AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: Applied

ISSN: 1076-898X

2022, Vol. 28, No. 1, 85–99
https://doi.org/10.1037/xap0000410

Is the Key to Phishing Training Persistence?: Developing a Novel Persistent Intervention

Dawn M. Sarno[1], Rachel McPherson[2], and Mark B. Neider[3]
[1] Department of Psychology, Clemson University
[2] School of Nursing, University of Maryland Baltimore
[3] Department of Psychology, University of Central Florida

Most previous phishing interventions have employed discrete training approaches, such as brief instructions aimed at improving phishing detection. However, these discrete interventions have demonstrated limited success. The present studies focused on developing an alternative to discrete training by providing college-age adults with a persistent classification aid that guided them on what characteristics a phishing email might include. Experiment 1 determined if this classification aid improved email categorization performance relative to feedback and control. Experiment 2 continued the evaluation of the classification aid to determine whether performance improvements were due to increased systematic processing of emails. Experiment 3 explored whether the classification aid would be more effective when embedded directly into the email interface. The results suggested three major findings. (a) Persistent interventions can improve phishing email detection. (b) Performance improvements were largest when the classification aid was embedded into the task. (c) These benefits were likely driven by an improved systematic processing of the emails. This novel phishing classification aid serves as a promising persistent intervention that can be adaptable to specific email environments and individuals.

Public Significance Statement
The present studies developed a persistent phishing intervention as an alternative to standard discrete methods. The results indicate that persistent interventions may be a promising strategy for improving phishing detection, particularly when embedded into the task, for both organizations and researchers.

Keywords: phishing, intervention, phishing classification aid, embedded interventions

Supplemental materials: https://doi.org/10.1037/xap0000410.supp

Cybercrime has become a pervasive problem in modern society. Email systems, like Google's GMAIL, block over 100 million phishing emails every day (Pegoraro, 2019). However, even though spam filters block countless emails, cyberattacks are ever-evolving and it has become virtually impossible to prevent every single phishing attempt from entering users' inboxes. This reality, that users are required to determine the authenticity of some of the phishing emails they receive, translates into an economic loss of over $17,700 per minute (Urrico, 2019). With losses of more than $9 billion a year, imperfect spam filters are simply not enough. It is critical that human users can identify fraudulent emails when spam filters inevitably fail to identify them. Recently, cybersecurity research has begun to investigate why certain individuals are more vulnerable to phishing attacks (e.g., Cain et al., 2018; Sarno et al., 2020; Sheng et al., 2011) and how they can be trained to detect them (e.g., Kumaraguru et al., 2007; Mayhorn & Nyeste, 2012; Sawyer et al., 2015). Somewhat surprisingly, several studies have indicated that younger adults might be the most vulnerable age group to phishing attacks. Furthermore, there has been minimal success in training younger users to detect email scams. The present studies explore a novel intervention aimed at assisting young email users in their evaluation of potentially fraudulent emails.

Younger Adults and Phishing Susceptibility

College-age adults appear to be a particularly vulnerable age group for phishing attacks. While investigating the differences in phishing vulnerability across the lifespan, Sarno et al. (2020) discovered that younger adults may be more susceptible to phishing emails than their older adult counterparts. Both age groups demonstrated similar overall classification accuracies, but younger adults were more conservative in their ratings, rating fewer emails as spam or not safe. This is particularly concerning given that younger adults were exhibiting behaviors that reflected a sense of complacency for fraudulent emails. This complacency is congruent with findings from Sheng et al. (2011) who determined that individuals ages 18–25 are the most susceptible age group to phishing attacks. One potential explanation for these findings is that younger adults are

This article was published Online First February 14, 2022.
Dawn M. Sarno https://orcid.org/0000-0001-5605-5957
Rachel McPherson https://orcid.org/0000-0002-0159-6120
Mark B. Neider https://orcid.org/0000-0001-7481-0960
Correspondence concerning this article should be addressed to Dawn M. Sarno, Department of Psychology, Clemson University, 321 Calhoun Dr., Clemson, SC 29634, United States. Email: dmsarno@clemson.edu

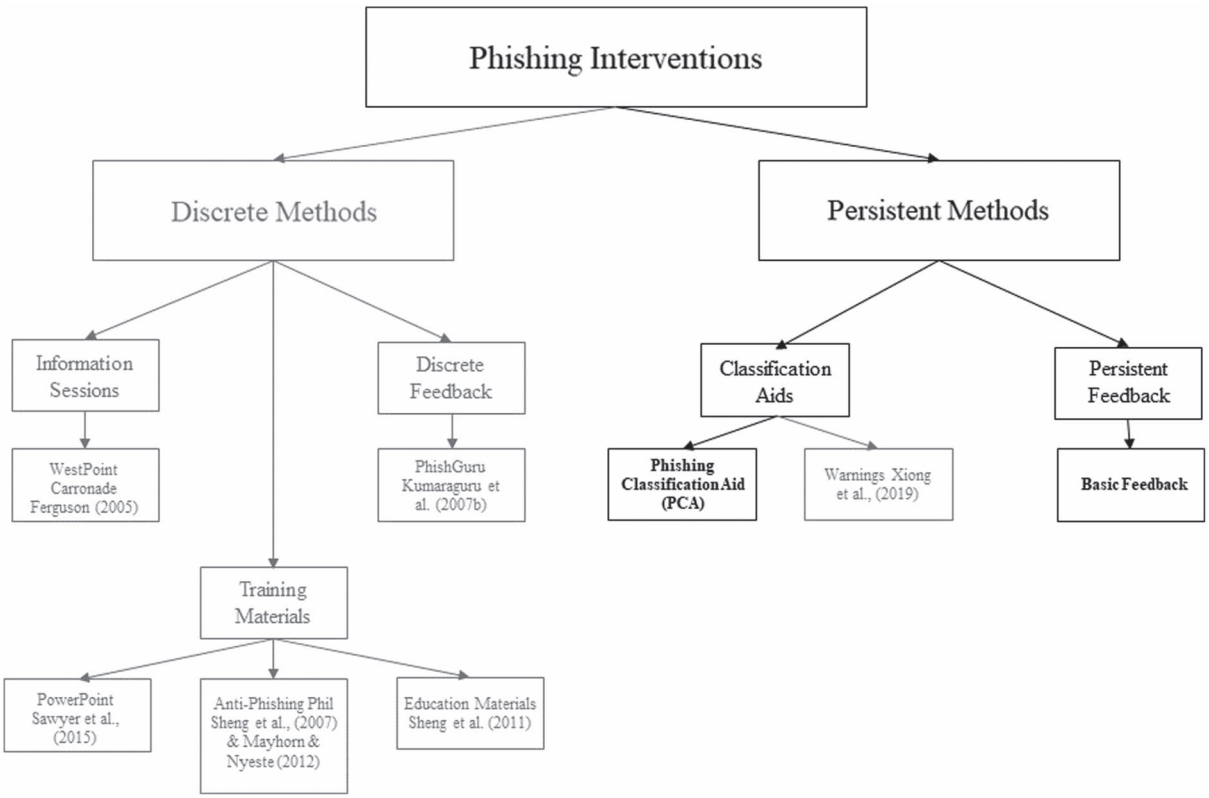
more complacent in online environments, possibly due to more exposure to technology and/or poor cyber habits. Cain et al. (2018) found that younger adults do indeed behave less securely online in general compared to older adults. Thus, younger adults are a population that might benefit from additional instruction to decrease their likelihood of interacting with phishing emails. Such training is potentially most important for when this younger adult population enters the workforce and these younger employees become potential liabilities for organizational information security.

Previous Training Attempts

There have been several attempts to develop training interventions for younger adults to bolster their ability to detect phishing attacks (see Figure 1). One such attempt by Sawyer et al.'s (2015) trained email users with a single PowerPoint of information and assessed performance within clerical emails. Although they saw robust training benefits, they were only tested on emails with slight variations from those they trained on and there were no data to indicate if the participants retained this information for any length of time following training. Phishing emails can be quite diverse in nature, not only in the email content (e.g., banking, social media, shopping; Sarno et al., 2020) but in the specific deception triggers that could indicate to users that the email is fraudulent (Drake et al., 2004). Other training methodologies have focused on more diverse email sets and assessed the time-based atrophy of any accrued benefits. Kumaraguru et al. (2007) compared two methods of feedback after interacting with phishing emails. The first method had feedback embedded into the web browser and the second method had the same information sent in a secondary email (i.e., not embedded). Accuracy for detecting phishing emails was increased by more than 50% in the embedded condition relative to the nonembedded condition. However, even in the embedded condition participants only exhibited a 68% detection rate for phishing emails. This low accuracy following training is alarming given that it only takes one email to compromise the integrity of a system. Mayhorn and Nyeste (2012) conducted a similar study investigating the benefits of delivering phishing information through a comic or video training. Although their results showed that individuals benefited from the phishing materials immediately, participants were more likely to fall for a phishing attack 2 weeks following training than they had been prior to any intervention at all. Similar patterns have been observed in highly motivated groups of younger adults as well. The West Point Carronade is perhaps the most impactful example of just how poor younger adults can be at detecting phishing emails. After several security breaches due to clicking links and opening attachments in fraudulent emails, West Point decided to test their cadets with a simulated phishing email (Ferguson, 2005). Shockingly, over 80% of the cadets clicked a link within the email, and over 90% of freshmen cadets clicked the link, despite receiving computer security education just 4 h prior. Overall, current cybersecurity training approaches have been very limited

Figure 1

Phishing Interventions: Discrete Versus Persistent Methods



Note. The persistent methods proposed in the present studies are in bold.

in engendering any meaningful long-term benefits for accurate phishing detection.

Discrete Versus Persistent Training Interventions

One key aspect that most phishing detection training paradigms share is that they employ discrete interventions that are no longer present following the training period (see Figure 1). One could, however, envision that email systems, such as GMAIL, could permanently embed some type of persistent cyber intervention to encourage their users to be more resilient to cyberattacks. Byrne et al. (2016) suggested something similar in the context of facilitating safe online behaviors within organizations. Specifically, they examined participants' awareness of risky online behaviors and enjoyment while utilizing the internet. Participants indicated the types of actions they take on the internet, why they take those actions and how risky those actions are. The results demonstrated that similar to phishing emails, when individuals are online, they typically make poor actions due to their lack of awareness of the risks associated with those actions. The authors suggested that organizations could develop aids or lists of information that provide an easy and simple view of safe online behaviors. A similar type of classification aid could be developed for email interfaces. Individuals could have phishing classification aids embedded into their email interface that highlight potential phishing indicators, and this could facilitate safer email habits. This type of persistent intervention may be more effective than previous attempts because it is not removed following training and does not require the user to recall any training information. In the discrete training methods, email users are tasked with not only classifying emails, but with remembering the training information as well. This imposes unnecessary workload on the email user and potentially impairs their performance (Wickens, 2008). Additionally, one key limitation to email users learning how to detect phishing emails is the absence of immediate feedback on their performance. Schmidt and Bjork (1992) detail the importance of providing feedback instantaneously to aid in the learning process by not only encouraging correct behavior but also increasing efficient behavior. Although it is likely impossible for email systems to provide feedback for users on real fraudulent attacks, it is possible for systems to email users "test" fraudulent attacks or provide email simulations to evaluate how vulnerable they are. Thus, a potential, albeit still challenging, avenue of a persistent phishing intervention is the inclusion of periodic feedback. In the meantime, comparing other interventions to feedback in laboratory settings may allow researchers to develop more feasible methods (e.g., classification aids) for real-world use.

Persistent interventions could be an effective alternative to discrete training methodologies not only due to decreased workload but because of how they influence task performance. Vishwanath et al. (2016) recently developed the Suspicion, Cognition, and Automaticity Model (SCAM) of phishing susceptibility (see Vishwanath et al., 2016, for a full description/figure). This model details two key paths that lead to suspicion of phishing emails, the automaticity path (includes deficient self-regulation and email habits) and the cognition path (includes cyber risk beliefs and heuristic/systematic processing). The automaticity path in the SCAM focuses on how deficient self-regulation can cause poor email habits, ultimately leading to less suspicion of a phishing email. Since previous attempts at bolstering phishing detection have removed the intervention following training, individuals often slip back into their old email habits due to poor self-regulation. This is evidenced by the poor retention data in these types of studies (e.g., Mayhorn & Nyeste, 2012). However, persistent interventions may permanently change email users' evaluation of emails by introducing a new component (e.g., phishing classification aid) to the task. This serves to break the path between deficient self-regulation and email habits by introducing new, safer email practices. These types of interventions may also interact with the cognition path of the model by increasing more systematic processing of all emails. It is reasonable that if users are given a classification aid with phishing characteristics to look for they will go through the email more methodically, searching for those characteristics. Ultimately, changing both paths should result in increased suspicion of phishing emails and improve performance. Additionally, based on the previously mentioned work by Kumaraguru et al. (2007) persistent interventions should be even more effective if embedded into the task. In the context of Kumaraguru et al.'s (2007) work and of the SCAM, an embedded phishing classification aid (PCA) should encourage email users to practice safer email habits and to more systematically process emails.

Signal Detection Theory and Training

Most previous cybersecurity studies have focused on how phishing interventions can decrease interactions with phishing emails (e.g., Kumaraguru et al., 2007). However, only capturing whether a user interacts with a phishing email obscures important facets of performance. More specifically, these interventions may only cause users to become more cautious rather than improve their ability to discern phishing emails from legitimate ones. By prioritizing the detection of phishing emails users may be deleting or ignoring important legitimate emails. Ideally, phishing intervention studies need to capture whether changes in performance are associated with an improved ability to detect and categorize legitimate and illegitimate emails, or some sort of learned decision bias (e.g., users might lower their threshold for classifying an email as phishing because that seems like a safer choice than misclassifying a phishing email as legitimate). Signal Detection Theory (Green & Swets, 1988) incorporates both this detection ability (i.e., sensitivity) and how cautious someone is (i.e., response criterion). These signal detection measures have proven useful in recent cybersecurity studies (Canfield et al., 2016; Sarno et al., 2020) in determining whether poor phishing performance is due to email users' ability to detect phishing emails, them being too liberal (or risky) in their classifications, or a combination of both. This distinction is important for phishing training paradigms because improvements in sensitivity mean that users are improving their ability to detect phishing emails without the cost of missing legitimate emails. Additionally, changes in sensitivity are likely to be more resilient to time-based atrophy compared to changes in response criterion. As detailed by the SCAM (Vishwanath et al., 2016), email users often return to their poor habits relatively quickly following cybersecurity training. Thus, if training interventions are only changing response criterion, the benefits may be short lived. If the intervention improves sensitivity, email users should be more likely to retain training benefits. In situations where the interventions struggle to meaningfully improve sensitivity (as in past research), it may be necessary for the intervention to make users more cautious to avoid them from

engaging with fraudulent emails, at least until sensitivity improvements can be attained. For example, Sheng et al. (2007) and Xiong et al. (2019) have found some success with interventions that improved phishing sensitivity but also made users more cautious for both emails and webpages. Thus, the present studies will analyze how our persistent interventions influence both phishing sensitivity and response criterion.

Metacognition and Training Performance

As previously mentioned, prior attempts to improve phishing detection performance have demonstrated mixed results (e.g., Ferguson, 2005; Sawyer et al., 2015). One reason why participants may not demonstrate improved detection abilities is due to miscalibrated metacognition. Canfield et al. (2019) detailed the importance of metacognition, or awareness of one's thoughts and abilities, in the context of phishing interventions. Specifically, that participants demonstrate better metacognition for performance on legitimate emails compared to phishing emails. This may be because users don't always get feedback on phishing emails, particularly when they ignore or delete them. The authors suggest the importance of appropriately calibrating metacognition in phishing interventions. If users are not confident in their classifications, they may be less likely to interact with dangerous emails. However, the opposite is also true; if users are extremely confident in their misclassifications of phishing emails, they may be more likely to interact with them. Thus, the present studies measured metacognition (i.e., confidence ratings) to determine if persistent interventions can improve metacognition for email classifications.

The Present Studies

The present studies propose a PCA as a novel and persistent intervention to improve phishing detection (see Figure 1). The goal of the present studies was to develop and test this novel PCA. The persistent aid developed, like Byrne et al. (2016), includes tips aimed at assisting email users in categorizing phishing and nonthreatening emails. We investigated how effective this PCA, as well as basic feedback (correct/incorrect feedback after each trial), were at improving phishing detection. Basic feedback was included to serve as a comparison for the success of our intervention. We hypothesized that both persistent interventions would improve phishing detection, but that the PCA would induce the most robust performance benefits (e.g., higher sensitivity), particularly when embedded into the task. Additionally, we predicted that all interventions would improve participants' metacognition resulting in higher confidence for correct responses, and lower confidence for incorrect responses relative to the control group. Experiment 1 investigated differences in email classifications for the physical classification aid (i.e., a physical list of tips) and basic feedback groups compared to a control group that completed the task without any intervention. Experiment 2 examined these differences again but controlled the presentation of the emails for all participants to 15 s to ensure that any differences in performance were due solely to the interventions and not due to differences in how long they viewed the emails. Finally, Experiment 3 evaluated how performance differs for the physical PCA versus the same aid embedded into the task.

Experiment 1

Method

Participants

A total of 81 participants (M_{age} = 19.23, 44 females) were recruited from the University of Central Florida in exchange for course credit. All participants had normal or corrected-to-normal vision and provided informed consent prior to participating in the study. After the informed consent, participants were prescreened for normal vision (visual acuity (20/32 or better-corrected vision on a Snellen eye chart)) and color vision (Ishihara's test for color blindness; 13 plates). A total of six participants had accuracies that fell more than two standard deviations below the mean and were removed from all further analyses. This left a total sample size of 75 participants (25 participants per group). This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the University of Central Florida.

A power analysis was conducted in G * Power (Faul et al., 2007) to ensure that Experiment 1 would have enough participants to detect an effect of the phishing interventions. Sawyer et al. (2015) found an effect size of η_p^2 = .23, for the influence of training on email performance. Thus, an ANOVA power analysis was calculated using Cohen's *f* of .55, power of 0.95, an α probability of 0.05, and 3 groups. Based on this analysis, 57 participants (19 per group) should be sufficient to find a moderate effect size exploring the impacts of various intervention types. Given the present interventions were different than that of Sawyer et al. (2015) an additional six participants per condition were recruited to ensure sufficient power for analyses.

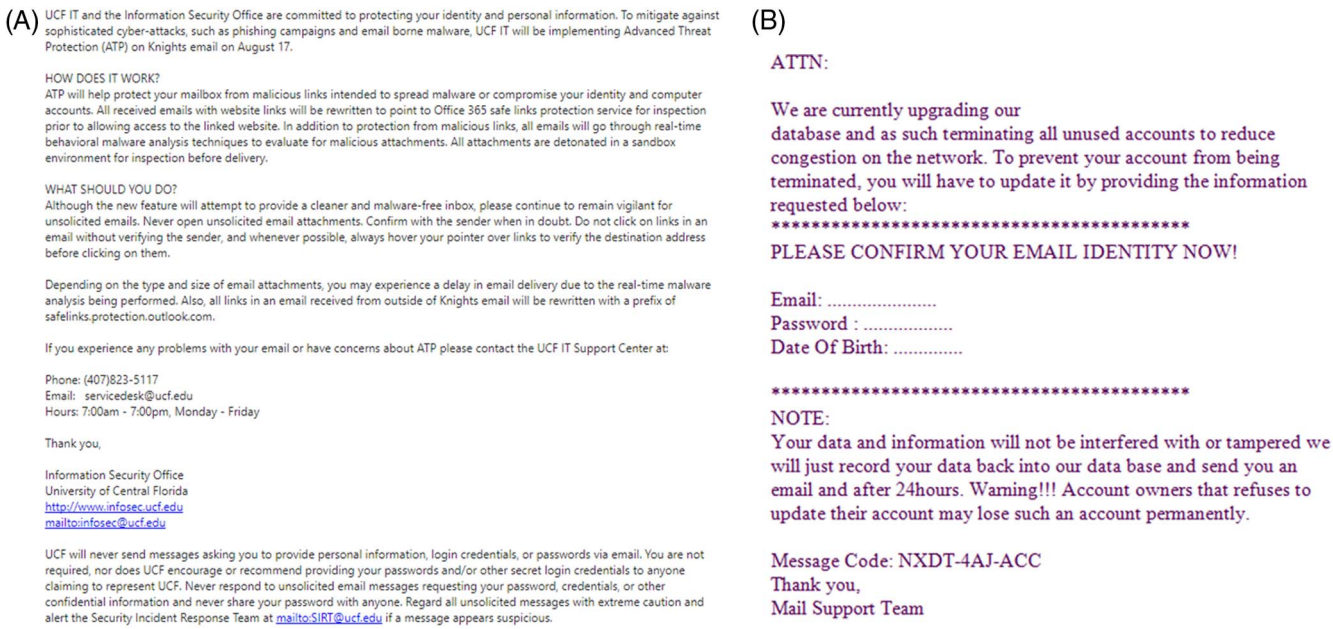
Stimuli and Procedure

The entire experiment was programmed and run in SR Research Ltd's Experiment Builder. Each participant viewed 100 real emails (Figure 2) embedded into a Gmail interface (see Figure 3, for an example) on a Dell Professional P190S monitor with a resolution of 1,280 × 1,024. Participants sat approximately 23 in. away making the visual angle of the screen roughly 36° × 29°. All emails were real emails that were obtained either through internet searchers or from the experimenter's personal inboxes or spam folders. Half of these emails were from legitimate sources; the other half were phishing attempts (see Figure 2, for examples). This email set has been utilized previously in similar studies and is diverse in both content (e.g., banking, social media, shopping) and the phishing themes utilized (e.g., threats to delete/suspend accounts, requiring a quick response; see Sarno et al., 2020, for a full description of the email set).

All participants provided informed consent, were prescreened for normal vision, and then were seated at a computer station for the remainder of the study. Participants were assigned to one of three conditions. The first condition was a PCA condition where participants were provided with a physical list of seven questions aimed at indicating whether an email was legitimate or not (see Figure 4A). This list of questions was determined by the most predictive phishing themes identified in previous research (Sarno et al., 2020) and participants were instructed to use the questions to inform their evaluations. The classification aid was provided to participants on a standard 8.5 in. × 11 in. piece of paper in a plastic sleeve. The second condition evaluated the benefit of basic feedback on phishing detection performance (see Figure 4B). Specifically, after each trial

Figure 2

Example Emails (A) Legitimate Email and (B) Phishing Email



Note. See the online article for the color version of this figure.

participants would be reminded of what answer they selected and whether it was correct or incorrect. The last condition served as a control condition where participants completed the email classifications without feedback or a classification aid.

All participants were asked to view emails and indicate which were legitimate and which were not legitimate via button press. Each participant was free to view the email as long as they wished (M_{RT} = 14.10, SD_{RT} = 6.66). After the main classification, participants were asked what action they would take next (e.g., click a link/open an attachment, reply, need more information, delete, or report as suspicious). If they chose "need more information" they were also able to indicate what specific information they needed. We included this open response portion of the "need more information" action to ensure participants did not always default to choosing "need more information." Finally, participants indicated how confident they were in their response on a sliding scale ranging from *not confident* (0) to *confident* (100). After evaluating all 100 emails, participants completed a brief demographic questionnaire and survey which included questions about previous cybersecurity and technology experience, impulsivity (i.e., Cognitive Reflection Task; Frederick, 2005), the Multimedia Index (i.e., MMI; Ophir et al., 2009) and personality (i.e., HEXACO openness to experience and extraversion; Ashton & Lee, 2009).

Results and Discussion

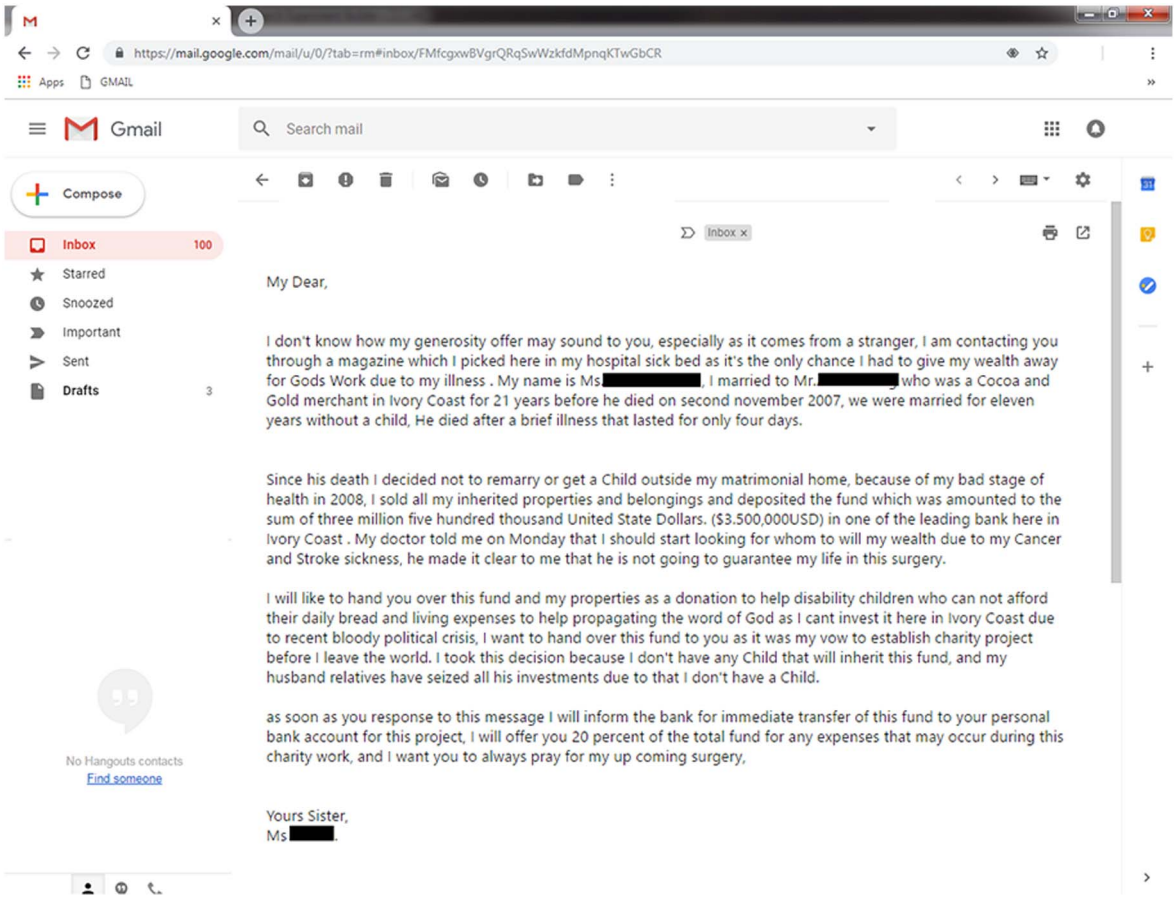
To determine how the efficacy of our interventions, we conducted a two-factor mixed ANOVA with an α level of .05, with intervention (control, PCA, basic feedback) and email type (legitimate, phishing) as the independent variables, on response times. Additionally, we conducted separate one-way between-subjects ANOVAs with α levels of .05, with intervention (control, PCA, basic feedback) as the independent variable, on signal detection measures and action accuracy. Confidence was also analyzed in a three-factor mixed ANOVA with an α level of .05, with intervention (control, PCA, basic feedback), accuracy (correct, incorrect), and email type (legitimate, phishing) as the independent variables. The individual difference variables (i.e., personality, previous cyber experience, impulsivity) are not reported here. Materials and analysis code for this study are available by emailing the corresponding author.

Signal Detection Theory Measures

Performance was analyzed within the context of Signal Detection Theory (Green & Swets, 1988). Hits were considered on trials where participants correctly identified not legitimate (i.e., phishing) emails; false alarms were on trials where participants incorrectly classified legitimate emails as not legitimate ones. Response criterion (c) was utilized in the present studies over response bias (β) because of the more balanced distribution between conservative and liberal responses. Response bias (β) scores are constricted to 0–1 for liberal responders, and 1–∞ for conservative responders (Green & Swets, 1988). For response criterion, liberal responders have scores that are <0, conservative responders have scores that are >0, and unbiased responders have scores equal to 0 (Stanislaw & Todorov, 1999). In the present studies, liberal responders classified more emails as not legitimate, and conservative responders classified more emails as legitimate. Hit rate and false alarms are presented in the Supplemental Materials.

Sensitivity. There was a main effect of intervention on sensitivity, $F(2,72) = 3.64$, $p = .031$, $\eta_p^2 = .09$ (see Figure 5A). Pairwise comparisons revealed that participants in the PCA condition ($M = 1.47$, $SD = .50$) were more sensitive than the control group ($M = 1.09$, $SD = .62$, $p = .013$), but not more sensitive than the feedback group ($M = 1.40$, $SD = .44$, $p = .631$). Participants in the

Figure 3
Example Email Embedded in Gmail Interface

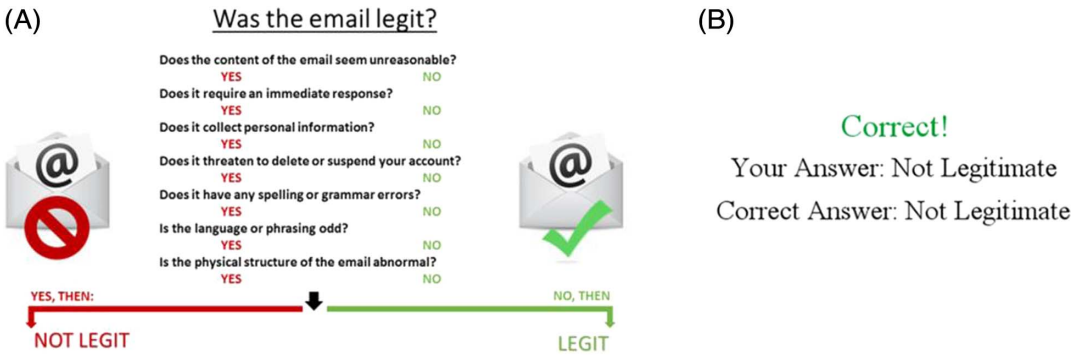


Note. See the online article for the color version of this figure.

feedback group were also more sensitive than the control group ($p = .043$).Overall, these findings indicate that the two persistent interventions improved the participants' ability to discriminate between the legitimate and phishing emails. Although our persistent interventions improved phishing sensitivity, sensitivity was generally low.

Response Criterion. Unlike sensitivity, there was no main effect of intervention on response criterion, $F(2,72) = 0.50$, $p = .554$, $\eta_p^2 = .02$ (see Figure 5B). Additionally, to classify participants' response criteria as liberal, conservative, or unbiased, each intervention group's response criterion was submitted to a one-samples *t*-test to determine if

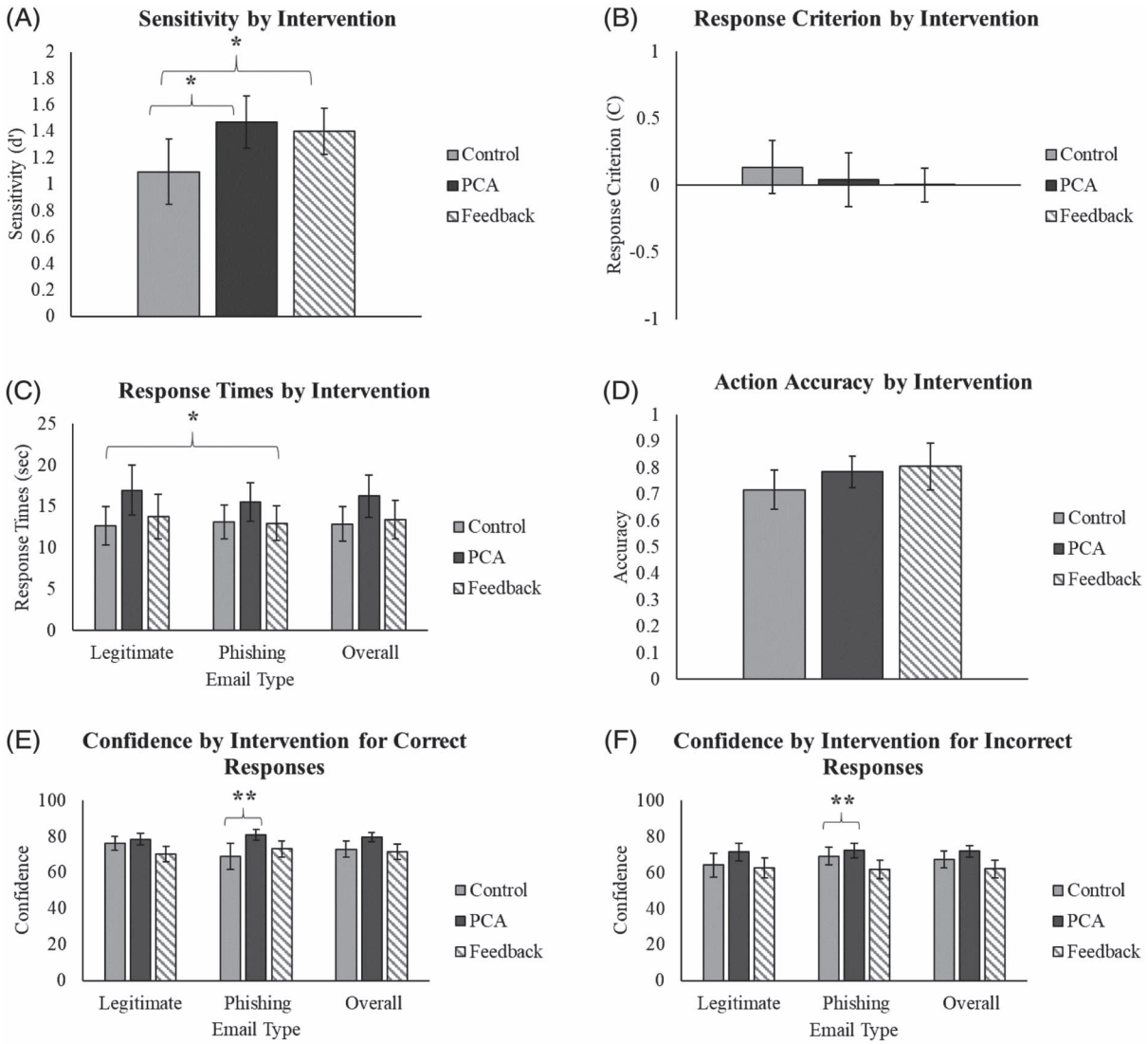
Figure 4
Example Interventions (A) Phishing Classification Aid (PCA) and (B) Basic Feedback



Note. See the online article for the color version of this figure.

Figure 5

Experiment 1 Results: (A) Sensitivity (d′), (B) Response Criterion (c), (C) Response Times, (D) Action Accuracy, (E) Confidence for Correct Responses, and (F) Confidence for Incorrect Responses



Note. Error bars represent two standard errors of the mean.
* $p < .05$. ** $p < .01$.

their scores were different from zero. Each groups' response criteria were not different from zero in this case ($p > .182$), suggesting that all participants were relatively unbiased in their responses. Taken together with the sensitivity results, these findings suggest that our persistent interventions improved email classification without changing our participants' response criterion.

Response Times

Response times were collapsed across both correct and incorrect responses. There was no significant interaction between email type and intervention, $F(2,72) = 1.23$, $p = .297$, $\eta_p^2 = .03$, nor a main effect of intervention on response times, $F(2,72) = 2.13$, $p = .127$, $\eta_p^2 = .06$. It is important to note that although there was no main effect of intervention on response times, pairwise comparisons indicated that the PCA group ($M = 16.14$ s, $SD = 6.89$) was slower to evaluate the emails than the control group ($M = 12.34$ s, $SD = 5.96$, $p = .044$). This result is potentially important given the benefits for the PCA group relative to the control group and suggest those benefits may be related to a speed/accuracy tradeoff. There was a main effect of email type on response times, $F(1,72) = 4.75$, $p = .033$, $\eta_p^2 = .06$ (see Figure 5C), such that participants were slower to classify legitimate emails ($M = 14.52$ s, $SD = 7.36$) compared to phishing emails ($M = 13.70$ s, $SD = 6.11$). Overall, these results suggest that the largest response time differences are due to the email's legitimacy.

Action Accuracy

Determining the correct actions for legitimate emails is challenging since any of our action choices may be considered correct for legitimate emails. For instance, it is not always necessary to respond to a legitimate banking email (e.g., legitimate credit card use alerts). Thus, legitimate emails were not included in the action accuracy analyses. Incorrect actions are clearer for phishing emails, as it is always inappropriate to engage with a phishing email. Actions were only considered correct for the phishing emails if participants selected need more information, delete, or report as suspicious. It would be inappropriate for users to respond to or click a link/open an attachment in a phishing email. There was no main effect of intervention on action accuracy, $F(2,72) = 0.38$, $p = .685$, $\eta_p^2 = .01$ (see Figure 5D), suggesting that our persistent interventions did not improve our participants' ability to select appropriate actions for the emails despite their increased sensitivity. Overall, participants appear to select dangerous actions on 20–30% of phishing emails.

Confidence

There was a main effect of intervention on confidence, $F(2,72) = 3.15$, $p = .049$, $\eta_p^2 = .08$ (see Figure 5). Pairwise comparisons revealed that participants in the PCA group ($M = 74.81$, $SD = 18.01$) were more confident than the control group ($M = 68.09$, $SD = 18.01$, $p = .025$) and the feedback group ($M = 68.78$, $SD = 18.01$, $p = .044$). There was no difference between the control and feedback groups ($p = .816$). Additionally, there was a main effect of accuracy on confidence, $F(1,72) = 87{,}337$, $p < .001$, $\eta_p^2 = .55$, such that participants were more confident for correct ($M = 74.16$, $SD = 10.22$) versus incorrect ($M = 66.60$, $SD = 11.60$) responses. There was an interaction of accuracy and intervention, $F(2,72) = 3.76$, $p = .028$, $\eta_p^2 = .10$, but there were no other main effects or interactions ($p > .089$).

Further analysis of the simple effects explored the accuracy/intervention interaction by examining the effects of intervention and email type for both correct and incorrect responses separately. These analyses revealed that confidence differences existed among the interventions when correctly identifying phishing emails, $F(2,72) = 5.12$, $p = .008$, $\eta_p^2 = .12$. Specifically, pairwise comparisons indicated that the PCA group ($M = 79.83$, $SD = 8.14$) was more confident than the control group ($M = 66.75$, $SD = 21.53$, $p = .002$), but not the feedback group ($M = 74.83$, $SD = 10.41$, $p = .230$) and there was no difference between the feedback group and the control group ($p = .054$).

Taken all together these results suggest that confidence differences between our interventions are largely driven by the PCA group being more confident, relative to the control group, when they correctly detected phishing attempts. Additionally, although participants were generally more confident for correct responses, they appear to be overly confident for incorrect responses.

Experiment 2

Experiment 1 determined that both our PCA and feedback interventions improved email classifications. However, participants who had the PCA were slower to classify emails compared to the control group. Taken together, these results suggest that the PCA may cause users to engage in a speed/accuracy tradeoff, where they are prioritizing accuracy over speed. The benefit of this tradeoff may be a more systematic processing of the email. However, it is unclear if this improved processing is due to the information contained in the PCA alone or just increased time on task. Experiment 2 investigated this possibility by controlling the amount of time each participant could view the emails to 15 s. This specific time window was selected because previous research suggests it generally encapsulates the longest response times across conditions (Sarno et al., 2020). If the benefits of the PCA disappear under controlled presentation, then the differences in Experiment 1 were likely due to the control group evaluating emails too quickly.

Method

Participants

A total of 75 participants ($M_{age} = 18.97$, 42 females) were recruited from the University of Central Florida in exchange for course credit. Prior to participating in the study all participants provided informed consent and had normal or corrected-to-normal vision. All participants were prescreened for vision (visual acuity (20/32 or better-corrected vision on a Snellen eye chart) and color vision (Ishihara's test for color blindness; 13 plates). The same sample size was utilized from Experiment 1 to determine the benefits of the persistent interventions under controlled time presentations.

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1 with the following exceptions. All participants viewed the emails on a Samsung Syncmaster 2,233 with a resolution of 1,680 × 1,050 making the visual angle of the screen roughly 45° × 29°. Due to the potential speed–accuracy tradeoff for the PCA group in Experiment 1, all participants viewed each email for 15 s, after which the email would disappear, and participants would be asked to indicate via mouse click if the email was legitimate or not legitimate. Additionally, after making their classifications participants were asked how confident they were in their classification, and what action they would take next (i.e., click a link, respond, need more information, ignore, or delete). If they selected "need more information" they were told to detail what that information might be.

Results and Discussion

Analyses were similar to those conducted in Experiment 1, with exception of response time analyses. Specifically, given that all participants were constrained to 15 s to view the emails, response times are not particularly informative and were not analyzed.

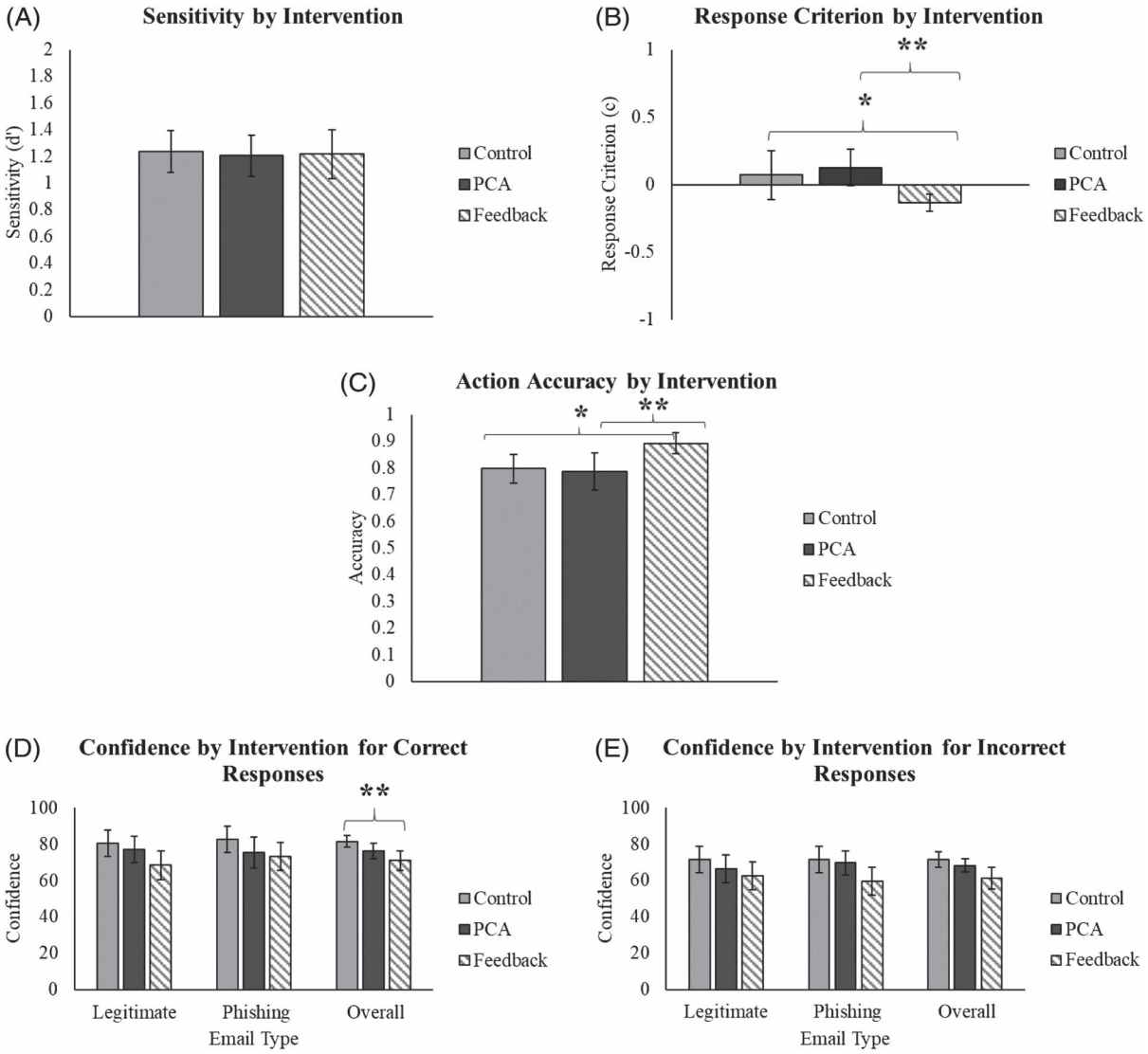
Signal Detection Theory Measures

Sensitivity. There was no main effect of intervention on sensitivity, $F(2,72) = 0.04$, $p = .962$, $\eta_p^2 < .01$, suggesting that our persistent interventions did not impact participants' abilities to classify emails under similar time constraints (see Figure 6A).

Response Criterion. There was a main effect of intervention on response criterion, $F(2,72) = 4.00$, $p = .022$, $\eta_p^2 = .10$ (see Figure 6B). Pairwise comparisons revealed that the basic feedback group ($M = -0.13$, $SD = 0.16$) was more liberal in their responses than the control group ($M = 0.07$, $SD = 0.45$, $p = .039$) and the

Figure 6

Experiment 2 Results: (A) Sensitivity (d′), (B) Response Criterion (c), (C) Action Accuracy, (D) Confidence for Correct Responses, and (E) Confidence for Incorrect Responses



Note. Error bars represent two standard errors of the mean.
* $p < .05$. ** $p < .01$.

PCA group ($M = 0.13$, $SD = 0.34$, $p = .009$). Additional one-samples *t*-tests were conducted on each groups response criterion to determine if they were different from zero. Neither the control group ($p = .448$) nor the PCA group ($p = .077$) were significantly different from zero. The feedback group was different from zero suggesting a liberal response criterion shift ($p < .001$). Overall, these results indicate that the feedback group demonstrated a criterion shift relative to the control and PCA groups where they were more likely to classify an email as not legitimate.

Action Accuracy

There was a main effect of intervention on action accuracy, $F(2,72) = 4.33$, $p = .017$, $\eta_p^2 = .11$ (see Figure 6C). Pairwise comparisons indicated that the feedback group selected more correct actions for phishing emails ($M = 89.20\%$, $SD = 0.10\%$) than the control group ($M = 79.68\%$, $SD = 0.14\%$, $p = .019$) and the PCA group ($M = 78.64\%$, $SD = 0.18\%$, $p = .009$). There was no difference between the PCA group and the control group ($p = .793$).

Confidence

There was a main effect of intervention on confidence, $F(2,72) = 5.06$, $p = .009$, $\eta_p^2 = .12$ (see Figure 6). Pairwise comparisons revealed that the feedback group ($M = 65.80$, $SD = 18.71$) was less confident than the control group ($M = 75.49$, $SD = 18.71$, $p = .002$). There were no other differences between the groups ($p > .078$). There was a main effect of accuracy on confidence,

$F(2,72) = 119.41$, $p < .001$, $\eta_p^2 = .62$, where participants were more confident for correct responses ($M = 75.43$, $SD = 10.96$) than incorrect responses ($M = 66.27$, $SD = 11.83$). There were no other significant effects ($p > .120$). Like Experiment 1, although participants were generally more confident when they were correct, they appear to still be overly confident for incorrect classifications.

Experiment 3

The first two experiments examined how a persistent PCA and a persistent feedback intervention can improve phishing detection. The results from Experiment 2 suggested that the PCA benefits were linked to time on task and a more systematic processing of the emails as predicted by the SCAM (Vishwanath et al., 2016). Thus, Experiment 3 aimed to develop a more robust PCA that might assist email users beyond increasing systematic processing by embedding the PCA into the task (Kumaraguru et al., 2007). The embedded PCA included the same information from the physical PCA. Based on the previous studies, it was expected that the physical PCA would improve performance relative to the control group, but the highest phishing detection was expected to occur in the embedded PCA condition. Additionally, since the goal of Experiment 3 was to develop a more robust PCA, the aid was tested in a more realistic and challenging environment. Specifically, we decreased the prevalence of phishing emails (Sawyer & Hancock, 2018) and increased the perceived load of emails (Vishwanath et al., 2011). By increasing the difficulty of the task, we hypothesized that the participants would be more likely to utilize the PCAs in their task. Finally, since the physical PCA provided similar benefits to the email-by-email feedback in Experiment 1, and it would be difficult, if not impossible, to provide this type of persistent feedback in the real world, it was not included as a condition in Experiment 3.

Method

Participants

Given Kumaraguru et al. (2007) already found benefits for embedded training compared to the same intervention nonembedded into the task, Experiment 3 recruited 57 participants based on Experiment 1's power analysis. Fifty-seven participants (M_{age} = 18.35, 19 males, 38 females) from the University of Central Florida participated in this study in exchange for course credit. All participants had normal or corrected-to-normal vision (20/32 or better-corrected vision on a Snellen eye chart) and color vision (Ishihara's test for color blindness; 13 plates).

Stimuli and Procedure

The stimuli and procedure were identical to Experiment 1 with the following exceptions. To develop a more robust intervention, Experiment 3 utilized challenging task conditions. Previous research has indicated that more emails (i.e., high email load) and low phishing prevalence (i.e., few phishing emails) make the task more difficult (Sawyer & Hancock, 2018; Vishwanath et al., 2011, respectively). To accomplish higher email load without changing the actual number of emails participants evaluated, all participants were deceived into believing they needed to evaluate 300 emails (via an inbox counter), when in reality they only viewed 100 emails. Specifically, after they had viewed emails 300–200, the experiment ended. To exacerbate this effect of email load, participants were also provided with a timer to keep track of how much time they had left. The timer was set to 1 h and participants were told they needed to classify all emails in that timeframe. If participants ran out of time, they were told to inform the researcher. This only happened in a few cases and participants were instructed to finish the task. Low phishing prevalence was accomplished by only including five phishing emails in the email set (i.e., 5%). These 5 phishing emails were randomly selected from the set of 50 emails utilized in Experiments 1 and 2.

Participants who received an intervention either saw a PCA that included tips for detecting phishing emails (see Figure 7) or the same aid embedded in the GMAIL interface (see Figure 8). This information was the same as the first two experiments with the following exceptions. To fit into the embedded interface, only the top five phishing email characteristics were utilized. These five remaining characteristics included implausible premise, time pressure, collecting personal information, account deletion/suspension threats, and spelling or grammatical errors (see Figures 7 and 8). Additionally, the information was framed as tips rather than questions. Thus, there were three groups, a physical PCA group, an embedded PCA group, and a control group who received no intervention.

The procedure utilized in Experiment 3 was the same as in Experiment 1 with the following differences. Instead of participants' impulsivity (i.e., CRT), previous experience (i.e., MMI), extraversion, and openness to experience being measured, participants completed a modified Big Five inventory (John et al., 1991) to determine how they rated on conscientiousness and agreeableness. These two personality measures were of particular interest since they have been linked to the utilization of interventions and following cybersecurity protocols (McBride et al., 2012; Shropshire et al., 2006, 2015). However, they were not included in the present analyses. Additionally, participants were no longer given the option to select need more information for their next action with the email. Instead, participants had to select one of the following, click a link or attachment, reply, check sender's address, delete, or report as suspicious. Finally, to get a more complete picture of each participant's metacognition, participants were also asked how threatening they found each email, and how difficult their overall classification was. Like the individual differences, these additional metacognition classifications were not reported here.

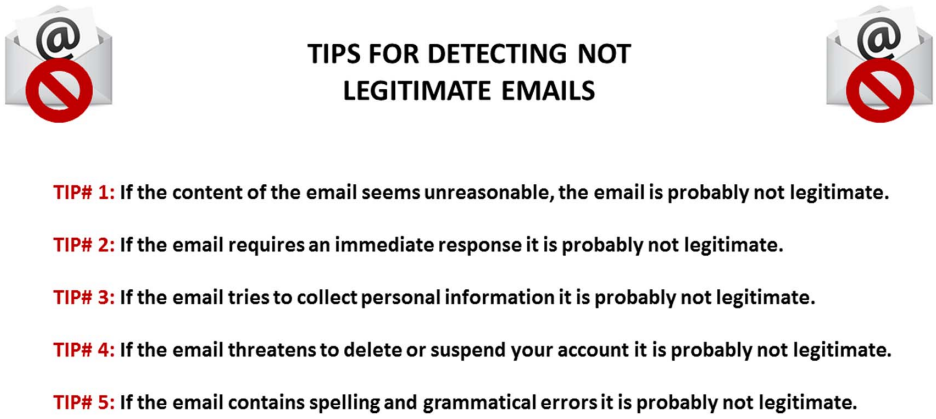
Results and Discussion

Identical analyses to Experiments 1 were conducted, with the following exceptions. The feedback condition was replaced by the embedded PCA condition. Due to the low frequency of phishing emails, sensitivity and response criteria were not analyzed. Instead, hit rate and false alarms were analyzed to determine how the interventions influenced performance.

Signal Detection Theory Measures

Hit Rate. There was a main effect of intervention on participants' hit rate, $F(2,54) = 4.30$, $p = .018$, $\eta_p^2 = .14$ (see Figure 9 A). Pairwise comparisons revealed that the embedded PCA ($M = .82$, $SD = .15$) engendered more hits than the control group ($M = .62$, $SD = .21$, $p = .005$). There were no other significant differences ($p > .067$). Overall, these results suggest that the embedded PCA

Figure 7
Physical (Nonembedded) Phishing Classification Aid (PCA)

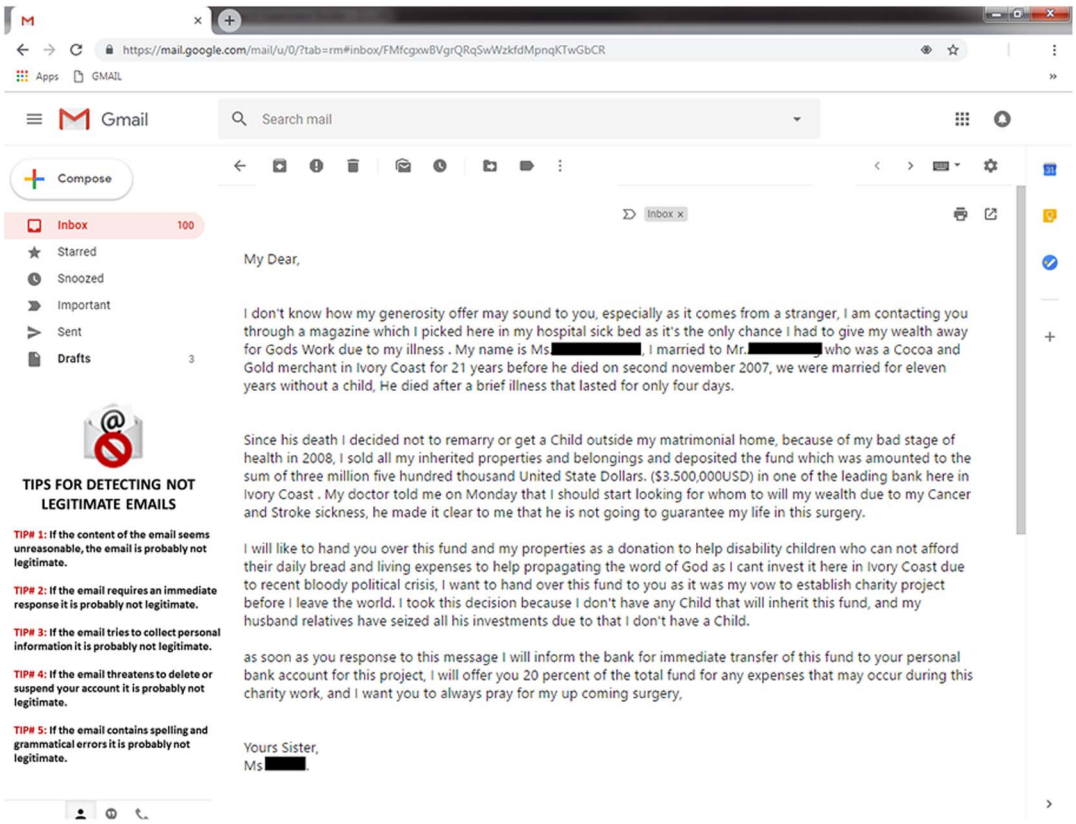


Note. Participants in this condition were given a phishing classification aid to assist in their classification of emails. The phishing classification aid included tips that indicated qualities of phishing (or not legitimate) emails (e.g., collecting personal information). See the online article for the color version of this figure.

group correctly identified more phishing emails. It is important to note that due to the low prevalence of phishing emails, this effect is based on only five trials. See the Supplementary Materials, for further comparison of hit rates for each participant by condition.

False Alarm Rate. There was no main effect of intervention on participants' false alarm rate, $F(2,54) = 1.95$, $p = .152$, $\eta_p^2 = .07$, suggesting that the interventions did not influence the participants' ability to classify legitimate emails (see Figure 9B). Although

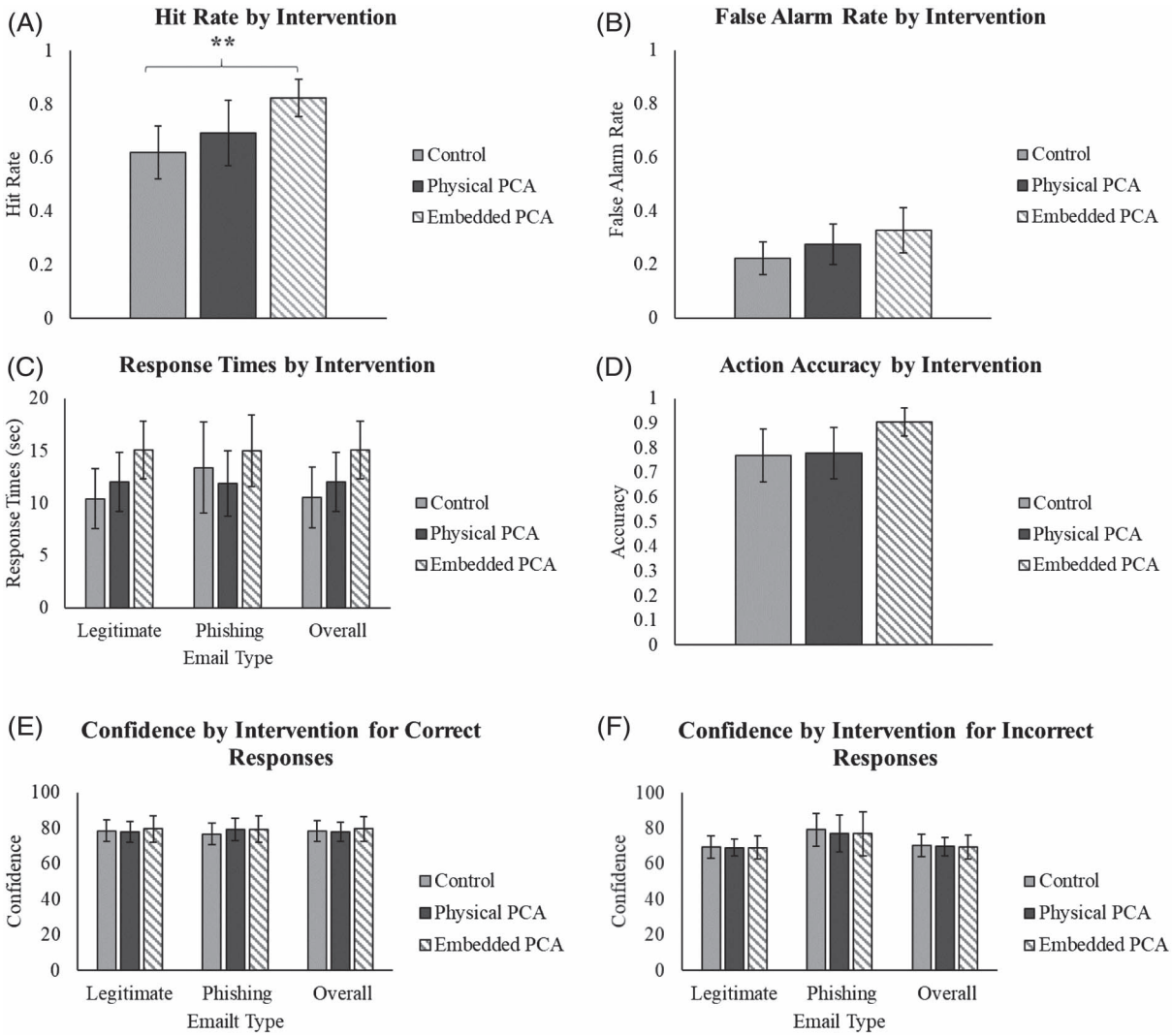
Figure 8
Embedded Phishing Classification Aid (PCA)



Note. Participants in this condition were given the same information from the physical (nonembedded) PCA condition but embedded into the GMAIL interface. See the online article for the color version of this figure.

Figure 9

Experiment 3 Results: (A) Hit Rate (B), False Alarm Rate (C), Response Times (D)



Note. Error bars represent two standard errors of the mean.
* $p < .05$. ** $p < .01$.

nonsignificant, the embedded PCA group appears to be trending toward an increase in false alarms. Coupled with their increase in hits, this may suggest that the embedded PCA participants were more liberal in their responses.

Response Times

There was no main effect of intervention, $F(2,54) = 1.27$, $p = .290$, $\eta_p^2 = .05$, suggesting that our interventions did not change the amount of time participants spent evaluating emails (see Figure 9C). There was no main effect of email type on response times, $F(1,54) = 2.39$, $p = .128$, $\eta_p^2 = .04$, suggesting that participants took the same amount of time to evaluate both phishing and legitimate emails. There was a significant interaction between email type and intervention, $F(2,54) = 3.24$, $p = .047$, $\eta_p^2 = .11$. Additional ANOVA's on each email type revealed that there were no differences between the intervention groups ($p > .082$). These analyses might be underpowered, or the interaction may be spurious. Overall, these results suggest that all participants took roughly the same amount of time to evaluate both types of emails regardless of the intervention group.

Action Accuracy

There was no main effect of intervention on action accuracy, $F(2,54) = 2.70$, $p = .076$, $\eta_p^2 = .09$, suggesting that although our embedded PCA improved participants' ability to detect phishing emails, it did not change the actions selected (see Figure 9D).

Confidence

There was a main effect of accuracy on confidence, $F(1,38) = 12.61$, $p = .001$, $\eta_p^2 = .25$, indicating that participants were more

confident when they were correct ($M = 76.61$, $SD = 13.53$) than when they were incorrect ($M = 71.93$, $SD = 16.22$; see Figure 9). There were no other significant effects ($p > .131$). Like Experiments 1 and 2, these results suggest that although participants have tuned their metacognition, with increased confidence when they are correct, they are still potentially too confident when they are incorrect.

General Discussion

The main goal of the present studies was to develop a novel persistent intervention that could improve the classification of phishing emails. Experiment 1 determined that our novel persistent intervention, the PCA, and basic feedback, increased participants' ability to discriminate phishing emails from legitimate ones. Additionally, neither group experienced a response criterion shift, suggesting that any observed benefits were the result of the participants' improved ability to discern phishing emails from legitimate ones. It is interesting that the PCA resulted in benefits similar to that of general feedback on performance. Since consistent feedback in the real world would be rather challenging to implement, it is encouraging that a persistent PCA can engender similar benefits.

It is important to note, that although the PCA group demonstrated higher sensitivity in their email classifications, they were slower than the control group to respond. Thus, it was difficult to determine if the PCA group was better because of the information contained in the aid, or the mere presence of the aid simply encouraged systematically processing of the emails. The SCAM (Vishwanath et al., 2016) suggests that it is this systematic processing in general that can often lead to the necessary suspicion to detect phishing emails. Thus, Experiment 2 investigated whether the benefits of the PCA were present when all three groups viewed the emails for the same amount of time. When we controlled the length of email presentations no meaningful benefits of the PCA (or the feedback) were found compared to the control group. These findings suggest that increasing the systematic processing of emails (i.e., slowing classifications down) may account for the benefits observed for the PCA in Experiment 1. It is possible that individuals just need more time to utilize the information in the PCA to make their decisions. However, the systematic processing hypothesis is further supported by comparing the data from the first two experiments. In Experiment 2, the null results were not due to the PCA group performing worse under time constraints, but rather the control group performing better. In Experiment 1, participants in the control group viewed the emails for roughly 12 s on average. Thus, in Experiment 2 when they were required to view the emails for roughly 3 s longer, they performed better. Taken together these results indicate that improving email classifications may be dependent upon encouraging a more systematic, deeper processing of the email. Classification aids may serve as a catalyst for inducing such behavior.

The main aim of Experiment 3 was to develop a more robust intervention than the physical PCA. Based on previous research (i.e., Kumaraguru et al., 2007), embedding the PCA into the task was expected to result in larger performance benefits. As predicted, the embedded aid resulted in the best performance relative to the control group as demonstrated by an increase in the detection of phishing emails (i.e., higher hit rates). This suggests that the embedded PCA group improved in their ability to detect phishing emails. This was coupled with an increase in false alarm rates (albeit non-significant), potentially suggesting a bias shift rather than an increase in sensitivity. Embedded training interventions have recently been linked to improved phishing detection in webpages (Xiong et al., 2019) and have been previously linked with better performance with emails (Kumaraguru et al., 2007). Overall, embedding helpful tips into the email interface may be a viable avenue for future persistent interventions. However, further research is necessary to develop a more robust email intervention that can improve classification accuracy to near ceiling performance.

How to Limit Dangerous Actions

Although our persistent interventions (i.e., PCA & feedback) generally improved classification accuracy, there seemed to be a disconnect between classification and the appropriate actions selected for phishing emails. In Experiment 1, participants still made inappropriate actions on roughly 20–30% of phishing emails. The feedback group from Experiment 2 did make safer actions in Experiment 2, however, this was likely due to their biased response criterion and not an increased sensitivity to phishing emails. In Experiment 3, although not significantly better, the embedded PCA group appeared to make safer actions on at least one more out of the five phishing emails compared to the other two groups. These results suggest that our persistent interventions did not consistently aid participants in what actions were selected for the phishing emails. Given that our classification aid and feedback did not highlight appropriate actions to take this finding is not entirely surprising. Previous research has indicated that cyber hygiene, or safe online behaviors, is often distinct from individual differences like previous cyber experience (Cain et al., 2018). Based on the SCAM (Vishwanath et al., 2016), it is possible that although our classification aid increased the systematic processing of the emails, thus increasing suspicion, it did not translate into actions. Future interventions may need to specifically indicate what are safe/dangerous actions in the email context. Regardless, further research is required to determine how to train users to take safer actions with emails.

Poor Metacognition Despite Training

What is perhaps the most surprising finding across previous phishing training studies is the overwhelmingly low classification accuracy. Although the present studies did find improvements for our persistent interventions, participants generally demonstrated low sensitivity toward the email stimuli. These low sensitivities suggest that individuals are potentially unaware of just how vulnerable they are. Unsurprisingly, across all three studies, participants were more confident for correct responses than incorrect responses. However, this difference was rather small, and even for incorrect responses participants were still fairly confident (70/100). This is particularly concerning for the phishing emails, given that their confidence could result in them being more likely to interact with a dangerous email. Future interventions should focus on increasing the confidence differences between correct and incorrect responses, allowing for participants to behave more cautiously when they might be vulnerable.

Conclusions

Overall, the present studies suggest that persistent interventions, such as phishing classification aids, can improve phishing detection. Additionally, these benefits may be due to increased systematic processing of the emails and are most effective when embedded into the task. These types of persistent interventions may be a viable alternative to discrete methods (see Figure 1). Not only are email users more likely to retain the information, since it is permanently added to the task, but it can be easily implemented into current email interfaces. Research focusing on the effectiveness of security warnings has shown that users tend to ignore security warnings after a prolonged use (Bravo-Lillo et al., 2013; Forget et al., 2016; Herley, 2009). This suggests that the effectiveness of persistent interventions may decrease over time. The present studies are an early attempt at developing an alternative to discrete training methodologies. Future studies should not only explore how persistent interventions influence performance over time, but also directly examine if they are better than discrete methods.

Perhaps most notably, the present phishing classification aids can (and should) be completely adaptable to new types of phishing emails, the specific email environment (e.g., work, school, personal), the person (e.g., age, personality), and an individual's performance (e.g., response bias, previous susceptibility). While promising, it remains unclear whether even persistent interventions will ultimately provide the efficacy necessary to insulate email users against phishing attacks to an optimal level. In application, it may well be the case that until training methods provide evidence of demonstrable improvements in email users' sensitivity to phishing emails, it may be best to focus on interventions that encourage users to be more cautious with their email classifications in general (i.e., response criterion shifts). Recent work from Canfield and Fischhoff (2018) suggests a similar notion that interventions that target response bias (or criterion) shifts may be more effective than those methods targeting sensitivity. Ultimately, organizations and researchers need to determine if detecting more phishing emails outweighs the cost of potentially misclassifying legitimate ones.

References

Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment*, *91*(4), 340–345. https://doi.org/10.1080/00223890902935878

Bravo-Lillo, C., Komanduri, S., Cranor, L. F., Reeder, R. W., Sleeper, M., Downs, J., & Schechter, S. (2013). *Your attention please: Designing security-decision UIs to make genuine risks harder to ignore* [Conference session]. Proceedings of the Ninth Symposium on Usable Privacy and Security, Newcastle, United Kingdom.

Byrne, Z. S., Dvorak, K. J., Peters, J. M., Ray, I., Howe, A., & Sanchez, D. (2016). From the user's perspective: Perceptions of risk relative to benefit associated with using the Internet. *Computers in Human Behavior*, *59*, 456–468. https://doi.org/10.1016/j.chb.2016.02.024

Cain, A. A., Edwards, M. E., & Still, J. D. (2018). An exploratory study of cyber hygiene behaviors and knowledge. *Journal of Information Security and Applications*, *42*, 36–45.

Canfield, C. I., & Fischhoff, B. (2018). Setting priorities in behavioral interventions: An application to reducing phishing risk. *Risk Analysis*, *38*(4), 826–838. https://doi.org/10.1111/risa.12917

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying phishing susceptibility for detection and behavior decisions. *Human Factors*, *58*(8), 1158–1172. https://doi.org/10.1177/0018720816665025

Canfield, C. I., Fischhoff, B., & Davis, A. (2019). Better beware: Comparing metacognition for phishing and legitimate emails. *Metacognition and Learning*, *14*(3), 343–362. https://doi.org/10.1007/s11409-019-09197-5

Drake, C. E., Oliver, J. J., & Koontz, E. J. (2004). *Anatomy of a phishing email* [Conference session]. Proceedings of the Conference on Email and Anti-Spam, Mountain View, California, United States.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Ferguson, A. (2005). *Fostering e-mail security awareness: The west point carronade*. https://er.educause.edu/articles/2005/1/fostering-email-security-awareness-the-west-point-carronade

Forget, A., Pearman, S., Thomas, J., Acquisti, A., Christin, N., Cranor, L. F., Egelman, S., Harbach, M., & Telang, R. (2016). Do or do not, there is no try: User engagement may not improve security outcomes. In *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, pp. 97–111.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, *19*(4), 25–42. https://doi.org/10.1257/089533005775196732

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Peninsula Pub.

Herley, C. (2009). So long, and no thanks for the externalities: the rational rejection of security advice by users. *Security*, 133–144. https://doi.org/10.1145/1719030.1719050

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. University of California, Institute of Personality and Social Research.

Kumaraguru, P., Rhee, Y., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Protecting people from phishing: The design and evaluation of an embedded training email system. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 905–914). ACM.

Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work (Reading, Mass.)*, *41*(Suppl 1), 3549–3552. https://doi.org/10.3233/WOR-2012-1054-3549

McBride, M., Carter, L., & Warkentin, M. (2012). *Exploring the role of individual employee characteristics and personality on employee compliance with cybersecurity policies*. RTI International-Institute for Homeland Security Solutions.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15583–15587. https://doi.org/10.1073/pnas.0903620106

Pegoraro, R. (2019, August 9). We keep falling for phishing emails, and Google just revealed why. *Fast Company*. https://www.fastcompany.com/90387855/we-keep-falling-for-phishing-emails-and-google-just-revealed-why

Sarno, D. M., Lewis, J. E., Bohil, C. J., & Neider, M. B. (2020). Which phish is on the hook?: Phishing vulnerability for older versus younger adults. *The Journal of Human Factors and Ergonomics Society*, *62*(5), 704–717. https://doi.org/10.1177/0018720819855570

Sawyer, B. D., Finomore, V. S., Funke, G. J., Mancuso, V. F., Miller, B., Warm, J., & Hancock, P. A. (2015). Evaluating cybersecurity vulnerabilities with the email testbed: Effects of training. *Proceedings 19th Triennial Congress of the IEA* (Vol. *9*, p. 14).

Sawyer, B. D., & Hancock, P. A. (2018). Hacking the human: The prevalence paradox in cybersecurity. *Human Factors*, *60*(5), 597–609. https://doi.org/10.1177/0018720818780472

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*(4), 207–218. https://doi.org/10.1111/j.1467-9280.1992.tb00029.x

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L., & Downs, J. (2011). Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. *Conference on Human Factors in Computing Systems Proceedings*, 373–382.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-phishing phil: The design and evaluation of a game that teaches people not to fall for phish. *Proceedings of the 3rd symposium on usable privacy and security* (pp. 88–99).

Shropshire, J., Warkentin, M., Johnston, A., & Schmidt, M. (2006). Personality and IT security: An application of the five-factor model. *AMCIS 2006 Proceedings*, 415.

Shropshire, J., Warkentin, M., & Sharma, S. (2015). Personality, attitudes, and intentions: Predicting initial adoption of information security behavior. *Computers & Security*, *49*, 177–191. https://doi.org/10.1016/j.cose.2015.01.002

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Urrico, R. (2019, August 7). Cybercrime reports: The costs & effects on financial institutions. *Credit Union Times*. https://www.cutimes.com/2019/08/07/cybercrime-reports-the-costs-effects-on-financial-institutions/?slreturn=20190720095129

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, *45*(8), 1146–1166.

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H. R. (2011). Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, *51*(3), 576–586. https://doi.org/10.1016/j.dss.2011.03.002

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Xiong, A., Proctor, R. W., Yang, W., & Li, N. (2019). Embedding training within warnings improves skills of identifying phishing webpages. *Human Factors*, *61*(4), 577–595. https://doi.org/10.1177/0018720818810942

Received August 21, 2020
Revision received September 17, 2021
Accepted September 29, 2021 ■