Chapter 6

Phishing Susceptibility Study

This chapter is joint work with Mandy Holbrook, Julie Downs, Lorrie Cranor, and Ponnurangam Kumaraguru. An earlier version of the content in this chapter was submitted to CHI 2010 [125].

Phishing attacks, in which scammers send emails and other messages to con victims into providing their login credentials and personal information, snare millions of victims each year [43]. A variety of efforts aim to combat phishing through law enforcement, automated detection, and end-user education. Researchers have studied why people fall for phishing attacks; however, little research has been done to study demographic factors in susceptibility to phishing. By determining which groups are most susceptible to phishing, we can determine how best to focus anti-phishing education.

In this paper, we present the results of our roleplay phishing study, administered to 1001 online survey respondents in order to study demographics and phishing susceptibility. The rest of the paper is organized as follows. In the next section, we present background and related work on why people fall for phishing. Then we describe the design of our experiment and present the results of our study, identifying several important demographic factors that affect phishing susceptibility and describing the effects of education in bridging these gaps. Finally we discuss the implications of our study for designing anti-phishing tools and improving public policy.

6.1 Background and related work

Research has shown that people are vulnerable to phishing for several reasons. First, people tend to judge a website's legitimacy by its "look and feel," which attackers can easily replicate [23]. Second, many users do not understand or trust the security indicators in web browsers [140]. Third, although some consumers are aware of phishing, this awareness does not reduce their vulnerability or provide useful strategies for identifying phishing attacks [26]. Fourth, the perceived severity of the consequences of phishing does not predict users' behavior [27].

6.1.1 Demographics and Phishing Susceptibility

To the best of our knowledge, there has been no study dedicated to understanding what demographic factors correlate with falling for phishing, and the effectiveness of educational interventions in bridging the demographic divide. We highlight here a few studies that have measured susceptibility to specific types of phishing attacks or have studied the effectiveness of anti-phishing education while reporting at least some data on gender and other demographic factors.

Jagatic et al. performed a spear phishing experiment at Indiana University to quantify how reliable social context would increase the success of a phishing attack. They launched an actual (but harmless) phishing attack targeting college students aged 18–24 years old by using information harvested from social networking sites. In their study of 487 participants, female students fell for 77% of the spear phishing attacks, while male students fell for 65% [53].

In a related study, Kumaraguru et al. conducted a real-world phishing study with 515 participants to study the long-term retention of PhishGuru anti-phishing training [69]. They did not find significant differences based on gender, but did find that participants in the 18-25 age group were consistently more vulnerable to phishing attacks. They also did not explain the reason behind this finding.

Finally, Kumaraguru et al. [71] conducted a study of 5182 Internet users measuring the effectiveness of Anti-Phishing Phil, an interactive game that teaches people not to fall for phish. They found that men were more likely to correctly distinguish phishing and legitimate websites than

women (75.5% correct vs. 64.4% correct). They collected only very coarse-grained information on the age of participants, but found that people under the age of 18 performed worse than those above 18.

Although past studies have found differences in phishing susceptibility based on gender and age, they generally did not collect enough information about study participants to isolate these variables from other potentially confounding factors. In addition, previous studies did not address why these demographic factors correlate with falling for phishing. In our paper, we address these research questions.

6.1.2 Susceptibility vs. Risk Behavior

The risk literature has shown reliable demographic differences in risk perceptions on various topics, with relatively oppressed groups (e.g., women, racial and ethnic minorities, and less wealthy people) perceiving more risk in the world around them [37, 129]. Such perceptions may be linked to these groups' experiences of a riskier world, perhaps due to lower degrees of control over risky processes. Age has also been linked to risky behavior, with adolescents tending to engage in riskier behaviors on average, perhaps as a function of their ongoing learning about the world around them [25,114]. Because real-world risk behaviors are complex and subject to such varied predictors as knowledge, goals, and benefits associated with what is perceived to be risky behavior, there have been relatively few studies with the power to assess multiple mediators of demographic effects on risky behavior. The current paper takes a specific, well-defined behavior as a context in which to identify content-specific factors that may explain effects of age, gender, and ethnic background.

6.1.3 Security User Education

Despite claims by some security and usability experts that user education about security does not work [48], there is evidence that well-designed user security education can be effective in the real world [67, 127]. Web-based training materials, contextual training, embedded training, and interactive games have all been shown to improve users' ability to avoid phishing attacks.

A number of organizations have developed online training materials to educate users about phishing [28, 32]. In a previous study, Kumaraguru et al. tested the effectiveness of some of these online materials and found that, while these materials could be improved, they are surprisingly effective when users actually read them [70].

Several studies have adopted a contextual training approach in which users are sent simulated phishing emails by the experimenters to test users' vulnerability to phishing attacks. At the end of the study, users are given materials that inform them about phishing attacks. This approach has been used in studies involving Indiana University students [53], West Point cadets [33], and New York State employees [104].

A related approach, called embedded training, teaches users about phishing during their regular use of email. This trainer sends phishing email to users and, if users click on phishing links, immediately presents an intervention designed to train them not to fall for phishing attacks. Kumaraguru et al. created several intervention designs based on learning sciences, and found that these interventions were more effective than standard security notices that companies email to their customers [68]. The researchers continued to refine the most successful intervention, a comic strip featuring a character named PhishGuru. A follow-up study showed that people were able to retain what they learned from this training [69].

Finally, Sheng et al. designed Anti-Phishing Phil, an online game that teaches users good habits to help them avoid phishing attacks. The researchers used learning science principles to design and iteratively refine the game. Their evaluation showed that participants who played the game were better able to identify fraudulent web sites compared to participants in other conditions [127].

We studied the effectiveness of several of these educational approaches in bridging the demographic divide. The materials we tested included a set of popular web-based training materials, Anti-Phishing Phil, a PhishGuru cartoon, and the combination of Anti-Phishing Phil and a PhishGuru cartoon.

6.2 Study Design

In this online study, participants provided demographic information, answered survey questions to assess their knowledge about phishing, and completed a roleplay task to assess their behavioral susceptibility to phishing, prior to receiving one of several possible forms of training. Participants then completed a second roleplay task to assess reductions in phishing susceptibility as well as any changes in participants' tendencies to be suspicious of legitimate emails. Participants were assigned randomly to a control condition or one of four experimental conditions. The conditions varied based on the type of training participants were exposed to (or no training). The ordering of the survey questions relative to the initial roleplay was also counterbalanced.

6.2.1 Recruitment

Participants were recruited through Amazon.com's Mechanical Turk (mTurk), a marketplace for work requiring human intelligence. In this online environment, requesters post tasks known as HITs (Human Intelligence Tasks), and workers are paid for completing these HITs. We offered to pay participants four dollars for those that qualified and twenty cents to those who did not. In total, 1001 participants qualified and completed the entire study as detailed in Table **??**.

To disqualify people who were hoping to earn money for completing the study without actually paying attention to the study tasks, we asked all participants a series of questions about an email message that discussed an upcoming meeting. We used two of these questions, both of which could be answered correctly by a careful reading of the email, to screen out those participants who were not paying attention to the email content. We also asked basic demographic questions (such as occupation and age) so that participants would not be able to easily identify qualifying questions.

6.2.2 Roleplay

Behavior was measured by performance in a roleplay task, with two equivalent exercises administered before and after training (the order of which was counterbalanced). This task is based

Table 6.1 Participant demographics by conditions. There is no statistical significant of demographics between different conditions.

Characteristics	**Control**	**Popular training materials**	**Anti-Phishing Phil**	**PhishGuru Cartoon**	**Anti-Phishing Phil with PhishGuru**
Sample Size	218	217	166	201	199
Gender					
Male	50%	48%	54%	45%	45%
Female	50%	52%	46%	55%	55%
Average age in years	30	30	29	30	31
Education					
High school or less	10%	8%	7%	7%	8%
Some college	33%	32%	37%	39%	36%
Competed 4-year college degree	29%	29%	30%	30%	27%
Some Post-graduate education	11%	12%	10%	6%	10%
Have master or Ph.d degree	17%	19%	16%	18%	17%
Percentage from US?	74%	71%	73%	78%	80%
Percentage student?	25%	26%	31%	20%	25%
Average years on the Internet	13	12	12	13	13
Average emails per day	44	44	32	57	43

on an established roleplay exercise that has been shown to have good internal and external validity. [27]. Participants were told to assume the role of Pat Jones, who works at Baton Rouge University and uses the email address patjones@bru.edu for both work and personal emails. Each roleplay showed participants fourteen images of emails along with context about Pat Jones that may help to interpret the emails. Images matched the participant's operating system and browser (e.g. Firefox on a Mac or Internet Explorer on a PC) so that all images and cues would be familiar

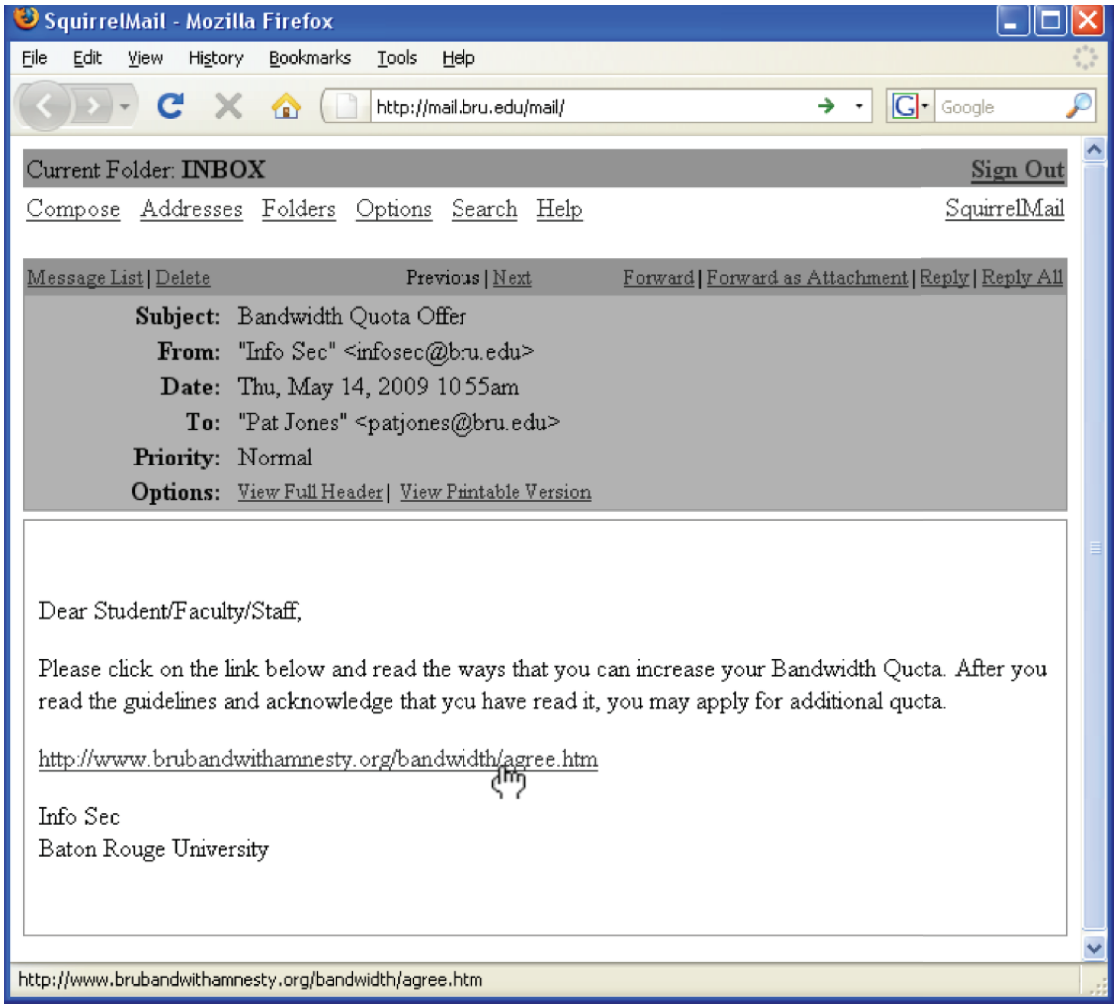

Figure 6.1 One of the emails that Pat encounters in her email box

to the participant. Participants were asked to indicate how they would handle the emails if they received them in their own email inbox, whether that would be forwarding the email to someone else, replying by email, or any other action from a list of responses generated through earlier qualitative work [26]. Table 6.2 details the list of possible responses.

The first email was created to familiarize the participant with the procedure. It was a short message from the same domain as Pat's email address. This message from the BRU Information Security Office announced a scavenger hunt for National Cyber Security month. The participants continued through the roleplay task by viewing a combination of real, phishing, malware and spam

email images. Table 6.3 lists a representative sample of the emails that Pat encounters in one of the roleplays.

Each email contained a link to a web page (e.g. Figure 6.1), shown with the mouse pointer positioned on the link and the actual URL destination displayed in the status bar, as it would be if users prepared to actually click on the link on their own computer. For individuals who indicated that they would click on the link or otherwise end up at the web page, an image of that web page was displayed. Each web page requested information to be entered and participants were asked to indicate if they would click on a link on the page, enter the requested information, bookmark the page, visit another related web page, close the website, or take other action. No matter what other actions the user indicated, those who said that they would enter the requested information

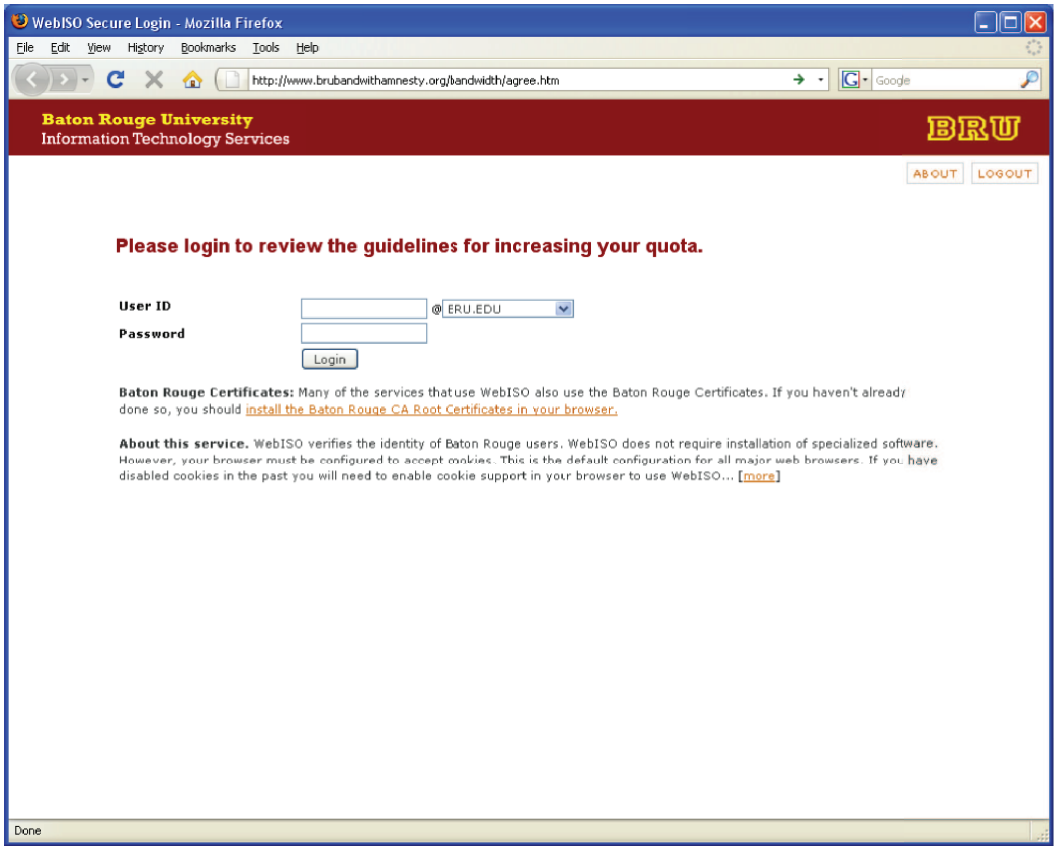

Figure 6.2 The corresponding website is shown when Pat chooses to the “click on the link” option in the email

Table 6.2 List of possible responses for emails in the role play survey

reply by email
contact the sender by phone or in person
forward the email to someone else
delete the email
keep, save or archive the email
click on the selected link in the email (the one that the browser hand is pointing to)
copy and paste the selected URL (the www address) from the email into a web browser, if a URL is selected in this email
type the selected URL into a web browser, if a URL is selected in this email
click on a different link in the email (please specify which link(s) you would click on)
Other (please specify)

were coded as having fallen for phishing or complied with a legitimate email, corresponding to the legitimacy of the email in question.

6.2.3 Education Materials

Participants were randomly assigned to the control condition, or to view one of four types of educational materials on ways to avoid falling for phishing attacks: a PhishGuru cartoon, Anti-Phishing Phil, several popular web-based training materials, and a combination of Anti-Phishing Phil plus a PhishGuru cartoon.

For popular web-based training, we selected three consumer oriented education materials from the first page of search results from google using keyword “phishing.” They are Microsoft Online safety [89], OnGuardOnline phishing tips [106], and National Consumer League Fraud tips [99]. In total, these materials have 3107 words, and would take roughly 15 minutes to complete reading with a scanning speed of 250 words per minute.

In the Anti-Phishing Phil conditions, participants were taken through three levels of the game and allowed to exit at any point. For the educational web page conditions, participants were asked at the end of each of three pages if they would like to read more information or move to the next

Table 6.3 A representative sample of emails in Pat's inbox from one of the roleplays

Email Subject	**Legitimacy**	**Relevant features of email and websites**
Earn Bonus Points #1	real	win a prize in an online scavenger hunt from BRU Information Security Office link: `https://www.bru.edu/iso/aware/ncsam/hunt/bonus`
Picture from last weekend's party	possible malware	impersonal greeting link:`http://picasaweb.google.com/stevewulitzer/Partypics/` actual url: `http://128.3.72.234/Partypics.jpg.exe`
No obligation bankruptcy consultation	spam	text of link: "Apply online now" actual url: `https://www.bankruptcylawyerfinder.com/...`
Bandwidth Quota Offer	phishing	misspelling in url and .org domain link `http://wwwbrubandwithamnesty.org/bandwidth/agree.htm` actual url: same
eBay Accounts Security	phishing	threatens account suspension link: `https://signin.eBay.com/ws/...` actual url: `http://www.security-validation-your-account.com/`
Your Amazon.com Order (#103-0607555-6895008)	real	problem with shipping link: `www.amazon.com/help/confirmation` actual url: same
Your eBay item sold!	real	text of link: "Send Invoice Now" actual url: `http://payments.ebay.com/eBayISAPI...`

part of the study. The PhishGuru conditions provided participants with one page of materials and then participants moved on to the next part of the study.

All participants who viewed any of the educational materials were asked how likely they would be to visit that specific educational tool again and how likely they would be to recommend it to someone else, on a scale ranging from 1 (not at all likely) to 7 (extremely likely).

6.2.4 Previous Experiences and Demographics

Along with asking participants extensive demographic related questions, all participants were asked to complete a series of questions about their online experiences, including questions about their choice of websites for recent purchases, their use of online banking and their prior exposure to anti-phishing educational materials. Participants also indicated any negative consequences such

as having information stolen or compromised in some way by entering it into a web site. Table 3 presents basic demographics of the sample.

6.2.5 Knowledge and Technical Background

Knowledge questions asked participants to choose the best definition for four terms related to computer security: 'cookie,' 'phishing,' 'spyware,' and 'virus.' Participants were given the same list of eight possible definitions to choose from for each, as well as choices to indicate lack of familiarity with the word. Each term had one correct answer on the list. The options included:

1. Something that protects your computer from unauthorized communication outside the network
2. Something that watches your computer and send that information over the Internet (*spyware*)
3. Something websites put on your computer so you don't have to type in the same information the next time you visit (*cookie*)
4. Something put on your computer without your permission, that changes the way your computer works (*virus*)
5. Email trying to trick you into giving your sensitive information to thieves (*phishing*)
6. Email trying to sell you something
7. Other software that can protect your computer
8. Other software that can hurt your computer
9. I have seen this word before but I don't know what it means for computers
10. I have never seen this word before
11. Decline to answer
12. Other (please specify)

To assess the level of their technology background, participants were asked if they had an Information Technology-related degree and any experience with programming languages, and they self-rated how technologically savvy they were on a scale ranging from 1(not at all savvy) to 7 (extremely savvy).

6.2.6 Risk Perceptions

To evaluate participants' risk perceptions, we presented them with a series of statements taken from the Domain-Specific Risk-Taking scale of adult populations (DOSPERT) [13], drawing on the categories of financial risk and health & safety risk. These questions asked participants to rate the risk associated with statements such as betting a day's income at the horses races and riding a motorcycle without a helmet, on a scale ranging from 1 (not at all risky) and 7 (extremely risk).

6.3 Results

6.3.1 Measuring User Performance

We measured participants' susceptibility to phishing by examining two kinds of errors before and after education interventions: falling for phish and false positives. A false positive is when a legitimate email or website is mistakenly judged as a phish and users refuse to follow the desired actions. Falling for phish occurs when a phishing email or website is incorrectly judged to be legitimate and users click on the email and submit information to the website. In our analysis, we consider falling for phishing as giving information to phishing websites, unlike previous studies that have used the close correlate of clicking on links in phishing emails. In previous studies and this one, around 90% of the participants who clicked on the phishing link end up giving information to the phishing website [68, 69]. We used giving information to phishing sites as a stricter measure for falling for phishing.

6.3.2 Regression Analysis

To explore factors that predict phishing susceptibility, we performed a multivariate linear regression. This section explains the steps we took to build the model and discusses the results from the linear regression.

We used factor analysis to reduce the dimensionality of our variables on participants' online experience (eight variables), participants' technical knowledge and experience (5 variables), and

Table 6.4 Regression analysis with parameters that are significant at p $<$ 0.01

Model Parameters	**Standardized Coefficients**
Ever seeing information on avoiding phish before this test	.19
Gender	.14
Age	-.12
Participants' technical knowledge	-.10
Risk perception of financial investment	-.08

participants' risk perception(12 variables). The factor analysis using principle component and varimax rotation reduced our list of variables from 40 to 22.

We then ran the regression predicting falling for phish from the 22 variables. In Table 6.4, we report variables that are statistically significant at $p <= 0.01$. Participants' degree of prior experience with phishing education significantly predicts how much phishing they will fall for (B = 0.189, $p <$0.01). Participants who have seen training material before (56.6% of total participants) fell for 2.4 phishing websites (40%), whereas those who have not seen training before fell for 3.6 phishing websites (60%), t = -9.02, $p < 0.001$. This factor had the most impact on phishing susceptibility, suggesting that exposure to education may play a larger role than other important factors.

Women fall for more phish than men (B = 0.140, t = 3.98, $p < 0.01$), an average of 53.1% phishing emails, compared to just 41% for men, $t(981)$ = -5.48, $p < 0.001$. We explore reasons for women's greater susceptibility in the next section.

Participants' age linearly predicts their susceptibility to phishing (B = -0.116, $p < 0.01$). An analysis of variance (ANOVA) comparing age groups found a significant overall effect, $F(4, 996) = 9.65$, $p < 0.001$, driven by participants aged 18 to 25 falling for phishing more than other age groups (all post-hoc tests comparing this group to other groups significant at $p<.01$; no other groups significantly different from one another).

Participants' self-rated knowledge about technology also significantly predicts whether they will fall for phishing. For each standard deviation higher the tech knowledge score, participants fell for [how many: raw number] fewer phish (3.6%).

Finally, participants' risk aversion, as measured by reactions to risks of financial investments, also predicts whether they will fall for phishing. The more risk-averse a participant is, the less likely he or she will fall for phish. For each standard deviation increase in their risk perception score, participants fell for [how many: raw number] fewer phish (2.8%).

6.3.3 Gender and Falling for Phish

In order to better understand why women appear to be more susceptible to phishing, we examined clicking on phish, giving information to phish, clicking on legitimate URLs, and giving information to legitimate websites with respect to gender.

We found that, before training, women were more likely than men to click on phishing links and enter information on phishing websites. On average, women clicked on 54.7% of phishing emails, compared to just 49% for men, t(981) = 2.69, $p < 0.01$. After clicking on a phishing link, women continued on to give information to the corresponding phishing website 97% of the time, compared to 84% for men, $t = 5.42$, $p < 0.001$. This further exacerbates the gender differences in clicking on links.

These results are consistent with previous real world phishing studies [67], where 52.3% of participants clicked on the simulated spear phishing emails they sent and subsequently 40.1% gave information to phishing sites. The similarity in our results suggested the validity of the roleplay survey instrument.

In an attempt to explain these gender effects, we did a mediation analysis using all the key predictors as potential mediators. Mediation analysis explains "how" an effect occurred by hypothesizing a causal sequence. The basic mediation model is a causal sequence in which the independent variable (X) causes the mediator(s) (M) which in turn causes the dependent variable (Y), therefore explaining how X had its' effect on Y [76, 77]. Mediational processes are common in basic and applied psychology.

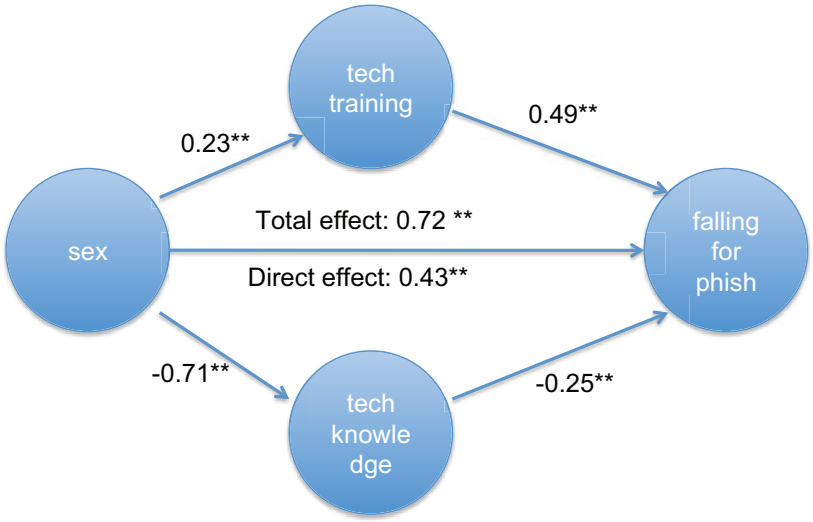

Figure 6.3 Mediation of the effect of gender on falling for phishing through participants' tech_knowledge and tech_training.

Table 6.5 Mediation analysis for gender. Each path is quantified with unstandardized regression coefficients. The direct effect of gender on phishing susceptibility (measured by number of phishing websites participants' giving information to) is calculated as total effect minus all the effect through each of the mediators, which is calculated as the product of coefficients in the paths.

	Point estimates	Percentile 95% CI	
		Lower	Upper
Total Effect of gender on falling for phishing	0.72		
Total effect of various mediators	0.29	0.18	0.42
tech knowledge	0.17	0.10	0.27
tech training	0.12	0.02	0.21

We used the multiple mediator model developed by Preacher and Hayers [63] for our mediation analysis. For gender, we used tech knowledge and tech training as mediators; our hypothesis is that women have less technical experience than men and therefore fall for phishing more. We report the mediation statistics in Table 6.5 and Figure 6.3 shows the results of the analysis, which are consistent with the hypothesis.

As shown in Figure 2, the effect of being female on falling for phishing drops from a total effect of 0.72, $p<0.01$, down to a direct effect of just 0.43, $p<0.01$. The difference between these effects represents the total indirect effect through the two mediators, with a point estimate of 0.29, and a 95% CI of 0.18 to 0.42 (see Table 6.5). Thus, women in our study have less technical training and have less technical knowledge than men, which appears to partially account for their greater susceptibility to phishing.

The mediation relationship is only partial, as the direct effect is still statistically significant. This partiality suggests that there are other factors that are not captured by our survey instruments; these factors might be explored in future work.

We included several other predictors that did not mediate this relationship. For example, women may fall for phishing more because they have fewer opportunities or are less motivated to learn about phishing. However, prior exposure to phishing education did not turn out to be significant mediator. In fact, in our sample, more women than men claimed to have seen phishing education before the study. Neither were income or education significant mediators for the effect of gender on phishing susceptibility.

Other factors that we did not measure might potentially explain the remaining tendency for women in our study to be more susceptible to phishing than men. Factors that may be worth further exploration include differences in the way men and women use the Internet, differences in the way men and women make trust decisions, and differences in the tendency of men and women to be cooperative or comply with instructions.

6.3.4 Age and Falling for Phish

As described above, people in the 18 – 25 age group were more likely to fall for phish than people of other ages. We used the multiple mediator model to determine why younger people are more frequently falling for phishing. We report the mediation statistics in Table 6.6 and Figure 6.4.

Taken as a set, participants' prior exposure to phishing, numbers of years on the Internet, financial risk perception, and education mediate the effect of age on falling for phishing. As can

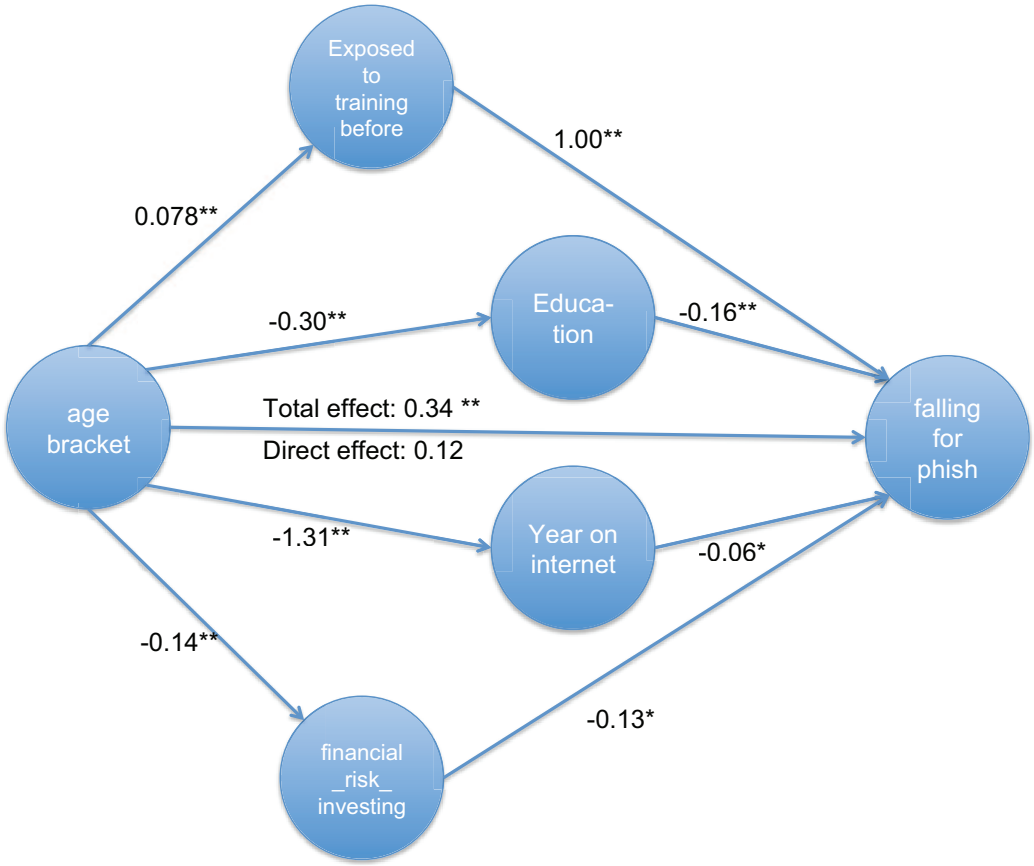

Figure 6.4 Mediating the effect of age with prior exposure to training, education, years on the Internet and risk perception for financial investment. Each of the paths is quantified with unstandardized regression coefficients.

Table 6.6 Total effect of age on falling for phishing and the effect of various mediators that are statistically significant at $p<0.01$.

	Point estimates	Percentile 95% CI	
		Lower	Upper
Total Effect of age on falling for phishing	0.34		
Total effect of various mediators	0.23	0.16	0.29
Prior exposure	0.08	0.04	0.12
years on Internet	0.08	0.03	0.13
education	0.05	0.02	0.08
risk financial investing	0.02	0.00	0.04

be seen in Figure 3, the total effect of age on falling for phishing fell from 0.34, $p < 0.01$, down to 0.12 (not significant). The difference between the total and direct effects is the total indirect effect through the four mediators, with a point estimate of 0.23, and a 95% CI of 0.16 to 0.29 (see Table 6). Because younger people have a lower level of education, fewer years of experience with the Internet, less exposure to training material, and less of an aversion to financial risks, they tend to be more susceptible to phishing.

6.3.5 Effects of Education

All of the training materials we tested reduced participants' tendency to click on phishing links in emails by 13-17 percentage points. There is no statistical difference between each education material, $F(3,779) = 1.28$, $p = 0.28$. The control group, which received no training during the study, showed no statistically significant improvement between the first and second roleplay. We also did not find the ordering of the knowledge survey affected the users' performance, so in our analysis we collapsed across orders.

All training materials reduced participants' tendency to enter information into phishing webpages by about 16-21 percentage points, and there is no statistically significant improvement for the control group.

Anti-Phishing Phil, Phishguru cartoon and Anti-Phishing Phil with Phishguru cartoon did not decrease participants' tendency to click on legitimate links and go to legitimate websites. However in the popular training condition, participants' tendency to click on legitimate links was slightly reduced, $t(216) = 2.01$, $p < 0.05$, suggesting that improvements in avoiding phish may merely reflect an avoidant strategy and not better detection.

Since the various education materials perform similarly in reducing people not falling for phishing, to study the effect of education in bridging the demographic gaps, we combined all the training conditions together.

Before the training, participants on average fell for 2.8 phishing websites out of 6, or 47%. After the training, this number is reduced to 1.6 out of 6, or 26%, a 21 percentage point improvement or 42% improvement. In terms of demographics, we found that women learned more than men

during the training about avoiding phishing links (t (767) = 5.63 $p < 0.01$); after training women and men perform equally well in not clicking on phishing links in emails $t(767)$ = -0.05, p = 0.96).

In entering information into phishing sites, women and men learned similarly, $t(767)$ = -1.51, p = 0.13). Women's higher rate of entering this information before the training carried over, and they still fell for more phish after the training than men, $t(767)$ = -4.22, $p< 0.001$).

Finally, people of different age groups learned similarly in training, leaving no statistical difference between age groups' performance increase, $F(4,778) = 1.66$, p = 0.16. Participants between ages 18 and 25 were the most susceptible group in pretest, and they remained more susceptible to phishing in posttest. People in different education groups also learned similarly, $F(5,763) = 1.4$, p = 0.20). We also found no significant effect for education or race.

We also analyzed the amount of time user spent on education materials. We found that users in the game conditions (Anti-phishing phil alone and Anti-Phihing Phil with Phishguru cartoon) spent the longest time, averaging 8.6 minutes. Although the popular education were designed to last as long as the game condition, users only spent 1.8 minutes on average (Table 6.7).

Table 6.7 Time user spent on education materials

Education Materials	**Estimate time user would spent**	**Average time user spent**
Popular training materials	12 min	1.80 min (SD = 2.09)
Anti-Phishing Phil	10 min	8.68 min (SD = 5.70)
PhishGuru Cartoon	2 min	.50 min (SD = 1.05)
Anti-Phishing Phil with PhishGuru Cartoon	12 min	8.55 min (SD = 5.50)

6.4 DISCUSSION

6.4.1 Limitations

There are several limitations to the current study. First, the sample was drawn from mTurk users and is not expected to be representative of the larger population of email users. Our sample of mTurk users tends to be younger, more educated and more tech savvy than the general public.

A second limitation of this study is the lack of direct consequences for user behavior. Participants might be more willing to engage in risky behavior in this roleplay if they feel immune to any negative outcomes that may ensue. Similarly, participants are not risking opportunity costs from being too conservative in their behavior. However, performance on this roleplay has been validated with real-world behavior, showing that, if anything, people are more conservative in their roleplay responses than they are with their actual email inboxes [121]. Furthermore, there is no reason to believe that the predictors described here should differ in their relationship to roleplay behavior compared to real-world behavior.

6.4.2 Summary of findings

Prior exposure to phishing education is associated with less susceptibility to phishing, suggesting that phishing education may be an effective tool. Also, more risk-averse participants tended to fall for fewer phish.

Gender and age are two key demographics that predict phishing susceptibility. Specifically, women click on links in phishing emails more often than men do, and also are much more likely than men to continue to give information to phishing websites. In part, this appears to be because women have less technical training and less technical knowledge than men. There is also a significant effect for age, in which participants aged between 18 and 25 are much more likely than others to fall for phishing. This group appears to be more susceptible because participants in this age group have a lower level of education, fewer years on the Internet, less of an exposure to training materials, and are less of an aversion to risks. Educators can bridge this gap by providing anti-phishing education to high school and college students.

All the education materials in our study reduce users' tendency to enter information into phishing webpages by about 16-21 percentage points. However, some education material decreased participants' tendency to click on legitimate links, this suggests that educator need to do a better job of teaching people how to distinguish phish from non-phish so that they avoid false positives.

Demographics such as age, gender, race, and education do not affect the amount of learning, suggesting that training can provide some benefit for all groups, if provided with the right materials. Although the 46% reduction in phishing susceptibility after training is substantial, even after training participants fell for 26% of the phishing messages in our roleplay. This finding shows that education is effective and needed but is not a cure all. In our study, 61% of the U.S participants have seen phishing education before; the task for the various stakeholders is to reach out to the 39% of the population who have not been exposed to training. However, even with the best educational materials, participants in our study still fell for around 28% of phish after training. Women and younger populations such as college students are especially vulnerable. These findings show that education should be complemented with other countermeasures such as filtering and law enforcement.

6.4.3 Role of education

As phishing continues to evolve, what is the role of education in combating it? Specifically, what problems can education solve, and how does education fit into a layered approach to combat phishing? We discuss these questions in the concluding section of this chapter.

Generally speaking, strategies for protecting people from phishing fall into three major categories: silently eliminating the threat, warning users about the threat, and training users not to fall for attacks. These categories of anti-phishing strategy mirror the three high-level approaches to usable security: build systems that just work" without requiring intervention on the part of users, make security intuitive and easy to use, and teach people how to perform security-critical functions [19].

Our view is that these three approaches should complement each other. Today, the majority of phishing emails are filtered at email gateways, and forwarding the future more efforts are needed to

filter as many phishing emails as possible, as quickly as possible, and with as few false positives as possible. Without this first layer of defense, even the best-educated users would be inundated with phishing messages that could paralyze their decision-making. It is also important to strengthen the browser, OS, and application security. Since it would be very difficult even for the experts to notice a compromised browser URL bar, user education would do little to alleviate the problem. In the same vein, users' computers can be infected with malware even without any user action. As a result, where possible, the first layer of defense should always be automated solutions to filter and increase the default security offered to users' computers and web applications.

However, we also need to acknowledge that these systems are not completely accurate in detecting phishing attacks. It is unlikely that any system will ever be completely accurate in detecting phishing attacks, especially when detection requires knowledge of contextual information. While it makes sense to use automated detection systems as one line of defense against semantic attacks, there will still remain many kinds of trust decisions that users must make on their own, usually with limited or no assistance. Thus, the second line of defense is to develop a complementary approach to *support* users so that they can make better trust decisions. There are two options for this approach: teach people not to fall for phish, or build easy-to-use software and interfaces that prevent users from falling for phishing.

User education is a low-hanging fruit. In our study, 61% of the U.S participants have seen phishing education before, and those who have seen education on average fell for 40-50% less phishing. Therefore efforts to reach out to the 39% of the population who have not been exposed to training would be likely to quickly reduce phishing susceptibility.

Finally, User education has its limits as well. Even with the best educational materials, participants in our study still fell for around 28% of phish after training. Women and younger populations such as college students are especially vulnerable. Therefore, the last step of defense is to build easy-to-use software and interfaces. Examples such as integrated web browser warnings [29] and foolproof phishing solutions are promising [108].

Appendix

Email	Legitimacy	Relevant features of email and sites
contest	real	Win a price in an online scavenger hunt From BRU Information Security Office link: `https://www.bru.edu/iso/aware/ncsam/hunt/bonus`
National City	real	Pat has an account. text of link: "view your statement" actual url: `http://www.nationalcity.com/statements`
party	possible malware	impersonal greeting link: `http://picasaweb.google.com/stevewulitzer/Partypics/` actual url: `http://128.3.72.234/Partypics.jpg.exe`
verify email account	phishing	threatens account deactivation asks for password in text of email no link in email
bankruptcy	spam	text of link: "Apply online now" actual url: `https://www.bankruptcylawyerfinder.com/`
bandwidth	phishing	misspelling in url link `http://wwwbrubandwithamnesty.org/bandwidth/agree.htm` actual url: same
eBay	phishing	threatens account suspension link: `https://signin.eBay.com/ws/eBayISAPI.dll...` actual url: `http://www.security-validation-your-account.com/...`
Amazon	real	problem with shipping link: `www.amazon.com/help/confirmation` actual url: same
National City	phishing	system upgrade link: `http://service-nationcity.org` actual url: `http://210.7.78.331/SITE/natcity/`
summary report	real	sender from bru.edu and in Pat's address book. summaryreport.doc attached
help desk	phishing	threatens account termination link: `http://bruwebmail.org/password/change.htm` actual url: same
eBay	real	text of link: "Send Invoice Now" actual url: `http://payments.ebay.com/eBayISAPI...`
networking:	phishing	.org domain link: `http://batonrougenetworking.org/summer09/register.html` actual url: same
As seen of Television	spam	dot com written out in email text

Table 6.8 Emails in Pat Jones' Inbox: Roleplay A

Rotated Component Matrix				
	Component			
	1	2	3	4
purchased anything on the web	.804	.024	-.057	-.039
online banking: ever used online banking	.258	.842	-.136	-.027
bills online: ever paid bills online	-.177	.885	.070	.113
credit card stolen: ever happen	.339	.115	.168	.670
ssn stolen: ever happen	.240	.086	.810	.081
info stolen: ever happen	.284	.162	-.705	.120
lose money: did you permanently lose money	-.113	-.009	-.147	.832
paypal account: ever had a paypal account	.754	.020	.049	.140

Table 6.9 Factor analysis for various Internet experience variables. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations. They are (1) **"web purchase experience"** (averaging purchasing at the web or whether had a paypal account); (2) **"online_banking"** by averaging ever used online banking and online bill pay; (3) **"ssn_stolen"** that is whether they had their ssn stolen, and (4) and **"creditcard_stolen"** that averages the credit card stolen and ever lose money.

Rotated Component Matrixa		
	Component	
	1	2
programming languages	-.254	.733
techology spectrum	.850	-.170
tech savvy	.820	-.288
security preference adjusted	-.569	.032
computers daily	-.153	.376
IT degree	.047	.861

Table 6.10 Factor analysis for various Internet experience variables. Rotation Method: Varimax with Kaiser Normalization. Rotation converged into two factors in five iterations. They are (1) **"tech_knowledge"** by averaging tech spectrum and tech savvy, and(2) called **"tech_training"** by averaging programming languages and IT degree (for tech_training, lower numbers mean more training).

Rotated Component Matrixa			
	Component		
	1	2	3
risk: Betting a days income at the horse races	.083	.901	.059
risk: Investing 10 of your annual income in a moderate growth mutual fund	-.008	-.030	.701
risk: Drinking heavily at a social function	.446	.415	-.042
risk: Betting a days income at a high stake poker game	.132	.911	.058
risk: Investing 5 of your annual income in a very speculative stock	.051	.082	.829
risk: Betting a days income on the outcome of a sporting event	.140	.894	.129
risk: Engaging in unprotected gender	.628	.179	-.045
risk: Driving a car without wearing a seat belt	.800	.062	.019
risk: Investing 10 of your annual income in a new business venture	.154	.142	.764
risk: Riding a motorcycle without a helmet	.681	.120	.218
risk: Sunbathing without sunscreen	.755	.042	.068
risk: Walking home alone at night in an unsafe area of town	.740	.065	.073

Table 6.11 Principle Component analysis for various Internet experience variables. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.440a	.194	.174	1.90464

Table 6.12 Regression statistics

ANOVA						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	799.840	22	36.356	10.022	.000a
	Residual	3330.196	918	3.628		
	Total	4130.036	940			
b. Dependent Variable: pre_test_phish_giveinfo						

Table 6.13: Complete list of variables for regression

Variable	Descriptions	Statistics
age_numeric	What is your age?	M = 30.1 SD = 10.6
sexsurvey	What is your gender?	
	1 = Male	475
	0 = Female	508
education_recode	What is your highest education?	
	1 = High school or less	79
	2 = Some college	349
	3 = Completed 4-year college degree	284
	4 = Some Post-graduate education	97
	5 = Have masters or Ph.D degree	169
	6 = Decline to answer	5
OCCU_student	Are you currently a student?	
	1 = YES	247
	0 = NO	736
hispanic	Are you Hispanic?	
	1 = YES	66
	2 = NO	842
racewhite	What's your race (white or Caucasian?)	
	1 = YES	641
	2 = NO	342
countryindia	Do you currently reside in India?	
	1 = YES	145
	2 = NO	838
countryusa	Do you currently reside in US?	
	1 = YES	739
	2 = NO	244
income	What is your annual household income?	
	1 = $<$ \$20,000	203
	2 = \$20,000 - \$39,000	196
	3 = \$40,000 - \$59,000	181
	4 = \$60,000 - \$79,000	99
	5 = \$80,000 - \$99,000	74
	6 = $>$100,000	74
	7 = Decline to answer	156
avoidphish	Have you ever seen information to avoid phish before this study?	
	1 = YES	556

Table 6.13: Complete list of variables for regression

Variable	Descriptions	Statistics
	1.5 = NOT SURE	85
	2 = NO	342
computersdaily	Do you use computers daily?	
	1 = YES	867
	2 = NO	116
emailperday_numeric	On average, how many emails do you receive a day?	M = 44, SD = 81
tech_knowledge	Tech_knowledge scale from Factor analysis (1 – 7)	M = 5.3 SD = 1.2
tech_training	Tech training scale from factor analysis (1 – 2)	M = 1.7 SD = 0.36
risk_health_safty	How do you perceive the following risks (1– 7)?	M = 5.5, SD = 1.0
risk_financial_betting	How do you perceive the following risks?	M = 5.8 SD = 1.3
risk_financial_investing	How do you perceive the following risks?	M = 4.1 SD = 1.1
magcomputer	What magazines do you frequently read (computers and electronics?)	
	1 = YES	335
	2 = NO	648
internet_numeric	At what year did you first use Internet?	M = 1996, SD = 3.7
online_banking	Online banking scale from Factor analysis (1 – 2)	M = 1.17 SD = 0.33
creditcard_stolen	Have you ever had your creditcard stolen online?	
	1 = YES	34
	1.5 = NOT SURE	25
	2 = NO	924
web_purchase	Web purchase experience scale from factor analysis (1 – 2)	M = 1.12 SD = 0.27

Coefficients						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-90.519	40.864		-2.215	.027
	age_numeric	-.023	.007	-.116	-3.142	.002
	sexsurvey	.586	.148	.140	3.964	.000
	education	-.126	.053	-.080	-2.391	.017
	OCCU_student	-.090	.170	-.019	-.533	.594
	hispanic	.068	.173	.012	.390	.697
	racewhite	-.324	.162	-.074	-2.005	.045
	countryindia	.074	.292	.012	.253	.801
	countryus	.018	.222	.004	.080	.936
	income	-.060	.031	-.060	-1.942	.052
	avoidphish	.851	.147	.189	5.787	.000
	computersdaily	.100	.201	.015	.495	.621
	emailperday_numeric	-.002	.001	-.073	-2.324	.020
	tech_knowledge	-.173	.061	-.103	-2.840	.005
	tech_training	.496	.208	.085	2.388	.017
	risk_health_safty	.103	.067	.050	1.530	.126
	risk_financial_betting	.110	.054	.067	2.055	.040
	risk_financial_investing	-.153	.061	-.080	-2.518	.012
	magcomputer	.213	.156	.048	1.366	.172
	internet_numeric	.046	.020	.083	2.275	.023
	online_banking	-.084	.222	-.013	-.377	.706
	creditcard_stolen	-.102	.324	-.010	-.315	.753
	web_purchase	-.250	.287	-.032	-.871	.384

Table 6.14 Complete output of the regression analysis.