because they misunderstood the URL. Not understanding the URL is the major cause for users to make wrong decisions on four of the five sites after the training (Table 5).

### *3.4.4 User Response to Training materials*

The amount of time that subjects spent on the training materials ranged from 4.30 to 11.00 minutes (mean = 6.99, s.d. = 2.34, var = 5.49). Among the participants tested in the training group, only three users clicked on some of the resources links (two in FTC and one in Mysecurcyberspace) to read more about phishing. All the participants completely read through the Microsoft and FTC materials, while only one subject completely read through the eBay materials and four subjects read through the Mysecurecyberspace materials. On average, subjects spent most of the time on the Microsoft and FTC materials. On average, our subjects spent less than 3.5 minutes on the eBay tutorial. Some of the subjects assumed that there was only one page - while there were 5 pages in the eBay tutorial. Except one subject, all others skimmed through the tutorial materials quickly. When asked to read the training materials, one of the subjects responded, "Is this a real website?" This shows that users would get suspicious about the websites they access just because they had been told that they had to find out which sites were legitimate and which were spoofed.

To summarize, the ability to identify phishing websites improved due to training. Subjects learned that companies do not request sensitive information or login credentials through email or websites. Users were able to unlearn some of their bad strategies, and learn good strategies. However, they still were unable to properly parse longer URLs with sub-domain.

## 4. ANALYSIS OF EXISTING TRAINING MATERIALS

In the previous section we discussed the content and the effectiveness of existing online anti-phishing training materials. Although training materials turned out to be surprisingly effective, in this section we discuss their presentation style as well as strategies to make them even more effective through principles derived from the learning science literature.

Learning science is the body of research that is involved in understanding the way people read, learn and understand to develop knowledge and skills. In general learning science principles has been used to educate people; basic learning

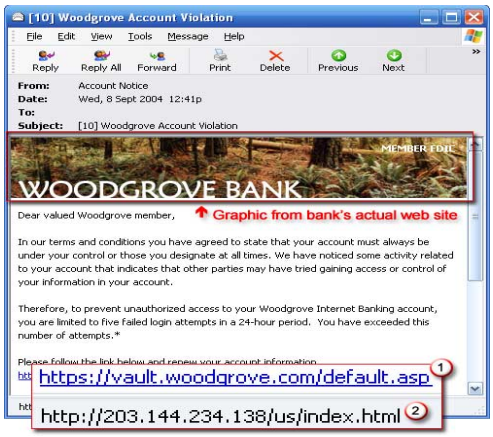


**Figure 2: One of the training images from the online training materials**

principles has been tried and evaluated in the context of e-learning and cognitive tutors. The principles that we used were fundamentally developed in the context of e-learning and cognitive tutors. E-learning is the field where educational materials are developed to be delivered on the Internet or computer. And cognitive tutors are computer-based interactive tutors training system that can adapt to the skill level of participants.

### 4.1 Multimedia Principle

This principle suggests that adding graphics to words can improve learning; in particular it is suggested not to use graphics that decorate the page (*decorative illustrations*) but to use graphics that aid learners to understand the material better (*explanative illustrations)* [8]. From Table 6 we can see that all the online training materials had graphics with words. When we analyzed further we found that some of the training materials use more of decorative than explanative illustrations. One of the training materials had Figure 2 to train users about the deceptive URL's, but did not support an explanation for the image [Microsoft training material]. The accompanying text also did not discuss the "Graphic from the actual website" that was provided in the image. The multimedia principle would postulate to design training materials so that the text and images are presented together, as discussed by Kumaraguru et al. [26].

### 4.2 Contiguity Principle

This principle suggests that placing corresponding words and graphics near each other can improve learning. Studies have shown that integrating text and graphics produce better learning than when they are separated [8]. One common violation that we found in the online materials on phishing was that the visual and explanation text were placed separately. In almost all online materials the violation of this principle was due to scrolling screens and information presented in different pages. We saw eBay providing a better integration of text and graphics than other tutorials. We can also see that instructions provided using Figure 2 did not integrate graphics and text. Table 6 also shows that none of the existing online training materials apply this principle completely.

### 4.3 Personalization and Story Based Instruction Principle

This principle suggests that using conversational style in comparison to formal style improves learning. Also, using

characters and stories can improve learning [8]. Most of the online materials on phishing do not implement this principle. From Table 6 we can see that only the FTC has implemented this principle. Kumaraguru et al. have shown that story-based material that has a character or a coach help users learn better than formal instructional materials [26].

## 4.4 Simplicity

Keeping the instruction simple and short is an essential principle for designing training materials. Research has shown that people learn better when their working load memory is minimized [3]. Other studies have shown that length of the instruction is one of the reasons why people don't read the training materials which are available through security notices. This principle suggests that short training materials will be most effective [27].

## 4.5 Provide Immediate Feedbacks on Errors

This principle suggests that providing immediate feedbacks to users when they make an error can induce better learning [3]. The above principles (Section 4.1 - Section 4.4) discuss how the training materials should be presented, but there is also the question of how to make users read the training materials. Kumaraguru et al. showed that users do not read the security notices that are sent through email. Providing training materials immediately after users fall for phishing emails offers immediate feedback. Online materials available on phishing do not make use of this principle: they are not designed to give feedbacks.

**Table 6: Availability of principles in different training materials / mechanisms; √ is Available, X is Not available, & is partially available**

| Principle | eBay | FTC | Microsoft | Our design |
|---|---|---|---|---|
| Multimedia principle | √ | √ | √ | √ |
| Contiguity | & | X | & | √ |
| Personalization and story | X | √ | X | √ |
| Simplicity | X | √ | & | √ |
| Immediate feedbacks | X | X | X | √ |

# 5. DISCUSSION

In the previous sections we have presented the results of a user study in which users spent 15 minutes reading web-based anti-phishing educational materials. Our results show that users demonstrated significant improvements in their ability to recognize fraudulent websites.

There are two questions that need to be addressed to make training more effective to people. The first question is how to better deliver training materials, so that people will read them. The second question is, given the current training materials, how can they be further improved. In this section we focus on the second question.

Based on the results of the user study, we think three additional things should be taught to users:

□ **Teach users that using the design and content of a website as a cue for determining its legitimacy is a bad strategy.** Phishers can fake the design and the content of websites easily, and our analysis shows that even after the training, users still use the design and the content of the webpage as one of the primary cues. The educational materials we examined do not teach users to avoid this strategy.

□ **Encourage the use of good alternative strategies such as search engines.**

□ **Focus on longer URLs, and some basics of domain name knowledge.** Our study results also show that user's lack of knowledge about URLs and domain names make them still vulnerable for phishing sites whose sub-domain name match the real organization's domain. Furthermore, an increase in awareness without enough knowledge increases the false positive rate. Therefore, we would recommend some materials to cover the basics of the domain name and URLs.

Apart from the content of the training material, how to deliver the material is also important. Our discussion of the learning science principles can help better design and deliver the training materials. Principles such as multimedia, contiguity, personalization, immediate feedback were grounded on learning science principles. We recommend designers to use these principles for designing and delivering of the training materials.

We believe that high level of learning takes place when the instructions relate to users' prior knowledge. We also found that users use specific strategies like "design and content" of the website to make their decision, so training materials have to address these myths in the instructions. From our user study we also found that FTC worked better among the users in providing training. We believe this is due to compliance of FTC to most of the design principles discussed in the framework next to our design.

As in other user studies, there are some limitations in our study also. Participants of our user study are more educated and younger than the general internet user population. So the result may not be generalizable to other groups. Another limitation of our study is that we tested websites, but the best online training materials available are to train user on emails.

# 6. CONCLUSIONS AND FUTURE WORK

In this paper we have presented the user study that we conducted to evaluate the effectiveness of the existing online training materials. We showed that if users are made to read the training materials they perform better in identifying phishing websites. We also showed the different strategies that users use in making their decision and how that changes due to the training. We presented the results from our analysis of the existing training materials based information provided in the training materials. We also analyzed the online training materials using the learning science principles. We provided some suggestions which can be used to develop training materials in the context of phishing.

We have not tested the relative importance of the learning science principles in the context of phishing education; we plan to do this as a future work. We plan to test whether these principles can be generalized to larger area of security. We are currently designing a more interactive training system that can adapt to the skill level of participants. We are also developing an interactive game to train users on identifying phishing URL's and websites.

# ACKNOWLEDGMENTS

We gratefully acknowledge support from National Science Foundation grant number CCF-0524189 entitled "Supporting Trust Decisions." The authors would like to thank all members of

the Supporting Trust Decisions project for their feedback. In particular we would like to thank Yong Woo Rhee and Elizabeth Nunge for conducting some of the user studies that were discussed in this paper. We would also like to thank CyLab, Carnegie Mellon University. The authors would also like to thank Dr. Anne Fay of Eberly Center for Teaching Excellence and Dr. Vincent Aleven for their advice on the learning science aspects of this study.

## REFERENCES

[1] Account Guard. Retrieved Nov 3, 2006, http://pages.ebay.com/ebay_toolbar/.

[2] Adams, A. and M. A. Sasse. Users are not the enemy: why users compromise security mechanisms and hot to take remedial measures. In Lorrie Cranor and Simson Garfinkel (Eds.) *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly. Aug, 2005.

[3] Anderson, J. R., A. T. Corbett, K. R. Koedinger and R. Pelletier. Cognitive Tutors: Lessons Learned. *The Journal of The Learning Science.* 4 (2), 167 – 207. 1995.

[4] Anti-Phishing Resources. Anti-Phishing Working Group. Retrieved on Sept 20, 2006. http://www.antiphishing.org/resources.html.

[5] Anti-Phishing Working Group. Phishing Activity Trends Report. August 2006. Retrieved Nov 9, 2006. http://www.antiphishing.org/reports/apwg_report_August_2006.pdf.

[6] Bransford, J., A. Brown and R. Cocking. How People Learn: Brain, Mind, Experience, and School. Retrieved Nov 18, 2006. http://www.nap.edu/html/howpeople1/.

[7] Camtasia Studio. Retrieved Nov 9, 2006. http://www.techsmith.com/camtasia.asp.

[8] Clark, R. C. and R. E. Mayer. *E-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning*. Pfeiffer. 2002.

[9] Cranor, L., S. Egelman, J. Hong and Y. Shang. Phinding Phish: An Evaluation of Anti-Phishing Toolbars. August 2006. Under review. Retrieved on Sept 27, 2006, http://lorrie.cranor.org/pubs/toolbars.pdf.

[10] Dhamija, R. and J. D. Tygar. The battle against phishing: Dynamic Security Skins. SOUPS '05: Proceedings of the 2005 symposium on Usable privacy and security. 2005. pp. 77 - 88. ACM Press. New York, NY, USA.

[11] Dhamija, R., J. D. Tygar. and M. Hearst, Why Phishing Works, In the Proceedings of the Conference on Human Factors in Computing Systems (CHI2006), 2006, Retrieved Feb 10, 2006, http://www.sims.berkeley.edu/~rachna/papers/why_phishing_works.pdf.

[12] Domain keys. Retrieved Nov 5, 2006. http://en.wikipedia.org/wiki/Domain_keys.

[13] Downs, J., M. Holbrook and L. Cranor. Decision Strategies and Susceptibility to Phishing. In Proceedings of the 2006 Symposium on Usable Privacy and Security, 12 - 14 July, 2006, Pittsburgh, PA.

[14] EarthLink. Retrieved Nov 3, 2006, http://www.earthlink.net/software/free/toolbar/.

[15] eBay. Spoof Email Tutorial. Retrieved March 7, 2006, http://pages.ebay.com/education/spooftutorial/.

[16] Emigh, A. Online Identity Theft: Phishing Technology, Chokepoints and Countermeasures. October, 2005. Retrieved Nov 3, 2006, http://www.antiphishing.org/Phishing-dhs-report.pdf.

[17] Federal Trade Commission. An E-Card for You game. Retrieved Nov 7, 2006, http://www.ftc.gov/bcp/conline/ecards/phishing/index.html.

[18] Federal Trade Commission. How Not to Get Hooked by a Phishing Scam. Retrieved Nov 7, 2006, http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt127.htm .

[19] Fette, I., N. Sadeh and A. Tomasic. Learning to Detect Phishing Emails. June 2006. ISRI Technical report, CMU-ISRI-06-112. Retrieved Sep 2, 2006, http://reports-archive.adm.cs.cmu.edu/anon/isri2006/CMU-ISRI-06-112.pdf.

[20] Ferguson, A. J. Fostering E-Mail Security Awareness: The West Point Carronade. EDUCASE Quarterly. 2005, 1. Retrieved March 22, 2006, http://www.educause.edu/ir/library/pdf/eqm0517.pdf.

[21] Google Toolbar. Google. Retrieved Nov 3, 2006, http://www.google.com/tools/firefox/safebrowsing/.

[22] Herzberg, A., and Gbara, A. 2004. TrustBar: Protecting (even Naïve) Web Users from Spoofing and Phishing Attacks. Cryptology ePrint Archive, Report 2004/155. http://eprint.iacr.org/2004/155.

[23] Jackson, J. W., A. J. Ferguson and M. J. Cobb. Building a University-wide Automated Information Assurance Awareness Exercise: The West Point Carronade. 35th ASEE/IEEE Frontiers in Education Conference. 2005. Retrieved March 22, 2006, http://fie.engrng.pitt.edu/fie2005/papers/1694.pdf.

[24] Jagatic, T.,N. Johnson, M. Jakobsson and F. Menczer. Social Phishing. To appear in the Communications of the ACM. Retrieved March 7, 2006, http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf.

[25] James, L. *Phishing Exposed*. Syngress. 2005.

[26] Kumaraguru, P., A. Acquisti and L. Cranor. Trust modeling for online transactions: A phishing scenario. In the proceedings of Privacy Security Trust, Oct 30 - Nov 1, 2006, Ontario, Canada.

[27] Kumaraguru, P., Y. Rhee, A. Acquisti, L. Cranor, J. Hong and E. Nunge. Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System. Technical Report CMU-CyLab-06-017, CyLab, Carnegie Mellon University, November 2006. Retrieved Nov 10, 2006, http://www.cylab.cmu.edu/default.aspx?id=2253.

[28] Lininger, R. and R. Dean. *Phishing: Cutting the Identity Theft Line*. Wiley, publishing Inc. 2005.

[29] Mail frontier. Mailfrontier Phishing IQ test. Retrieved Sept 2, 2006, http://survey.mailfrontier.com/survey/quiztest.html.

[30] McMillan, R. Consumers to lose $2.8B to phishers in 2006: Gartner says fewer, but bigger, attacks will gain more for criminals. November, 2006, Retrieved Nov 10, 2006, http://www.infoworld.com/archives/emailPrint.jsp?R=printThis&A=/article/06/11/09/HNgartnerphishing_1.html.

[31] Microsoft. Recognizing phishing scams and fraudulent emails. Retrieved Oct 15, 2006. http://www.microsoft.com/athome/security/email/phishing.mspx.

[32] Miller, R. C. and M. Wu. Fighting Phishing at the User Interface, In Lorrie Cranor and Simson Garfinkel (Eds.) *Security and Usability: Designing Secure Systems that People Can Use*. O'Reilly. Aug, 2005.

[33] MySecureCyberspace. Uniform Resource Locator (IRL). Retrieved Oct 15, 2006. http://www.mysecurecyberspace.com/encyclopedia/index/uniform-resource-locator-url-.html.

[34] Netcraft. Retrieved Nov 3, 2006, http://toolbar.netcraft.com/.

[35] New York State Office of Cyber Security & Critical Infrastructure Coordination. Gone Phishing… A Briefing on the Anti-Phishing Exercise Initiative for New York State Government. Aggregate Exercise Results for public release.

[36] Robila, S. A., J. James and W. Ragucci. Don't be a phish: steps in user education. ITICSE '06: Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education. 2006. pp 237-241. New York, NY, USA.

[37] Sender Policy Framework (SPF). Retrieved Nov 5, 2006. http://en.wikipedia.org/wiki/Sender_Policy_Framework.

[38] SpamAssasin. Retrieved Nov 5, 2006, http://spamassassin.apache.org/.

[39] SpoofStick. Retrieved Sept 2, 2006, http://www.spoofstick.com/.

[40] SpoofGuard. Retrieved Sept 2, 2006, http://crypto.stanford.edu/SpoofGuard/.

[41] Whitten, A and J. D. Tygar. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium, August 1999.

[42] Wu, M. Fighting Phishing at the User Interface. MIT PhD. thesis, 2006. Retrieved Nov 5, 2006, http://groups.csail.mit.edu/uid/projects/phishing/minwu-thesis.pdf.

[43] Wu, M., R. C. Miller and Little, G. Web Wallet: Preventing Phishing Attacks By Revealing User Intentions. In Proceedings of the 2006 Symposium On Usable Privacy and Security, 12 - 14 July, 2006, Pittsburgh, PA.

[44] Wu, M., R. C. Miller and S. L. Garfinkel. Do Security Toolbars Actually Prevent Phishing Attacks? In the Conference on Human Factors in Computing Systems (CHI 2006). Retrieved Feb 10, 2006, http://www.simson.net/ref/2006/CHI-security-toolbar-final.pdf.

[45] Ye, Z. and Sean S. Trusted Paths for Browsers. Proceedings of the 11th USENIX Security Symposium. 2002. pp. 263 - 279. USENIX Association. Berkeley, CA, USA.

[46] Yee, K. P. and Sitaker K. PassPet: Convenient Password Management And Phishing Protection. In Proceedings of the 2006 Symposium On Usable Privacy and Security, 12 - 14 July, 2006, Pittsburgh, PA.

[47] ZILLAbar. International Software Systems Solutions, Inc. Retrieved Nov 3, 2006, http://zillabar.com/home.do.