Knowledge and Capabilities that Non-Expert Users Bring to Phishing Detection

Rick Wash
Michigan State University

Norbert Nthala
Michigan State University

Emilee Rader
Michigan State University

Abstract

Phishing emails are scam communications that pretend to be something they are not in order to get people to take actions they otherwise would not. We surveyed a demographically matched sample of 297 people from across the United States and asked them to share their descriptions of a specific experience with a phishing email. Analyzing these experiences, we found that email users' experiences detecting phishing messages have many properties in common with how IT experts identify phishing. We also found that email users bring unique knowledge and valuable capabilities to this identification process that neither technical controls nor IT experts have. We suggest that targeting training toward how to use this uniqueness is likely to improve phishing prevention.

1 Introduction

Email is one of the most commonly used methods of communication, especially in large organizations and for e-commerce. Over 3.9 billion people have email accounts, and collectively they send and receive over 290 billion emails per day [11]. Email is one of the major methods that is used to communicate with strangers. However, because email is a global system where anyone can communicate with anyone, malicious actors send emails that pretend to be something that they are not, and trick people into taking actions that they otherwise wouldn't — which is known as phishing [34].

Phishing messages are an attack vector that has caused a large amount of damage in society. Phishing emails have been used to steal large amounts of money [22], install ransomware [31], or simply steal email contents that are later made public [21]. 32% of all corporate breaches in 2018 were due to phishing [33]. Spear-phishing – a variant where emails are custom targeted to the recipients – is used by 65% of groups doing targeted cyber-attacks, and is more commonly used than zero-day vulnerabilities (only 23% of such groups) [32].

Phishing is a socio-technical problem, and addressing the problem requires the coordinated work of both technological innovation and human intervention. Technologies are being developed that help identify and filter phishing messages, but these technologies do not work with 100% accuracy and can be slow to respond to new innovations by adversaries [14]. IT administrators and governments often try to stop phishing before it starts by disrupting phishing websites and bulk email sending [10]. But the last line of defense is the end user; phishing messages that go through these other defenses can still be detected or ignored by end users to prevent harm.

In this paper, we surveyed email end users without IT training or expertise and asked them about specific experiences with phishing emails they have received. Approximately half of survey respondents were able to identify a specific incident that they then answered detailed questions about. Building on Wash's [34] model of how IT experts detect phishing emails, we asked each person about what they noticed about the email, what they expected in the email, what made them suspicious of the email, what investigation they did, how they decided whether the email was legitimate, and what they finally did with the email.

From these questions, we are able to identify patterns in how email users who are not IT experts currently identify phishing scam emails in their own inboxes. Most research looks at phishing detection failures and what needs to be fixed; instead we compare non-experts with Wash's experts and identify what is working well that we can build upon. We find that email users often bring unique knowledge to this identification process that other phishing prevention methods do not have, such as whether the email was expected or not

USENIX Symposium on Usable Privacy and Security (SOUPS) 2021.
August 8–10, 2021, Vancouver, B.C., Canada.

and what emails like this typically look like and ask for. We also find that email users have valuable capabilities for investigation, such as asking other people for advice, or checking with senders for validity. Together, these findings suggest that email users can be an important part of the phishing prevention ecosystem, though phishing training can be improved to focus on how users can better use their unique knowledge and capabilities.

2 Previous Work

2.1 Preventing Harm from Phishing

Our society has three forms of defenses that help identify and limit the success of phishing scams. Technological defenses try to automatically detect known features of phishing emails and block or remove emails. Some defenses combine the work of computers and people by warning end users of the potential phishing message, which is then investiagted further by the end user to determine if it is a phishing email. And finally, there are human defenses, where the recipient of the email is relied upon to recognize the email as dangerous and act accordingly.

2.1.1 Automated Detection and Deletion

Automated detection and deletion approaches aim to classify emails as phishing or legitimate and block or remove them before the end user encounters them. Efforts in this space have focused on improving and finding new ways to identify outgoing and incoming phishing messages using blacklists [10], heuristics [3, 13, 16, 23], and machine learning [9, 29]. These approaches filter emails based on known features that conclusively identify emails as phishing.

Automated approches, however, rely on probabilistic algorithms which produce false positives, causing legitimate emails to be blocked or removed. In addition, automated approaches have limited ability to detect new permutations of phishing attacks [12] and cannot identify all older phishing emails.

2.1.2 Phishing Warnings

Phishing warnings augment automated detection techniques by warning end users of potential phishing emails, instead of blocking or removing them. Warnings are commonly used when automated detection cannot conclusively classify an email as phishing [25]. In practice, warnings have been reported to improve end users' ability to identify phishing emails [8, 26]. Ongoing research efforts in this area have focused on finding better ways to design and present warnings to the end user.

Despite their positive impact, warnings share the same limitations with automated detection and deletion approaches. They are prone to false positives (tagging legitimate emails as potentially dangerous) and false negatives (letting malicious emails through without warning, especially zero-hour phishing attacks). As Yang et al. argue, warnings and user training must complement each other to improve their effectiveness [37].

2.1.3 User Training

Security researchers and practitioners have developed various methods and materials for training users to identify and react to phishing emails accordingly. Kumaraguru et al. [19] and Caputo et al. [2] found that embedded training (i.e. instructional materials presented the moment a participant clicked on a URL in a phishing email), which is very commonly used in large organizations, improved user motivation to learn and enhanced knowledge acquisition. Rader et al. [27] found that people also learn about phishing scams and protective actions from stories about security incidents. Wash and Cooper [35] found that traditional facts-and-advice phishing training worked better when presented by an expert, while narrative security stories worked better when told by a peer.

The most widely shared phishing training messages across governments, businesses, and individuals teach people to identify certain cues (e.g. sender email address, URLs in emails, poor grammar or spelling) or apply a set of rules to detect, avoid and report phishing messages. Such training messages have been extensively studied and have shown potential to improve people's resistance to phishing attacks [4, 19]. Some messages focus on behavioral change, e.g., never click on a URL or open an attachment in an email from an unknown sender.

Other training messages focus on informing users of the common types of phishing threats and how to identify them, with the aim of manipulating the risk level and subsequently the level of fear in the users [5, 20]. Some researchers have argued that fear appeals increase end users' intentions to act securely. However, despite their ability to change behavioral intentions of end users [5], fear appeals do not predict or result in secure behavior [6].

User training typically focuses on aspects of the email message and tries to change the way people think about email messages so that they are paying attention to the features most associated with phishing. Studies have shown that this improves user knowledge, enhances their capabilities to identify phishing emails, and reduces the number of successful attacks [2, 19, 35]. However, the number of successful phishing attacks is still reasonably high, comprising 32% of all corporate breaches in 2018. More needs to be done to improve the capabilities of end users in identifying and preventing phishing attacks.

Most user training is developed from understanding how and why people fall for phishing [6]. We postulate that if training were to focus more on aspects of how people already

think about and deal with email in general, this can open up new avenues for phishing training. Unfortunately, we do not have a comprehensive understanding of how non-expert users do this. A similar problem was encountered in technical skills training where researchers investigated ways to improve the training of troubleshooters (technicians) [15]. They studied and identified a common conceptual process and strategies that technicians used when troubleshooting problems. This helped them to identify gaps in existing training methods and messages and subsequently helped them to identify areas of improvement. We argue that undertstanding the process(es) and strategies that non-experts use to identify phishing emails can reveal potential improvement areas for phishing training.

2.2 How Do People Identify Phishing Emails?

Downs et al. [7] investigated decision strategies of non-expert computer users when encountering suspicious emails. They identified three strategies that participants used to make sense of the emails they received: 1) this email appears to be for me; 2) it's normal to hear from companies you do business with and 3) reputable companies will send emails. Downs et al. [7] state that none of the strategies helped people to identify well-constructed phishing messages. The study, however, involved role-playing in a controlled environment. We do not know which of these strategies apply to and how prevalent they are in people's natural contexts and inboxes.

Wash [34] looked at how experts identify phishing emails by interviwing 21 IT experts about instances when they successfully identified emails as phishing in their inboxes. He identified a 3-stage process for identifying phishing emails. In the first stage, the email is received and treated like any other email — the content in the email is taken at face value and the person tries to make sense of the email and figure out what it is asking them to do. As they do this, they notice discrepancies — things that "feel off" about the email. Eventually, something triggers the person to think that this email is not legitimate — that it might be a phishing email that is not what it says it is. At this point, they become suspicious and begin explicitly looking for things that can help them determine if the email is legitimate or not. These new pieces of information often allow them to conclusively identify the email as phishing.

The work of Wash [34] demonstrates how some of the lessons from phishing training are applied in real-world contexts. However, Wash studied experts only. Experts might have more advanced skills, experience and knowledge about phishing and countermeasures compared to non-experts. We do not know which of the findings might apply to non-experts and can be used to improve their training.

2.3 Phishing: A Socio-Technical Problem

Phishing is a socio-technical problem. Automated solutions do not detect 100% of phishing emails. Hence end users must identify these emails in their inboxes. As Khonji et al. state, no single solution exists to mitigate phishing attacks [17]; thus automated / warning and user training techniques must be implemented to complement each other [19]. This is comparable to James Reason's Swiss Cheese Model (SCM) [28] of accident causation and response. SCM is a popular tool used to investigate or analyze the complexity of systems by showing that an incident is a result of a combination of active failures by operators and latent conditions of the system. SCM depicts socio-technical systems as multiple slices of Swiss cheese that are stacked together, each slice with a hole. Each slice depicts a layer of system defense against certain types of failures, while each hole represents failures in system defense at that particular layer. Bryans and Arief applied the model to understand security layers and fault-tolerance in computer systems [1]. They depict each layer as a protective mechanism against certain types of attacks, but has weaknesses (holes) against other types.

Both automated detection and deletion and warning techniques rely on the end user as the last line of defense against phishing. However, the number of recent successful phishing attacks suggests that more work needs to be done to improve user training. While most training focuses on teaching end users to identify known, conclusive features of phishing emails, Downs et al. [7] and Wash [34] found that end users rely on features other than conclusive distinguishers to identify phishing emails. We need to explore improved ways of keeping the user in the loop of defending against phishing attacks. More research needs to be done to understand how non-experts identify phishing emails, what aspects or information they rely on, and the kinds of things they do in the process. This understanding can help us to tailor and target phishing training and technologies that support human decision-making. Our study takes a first step in this direction by applying Wash's model in a survey to study the techniques that non-experts follow to identify phishing emails.

3 Methods and Sample

In this paper, we look at how non-expert users identify phishing emails, and look at whether some of the techniques that Wash [34] identified in experts continue to be present when non-experts identify phishing emails. To study this, we conducted a survey where we asked non-expert Internet users to remember a specific email that they received that was "suspicious or potentially harmful," and then answer questions about their experience with that email.

We asked questions to try to understand what they noticed and didn't notice about the emails respondents received and understand what kinds of things seemed important to them.

This is a retrospective account of a past email; we expect that respondents won't remember some of the details of what happened. We make the assumption that things they don't remember are most likely less important in their thinking about the email [18].

3.1 Survey

We started with a survey instrument that is loosely based on Rader et al. [27]. Near the beginning of the survey, we asked respondents to identify a specific "story" or incident where they received a suspicious or potentially dangerous email. We then asked them to answer a number of questions about that specific incident.

We included a screening question that asked potential respondents whether they could recall receiving the type of email we were interested in. The survey informed respondents that "In this survey, we are interested in hearing about emails you received that were suspicious or potentially harmful in some way." It then asked them to think back over their email, and told them it was OK to look back at their email if it would help. We asked "Can you remember any suspicious or potentially harmful email messages that you've received?" Only respondents who answered yes to this question proceeded on with the survey. 315 potential respondents that were otherwise qualified were excluded from the study because they did not answer "Yes" to this question.

Much like Rader et al. [27], we began the survey with an elicitation process to get respondents to identify a single "suspicious or potentially dangerous email" to answer questions about. The elicitation included three parts. First we asked respondents to write down in a short answer box "ways that an email message can be unsafe or cause security problems" and "ways you know of to recognize an email that is suspicious or potentially harmful." These prompts were intended to help trigger the respondent's memory of potential phishing emails. Respondents wrote an average of 12-14 words for each of these prompts.

Second, we asked the respondent to "think about times in the past when you personally received a suspicious or potentially harmful email" and "list as many of these emails as you can remember" in a text box. Respondents averaged 15 words in response to this prompt.

Third, we presented this list back to the respondent and asked the respondent to "Choose one email message from the list above that it's easy for you to recall details about." We asked them to briefly summarize that specific email. We presented this brief summary back to the respondent at the top of each subsequent page of the survey to help them remember which email they were answering questions about. These summaries averaged 21 words long.

The rest of the survey asked for more details about the specific email incident that was chosen by the respondents. Based on Wash's model [34], we identified six processes that experts use in phishing detection. We structured the questions around these six processes:

- **Noticing**: Things they noticed about the email, like when they received the email, what kind of mail (attachments, etc.), work or personal content, work or personal account, etc.
- **Expecting**: What they were expecting in the email; builds on noticing and compares what they noticed with what they expected. Have they received other emails like this, interacted with sender before, was the email expected, etc.
- **Suspecting**: What felt "off" about the email — subject, from, body, etc.. What in the email caused them to suspect the email. Did it contain links, attachments, etc.
- **Investigating**: What they went and explicitly looked for once they suspected the email (if anything) to figure out if the email was legit or fraud. Things like "did you look at headers, or hover over links, or try to contact the sender?"
- **Deciding**: How was the legit/phish decision made. Did you decide, and if so, how? How sure are you?
- **Acting**: After deciding, what did you do with the email? Report it? Just delete it? How did you feel about the email? Fear? Dread? Anxiety?

The complete survey instrument can be found in the supplementary materials.

3.2 Sample

We contracted with Qualtrics to field our survey to a panel of US participants in February 2020, which was just before the COVID pandemic. We excluded respondents who had had technical expertise or worked as technology professionals because we specifically wanted non-expert respondents. We placed quotas on age, gender, and ethnicity that roughly matched the US population, to try to get a more representative sample. We received a total of 297 valid responses. Respondents were compensated by Qualtrics with points that could be redeemed for items.

Table 1 summarizes the demographics of our sample. Our sample achieved the quotas and therefore roughly matches the US population along those lines. It also happened to come close to the US population in terms of education.

Only about 50% of our sample was currently employed either full-time or part-time. This is lower than in the US population (which was approximately 61% employed at the time of the survey [24]). This is the major way we believe our sample differs from the larger US population. We are not sure how this might affect responses about phishing emails.

	N	%		N	%
Age			**Employment**		
18-30	75	25%	Employed Full Time	105	35%
30-50	104	35%	Employed Part Time	42	14%
50-65	73	25%	Unemployed and looking for work	24	8%
Over 65	45	15%	Unemployed and not looking	25	8%
Gender			Retired	45	19%
Man	151	49%	Disabled	29	10%
Woman	156	50%	Student	16	5%
Other	2	1%	**Annual Household Income (USD)**		
Prefer not to answer	1	0%	Less than $25,000	66	22%
Ethnicity			$25,000 to $34,999	51	17%
White	202	64%	$35,000 to $49,999	35	12%
Hispanic, Latino, or Spanish	51	16%	$50,000 to $74,999	69	23%
Black or African American	37	12%	$75,000 to $99,999	33	11%
Asian	18	6%	$100,000 to $149,999	30	10%
American Indian or Alaska Native	8	3%	$150,000 to $199,999	7	2%
Education			$200,000 or more	6	2%
No College	71	24%			
Technical, Trade, or Vocational	22	7%			
Some college	102	34%			
College Degree	102	34%			

Table 1: Demographics of the survey sample. We received valid responses from a total of 297 respondents. Quotas were used on Age, Gender, and Ethnicity to approximately match demographics of the United States.

The majority of respondents in our sample had previous experience with cybersecurity incidents; only 17% of respondents indicated that they had not been a victim of a cybersecurity incident. About half of the sample reported having a virus (52%), and almost half reported having received a notification of a data breach (47%). Approximately one quarter (26%) had been the victim of credit card fraud, and 6% reported being a victim of identity theft more serious than credit card fraud. 18% reported having a device hacked. Interestingly, 16% of respondents reported having previously fallen for a phishing email or other scam email. These statistics suggest that our sample is also somewhat biased toward people who have had prior experience with cybersecurity incidents.

3.3 Analysis

Near the end of the survey, we asked respondents to "please write the story of the email as if you were telling it to a friend." We provided a large text box for the participant to enter in the story, and required that respondents enter at least 300 characters into this box. Respondents averaged over 400 characters (mean=411, min=300, max=1523), which is about 80 words per story on average (mean=81, min=41, max=288). We had two research assistants code these stories in parallel, meeting weekly to update the codebook, measure agreement, and resolve differences. We ended up with a codebook that coded stories for features organized in 5 categories: properties of the purported sender of the email; the action requested by the email; what felt off in the email; actions taken in the story; and final decision about the email.

After the training and codebook development, the two coders coded all 297 stories independently for a codebook of 39 distinct codes. After this initial coding, over half of the codes had a Cronbach's alpha above 0.7, and only 3 codes had an alpha below 0.5. We dropped the 3 codes with low agreement. The two coders then met and talked through all instances where there was disagreement and mutually agreed to a final decision about all codes for all stories.

In this paper, results from this manual coding will be explicitly labeled as such. Any results not labeled as resulting from manual coding are self-report data directly from questions in the main body of the survey. 13 (4%) of the stories were agreed to be "not a story" by both coders. These were instances where the participant filled out this text box for the whole survey, but did not describe an experience with a specific email, and instead described more general experiences. These responses are not included in statistics for the manual coding.

Replication materials for this analysis are available at https://osf.io/82sd9/. Additionally, all stories are presented exactly as they were entered by respondents, typos included.

4 Findings

In this survey, we asked respondents to identify "a suspicious or potentially harmful email message you received in the past." 315 otherwise qualified respondents were unable to identify an email, and 311 otherwise qualified respondents were able to do so. Quotas only applied to the qualified respondents who remembered such emails, and respondents were incentivized to remember such an email to participate in the survey and receive the incentive payment. Our goal was not to discover how prevalent phishing is among different demographic groups, and this sample should not be interpreted as measuring prevalence of phishing. However, it suggests that approximately 50% of the non-expert people in the Qualtrics subject pool have stories about specific phishing emails that they have received, which shows how widespread experience with these emails is.

Almost all of the remaining questions on the survey then asked the respondent for more details about the specific incident where they received that email that they chose to tell us about: what happened as they received it, what did they notice, and how did they handle it? In the majority of this paper, we report statistics about responses to multiple choice questions.

Based on findings from Wash [34], we organized the survey based on six different activities that a person needs to do to recognize a phishing email: 1) Noticing aspects of the email; 2) Forming expectations about what should and should not be in the email; 3) Becoming suspicious of the email; 4) Investigating the email; 5) Deciding whether the email is suspicious or not; and 6) Acting on that decision.

These six activities provide a way for us to describe what generally happens when a person receives a phishing email, and to look at patterns in what they notice and what they do. We organize our description of the findings in this paper around these six different activities.

4.1 Incidents

Each participant was asked to answer questions about a single incident that they experienced. We begin by describing the types of incidents that respondents reported on. Each incident was an email that the participant had received and decided was suspicious or potentially dangerous. All of these incidents represent emails that had made it through any technical defenses and into the participant's inbox, and so do not include phishing mails that were successfully filtered by technical phishing protections. Still, these emails were not uniform; respondents reported receiving a wide variety of different types of phishing scam emails.

We asked each respondent to identify a list of possible incidents / emails that would qualify, and then asked them to choose one that is "easy for you to recall details about" and then answer more questions about that one. We had a total of five questions that tried to understand broadly what these emails were about — one question near the beginning asking the respondent to summarize the incident, one question near the end asking the respondent to explain the whole incident, and then three questions asking for brief, 5-words descriptions of the chosen incident. Here we use these 5-word descriptions to describe the kinds of incidents that people reported on.

When asked to summarize the incident early in the survey, respondents responded with an average of 21 words (median: 17 words). In these summaries, respondents mostly reported facts about the email that they received, with the most common words being email (39% of respondents), account (17%), money (15%), link (13%) and received (11%).

In addition to the summary, we asked respondents, "In approximately five words" to describe what made the email suspicious, what made the email hard to figure out, and what the email was asking them to do. The respondents reported that they were suspicious mostly looking at the email / sender address or because it involved money. The emails were mostly asking respondents to click links (22%), for money (17%), or for "information" (14%). Together, these summaries suggest that most of the phishing stories were about economic issues (money) or asking for or providing information.

81% of respondents indicated that they found it easy to remember such an email. The emails that respondents chose to respond about were widely distributed in time: 24% of respondents received it within the last week; 30% within the last month (but not the last week); 25% within the last year (but not last month); and 15% more than a year ago.

In the manual coding, we coded the full incident stories for information about who the purported sender of the email was. This was not who actually sent the email, but who the email pretended to be from. 44% indicated that the email was from a group or organization, and 25% indicated that the email seemed to be from an individual. In 30% of the stories, the participant indicated that they had a pre-existing relationship with the purported sender, and 14% of the stories the participant explicitly stated that they did not have a pre-existing relationship. 76% of the pre-existing relationships were with a group or organization; suggesting that emails pretending to be from an organization were more likely to be seen as part of a pre-existing relationship.

As an example of a story about an email from an organization the participant had a pre-existing relationship with, consider the following story about an email from Amazon.com:

P233 Story: *I received an email that appeared to be from amazon. It had my name and address but said i owed money for a purchase. I hadn't purchased anything for a while so that seemed strange. Email had misspellings and an odd looking link. I looked closely at the email, then checked my amazon account on their website. There was nothing there about any orders or owing money.*

The actual senders varied widely across stories: about 12%

said it was a bank or financial institution, 8% said the email appeared to be from a foreign person, 4% from the government, and 2% from an IT support organization.

In the manual coding of stories, we also coded for what kind of information was being requested. 30% of the stories mentioned that the recipient of the email would receive some sort of valuable (money, award, gift, job offer, etc.), and 19% of the stories reported that the email asked the recipient to send money. 19% of the stories mentioned that the email was asking for personal information, 10% of the stories were asking for technical information such as usernames or passwords, and 10% of the stories were asking for financial information like bank account numbers, credit card numbers, etc. This suggests our respondents received emails with a wide range of requests, with no particular type of request being overwhelmingly common. What end users consider to be phishing is diverse, and training that focuses mostly on cues may miss classes of email messages that stand out to end users as potentially harmful.

4.2 Noticing

4.2.1 What people notice in an email

As a person reads an email, they cannot notice and remember everything about the email. Instead, the things in the email that the person can most easily make sense of and connect with are the easiest to notice and remember [18]. We asked respondents "What aspects of the email stood out to you?" and allowed them to check all that apply. The answers to this question show us, for these suspected phishing emails, what aspects of the email were most important to the respondents, because they were the most memorable.

By far, the aspect noticed by the largest number of people was that the email included a request for an action. 76% of respondents noticed this about the email. This corresponds well with past research that suggests that people tend to use email as a to-do list [36]; they quickly focus on what the email is asking them to do. It also corresponds with Wash's [34] finding that requests for actions (action links) were important triggers for experts.

The second most commonly noticed aspect of email was what the email was about, with 52% of respondents noticing this. The topic of the email, and whether that topic is relevant to the recipient of the email, is commonly seen as an important aspect of phishing. This data backs up that idea, and shows that this is something that people quickly are able to identify and remember about emails.

Much past work on phishing has focused on "conclusive distinguishers": aspects of an email that can help the recipient to conclusively distinguish legitimate emails from phishing emails, or at least strongly indicate phishing. For example, phishing training usually focuses on aspects such as inappropriate URLs in links, urgency in requests for action, or poor grammar/spelling. However, Wash emphasizes that when experts identify phishing emails in their own inboxes, they instead look for more minor discrepancies, which are things that seem off about the email, but don't necessarily indicate phishing and definitely are not enough on their own to conclusively identify phishing.

These first two things that respondents noticed — requests for action and topic of the email — do not conclusively indicate that the email is a phishing message, and are not normally part of phishing training. Instead, they simply indicate that there is something weird about the emails. However, for some people they might be enough. For example, consider this story:

P19 Story: *I got an email last Friday from one of the companies we work for that pays us to provide service for them and I immediately could tell it was a fake email because the company the email sender disguised themselves as is a company that pays us, we don't pay them.*

I called the company we work for and reported it to them so they would know someone was trying to disguise themselves as them

The next two most commonly noticed aspects of the email are much more commonly associated with phishing identification: links in the email (44%), mistakes or poor quality (41%). These are often found in phishing emails (especially the kinds of phishing emails that non-experts in our sample might be able to successfully detect).

38% of respondents reported that the sender's name stood out to them. The remaining aspects of email, such as attachments, images, formatting, or length of email, were noticed by less than 20% of respondents, though all of them were important to a non-trivial subset of users. This finding suggests that people seem to naturally notice actions and topics of email much more than they notice more conclusive distinguishers like URLs or typos. This is important, because a person cannot use a feature to detect phishing unless they first notice that feature.

4.2.2 Non-email features

In addition to noticing aspects of the email, there are a number of aspects of the situation that are not necessarily part of the email but nonetheless appear to be important and memorable to respondents.

90% of the respondents noticed that the email had come to their personal email account. None of our respondents chose the "I don't remember" option for which email account it arrived at. The account that the email arrived to is salient and memorable to respondents, and is possibly something that can be used to help identify suspicious email. Only 78% of respondents reported that the email was of a personal nature.

70% of the respondents reported that the email appeared to come from a company, business, or other organization(i.e.

from a person). Only 6% of respondents cannot remember who the email appeared to come from. The email sender appears to be a highly salient aspect of the email. It is interesting that 94% of respondents can remember who the email appeared to come from, but that fact only stood out to only 38% of them.

4.3 Expecting

When trying to understand and make sense of an email, people naturally fall back to what kinds of email they expect to receive, and to comparing the email with past emails that they have received [34].

Almost all of the suspicious emails arrived unexpectedly (95%). This seems to be one of the strongest aspects of phishing identification for our respondents. It is also something that users find relatively easy to identify, but is almost impossible to measure technically. That is, whether an email is expected or not is something that is a valuable piece of information that only the user has and computers do not.

However, just because the email was unexpected does not mean it was unfamiliar. 72% of respondents reported either "somewhat agree" or "strongly agree" to the statement "I felt like I had received other email messages like this one before.". That is, almost three quarters of the emails felt familiar to the recipients.

This fact both helps and hinders phishing detection. On the one hand, since the emails are familiar, people can easily integrate these into their lives and might not read them very carefully. On the other hand, as Wash [34] points out, when the email is similar to other, past emails, then it is possible to form expectations about what is typical in those past emails, and then compare this email to the past, similar emails and notice more things that are different or wrong about this email.

While respondents reported receiving emails similar to the suspicious email, the suspicious email was not a typical email. 86% of respondents chose "somewhat agree" or "strongly agree" about the statement "This email message seemed different from the email messages I typically receive."

Putting these findings together, suspicious emails that people remember are generally emails that are unexpected, different than the emails typically received, but often are like other emails that have been received before. The feeling that an email is suspicious, or unexpected, represents intuition, or a "gut feeling" about an email, and such intuitions are often important aspects of human decision-making [18].

Only 19% of respondents remembered receiving an email from this sender before. The remaining either had never received an email from the sender (45%) or were not sure (33%). So while the email felt familiar, the sender generally was not. Even more telling, only 12% of respondents had actually interacted with the sender before reading this email, and 80% of respondents checked "No" to having previously interacted with the sender. This suggests that non-experts remember and pay attention to who they interact with via email and that this piece of information is important to them as they process new emails.

4.4 Request

The definition of a "phishing" message in this paper is a message (email) that pretends to be something that it is not, in order to get the user to do something they wouldn't normally be willing to do. The second part of that definition is important; phishing isn't just fake email, but it is fake email that requests action.

We wanted to see what kinds of actions were being requested in the suspicious emails that people received and remembered. We asked the respondent whether the email was asking them to do any of a common set of actions. The most common action requested was clicking on a link, which was requested in 57% of the emails reported. This is unsurprising, as this is the stereotypical phishing email, though if anything the surprise was that 40+% of respondents did not remember a requesting link. Only 19% of emails reported asked the user to open an attachment.

46% of emails asked the recipient to respond to the email with some kind of information. That is, rather than using a webpage to collect information or attaching malicious code to the email, the email asked for a response. Responding to emails is a very normal, everyday activity. As an example, consider this story:

P20 Story: *I got an email and it was from an unknowns sender and it was from a different country. As for the country I am unsure of what country it came from. I did not recognize the sender at all. They told me that I won some type of lottery and that all I needed to do was verify my name address date of birth and I could get the money. Then they also said in order to get paid the money all I had to was verify the information and then they would send me the money into my bank account. Then in order for them to send it they needed me to provide them my bank account information my routing number and account number and the banks name and address. I found all of this very concerning and was always told to never give out my social security number or any other personal information to anyone asking for it.*

Almost a third of emails, or 32% of emails, asked the user to take some sort of action outside of the context of email. P39 was asked to make a phone call, for example:

P39 Story: *After receiving a fraud alert email requesting me to call a company I do business with, I checked the phone number, and it was not what I had on file. I also was unaware of any fraudulent activities involving me; however I had my doubts. Therefore I called the number requested, and they started to ask me questions to corroborate my identity. I was reluctant to provide any information, and they told me that*

they would not provide information to me because they were concerned about my identity.

After a bit of a discussion I terminated the call. Subsequently I called the firm at a number that was familiar to me. They wound up transfering me to the fraus department internally. The end result is that the email was legitimate, just poorly constructed). The good news is that therre was no fraus regarding my account.

Most summaries of phishing focus on technical means of information extraction (malicious links, malware attachments) [32], but this suggests that we should also examine non-technical means like simply replying to the email. These incidents that ask for responses or actions outside of email are important reminders that email is a small piece of much larger systems of work, and that email can often be a thing that triggers other types of work to be done. Anti-phishing systems cannot just focus on email; they also need to watch the other non-email work that people do in response to email.

Interestingly, 94% of respondents were able to identify at least one requested action by the suspicious emails. Requesting actions is part of the definition of phishing because it is these actions that the attackers are most interested in. It is good news that users seem to be quite attentive to what actions are being requested, which means this is something that is necessarily present in all phishing emails, and also something that users are good at identifying, which makes it a good place to focus training.

4.5 Suspecting

Our definition of phishing includes that the email is fraudulent — it either explicitly lies or lies by omission about some important aspect of the email. In order to become suspicious of the email, though, it isn't enough to just notice those aspects of the email. The recipient of the email also has to suspect that something is not right about the email.

We asked respondents about each part of the email and whether it felt normal or whether it felt "off" in some way. 59% of respondents reported that the subject line of the email felt "off" in some way. 70% of respondents reported that the sender information felt "off", and 75% of respondents said that the body of the email was "off" in some way. This suggests that all three aspects of an email can provide important clues to end users that an email might be phishing, though the body (content) of an email tends to help users more.

When a respondent felt that the sender was off, they were about twice as likely to indicate that the email address felt off than they were to indicate that the sender's name was the thing that felt wrong. Though, as P99's story shows, the name can also be important:

P99 Story: *Upon strolling through my email account I notice this bogus looking email from what should have been Social Security Administration.*

Except the administration was replaced with bureau & immediately I knew it was bogus.I politely pulled the lil trashcan up for a good old fashion delete session. I usually don't open up anything deemed be to good to be true or bogus or otherwise.

When a respondent felt that the body of the mail felt off, we provided a number of options to them for indicating what in the body felt off. 32% of respondents indicated that the body included unexpected typos or other similar issues. 28% indicated that the body included something strange that isn't normally seen in emails like this. These two aspects suggest that typos are definitely triggers for suspicion, but other strange aspects of emails are almost as common as a trigger.

15% indicated that the email was missing something important. 14% indicated that the email included less information than they would expect. And only 7% indicated that the email included more information than they would expect. To our respondents, phishing emails including less information or missing something triggered suspicions much more often than including too much information. This means that for non-expert end users, their expectations for how much information the emails in their inbox typically include is an important aspect of suspecting an email might be phishing.

4.6 Investigating

Wash [34] points out that people rarely go directly from treating an email as a real email to believing that it is a phishing email. Instead, there is an intermediate stage of "suspicion." When a person is suspicious of the email, they are not sure whether it is legitimate or fraudulent. During this suspicious stage, Wash [34] describes people as taking investigative steps to figure out whether the email is legitimate or not.

We asked respondents about the investigations that they did of their suspicious email. 24% of respondents indicated that they did not do any kind of investigation, and an additional 3% did not remember if they did. That means that 73% of respondents undertook at least one extra step to investigate the email to determine if it was legitimate or not.

The most common investigative step taken was to look more closely at the email address. 36% of respondents in this study indicated that they did this. Looking at the email address seems to be an important everyday step that non-expert users try when they are suspicious of an email.

P66 Story: *An email came in from Paypal describing that a subscription had been purchased with the amount and name of the company/person. I have never seen or heard of the indicated party and at first thought, it may have been a legitimate email. After debating to click the link to login to Paypal and stop the transaction, I hovered over the sender's information and saw the email address had absolutely nothing to do with PayPal's contact information.*

Only 12% of respondents indicated that they looked more closely at a link the email. 7% hovered over the link to see where it went, and 5% actually clicked on the link to see where it went. Link investigation is often mentioned in much phishing training, and it is disappointing that only 12% of respondents investigate links. It is especially disappointing that over a third of those respondents clicked the link as the investigative step.

On the other hand, 16% of respondents reported looking at the headers of the email. This was more common than we expected.

4.6.1 Investigating outside of the email

As mentioned above, emails are frequently just small parts of larger systems. During the investigation, it is possible to look outside of the email for additional information that can inform the decision. In one common method, 18% of respondents reported seeking out a second opinion about the email and asked someone else.

We specifically asked respondents about steps they took to learn more about the purported sender of the email. 82% of respondents reported that they did not take any steps to learn more about the sender, but the remaining 18% did. 9% went to the purported sender's website to get more information about the email. 6% tried to contact the sender via phone. And 1% talked to the sender face-to-face, such as P220:

P220 Story: *I got an email from my work email account from what I thought was my coworker. The body of the email was worded strangely and asked me to click on a suspicious link. I looked closely at the email address it was sent from and it was not exactly correct given my work email addresses. I went to who I thought was the sender face-to-face and asked if he sent the email. He said no and I went ahead and deleted the email.*

Too much phishing training focuses on teaching people to investigate suspicious emails by looking at features internal to the email, such as the sender's email address and links [19,30]. It is surprising that as many as 18% of our respondents took investigative steps outside of the email.

4.7 Deciding

Wash [34] found that after investigating the email, his expert participants would frequently come to a final decision about whether the email was legitimate or phishing. We asked our respondents whether they did come to a final decision, and if so, what that decision was. 80% of respondents did come to a final decision, and almost all of them decided that the email was definitely not safe (78% not safe, 2% safe). The remaining 20% were either still not sure (17%) or don't remember if they came to a decision (3%).

We asked respondents how confident they were in their final decision on a scale of 0 to 10. 69% of respondents chose the highest confidence option (10), and the average confidence was 8.9. Respondents reported very high levels of confidence in their decision about whether the email was safe or not.

4.8 Acting

After deciding whether the email is legitimate or phishing, one decision still remains: what should be done about the email? By far, the most common action was simply deleting the email. 78% of respondents reported that they deleted the email and moved on after deciding it was not safe. 32% indicated that they clicked a button in their interface to report the email as spam or as phishing. Only 4% left it in their inbox.

The survey only asked about actions we knew about ahead-of-time. In the manual coding, we were able to code for more actions. 43% of the respondents mentioned deleting the email, and 15% mentioned clicking a button to mark as spam or phish. Additionally, 9% discussed reporting it to authorities in their story in another way, such as calling an IT help desk.

32% explicitly mentioned a "negative action": that they intentionally chose to not do something (like open the email, or respond). These negative actions are often very strongly worded, and respondents seemed to feel strongly about them, often using language describing bad things to justify not doing things in the future. Consider, for example, how P115 justifies not answering phone calls:

P115 Story: *Computer was shut down because of inappropiate access to a potentially dangerous website. I was telling me that i had to pay a fine of $200 to gain access to my computer. I received a phone call about going to a local store to purchase gift cards. I went so far as going to the store to purchase the gift cards and upon checking out. the clerk at the register informed me that I was being scammed and not to buy these cards. In the meantime I had an open line to this scammer, which I promptly hung up on. Upon arriving home I kept getting phone calls from this person, which I never talked with again.*

An additional 9% of respondents reported taking increased precautions in the future, such as installing a virus scanner or being more careful with emails.

People also have emotional reactions to the email. We asked respondents about their experience of a set of emotions, including "nervous," "fear," "terror," "dread," "worry," and "anxiety". All emotions had very low scores, and no emotion averaged higher than 2.2 out of 5. Despite being unsafe, these emails did not evoke strong emotions from our respondents. Past phishing training, especially that derived from Protection Motivation Theory, has used fear appeals to motivate users [5, 20]. Based on this data, phishing emails generally do not lead to strong emotions, and this could explain why fear appeals do not motivate changes in behavior [6].

5 Discussion

5.1 Humans Identify Phishing Differently

Modern email systems involve multiple layers of protection against phishing attacks. Many email senders include checks for phishing as emails get sent. Most email systems include at least one, and often more than one technical system that filters out emails that are believed to be spam or phishing. Many of these systems also label emails as possibly phishing, as a warning to users (e.g., Google's email system [25]). And end users read emails and make legitimacy determinations on their own.

Reason's Swiss Cheese Model of filtering [28] suggests that when there is a chain of filters like this, the filters work best when each filter works on different principles or using different information than other filters in the chain. If two filters use the same information (e.g. sender from email address) in similar ways, then the holes in the cheese line up and malicious emails that get through one filter are also likely to get through the other. However, if two filters use different information, or operate on the information in fundamentally different ways, then each filter is likely to catch messages that the other filter misses, and including both filters makes the system more resilient to attacks than only including one.

In this paper, we present evidence that this final filter – humans reading emails and determining if an email is legitimate – operates in a very different way, using different knowledge and capabilities, than almost all of the technical filters. We found that humans possess important information that technical phishing filters do not have. They rely on their familiarity with related emails received in the past (72%) and their expectations of incoming emails (95%) to make sense of and become suspicious of phishing emails. This knowledge is highly contextual and very unique to each individual and their experiences. In addition, humans use their knowledge of what was typical in emails they received in the past to spot unexpected and missing important pieces of information in new emails. This information is critical for detecting zero-day phishing attacks, which technical solutions rarely detect [12].

Our respondents were able to notice the nature of the email (e.g. 78% noticed it was personal) and the email account in which the email was received. This requires knowledge of all email accounts a person has and the kinds of communications expected in each account based on how and what the person chooses to use each account for. It is very complex and challenging for technical filters to acquire such knowledge and apply it accordingly, lest they surveil individuals.

Second, we found that humans possess unique capabilities that they use to identify phishing messages, which technical filters do not have. 94% of the non-expert respondents were able of identify what action the email was asking them to do, and over three quarters said they explicitly noticed this about the email. Requests for action are not commonly part of many spam and phishing filters, and when they are, they are often limited in scope mostly by language issues (e.g. checking if the email contains a link to a login page and verifying if the login page is legitimate [23]). Even non-experts are highly attuned to these requests and can confidently identify them.

When filtering, humans also have an investigative ability that technical filters lack: they can choose to take additional time and look up more information from third party sources. A number of our respondents indicated that they would ask colleagues for advice or try to contact the purported sender of the email.

The above are capabilities and knowledge that humans have, but technical phishing filters lack. Following the logic of the Swiss Cheese Model, relying on both humans and technical filtering in combination is better than just relying on one or the other. In recent years, organizations have been relying more heavily on automated phishing detection. Our findings suggest that reducing the diversity of filters may leave systems vulnerable to phishing, and that approaching end user training differently could strengthen strategies for preventing harm from phishing.

Much of the advice about phishing in the IT community involves preventing messages from ever getting to end users [14], rather than trying to educate end users. Because end users are able to filter messages in fundamentally different ways than technical filters, it would be more valuable to spend some money and resources improving the ability of end users to have a significant role in detecting phishing messages. Too much phishing training focuses on technical details (like url parsing [19, 30]) or behavioral changes (like not clicking [20, 35]), instead of trying to strengthen the capabilities that are unique to humans. In this paper, we have presented evidence of some of the knowledge and capabilities that humans have which can be leveraged to enhance phishing training and detection, e.g. forming expectations for emails and asking other people for information.

As the Swiss Cheese Model points out, in a series of filters, putting all of your resources into one layer of filters in exclusion to others removes the benefits you get from a defense in depth strategy. It is often better to have two imperfect filters that operate on different principles or information than it is to have one filter that is highly optimized but limited.

5.2 Similar to Expert Phishing Detection?

Our findings also have implications for identifying similarities between expert and non-expert user phishing email detection. Wash [34] conducted a detailed study of how people detect phishing emails. That study was conducted with IT experts – people with IT training and professional experience that allows them to successfully detect phishing emails. We extended that model, and based many of our questions on that extended model, partially to try to determine if features of that model are also present in how non-experts detect phishing.

In this paper, we are able to validate parts of his model with a non-expert population. Wash also pointed out that in addition to IT expertise, being a knowledge worker can provide expertise in managing email that is relevant to phishing detection. Our sample is not IT experts, and it is also not primarily knowledge workers who deal with email constantly.

In particular, we are able to validate that non-experts do have expectations about what should be present in emails and notice when those things are different. We are also able to validate that even in non-experts, people's attention is focused on what the email is requesting that they do; almost everyone in our study was able to identify what request the email was making. We validated that our non-experts self-reported that they frequently had gut feelings that something was off about the emails, helping them become suspicious. We were able to validate that people would frequently take explicit steps to investigate an email that they found to be suspicious. And we were able to validate that non-experts were able to conclusively decide whether an email was a phishing email or not. This lends support to the implication that expertise about one's own email inbox is an important and yet underutilized aspect of phishing detection training.

We were not able to validate all aspects of Wash's model with non-experts. In particular, Wash's model includes a chronological ordering of stages – first sensemaking, then suspicion, then acting. Our study is a survey and was unable to determine a chronological ordering that things happened in, and as such, we are not sure that things necessarily happen for non-experts in the order that Wash proposes.

5.3 Implications for Phishing Prevention

Email users engage in complex investigations of suspicious emails before they determine if the email is phishing, but current training and technologies do not support these investigations. Our findings suggest that phishing training could support user investigations better by encouraging users to delay taking actions until finalizing their investigation and encouraging email users to leverage peer capabilities (such as asking a friend for help). Additionally, companies that send email can provide helpdesk-style support to help users determine if the company actually sent the email to the user. Email clients could better support investigations by including a "help me troubleshoot this email" button, with contextualized suggestions for investigation.

6 Limitations

This paper is about people, their cognition, and how they successfully detect phishing. It is not about phishing emails. A survey is not a good method for collecting underlying ground truth data on the actual phishing emails or detection failures, because of selection bias and imperfect memory.

Recalling a phishing email prompted recollection of a specific instance, allowing the survey to investigate the processes that people use to detect phishing emails in their inbox. The answers we received were only about this one specific incident, and do not necessarily represent other incidents that the person was involved in; however, across respondents, these answers do represent a variety of the types of phishing incidents that non-experts encounter. Past research has focused almost exclusively on detection failures and fixing those failures; we instead look at what is working well in phishing detection and what should be supported.

Since this is a survey, we can only ask detailed questions about things we know about ahead-of-time. We based our survey questions on Wash's investigation of expert phishing detection [34]. We are not able to determine if the non-experts also use additional methods that were not present in Wash's experts. That is, we seek to learn which of these experts' methods are also used by non-experts, but we cannot learn anything about non-expert methods that are unique to non-experts. Therefore, we do not claim that these methods are a comprehensive description of how non-experts identify phishing; instead, we characterize some methods that they do use.

7 Conclusion

Phishing is a cybersecurity threat that many people experience; almost half of the people eligible for our survey could identify at least one specific phishing email that they received. These people have stories about phishing experiences that they can share with others, and we suspect these stories form an important part of how email users learn about phishing.

We found that many of the techniques that experts use to identify phishing [34], such as noticing minor discrepancies, forming expectations about what the email should look like and noticing differences from those expectations, and becoming suspicious and investigating the email more closely, are also present in how non-experts detect phishing emails.

We also found that much of the information that non-experts use when identifying phishing cannot be replicated by technical phishing detection systems. End users know the purpose (business, personal) of email accounts they receive emails at, and pay attention to that fact. They know whether an email is expected, and are able to compare it against other, similar emails they have received in the past (phishing emails often feel familiar). Additionally, these non-experts have investigative abilities, such as delaying responding to emails and asking the sender for confirmation or more information, that technical phishing filters don't possess. Targeting future phishing training at improving the use of this unique knowledge and expanding the use of these abilities is likely to yield improvement in phishing protection.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1714126. We would like to thank Faye Kollig and Abrielle Mason for assistance with coding the stories and copy editing. All members of the MSU BITLab provided valuable feedback on this study and paper.

References

[1] Jeremy Bryans and Budi Arief. Security implications of structure. In *Structure for Dependability: Computer-Based Systems from an Interdisciplinary Perspective*, pages 217–227. Springer, 2006.

[2] Deanna D Caputo, Shari Lawrence Pfleeger, Jesse D Freeman, and M Eric Johnson. Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 12(1):28–38, 2013.

[3] Debra L. Cook, Vijay K. Gurbani, and Michael Daniluk. Phishwish: a simple and stateless phishing filter. *Security and Communication Networks*, 2(1):29–43, 2009.

[4] Lorrie Faith Cranor. Can phishing be foiled? *Scientific American*, 299(6):104–111, 2008.

[5] Nicola Davinson and Elizabeth Sillence. It won't happen to me: Promoting secure behaviour among internet users. *Computers in Human Behavior*, 26(6):1739–1747, 2010.

[6] Julie S. Downs, Mandy Holbrook, and Lorrie Faith Cranor. Behavioral response to phishing risk. In *Proceedings of the Anti-Phishing Working Groups 2nd Annual ECrime Researchers Summit*, eCrime '07, pages 37–44, New York, NY, USA, 2007. Association for Computing Machinery.

[7] Julie S Downs, Mandy B Holbrook, and Lorrie Faith Cranor. Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security*, pages 79–90, 2006.

[8] Serge Egelman, Lorrie Faith Cranor, and Jason 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*, CHI '08, pages 1065–1074, New York, NY, USA, 2008. Association for Computing Machinery.

[9] Ian Fette, Norman Sadeh, and Anthony Tomasic. Learning to detect phishing emails. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 649–656, New York, NY, USA, 2007. Association for Computing Machinery.

[10] Joshua T Goodman, Paul S Rehfuss, Robert L Rounthwaite, Manav Mishra, Geoffrey J Hulten, Kenneth G Richards, Aaron H Averbuch, Anthony P Penta, and Roderict C Deyo. Phishing detection, prevention, and notification, October 16 2012. US Patent 8,291,065.

[11] The Radicati Group. Email statistics report 2019-2023 executive summary. Technical report, The Radicati Group, 2019.

[12] Ryan Heartfield and George Loukas. A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks. *ACM Computing Surveys (CSUR)*, 48(3):1–39, 2015.

[13] Thorsten Holz, Christian Gorecki, Konrad Rieck, and Felix C Freiling. Measuring and detecting fast-flux service networks. In *The Network and Distributed System Security Symposium (NDSS)*, 2008.

[14] Jason Hong. The state of phishing attacks. *Communications of the ACM*, 55(1):74, Jan 2012.

[15] Scott D Johnson, Jeffrey W Flesher, and Shih-Ping Chung. Understanding troubleshooting styles to improve training methods. In *American Vocational Association Convention*. ERIC, Dec 1995.

[16] Y. Joshi, S. Saklikar, D. Das, and S. Saha. Phishguard: A browser plug-in for protection from phishing. In *2008 2nd International Conference on Internet Multimedia Services Architecture and Applications*, pages 1–6, 2008.

[17] Mahmoud Khonji, Youssef Iraqi, and Andrew Jones. Phishing detection: a literature survey. *IEEE Communications Surveys & Tutorials*, 15(4):2091–2121, 2013.

[18] Gary Klein. *Sources of Power: How People Make Decisions*. MIT Press, 1998.

[19] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, 10(2):1–31, 2010.

[20] Robert LaRose, Nora J. Rifon, and Richard Enbody. Promoting personal responsibility for internet safety. *Communications of the ACM*, 51(3):71–76, March 2008.

[21] Eric Lipton, David E Sanger, and Scott Shane. The Perfect Weapon: How Russian Cyberpower Invaded the U.S. *The New York Times*, dec 2016.

[22] MacEwan University. University Discovers Online Fraud. Press Release, 2017. https://www.macewan.ca/wcm/MacEwanNews/PHISHING_ATTACK.

[23] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguyen. A novel approach for phishing detection using url-based heuristic. In *2014 International Conference on Computing, Management and Telecommunications (ComManTel)*, pages 298–303, 2014.

[24] US Bureau of Labor Statistics. Employment–population ratio, Retrieved Feb, 2021. https://www.bls.gov/charts/employment-situation/employment-population-ratio.htm.

[25] Rob Pegoraro. We keep falling for phishing emails, and google just revealed why. *Fast Company*, 2019. https://www.fastcompany.com/90387855/we-keep-falling-for-phishing-emails-and-google-just-revealed-why.

[26] Justin Petelka, Yixin Zou, and Florian Schaub. Put your warning where your link is: Improving and evaluating email phishing warnings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–15, New York, NY, USA, 2019. Association for Computing Machinery.

[27] Emilee Rader, Rick Wash, and Brandon Brooks. Stories as informal lessons about security. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)*, pages 1–17, 2012.

[28] James Reason. *Human Error*. Cambridge University Press, 1990.

[29] Ozgur Koray Sahingoz, Ebubekir Buber, Onder Demir, and Banu Diri. Machine learning based phishing detection from urls. *Expert Systems with Applications*, 117:345 – 357, 2019.

[30] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS)*, pages 88–99, 2007.

[31] Rebecca Smith. How a U.S. Utility Got Hacked. *Wall Street Journal*, Dec 2016.

[32] Symantec. Internet Security Threat Report. Technical Report February, 2019.

[33] Verizon. 2019 Data Breach Investigations Report. Technical report, 2019.

[34] Rick Wash. How experts detect phishing scam emails. *Proceedings of the ACM: Human Computer Interaction*, CSCW(160), October 2020.

[35] Rick Wash and Molly M Cooper. Who provides phishing training? facts, stories, and people like me. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[36] Steve Whittaker, Victoria Bellotti, and Jacek Gwizdka. Email in personal information management. *Communications of the ACM*, 49(1):68–73, January 2006.

[37] Weining Yang, Aiping Xiong, Jing Chen, Robert W. Proctor, and Ninghui Li. Use of phishing training to improve security warning compliance: Evidence from a field experiment. In *Proceedings of the Hot Topics in Science of Security: Symposium and Bootcamp*, HoTSoS, pages 52–61, New York, NY, USA, 2017. Association for Computing Machinery.

A Survey Instrument

A.1 Consent Form

Thank you for your interest in this research study. After reviewing the consent form below, please select the "I Agree" button if you would like to participate.

What is the purpose of this study? You are being asked to participate in a research study that is being conducted by Dr. Rick Wash and members of the Behavior, Information and Technology Lab (BITLab) at Michigan State University. The purpose of this study is to better understand how people think about and react to email messages they receive that seem suspicious or potentially harmful. You must be 18 years old to participate in this study.

What will I do if I choose to be in this study? Completing this survey should take approximately 20 minutes. The survey consists of multiple choice and fill in the blank questions. You will be asked questions about yourself, and about email messages that you have received. You will then be asked to remember specifics about a suspicious or potentially harmful email message you received in the past, and answer questions about that particular email message.

What are my rights as a participant in this study? You have the right to stop participating at any time. Your decision regarding participating will have no adverse consequences. You have the right to contact the researchers to ask questions about the purposes and procedures of this research after you have finished the survey. You may request that any information you give be ignored, or that any or all data from your survey be destroyed.

What are the risks and benefits of participating? Your participation in this study does not involve any physical or emotional risk to you beyond that of normal, everyday use of the Internet and email. You may not directly benefit from your participation in this study. However, your participation in this study may contribute to the understanding of how people think about suspicious email messages they receive. This will help researchers to develop tools and training that could prevent email messages from causing harm in the future.

How will I be compensated? If you successfully complete the entire survey, you will receive the incentive stated in your invitation in return for your participation.

What about the confidentiality and privacy of my information? Your survey responses will be assigned an anonymous code number, and researchers will save all survey responses by this code number. Any personally identifying information that you may provide in your answers to the survey questions will be removed by researchers before analyzing the data, so your answers cannot be linked with your name or identity in any way.

Survey responses and aggregate results of this research may be used for teaching, research, publications, or presentations at professional or scientific meetings. They may also be used for future research studies or shared with other researchers for secondary analysis or use in other research without additional informed consent from you. This means researchers may publish, present and share with other researchers summaries of data from multiple people, and direct quotations from individual responses.

No potentially sensitive, incriminating, or identifying information about you or others mentioned in the survey responses will be used in any publication or presentation, or shared outside the research team, except as required by Michigan State University's Human Research Protection Program or by law. Any use of your responses for public consumption will be carefully anonymized so it does not contain any identifying information.

Please note that the data will be retained at Michigan State University for a minimum of 5 years after all analyses and publications related to this project have been completed. Data will be stored on a secure, password-protected computer.

Whom should I contact if I have questions or concerns about this research study? If you have concerns or questions about this study you may contact Dr. Rick Wash, who is in charge of this research study, at telephone number 517-355-2381 or by email at wash@msu.edu.

If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Rd, Suite 136, Lansing, MI 48910.

Consent to participate Your participation in this study is completely voluntary. By clicking "I agree" below you are voluntarily agreeing to participate.

Q: Please select "I agree" below if you would like to participate.
○ I agree
○ I do not agree

A.2 Screening

Q: Have you ever received formal training in computer science, software engineering, IT, computer networks, or a related technical field?
○ Yes
○ No
○ I'm not sure

Q: Have you ever worked in a "high tech" job such as computer programming, IT, or computer networking?
○ Yes
○ No
○ I'm not sure

Q: What is your age in years?

Q: In this survey, we are interested in hearing about emails you received that were suspicious or potentially harmful in some way. This can be any email that you were suspicious about, including emails that you were concerned about but ended up not being a problem.

We are very interested in hearing about emails where it was hard for you to figure out what to do. For example, this could be an email message that you were unsure of and had to look closely at it to figure out if it could be harmful. Many of these emails ask you to do something, like click a link, open an attachment, or respond to the email with information.

Can you remember any suspicious or potentially harmful email messages that you've received? It is OK to go look through your email account and then continue with the survey, to help you recall if you've ever received email messages like this.
○ Yes, I have received email messages like this in the past.
○ No, I do not remember receiving any email messages like this.
○ I'm not sure

Q: What gender do you identify as?
○ Man
○ Woman
○ Other (fill in the blank)
○ Prefer not to answer

Q: Which categories below best describe you? Select all that apply:
☐ White
☐ Hispanic, Latino or Spanish
☐ Black or African American
☐ Asian
☐ American Indian or Alaska Native
☐ Middle Eastern or North African
☐ Native Hawaiian or Other Pacific Islander
☐ Some Other Race, Ethnicity or Origin (please specify)

A.3 Elicitation

Q: First, to help you to remember emails that were suspicious or potentially harmful, please list some different ways that an email message can be unsafe or cause security problems:

Q: Next, think about different ways you know of to recognize an email that is suspicious or potentially harmful, and make a list of these below:

Q: Take a moment to think about times in the past when you personally received a suspicious or potentially harmful email. Please list as many of these emails as you can remember, using only a couple of words to describe each one. You may want to re-read your answers to the previous questions to jog your memory.

Q: On the previous page, you made a list of emails that you personally received that were suspicious or potentially harmful. For reference, here is the list:

Q: Choose one email message from the list above that it's easy for you to recall details about. You will be answering questions about this email in the rest of the survey. Briefly summarize that email, and what happened when you received it.

Q: In approximately 5 words, please describe what made this email seem suspicious:

Q: In approximately 5 words, please describe why it was hard for you to figure out how to deal with this email:

Q: In approximately 5 words, please describe what the email was asking you to do:

A.4 Noticing

For your reference, here is what you said about the email you will be answering questions about on this page:

Q: How long ago did you receive the email?
○ Within the last day
○ Within the last week
○ Within the last month
○ Within the last year
○ Longer than one year ago
○ I don't remember

Q: To help us monitor the quality of our data, please select "Somewhat disagree" from the choices below.
○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Q: At which of your email accounts did you receive the email?
○ Work Email account
○ Student Email account
○ Personal Email account
○ Other (please describe)
○ I don't remember

Q: What was the context of the email?
○ This email was related to work
○ This email was of a personal nature
○ Other (please describe, briefly)
○ I don't remember

Q: Who did the email appear to come from?
○ A work colleague
○ A close friend or family member
○ An acquaintance from outside work
○ A company, business or other organization

○ Other (please describe)
○ I don't remember

A.5 Expecting

For your reference, here is what you said about the email you will be answering questions about on this page:

Q: Please indicate your agreement or disagreement with the following statement:

When I read the email message, I felt like I had received other email messages like this one before.
○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

Q: Before receiving this email, had you ever received an email message from the sender?
○ Yes
○ I'm not sure
○ No
○ I don't remember

Q: Before receiving this email, had you ever interacted with the sender in some other way than email? (For example, if you had previously talked to the sender face-to-face, or visited their website.)
○ Yes
○ I'm not sure
○ No
○ I don't remember

Q: Before receiving this email, how long had you known the sender?
○ One month or less
○ Between one month and one year
○ One to two years
○ Two to five years
○ Five to ten years
○ More than 10 years
○ I don't remember
○ I did not know the sender

Q: Did you expect to receive this specific email?
○ Yes
○ I'm not sure
○ No
○ I don't remember

Q: Please indicate your agreement or disagreement with the following statement:

This email message seemed different from the email messages I typically receive.
○ Strongly agree
○ Somewhat agree
○ Neither agree nor disagree
○ Somewhat disagree
○ Strongly disagree

A.6 Suspecting

For your reference, here is what you said about the email you will be answering questions about on this page:

Q: Many suspicious emails ask you to do something. Was the email asking you to do any of the following? Please check all that apply.
☐ Click on a link or button
☐ Open something that was attached to the email
☐ Respond to the email with some information
☐ Take some action outside of the email
☐ None of the above
☐ I don't remember

Q: Think about the subject line of the email. Did the subject line feel normal, or did it feel "off" in some way?
○ I didn't notice anything that felt off about the subject line
○ The subject line was different than I would expect
○ I don't remember much about the subject line of the email

Q: Think about who the email said it was from. Did this sender information make sense, or did it feel "off" in some way?
○ I didn't notice anything that felt off about the sender
○ The sender's name looked different than I would expect
○ The sender's email address looked different than I would expect
○ I don't remember who the email said it was from

Q: Think about the main body of the email. Did the main body of the email seem normal, or did you notice anything that felt "off" about it? Please check all that apply.
☐ I didn't notice anything that felt off about the main body of the email
☐ The main body of the email included typos or other issues that I didn't expect to be in an email like this
☐ The main body of the email was missing something that I would expect to be in an email like this
☐ The main body of the email included something strange that I do not normally see in an email like this
☐ The main body of the email included more information than I expect to be in an email like this
☐ The main body of the email included less information than I expect to be in an email like this
☐ I don't remember much about the main body of the email

Q: When you read the email, did you believe that the email was harmful?
○ Yes, I thought it was harmful
○ I was not sure about whether it was harmful or not
○ No, I did not think it was harmful

○ I don't remember

Q: How sure or unsure are you about your answer to the previous question?

Please indicate your answer below on a scale from 0-100, where 0 means COMPLETELY UNSURE and 100 means COMPLETELY SURE.

A.7 Investigating, Deciding, and Acting

For your reference, here is what you said about the email you will be answering questions about on this page:

Q: What actions did you take to learn more about the email? Please check all that apply.
☐ Hovered over one or more of the links in the email to see where it went
☐ Clicked on one or more of the links to see where it went
☐ Looked more closely at the the email address the email came from
☐ Opened the attachment
☐ Looked at email headers
☐ Asked someone else about the email
☐ None of the above
☐ I don't remember
☐ Other

Q: In what ways did you attempt to learn about the sender of the email? Please check all that apply.
☐ I went to the website of the sender
☐ I contacted the sender via phone
☐ I contacted the sender through another communications medium (texting, chat, social media)
☐ I talked to the sender face-to-face about the email
☐ I did not try to contact the sender
☐ I don't remember

Q: After you learned more about the email, did you decide that the email was safe or not?
○ Yes, the email was safe
○ I was still not sure whether the email was safe or not
○ No, the email was definitely not safe
○ I don't remember

Q: How sure or unsure are you about your answer to the previous question?

Please indicate your answer below on a scale from 0-100, where 0 means COMPLETELY UNSURE and 100 means COMPLETELY SURE.

Q: What action(s) did you take with this email? Please check all that apply.
☐ Deleted the email
☐ Clicked a button to report the email as spam
☐ Sent the email to someone
☐ Responded to the email
☐ Left the email in my inbox
☐ None of the above
☐ I don't remember

Q: At any point while handling this email, to what extent did you experience these emotions?
Scale:
○ Not at all
○ Somewhat
○ Moderately
○ Quite a bit
○ An extreme amount
Emotions to be rated on that scale:
○ Dread
○ Terror
○ Anxiety
○ Nervous
○ Scared
○ Panic
○ Fear
○ Worry

Q: Please indaicate your agreement or disagreement with the following statement:

I feel like something harmful happened because of this email message.
○ Strongly disagree
○ Somewhat disagree
○ Neither agree nor disagree
○ Somewhat agree
○ Strongly agree

A.8 Full Story

You are almost done! You have now answered a number of questions about an email message that seemed suspicious or potentially harmful, and hopefully you have recalled quite a few important details. For reference, here is the short description of the email you provided at the beginning of the survey:

Q: Below, please write the story of the email as if you were telling it to a friend. Use as much detail as you can, including any thoughts or recollections about what happened you might have had as you were filling out the survey. Your story should be at least 4 or 5 sentences long (minimum 300 characters).

Q: How easy or difficult was it for you to remember a suspicious or potentially harmful email to answer questions about in this survey?
○ Extremely easy
○ Somewhat easy
○ Neither easy nor difficult
○ Somewhat difficult
○ Extremely difficult

A.9 Demographics

Q: What is the last grade or class you completed in school?
○ None, or grades 1-8
○ Some high school
○ High school graduate or GED certificate
○ Technical, trade, or vocational school AFTER high school
○ Some college, no 4-year degree
○ 4-year college degree
○ Some postgraduate or professional schooling, no postgraduate degree
○ Postgraduate or professional degree, including master's, doctorate, medical or law degree

Q: What is your current employment status?
○ Employed full time
○ Employed part time
○ Unemployed looking for work
○ Unemployed not looking for work
○ Retired
○ Student
○ Student and employed part time
○ Disabled

Q: What was your total household income before taxes during the past 12 months?
○ Less than $25,000
○ $25,000 to $34,999
○ $35,000 to $49,999
○ $50,000 to $74,999
○ $75,000 to $99,999
○ $100,000 to $149,999
○ $150,000 to $199,999
○ $200,000 or more

Q: How familiar are you with the following Internet-related terms?

Please rate your understanding of each term below from None (no understanding) to Full (full understanding):

Scale:
○ None
○ Little
○ Some
○ Good
○ Full

Terms to be rated:
○ Wiki
○ Meme
○ Phishing
○ Bookmark
○ Cache
○ SSL
○ AJAX
○ RSS
○ Filtibly

A.10 Thank You

Thank you for participating! If you have any questions or concerns, please contact Dr. Rick Wash at telephone number 517-355-2381 or by email at wash@msu.edu.