Which Phish Is on the Hook? Phishing Vulnerability for Older Versus Younger Adults

Dawn M. Sarno, University of Central Florida, Orlando, USA,
Joanna E. Lewis, University of Northern Colorado, Greeley, USA, and
Corey J. Bohil, **Mark B. Neider**, University of Central Florida, Orlando, USA

Objective: To determine if there are age-related differences in phishing vulnerability and if those differences exist under various task conditions (e.g., framing and time pressure).

Background: Previous research suggests that older adults may be a vulnerable population to phishing attacks. Most research exploring age differences has used limiting designs, including retrospective self-report measures and restricted email sets.

Method: The present studies explored how older and younger adults classify a diverse sample of 100 legitimate and phishing emails. In Experiment 1, participants rated the emails as either spam or not spam. Experiment 2 explored how framing would alter the results when participants rated emails as safe or not safe. In Experiment 3, participants performed the same task as Experiment 1, but were put under time pressure.

Results: No age differences were observed in overall classification accuracy across the three experiments, rather all participants exhibited poor performance (20%–30% errors). Older adults took significantly longer to make classifications and were more liberal in classifying emails as spam or not safe. Time pressure seemed to remove this bias but did not influence overall accuracy.

Conclusion: Older adults appear to be more cautious when classifying emails. However, being extra careful may come at the cost of classification speed and does not seem to improve accuracy.

Application: Age demographics should be considered in the implementation of a cyber-training methodology. Younger adults may be less vigilant against cyber threats than initially predicted; older adults might be less prone to deception when given unlimited time to respond.

Keywords: signal-detection theory, cybersecurity, decision making, designing for the elderly, age

Address correspondence to Mark B. Neider, Department of Psychology, University of Central Florida, 4111 Pictor Lane, Psychology Bldg 99 Ste. 320, Orlando, FL 32816-1390, USA; e-mail: mark.neider@ucf.edu.

HUMAN FACTORS
Vol. 62, No. 5, August 2020, pp. 704–717
DOI: 10.1177/0018720819855570
Article reuse guidelines: sagepub.com/journals-permissions

Cybersecurity threats have become a pervasive concern within our day-to-day lives. These types of attacks range from massive hacks on corporations that compromise thousands of employees' personal information to smaller scale infiltrations where an individual's identity is stolen (Elkind, 2015; Yang & Jakakumar, 2014). Both often arise from phishing attacks, which can be defined as "an email scam that attempts to defraud people of their personal information" (Drake, Oliver, & Koontz, 2004, p. 1). Despite their prevalence, there has been limited research investigating how different demographic groups may be more (or less) vulnerable to phishing attacks. One group that may be especially susceptible to cyber threats are older adults. Older adults report receiving the same amount of spam (i.e., junk mail) as younger adults despite their lower overall computer use and are more likely to make a purchase as a result of engaging with fraudulent emails (Grimes, Hough, & Signorella, 2007; Kircanski et al., 2018). Furthermore, older adults appear to be particularly vulnerable to the types of deception, upon which phishing attempts rely, such as when presented with the opportunity for financial gain or with emails from perceived authority figures (e.g., lawyers and politicians) (Oliveira et al., 2017). Older adults often suffer real consequences from these types of fraudulent attacks. Recent studies of financial exploitation among older adults put prevalence rates at nearly 5% (see Lichtenberg, 2016, for a review) with annual estimated losses to victims of over $3 billion (National Committee for the Prevention of Elder Abuse, 2011). Despite the fact that older adults have been identified as a vulnerable population, few studies have explored their susceptibility to fraudulent emails.

There are numerous reasons why older adults may be more vulnerable to fraudulent email

attacks compared to their younger adult counterparts. Among these are changes that occur over the course of normal healthy aging. More specifically, consistent cognitive declines are typically observed after 65 years of age in a variety of tasks involving attention, decision making, and working memory, with impairments magnified in the presence of time pressure (for a review, see Salthouse, 2010). After the age of 60, most individuals start seeing decrements in at least one of five primary mental abilities: verbal meaning, spatial orientation, inductive reasoning, numerical ability, and world fluency (Schaie, 1994). All of these areas of cognitive decline may be involved in a complex decision-making task such as detecting phishing emails. For instance, verbal meaning and inductive reasoning are critical components in understanding if the content of an email was from a legitimate source or a phisher. However, deception detection is a crucial ability that may deteriorate across the lifespan (e.g., Gavett et al., 2017).

The ability to reason regarding the authenticity of an email may also be influenced by deception-specific factors that change during aging: susceptibility to deception, increased vulnerability to emotional claims regarding fraud, and an overall increase in trust (Kircanski et al., 2018; Li & Fung, 2013; Ruffman, Murray, Halberstadt, & Vater, 2012). Relatedly, focusing on task relevant information is often more challenging for older adults due to declines in inhibitory control processes (Hasher & Zacks, 1988). Many cybersecurity threats are detected by relevant information contained within the email; however, scammers, or "phishers," often include other extraneous information to lure individuals to engage with the email (Drake et al., 2004). Thus, older adults may be distracted by erroneous information in fraudulent emails, such as the emotional plea of a spoofed family member (Oberauer, 2001). In addition, as we age, we experience a shift from focusing our attention on negative information to more positive information (e.g., Carstensen & Mikels, 2005; Mather & Carstensen, 2005). As such, older adults' tendency to focus on more positive information may lead them to ignore the negative information necessary to determine that an email is fraudulent. For instance, older adults may focus on the potential to win $10,000 dollars from a fraudulent sweepstakes more so than the potential negative consequences of sharing their social security number through an email link, particularly if in financial need. With cognitive decline in all these crucial areas, older adults are an inherently vulnerable target population to fraudulent attacks.

Previous cyber experience can also influence the decision-making process in older adults. When determining the authenticity of an email, users are often required to use their prior experience with, or knowledge of, fraudulent emails. However, older adults often are considered to have a lack of computer experience and computer literacy (Czaja, 1996; Czaja et al., 2006; Gatto & Tak, 2008), requiring various aids to help them in computer tasks (Charness & Boot, 2009; Hawthorn, 2000). Even though the current cohort of older adults may have more computer experience than 20 years ago because technology is ever evolving, there seems to be a consistent gap between older adults and modern technology adoption (Charness & Boot, 2009; Lee & Coughlin, 2015; Wu, Damnée, Kerhervé, Ware, & Rigaud, 2015). Thus, although older adults who are aged 65–75 now may have more technological experience than older adults who are 85 and older, all older adults are likely to have less modern technology experience than the current cohort of younger adults. This lack of experience may produce challenges for older adults when classifying emails.

Limited research has explored susceptible populations in cybersecurity contexts. Grimes and colleagues (2007) collected self-reported data regarding attitudes and experience with spam across three age groups (i.e., college-age, working-age, and retirement-age). Older adults were more likely to report making a purchase as a result of receiving spam emails despite their lower overall computer use. As these findings are based on self-report data, they are limited in their conclusions. It is possible that older adults were just more likely to report their exploitations, but in reality, all groups fell victim to scammers equally. Nonetheless, the results still suggest that continued research on age differences in cyber contexts is important for developing useful intervention strategies.

Sheng, Holbrook, Kumaraguru, Cranor, and Downs (2010) completed a similar study investigating how various demographic groups responded to an online email role-play survey. Interestingly, younger adults (18–25 years) were found to be more susceptible to phishing attacks than any other age group prior to training. Both studies are limited because they either used retrospective, self-report measures or narrow sets of emails. This limits the generalizability of the studies because users are not actively engaging with emails, are required to remember their previous experiences, and the restricted email sets may not accurately portray the types of emails users may engage with daily. Oliveira and colleagues (2017) directly examined the relationship between age and phishing attacks by periodically emailing participants artificial spam to their personal email. Older adults were found to be more likely to click on a link in a phishing email compared to younger adults. It is possible that the older adults' decision to interact with the phishing emails was simply due to larger interest in the content of the email rather than purely a misclassification. Specifically, researchers have shown that older adults enjoy emailing (Gatto & Tak, 2008) and this may have led to an increased curiosity regarding the email compared to younger adults, who may have been uninterested in the email. These findings highlight the importance of having a diverse set of emails to avoid some emails being more interesting to a specific age group. Gavett and colleagues (2017) have also explored phishing suspicion and aging, but within web pages. They found that older adults may be more suspicious of phishing, but that age differences may be related to other factors such as previous experience and education. Thus, further research exploring how users classify emails is required to draw more accurate conclusions regarding the detection of cyber threats across the lifespan.

These studies aim to explore if older adults may be more vulnerable to phishing emails than younger adults. Experiment 1 explored if older adults are more accurate than younger adults in detecting whether an email is spam (i.e., phishing). As framing has been shown to influence classifications (Tversky & Kahneman, 1989), even in cyber tasks (Carpenter, Zhu, & Kolimi, 2014; Sarno, Lewis, Bohil, Shoss, & Neider, 2017), Experiment 2 determined if the phrasing of the classification influences detection performance. Finally, since we know time pressure can impact older adults' performance (Salthouse, 2010), Experiment 3 incorporated a fixed time to view emails to evaluate whether time pressure affects an individual's ability to judge an email's authenticity. Previous research has suggested that signal-detection measures can be a useful tool for elucidating performance in a similar email classification task, as they allow for a dissociation between an individual's sensitivity to a signal and their inherent bias in responding (Canfield, Fischoff, & Davis, 2016). Thus, all three experiments used signal-detection measures to examine how individuals judged the emails.

EXPERIMENT 1

Method

Participants. A total of 10 (M_{age} = 19.10 years, 10 female) younger adult participants were recruited from the University of Central Florida in exchange for course credit. A total of 10 (M_{age} = 74.10 years, 7 female) older adult participants were recruited from the local community in exchange for $10/hr. All participants 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). Older adults were only included if they were 65 or older and exhibited normal cognitive decline, scoring a 23 or above on the Folstein Mini-Mental State Exam (Folstein, Folstein, & McHugh, 1975). This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at University of Central Florida.

Stimuli and procedure. The experiment was programmed and run in SR Research Ltd's Experiment Builder. Stimuli were 100 real emails, obtained either from the researchers' inbox/spam folders or through web searches. To determine how users classify emails in general, a diverse sample of emails were used. The emails included were selected because of their

TABLE 1: Email Content Categories

Banking (18%)	Emergency (1%)	Job ad (5%)	Shopping (24%)
Charity (1%)	Entertainment (8%)	Lawyer (e.g., will) (2%)	Social media (6%)
Contest (2%)	Family member (1%)	Scholarship (2%)	Taxes (2%)
Cloud storage (2%)	Food (2%)	Security (4%)	Travel (5%)
Email (e.g., Google) (7%)	Health insurance (2%)	Shipping (2%)	Utilities (4%)

Note. Email content was matched for both legitimate emails and phishing emails. For an example please see Figure 1. Values in the parentheses represent the percentage of emails that were in that category.

potential interest to both younger and older adults (e.g., banking, social media, and shipping; please see Table 1 for a full list).

Half of the emails used were fraudulent phishing emails and half were from a trusted source (i.e., legitimate; see Figure 1). Phishing emails also had consistent themes that could assist participants in their classifications. These themes have been identified in previous studies and were assessed for each of our emails (Bergholz et al., 2010; Chandrasekaran, Narayanan, & Upadhyaya, 2006; Drake et al., 2004). Accuracy for each phishing theme was assessed across experiments to determine which was the most predictive of correct detection. Threats to delete or suspend accounts were the most predictive, where implausible premises were the least predictive (see Table 2). Interestingly, older and younger adults did not differ in the order of how predictive each phishing theme was for correct detection. Phishing emails and legitimate emails were matched in content as best as possible. For instance, it is impossible to match a phishing attack if the sender claims they are the prince of Zimbabwe and they want to give you 1 million dollars. In these situations, approximate matches were found, for instance, a contest to win a house from HGTV. The experiment was presented on a Dell Professional P190S, 19-inch monitor at a resolution of 1280 × 1050 pixels with participants seated approximately 23 inches away, making the visual angle of the display roughly 36° by 29°.

All participants provided informed consent upon entering the lab were prescreened for vision and then were seated at a computer station for the remainder of the study. Participants were instructed to rate each email as spam or not spam via button press. Participants were instructed to classify the emails quickly while maintaining their accuracy and did not receive any feedback after their response. Participants were free to view each email for as long as they desired and proceeded to the next trial following their response. There were optional breaks between experimental blocks (25 trials each). Following the study, participants filled out a brief survey which included a demographics questionnaire, a modified Media Multitasking Index (Ophir, Nass, & Wagner, 2009) to assess their previous experience with technology, and the Cognitive Reflection Task to assess impulsivity (Frederick, 2005). The entire experiment took approximately an hour with breaks.

Results

To compare cyber performance across the lifespan, we conducted several analyses on accuracy, response times, and signal-detection measures using one-way (age: younger adults vs. older adults) between subjects' analyses of variance (ANOVAs). Due to several marginal effects, we also included posterior probabilities (p_{BIC}[H1|D]), which indicate a graded probability of whether the null hypothesis or alternative hypothesis is better supported by the present data (Masson, 2011). Specifically, a p_{BIC}(H1|D) < .50 indicates more support for the null hypothesis, whereas a p_{BIC}(H1|D) > .50 suggests more evidence for the alternative hypothesis. In addition, no significant or interpretable relationships were observed between the Media Multitasking Index, the Cognitive Reflection Task, or the email characteristics (i.e., content and phishing themes) and they were excluded from further analyses in all three experiments.

Accuracy and response times. A one-way between subjects' ANOVA was conducted on

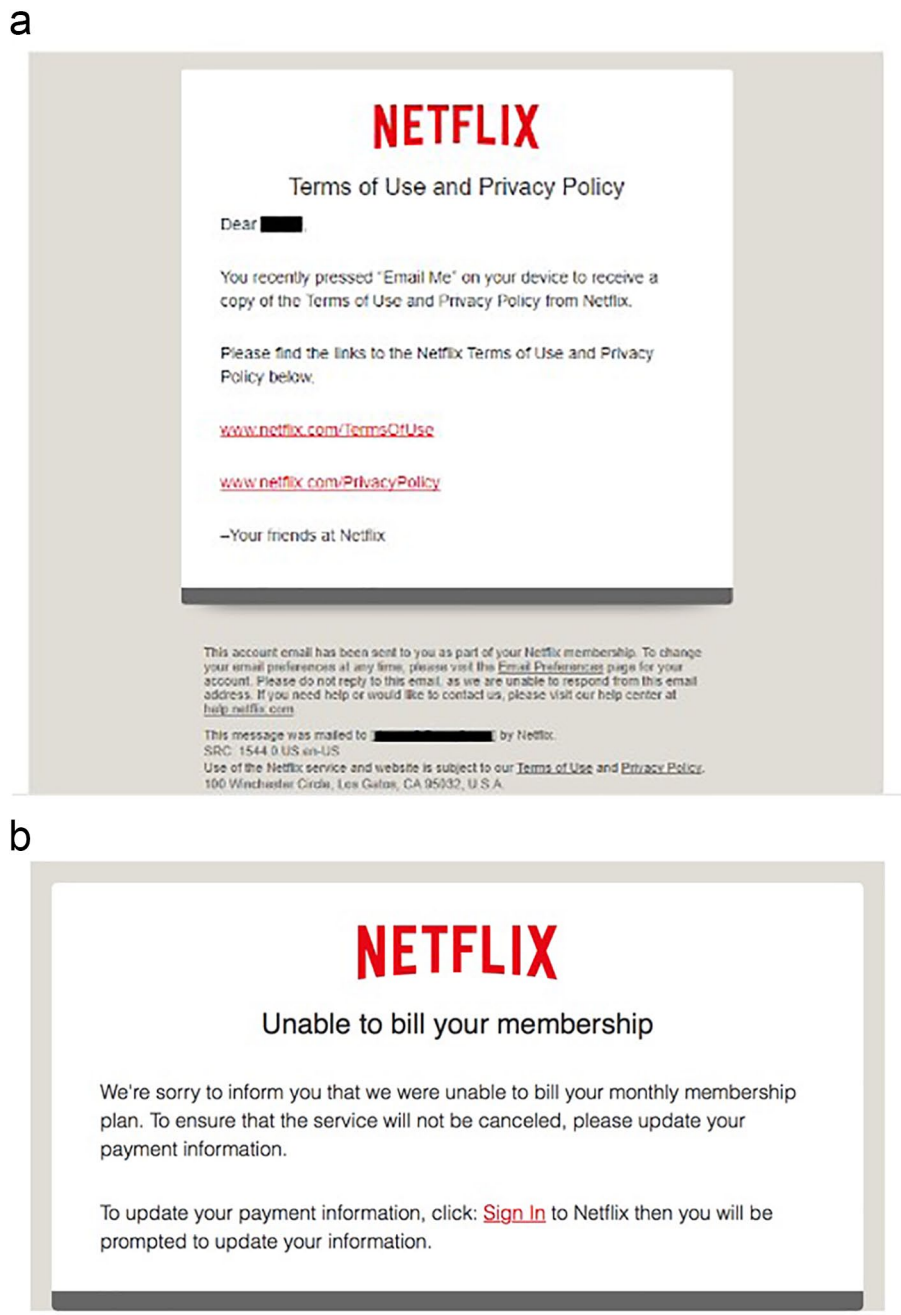

Figure 1. Example emails: (a) legitimate email and (b) phishing email.

both accuracy and response times to determine if there were any significant differences between younger and older adults in fraudulent email detection. The analyses revealed no significant effect of age on accuracy, $F(1,18) = 2.19$, $p = .156$, $\eta_p^2 = .11$, $p_{BIC}(H1|D) = .40$, indicating that younger and older adults did not differ in their ability to correctly detect fraudulent emails (see Table 3). However, we did find a significant effect of age on response time, $F(1,18) = 30.07$,

TABLE 2: Phishing Themes

1. Threats to delete/suspend accounts (37%)	5. Requiring quick response (55%)
2. Spelling and grammatical errors (55%)	6. Abnormal physical structure (53%)
3. Collecting personal information (48%)	7. Implausible premise (29%)
4. Abnormal language/phrasing (66%)	

Note. Themes are listed in order from the most predictive of correct detection to the least predictive across all three experiments. Numbers in the parentheses indicate the percentage of phishing emails that contained that theme.

TABLE 3: Experiment 1 Results

Age	Accuracy	Response Times (sec)	Hit Rate	False Alarms	Response Bias (β)	Sensitivity (*d'*)
Younger adults	0.67 (0.03)	26.77	0.60 (0.05)	0.26 (0.03)	1.26 (0.17)	0.94 (0.15)
Older adults	0.72 (0.02)	8.64	0.82 (0.02)	0.39 (0.04)	0.72 (0.08)	1.27 (0.13)

Note. Values displayed represent the group mean for each age group. Parenthetical values represent the standard error of the mean.

$p < .001$, $\eta_p^2 = .63$, $p_{BIC}(H1|D) = .99$, with older adults taking significantly longer to respond than younger adults (see Table 3). Taken together, these results suggest that although older adults can perform as accurately as younger adults, it comes at the cost of time. In the case of the current data, older adults took nearly 18 seconds longer to reach their judgments than younger adults. It is also worth noting that both our age groups had surprisingly low classification accuracy (ranging from 28% to 33% error rates), suggesting that while our age groups did not differ significantly from one another, they were both fairly poor at the task.

Signal-detection measures. Signal-detection measures were analyzed to fully understand how both age groups made their decisions regarding the emails. Four main measures were examined, hit rate, false alarms, response bias (β), and sensitivity (*d'*) (Green & Swets, 1966/1988). Each measure was submitted to a one-way between-subject ANOVA to determine if there were any age-related differences. Interestingly, older adults had a significantly higher hit rate than younger adults (see Table 3), $F(1,18) = 20.32$, $p < .001$, 1 $\eta_p^2 = .53$, $p_{BIC}(H1|D) = .99$, indicating that older adults were able to correctly detect spam emails at a higher rate than younger adults. However, older adults' increased detection of spam emails came at the cost of significantly more false alarms, $F(1,18) = 6.06$, $p = .024$, $\eta_p^2 = .25$, $p_{BIC}(H1|D) = .80$. This pattern of higher hit and false alarm rates for older adults is born out in the broader signal-detection measure of β or response bias. As it relates to this study, response bias scores above 1 can be considered more conservative for categorizing an email as spam and scores below 1 as more liberal for categorizing an email as spam (Green & Swets, 1966/1988). Here, older and younger adults had significantly different response biases, $F(1,18) = 8.50$, $p = .009$, $\eta_p^2 = .32$, $p_{BIC}(H1|D) = .91$, where older adults were more liberal in classifying an email as spam and younger adults more conservative in their spam classifications. There were no significant differences in younger and older sensitivities or *d'*, $F(1,18) = 2.68$, $p = .119$, $\eta_p^2 = .13$, $p_{BIC}(H1|D) = .47$, indicating that older adults and younger adults did not differ in their ability to distinguish authentic emails from spam emails. Combined, the response bias and sensitivity measures suggest that although older and younger adults did not differ in overall accuracy, older adults were more likely to classify an email as spam in general, perhaps out of an abundance of caution. In

TABLE 4: Experiment 2 Results

Age	Accuracy	Response Times (sec)	Hit Rate	False Alarms	Response Bias (β)	Sensitivity (*d'*)
Younger adults	0.67 (0.32)	13.04 (2.71)	0.82 (0.04)	0.47 (0.08)	0.83 (0.17)	1.10 (0.18)
Older adults	0.75 (0.02)	29.08 (1.91)	0.76 (0.03)	0.27 (0.05)	1.06 (0.13)	1.40 (0.09)

Note. Values displayed represent the group mean for each age group. Parenthetical values represent the standard error of the mean.

contrast, younger adults were biased toward classifying a given email as not spam.

EXPERIMENT 2

Experiment 1 indicated that there are no differences between younger and older adults in their overall classification accuracy but rather they approached the task differently (e.g., response biases). Numerous studies have shown that the framing of a task can affect how participants make decisions (e.g., Tversky & Kahneman, 1989). Notably, previous research has indicated that the framing of a classification task (e.g., spam, dangerous, and inauthentic) can affect an individual's classification specifically in email contexts (Sarno et al., 2017). In addition, the wording of cyber alerts (e.g., warning, caution, and hazard) has also been found to impact performance (Carpenter et al., 2014). Thus, Experiment 2 explored the same task but with a slightly different framing for the classification. Specifically, participants were asked to classify emails as safe or not safe.

Method

Participants. A total of 10 (M_{age} = 18.5 years, 9 female) younger adult participants were recruited from the University of Central Florida in exchange for course credit. A total of nine (M_{age} = 71.11 years, 6 female) older adult participants were recruited from the local community in exchange for $10/hr. All participants had normal or corrected-to-normal vision.

Stimuli and procedure. The experimental stimuli and procedure were identical to that of Experiment 1 with the following exception. To determine how participants classify emails, under various framings, Experiment 2 asked participants to rate emails as either safe or not safe via button press.

Results

Accuracy and response times. Accuracy and response times were again submitted to a one-way between-subjects ANOVA to examine the possible differences in email classifications between younger and older adults. The analyses indicated that there were no significant differences in accuracy between our two age groups (see Table 4), $F(1,18) = 3.98$, $p = .062$, $\eta_p^2 = .19$, $p_{BIC}(H1|D) = .64$. The results approached significance, suggesting that older adults may be better at determining if an email is safe. However, we are cautious to overinterpret this pattern, given the marginal significance coupled with a posterior probability close to chance. Similar to Experiment 1, both groups demonstrated surprisingly low overall classification accuracy (25%–33% errors). Again, we found that older adults took significantly longer to make their decisions than younger adults (see Table 4), $F(1,18) = 22.44$, $p < .001$, $\eta_p^2 = .57$, $p_{BIC}(H1|D) = .99$. Although we are cautious in interpreting non-significant results, this increase in time may be why older adults appear to be more accurate in their task.

Signal-Detection Measures. Signal-detection measures were calculated again to examine the differences in email classification between older and younger adults. Interestingly, with the classification as safe versus not safe, the only significant difference between younger and older adults was with false alarms (see Table 4), $F(1,18) = 4.55$, $p = .048$, $\eta_p^2 = .21$, $p_{BIC}(H1|D) = .68$, where younger adults had significantly more false alarms than older adults. It is important to note that because we flipped the phrasing of the classification this result is still consistent with the first experiment; younger adults classified more emails as safe (not spam in Experiment 1) and older adults classified more emails

as not safe (spam in Experiment 1). No other significant differences were found for the other signal-detection measures (p values > .176). Note that although the difference between the groups was not significant, younger adults' response biases (i.e., beta) appear to be liberal and older adults appear slightly more conservative (see Table 4). While the results from Experiment 2 were not significant, they suggest a similar pattern to Experiment 1 that older adults may be more biased toward classifying an email as spam or not safe. In addition, the results from Experiment 2 highlight the issue of wording and that slight modifications in the classification affect how users respond.

EXPERIMENT 3

In both Experiments 1 and 2, older adults tended to classify more emails as spam/not safe and take significantly longer to make their decision; however, there is substantial evidence that older adults tend to perform more poorly when under time pressure (Salthouse, 2010). In addition, previous research has established that perceived urgency is a key aspect of phishing attacks (Zielinska, Welk, Mayhorn, & Murphy-Hill, 2016). In the context of this study, it is possible the relationship between age and email classification performance might be dependent on the amount of time available to complete the task. That is, older adults may perform well at classifying emails as spam or not safe when they have unlimited time but may do worse when time is limited. More specifically, while older and younger adults both demonstrated similar accuracy in Experiments 1 and 2, it is possible that older adult performance might decrease under time pressure. Experiment 3 explored this possibility and presented the emails for a controlled time to both younger and older adults.

Method

Participants. A total of 15 (M_{age} = 19.60 years, 7 female) younger adult participants were recruited from the University of Central Florida in exchange for course credit. A total of 15 (M_{age} = 72.87 years, 9 female) older adult participants were recruited from the local community in exchange for $10/hr. All participants had normal or corrected-to-normal vision.

Stimuli and procedure. The stimuli and procedure in Experiment 3 were identical to that of Experiment 1 with the following exceptions. To explore how time constraints may affect performance across age groups, all participants viewed emails for 15 seconds. After 15 seconds, a display would appear instructing them to classify the email as spam or not spam. The length of time was determined using the average response time of the younger adults from the first two experiments. Since we controlled the length of the email presentation, response times here are representative of the time it took for participants to classify the email after the email was presented for 15 seconds. As with the other two experiments, participants were told to classify the emails quickly while maintaining accuracy. The goal was to see if older adults would still perform comparably to younger adults if they viewed the emails for the same amount of time.

Results

Accuracy and response times. As with the first two experiments, a one-way between-subjects ANOVA was performed on accuracy and response times to determine if there were any age differences in correctly detecting the presence of a spam email. The analyses indicated that under time pressure, there were no significant differences for the correct detection of spam emails between younger and older adults (see Table 5), $F(1,28) = 1.88$, $p = .181$, $\eta_p^2 = .06$, $p_{BIC}(H1|D) = .32$. There were no significant response time differences between younger and older adults for their email classification (see Table 5), $F(1,28) = 2.20$, $p = .11{,}050$, $\eta_p^2 = .07$, $p_{BIC}(H1|D) = .36$. As with the first two experiments, participants showed surprisingly high overall error rates (30%–36%), indicating just how poor individuals are at this task.

Signal-detection measures. Once again, signal-detection measures were derived for the two groups to illuminate possible differences in responses for their email classification task (see Table 5). Under time pressure, no significant differences were seen for any of the signal-detection measures between younger and older

TABLE 5: Experiment 3 Results

Age	Accuracy	Response Times (sec)	Hit Rate	False Alarms	Response Bias (β)	Sensitivity (*d*′)
Younger adults	0.64 (0.03)	1.70 (0.22)	0.60 (0.06)	0.32 (0.04)	1.07 (0.11)	0.84 (0.19)
Older adults	0.70 (0.03)	2.25 (0.30)	0.71 (0.04)	0.32 (0.04)	1.12 (0.19)	1.14 (0.15)

Note. Values displayed represent the group mean for each age group. Parenthetical values represent the standard error of the mean.

adults (p values > .115). Both groups were slightly conservative with their classifications. Overall, these results suggest that under time pressure older adults' performance becomes even more similar to that of younger adults and that their bias toward reporting an email as spam disappears.

DISCUSSION

Based on both the cognitive aging and cybersecurity literatures, there is reason to believe that older adults are a particularly susceptible population to phishing emails. Most of these studies in the cybersecurity domain, however, have used retrospective, self-reported designs or limited sets of emails and are thus limited in their ability to generalize to daily email classifications. To remedy this, the present studies compared performance for older and younger adults in an email classification task with a wide variety of emails. Overall, we found that when under similar time constraints, older and younger adults do not differ in their ability to accurately classify emails, but rather given unlimited time older adults tend to be biased toward classifying an email as spam.

Our findings are somewhat contrary to previous work that has shown older adults to be less accurate in their detection of phishing emails (Grimes et al., 2007; Kircanski et al., 2018). In Experiment 1, older adults were just as accurate as younger adults in classifying emails as spam or not spam. Since previous work has demonstrated that the wording of the classification matters (Sarno et al., 2017), Experiment 2 explored whether our age groups would differ in performance if we asked them to classify the emails as safe or not safe. Experiment 2 again confirmed that older adults are no worse than younger adults at detecting fraudulent emails. However, in both Experiments 1 and 2, older adults took an average of 18 seconds longer than younger adults to make their classification. Experiment 3 (in which all participants had 15 seconds to view each email) determined that this increased response time was not directly related to performance for older adults; even under time pressure no significant differences were seen between older and younger adults in accuracy. The results of our study suggest that older and younger adults do not differ in their overall ability to detect fraudulent emails, at least given the emails we used in our task. It is interesting, however, to consider whether our findings are necessarily related to older adults performing better than expected, or younger adults performing more poorly than expected. Generally speaking, accuracies were much lower than desirable for both age groups (~66% for younger adults and 72% for older adults across all experiments, respectively). Although our initial motivation for the present studies was to explore age effects, the surprisingly low email classification accuracy is worth noting. Across all three of our experiments participants missed between 9 to 24 phishing emails on average. This means that many individuals missed more than half of the phishing emails presented. This finding is particularly concerning, given that users only need to fail to identify one fraudulent email to allow a phisher to steal their personal information or gain access to an organization's system; our participants were fooled by far more than one phishing email. Other studies have also found that younger adults are poor at detecting fraudulent emails (Cain, Edwards, & Still, 2018), especially under low phishing prevalence (Sawyer & Hancock, 2018), and even after training (Ferguson, 2005; Kumaraguru et al., 2007; Mayhorn & Nyeste, 2012). Overall, our results demonstrate

that poor detection exists across the lifespan and highlights the importance of continued efforts to improve detection regardless of age.

Although accuracies were generally similar across age groups, our signal-detection measures allowed us to examine differences between older and younger adults in email classifications at a more detailed level. Overall, these measures indicated that older adults are more suspicious of all emails and biased toward classifying them as spam or not safe. In Experiment 1, this bias resulted in older adults classifying more emails as spam, and although this resulted in more accurate detection of spam emails (i.e., hits), it also produced an increase in incorrect classifications of legitimate emails as spam (i.e., false alarms). These results are interesting because they converge with previous research, suggesting that younger adults are actually poorer at detecting spam emails (Kumaraguru et al., 2007); in Experiment 1, younger adults were indeed biased toward classifying an email as not spam. In addition, although older adults were better at detecting spam emails, their bias might lead to them missing out on real (and possibly important) emails as a consequence; tradeoffs in response biases come with a cost. Experiment 2 found similar results; younger adults demonstrated more false alarms by incorrectly classifying more not safe emails as safe compared to their older adult counterparts. Furthermore, although there were no significant differences, raw response bias (i.e., beta) values indicated that older adults were more biased toward categorizing an email as not safe whereas younger adults were more biased toward saying an email was safe. Combined with Experiment 1, these results suggest that older adults are warier of fraudulent emails and are biased toward classifying emails as such. This inference is consistent with the finding that older adults often develop strategies to compensate for cognitive decline (e.g., Cabeza, Anderson, Locantore, & McIntosh, 2002). In this study, that strategy may be to mitigate risk through cautious decision making. This strategy is also consistent with recent research that suggests that older adults engage in more secure cyber behaviors than younger adults (Cain et al., 2018). This finding is surprising, given that older adults may be developing a strategy that does not aid them in their task. Specifically, when they are under time pressure and cannot use this strategy, their accuracy does not decrease. This indicates that older adults may be overly cautious without exhibiting any performance benefits. It is reasonable to surmise that older adults are aware that they are potentially vulnerable, possibly due to normal cognitive decline or previous fraud experience, resulting in a more cautious response profile. This is consistent with findings that show older adults are more cautious in initial perceptions of risk (Rolison, Hanoch, & Wood, 2012) and avoid negative outcomes (Frank & Kong, 2008) in various decision-making tasks. Older adults have also been shown to exhibit more task-related interference (e.g., mind wandering about the task) than younger adults, and this is more exacerbated when older adults' age stereotype threats are activated (Jordano & Touron, 2017). Age stereotype threats do not always directly impact task performance, as in our study. Studies that have explored stereotype threat in hazard perception while driving show that while age stereotype threats do not impact performance, they do negatively impact confidence (Chapman, Sargent-Cox, Horswill, & Anstey, 2016). Our older adults bias shift may be due to something similar, possibly due to a lack of confidence with technology or email. Although we did not find any meaningful differences in technology experience (from the MMI) between our age groups, it is possible that our older adults were less confident in their email abilities, even if this lack of confidence is unfounded. Though subtle, this difference between older and younger adult performance is one that warrants further scrutiny, as there might be circumstances in which such strategies are more difficult to apply, such as when emails are more personally relevant or situational task demands requiring even more rapid classifications than were explored in our experiments.

Time pressure is a task demand that has been a well-documented inhibitor of performance in older populations (Salthouse, 2010). Surprisingly, we found no differences in any of our measures between the younger and older adults when the time presentation of the email was limited. This finding suggests at least two possibilities:

(a) time pressure is not particularly detrimental to older adults in this sort of email classification task, or (b) the email classification task could still be comfortably completed by the older adult participants in the time provided. Further studies could empirically distinguish between these possibilities.

Broadly speaking, the absence of general age effects on overall classification accuracy in our studies were somewhat surprising. However, age-related differences are not ubiquitous across all processing domains. For instance, older adults do not typically demonstrate deficits in implicit memory tasks (Light & Singh, 1987), and it is possible that judgments regarding email authenticity rely more heavily on implicit mechanisms. Specifically, classifying emails may be consistent with a procedural memory task and not involve the explicit retrieval of previous experiences with email. In addition, within the context of fraud, emotional arousal has not been found to increase vulnerability to phishing attacks in older adults compared to younger adults' emotional arousal (Kircanski et al., 2018). Phishing attacks often prey on people's emotions (e.g., someone was robbed at gunpoint and needs $200), and previous research indicates that older adults are especially susceptible to these emotional appeals (Kircanski et al., 2018; Wang, Herath, Chen, Vishwanath, & Rao, 2012). As it happens, older adults may be more resilient in this context than initially predicted. This finding adds to our understanding of the technological abilities of current cohort of older adults. As Charness and Boot (2009) suggested, the current cohort of older adults may be more comfortable with technology than previous generations. However, older adults still may vary from younger adults in their utilization of technology. Specifically, older adults may be more cautious when evaluating emails.

It is important to note that all three of our experiments have some limitations that must be considered when interpreting the data, both in terms of age-related differences and in terms of broader demographic populations. First, it is possible we did not observe age-related differences in overall accuracy because of our older adult sample. Our older adult participants were recruited from a rather active local learning community, and it is possible that another older adult sample may perform drastically differently. High-performing older adults are often able to compensate for typical age-related decline through a plastic reorganization of their neurocognitive networks (Cabeza et al., 2002). Our sample may be representative of this, as our older adults appeared to demonstrate a strategy of classifying more emails as spam or not safe, although it is unclear if this strategy results in any meaningful performance benefits. Second, our studies did not include a middle-age group. Previous research has shown that both younger and older adults may both be susceptible populations (Ebner et al., 2018; Grimes et al., 2007; Kumaraguru et al., 2007; Sheng et al., 2010). Indeed, our data do not contradict this finding, as both age groups demonstrated relatively poor overall accuracy in all three experiments. However, it is possible that there may be an inverted U-shaped relationship in regard to email classification performance and age, such that both younger and older adults may be poor at classifying emails with middle-aged adults being the best at the task. In other words, while our study might tell an important part of the story relating normal aging to phishing vulnerability, middle-age adults might represent a yet-to-be-written chapter in the tale. In addition, our sample sizes may be underpowered for detecting small differences across experimental factors. We also did not control for various age differences (e.g., socio-economic background, employment history). This lack of control may have influenced our results; however, it is unlikely that these types of differences are the driving force behind our findings. Finally, although it is unlikely that differences in compensation produced our results, it is possible that older adults took the task more seriously because they were financially compensated rather than compensated with course credit. Broadly, it would be helpful for future research to explore a more diverse sample of older adults and also examine a middle-age group.

Overall, the present studies made progress toward understanding the susceptibility of individuals to phishing emails across the lifespan.

Experiment 1 determined that younger and older adults do not differ in their overall accuracy for the task, but do differ in the patterns of their responses. Specifically, older adults were more biased toward classifying emails as spam; younger adults were biased toward classifying emails as not spam. Experiment 2 found supporting results with a different classification, where older adults did not differ in overall accuracy, but in their bias toward classifying an email as not safe. Our third and final experiment demonstrated that under time pressure older adults perform more comparably to younger adults in both accuracy and in their decision profiles. Our experiments suggest that understanding vulnerability to cyber threats across the lifespan may not be as simple as articulating the likelihood that a given individual may successfully classify an email, particularly given the fairly low accuracies we observed across both age groups (one wrong choice might be enough to compromise an individual's security). Instead, subtle indicators, such as the general classification strategies (e.g., response biases) that individuals employ, and the circumstances under which those strategies can or cannot be engaged (e.g., time pressure and task framing), may better explain performance and shed light on avenues that can improve the efficacy of the present cybersecurity training and interventions.

KEY POINTS

- Given unlimited time to make their classifications, older adults exhibit a bias toward classifying emails as "spam" or "not safe."
- Under time constraints, older and younger adults do not differ in their classification of emails.
- Together, the results suggest that older adults are more cautious with emails, but they cannot exhibit the same level of cautiousness when under time pressure. Interestingly, this strategy does not seem to impact their overall accuracy in phishing identification.
- Accuracy is poor for both older and younger adults with both groups missing 20%–30% of emails.

REFERENCES

Bergholz, A., De Beer, J., Glahn, S., Moens, M. F., PaaB, G., & Strobel, S. (2010). New filtering approaches for phishing email. *Journal of Computer Security*, *18*, 7–35.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging gracefully: Compensatory brain activity in high-performing older adults. *NeuroImage*, *17*, 1394–1402.

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

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

Carstensen, L. L., & Mikels, J. A. (2005). At the intersection of emotion and cognition: Aging and the positivity effect. *Current Directions in Psychological Science*, *14*, 117–121.

Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006). Phishing email detection based on structural properties. In *Proceedings of the NYS Cyber Security Conference Annual Meeting*. Retrieved from https://pdfs.semanticscholar.org/d6e6/d07f47245974ebc9017c5295a6574286d8f1.pdf

Chapman, L., Sargent-Cox, K., Horswill, M. S., & Anstey, K. J. (2016). The impact of age stereotypes on older adults' hazard perception performance and driving confidence. *Journal of Applied Gerontology*, *35*, 642–652.

Charness, N., & Boot, W. R. (2009). Aging and information technology use: Potential and barriers. *Current Directions in Psychological Science*, *18*, 253–258.

Czaja, S. J. (1996). Aging and the acquisition of computer skills. In W. A. Rogers, A. D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 201–220). Mawwah, NJ: Lawrence Erlbaum Associates Inc.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging*, *21*, 333–352.

Drake, C. E., Oliver, J. J., & Koontz, E. J. (2004). Anatomy of a phishing email. In *Proceedings of the Conference on Email and Anti-Spam* (pp. 1–8). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.9431&rep=rep1&type=pdf.

Ebner, N. C., Ellis, D. M., Lin, T., Rocha, H. A., Yang, H., Dommaraju, S., . . . Oliveira, D. S. (2018). Uncovering susceptibility risk to online deception in aging. *The Journals of Gerontology: Series B, Psychological Sciences & Social Sciences*, 1–12. doi:10.1093/geronb/gby036

Elkind, P. (2015, July). Inside the hack of the century. *Fortune*. Retrieved from http://fortune.com/sony-hack-part-1/

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

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.

Frank, M. J., & Kong, L. (2008). Learning to avoid in older age. *Psychology and Aging*, *23*, 392–398.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Gatto, S. L., & Tak, S. H. (2008). Computer, internet, and e-mail use among older adults: Benefits and barriers. *Educational Gerontology*, *34*, 800–811.

Gavett, B. E., Zhao, R., John, S. E., Bussell, C. A., Roberts, J. R., & Yue, C. (2017). Phishing suspiciousness in older and younger adults: The role of executive functioning. *PLoS ONE*, *12*(2), e0171620.

Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing. (Original work published 1966)

Grimes, G. A., Hough, M. G., & Signorella, M. L. (2007). Email end users and spam: Relations of gender and age group to attitudes and actions. *Computers in Human Behavior*, *23*, 318–332.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower & G. H. Bower (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). San Diego, CA: Academic Press.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers*, *12*, 507–528.

Jordano, M. L., & Touron, D. R. (2017). Stereotype threat as a trigger of mind-wandering in older adults. *Psychology and Aging*, *32*, 307–313.

Kircanski, K., Notthoff, N., DeLiema, M., Samanez-Larkin, G. R., Shadel, D., Mottola, G., . . . Gotlib, I. H. (2018). Emotional arousal may increase susceptibility to fraud in older and younger adults. *Psychology and Aging*, *33*, 325–337.

Kumaraguru, P., Rhee, Y., Sheng, S., Hasan, S., Acquisti, A., Cranor, L. F., & Hong, J. (2007). Getting users to pay attention to anti-phishing: Evaluation of retention and transfer. In *Proceedings of the Anti-Phishing Working Group's Annual eCrime Researchers Summit* (pp. 70–81). New York, NY: ACM.

Lee, C., & Coughlin, J. F. (2015). Perspective: Older adults' adoption of technology: An integrated approach to identifying determinants and barriers. *Journal of Product Innovation Management*, *32*, 747–759.

Li, T., & Fung, H. H. (2013). Age differences in trust: An investigation across 38 countries. *The Journals of Gerontology: Series B, Psychological Sciences & Social Sciences*, *68*, 347–355.

Lichtenberg, P. A. (2016). Financial exploitation, financial capacity, and Alzheimer's disease. *American Psychologist*, *71*, 312–320.

Light, L. L., & Singh, A. (1987). Implicit and explicit memory in young and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 531–541.

Masson, M. E. (2011). A tutorial on practical Bayesian alternative to null-hypothesis significance testing. *Behavioral Research Methods*, *43*, 679–690.

Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in Cognitive Sciences*, *9*, 496–502.

Mayhorn, C. B., & Nyeste, P. G. (2012). Training users to counteract phishing. *Work*, *41*, 3549–3552.

National Committee for the Prevention of Elder Abuse. (2011). *The MetLife study of elder financial abuse: Crimes of occasion, desperation and predation against America's elders*. Westport, CT: Virginia Tech, MetLife Mature Market Institute.

Oberauer, K. (2001). Removing irrelevant information from working memory: A cognitive aging study with the modified Sternberg task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 948–957.

Oliveira, D., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., . . . Ebner, N. (2017). Dissecting spear phishing emails for older vs young adults: On the interplay of weapons of influence and life domains in predicting susceptibility to phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6412–6424). New York, NY: ACM.

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*, 15583–15587.

Rolison, J. J., Hanoch, Y., & Wood, S. (2012). Risky decision making in younger and older adults: The role of learning. *Psychology and Aging*, *27*, 129–140.

Ruffman, T., Murray, J., Halberstadt, J., & Vater, T. (2012). Age-related differences in deception. *Psychology and Aging*, *27*, 543–549.

Salthouse, T. A. (2010). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*, 754–760.

Sarno, D. M., Lewis, J. E., Bohil, C. J., Shoss, M. K., & Neider, M. B. (2017). Who are phishers luring? A demographic analysis of those susceptible to fake emails. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*, 1735–1739.

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

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304–313.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L., & Downs, J. (2010). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Conference on Human Factors in Computing Systems Proceedings* (pp. 373–382). New York, NY: ACM.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In B. Karpak and S. Zionts (Eds.), *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81–126). Berlin, Germany: Springer.

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, R. (2012). Phishing susceptibility: An investigation into the processing of a targeted spear phishing email. *IEEE Transactions on Profession Communication*, *55*, 345–362.

Wu, Y. H., Damnée, S., Kerhervé, H., Ware, C., & Rigaud, A. S. (2015). Bridging the digital divide in older adults: A study from an initiative to inform older adults about new technologies. *Clinical Interventions in Aging*, *10*, 193–200.

Yang, J. L., & Jakakumar, A. (2014, January). Target says up to 70 million more customers were hit by December data breach. *The Washington Post*. Retrieved from https://www.washingtonpost.com/business/economy/target-says-70-million-customers-were-hit-by-dec-data-breach-more-than-firstreported/2014/01/10/0ada1026-79fe-11e3-8963-b4b654bcc9b2_story.html

Zielinska, O., Welk, A., Mayhorn, C. B., & Murphy-Hill, E. (2016). The persuasive phish: Examining the social psychological principles hidden in phishing emails. In *Proceedings of the Symposium and Bootcamp on the Science of Security* (pp. 126). New York, NY: Association for Computing Machinery.

Dawn M. Sarno is currently a graduate student working toward completing her PhD in human factors and cognitive psychology at the University of Central Florida. She received her MA in human factors and applied psychology from the University of Central Florida in the spring of 2018.

Joanna E. Lewis is an assistant professor at the University of Northern Colorado in the Department of Psychological Sciences. She received her PhD in human factors and cognitive psychology from University of Central Florida in 2018.

Corey J. Bohil is an associate professor at the University of Central Florida in the Department of psychology. He received his PhD in cognitive psychology from the University of Texas at Austin in 2002.

Mark B. Neider is an associate professor at the University of Central Florida in the department of psychology. He received his PhD in cognitive/experimental psychology from Stony Brook University in 2006.

Date received: January 9, 2019
Date accepted: May 15, 2019