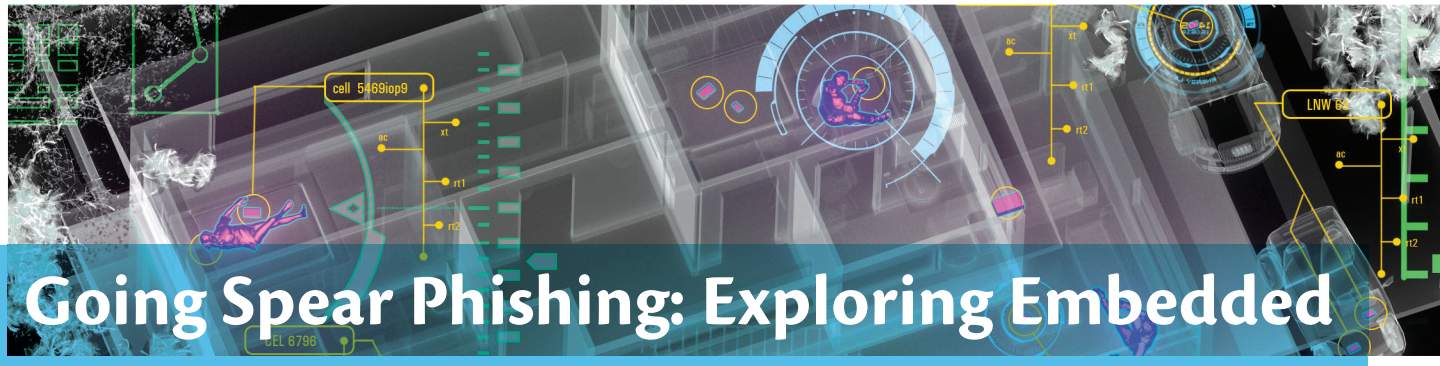

Going Spear Phishing: Exploring Embedded Training and Awareness

Deanna D. Caputo | MITRE
Shari Lawrence Pfleeger | I3P Dartmouth College
Jesse D. Freeman | MITRE
M. Eric Johnson | Vanderbilt University

To explore the effectiveness of embedded training, researchers conducted a large-scale experiment that tracked employees' reactions to spear phishing emails and immediate training activities. The results showed that many participants did not read the training, and clicked either all links or none across three trials.

Our adversaries have long understood the ease, effectiveness, and efficiency of *social engineering*: attacking technology by directly compromising the user.[1] Using email as a vector has proven particularly successful. As Joshua Perrymon noted,[2]

> Email-based attacks are probably one of the most effective in today's hacker bag of tricks. … The problem lies in a directed, under-the-radar, spear-phishing attack—the type where the attacker spends time to understand the target, create an effective spoofed email and phishing site, [and] then attacks.

The Anti-Phishing Working Group's (APWG) 2011 analysis reported, "In the latter months of 2010, APWG witnessed an increase in so-called 'spear-phishing' attacks. … This trend is accelerating in 2011, and is responsible for some high-profile corporate data breaches."[3] Spear phishing is a form of cyberattack attempting to infiltrate a system or organization for cybercrime or espionage purposes. Cyberattackers find inside information specifically relevant to users and craft fake email messages, usually impersonating well-known companies, trusted relationships, or contexts. For the attack to succeed, the user must take action. For example, by clicking a link in an email message, users could install malicious software on their system, or they might be asked to provide personal information, such as a username, password, or credit card number.

Spear phishing volume climbed dramatically from 2009 to 2011.[4] Indeed, "[t]hese spear phishing attacks are a key part of the Advanced Persistent Threats (APTs) that companies and governments are facing today. Responders, industries and governments engaging these threats need new ways to detect them, measure their proliferation, and defend against them."[5]

A recent Cisco report shows why spear phishing need not occur on a massive scale to be effective.[6] Table 1 illustrates that the financial benefit of mass attacks dropped by more than half during the same time that the financial benefit of spear phishing attacks tripled. Table 2 compares the results of

 Copublished by the IEEE Computer and Reliability Societies

Table 1. Total cybercrime benefit.

Attack	Benefit one year ago (in millions of dollars)	Current benefit (in millions of dollars)
Mass attacks	$1,050	$500
Spear phishing attacks	$50	$150
Targeted attacks	Varies	Varies
Total	$1,100	$650

Table 2. Relative cost and benefit from spear phishing campaigns.

Example of a typical campaign	Mass phishing attack (single campaign)	Spear phishing attack (single campaign)
Total messages sent	1,000,000	1,000
Block rate	99%	99%
Open rate	3%	7%
Click-through rate	5%	50%
Conversion rate	50%	50%
Victims	8	2
Value per victim	$2,000	$80,000
Total value from campaign	$16,000	$160,000
Total cost for campaign	$2,000	$10,000
Total profit from campaign	$14,000	$150,000

mass phishing versus targeted spear phishing in more depth. Using far fewer emails than mass phishing attackers, spear phishing attackers need only a quarter of the victims to click to yield more than 10 times the financial benefit.

During a July 2010 workshop held in Washington, DC, industry participants identified spear phishing as one of the top problems companies face.[7] Consequently, the Institute for Information Infrastructure Protection (I3P) initiated research that addressed ways to accomplish two key industry goals: reduce the vulnerability to spear phishing by training employees to recognize it, and improve an organization's "security culture" by encouraging employees to report spear phishing incidents. This article describes the study conducted, the reasoning underlying the methodology, and the surprising results.

Standing on the Shoulders of Others

Our study began with a hypothesis suggested by previous research findings: that embedded training offers an effective way to increase security awareness and reduce the dangers posed by spear phishing. Most organizations provide security awareness training to their employees. Often performed only once a year, the training describes threats to the organization as well as steps each employee can take to address them. Research shows that such training is mostly ineffective because information isn't recalled and practiced enough.

As an alternative approach, David Bank described attempts to make employees aware of spear phishing using "gotcha" exercises:[8]

- In 2004, more than 500 cadets at the US Military Academy were spear phished, and 80 percent clicked the embedded link and were subsequently warned about the risks.
- In 2005, nearly 10,000 New York State employees were phished. Seventeen percent clicked the link, and 15 percent began to enter personal information on the site before being "caught." Those who clicked received a "slap on the wrist" from the state's chief information officer—reprimands and a message about the dangers of phishing. Two months later, the same group was phished again; 14 percent still clicked, but only 8 percent entered personal information.

For some of these and similar studies, the organizations later offered additional training to test understanding and help employees identify the best cues for recognizing a phishing attack. After such training, fewer people clicked the next time they were phished, but the question remained—would these changes last over time?

Jason Hong reviewed the considerable body of

literature that explores the effectiveness of awareness training in reducing both phishing and spear phishing.[9] He found that many previous studies had significant methodological drawbacks. For example, many studies involved students placed in unrealistic situations, so it wasn't clear how the results would apply in an industrial setting. For instance, Ponnurangam Kumaraguru and his colleagues found that embedded training worked better than standard security notices, but their studies involved students role-playing and imagining what they would do if the message appeared in their email inbox.[10] In addition, many studies did not use controlled sampling techniques, so it was unclear which population was represented in the study. For instance, Kumaraguru and his colleagues used a convenience sample, obtained by posting flyers and asking for volunteers.

One of the most valuable elements of scientific research is the emphasis on replication of findings before achieving high confidence in data results. Most research findings are expected to be replicated by different researchers using varied samples and contexts. We chose to replicate aspects of Kumaraguru and his colleagues' embedded antiphishing training work.[10,11] Our work in this article closely follows theirs, with some differences. In their 2007 paper, they found that providing training text in the form of a comic strip was more compelling for their student sample. However, we chose to use their two-column text training materials because senior members of the corporation didn't feel that a comic strip was the appropriate format for corporate employee training. In the present study, we tested the retention effects over approximately 90 days instead of 28 days, because training every month in a corporate setting isn't manageable. In addition, we tripled the sample size and used stratified sampling to determine whether training effects would carry across all levels of staff in a corporate setting. Lastly, our study focused on APT techniques, in which adversaries use spear phishes that don't ask for personally identifying information (which can lead to greater suspicion), but instead sought to have people click links that released malicious code that gained entry to the users' network.

Our goal was to explore the embedded training's effectiveness by using more rigorous methods that included

- controlled sampling,
- realistic situations,
- scientific and documented processes,
- clearly stated hypotheses,
- data analysis to support evidence-based cybersecurity decisions, and
- data, tools, and techniques made available for others to use.

We also applied behavioral science principles that were likely to make training more effective. For example, Amos Tversky and Daniel Kahneman showed that framing a situation in two different but logically equivalent ways can lead decision makers to make very different choices.[12] Thus, our experimental methodology used portions of materials developed in previous studies but incorporated additional conditions and applied them in an industrial setting using careful sampling and control of significant variables.

Our approach also built on previous behavioral science findings about embedded training. An embedded training scheme that combines testing employees' behavior in their normal work environments with instant corrective performance feedback produces more lasting change to attitudes and behaviors. This approach, proven successful in situations ranging from military training operations to learning centers, can likely aid in changing employees' security-relevant behaviors. Although specific messages might change, the general principles for testing different training options' effectiveness are transferable to a wide variety of desired behaviors. The remainder of this article describes our approach and findings.

Setting the Hook

The better a study can identify and control variables, the more effectively its results can be extrapolated to the wider world. For this reason, we performed a controlled experiment in an actual industrial environment using a very large sample. With the cooperation of senior management, we worked within the organization's network and email system so we could send and track spear phishing emails, and then monitor each employee's reactions to them.

Methods

We hypothesized that if users are provided with training immediately following an error in judgment, they will be less likely to make the same error when presented again with a similar judgment. In this case, we wanted to see two changes in behavior based on the training: a lower rate of clicking spear phishing links and an increase in reporting suspicious emails. Both these steps reflect an improved security culture at the organization—the ultimate goal.

We drew participants from a medium-sized Washington, DC-based organization that uses email as a common communication medium. The combination of large sample size and stratified sampling ensured not only that results represented all types of workers but also that the results could be generalized to similar organizations. We categorized each of 6,000 workers by cumulative job experience (thereby defining the strata), then randomly sampled from each stratum

proportionally. All participants (n = 1,500) were notified by email that they had been selected to take part in a security study; they would receive further information unless they opted out in an email reply. The study was performed with management permission, and participants were debriefed about the study's true nature after the study was complete.

We removed 141 participants from the sample for various reasons. These included opting out, leaving the organization before study completion, technical problems with email, or participation in preparatory study activities, such as a pilot deployment of the spear phishing emails and training pages.

The final sample included 1,359 participants, each of whom was randomly assigned to one of five conditions: one control group or four treatment groups. All participants received the organization's annual information security training, and all received the same spear phishing emails. Control group participants received no training if they clicked the link in any of the three spear phishing emails sent to them; instead, they were informed only that they had received a spear phish. By contrast, treatment group participants who clicked the link received a Web-based training page that was linked to the spear phishing email; using color-coding, it pointed out items in the spear phishing email that should have made the reader suspicious. The four treatments corresponded to four different experimental training pages, each designed to test the impact of message framing (gain versus loss) and the nature of the impact (self versus other). The four experimental groups were

- gain-framed and individually focused embedded training (you kept yourself from harm),
- loss-framed and individually focused embedded training (you put yourself at risk),
- gain-framed and other-focused embedded training (you kept your coworkers from harm), and
- loss-framed and other-focused embedded training (you put your coworkers at risk).

To test the persistence of the training's effect, each of the spear phishing emails had to be realistic and equally difficult to recognize as suspect. Thus, we used 50 personnel (excluded from our sample) to perform an initial test of six carefully crafted spear phishing emails tailored to tempt them. The recipients were told the following:

> Some of these emails are real emails and some are fake emails. Your task is to rate each of these emails on the following five-point scale:
>
> 1 = "Extremely Fake"
> 2 = "More Fake than Real"
> 3 = "Somewhat Fake/Real"
> 4 = "More Real than Fake"
> 5 = "Extremely Real"

Each of the three emails chosen for the experiment received mean ratings of 3.2; the other three emails in the pilot received much higher or much lower scores. We used the three emails that pilot testing determined were equally difficult to detect as malicious so that we could increase our confidence that any data findings would be due to training effects and not the quality of the emails used.

Materials

By emulating typical adversary actions, the emails aimed to entice participants to click a suspicious link. Each spear phishing email contained five elements that should have helped participants identify it as a spear phishing attempt (see Figure 1):

- mismatched name and address in the From field;
- errors such as a misspelling, incorrect grammar, or odd spacing;
- encouragement to take immediate action;
- mismatch between link text and the link address displayed when hovering the mouse over it; and
- intuition—an overall feeling that something isn't right (for example, you weren't expecting it).

Participants who clicked the link were sent to pages containing information displayed as shown in Figures 1 and 2: a training webpage (treatment condition—spear phishing notification plus training) or a no-training webpage (control condition—spear phishing notification only). These materials were based on the student studies by Kumaraguru and colleagues.[10] All the training pages displayed to the treatment groups explained spear phishing, illustrated how to recognize the spear phishing attempt from the actual email, and outlined how to avoid becoming a victim of spear phishing in the future. Each treatment employed a differently framed approach for communicating the training information. The no-training page informed participants only that they had clicked a spear phishing email, with no further guidance about recognizing or avoiding further attempts. In addition, we purchased real Web domains and developed software to customize the training pages, create senders, track link clicks, and manage email logs.

The Spear Phishing Process

We sent three spear phishing trials to our stratified sample. The email in Trial 1 (February 2011) ostensibly came from the company's timecard system. It asked the recipients to acknowledge changes made to their timecards by someone else and provided a

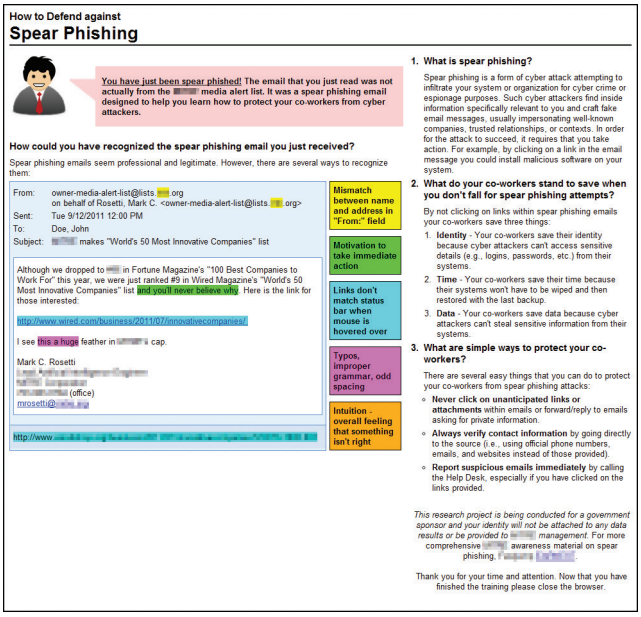

Figure 1. The treatment group training page shows users why they should have been suspicious and not clicked a link. Each colored area highlights one of the five indicators, such as mismatched names or spelling errors.

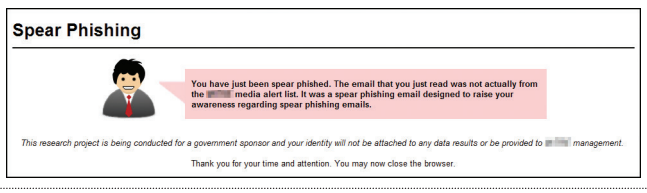

Figure 2. The control group notification simply alerts users, providing no training.

link for doing so. Although pilot testing had explored the relative difficulty of detecting the spear phishing attempts in each email, the actual study showed that the timecard email was significantly more difficult to detect than expected, perhaps because it appeared to come from inside the organization. After consultation with corporate management, we redesigned the subsequent trials with emails appearing to come from an outside entity. Thus, we created three new emails and did another pilot test.

The email used in Trial 2, sent in May 2011, contained a link to an apparent *Washington Post* article reporting on the government ranking of similar companies. Trial 3, sent in September 2011, appeared to come from an employee posting a message on an internal corporate listserv that didn't really exist. The email announced the company's apparent ranking in *Wired*'s World's 50 Most Innovative Companies and provided a link to the online article.

During each trial, we recorded clicks on the spear phishing links and any reports made to the organization's information security office or help desk (via email or telephone calls).

After the email trials were complete, all participants received a debriefing email describing the study hypothesis, design, and results. At that time, participants were asked to consent to a follow-up interview designed to help us better understand their experiences and perceptions of the emails, the particular training page they saw, and the embedded training approach. Of the 1,359 participants, 327 agreed to be interviewed. Within this group, we focused on three subgroups to help us understand possible strategies for mitigating the threat:

- all-clickers (those who clicked in all three trials),
- nonclickers (those who clicked in none of the trials), and
- two-clickers (those who clicked in the first two trials but not in the last).

Results

In Trial 1, because no differential training had yet occurred, participants demonstrated their baseline ability to identify a spear phishing attempt. Table 3 shows no statistically significant differences in click rates between conditions. Nevertheless, Trial 1 revealed a very high overall click rate compared to previous studies. Whereas other studies reported an average click rate of 30 to 35 percent, ours showed almost twice that rate. This difference might reflect the difficulty of detecting the spear phishing elements in our particular messages where no personal information is requested and clicking a link is enough for the adversary—something almost impossible to compare across studies.

Trial 2 let us see which of the 813 participants who clicked the link in Trial 1 would click again, whether they had received training or not. Table 4 shows that training had no significant effects as compared to the control condition. In addition, no differences in performance appeared among the treatment conditions, so framing had no effect.

Table 5 provides a different perspective by comparing click rates for participants who clicked and received training (training), participants who clicked and received no training but were told that they had been spear phished (awareness), and participants who did not click in Trial 1 (nonclickers). In Trial 2, the nonclickers of Trial 1 were again significantly less likely to click than the others.

Although there were no differences between conditions in Trial 2, the overall click rate was significantly

Table 3. Summary of Trial 1 clicks by condition.

Condition	No. of participants in each group	Trial 1 clicks
Other gain	266	156 (59%)
Other loss	263	156 (59%)
Individual gain	274	153 (56%)
Individual loss	275	175 (67%)
Control	281	173 (62%)
Total	1,359	813 (60%)

Table 4. Summary of Trial 2 clicks by condition.

Condition	No. of participants in each group	Trial 2 clicks
Other gain	156	53 (34%)*
Other loss	156	59 (38%)*
Individual gain	153	46 (30%)*
Individual loss	175	59 (34%)*
Control	173	63 (36%)*
No Trial 1 click	546	154 (28%)*
Trial 1 clicker total	813	280 (34%)

**Chi-Square (5) = 8.378, p = .137*

Table 5. Summary of Trial 2 clicks by experience.

Condition	No. of participants in each group	Trial 2 clicks
Trial 1 training	640	217 (34%)*
Trial 1 awareness	173	63 (36%)*
Trial 1 nonclickers	546	154 (28%)*
Total	1,359	434 (32%)

**Chi-Square (2) = 6.237, p = .044*

lower: 34 versus 60 percent. This difference in click rate suggests a difference in difficulty between an apparently internal email and an apparently external one.

Four possible reasons might explain Trial 2's nonsignificant findings:

- Our hypothesis might be incorrect; perhaps embedded training isn't an effective means for improving identification of spear phishes.
- Repetition might be required to achieve the training effects; participants might need repeated exposure to the training before their behavior is changed.
- The training page was ineffective; participants might have perceived the information as not credible, relevant, or interesting.
- Participants didn't read the training page so they didn't actually receive any training.

Each reason could explain why treatment participants performed no differently than control participants. The fourth reason is particularly intriguing. Might participants have ignored the training page because they misinterpreted its purpose and found it threatening? In reports to the information security office and help desk, participants expressed concern that the training webpage might have been part of the spear phishing attempt; consequently, many participants closed the training page immediately without reading any text on the page. Several participant comments reflect this possibility:

- "[I] clicked on it inadvertently without thinking and exited Explorer without reading the link."
- "I just opened this. Then followed link like an idiot. Then killed the process using Task Manager. Please advise as what to do."

Table 6. Summary of Trial 3 clicks by experience.

Condition	No. of participants in each group	Trial 3 clicks
Clicked in Trial 1 only	533	199 (37%)
Clicked in Trial 2 only	154	58 (38%)
Clicked in Trial 1 and Trial 2	280	146 (52%)*
Did not click in Trial 1 or Trial 2	392	97 (25%)*
Total	1,359	500 (37%)

**Chi-Square (3) = 52.955, p = .001*

Table 7. Trial 3 median viewing times by condition.

Condition	Seconds
Group gain	29.25
Group loss	26.75
Individual gain	33.50
Individual loss	37.50
Control	14.00

- "I just got this email and clicked on the link. A webpage came up, but it seems suspicious."
- "I can't believe I actually clicked on the link! Let me know if there's something I need to do to ensure my laptop isn't infected, or if this is just a prank."

To determine whether participants read the training page, we recorded the length of time that the training page was open for each participant who clicked the email link in Trial 3. The amount of time a webpage was open doesn't tell us whether a participant was actually reading the training, but it does indicate whether the page was open long enough to be read completely.

Table 6, displaying Trial 3 clicks based on performance during previous trials, reveals two statistically significant findings. Participants who clicked in the first two trials were more likely to click in the last trial, and participants who didn't click in the first two trials were less likely to click in the last trial. These findings told us little about the embedded training, but they suggest that we should study three groups of people to better understand spear phishing clicking behaviors:

- all-clickers—people who click regardless of previous training, awareness, or information about spear phishing;
- nonclickers—people who don't click links in spear phishing emails and might simply not click links at all; and
- the rest—people who exhibit no consistent clicking behavior, for reasons we don't yet understand.

Training might be most effective for this last group.

Table 7 shows the median viewing time by condition for Trial 3. There was no statistical significance between experimental conditions, but as we expected, the control condition spent significantly less time viewing the four-sentence notification page.

To explore the significance of viewing time for the training page, we asked small samples of colleagues who had never seen the page to scan the title and headers but not to read the content, to read half a page, or to read the entire page while we timed them. This small test suggested that participants spending less than 16 seconds on the training page had looked only at the header information. Similarly, participants spending 16 to 60 seconds either read only one column or skimmed the content quickly. Participants who spent more than one minute on the training page could have read all the content. Figure 3 compares these test times with Trial 3 viewing times for the control page.

We reached four main conclusions:

- The training's framing had no significant effect on the likelihood that a participant would click a subsequent spear phishing email.
- It's unlikely that many participants read all the information presented on the training page; therefore, they weren't actually trained.
- Many participants were all-clickers (11 percent) or nonclickers (22 percent), regardless of whether and what kind of training they received.
- Employees who clicked an initial spear phishing email were more likely to click subsequent spear phishing emails; similarly, those who didn't click an initial spear phishing email were less likely to click subsequent spear phishing emails.

Diving Deeper

From the 27 percent of participants who agreed to be interviewed after the study, we selected interviewees based on the number of times they clicked. We considered three groups: the 31 all-clickers (who clicked links

in all three trials), the 20 nonclickers (who clicked no links in all three trials), and the 17 two-clickers (who clicked links in the first two trials but not the third). We wanted to discover what nonclickers did right (are they more security conscious, are they simply lucky, or do they simply not pay attention to their emails?), what the clickers did wrong, and whether getting caught clicking twice increased awareness in the last trial. The primary goal was to determine how training could help any or all of them.

Using the follow-up questions shown in the sidebar, we conducted semistructured interviews, allowing respondents to answer in their own voices and to discuss related topics. When interviewees deviated from the questions, their responses were assigned to a relevant section of the transcript for later analysis. We employed content analysis to identify the major themes and topics in each transcript. Because the sample was small and not representative, we performed no quantitative analysis. However, we identified trends based on the relative proportions of responses that exhibited a particular viewpoint or opinion. Where we saw little agreement, we instead sought to determine major themes.

The majority of interviewees believed that the kind of embedded security awareness and training employed in the study is more effective than the once-yearly mandatory training they receive. However, they admitted that they ignored or skimmed the bulleted text on the right-hand side of the training page presented after they clicked a spear phish. This finding could explain the lack of significant difference among study treatments, because the majority of the framing words were in the text on the right. Interviewees seemed to be security aware, showing familiarity with standard techniques for identifying a trusted email, such as knowing the sender and expecting the communication. Thus, although respondents knew some proper cyber behavior, in practice, they didn't apply it appropriately.

When asked to recall clicking events, almost all interviewees who had clicked an email remembered at least one, most often the email in Trial 3. When the system notified them that they had clicked inappropriately, most remembered feelings of fear or shame, plus relief that they hadn't actually infected their computers.

Why did they click? The two most cited reasons were an interest in the subject matter and lack of careful attention. Indeed, many interviewees said they were working quickly, distracted by other tasks, or simply trying to deal quickly with email volume. If an email looked interesting, they would click a link without thinking. Two-clickers had no clear memory of identifying an email as suspicious. Nevertheless, some interviewees reported being more aware of spear phishing after the study; they

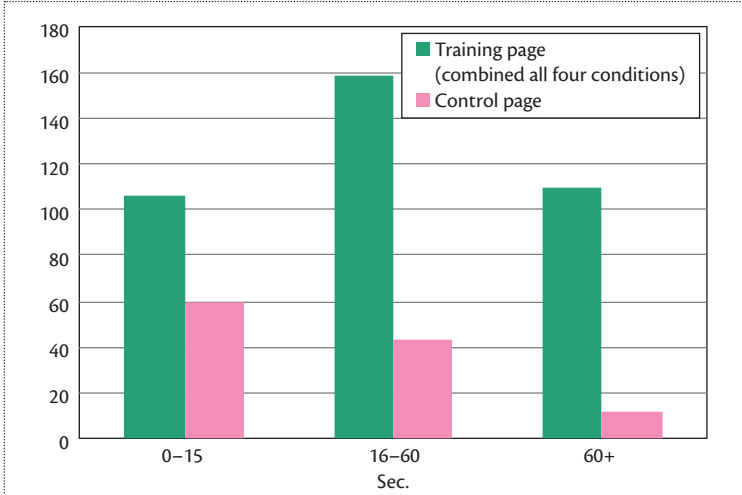

Figure 3. Trial 3 viewing times. Readers who kept the page open for less than 16 seconds had no time to read it. Those who kept the page open between 16 and 60 seconds could skim or read one column. Only those who had the page open for more than 60 seconds had time to read it completely.

subsequently sent other suspicious emails to the organization's information security office.

Most nonclickers remembered none of the emails. In fact, they probably deleted the emails immediately without reading them or generally don't click links. Instead, they seek relevant information using their own Web searches. For example, the majority of nonclickers who remembered the timecard email went to the timecard website and checked their required actions directly. They recounted similar examples, such as bank notification emails, that spurred them to visit a site directly instead of using a supplied link.

Almost all interviewees, regardless of group, mentioned strategies for determining whether to click a link. These strategies, similar to those on the training page, included checking the sender, determining whether the sender is known, and if not, verifying the sender's existence through the company's internal database.

Many participants misinterpreted the training page when it popped up immediately after clicking a spear phish and feared it represented a further attempt to entrap them. We asked about other warning techniques that might be more helpful, such as an icon indicating suspicious emails. Almost all liked the idea of an assistive icon but expressed concern that many false positives would, after a while, result in lack of attention to it.

Most all-clickers remembered a page popping up when they clicked, but almost none remembered specifics. Some remembered only the notification message but not the training page; they thought they were in a control group because they didn't remember any training. The majority of all-clickers felt shock, surprise, and

Follow-Up Questions

1. Do you remember clicking any links contained in the spear phishing emails? If so, why do you remember the experience?
2. Do you recall finding the emails suspicious or unusual?
3. Why do you think you clicked or did not click the links?
4. How do you decide whether to click a link or open an attachment?
5. How would you feel about an icon (such as a question mark) that would identify possible suspicious emails?
6. What happened after you clicked?
   a. If nothing, do you remember seeing anything pop up?
   b. If something, what were your reactions or overall impressions?
   c. Do you recall your initial reaction to the training page: its presence, quality, or content?
7. How much time would you estimate you spent reading the training page?
8. Now looking at the training page, what do you think of the page? What did you like about it? What did you not like about it? Did you learn something from it?
9. What do you think of the five things to look for in emails? How intuitive and/or easy are they?
10. Some people didn't read the training page, or didn't read it completely. How can we make the page more credible, nonthreatening, and informative for you?
11. What would make the training page more interesting to you?
12. How many security trainings (corporate, government, and so forth) have you taken in the last year that offered information on spear phishing?
13. How do you think these trainings impact your behavior?
14. How effective do think [the company's] firewall is at protecting you from spear phishing and other attacks? If effective, do you think it's so good that you don't worry about what you click?
15. When you click a link and get [the company's] Uncategorized URLs notice, what do you do next?
16. Are you aware of [the company's] security awareness programs?
17. In what format and how often would you like [the company] to remind you of security risks?
18. What information would you like to know in these communications?
19. Would you like direct feedback on your security-related behavior from [the company's] information security office (for example, a report assessing your "cyberactivity")?
    a. If yes, how and how often?
    b. If no, why not?
20. Is there anything else you want to share?

even anger at themselves; some also were angry at the study for having deceived them. Many interviewees expressed disappointment in themselves and felt they should have known better by Trial 3.

Some interviewees considered the training page's image too "cartoonish" and insulting, but others thought it encouraged attention to the page. The majority of emotional responses came from the all-clickers who, unsurprisingly, had more visceral reactions to the training than the nonclickers. Indeed, many nonclickers ignored the images and instead discussed the sample email, the colored text boxes, and their overall positive opinions of the design.

During the interviews, most respondents preferred the training page that reflected the loss-framed, individual-focused condition. In addition, several respondents believed that loss should be described more specifically, such as lost time for cleaning an infected laptop or for mandatory remedial training.

Although interviewees liked the color-coding aligning the content of the phishing email and the warning signs, they offered a few constructive comments. First, because red, yellow, and green (and sometimes blue) have very distinct meanings in other contexts, many employees might associate the colors with traditional meanings (bad, caution, and good). Therefore, a different color palette might be better for the training page. Alternatively, the training page could be color-coded with red as dangerous, and green or blue as benign. Some participants found the orange box (Intuition) confusing, because it wasn't tied to any specific item in the spear phishing email.

Most interviewees said they ignored the text and paid attention only to the email and colored text boxes. Commenting that the training page was too dense with text, many suggested replacing text with a link to more information—surprising, because those reading the page must already have clicked a bad link! Good visual design of training is very important and needs to be well tested.

Previous training might also have affected participants' responses. Most interviewees had taken only the company's annual training, plus one or two courses required by their projects. Although they assumed that spear phishing awareness was part of this standard training, few could remember it with any certainty. Almost all nonclickers considered annual company training ineffective; it repeated information that they already knew and instructed them to act in ways they already did. The all-clickers felt the same, but they made fewer connections between known information and appropriate performance.

Company security might offer a false sense of confidence. Almost all interviewees believed that company firewall and information security programs prevented malicious emails from reaching them. Long-term employees had noticed a positive change in the past few years, recalling the days when inboxes were inundated with spam and unsolicited emails.

Most interviewees reported different behavior when working behind the corporate firewall, because a dedicated security staff and cybersecurity tools, such as firewalls and monitored antivirus software, provide a level of

security usually not available on home computers. These controls make interviewees more likely to click links in emails or on the Web while using their corporate computers, because they feel more protected within the company firewall. All participants said they double-check the URL when the warning for unrecognized websites appears on their corporate computers. Almost all then quickly evaluate whether they had already visited the site or need it for work, in which case they click through. If the site is questionable but they still want to visit it, many said they might use a personal computer instead.

Interviewees had different opinions about risk notifications. The half who preferred risk warnings on the corporate intranet said they already received too many emails; they often ignore regular company emails. Ironically, the half who reported paying no attention to risk notifications suggested that warnings be included in a regular company email. However, almost all commented that notification more frequently than monthly would create information overload; people would stop paying attention.

On the other hand, almost all wanted specific information about their personal security posture on the corporate network, something like a personal report card. Some favored visualization (for example, you clicked X number of bad links), whereas others preferred specific anecdotes (for example, everyone who clicked X link lost personal information), particularly those that highlight risks to the company (for example, X number of employees seriously compromised the company's network this month, and there were consequences). They wanted feedback to help maintain awareness, especially about possible privacy issues. Although interviewees knew that activities on the corporate network could be monitored, they didn't like reminders about regular monitoring. They preferred monthly or quarterly reporting, or notification only when they had violated some information security standard or otherwise caused a potential problem.

The Catch and the Takeaway

Making embedded training effective in a corporate setting is more difficult than earlier studies suggest. Our results indicate that immediate feedback and tailored framing don't suffice to reduce click rates or increase reporting. These results add to the growing body of research on spear phishing by replicating aspects of previous work and demonstrate that training effects might be lost somewhere between Kumaraguru and his colleagues' 28 days and our longer testing intervals. However, this research had several limitations that can be addressed in replications of this study:

- All trials must use spear phishing emails that are equally difficult to recognize.
- To determine how long the training effect lasts, the study must incorporate repeated trials over a significant period of time. However, such studies are time-consuming and expensive, and in an organization with substantial turnover, it might be impossible to involve the same people for the entire study period.
- This article reports on only one corporate study. The researchers conducted a second study, to be reported in the future, using MBA students with business experience. The study yielded similar results, reinforcing our conclusions. Nevertheless, for widespread validity, the study must be replicated at other institutions.
- There is no practical way to confirm that a participant actually read the training page completely. By using window opening time as a surrogate measure, we can confirm only that the window was open long enough to permit reading. Similarly, we can monitor whether a participant scrolled to the page's bottom; if not, the participant didn't read the page completely.

Changing security behavior is challenging, and it takes only one misstep to seriously compromise a system. Our study suggests that effective embedded training must take into account not only framing and security experience but also perceived security support, information load, preferred notification method, and more.

Although our study adapted findings and materials from prior research, it was the first of its kind in several dimensions:

- It used a very large, stratified sample, with random assignment to one control group or one of four treatment groups.
- Employees participated as part of their normal workday and in their normal working environment.
- The study materials are available from the I3P for replication in other environments.

We invite you and your organization to use our materials, administer a similar study in your own context, and build a corpus of carefully designed experiments that can help us design more effective security training. The study materials, including the webpages used and the software to create senders, track link clicks, and manage email logs, are available from the I3P for replication in other environments. ■

Acknowledgments

The authors prepared this work under award 2006-CS-001-000001 from the US Department of Homeland Security. The statements, findings, conclusions, and recommendations are

those of the authors and do not necessarily reflect the views of the US Department of Homeland Security.

References

1. B. Landreth, *Out of the Inner Circle*, Microsoft Press, 1985.
2. K.J. Higgins, "Spear-Phishing Experiment Evades Big Name Email Products," Dark Reading, 5 Jan. 2010, www.darkreading.com/end-user/spear-phishing-experiment-evades-big-nam/222200326.
3. "Phishing Activity Trends Report 2nd Half 2010," Anti-Phishing Working Group, July 2011; www.antiphishing.org/reports/apwg_report_h2_2010.pdf.
4. M.J. Schwartz, "Spear Phishing Attacks on the Rise," *Information Week*, 8 June 2011; www.informationweek.com/news/security/attacks/230500025.
5. "Phishing Activity Trends Report 1st Half 2011," Anti-Phishing Working Group, 2011; http://docs.apwg.org/reports/apwg_trends_report_h1_2011.pdf.
6. "Email Attacks: This Time It's Personal," Cisco, June 2011, www.cisco.com/en/US/prod/collateral/vpndevc/ps10128/ps10339/ps10354/targeted_attacks.pdf.
7. S.L. Pfleeger, D. Caputo, and M.E. Johnson, *Workshop Report: Cyber Security through a Behavioral Lens II,* Inst. Information Infrastructure Protection, July 2010; www.thei3p.org/docs/publications/442.pdf.
8. D. Bank, "Spear Phishing Tests Educate People about Online Scams," *The Wall Street J.*, 17 Aug. 2005; http://online.wsj.com/public/article/0,,SB112424042313615131--z_8jLB2WkfcVtgdAWf6LRh733sg_20060817,00.html?mod=blogs.
9. J. Hong, "The State of Phishing Attacks," *Comm. ACM*, vol. 55, no. 1, 2012, pp. 74–81.
10. P. Kumaraguru et al., "Protecting People from Phishing: The Design and Evaluation of an Embedded Training Email System," *Proc. Conf. Human Factors in Computing Systems* (CHI 07), ACM, 2007, pp. 905–914.
11. P. Kumaraguru et al., "School of Phish: A Real-World Evaluation of Anti-Phishing Training," *Proc. Symp. Usable Privacy and Security* (SOUPS 09), ACM, 2009, article 3.
12. A. Tversky and D. Kahneman, "The Framing of Decisions and the Psychology of Choice," *Science*, vol. 211, no. 4481, 1981, pp. 453–458.

Deanna D. Caputo is a principal social psychologist at the MITRE. She investigates questions addressing the intersection of social science and computer science, such as insider threat, cybersecurity, and effective ways to change behavior. Caputo received a PhD in social and personality psychology from Cornell University. Contact her at dcaputo@mitre.org.

Shari Lawrence Pfleeger is an I3P Research Fellow and is editor in chief of *IEEE Security & Privacy*. Contact her at pfleeger@dartmouth.edu.

Jesse D. Freeman is a lead systems engineer at MITRE. He focuses on cybersecurity research and object-oriented real-time development. Contact him at jfreeman@mitre.org.

M. Eric Johnson is dean of Vanderbilt's Owen Graduate School of Management and the Bruce Henderson Professor of Strategy. His teaching and research focus on digital strategies within the extended enterprise and information security's impact on business performance. Johnson has a PhD in engineering from Stanford University. Contact him at m.eric.johnson@owen.vanderbilt.edu.

Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.