Experimental Investigation of Demographic Factors Related to Phishing Susceptibility

Wanru Li
George Mason University
wli15@gmu.edu

James Lee
George Mason University
jlee194@gmu.edu

Justin Purl
Google
justin.purl@gmail.com

Frank L. Greitzer
PsyberAnalytix LLC
frank@psyberanalytix.com

Bahram Yousefi
George Mason University
byousefi@gmu.edu

Kathryn B. Laskey
George Mason University
klaskey@gmu.edu

Abstract

This paper reports on a simulated phishing experiment targeting 6,938 faculty and staff at George Mason University. The study examined various possible predictors of phishing susceptibility. The focus of the present paper is on demographic factors (including age, gender and position/employment). Since previous studies of age and gender have yielded discrepant results, one purpose of the study was to disambiguate these findings. A second purpose was to compare different types of email phishing exploits. A third objective was to compare the effect of different types of feedback given to those who clicked on one or more of three simulated phishing exploits that were deployed over a three-week period. Our analysis of demographic factors, effects of phishing email content, and effects of repeated exposure to phishing exploits revealed significant age effects, marginally significant gender differences, and significant differences in email type. A multi-level model estimated effects of multiple variables simultaneously.

1. Introduction

Phishing attacks—social engineering exploits using digital means—are unintentional insider threats [6] that can result in serious financial impacts and/or losses of confidential information. These exploits cause grave damage to both commercial and US government entities—such as the now infamous cyber/phishing attack against the US Government Office of Personnel Management that gave attackers access to sensitive data on millions of government employees and contractors. Phishing attacks targeted on US organizations increased more than 40% in 2018 [12]; the FBI Internet Crime Complaint Center reported 2018 complaints amounting to losses of over $1.2 billion for business email compromises, and, more generally, $2.7 billion in Internet crime losses – nearly twice the financial impact of 2017 [3].

Research on factors that affect people's susceptibility to phishing is an essential step in improving cybersecurity awareness and designing protective strategies. Research has revealed numerous personal or demographic factors that are related to phishing susceptibility [7]. However, there is a lack of agreement among studies that differ in methods and populations studied. A primary purpose of this paper was to disambiguate some of the discrepant findings on demographic factors (particularly age and gender) and to compare effects of different phishing email content on phishing susceptibility. The study also examined possible effects of different types of feedback given to those who clicked on one or more of three simulated phishing exploits that were deployed over a three-week period.

The rest of the paper is organized as follows. Section 2 reviews previous phishing susceptibility studies; Section 3 presents the research questions for this study; Section 4 describes the design of the phishing study; Section 5 discusses the results of the phishing susceptibility analysis; Section 6 discusses the key findings, contributions and limitations of this study; and Section 7 concludes the paper.

2. Related research

2.1. Previous phishing study results

A typical phishing study includes simulated phishing campaigns, surveys, or both. Simulated phishing attacks generally do not inform users that they are participating in a phishing study. A study may provide a warning of possible phishing exploits

URI: https://hdl.handle.net/10125/64015
978-0-9981331-3-3

to evaluate the effectiveness of warnings. For example, Mohebzada et al. [11] conducted two phishing email experiments targeting 10,568 faculty, staff and students at the American University of Sharjah during the spring semester in 2010. They sent out warning notifications after running 18 hours of the campaign and found the warning messages were largely ignored, suggesting that warnings may not be sufficient to prevent users from falling for phishing.

Survey studies typically inform participants of the study's purpose before distributing the survey. For example, Sheng et al. [13] recruited 1001 online participants through Amazon.com's Mechanical Turk, who then answered survey questions and completed a roleplay task; the study examined demographic factors and the effectiveness of different anti-phishing training materials.

Two recent phishing studies included both experiments and surveys. In spring 2018, Diaz et al. [1] launched a phishing email experiment to study demographic factors related to phishing at the University of Maryland, Baltimore County (UMBC). They sent out simulated phishing attacks targeting 1,350 undergraduate students who were not informed before the experiment. A survey was sent after the experiment to analyze some additional demographic factors such as computer usage time and anti-phishing training experience. Another recent example is a study by Greene et al. [5], who launched three phishing exercises (Mar, Aug, Dec) targeting approximately 70 staff at the National Institute of Standards and Technology. The purpose was to study the reasons why email users were clicking or not clicking on phishing links and attachments. They also conducted three post-exercise surveys corresponding to each phishing exploit and compared survey responses between clickers and non-clickers.

Below we describe findings associated with phishing and, as motivation and background for the present study, point out several issues that may contribute to discrepant or ambiguous results.

2.2. Factors related to phishing susceptibility

2.2.1. Gender. Inconsistent results have been reported in previous studies of the relationship between gender and phishing susceptibility. Jagatic et al. [9], Sheng et al. [13], and Halevi et al. [8] indicated that women were significantly more likely to fall for phishing than men. In contrast, Mohebzada et al. [11] found males and females were equally likely to fall for the first phase of a phishing attack, but more males (60.9%) were deceived in the second phase of their phishing attack than females (39.1%). Diaz et al. [1] found that 3% more males than females clicked on the phishing email, although this difference was not statistically significant. Further study of gender effects is warranted to deconflict these results.

2.2.2. Age. Inconsistent results have also been reported on the association between age and phishing susceptibility. Some key methodological differences in relevant studies may account for this. In an online survey and role-playing study involving a university population, Sheng et al. [13] found that subjects in the 18-25 age range were more likely to click on the phishing emails than people in other age groups (26-35, 36-45, 46-55, and over 55). In contrast, the university study by Downs et al. [2] found no significant association between age and behaviors exhibited in their email role playing study, although they found that younger people engaged in more risky online activities. Their sample included students, faculty and staff ranging from 18-45 years old.

Several other studies reported findings on age factors that appear to differ from the above studies, but there are important differences in how these studies categorized age groups. In the Mohebzada et al. [11] simulated phishing experiment targeting university faculty, staff and students, no relationship was found between "age" and phishing susceptibility--but the age range in this study was defined by undergraduate level (freshman, sophomore, junior, senior), with a "typical" age range from 18-21 years old. Similarly, in the Jagatic et al. [9] phishing email experiment conducted at Indiana University targeting students aged 18-24 years old, younger users were slightly more likely to be successfully phished. The age categories in [11] and [9] correspond to the lowest age category examined by Sheng et al. [13] and Downs et al. [2], and therefore these results are not necessarily inconsistent.

With regard to age effects, the evidence seems to suggest that the younger age categories are more susceptible to phishing than older age groups. However, the discrepancies in methods and populations studied, as well as possible confounding factors that were not addressed (e.g. experience), point to a need for a more careful study of age effects that controls for possible confounding variables.

2.2.3. Phishing email content. Previous research indicates that the address of the sender and email content affect user response to email. Furnell [4] indicates that content characteristics, such as visual factors (logos and banners) may entice people to click on a phishing email. Jagatic et al. [9] suggest that a sender address from the university domain lowers students' guard. Greene et al. [5] argue that the alignment of user context and the phishing attack premise is a significant factor in phishing susceptibility. Vishwanath et al. [14] found the level of attention to urgency cues or to email subject lines significantly affects clicking response to phishing emails; however, levels of attention to grammar or spelling were significantly less likely to affect users being phished. The importance of visual and other cues is clearly a topic for further investigation.

2.2.4. Feedback type. A limited amount of research has been directed toward the impact of previous experience with phishing. Sheng et al. [13] reported that their survey participants who had previous anti-phishing training experience were less likely to fall for phishing attacks. Effects of feedback or learning effects may be studied in surveys, through self-reporting, or they may be examined in more longitudinal approaches that determine if prior exposure to phishing impacts future behavior. This is one of the questions addressed in the present study.

3. Research questions

As we noted in the previous section, the various studies relating to demographic (and some contextual) factors have yielded somewhat inconsistent results. A more complete list of relevant research topics ([6], [7]) includes the effects of demographics factors, email content/visual cues, previous experience with phishing, level of sophistication in using computers or internet experience, and human behavioral/psychological factors. The study reported here examined each of these topics to some degree, but the present paper focuses on the first three research needs; findings relating to the other topics will be reported in future papers. Thus, the present paper reports on the following research questions:

Research Question 1: *Will the experience of succumbing to a phishing email and subsequent feedback impact future behavior?* We ask if users who obtain explanatory feedback after clicking on a phishing email will be less likely to click on a phishing email in the future, and whether the nature of the feedback (a brief message or a video landing page) will affect the likelihood of succumbing to a subsequent phishing attack.

Research Question 2: *Regarding effects of demographic factors,*

(a) *will there be differences in susceptibility to phishing (as measured by likelihood of clicking on a phishing email) based on age?* We are particularly interested in whether there will be significant age effects, after statistically adjusting for other possible factors.

(b) *will there be gender effects?* We suspect that the effect of gender will be minimal, if at all. We do not expect to find large differences in clicking behavior in response to phishing emails, after statistically adjusting for other possible factors.

(c) *will there be differences in susceptibility to phishing based on employment category (position/department)?* We are particularly interested in the nature of this possible effect, after statistically adjusting for other factors.

Research Question 3: *Will the content of phishing emails (source of message, visual cues) impact the likelihood of responding to the exploit?* Previous research suggests that certain characteristics of a phishing email may affect clicking behavior (e.g., [5]). We focus on the type or source (IT/tech support, package delivery, credit card warning) that presents different message content or context. Since the IT tech support context may be more relevant to users, we expect that more users will click on the IT tech support email than the other two types of email.

4. Method

We designed and conducted an experimental study targeting 6,938 faculty and staff at George Mason University to identify the characteristics of users who are susceptible to phishing. We distributed three simulated phishing emails over a period of three weeks, from October 30 to November 21 of 2018. Study weeks started on Tuesdays and ended on Mondays. All data in this study was de-identified to protect personally identifiable information [10].

Our experimental design included varying types of simulated email scams: one related to IT/tech support, one related to finance/banking, and one related to e-commerce/package delivery. This allowed us to examine possible differences in vulnerability across these phishing email types. Users in our study received three different phishing emails – IT/help desk (IT), Package Delivery (PD), and Credit Card Warning (CC)-- each in a different week and on a different day of the week. People who clicked on the simulated phishing link were taken to a randomly chosen landing page (LP). Three different LPs were designed: a “page not found,” a brief message informing users that the email was part of a phishing study, and a similar message with a short anti-phishing training video. In addition to the simulated phishing attack, we collected Human Resource (HR) data to enable an analysis of demographic factors.

Prior to initiating the campaign, the university administration required that a pre-phish email be sent notifying users that they may receive email messages as part of a phishing test. This email also cautioned users against clicking links or visiting URLs if a message is suspected to be a phish. Because there was no warning on individual phishing emails, the IRB approved a waiver of informed consent but required a deception notice to be sent after study completion. Therefore, we sent subjects a debriefing email to explain that the simulated phishing emails were part of an experiment for research purposes with no security risks that would make them vulnerable to any threats. Users could indicate a desire not to have their data used for the study.

4.1. Study population

Of the 6,938 participants in this study, 46% were male and 54% were female. We excluded 17 people who were involved with the design of the study and 11 who opted out of the study. The 27-41 year old age group had the highest proportion of participants (31%), while the youngest age group (less than 27 years old) had the smallest proportion (9.3%). The proportions of technical faculty, other faculty, and administrative staff were 15.5%, 45.1%, and 39.4%, respectively. The most notable gender difference was the much higher proportion of males than females among technical faculty over age 59.

4.2. Data overview

We used Human Resource records for faculty and staff to identify demographic factors, including age, gender, position, and department type. Age groups were defined so that there would be no singly identifiable personnel using other demographic information. We broke down the position factor into adjunct faculty, full-time faculty, wages staff and other staff. The department type is broken down into administration, technical college, and other college. Technical college includes employees in engineering and science; remaining non-administration employees are categorized as other college.

To collect users' click behavior reflecting their susceptibility to phishing emails, we used an open source phishing framework called Gophish [15], intended to help organizations test their own exposure to phishing. The Gophish application sent simulated phishing emails, directed clickers to the appropriate LP, and recorded the data.

Other data collected in this study included technical data such as VPN and firewall logs; a pre-campaign survey on technical/cybersecurity-related experience and psychological/behavioral/personality factors; and a final survey, after the conclusion of the phishing campaigns, asking more in-depth questions about the phishing emails, reasons for clicking and/or not clicking the email links, and the user's usual behavior when receiving or reacting to such emails. Since here we are focused primarily on demographic factors, the analysis and reporting on these relationships are planned for a future publication.

4.3. Simulated phishing campaign

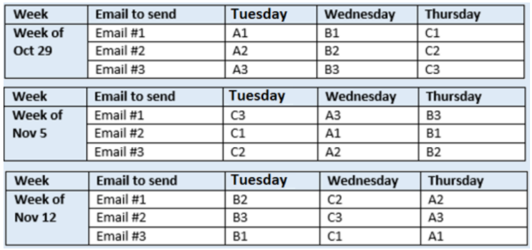
We designed three phishing emails in different contexts with urgency cues to stimulate users to click on a link:

- **IT Helpdesk (IT)**. An IT helpdesk email notifies the user that there had been suspicious activity overnight, which caused the account to be deactivated. The user is instructed to click a link to review the activity and reactivate the account.
- **Package Delivery (PD)**. A package delivery service email is sent to users describing a failed package delivery due to invalid postal code. The user is instructed to click on the link to download the shipping label that must be brought to the post office to pick up the package.
- **Credit Charge Warning (CC)**. An email notifies users of a suspicious charge on a credit card, for which large purchase notifications are enabled. The user is asked to click a link to review the charge and change notification settings.

We purchased domain names for each of the sender accounts. We purposefully assigned plausible names that would be somewhat suspicious to careful users. The domain name for the IT helpdesk email was "support@masonhelpdesk.com" as opposed to the actual university IT email account, and the package delivery email was sent from the fictitious "pkginfo@vapostal.com." The credit card email was sent from "service@acubank.co", which has a ".co" instead of a ".com" address.

The goal was to send each user all three emails. However, there was a concern that if a user receives all three emails or if all users received the same email on the same day (or week), it would raise suspicion about the emails and be less effective as a result. To minimize potential suspicion and to counterbalance potentially confounding factors such as day of week and order of receipt, we created nine user groups (A1, A2, A3, B1, B2, B3, C1, C2, C3) using stratified sampling to make sure each group has similar age, gender, and department type composition. We sent each group one email per week on a different day of the week. This way, each group would receive all three emails, but on a different day of each week (Table 1). In this paper, we consider week 1, week 2, and week 3 as the sequence of three weeks.

Table 1. Phishing campaign schedule



Week	Email to send	Tuesday	Wednesday	Thursday
Week of Oct 29	Email #1	A1	B1	C1
	Email #2	A2	B2	C2
	Email #3	A3	B3	C3
Week	Email to send	Tuesday	Wednesday	Thursday
Week of Nov 5	Email #1	C3	A3	B3
	Email #2	C1	A1	B1
	Email #3	C2	A2	B2
Week	Email to send	Tuesday	Wednesday	Thursday
Week of Nov 12	Email #1	B2	C2	A2
	Email #2	B3	C3	A3
	Email #3	B1	C1	A1

The email campaigns were terminated on November 21, giving each email at least a full week to be opened and clicked by each user. We recorded the operating system and time of the clicks so that we could link the click behavior to IT data and identify technical indicators that suggest susceptibility to phishing. If a user made multiple clicks on an email, we recorded the time of the first click.

To examine the impact of feedback given after clicking on a phishing link, we varied the Landing Page (LP) to which the user was redirected after clicking the link. We were interested in any differences in the impact of LP on subsequent behavior (i.e., the likelihood of clicking on a subsequent phishing email). Users who clicked on a phishing link were redirected at random to one of three LPs: (a) a standard 404 (“Page Not Found”) error that does not notify the user that he or she has clicked on a phishing link; (b) a webpage that displays a simple message notifying the user that he or she has clicked on a simulated phishing link from the study; (c) a webpage that notifies the user that he or she has clicked on a phishing link from the study, explains the study, and provides a training video on how to identify suspicious emails. Thus, the training video provided the most educational feedback, and the “Page Not Found” 404 message provided the least informative feedback. For those who clicked on the link multiple times, we used HTTP cookies to implement a script that would ensure that they would see the same LP each time, as long as they were using the same device or browser. For the few users who used different devices and therefore saw different LPs for the same email, we recorded the “strongest” LP that was experienced.

Our LP research question may be described using two hypotheses: 1) users who receive notification about clicking on a phishing link would be *less* likely to click on a future link, and 2) users who receive a stronger notification (i.e., training video LP) would be *less* likely to click on a future link than those who received a simple message notification. To ensure a sufficient sample size for the first hypothesis mentioned above, we set the probability distribution for simple message LP to be 25%, training video LP to be 25%, and standard 404 LP to be 50%.

5. Results

The statistical methodologies applied in this study are the Chi-square test for independence at significance level α=0.05, Cramer’s V to test strength of that significance, and multiple pairwise comparisons for proportions with Bonferroni correction. For the Chi-square test, the null hypothesis is that there is no association between the test variable and the clicking result. For the proportions test, our null hypothesis is that there is no difference between the test proportions.

5.1. Click behavior and landing page

5.1.1. Landing page analysis. We examined the LP data to assess whether clickers learned from the LP to be more alert when receiving the next simulated phishing email. We conducted two types of LP analysis to investigate our hypothesis: First we checked to see if clickers who received a brief message or a video LP from the first two weeks are less likely to click on email link in the third week (results shown in Table 2). For the second analysis, we explored the effectiveness of the video LP by comparing the click rates in the future week by different LPs that people received from the first two weeks (results shown in Table 3).

For the first analysis (Table 2), we considered the strongest LP variable with three levels: 404 page, brief message and video page, and no LP. No LP indicates the user did not click the link in the email from the previous week. This would mean they did not see any LP. Although we hypothesized that users may learn from the notification on the LP, we found no statistically significant difference in the week 3 click rate between users who received the 404 LP versus those who received some form of notification (brief message or video) in previous weeks. On the other hand, comparing the click rate for 404 (or combined message and video) with no LP, we found a highly significant difference: clickers were more likely to be repeat-clickers than non-clickers were to become clickers (Table 3). In other words, previous week non-clickers are significantly less likely to click than previous week clickers.

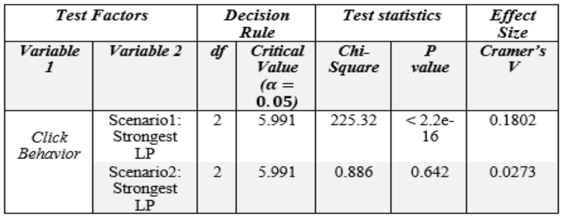
The second LP analysis focused on the effectiveness of the video. We found no significant difference between click behavior across the strongest LP variable with three levels, 404 page, brief message, and video page (Table 3 and Table 4). Specifically, the video LP did not contribute to a lower click rate, contrary to our initial hypothesis.

5.1.2. Week-to-week click rate. In another analysis, we observed a decreasing trend for the week-to-week click rate. 719 users (10.4%) clicked on the week 1 simulated phishing email. 617 users (8.9%) clicked on the week 2 email. 539 users (7.8%) clicked on the week 3 email. However, the decreases in click rate from one week to the next were significant *only* for users who *did not click* in the previous week (Table 5). In other words, previous week non-clickers are less likely to click than previous week clickers. We hypothesize that this occurs because removing users who were successfully phished in the previous week results in a population less susceptible to being phished. Lack of a statistically significant decrease in click rate among those who *did click* in a previous week may reflect lack of an effect or insufficient power/sample size.

Table 2. Two types of landing page analysis

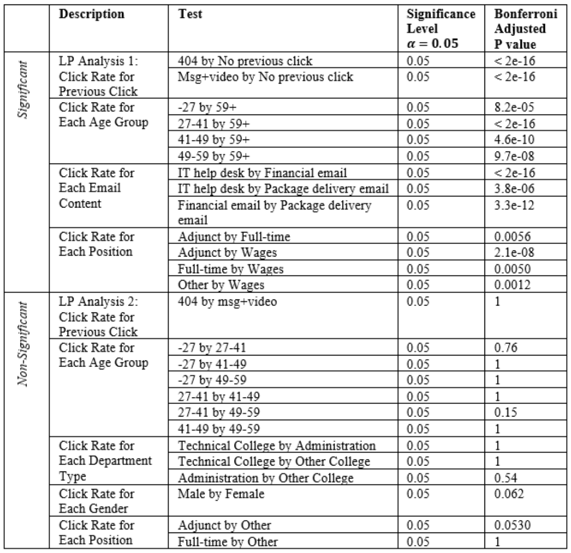
LP Analysis Type 1		Strongest Landing Page		
		404	Msg + video	No LP
W3 Clicked		99	119	321
W3 Did not click		408	563	5428
Total		507	682	5749
Click Rate		**19.53%**	**17.45%**	**5.58%**
95% Confidence Interval for Click Rate (%)	Lower Bound	16.31	14.78	5.02
	Upper Bound	23.20	20.48	6.21
LP Analysis Type 2		**Strongest Landing Page**		
		404	**Msg**	**Video**
W3 Clicked		99	57	62
W3 Did not click		408	276	287
		507	333	349
Click Rate		**19.53%**	**17.12 %**	**17.77%**
95% Confidence Interval for Click Rate (%)	Lower Bound	16.31	13.45	14.11
	Upper Bound	23.20	21.53	22.12

Table 3. LP analysis: Result of Chi-square test for independence



Test Factors		***Decision Rule***		***Test statistics***		***Effect Size***
Variable 1	***Variable 2***	***df***	***Critical Value*** ***(α = 0.05)***	***Chi-Square***	***P value***	***Cramer's V***
Click Behavior	Scenario1: Strongest LP	2	5.991	225.32	< 2.2e-16	0.1802
	Scenario2: Strongest LP	2	5.991	0.886	0.642	0.0273

Table 4. Multiple pairwise comparisons for proportions with Bonferroni correction



	Description	**Test**	**Significance Level** $\boldsymbol{\alpha = 0.05}$	**Bonferroni Adjusted P value**
Significant	LP Analysis 1: Click Rate for Previous Click	404 by No previous click	0.05	< 2e-16
		Msg+video by No previous click	0.05	< 2e-16
	Click Rate for Each Age Group	-27 by 59+	0.05	8.2e-05
		27-41 by 59+	0.05	< 2e-16
		41-49 by 59+	0.05	4.6e-10
		49-59 by 59+	0.05	9.7e-08
	Click Rate for Each Email Content	IT help desk by Financial email	0.05	< 2e-16
		IT help desk by Package delivery email	0.05	3.8e-06
		Financial email by Package delivery email	0.05	3.3e-12
	Click Rate for Each Position	Adjunct by Full-time	0.05	0.0056
		Adjunct by Wages	0.05	2.1e-08
		Full-time by Wages	0.05	0.0050
		Other by Wages	0.05	0.0012
Non-Significant	LP Analysis 2: Click Rate for Previous Click	404 by msg+video	0.05	1
	Click Rate for Each Age Group	-27 by 27-41	0.05	0.76
		-27 by 41-49	0.05	1
		-27 by 49-59	0.05	1
		27-41 by 41-49	0.05	1
		27-41 by 49-59	0.05	0.15
		41-49 by 49-59	0.05	1
	Click Rate for Each Department Type	Technical College by Administration	0.05	1
		Technical College by Other College	0.05	1
		Administration by Other College	0.05	0.54
	Click Rate for Each Gender	Male by Female	0.05	0.062
	Click Rate for Each Position	Adjunct by Other	0.05	0.0530
		Full-time by Other	0.05	1

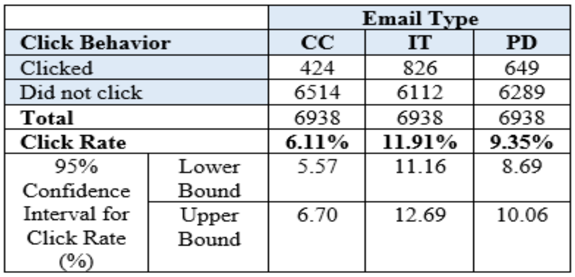
Table 5. Week-to-week click rate by LPs

		W1 404 (W1 Clicked)	**W1 msg+video (W1 Clicked)**	**W1 No LP (W1 Did not click)**	**Total**
W2 Clicked		72	85	460	617
W2 Did not click		262	305	5754	6321
Total		334	390	6214	6938
W2 Click Rate		**21.56%**	**21.79%**	**7.40%**	**8.89%**
95% Confidence Interval for Click Rate (%)	Lower Bound	17.48	17.98	6.78	8.25
	Upper Bound	26.28	26.16	8.08	9.59
		W2 404 (W2 Clicked)	**W2 msg+video (W2 Clicked)**	**W2 No LP (W2 Did not click)**	**Total**
W3 Clicked		62	65	412	539
W3 Did not click		225	273	5901	6399
Total		287	338	6313	6938
W3 Click Rate		**21.60%**	**19.23%**	**6.53%**	**7.77%**
95% Confidence Interval for Click Rate (%)	Lower Bound	17.23	15.38	5.94	7.16
	Upper Bound	26.72	23.77	7.16	8.42
		W1 or 2 404 (W1 or 2 Clicked)	**W1 or 2 msg+video (W1 or 2 Clicked)**	**W1 or 2 No LP (W1 or 2 Did not click)**	**Total**
W3 Clicked		99	119	321	539
W3 Did not click		408	563	5428	6399
Total		507	682	5749	6938
W3 Click Rate		**19.52%**	**17.45%**	**5.58%**	**7.77%**
95% Confidence Interval for Click Rate (%)	Lower Bound	16.31	14.78	5.02	7.16
	Upper Bound	23.20	20.48	6.21	8.42

5.2. Click behavior and email content

5,421 users (78.1%) did not click on any simulated phishing emails. 1,517 users (21.7%) clicked at least one email. 1,192 users (17.2%) clicked on only one email, 268 users (3.9%) clicked on two emails, 57 users (0.8%) clicked on all three. Of the 1,517 users who clicked on at least one email, 424 (6.11% of all users) clicked the CC email, 826 (11.91% of all users) clicked the IT email and 649 (9.35% of all users) clicked the PD email (Table 6). There is a significant difference between the click rates for any two emails. As we hypothesized, the IT help desk email tricked the largest proportion of users, followed by the package.

Table 6. Click behavior by email type



		Email Type		
Click Behavior		**CC**	**IT**	**PD**
Clicked		424	826	649
Did not click		6514	6112	6289
Total		6938	6938	6938
Click Rate		**6.11%**	**11.91%**	**9.35%**
95% Confidence Interval for Click Rate (%)	Lower Bound	5.57	11.16	8.69
	Upper Bound	6.70	12.69	10.06

5.3. Demographic variables

We examined click rates for each demographic category directly. Figures 1 and 2 show 95% confidence intervals for click rate by demographic (age and gender) and employment (department type and position) factors.

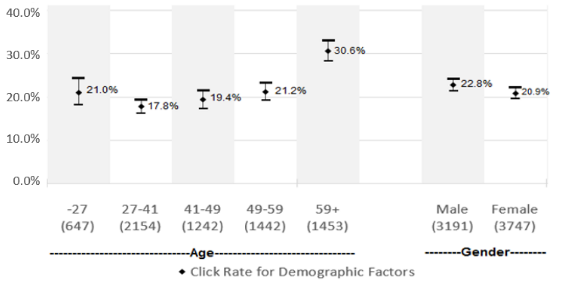


Figure 1. Click rate and 95% confidence interval by demographic factors

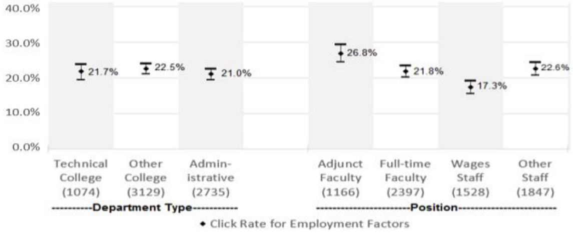


Figure 2. Click rate and 95% confidence interval by employment factors

The Chi-squared test shows a significant relationship between age or position and click behavior (Table 7). Table 4 shows that people over 59 years old are more likely to fall for phishing but contrary to our initial hypothesis, we did not find that people in the youngest age group were significantly more likely to fall for the phishing emails.

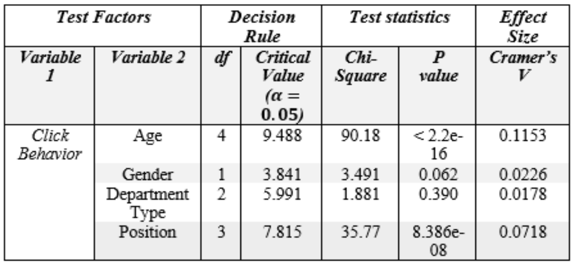
We found adjunct faculty are significantly more susceptible to phishing than full-time faculty and wages staff, and marginally more likely to click than other staff at significance level 0.05 (Table 4), which may be explained by that a higher proportion of older people (29.5%) is identified in adjunct faculty than other job categories (full-time: 24.5%; other: 14.5%; wages: 16.6%). Wages staff are significantly less likely to be phished than people in other positions. We found there is a marginally significant association between gender and phishing susceptibility at significance level 0.05 (Table 7).

Males had a slightly higher click rate. Since the sample size is large, the Cramer's V values show a relatively weak relationship between the click behavior and each demographic variable, but the differences are statistically significant. There is no statistically significant association between department type and phishing susceptibility. We found no evidence indicating that employees from technical colleges are less likely to fall for phishing than from other departments.

5.4. Demographic variables and email content

To investigate the relationship between demographic variables and email content, we analyzed the users who clicked on each type of phishing link. Figs. 3 and 4 show click rates and 95% confidence intervals for each type of email content by demographic and employment factors. Among the clickers, we observed that males are more likely to click on the credit card email than females. Pairwise comparison of proportions (Table 4) suggests users in the oldest group (above 59 years old) are more likely to click on

Table 7. Demographic variables analysis: Result of Chi-square test for independence



Test Factors		***Decision Rule***		***Test statistics***		***Effect Size***
Variable 1	***Variable 2***	***df***	***Critical Value (α = 0.05)***	***Chi-Square***	***P value***	***Cramer's V***
Click Behavior	Age	4	9.488	90.18	< 2.2e-16	0.1153
	Gender	1	3.841	3.491	0.062	0.0226
	Department Type	2	5.991	1.881	0.390	0.0178
	Position	3	7.815	35.77	8.386e-08	0.0718

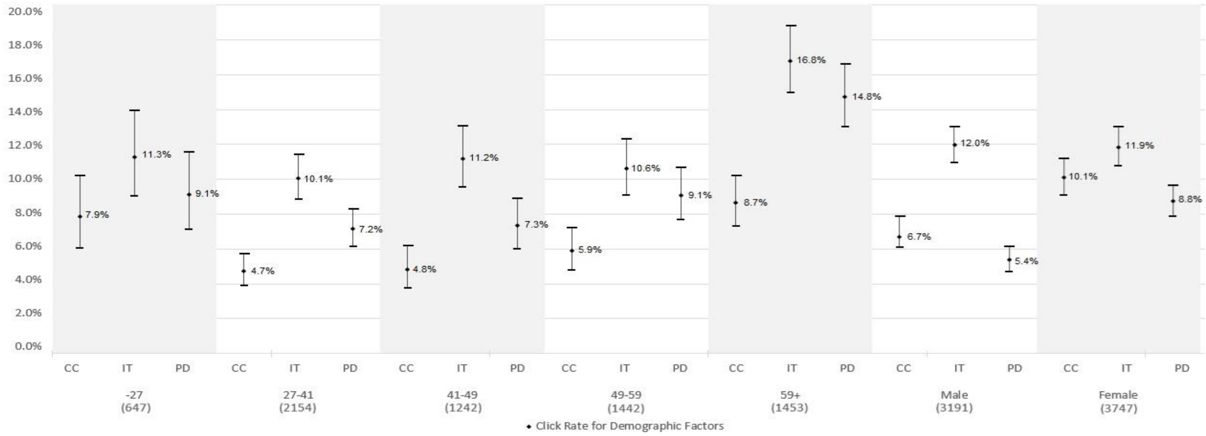


Figure 3. Click rate and 95% confidence interval for each email content by demographic factors

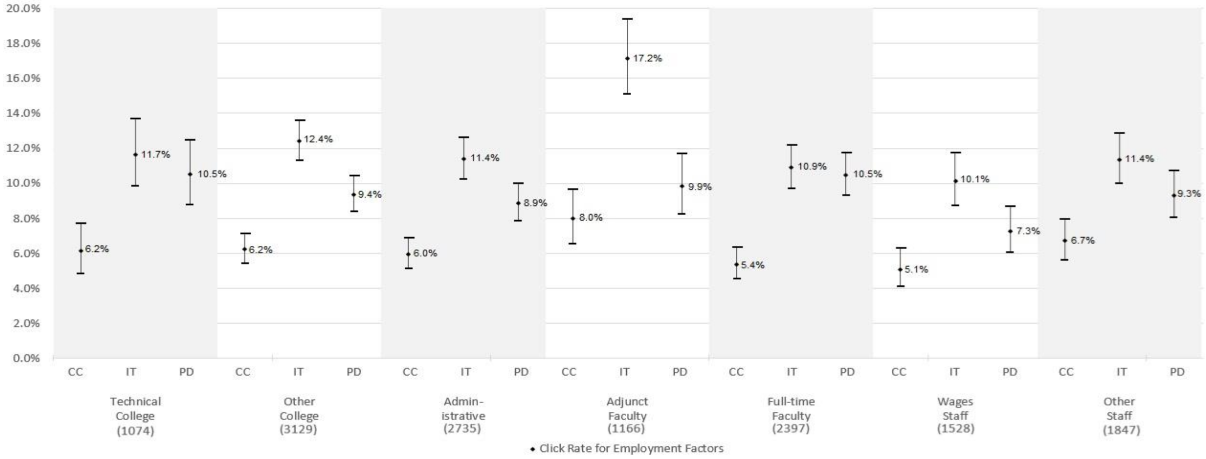


Figure 4. Click rate and 95% confidence interval for each email content by employment factors

the credit card email than people aged 27 to 59.

Users in the youngest age group (less than 27 years old) are significantly more likely to click on the credit card email than people aged between 27 and 41. Our analysis also suggests users in the oldest age group are significantly more likely to fall for IT help desk email and package delivery email than other groups. Our analysis shows that adjunct faculty are significantly more likely to click on the credit card email than the full-time faculty and wages staff. Adjunct faculty are significantly more susceptible to the IT help desk email than people in other positions. Full-time faculty are more likely to click on the package delivery email than wages staff. We found no other differences that were statistically significant. There is no significant difference in proportion of people who clicked on a particular email by department type.

5.5. Multi-level model

Univariate and bivariate analyses have the potential for confounds. For example, in samples where males tend to be younger, the univariate effect of age may be caused by gender differences (or vice versa). Multivariate statistics can separate the effect of variables when entered into the same model. Another potential confound for the univariate statistical analysis used in this study is the effect of time. Extending our hypothetical example, if young male staff happen to respond to email more quickly than older female faculty, and if responding quickly is related to phishing susceptibility, then a univariate effect for age, gender, or position would actually be explained by time (response latency). We conducted a separate analysis of the factors examined in this study using a multi-level logit model including all variables simultaneously to better isolate the effects of these factors. Using this approach, the effects of time, age, gender, position, and LP are estimated for each of the email types (Table 8). Two conclusions are consistent across email type: university employees are less likely to click as time passes and those who have clicked before are more likely to click again. This second finding is consistent with univariate results. The effect of age does seem to vary by email type (as in the bivariate statistics), such that the over 59 year old age group tend to be more susceptible to the IT email, whereas the youngest age group seems to be more susceptible to the banking email. The effect of gender (male susceptibility) does seem to be independent from the effect of age for the banking email, where the other email types have no significant gender effects. Adjunct faculty seem to have a significant susceptibility to the IT email with the added effect of adjunct faculty and other staff having somewhat higher susceptibility to the banking email. Finally, the feedback intervention was only successful for the banking email: those who saw a LP were less likely to click again than those having seen a 404 error.

Table 8. Fixed effects for a multi-level logit model predicting susceptibility for three email types.

	IT	PD	CC
(Intercept)	-5.90***	-6.53***	-6.61***
Time	-0.08***	-0.09***	-0.09***
27-41	-0.16	-0.36*	-0.52**
41-49	-0.06	-0.37*	-0.51*
49-59	-0.15	-0.15	-0.33†
59+	0.28*	0.30†	-0.04
Female	0.06	-0.06	-0.21*
Full-time	-0.46***	0.15	-0.34*
Other	-0.38***	0.07	-0.05
Wage	-0.50***	-0.21	-0.40*
Prior click	1.06***	1.16***	1.32***
Message	-0.26	0.02	0.09
Video	0.05	-0.04	-0.78**

These results further aid interpretation of findings regarding demographic variables predictive of phishing susceptibility. Statistical control supported the independent effect of age, position, and gender though these effects depended on email type. The analysis further supported the effect of the video feedback intervention for certain email types.

6. Discussion

The best predictor of phishing susceptibility may be *having been phished before*. Individuals who clicked on a previous week's email are significantly more likely than non-clickers to click on the next week's email. Contrary to our initial hypothesis, we found varied effects of LP on subsequent week behavior. It is possible that those most susceptible to phishing are also those unlikely to patiently read/view feedback. The positive finding for the banking email may reflect that any feedback effect may only apply to high stakes scenarios (i.e., falling for a banking scam [fiscal damage] is likely more impactful than an IT scam [computer damage]). This should be investigated in the future.

There was a statistically significant association between age or position and phishing susceptibility. Department type is not significantly related to phishing susceptibility. The effect of gender was less consistent with previous literature. We found that gender is a small significant factor, which contradicts the result in Mohebzada et al. [11] and Diaz et al. [1], although they found a non-significant higher rate for males. Additionally, even though Jagatic et al. [9], Sheng et al. [13], and Halevi et al. [8] indicated that gender is a significant factor in phishing susceptibility, the direction of our result is inconsistent with their results claiming that females are more likely to click [8][9][13]. Furthermore, considering the gender click behavior by each email content, our result shows that males are significantly more likely to click on the financial email than females. Individuals in the youngest age group (less than 27 years old) are significantly more likely to click on the financial email than people aged between 27 and 41. Those in the oldest age group (greater than 59 years old) are significantly more likely to click on the financial email than people aged 27 to 59, and moreover, they are more likely to fall for IT help desk email and package delivery email than other groups. Our findings suggest that people over 59 years old may be the most vulnerable group to all three phishing email content types.

There were more clickers on the IT help desk email than the other two emails. This result may suggest that university employees pay more attention to emails related to their work context, which is consistent with findings from Greene et al. [5]. The financial email fooled the smallest proportion of users, which may suggest that people are more alert to the emails that come from an unfamiliar bank that they were not enrolled. As the domain name of the financial email ended by ".co" instead of ".com", this may also explain the smaller proportion of users who were deceived by the financial content. The urgency cues [10] in the three emails might triggered users to believe they are genuine emails. This phenomenon will be investigated in our post-experiment survey.

Our study design had several important aspects that help to disambiguate results of previous studies and clarify implications for IT policy and practice: (a) *Varying phishing email types*. We used three types of simulated phishing emails (one related to IT/tech support, one related to finance/banking, and one related to e-commerce/package delivery) to increase the generality of findings and assess any differences in vulnerability to different types of email phishing exploits. (b) *Stratified sampling of users.* Users were grouped by stratified sampling to ensure each user would receive all emails but in different days in three weeks, to reduce influence of possible confounding factors. (c) *Large-scale study*. This study used a large number of subjects and a wide age range of users, enabling us to disambiguate some of the discrepant findings previously reported on demographic factors.

Some possible limitations of this study should be considered in planning future research. While the multivariate statistical models allowed variables to be tested while controlling for all other variables, it is still possible that unmeasured variables may be the underlying cause of some the relationships. First, for the clickers who received the message or the video LP in the previous week, we did not evaluate how carefully they read the message or watched the embedded anti-phishing training video. In the post-study survey, we asked if people watched the video and found that click behavior has no significant relationship with whether people viewed the video (although this result is restricted to survey respondents). A second limitation is that the results related to the email content analysis are restricted to the specific email designs that we used; since we didn't include multiple versions of each type, it would be risky to generalize these results.

7. Future work

Future research should explore the effect of email type on phishing susceptibility; and plans also include analysis of other behavioral factors that were collected in this study but not yet examined. Understanding these factors and characteristics will enable development of IT policies and practices, better defensive software tools, and more effective, perhaps tailored, awareness training for the most susceptible users.

Acknowledgments

This research was supported under IARPA contract 2016-16031400006. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Government. The research study was reviewed and approved by the George Mason University IRB. The authors gratefully acknowledge contributions from team members: Dr. James Jones, Dr. Edward Huang, Dr. Syed Abbas Zaidi, Dr. Yung Mei Leong, Dr. Shou Matsumoto, Mr. Mohanad Ajina, Mr. Abdul Alsayed, for their valuable technical support; Dr. Dennis Buede, Dr. Paul Sticha, and Dr. David Brown, for their support and helpful feedback.

8. References

[1] A. Diaz, A. T. Sherman, & A. Joshi, "Phishing in an Academic Community: A Study of User Susceptibility and Behavior," 2018, arXiv preprint arXiv:1811.06078.

[2] J. S. Downs, M. B Holbrook, & L. F Cranor. "Decision Strategies and Susceptibility to Phishing." *Symposium On Usable Privacy and Security (SOUPS)*, July 12-14, 2006, Pittsburgh, PA, USA.

[3] Federal Bureau of Investigation's Internet Crime Complaint Center (IC3). 2018 Internet Crime Report. Retrieved from: https://pdf.ic3.gov/2018_IC3Report.pdf

[4] S. Furnell, "Phishing: can we spot the signs?," Computer Fraud & Security, 2007(3), 10-15.

[5] K. K. Greene, M. P. Steves, M. F. Theofanos, & J. Kostick, "User Context: An Explanatory Variable in Phishing Susceptibility," In in *Proc. 2018 Workshop Usable Security*.

[6] F. L. Greitzer, J. Strozer, S. Cohen, J. Bergey, J. Cowley, A. Moore, & D. Mundie, Unintentional insider threat: contributing factors, observables, and mitigation strategies. In *2014 47th Hawaii International Conference on System Sciences*, January, 2025-2034. IEEE.

[7] F. L. Greitzer, J. R. Strozer, S. Cohen, A. P. Moore, D. Mundie, & J. Cowley, Analysis of unintentional insider threats deriving from social engineering exploits. In *2014 IEEE Security and Privacy Workshops*. 236-250

[8] T. Halevi, J. Lewis, & N. Memon. Phishing, Personality Traits and Facebook. Cornell University Library. 2013. http://arxiv.org/abs/1301.7643

[9] T. N. Jagatic, N. A. Johnson, M. Jakobsson, & F. Menczer, "Social phishing," Communications of the ACM, 50(10), 94-100, 2007.

[10] J. D. Lee, W. Li, S. Matsumoto, M. Ajina, B. Yousefi, J. Jones & K. B. Laskey. "GMU Phishing Study Technical Report", George Mason University Technical Report, Fairfax, Virginia, Feb – May 2019.

[11] J. G. Mohebzada, A. El Zarka, A. H. BHojani, & A. Darwish, "Phishing in a university community: Two large scale phishing experiments," *Innovations in Information Technology (IIT), 2012 Int'l Conference*, 249-254, IEEE.

[12] PhishLabs, 2019 Phishing Trends and Intelligence Report. Retrieved Sep. 2019, from https://info.phishlabs.com/hubfs/2019%20PTI%20Report/2019%20Phishing%20Trends%20and%20Intelligence%20Report.pdf

[13] S. Sheng, M. Holbrook, p. Kumaraguru, L. F. Cranor, & J. Downs, "Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 373-382, April 2010.

[14] A. Vishwanath, T. Herath, R. Chen, J. Wang, & H. Rao. Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. *Decision Support Systems*, 51(3), 576-586, 2011

[15] J. Wright, Open-Source Phishing Framework. Retrieved March 6, 2019, from https://getgophish.com/