Does Anti-Phishing Training Work?

ABSTRACT

Phishing attacks exploit users' inability to distinguish legitimate websites from fake ones. Strategies for combating phishing include the prevention and detection of phishing scams, tools to help users identify phishing websites, and training users not to fall for phish. While a great deal of effort has been devoted to the first two approaches, less research has been done in the area of training users. Some research even suggests that users cannot be educated. However, previous studies have not evaluated the quality of the training materials used in their user studies. In this paper we present the results of a user study we conducted to test the effectiveness of existing online training materials that teach people how to protect themselves from phishing attacks, and an analysis of those materials through the lenses of principles derived from learning science. We show that existing training materials are surprisingly effective when users actually read them. Their effectiveness can be attributed not just to their ability to raise users' awareness of the phishing problem and make them regard unknown web sites with suspicion, but also to their ability to teach users how to identify fraudulent web sites. We then present our analysis of the training materials based on principles from learning sciences, and provide some suggestions on how to improve training materials based on those principles.

Categories and Subject Descriptors

D.4.6 Security and protection, H.1.2 User / Machine systems, K.6.5 Security and protection education.

General Terms

Security, Human factors

Keywords

Phishing, email, usable privacy and security, user education.

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 had been sent by trusted companies. These emails lead to websites that are similar or virtually identical to the legitimate ones, but lure people into disclosing sensitive information to criminals. Phishers use that information for criminal purposes, such as identity theft [26], [30].

People are vulnerable to phishing attacks because spoofed websites look very similar to legitimate websites. 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], [25].

Researchers have developed several technical approaches to countering phishing attacks, including toolbars, email filters, and verified sender addresses [15]. However, these approaches are not foolproof. In a recent study of 10 anti-phishing tools, only one tool was able to correctly identify over 90% of phishing websites, and that tool also incorrectly identified 42% of legitimate websites as fraudulent [51]. Furthermore, while automated phishing detection is improving, phishers continuously adapt their attack techniques to improve their chances of success. Finally, contextual information known to the recipient may be needed to determine whether some email messages are legitimate. For example, the recipient may recognize whether a message comes from a business where they have made a purchase, or whether an email purportedly from a friend is written in that friend's writing style. In short, while automated detection systems should be used as a first line of defense against phishing, they are unlikely to be perfect, and should be complemented with training strategies to improve the ability of users to recognize fraudulent email and websites.

Some experts have cast doubts on whether user training can be an effective way of preventing users from falling for phishing attacks. Jakob Nielsen, a web usability guru, has argued that educating users about security does not work [36]. Two recent papers tested user education in the form of Phishing IQ tests [3] and documentation for the extended validation feature in IE7 [23] and concluded that user education is ineffective. One security expert was recently quoted in a press report as saying, "[u]ser education is a complete waste of time. It is about as much use as nailing jelly to a wall," [16]. In short, the conventional wisdom seems to be that training users not to fall for phishing attacks is pointless.

Contrary to that view, this paper makes two research contributions. The first contribution is a user study that demonstrates that existing anti-phishing educational materials are surprisingly effective if people actually read them. Our participants spent at most 15 minutes reading anti-phishing educational materials and then demonstrated significant improvements in their ability to recognize fraudulent websites when compared to a control group. The second contribution is an analysis of existing anti-phishing educational material through the lenses of principles derived from learning sciences. Based on those principles, we provide some suggestions on how to further improve training materials.

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 February 2007 was 16,463, compared to 7,197 in December 2005 [6]. Gartner estimates the aggregate financial loss in 2006 due to phishing to be $2.8 billion [32]. Not only do victims lose their money and identities, but they also undergo significant emotional stress [26].

Many solutions to this problem have been proposed. They can be classified into three categories, that we further discuss below: (1) preventing and detecting phishing scams; (2) tools to help users identify phishing websites; and (3) user training.

2.1 Preventing and Detecting Phishing Scams

One way to combat phishing scams is to prevent spoofed emails and web pages from reaching end users. This can be achieved in a number of ways: (1) implementing filters to detect and delete emails automatically at the server [19], [41]; (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 [11], [40].

However, 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. Due to cross-border jurisdictional problems, it is often difficult to shutdown phishing websites quickly: according to APWG, phishing sites stay online on average for 4.5 days [6]. 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, thus raising the need for training users to identify phishing emails and websites.

2.2 Tools to Help Users Identify Phishing Websites

Dozens of tools are available that provide visual indicators to help users identify potential phishing scams. For example, some anti-phishing toolbars display colored icons to indicate the degree of danger of a website, while others provide risk ratings, information about the age and physical location of a website, and other information designed to inform users about potentially fraudulent sites. Some of the toolbars available are Account Guard [1], EarthLink [13], Google Toolbar [21], Netcraft [37], SpoofGuard [44], SpoofStick [43], and Zillabar [52]. In addition, anti-phishing tools are now built into the Microsoft Internet Explorer, Firefox, and Netscape Navigator web browsers.

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 reduce misconceptions about what the system is doing and help users make better decisions. However, studies have shown that users often do not understand or act on the cues provided by toolbars [34], [48]. In addition, a recent study shows that some anti-phishing toolbars are not very accurate, and even the best toolbars may miss over 20% of phishing websites [51].

Other tools, such as PassPet and WebWallet, try to engage users by requiring them to interact actively with the tool before giving out sensitive information [46], [47], [50]. However, even these solutions ultimately rely on the users' ability to make the right decision.

Ye et al. [49] and Dhamija and Tygar [9] 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]. More user studies are needed to assess the effectiveness of these approaches.

In all of the above systems, users are still involved in the decision-making process. These tools aid users in making a decision, but they do not make the decision for users. Studies have shown that users frequently disregard the information presented by anti-phishing tools, often due to inaccurate beliefs about the nature of phishing attacks [48]. This suggests a need to raise users' awareness about phishing and to train users on how to avoid falling for these attacks.

2.3 User Training

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 [5] and businesses [14]. Another approach allows users to take tests on phishing websites and emails. For example, Mail Frontier [31] 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. One study examined the effectiveness of such phishing tests at educating users, and concluded that they are not effective [3]. In another study, Robila et al. trained students in a class room setting, and demonstrated that class discussion and exercises made students more aware of phishing and better at recognizing phishing attacks [39].

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 with students about contextual attacks making use of personal information (also known as *spear-phishing*) [25]; at West Point [20], [23]; and at a New York State Office [38].

In related work, we presented the design and evaluation of an email-based approach to train people to avoid phishing attacks [28], [29]. 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 and contrasted several designs, and found that our interventions were more effective than standard security notices. We have also shown that users can be trained to make better decisions when the training is presented in an interactive game form [41]. However, previous work evaluated the effectiveness of phishing tests, classroom instruction, and email-based training. The study presented in this paper focuses on the effectiveness of existing web-based training materials. This paper presents a study that helps us understand how effective current training materials are, and takes us a step closer to our goal of developing more effective anti-phishing training materials.

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, details about its participants, and its results.

3.1 Study Design

We based the design of our user study on Dhamija et al.'s study of users' ability to identify phishing websites [10]. 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, and to tell us how confident they were in their judgments (on a scale of 1 to 5, where 1 means not confident at all, and 5 means very confident).

We used 20 websites for the study. Ten of them were phishing sites from the APWG database. The other ten were legitimate websites from popular financial institutions and online merchants, as well as random websites. We divided the twenty websites into two groups (A and B), with 5 phishing sites and 5 legitimate sites in each group. In our test, participants were asked to view one group of sites (*pre test*), take a fifteen minute break to complete a task prescribed by the conditions below, and then view the second group of websites (*post test*). We randomized the order of pre test 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 provided in the appendix.) We hosted the phishing websites on a local computer by modifying the host DNS file. Thus, our participants were not actually at risk while we were able to show them realistic phishing sites even after the actual sites had been taken down.

Before the study, we informed participants that they could use any means of their choice to determine the websites' legitimacy, other than calling the institution. We also let participants use a separate web browser, if they elected so. We did not prompt them about how or why this might be useful. Some participants used this other web browser to access a search engine to help determine whether a website was legitimate or not. We used Camtasia Studio [7] to record our participants' computer screens and spoken comments during the study.

We used a between-subjects design with two conditions:

- **Control condition**: In this condition, participants were asked to play a simple computer game (such as solitaire or minesweeper) in the break between the pre and post test.
- **Training material condition**: During the break between pre and post test, participants were asked to read what we judged to be the best web-based educational material on phishing currently available. The rest of the setup was identical to that used for the control condition.

3.2 Training Materials

We compiled and evaluated a list of 24 online anti-phishing training materials to select the materials for our study. Our final selections were eBay's tutorial on spoofed emails [14], Microsoft's Security tutorial on Phishing [33], and Phishing E-card from the U.S. Federal Trade Commission, [17]. We also included a URL tutorial from MySecureCyberspace [35], which is a portal for educating people about security risks and countermeasures on the Internet. In Table 1 we present information about the format of the training materials, their length in words and in pages, the number of graphic examples they included, and what concepts they tried to teach. Almost all the training materials we used were geared primarily towards identifying phishing emails. We were unable to find good training materials that teach users to identify phishing websites. However, many of the techniques and principles in the training materials we selected were generalizable to the website scenario. All the training materials that we selected for the study contained some link to other resources about phishing and personal information security.

Almost all the training materials started with some basic definition of phishing, such as "[c]laiming to be sent by well-known companies, [phishing] emails ask consumers to reply with personal information, such as their credit card number, social security number or account password" [14], or variations 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 about how to avoid falling for such scams. Table 1 also presents the characteristics of the emails discussed in the training material and the suggestions provided therein. Almost all the materials mentioned instructed the reader that "organizations do not request personal information through emails." Finally, the training 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, and reporting them to the FTC.

It is worth noting that not all the tips presented in the training materials were correct: for example the Microsoft tutorial suggests that mouse over the link in the email can show the real URL of the link. However, an attacker can easily obfuscate the correct information with JavaScript.

Table 1: Information about training materials used in our study.

Source	Content format	Length in words	# of printed pages	# of graphic examples	Cues to look for in the email	Suggestions to avoid falling for scams
Microsoft [33]	Webpage	737	3	2	- Urging urgent action - Non-personalized greeting - Requesting personal information through email	- Mouse over the link to see what website it really goes to
eBay [14]	Webpage	1276	5	8	- Urging urgent action - Non-personalized greeting - Requesting personal information through email - Sender email address - Links in the email - Legitimate eBay address versus fake eBay address	- Open a new browser to type in the URL - Never click on the link in an email - How to identify legitimate eBay address
FTC Phishing E-card [17]	Video	N/A	N/A	N/A	- Requesting personal information through email	- Do not provide personal information requested through email
My Secure Cyberspace [35]	Webpage	236	1	0	N / A	N / A

3.3 Participants

We recruited 28 participants for our experiment, and randomly assigned 14 to each condition. To recruit participants, we posted flyers around our campus and posted recruitment messages on university bulletin boards and craigslist.com.

Table 2: Summary of Participant demographics

	Participant Demographics	Correlation
Age	18-34: 93%; 35-44: 7%	+0.25
Gender	Male: 33%; female: 67%	+0.31
Education	high school: 9% ; undergraduates: 48% ; college graduates: 22%; graduate degrees:22%.	+0.20
Race	Asian: 56%; white: 37%; African American: 4%, declined to answer: 7%	+0.33
Years on Internet	< 5 years: 15% 6 - 10 years: 70%, 11-15 years: 15%	+0.11
Hours online (per week)	< 5 hours: 4%, 6-15 hours: 19% 16-30 hours: 50%, 31-50 hours: each week, 12% ; > 51 hours: 15%	-0.12

We screened participants with respect to their knowledge of computers in general, aiming to recruit only participants who could be considered "non-experts." To that goal, 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 phishing-related studies [12], [27].

The participants' demographics are shown in Table 2. In summary, they are younger, more educated, and spend more time online than the average internet population. However, we found no statistically significant correlation between the participants' scores in our experiment and these demographics. Other studies have also found no statistically significant correlation between these demographics and susceptibility to phishing [10], [12].

3.4 Results

In this section we present the results of our study. Our main finding is that subjects in the training condition demonstrated significant improvements in their ability to recognize fraudulent websites.

3.4.1 Effectiveness of Training

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

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 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 was a significant reduction in the false negative rate after the training — from 0.40 to 0.11 (paired t-test: µ1=0.40, µ2=0.11, p = 0.01, DF = 13). There was no statistically significant change in the false negative rate for the control group (paired t-test: µ1=0.47, µ2=0.43, p=0.29, DF=13). Figure 1 shows the comparison of false negatives in both the conditions in pre and post test evaluation.

We tabulated the training group's performance by website. In the appendix we show, for each website, the percentage of correct answers before and after training. We found that users made better decisions on eleven of the twenty sites, did not improve on four sites, and performed worse on five 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: µ1=0.31, µ2=0.41,p=0.12, DF = 13).

We defined total correctness as the ratio of the number of correctly identified websites to the total number of websites shown to the participants. The total correctness for the control group changed from 0.59 in the pre test to 0.61 in the post test (see Figure 2). This change was not statistically significant. The total correctness of the training group changed from 0.65 to 0.74. This change was not strongly statistically significant either (p=0.11): the increase in false positive rate offsets the improvements from finding phishing websites. We discuss the interpretation and significance of this finding further below in Section 3.5.

3.4.2 User Strategies for determining website legitimacy

To gain insight into users' decision process, we asked users to think aloud of their reasons for determining website legitimacy. We recorded these reasons and categorized them into seven categories: design and content, URL, information requested by the website, consistency, search engine, prior knowledge, and security indicators. Table 3 explains these strategies in detail and shows the percentage of websites where they were used by the participants in the control group.

Our categorization extended previous user studies' [9], [12] discussion on strategies for determining website legitimacy. For example, we extended Dhamija et al.'s categorization with contextual clues such as consistency of different pages, and prior knowledge with the site.

During our analysis we reviewed the recordings of each user's session and coded the strategies used at each site into one of these seven categories described above. As noted earlier in Section 3.1, we let participants in both conditions use a second web browser for whatever purpose they desired, and some used this extra web browser to load a search engine.

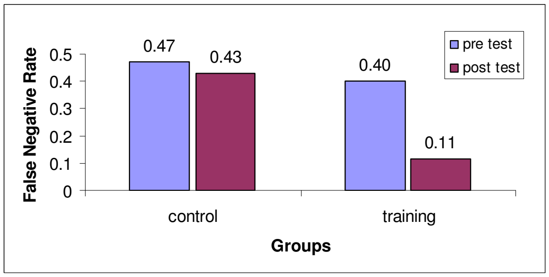


Figure 1: False negative rates. N(control) = 14, N(training) = 14. We found no significant change in the false negative rate for the control group, but did find a statistically significant reduction in the false negative rate for the training group.

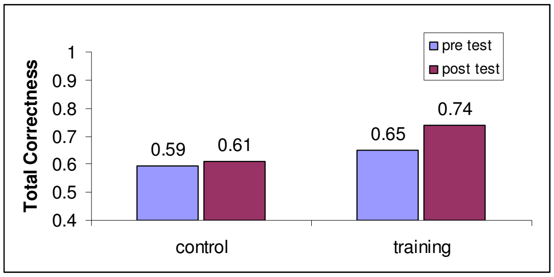


Figure 2: Total correctness. N(control) = 14, N(training) = 14. We found no significant change in the total correctness for the control group, and no significant difference for the training group either.

To ensure we did not bias participants, study administrators only prompted participants to speak about their decisions if they did not do so (which usually only happened at the beginning of a study). At no point did the test administrators provide hints or ask participants to look at certain cues.

We found that forty percent of the time participants in the control group used the design and the content of the website as a strategy to identify phishing sites in the pretest. This result is in agreement with previous research. Also in agreement with previous research, only 3% of the users notice the security indicators on the site (padlock, https). About 1/3 of the time, participants noticed the URL of the site, and use this as a clue. Finally, the choice of strategies in the pretest and post test did not change in a statistically significant way.

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 (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 training than before training. Furthermore, they examined the URLs of the webpage and the amount of information requested more often during post training than pre training. Both of these results are encouraging, as they show 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.

Table 3: Strategies used by the control group: The percentage was calculated across all websites and all participants. A user could use multiple strategies together, which means the percentage adds up more than 100%.

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

Table 4: Percentage change in strategies that participants used

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

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

As for URLs, the Microsoft and eBay training materials teach (1) the correct URL for their respective sites, and (2) some example URLs that phishers use to trick people. However, the training materials do not provide general information about identifying phishing URLs.

For identifying IP-address-based scams (which use IP addresses in the URL instead of a human-readable domain name), participants in the training group seemed to perform quite well, as only one user failed to recognize them (and failed twice on it). This participant's 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 also use deceptive URLs that are hard to detect. In Dhamija et al.'s study, 92% of the users fell for www.bankofthevvest.com (two *v*'s, instead of a *w*). In our study, none of the participants in the training group fell for the deceptive domain halifax-cnline.com (change of "o" to "c" in halifax-online.com) after training. Our participants noticed the typo immediately.

However, our participants had a hard time interpreting longer URLs, especially URLs using sub-domains. For example, many of the participants in the training condition labeled wellsfargo.com.wfcnet.net as legitimate because the word *wellsfargo.com* appeared in the name. Similarly, they labeled chaseonline.chase.com and web-da.citibank.com as phishing sites because they misunderstood the URL. Not understanding the URL was the primary cause of errors after training

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 resource links to read more about phishing (two in the FTC materials and one in MySecureCyberspace). All of our participants in the training condition completely read through the Microsoft and FTC materials, while only one completely read through the eBay materials and four read through the MySecureCyberspace materials. Participants spent most of the time on the Microsoft and FTC materials and less than 3.5 minutes on the eBay tutorial. Some of the participants assumed that the eBay tutorial had only one page, while it actually had five. Except for one person, all others skimmed through the tutorial materials quickly.

Participants generally responded positively to the educational material. When asked to rate the materials in terms of the educational value and fun level, 93% of the participants rated the materials as very or extremely educational, while 29% rated the materials as very or extremely fun. Many of the participants highlighted the FTC E-card animation as the best among the materials.

To summarize, the ability to identify phishing websites improved due to training. Subjects learned that legitimate companies do not request sensitive information or login credentials through email. 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-domains.

3.5 Effect of Training: Learning versus Increasing Alertness

We have observed how the increase in false positive rate in the training group offsets the improvements from finding phishing websites, raising the issue of whether our results are only driven by a heightened alertness among participants in the training condition, as opposed to actual learning. To gain insights into this issue we made use of Signal Detection Theory (SDT).

SDT is a means to quantify the ability to discern between signal (in this case, phishing websites) and noise (in this case, legitimate websites). SDT relies on two measures: sensitivity (d') and criterion (C). Sensitivity refers to how hard or easy it is to detect if a target stimulus is present in the background event. In our user study, sensitivity relates to the ability to distinguish phishing websites (signal) from legitimate websites (noise), expressed as the distance between the mean of the signal and the noise distributions. The larger the distance d', the better is the user at separating the signal from the noise. Criterion is defined as the *tendency* of the users while making a decision. Figure 3 shows that the criterion line divides the graph into four sections that correspond to: true positives, true negatives, false positives and false negatives.

In our user study, tendency relates to the participants' alertness about phishing attacks. When presented with training materials that prime participants about the existence and risks of phishing attacks, the participant may shift the decision criterion (C) to the left, to be more cautious. This might lead to fewer false negatives at the expense of false positives. This would be the effect of the heightened alertness. However, at the same time a participant in the training condition may learn to better identify phishing websites - in which case there will be an increase in d'.

We calculated C and d' for the participants in our user study, and we report the results in Table 5. We found that in the control condition, neither the sensitivity nor the decision criterion changed in a statistically significant manner. However, in the training condition, the sensitivity increased significantly ($p < 0.05$) and the criterion also shifted to the right significantly ($p <0.025$). Thus, training does raise awareness of the phishing problem and tends to make people suspicious. However, it also increases users' ability to distinguish legitimate and fraudulent web sites.

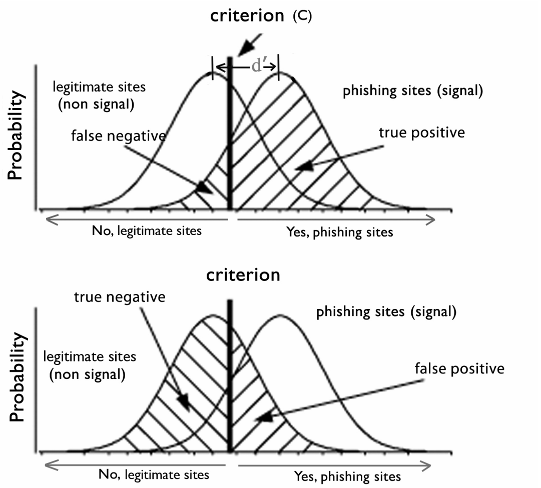


Figure 3: Illustration of Signal Detection Theory (SDT) in our application. We treat legitimate sites as "non signal," and phishing sites as "signal." The sensitivity (d') measures users' ability to discern signal from noise. Criterion (C) measures users' decision tendency. The effects of training could be to a) make the user shift the decision Criterion and thus increasing alertness; b) make users increase sensitivity, separating the two distributions better and thus improving people's ability to distinguish between phishing and legitimate sites; or c) a combination of both.

Table 5: Results from the Signal Detection Theory analysis. The results show that participants gained greater sensitivity with the training condition, meaning that they were better able to distinguish between phishing and legitimate sites. They also become more cautious in their decision. The control condition did not see any significant changes in either sensitivity or criterion.

	Sensitivity (d')			Criterion (C)		
	Pre test	Post test	Delta	Pre test	Post test	Delta
Control Condition	0.48	0.57	0.09	0.17	0.11	0.06
Existing training materials	0.81	1.43	0.62*	0.03	-0.51	-0.54 **

In our application of the SDT, we assumed that false positives are as bad as false negatives. Arguably, false negatives are actually much more dangerous than false positives in this context. Therefore, our analysis puts a lower bound on the effect of training.

4. ANALYSIS OF EXISTING TRAINING MATERIALS

In the previous section we discussed the content and the effectiveness of existing online anti-phishing training materials. The training materials used turned out to be effective (as proven by SDT), yet not perfect (as demonstrated by the undesirable increases in false positives), and therefore improvable. In this section we discuss their presentation style as well as strategies to make them even more effective through principles derived from the learning sciences literature.

Learning science is the body of research that examines the foundations of how people develop knowledge and learn new skills. Learning science researchers have developed learning science instructional design principles that can be applied to impart effective learning [4], [8]. These principles have been applied and evaluated in the context of e-learning and intelligent tutors. Here we discuss some of the learning science instructional design principles that we used in evaluating the existing online training materials.

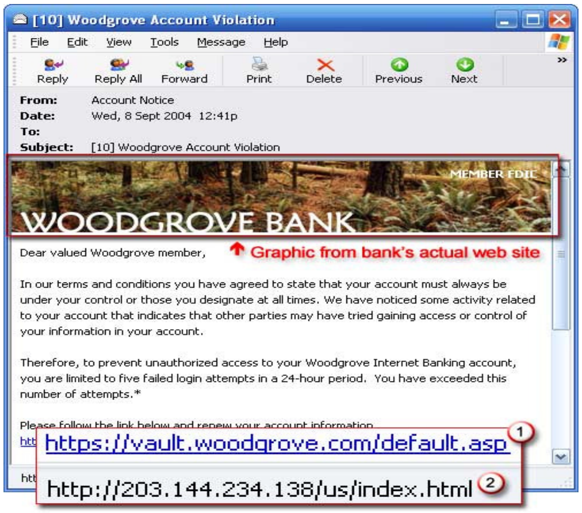
Table 6 presents our analysis of the training materials with respect to the learning science instructional design principles we examined. The rest of this section examines each of these principles in depth.

4.1 Multimedia Principle

This principle states that adding graphics to words can improve learning. In particular, explanative illustrations should be used to help people understand the material better, while the use of purely decorative illustrations should be minimized [8]. Table 7 presents our analysis of all the online training materials, showing that all of them except MySecureCyberspace used images along with text. In examining the training materials, we found that some illustrations were used more for decorative purposes than for explanative purposes. Another issue is that one set of training materials included a graphical example of a deceptive URL (Figure 4) but did not provide an explanation for the image in a caption or in the body of the text [33]. The multimedia principle would suggest designing training materials so that text and images are presented together, as discussed by Kumaraguru et al. [28].

Table 6: Availability of principles in different training materials; √ is available (if it applies this principle consistently through out the material), X is not available, & is partially available (it only applies the principle in some instances).

Principle	eBay	FTC	Microsoft
Multimedia	√	√	√
Contiguity	&	√	&
Personalization and story	X	√	X
Simplicity	X	√	&
Immediate feedback	X	X	X



From: Account Notice
Date: Wed, 8 Sept 2004 12:41p
To:
Subject: [10] Woodgrove Account Violation

Dear valued Woodgrove member,

In our terms and conditions you have agreed to state that your account must always be under your control or those you designate at all times. We have noticed some activity related to your account that indicates that other parties may have tried gaining access or control of your information in your account.

Therefore, to prevent unauthorized access to your Woodgrove Internet Banking account, you are limited to five failed login attempts in a 24-hour period. You have exceeded this number of attempts.*

Please follow the link below and renew your account information.

https://vault.woodgrove.com/default.asp

http://203.144.234.138/us/index.html

Figure 4: One of the training images from the online training materials [33]

4.2 Contiguity Principle

This principle states that placing corresponding words and graphics near each other can improve learning. Studies have shown that integrated text and graphics produce better learning than when they are separated [8]. One common violation that we found in the online training materials was that example images and corresponding text were located far apart from each other.

Table 7: Reasons for post training failures. A primary cause was a misleading or confusing URL.

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%	URL onlineast.bankofamerica.com, users were expecting www.bankofamerica.com
Chase online (real)	100 (4.5)	71 (2.8)	-29%	URL chaseonline.chase.com, user expecting www.chase.com
Citibank (real)	71 (4.0)	42 (4.0)	-29%	URL web-us.da.citibank.com, user expecting www.citibank.com
US Bank (real)	100 (4.2)	57 (4.2)	-43%	URL www4.usbank.com, user expecting www.usbank.com

In the majority of cases, the cause was long web pages that required scrolling. In a few instances, the cause was information being presented on different web pages. Our analysis suggests that eBay did the best job in terms of integrating text and graphics. However, Table 7 also shows that none of the existing online training materials apply this principle consistently.

4.3 Personalization and Story Based Instruction Principle

This principle states that using a conversational style can be more effective for learning than a formal style. Using characters and a story line can also improve learning [8]. Most of the online materials on phishing do not implement this principle. From Table 7, we can see that only the FTC has implemented this principle.

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 [4]. Other studies have shown that length of the instruction is one of the reasons why people do not read the training materials that are available through security notices. This principle suggests that short training materials will be most effective [28].

4.5 Provide Immediate Feedback on Errors

This principle suggests that providing immediate feedback to users when they make an error can induce better learning [4]. Providing training materials immediately after users fall for phishing emails offers immediate feedback [28]. Most online materials do not make use of this principle: they are not designed to give feedback. Materials that include game or test components can provide immediate feedback.

5. DISCUSSION

In the previous sections we presented the results of a user study in which users spent an average of seven minutes reading existing web-based anti-phishing educational materials, and our analysis of those materials based on learning science principles. Our results show that users demonstrated significant improvements in their ability to recognize fraudulent websites after reading the online training materials.

These findings suggest a novel perspective on anti-phishing training when contrasted to the results of previous user studies [3], [23]. For example, Anandpara, et al. suggested that "education only increases awareness, but not real ability." However, we have demonstrated using signal detection theory that both effects were actually in action in our user study. Such difference in the results may be due to the quality of the training materials used in the two studies. Anandpara, et al. employed only a one-page FTC phishing alert to train participants. This alert does not provide clear advice about identifying phishing emails, and does not embodies desirable principles prescribed by learning science.

While our participants were more educated and younger than the general Internet user population (so the results may not be generalizable to other groups), and while we focused on the ability to identify phishing websites (disregarding the phishing email messages that would typically lead to such websites), our results appear to contradict the conventional wisdom that training users to avoid phishing attacks does not work. This wisdom is generally based on the assumptions that (1) computer security concepts are difficult to teach to non-experts; and (2) because security is a secondary task for users, they will not spend time reading training materials. Although we generally agree with these assumptions, we believe the obstacles they pose can be overcome. People can be taught to identify phishing scams without the need for them to understand complicated computer security concepts: we demonstrated that by teaching a few simple concepts to our user study participants, they were able to identify most of the phishing websites. The second obstacle may be more difficult to overcome outside of a situation where people are required to read training materials. However, it seems worthwhile to explore ways of getting people to read them. For example, our group is developing an embedded training system [28] as well as a web-based anti-phishing game.

Further work is needed to determine the most effective way of delivering training materials so that people will read them, as well as ways to improve existing training materials to make them even more effective, and how effective this kind of training can be over a period of time. We also believe that better learning can occur when training materials relate to users' prior knowledge: we found that users adopt counter productive strategies (like examining the design and content of the website to make their decision) - hence, future training materials will also have to address these myths.

6. CONCLUSIONS

In this paper we have presented the results of a user study that evaluated the effectiveness of existing online anti-phishing training materials. We demonstrated that—contrary to popular wisdom—anti-phishing user education can be effective: users get significantly better at identifying phishing websites when they actually read training materials. We also showed the different strategies that users adopt to recognize phishing sites, and how those strategies evolve due to the training.

7. ACKNOWLEDGMENTS

8. REFERENCES

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

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

[3] Anandpara, V., Dingman, A., Jakobsson, M., Liu, D., and Roinestad, H. 2007. Phishing IQ tests measure fear, not ability. In Usable Security (USEC'07). http://usablesecurity.org/papers/anandpara.pdf.

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

[5] Anti-Phishing Working Group (APWG). Retrieved on Sept 20, 2006. http://www.antiphishing.org/.

[6] Anti-Phishing Working Group. Phishing Activity Trends Report. 2007.

http://www.antiphishing.org/reports/apwg_report_february_2007.pdf.

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

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

[9] Dhamija, R. and J. D. Tygar. 2005. The battle against phishing: Dynamic Security Skins. In Proceedings of the 2005 Symposium on Usable Privacy and Security (Pittsburgh, Pennsylvania, July 06 - 08, 2005). SOUPS '05, vol. 93. ACM Press, New York, NY, 77-88. DOI= http://doi.acm.org/10.1145/1073001.1073009.

[10] Dhamija, R., J. D. Tygar. and M. Hearst. 2006. Why phishing works. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006). R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson, Eds. CHI '06. ACM Press, New York, NY, 581-590. DOI= http://doi.acm.org/10.1145/1124772.1124861.

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

[12] Downs, J., M. Holbrook and L. Cranor. 2006. Decision strategies and susceptibility to phishing. In Proceedings of the Second Symposium on Usable Privacy and Security (Pittsburgh, Pennsylvania, July 12 - 14, 2006). SOUPS '06, vol. 149. ACM Press, New York, NY, 79-90. DOI= http://doi.acm.org/10.1145/1143120.1143131.

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

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

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

[16] Evers, J. Security Expert: User education is pointless. Retrieved, Jan 13, 2007, http://news.com.com/2100-7350_3-6125213.html.

[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. 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. 2005. 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, C., Simon, D., Tan, D., and Barth, A. 2007. An evaluation of extended validation and picture-in-picture phishing attacks. In Usable Security (USEC'07). http://usablesecurity.org/papers/jackson.pdf.

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

[25] 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.

[26] James, L. 2005. *Phishing Exposed*. Syngress, Canada.

[27] Anonymous.

[28] Anonymous.

[29] Anonymous. Under review.

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

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

[32] 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.

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

[34] Miller, R. C. and M. Wu. 2005. 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.

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

[36] Nielsen, J. 2004. User education is not the answer to security problems. http://www.useit.com/alertbox/20041025.html.

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

[38] 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.

[39] Robila, S. A. and J. W. Ragucci. 2006. 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.

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

[41] Anonymous.

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

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

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

[45] Whitten, A and J. D. Tygar. 1999. Why Johnny Can't Encrypt: A Usability Evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium. http://www.cs.berkeley.edu/~tygar/papers/Why_Johnny_Cant_Encrypt/USENIX.pdf.

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

[47] Wu, M., R. C. Miller and Little, G. 2006. Web Wallet: Preventing Phishing Attacks By Revealing User Intentions. In Proceedings of the Second Symposium on Usable Privacy and Security (Pittsburgh, Pennsylvania, July 12 - 14, 2006). SOUPS '06, vol. 149. ACM Press, New York, NY, 79-90. DOI= http://doi.acm.org/10.1145/1143120.1143133.

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

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

[50] Yee, K. P. and Sitaker K. PassPet: Convenient Password Management And Phishing Protection. In Proceedings of the Second Symposium on Usable Privacy and Security (Pittsburgh, Pennsylvania, July 12 - 14, 2006). SOUPS '06, vol. 149. ACM Press, New York, NY, 79-90. DOI= http://doi.acm.org/10.1145/1143120.1143126.

[51] Zhang, Y., S. Egelman, L. Cranor, and J. Hong. 2007. Phinding Phish: Evaluating Anti-Phishing Tools. In Proceedings of the 14th Annual Network and Distributed System Security Symposium (NDSS 2007), San Diego, CA, 28 February -2 March, 2007.

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

Appendix: List of websites tested and user performance

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 much information	14% (4.0)	71% (4.4)	+57%
PNC Bank	Spoof	Bank account update; pop-up window over the real PNC Bank web site; 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 has no resemblance with the bank.	42 (3.3)	85 (4.8)	+ 43%
HSBC	Spoof	Internet banking login page; layered information request; IP address URL	50 (4.0)	85 (4.8)	+ 35%
Chase Student	Real	Primitive looking page with few graphics and links	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; IP address URL	85 (4.1)	100 (4.4)	+15%
AOL	Spoof	AOL account update, deceptive domain myaol.com	85 (4.0)	100 (4.7)	+15%
Halifax Bank	Spoof	Halifax bank login page; deceptive domain halifax-cnline.co.uk.	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; IP address URL	85 (4.8)	85 (4.8)	0%
Wellsfargo bank	Spoof	Faked Wellsfargo home page; layered information request; sub domain deception with URL online.wellsfargo.wfosec.net	71 (4.0)	71 (3.8)	0%
Desjardins	Real	Account login page; unfamiliar foreign bank	57 (3.0)	57 (3.5)	0%
Card Financials Online	Real	Card Financial Online (part of MBNA); domain name has nothing to do with MBNA.	42 (4.3)	28 (3.5)	-14%
Bank of America	Real	Bank of America home page; URL: onlineast.bankofamerica.com	83 (4.2)	57 (3.7)	-26%
Chase online	Real	Online banking login page; URL: chaseonline.chase.com	100 (4.5)	71 (2.8)	-29%
Citibank	Real	Citibank login Page; URL: web-da.us.citibank.com	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%

appendix: Percentage of correct answers for the training group before and after training