Cross-National Study on Phishing Resilience

Shakthidhar Reddy Gopavaram
Indiana University
sgopavar@iu.edu

Jayati Dev
Indiana University
jdev@iu.edu

Marthie Grobler
CSIRO's Data61
marthie.grobler@data61.csiro.au

DongInn Kim
Indiana University
dikim@indiana.edu

Sanchari Das
University of Denver
Sanchari.Das@du.edu

L. Jean Camp
Indiana University
ljcamp@indiana.edu

***Abstract*—Phishing is a ubiquitous global problem that is both the simple crime of theft of authenticating information and the first step in advanced persistent attack chains. Despite receiving worldwide attention and investments in targeted anti-phishing campaigns, a large proportion of people are still vulnerable to phishing. This is not only due to the evolution of phishing attacks, but also due to the diversity of those exposed to phishing attacks in terms of demographics, jurisdiction, and technical expertise. To explore phishing resilience, we conducted a cross-national study to identify demographic and other factors that might have an impact on phishing resilience across nations. Specifically, we recruited 250 participants from the United States, Australia, New Zealand, Canada, and the United Kingdom to observe their responses to phishing websites in a simulated environment. We identified how factors including demographics, knowledge, skills, website familiarity, and self-reported risk assessment behaviors relate to efficacy in phishing detection. While participants' phishing knowledge, familiarity with the target website, and their reported use of the lock icon as a phishing indicator increases participants' probability of correctly identifying a legitimate website, we found that these factors did not specifically make them more resilient to phishing attacks. Our results further show that computer expertise has a significant positive impact on phishing resilience and that increased age correlates with the probability of misconstruing a phishing site as legitimate. These findings were applicable across all five countries in our study.**

***Index Terms*—Phishing, Risk Assessment, Resilience, Cyber Security, Usability, Cross-National, Socio-Technical.**

I. Introduction

Phishing attacks are recognised globally as one of the leading factors in cyber insecurity. The Anti-Phishing Working Group (APWG) reported that the number of phishing attacks grew through 2020 and doubled over the course of the year, with October 2020 seeing an all-time high reporting of $225,304$ unique phishing websites [1]. Despite continuous anti-phishing campaigns, training, and other interventions run by many organizations across the world, the constantly evolving approaches used by phishers still manage to trick a large proportion of the global population [2], [3]. To determine whether there are common factors affecting resilience to phishing attacks in countries with similar socio-technical profiles, and to explore whether there could potentially be a global approach to limit the number of people that fall victim to phishing attacks, our study evaluates phishing resilience across the Five Eyes countries: United States (US), Australia (AU), New Zealand (NZ), Canada (CA) and the United Kingdom (UK). These countries are largely all English speaking nations with perceived similar linguistic and high level cultural similarity. Due to information sharing, they are collectively referred to as the Five Eyes. Through our cross-national phishing resilience study, we aim to address the following research questions.

- RQ1: How do demographic factors relate to phishing resilience, specifically gender and age?
- RQ2: How is computer expertise related to phishing resilience, specifically knowledge and skills?
- RQ3: How does familiarity with websites affect phishing resilience?
- RQ4: How does risk assessment behaviour vary between countries in the Five Eyes and how does that behavior relate to phishing resilience?

In this paper, we define phishing resilience as including two dimensions of efficacy in phishing detection: the ability to correctly identify a phishing site and the ability to correctly identify a legitimate site, thus avoiding the associated harm of loss of access to legitimate resources.

By addressing the above research questions and further exploring the socio-technical components of traditionally understood cyber security awareness, our research aims to address a gap in the application of anti-phishing theory and contribute to global phishing resilience. Our work aims to identify commonalities amongst comparable populations, whilst identifying the factors that would result in differentiating behavior. This work is an implementation of the epidemiological approach proposed in [4], recognising that Internet users are individualistic and that human behavior in specific situations would not always be exactly the same.

This paper is structured as follows, with Section II presenting a brief overview of related works in the phishing domain.

Workshop on Usable Security and Privacy (USEC) 2021
7 May 2021, Auckland, New Zealand
ISBN 1-891562-73-8
https://dx.doi.org/10.14722/usec.2021.23017
www.ndss-symposium.org

Section III details the methodology followed in our global experiment, followed by Section IV that details the results. Section V provides a discussion, and Section VI concludes the paper.

II. Related Work

Phishing is a crime employing both social engineering and technical subterfuge to steal consumers' personal identity data and account credentials. Social engineering schemes target people with attackers masquerading as trusted, legitimate parties, such as by using fake email addresses and deceptive email messages. Phishing has a significant impact on the global economy, with on average 200,000 unique phishing websites and 130,000 unique phishing email subjects detected per month. Phishing is also often a critical initial component of complex attack chains [1].

Many studies have been conducted on phishing focusing on different behavioral factors, demographical factors, susceptibility and resilience, as well as technical aspects of phishing site identification [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. Despite being an ongoing research topic, little is confirmed about the individualistic and behavioral component that would make some individuals more regularly fall prey to phishing attempts as there are few accepted instruments for measuring resilience and few meta-studies that compare such results [9]. A study by Graves *et al.* [15], experimentally manipulated key variables in measuring attitudes to cyber crime judgements, including attack motivation, scope, and value. Canetti *et al.* [16] used simulated cyber attacks to measure changes in cortisol levels of victims, but focused on a small population. Cox *et al.* [17] investigated how people assess the likelihood of personal risk on online activity, whilst Ramkumar *et al.* [18] investigated how people use visual interpretation of security cues to drive their security behavior.

A number of studies have explored how national culture and location affect phishing resilience. Results from these studies show that national culture has a significant moderating effect on cyber security related activities. Professors Flores and Ekstedt identified a dominant 'clan mentality' that encompasses and integrates various aspects of the economy and society [19]. Organizational and cultural variables effect the perceived norms, attitudes, and behaviors associated with cyber security [20], [21]. In addition, the business and political landscape surrounding cyber security is also influenced by jurisdiction [22]. In this study, we examine participants from five nations, but do not further divide them into cultural groups. The distinction between cultural, organizational, and national influences is profoundly important and deeply nuanced. [23] This discussion is beyond the scope of this paper; however, given that we did identify differences between the participants' phishing resilience investigation of these cultural dimensions is an area of future work.

Global studies have not yet converged on basic cyber risk indicators. For example, the study by Van De Weijer and Leukfeldt [24] found no relationship between personality factors and cyber crime victimization, beyond general crime victimization. Studies in cultural susceptibility also do not provide consistent results, with multiple factors contributing to online risk resilience [25]. Demographics have been shown to have different implications for different populations in privacy studies, and these are often confounded by correlations with expertise.

For example, an evaluation of high school students and adults in the same community in the United States found no difference in resilience; however, the students were far more confident of their choices [26]. A larger scale study in Europe found that adolescents differ from other populations in technology use [27], with a younger population (18–26) being the most susceptible to online deception. Other researchers hypothesized that younger cohorts would be more resilient, being the most technologically enhanced generation [28]. Since phishing resilience may be a function of risk acceptance, and women are found to be more risk averse [29] it is reasonable to assume a correlation with gender and phishing decision-making. Yet empirical investigations have found gender to have no impact [30], whilst both men [31] and women are susceptible to fraud [3], according to different studies. A similar range of results have been found with age; with no consistent conclusion. One difficulty in evaluating these differences is that there were not measures of expertise beyond level of education.

It is critical to identify the best approach to communicate to diverse user populations about appropriate, safe, and secure Internet behavior. Although customisation of anti-phishing mechanisms are required to cater for regional and cultural differences, a single global anti-phishing mechanism would be advantageous in terms of resource availability. Our research works towards such a goal by establishing a better understanding for how specific factors are related to phishing resilience.

III. Experiment Design

The primary goal of this study is to understand how factors like demographics, knowledge, skills, website familiarity, and risk assessment behaviors relate to phishing resilience across the nations in the Five Eyes. Therefore, we conducted an experiment where the evaluated phishing resilience of participants from Australia, Canada, New Zealand, United Kingdom, and the United Sates in a simulated environment and assessed their relationship with the above mentioned factors. In this section, we provide detailed information about our choice of countries, the simulated environment, performance bonus, and experiment procedure.

A. Choice of Countries

We selected the countries form the Five Eyes for the following reasons:

- Firstly, because of their linguistic and arguably cultural similarity [23], [32].
- Secondly, because these allies are likely to be targeted by the same geopolitical opponents.
- Thirdly, because they share intelligence pertaining to national security.

B. Simulated Environment

In this study, we wanted to evaluate participants' resilience to phishing without putting them in actual risk. Therefore, we created a simulated environment to present participants with a series of legitimate and phishing websites and recorded their responses. To build the simulator, we created a dataset of website images which consisted of screenshots of actual websites appearing on a Firefox browser for legitimate variants and modified screenshots for phishing variants. We then presented these images in the browser with the user interface chrome disabled. This was done to give participants the appearance that they were actually viewing the website on a browser. Finally, we bitmapped the images to make the back button on the browser chrome and login button for the website clickable. Participants were asked to click on the back button if they thought the website was a phishing website and click on the login button if they perceived the website to be a legitimate one.

C. Performance Bonus

Each participant was presented with 13 phishing websites and 13 legitimate websites. The order of presentation was randomized for each participant. Participants were told that they could receive up to $1 in bonus pay for completing the experiment fast. However, there was a time penalty associated with each incorrect response. For example, when a participant correctly identifies a website and clicked on the appropriate button, they were presented with the next website in the randomized series. But if they were to misidentify the website and respond by clicking on the incorrect button, then they would receive a 15 second time penalty before being redirected to the next website. Introducing this performance bonus for speed and accuracy is a useful way to study risky environments without putting participants in actual risk [10].

D. Experiment Procedure

For this experiment, we recruited a total of 250 participants on Prolific [33] (50 participants per country). The study was conducted on the same day simultaneously across the five countries. To participate in this study participants had to be 18 years or older and be nationals of one of the five countries in the Five Eyes. Additionally, participant could only take this study on their desktop or laptop as the simulated environment was designed for larger screens. Each participant received $4 base pay and up to $1 bonus pay based on their performance.

Upon clicking on the Prolific task participants were presented with a Study Information Sheet (SIS). Upon reading the study information sheet and agreeing to take part in the study we requested *basic demographic information* from the participants, including their Prolific ID to track their responses throughout the experiment. The Prolific ID was used to make sure that participants could only participate in the study once.

After responding to the demographics questions, participants were presented with a *BART (Balloon Analog Risk Task) experiment* [34]. BART was used to measure participants' baseline risk-taking behavior before they started the phishing experiment[1].

After completing the BART experiment, participants were provided with instructions for the phishing experiment. We included questions to test their understanding of the instructions and to make sure they were aware of the controls, bonus pay, and the time penalty. Upon answering all the comprehension question correctly participants were presented with a series of 26 phishing and legitimates websites. The order of presentation of the websites was randomized for each participant.

After completing the phishing experiment, participants were asked to complete a survey which included questions about *website familiarity, security knowledge, computer expertise, and website risk assessment behavior*. The security knowledge and computer expertise questions were adopted from Rajivan *et al.* [35]. These questions are included in the Appendix.

E. Study Limitations

In our experiment, we specifically focused on controlling participants' risk assessment and resilience towards changed domain name in the website URL. We did not consider other possible modifications within the website design. This decision was specifically made in terms of our experiment design to exclude multiple variables from interacting and potentially interfering with each other, there is a possibility that participants might have looked at other website indicators to inform their choice. This is measured through a choice of indicator question the survey.

Upon reviewing the responses of 250 participants recruited for the study, we found that 25 responses were either incomplete or had multiple responses. We excluded them from our study which reduced our survey response pool to 225 participants. The study is therefore limited to between 43 and 48 responses for each of the countries, after exclusion of poor quality responses. Thus, our findings are specific to this participant group and not to the entire population. Nonetheless, our study provides insights in terms of security knowledge, computer expertise, and nationality based findings, which can be used to create more holistic phishing training practices. Additionally, survey respondents on Prolific tend to be more technologically adept [33] than the general population, which may skew our sample population to some extent.

IV. Results

In this section, we detail the analysis and findings to answer the four research questions presented in Section I. We first look at descriptive statistics for demographic and technical expertise variables and conduct statistical tests to identify differences between countries. We then examine distributions for phishing resilience and conduct statistical tests to identify significant difference between countries. Finally, we look at participants' ability to correctly identify legitimate websites and how they differed between countries.

[1]Although BART was included as a pre-test in our cross-national coordinated phishing resilience evaluation, the details and design of BART is outside the scope of this paper.

A. Difference in Demographics

The premise of many research studies is that people belonging to certain demographics would be more or less vulnerable to phishing attacks [35]. In order to operationalize the independent variables for RQ1, *How do demographic factors relate to phishing resilience, specifically gender and age?*, we wanted to identify statistically significant differences in demographic factors between countries and determine if those difference may have contributed to differences in phishing resilience. In this study, we recorded two demographic factors, age and gender. The distribution for these demographics is shown in Table I.

		AU	CA	NZ	UK	US
Age (years)	18-30	21	28	16	26	29
	31-40	11	10	16	14	11
	41-50	5	7	10	0	2
	51+	6	3	4	4	2
Gender	Other/ self-described	1	1	0	1	2
	Female	29	27	25	24	22
	Male	13	20	21	19	20
Total		43	48	46	44	44

TABLE I
PARTICIPANT DEMOGRAPHICS ACROSS SURVEY RESPONDENTS

One-way ANOVA is a test which helps us determine if there are statistically significant differences in means between groups. However, it does not specify which two groups are significantly different. A t-test on the other hand, compares two independent groups and helps us determine if the means of the two groups are significantly different from each other. In this paper, we we first conduct one-way ANOVA test to determine if there are statistically significant differences in means between countries. If the test shows significant differences between countries, we conduct pair-wise analysis to identify pairs of countries that are significantly different from each other. In this paper, we adjust the p-values for multiple testing using Benjamini & Hochberg method [36].

The results from our analysis show that the differences in distributions of age are statistically significant between countries (p-value: 0.021, F-value: 2.92). P-value is the probability of obtaining the observed result when the null hypothesis (no differences in distributions of age between groups) is true. The F-score is the ratio of variance between groups to the variance within groups. F-Score is close to 1 if the null hypothesis is true. Pair-wise comparisons only found the differences in distributions of age between New Zealand and United States to be statistically significant (p-value: 0.026, t: 3.095). Here, participants from the United States have a lower mean age when compared to New Zealand (USA: $\mu = 28.2$, NZ: $\mu = 35.28$).

We also conducted Kruskal–Wallis test (a non-parametric alternative for one-way ANOVA) to identify differences in distributions of gender between countries. The results show that there is no statistically significant difference in gender between countries (chi-squared: 5.154, p-value: 0.2718).

B. Difference in Security Knowledge and Computer Expertise

To address the independent variables for RQ2, *How is computer expertise related to phishing resilience, specifically knowledge and skills?*, we asked participants three sets of questions to measure their level of knowledge. We included these questions to determine if participants with security knowledge and computer expertise were more or less resilient to phishing attacks. We asked these questions after participants completed the phishing experiment section of the study to avoid introducing bias in responses.

The first set of questions were aimed at measuring participants' knowledge on phishing and certificates (Questions 1 and 2 in the Appendix). These were multiple choice questions and required participants to select all the correct answers. We computed the scores for participants' phishing and certificate knowledge by using the following formula: $(\#CorrectOptionsSelected + 1)/(\#WrongOptionsSelected + 1)$. We chose this method of scoring because it gave us the ability to award partial points to participants who selected both correct and wrong options. For visualization purposes we normalized the scores for these questions to be on the same scale, as shown in Figure 1. The distributions show that majority of the participants from all five countries had low scores for phishing and security knowledge. The median phishing knowledge was higher than the median certificate knowledge for all countries. We conducted one-way ANOVA tests to see if the distributions for phishing knowledge and certificate knowledge were significantly different between countries. The results from these tests show that the phishing knowledge (F-value: 1.86, p-value:0.119) and certificate knowledge (F-value: 0.763, p-value:0.55) are not significantly different between countries.

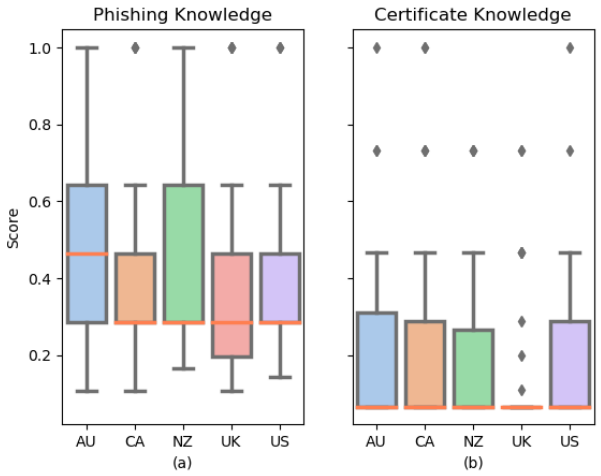

Fig. 1. Boxplots showing the country-wise distribution of (a) phishing knowledge scores and (b) certificate knowledge scores.

The second set of questions related to security knowledge in terms of SQL injection, Intrusion Detection Systems (IDS), and network operations, and were not specifically connected

with phishing (Questions 3, 4 and 5 in the Appendix). General security knowledge was measured on a three point scale, i.e. participants received one point for each correct answer on these second set of questions. Figure 2 provides country-wise distributions for general security knowledge. The distributions show that majority of the participants from all countries scored low on security knowledge. Given that security expertise is not common in the general population, it is not surprising that 75% of the participants from Canada, New Zealand, United Kingdom, and the United States had a general security knowledge score of zero. To see if distributions of general security knowledge were different between countries we conducted a one-way ANOVA test. The results showed that the general security knowledge is not significantly different between countries (F-value: 1.335, p-value: 0.258).

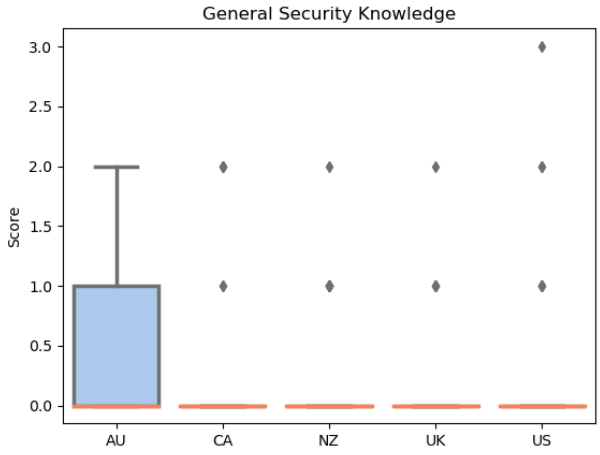

Fig. 2. Boxplot showing distributions for general security knowledge.

Finally, the third set of questions were aimed at measuring computer expertise. Here, we provided participants with a series of computer related tasks and asked them to select all the tasks that they have performed in the past (Questions 6 to 12 in the Appendix). These tasks varied from simple (such as installing a computer program) to complex (such as writing a computer program). Computer expertise was calculated by computing the total number of tasks performed by each participant. Figure 3 shows country-wise distributions for computer expertise. The distributions are similar for all countries, except for the United Kingdom which has a lower median compared when compared to the other countries. To see if computer expertise is significantly different between countries we conducted one-way ANOVA and pair-wise t-tests. The results from these tests show that the differences in computer expertise are statistically significant (F-value: 2.902, p-value: 0.022). Pair-wise comparisons showed that the distributions for computer expertise are significantly different between Australia and the United Kingdom (t:3.05, p-value: 0.002). Here, Australia has a higher mean ($\mu = 2.72$) when compared to the United Kingdom ($\mu = 1.52$).

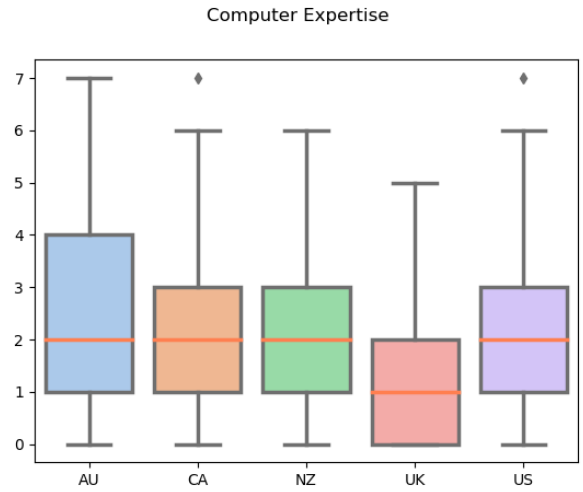

Fig. 3. Boxplot showing distribution of computer expertise scores for all five countries.

C. Differences in Website Familiarity

We collected information about participants' familiarity with the websites presented to them at the end of the experiment. We collected this information to observe the relationship between familiarity and phishing resilience. To operationalize the independent variable for RQ3, *How does familiarity with websites affect phishing resilience?*, we wanted to identify statistically significant differences in distributions of website familiarity between countries. Figure 4 provides country-wise distributions for website familiarity. To evaluate if the distributions for website familiarity varied between countries we conducted a one-way ANOVA test. Our results show that distributions of website familiarity are not significantly different between countries (F-value: 0.569, p-value: 0.68).

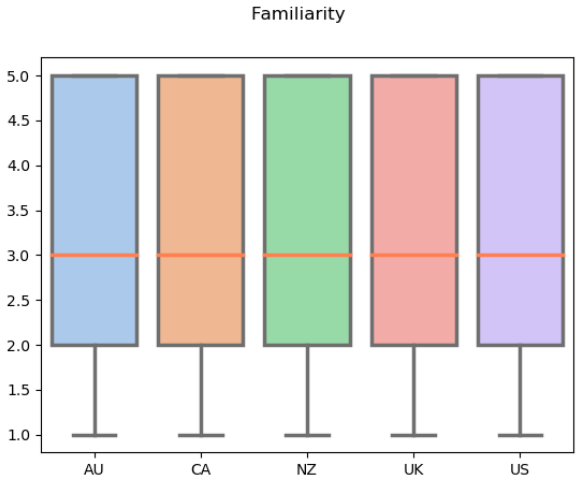

Fig. 4. Boxplot showing distribution for website familiarity.

D. Differences in Risk Assessment Behavior

To address RQ4, *How does risk assessment behaviour vary between countries in the Five Eyes and how does that behavior relate to phishing resilience?*, we asked participants to state the indicators that they used to determine if it is safe to enter their username and password on a particular website. To this, participants from all countries overwhelming reported that they use the lock icons and the HTTPS protocol as indicators. A good proportion of participants from all countries also reported using certificates, website type, and professional appearance to determine if it was safe to enter their login credentials. Relatively few participants reported using privacy policies as indicators for determining if a website is safe. Figure 5 shows a count plot representing the indicators used by participants to determine if a website is safe.

To see if the indicators used by participants are significantly different between countries we conducted Kruskal-Wallis tests (non-parametric alternative for one-way ANOVA). The test measured differences in frequencies of risk assessment indicators between countries. The results from our analysis did not find any statistically significant difference in risk assessment behavior.

E. Differences in Phishing Resilience

We define phishing resilience as including two dimensions of efficacy in phishing detection: the ability to correctly identify a phishing website and the ability to correctly identify a legitimate website, thus avoiding the associated harm of loss of access to legitimate resources. In this section, we identify statistically significant differences in the ability of participants from different countries to accurately identify phishing and legitimate websites.

1) Difference in Accuracy For Identifying Phishing Websites: Please recollect that in this experiment we presented each participant with 13 phishing websites and 13 legitimate websites in a simulated environment. For each website, we asked participants to take appropriate action based on if they thought the website was a phishing website or a legitimate one. To determine participants' ability to correctly identify phishing websites, we analyzed participants' responses to the 13 phishing websites. Specifically, we computed the fraction of phishing websites that participants accurately identified out of the total phishing websites presented to them. The higher the score, the better the ability of the participant to correctly identify a phishing website. Figure 6(a) provides the distributions for correctly identifying phishing websites for all five countries. The boxplots show that New Zealand and the United Kingdom have a lower median when compared to the remaining three countries. To identify if these differences are statistically significant we conducted one-way ANOVA test followed by pair-wise t-tests between countries. We found the differences between countries to be statistically significant (F-value: 2.902, p-value: 0.029). Pair-wise comparisons showed that the differences between AU and UK are statistically significant (t: 3.38, p-value:0.01). We adjusted the p-values for multiple testing. We elaborate on the factors that could explain these differences in Section IV-G.

2) Difference in Accuracy For Identifying Legitimate Websites: We further look at participants' resilience to incorrectly identifying legitimate websites as phishing websites. Specifically, we wanted to identify the fraction of websites that participants correctly identified as legitimate websites out of all the legitimate websites presented to them. The higher the score, the more resilient the participant is to incorrectly identify a legitimate website as a phishing website. Figure 6(b) provides country-wise distributions of accuracy for identifying legitimate websites. The boxplots show that Australia has the highest median accuracy followed by Canada and the United States. Once again, New Zealand and the United Kingdom have a lower median for accuracy when compared to the other three countries. To evaluate the statistical significance of these differences we conducted a one-way ANOVA test followed by pairwise t-tests. The results were adjusted for multiple testing. We found that the differences in the ability to correctly identify legitimated websites to be statistically significant (F-Value: 8.357, p-value $<$0.001). The results from pair-wise analysis are shown in Table II. The results show that Australia ($\mu = 0.93$) is significantly different from United Kingdom ($\mu = 0.79$) and New Zealand ($\mu = 0.84$). Additionally, Canada ($\mu = 0.89$) and the United States ($\mu = 0.88$) are significantly different from the United Kingdom ($\mu = 0.79$).

	p-value	t
AU-NZ	0.001	3.83
AU-UK	$<$0.001	4.60
CA-UK	0.003	3.40
UK-US	0.007	-3.038

TABLE II
PAIR-WISE ANALYSIS TO IDENTIFY SIGNIFICANT DIFFERENCES IN ACCURACY FOR IDENTIFYING LEGITIMATE WEBSITES BETWEEN COUNTRIES

In the next section, we consider to what extent the significant differences may be a result of differences in expertise or demographics among the participants.

F. Linear Mixed-Effects Models Analysis

We presented each participant with 26 websites (13 phishing and 13 legitimate) and recorded their responses for each of these websites. It is reasonable to expect the responses from the same participant to be correlated, i.e. one participant may be more or less accurate than the other. Additionally, it is also very likely that responses of participants from the same country will be more similar to each other than they would be to participants from a different country. While some of the similarity could be explained by measured characteristics such as demographics, expertise, and website familiarity, there may be other unmeasured attributes such as cultural characteristics that affect participants' behavior. To identify relationships that

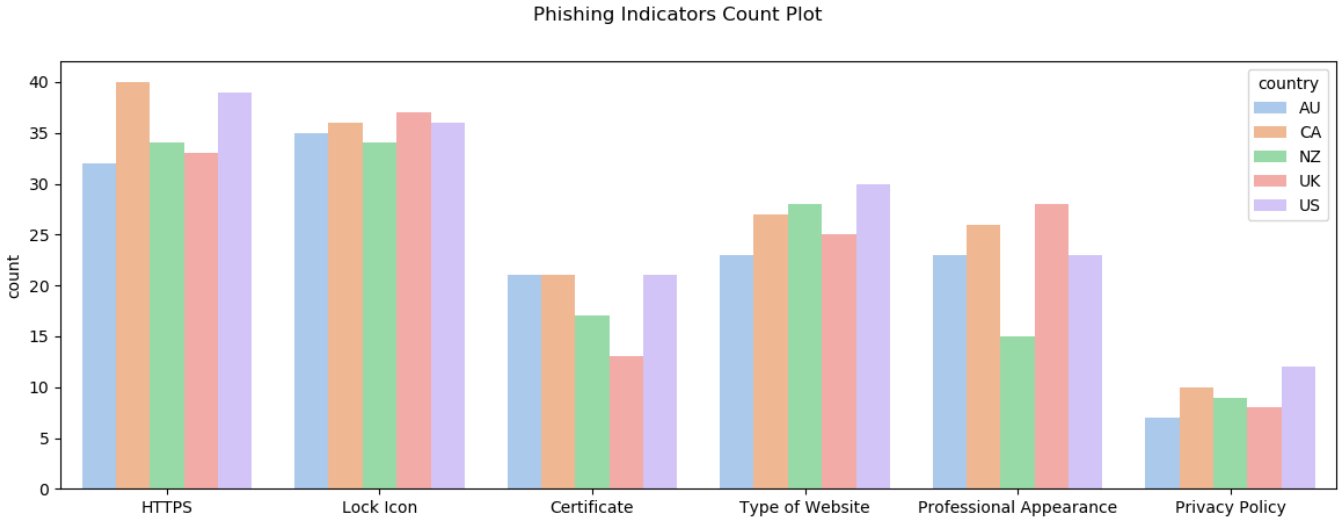

Fig. 5. Count plot showing the distribution of phishing indicators used by participants to determine if a website is safe.

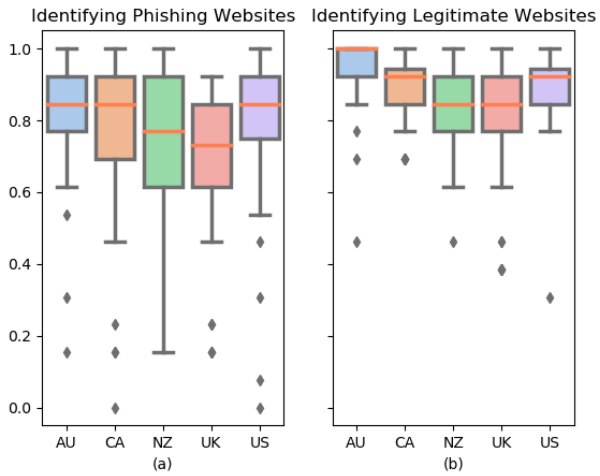

Fig. 6. Boxplots showing the distribution for (a) correctly identifying phishing websites, and (b) accuracy associated with identifying legitimate websites.

are generalizable across countries it is important to model the error variance arising from both multi-stage sampling and repeated measures. Linear Mixed-Effects Models (LMMs) provide us with the ability to model these error variances [37], [38]. Therefore, we use LMMs to understand the relationship of the measured factors with phishing resilience.

In this study, we use R and lme4 [39] to perform a linear mixed effects analysis. As fixed effects, we entered gender, phishing knowledge, certificate knowledge, general security knowledge, computer expertise, website familiarity, and phishing indicators. As random effects we had intercepts for participants and country. P-values were obtained by using normal approximation. The following subsections answer RQ1, RQ2, and RQ3 by performing LMMs analysis to identify factors impacting the ability to correctly identify phishing and legitimate websites.

G. Factors Impacting Accuracy For Correctly Identify Phishing Websites

Table III provides the results from the linear mixed-effects model analysis. The results indicate that age and computer expertise have a significant impact on the accuracy for correctly identify phishing websites. Age has a negative estimate. This indicates that older participants tend to be less accurate when compared to younger participants. Specifically, for every unit increase in age the accuracy for identify a phishing website drops by 0.2%. So a 20-year old person is expected to be 2% more accurate than a 30-year old and 4% more accurate than a 40-year old.

Computer expertise has a positive estimate. This indicates that a participant with a higher computer expertise is more accurate when compared to a participant with lower computer expertise. Specifically, for every unit increase in computer expertise the accuracy increases by 2.7%.

In Section IV-E1 we reported that the distributions of accuracy for identifying phishing websites are significantly different between Australia and the United Kingdom. This difference in accuracy could be due to the statistically significant differences in computer expertise between the two countries. Specifically, more participants from Australia had higher computer expertise when compared to those in the United Kingdom. This could have led Australia to have a higher accuracy when compared to the United Kingdom.

We also reported that the distributions of age are significantly different between New Zealand and the United States. This may have resulted in New Zealand having a lower median for accuracy when compared to the United States.

	Estimate	Std. Error	t-value	p-value
(Intercept)	0.667	0.074	8.947	<0.001
Male	-0.027	0.030	-0.915	0.359
Other	0.002	0.095	0.030	0.975
Female	0.	-	-	-
Age	**-0.002**	**0.001**	**-2.019**	**0.043**
Certificate Knowledge	-0.002	0.018	-0.123	0.901
Phishing Knowledge	-0.000	0.020	-0.007	0.994
General Security Knowledge	-0.023	0.029	-0.820	0.412
Computer Expertise	**0.027**	**0.009**	**2.988**	**0.002**
Familiarity	0.008	0.005	1.693	0.09
Phishing Indicators:				
HTTPS	0.030	0.034	0.869	0.384
Lock Icon	0.061	0.034	1.771	0.076
Certificate	0.014	0.031	0.466	0.640
Type of Website	0.057	0.032	1.772	0.076
Professional Looking Website	0.017	0.032	0.526	0.598
Privacy Policy	-0.035	0.036	-0.974	0.329

TABLE III
RESULTS FOR LMMS ANALYSIS SHOW THAT PARTICIPANTS' AGE AND COMPUTER EXPERTISE HAVE A SIGNIFICANT IMPACT ON ACCURACY FOR CORRECTLY IDENTIFYING PHISHING WEBSITES. SIGNIFICANT P-VALUES ARE HIGHLIGHTED

	Estimate	Std. Error	t-value	p-value
(Intercept)	0.716	0.047	15.249	<0.001
Male	0.016	0.017	0.967	0.333
Other	-0.031	0.053	-0.594	0.552
Female	0			
Age	**-0.002**	**0**	**-3.662**	**<0.001**
Certificate Knowledge	-0.005	0.0101	-0.565	0.571626
Phishing Knowledge	**0.023**	**0.011**	**2.012**	**0.044**
General Security Knowledge	**-0.042**	**0.016**	**-2.585**	**0.009**
Familiarity	**0.043**	**0.004**	**10.697**	**<0.001**
Computer Expertise	0.0056	0.005	1.076	0.281
Phishing Indicators				
HTTPS	0.0173	0.019	0.891	0.372
Lock Icon	**0.046**	**0.019**	**2.364**	**0.018**
Certificate	0.004	0.017	0.242	0.808
Type of Website	0.03	0.018	1.685	0.091
Professional Looking Website	-0.027	0.018	-1.508	0.131
Privacy Policy	0.0003	0.02	0.0176	0.985

TABLE IV
RESULTS FOR LMMS ANALYSIS SHOW THAT AGE, PHISHING KNOWLEDGE, GENERAL SECURITY KNOWLEDGE, WEBSITE FAMILIARITY, AND SELF-REPORTED ATTENTION TO THE LOCK ICON HAVE A SIGNIFICANT IMPACT ON ACCURACY FOR CORRECTLY IDENTIFYING LEGITIMATE WEBSITES. SIGNIFICANT P-VALUES ARE HIGHLIGHTED.

H. Factors Impacting Accuracy for Correctly Identifying Legitimate Websites

Table IV provides the results from the linear mixed-effects model analysis. The results show that age, phishing knowledge, general security knowledge, familiarity, and self-reported attention to the lock icon have a significant impact on accuracy for identifying legitimate websites. Both age and general security knowledge have a negative estimates. This shows that older participants and participants with higher general security knowledge are more likely to misidentify legitimate websites as phishing websites.

The misidentification of legitimate websites is a mirror of misidentification of a fraudulent website: for every unit increase in a age the chances for misidentification of legitimate website increase by 0.2%. For every unit increase in general security knowledge the chances for misidentification of a legitimate website increase by 4.2%. One possible explanation for this could be that older participants and participants with higher security knowledge are more sceptical when browsing the web and lack of effective risk indicators in the browser [5], [6], [7] is causing them to misidentify these legitimate websites as phishing websites.

The factors website familiarity, phishing knowledge, and self-reported attention to the lock icon are significant in this analysis, and have positive estimates. This suggests that participants with higher scores on these factors are less likely to misidentify legitimate websites as phishing websites. For every unit increase in website familiarity, phishing knowledge, and self-reported attention to the lock icon participants' likelihood for misidentifying legitimate websites drops by 4.3%, 2.3%, and 4.5% respectively.

In Section IV-E2, we reported that Australia's distribution of accuracy for identifying legitimate websites is significantly different form that of New Zealand and United Kingdom. However, other than Australia having a higher median for phishing knowledge we did not find any other significant difference between these countries. Therefore, the differences in the distributions of accuracy between these countries could be due to some factors unmeasured in our study.

United Kingdom's distribution accuracy was also significantly different from that of Canada and the United States. However, none of the measured factors were significantly different between these countries either. Therefore, the difference between these countries could also be due to unmeasured factors.

V. DISCUSSION AND IMPLICATIONS

Our primary goal was to understand how factors like demographics, knowledge, skills, website familiarity, and risk assessment behaviors relate to phishing resilience across the nations in the Five Eyes. Working towards this goal, we conducted a cross-national study on phishing resilience. After analyzing the results from our study, we found that:

- RQ1: Age had a significant negative relationship with phishing resilience. Specifically, older participants were more likely to misidentify both phishing and legitimate websites. We did not find any significant relationship between gender and phishing resilience.
- RQ2: Higher phishing knowledge increased participants' odds to correctly identify legitimate websites. However, this does not have a significant impact on correctly identifying phishing websites. Participants with high computer Expertise were more likely to correctly identify a phishing website.
- RQ3: Participants' familiarity with websites increased their chances for correctly identifying legitimate websites. However, their familiarity with a website did not increase their chances for identifying phishing websites.
- RQ4: We did not find any statistically significant differences in risk assessment behavior between countries.

The lock icon on websites was the only indicator that had a significant relationship with phishing resilience. Specifically, participants who reported paying attention to the lock icon were more likely to accurately identify a legitimate website.

While we found significant difference in the distribution of accuracy for correctly identifying legitimate websites between countries, we did not find any significant differences in distributions of phishing knowledge, general security knowledge, risk assessment indicators, and website familiarity. This indicates that there may other unmeasured factors that may be impacting phishing resilience.

Once we can understand the relationship between similar demographical participants in the global context, we can establish a global benchmark for an epidemiological approach to fully countering phishing attacks. A benchmark that incorporates demographic and expertise factors from an individual Internet user's perspective could prove to be invaluable in terms of guiding the design of effective anti-phishing mechanisms; and in targeted investments in stopping epidemics of global ecrime. Yet such a benchmark should integrate the differences across cultures as well as similarities. Our approach is targeted at identifying commonalities to leveraging these in behavioral expectations, as well as identifying the differences between the national groupings. This will support the goal of enabling complementary, global research that expands the current extant understanding of resilience to social engineering. Comparative global studies can also augment evaluations of national or regional studies, providing a baseline for comparison.

It is undeniable that phishing and malware are global challenges; and thus cross-national studies are a critical component of meeting this challenge. Studies such as this one offer an approach that could be used as a component to inform models of attack efficacy and diffusion that combine social engineering and technical components, or as a prerequisite for informing experiments related to interventions that lead to change, or as a source for analytic nuance when populations are known. The ultimate goal is to be able to identify the most vulnerable populations, and use that to craft interventions that can limit the spread of malware via the human agent [4].

VI. Conclusion

We present a cross-national evaluation of phishing resilience in the Five Eyes. All of the countries from which we selected participants are predominantly English-speaking and have populations that are Western, Industrialized, Educated, Rich, and the nations are Democratic (WEIRD) [40]. We have no expectation that the results could be generalized to other non-WEIRD nations. We explored the degree to which results in one of these countries may be applicable to others. To what extent might policies made for one nation, or technologies developed in one population, apply to the other?

One of the interesting findings of this study is that website familiarity, phishing knowledge, and attention to the lock icon only increases the odds that a person would not misidentify a legitimate website as a phishing website but it does not increase their accuracy for identifying phishing websites. The expansion of https to be ubiquitous on the HTTPS Everywhere [41] campaign may therefore make a human-centered contribution to phishing resilience beyond the protection of privacy provided by pervasive encryption of traffic.

In our analysis, for lack of adequate statistical sample size beyond that needed between group comparisons, we did not address the variance within the populations. Future work can be expanded to include representative populations for more in-depth investigation of the areas where the participants of different countries varied. Without this larger study it is necessary to be modest about our results. However, even with small sample we found significant differences and patterns of similarity. We also discovered that in every case different factors influenced sensitivity to false positives and false negatives. Before that, we would implement more detailed hierarchical analysis to identify differences in the weight of the various factors in different countries.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grant No. CNS 1565375 CNS 1814518; support from a Cisco Research Award No. 1377239, and funds from the Comcast Innovation Fund. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF, Cisco, Comcast, Indiana University, University of Denver, or CSIRO's Data61.

References

[1] Anti Phishing Working Group, "Phishing activity trends report: 4th quarter 2020," in *Activity October-December 2020*.

[2] S. Purkait, "Phishing counter measures and their effectiveness–literature review," *Information Management & Computer Security*, 2012.

[3] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and 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*, 2010, pp. 373–382.

[4] L. Camp, M. Grobler, J. Jang-Jaccard, C. Probst, K. Renaud, and P. Watters, "Conceptualizing human resilience in the face of the global epidemiology of cyber attacks," 2019.

[5] S. Egelman, L. F. Cranor, and J. Hong, "You've been warned: An empirical study of the effectiveness of web browser phishing warnings," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 1065–1074. [Online]. Available: https://doi-org.proxyiub.uits.iu.edu/10.1145/1357054.1357219

[6] J. S. Downs, M. B. Holbrook, and L. F. Cranor, "Decision strategies and susceptibility to phishing," in *Proceedings of the Second Symposium on Usable Privacy and Security*, ser. SOUPS '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 79–90. [Online]. Available: https://doi-org.proxyiub.uits.iu.edu/10.1145/1143120.1143131

[7] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 581–590. [Online]. Available: https://doi-org.proxyiub.uits.iu.edu/10.1145/1124772.1124861

[8] S. Das, A. Kim, Z. Tingle, and C. Nippert-Eng, "All about phishing exploring user research through a systematic literature review," in *Proceedings of the Thirteenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2019)*, 2019.

[9] A. Das, S. Baki, A. El Aassal, R. Verma, and A. Dunbar, "Sok: a comprehensive reexamination of phishing research from the security perspective," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 671–708, 2019.

[10] T. Kelley and B. I. Bertenthal, "Attention and past behavior, not security knowledge, modulate users' decisions to login to insecure websites," *Information & Computer Security*, 2016.

[11] S. Das, J. Abbott, S. Gopavaram, J. Blythe, and L. J. Camp, "User-Centered Risk Communication for Safer Browsing," in *Financial Cryptography and Data Security*, M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne, and M. Sala, Eds. Cham: Springer International Publishing, 2020, pp. 18–35.

[12] H. Abroshan, J. Devos, G. Poels, and E. Laermans, "Phishing Happens Beyond Technology: The Effects of Human Behaviors and Demographics on Each Step of a Phishing Process," *IEEE Access*, vol. 9, pp. 44 928–44 949, 2021.

[13] K. Althobaiti, N. Meng, and K. Vaniea, "I Don't Need an Expert! Making URL Phishing Features Human Comprehensible," *CHI Conference on Human Factors in Computing Systems (CHI '21)*, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445574

[14] S. Albakry, K. Vaniea, and M. K. Wolters, "What is This URL's Destination? Empirical Evaluation of Users' URL Reading," ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–12. [Online]. Available: https://doi-org.proxyiub.uits.iu.edu/10.1145/3313831.3376168

[15] J. Graves, A. Acquisti, and R. Anderson, "Experimental measurement of attitudes regarding cybercrime," in *13th Annual Workshop on the Economics of Information Security. Pennsylvania State University*, 2014.

[16] D. Canetti, M. Gross, I. Waismel-Manor, A. Levanon, and H. Cohen, "How cyberattacks terrorize: Cortisol and personal insecurity jump in the wake of cyberattacks," *Cyberpsychology, Behavior, and Social Networking*, vol. 20, no. 2, pp. 72–77, 2017.

[17] E. Cox, Q. Zhu, and E. Balcetis, "Stuck on a phishing lure: Differential use of base rates in self and social judgments of susceptibility to cyber risk," *Comprehensive Results in Social Psychology*, vol. 4(1), pp. 25–52, 2020.

[18] N. Ramkumar, V. Kothari, C. Mills, R. Koppel, J. Blythe, S. Smith, and A. Kun, "Eyes on URLs: Relating visual behavior to safety decisions," *Eye Tracking Research and Applications Symposium (ETRA)*, 2020.

[19] W. Rocha Flores and M. Ekstedt, "A model for investigating organizational impact on information security behavior," in *Proceedings of the Seventh Pre-ICIS Workshop on Information Security and Privacy, Orlando, December 15, 2012.*, 2012.

[20] C. W. Choo, "Information culture and organizational effectiveness," *International Journal of Information Management*, vol. 33, no. 5, pp. 775–779, 2013.

[21] D. Henshel, C. Sample, M. Cains, and B. Hoffman, "Integrating cultural factors into human factors framework and ontology for cyber attackers," in *Advances in Human Factors in Cybersecurity*. Springer, 2016, pp. 123–137.

[22] A. Onumo, A. Cullen, and I. Ullah-Awan, "An empirical study of cultural dimensions and cybersecurity development," in *2017 IEEE 5th International Conference on Future Internet of Things and Cloud (FiCloud)*. IEEE, 2017, pp. 70–76.

[23] G. Hofstede, "Dimensionalizing cultures: The hofstede model in context," *Online readings in psychology and culture*, vol. 2, no. 1, pp. 2307–0919, 2011.

[24] S. G. van de Weijer and E. R. Leukfeldt, "Big five personality traits of cybercrime victims," *Cyberpsychology, Behavior, and Social Networking*, vol. 20, no. 7, pp. 407–412, 2017.

[25] S. Ibrahim, "Social and contextual taxonomy of cybercrime: Socioeconomic theory of Nigerian cybercriminals," *International Journal of Law, Crime and Justice*, vol. 47, pp. 44–57, 2016.

[26] P. Unchit, S. Das, A. Kim, and L. J. Camp, "Quantifying susceptibility to spear phishing in a high school environment using signal detection theory," in *International Symposium on Human Aspects of Information Security and Assurance*. Springer, 2020, pp. 109–120.

[27] E. Hargittai and A. Hinnant, "Digital inequality: Differences in young adults' use of the internet," *Communication Research*, vol. 35, no. 5, pp. 602–621, 2008. [Online]. Available: https://doi.org/10.1177/0093650208321782

[28] R. Dodge, K. Coronges, and E. Rovira, "Empirical benefits of training to phishing susceptibility," in *Information Security and Privacy Research*, D. Gritzalis, S. Furnell, and M. Theoharidou, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 457–464.

[29] C. C. Eckel and P. J. Grossman, "Chapter 113 men, women and risk aversion: Experimental evidence," ser. Handbook of Experimental Economics Results, C. R. Plott and V. L. Smith, Eds. Elsevier, 2008, vol. 1, pp. 1061–1073. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574072207001138

[30] T. Lin, D. E. Capecci, D. M. Ellis, H. A. Rocha, S. Dommaraju, D. S. Oliveira, and N. C. Ebner, "Susceptibility to spear-phishing emails: Effects of internet user demographics and email content," *ACM Trans. Comput.-Hum. Interact.*, vol. 26, no. 5, Jul. 2019. [Online]. Available: https://doi.org/10.1145/3336141

[31] V. Garg and S. Nilizadeh, "Craigslist scams and community composition: Investigating online fraud victimization," in *International Workshop on Cyber Crime*. IEEE, 2013.

[32] H. Insights, "Compare countries," Jun 2020. [Online]. Available: https://www.hofstede-insights.com/product/compare-countries/

[33] S. Palan and C. Schitter, "Prolific.ac—a subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, vol. 17, pp. 22–27, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214635017300989

[34] C. W. Lejuez, J. P. Read, C. W. Kahler, J. B. Richards, S. E. Ramsey, G. L. Stuart, D. R. Strong, and R. A. Brown, "Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart)." *Journal of Experimental Psychology: Applied*, vol. 8, no. 2, p. 75, 2002.

[35] P. Rajivan, P. Moriano, T. Kelley, and L. J. Camp, "Factors in an end user security expertise instrument," *Information & Computer Security*, 2017.

[36] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[37] L. Meteyard and R. A. Davies, "Best practice guidance for linear mixed-effects models in psychological science," *Journal of Memory and Language*, vol. 112, p. 104092, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0749596X20300061

[38] B. Winter, "A very basic tutorial for performing linear mixed effects analyses," *arXiv preprint arXiv:1308.5499*, 2013.

[39] D. Bates, M. Maechler, B. Bolker *et al.*, "lme4: Linear mixed-effects models using s4 classes. r package version 0.999999-2," 2013.

[40] M. Muthukrishna, A. V. Bell, J. Henrich, C. M. Curtin, A. Gedranovich, J. McInerney, and B. Thue, "Beyond western, educated, industrial, rich, and democratic (weird) psychology: Measuring and mapping scales of cultural and psychological distance," *Psychological Science*, vol. 31, no. 6, pp. 678–701, 2020, pMID: 32437234. [Online]. Available: https://doi.org/10.1177/0956797620916782

[41] "Https everywhere," Jan 2020. [Online]. Available: https://www.eff.org/https-everywhere

APPENDIX A
POST EXPERIMENT SURVEY QUESTIONS

1) What is phishing?
2) What is the purpose of an X.509 certificate for websites?
3) SQL injection is a technique to (Select all that apply):
4) The difference between a passive and reactive Intrusion Detection System is?
5) Without any other changes in the default settings of a web server, what can be the motivation to close port 80?
6) Have you ever designed a website?
7) Have you ever registered a domain name?
8) Have you ever used SSH?
9) Have you ever configured a firewall?
10) Have you ever created a database?
11) Have you ever installed a computer program?
12) Have you ever written a computer program?
13) How many computer programming languages do you know (not including HTML)?

14) How many years of working experience do you have in network operation and security?
15) On average, how many times do you have to deal with computer security related problems?
16) What information and network security tools do you use regularly? (Firewall, Anti-virus, Intrusion Detection System (IDS), Secure Shell (SSH), Pretty Good Privacy (PGP), Access control (AC))
17) Which of the following indicators do you use to decide if it is safe to enter your username and password on a particular website? (https, lock icon on the page, certificate, website privacy statements, type of website, professional-looking website, other)