How Experts Detect Phishing Scam Emails

RICK WASH, Michigan State University, USA

Phishing scam emails are emails that pretend to be something they are not in order to get the recipient of the email to undertake some action they normally would not. While technical protections against phishing reduce the number of phishing emails received, they are not perfect and phishing remains one of the largest sources of security risk in technology and communication systems. To better understand the cognitive process that end users can use to identify phishing messages, I interviewed 21 IT experts about instances where they successfully identified emails as phishing in their own inboxes. IT experts naturally follow a three-stage process for identifying phishing emails. In the first stage, the email recipient tries to make sense of the email, and understand how it relates to other things in their life. As they do this, they notice discrepancies: little things that are "off" about the email. As the recipient notices more discrepancies, they feel a need for an alternative explanation for the email. At some point, some feature of the email — usually, the presence of a link requesting an action — triggers them to recognize that phishing is a possible alternative explanation. At this point, they become suspicious (stage two) and investigate the email by looking for technical details that can conclusively identify the email as phishing. Once they find such information, then they move to stage three and deal with the email by deleting it or reporting it. I discuss ways this process can fail, and implications for improving training of end users about phishing.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; *Computer supported cooperative work*; Empirical studies in HCI; HCI theory, concepts and models; User studies; • **Security and privacy → Social aspects of security and privacy**; *Usability in security and privacy*; **Phishing**; • **Social and professional topics → Phishing**; • **Applied computing** → *Enterprise computing*.

Additional Key Words and Phrases: phishing, security, email

ACM Reference Format:
Rick Wash. 2020. How Experts Detect Phishing Scam Emails. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW, Article 160 (October 2020), 28 pages. https://doi.org/10.1145/3415231

1 INTRODUCTION

Email is one of the most common methods of telecommuncation. Over 3.9 billion people have email accounts, and collectively they send and receive over 290 billion emails per day [19], including both business emails and consumer emails. Email allows people to easily communicate with others anywhere on the planet, which supports both business collaboration and personal communications. Email represents one of the largest CSCW systems on the planet, and fundamental to its design is the idea that anyone, anywhere can send an email to anyone else [53].

Unfortunately, some emails are phishing scams. A phishing email is an email that pretends to be something that it is not, in order to get its recipient to take an action that they otherwise wouldn't do. Phishing emails are currently causing large scale damage in society. Many current cybersecurity attacks have been tracked back to phishing emails. Phishing attacks have caused people to transfer

Author's address: Rick Wash, Michigan State University, Media and Information, East Lansing, MI, USA, wash@msu.edu.

2573-0142/2020/10-ART160
https://doi.org/10.1145/3415231

large amounts of money [33], to install ransomware that disables computers [47], or to have their emails stolen and posted publicly on Wikileaks [32].

Phishing is a semantic attack [43] that exploits the way that humans assign meaning to content that they read or hear. That is, semantic attacks like phishing take advantage of the fact that people generally believe what they see and hear, by fraudulently pretending to be something that they are not in order to get someone to undertake actions that they normally wouldn't be willing to take.

Because semantic attacks exploit the way that humans assign meaning, it is difficult to develop purely technical approaches to preventing them. Asking humans to be able to identify and recognize phishing messages is an important component of how we defend against phishing.

Much current research has focused on ways that humans are limited in their ability to detect phishing [12]; researchers have focused on trying to find ways to better train people [51]. Much research has gone into developing training materials such as comics [30] or games [3, 45] that try to help people recognize phishing messages.

Most of this training material focuses on one specific aspect of recognizing phishing, such as training people what fraudulent URLs look like [30]. However, we don't really understand what the full recognition process looks like. What cognitive process allow a human to detect that an email is phishing? What is a complete set of steps that people can realistically follow to recognize that an email is phishing? When people do successfully identify an email as phishing, how do they do it?

To examine this, I conducted a series of 21 interviews with professional IT experts. These experts regularly encounter phishing messages, both sent to them and sent to others that they support, as part of their job, have developed skills at recognizing phishing messages, and are likely to have successfully detected phishing emails in the past. In these interviews, I used the Critical Decision Method [11, 27] to get these experts to describe in great detail specifically how they identified and dealt with a single phishing message in a natural, real setting. By analyzing these descriptions, I am able to understand and describe the complete process by which experts identify phishing emails.

IT experts are a particularly important population for phishing. Many people rely on experts to help them detect phishing emails (for example, John Podesta asked his IT expert before clicking on the phishing email that lead to Clinton's emails on Wikileaks [32]). Also, many phishing campaigns target IT administrators because their accounts have access to many important functions [50]. Additionally, experts possess the skills that we hope to train in end users; examining experts allows us to set realistic goals for end user phishing training.

I found that experts go through a three-stage process to identify phishing messages. First, for all emails, the experts engage in a sensemaking process where they try to understand the email. As part of this process, they use their expertise in the topic of the email to notice discrepancies, or things about the email that aren't quite right. These discrepancies then trigger a cognitive shift, where the experts become suspicious that the email might not be legitimate. During this second stage, suspicion causes the expert to investigate the email by explicitly seeking information (for example, by hovering over links to see the URL) that can conclusively identify the email as phishing. Once the person makes a decision about whether the email is legitimate or phishing, then they enter the third stage, where they deal with the email, usually by either deleting it or reporting it.

Now that we understand this process and can describe it in detail, we can better understand ways that this process can fail and that people can fall victim to phishing. Current phishing training, like being suspicious or identifying fraudulent URLs, mostly helps during the second stage (suspicion) and focuses on information that conclusively distinguishes phishing from legitimate email. Experts found this stage to be relatively easy; instead, the most difficult and important part of the process was during the first stage, where people were noticing discrepancies as they tried to make sense of emails and then tried to find alternative explanations (such as phishing) that could explain those discrepancies.

2 PHISHING

Phishing attacks have been on the rise over the last few years. 32% of all breaches in 2018 were due to phishing, which was the second most common cyber threat action after denial of service [50]. The overall phishing rate has declined each year over the last four years. However, it is still a relatively high rate, with 1 in every 3200 emails received being a phishing message [48]. Recently, attackers have moved away from using malicious URLs and toward attachments in malicious emails [48].

Phishing attacks are particularly a problem when attackers are targeting specific individuals or organizations – which is often called "spear-phishing". People are falling for spear phishing at high rates [31], and attackers are getting better at using persuasive techniques to target individuals [35]. Spear-phishing remains the most popular targeted attack method, used by 65% of all groups doing targeted attacks [48]. (Compare this to only 23% of groups using zero-day vulnerabilities.)

Nation-states particularly rely on phishing attacks, which have been on the rise [50]. Inside large organizations, 36% of external information breaches were state-affiliated, and most of these cyber-espionage attacks begin with phishing [50]. In the public sector, 79% of all external attacks are cyber-expionage. Spear-phishing attacks were particularly being used to target elections in the US in 2018 [32, 48] and in other democracies around the world.

2.1 Why People Fall for Phish

There is little work on when and why experts fall for phishing emails. However, the existing research literature on phishing has a healthy number of papers trying to understand patterns in which phishing emails end users are more likely to believe as real. The literature calls this "falling for phish" [6].

Phishing emails are crafted by adversaries intentionally trying to deceive users. Dhamija et al. [13] found that intentional visual deception — making a phishing email look similar to a legitimate email, such as using similar domain names and visual design – was one of the biggest predictors of which phishing emails would work. Blythe et al. [6] found that emails with logos (which are easy to copy) are significantly harder to detect than emails without logos. Blythe et al. [6] also found that blind users were much better at detecting phishing emails, which also indicates that visual design is important in the deceptive aspects of phishing.

Oliveira et al. [35] looked at how marketing theories such as Cialdini's seven principles of influence [8] can be used to craft deceptive emails. They found that all six principles could increase phishing click rates, with 43% of users falling for the deception, but that scarcity and authority had disproportionate effects on older users while reciprocation and liking had disproportionate effects on younger users. Benenson et al. [5] found that emails that fit people's pre-existing expectations or that made people curious were more likely to be clicked.

A common finding in the literature is that women are more susceptible to phishing emails than men [31, 35, 44] and that older users are more susceptible than younger users [31, 35]. Sheng et al. [44] used a mediation analysis to confirm the hypothesis that this this is at least partially due to a difference in technical skill and expertise (even in non-experts), though technical skills only explain part of the difference.

Almost all of this work looks at properties of the phishing email or properties of the user to explain phishing susceptibility. This does not explain, though, why some people may fall for a specific email at some times but not others [7]. In this paper, I examine the *process* that expert users follow to identify a phishing message; by understanding this process I am able to hypothesize additional factors that may lead to phishing susceptibility, and particularly how personal background and situation can play an important role.

2.2 Technical Phishing Prevention

A number of technical measures have been created to help defend against phishing attacks. Technical solutions frequently try to "make [phishing] invisible" [23] to end users by using 1) blacklists to block known URLs, and 2) taking down known phishing landing pages.

In using blacklists, a number of solutions have been proposed that use machine learning techniques to detect aspects of phishing attacks [1]. They differ in what is being detected and what machine learning techniques are being used. Some methods focus on detecting malicious URLs, for example using GANs [2] or RNNs [4]. Other methods examine websites to detect when a website might be malicious [54] (ensemble classifier). Both of these methods fail to work on non-web based attacks like malicious attachments (which, as noted above, are on the rise [48]). Other techniques try to classify individual email messages, solving the training data problem by learning email over time [46]. Yet another approach involves basing machine learning on human reports of phishing messages [20]. Other technical solutions include designing better interfaces [23] that provide warnings to users like the SSL warnings in Chrome [15]. For example, one proposal provides warnings about phishing messages in an email client near the URL, but still allow users to click the URL if needed [38].

Technical phishing tools are able to detect many phishing URLs. However, these tools work best for mass phishing emails where many identical or similar emails are sent to large numbers of potential victims. These report-and-blacklist solutions do not help the first few people who receive the message. More importantly, they also do not help protect against spear-phishing attacks that are customized to their targets and only sent to a small number of targets [14]. This may be why attackers seem to be moving toward more spear-phishing attacks and fewer mass attacks [48].

2.3 Human-Centered Phishing Prevention

Purely technical solutions are not enough to prevent phishing. Defending against phishing attacks is difficult because, fundamentally, phishing attacks are socio-technical attacks [43]. They exploit modern communications infrastructure to send forged and fake messages. However, determining whether a given email is a phishing email requires knowing important social information that is only available to the recipient, such as what communications they are expecting and from whom [43].

One of the most common anti-phishing techniques used today is a social reporting mechanism. Emails that are suspected to be phishing emails are reported to a central location, which then investigates and verifies that the email is not what it purports to be. Once determined, that email (and others similar to it) can then be blacklisted and/or removed from inboxes of other recipients [20]. Most large organizations have IT security teams who do this work. Many tools exist to aggregate these reports across organizations (e.g Google's Gmail, or Proofpoint's phishing services [16]).

Many organizations train potential human recipients to recognize and identify phishing messages as a regular part of IT "security awareness" training efforts [17], and much research has gone into trying to identify ways to effectively train people to recognize phishing messages. Broadly speaking, there are three styles of training currently employed [17]: general-purpose training messages that communicate "best practices"; fake phishing campaigns [51]; and in-the-moment warning messages [38].

For all three of these methods of communication, there is an important question about *what* we should be teaching end users. Current training focuses on two important messages. First, the training communicates awareness of phishing threats: it tries to make users aware that these threats exist and that users have a role to play in detecting and addressing the threats.

Second, the training usually communicates a set of “best practices”: actions and behaviors that are recommended to users. These best practices are usually focused on helping users accurately determine if an email actually does what it says it does; for example, if the link in the email actually goes to the organization or website that it claims to, or if the email headers indicate the email is actually from who is says it is from. Blythe et al. [6] summarizes many of these best practices as “Never click on a link in an email” and “Never respond to an email asking for banking details”. Both Sheng et al. [45] and Kumaraguru et al. [30] emphasize understanding URLs and checking that URLs match the purported email sender; their training has become the basis for much real-world training.

Sasse has recently argued that this training, and particularly the attempts to make warnings more salient, are misguided and are causing more harm than good [42]. She points out that user time is valuable, and that distracting the user and making them take up their time dealing with increasingly intrusive warnings and difficult tasks (like hovering over every email link received) causes many problems for users, including lost time, distraction, and increased fearfulness. She recommends, instead, that security researchers should be focusing on reducing the number of decisions that users need to make so user time is used more efficiently.

This argument is particularly important for phishing. Much of the current phishing training encourages users to undertake actions that are very time-consuming when you multiply them by the number of emails a person receives in a day.

2.4 The Phishing Decision

We do not currently have good theoretical models of how people make the decision about whether an email is phishing or not. The closest model we have is Cranor’s Human-in-the-Loop model [12]. This model is based on Conzola and Wogalter’s Communication-Human Information Processing model [9] of warning effectiveness. These models assume that there is some initial communication (a warning message, or a training message) intended to modify behavior, and then models impediments to humans heeding that communication (including sender credibility, beliefs and attitudes, and prior knowledge and experience). The goal of these models is to improve the initial communication’s effectiveness at behavior change. Phishing detection, however, is substantially different because the initial communication – the phishing message – is not under our control. Indeed, the phishing message is created by adversaries with the specific intention of evading detection. Additionally, neither Cranor’s model nor Conzola and Wogalter’s model is a process model that describes the cognitive processes involved in detection.

To study the process that people follow to make decisions about phishing, it is important to clearly and accurately identify what that decision is. Traditionally, decisions are often characterized as *comparative*: there are two or more options, and a decision-maker needs to choose which of the options to go with [24]. For phishing, for example, a decision-maker could need to decide whether the email in front of them is a real email or a phishing email.

To study this, I looked to research on naturalistic decision making in real-world contexts, which suggests that this definition of decision is too narrow. In many situations, people who are experts at making a given decision often do not see the decision in front of them. Instead, they “just know what to do” and then do it [24]. Klein, Calderwood, and Clinton-Cirocco, in their work studying firefighters, found that most of the fire fighting decisions made by expert fire commanders were not made by comparing multiple options, but instead through a process of recognizing the situation as similar to some past situations, remembering what worked in those past situations, and then going forward with a similar action in this situation [26]. Klein instead then defines a decision as “any point where more than one reasonable option exists, even if those options are not considered by the decision-maker” [24, 41].

Following this definition, I define the phishing decision that people have to make as *for each email received, the recipient must decide whether the email is legitimate and statements it makes can be taken at face value, or whether the email is phishing and cannot be trusted.* Notice that this decision applies to each and every email that a person receives in their inbox, whether they explicitly consider the decision or not. It is possible that email recipients "just know what to do", much like the fireground commanders in Klein et al. [26], and do not have to explicitly consider each decision.

Also, notice that "just knowing what to do" is actually an ideal case. Email recipients should not have to spend time and effort taking complicated investigative steps, such as hovering over links or viewing email headers, for every email they receive. Instead, it would be ideal if they "just know what to do" and distinguish legitimate emails from phishing without being distracted from the main task at hand. Following the critique of Sasse [42], minimizing the work required from users to make accurate decisions about emails is better than adding work to users or inciting fear on a regular basis. Simply recognizing emails as legitimate or phishing and just knowing what to do is actually a sign of expertise in the user, and an ideal to be strived for.

To better understand what this ideal looks like, I chose to study *experts*: people with the knowledge, skills, and experience to accurately identify phishing emails, to better understand how they go about identifying phishing emails. I suspected that I would find that they often "just know what to do". Experts are people too, and non-experts should be able to develop some of the skills that experts have in detecting phishing, much like most people are able to learn to drive automobiles without needing to be expert auto mechanics.

2.5 Experts and Expertise

Hoffman [21] describes a number of ways to define expertise, and makes a distinction between "cognitive/perceptual" approaches to defining expertise and "sociological" approaches. Cognitive approaches to expertise look at an individual person and the skills and knowledge that person has. A person can be defined as an expert in a given domain if they have the skills and knowledge to excel in the domain. Sociological approaches, on the other hand, define experts as part of a larger social order, often relative to other people. Experts can be defined sociologically by examining job titles, certifications, awards, or other social recognition.

In this paper I focus on cognitive/perceptual definitions of expertise; I am examining the cognitive processes that experts use to detect phishing emails, so I focus on expertise defined by those cognitive processes rather than social indicators of expertise like job titles.

There are multiple ways to cognitively define expertise; Hoffman [21] describes three broad categories for defining expertise cognitively. First, experts are often described developmentally, as they develop through a series of levels or stages, often given names based on medieval guilds (novice, apprentice, journeyman, expert, master). This conceptualization of expertise is based on improving over time through the accumulation of knowledge, experience, and skill. Second, experts can be described in terms of knowledge structure. Experts possess more knowledge about the domain including being able to recite more facts; organize knowledge about the domain at a higher, more abstract level; and have the ability to model the world conceptually and use mental models to envision things that have not yet happened. Third, expertise can be defined in terms of process; experts reason about the world differently than non-experts.

Since this paper is primarily interested in cognitive processes, I focus on this third cognitive/perceptual definition of experts: a person is an expert in a domain if they reason about it like an expert. Klein and Hoffman [28] describe three major ways that expert reasoning is different than non-experts. First, experts are able to see typicality. Experts know what is typical in a situation and what is not typical, and use that knowledge to classify situations and to focus on what is important in a situation. Second, experts are able to see fine distinctions that non-experts miss. One example of this is in TV broadcasts of the Olympics, where broadcasters (who are experts) can see small details about the athletes in real-time that only become evident to the viewers upon slow-motion replay. Third, experts can see antecedents and consequences; that is, they use their knowledge of cause-and-effect to be able to see what happened in the past and what is likely to happen in the future.

3 METHODS

The goal of this study is to examine how people who are experts at identifying phishing messages do this for emails in their own inbox. I identified 21 full-time IT professionals from a large organization, and interviewed each one separately about a specific email that they had personally received, which they suspected to be phishing. This interview structure was intended to investigate how these experts identify phishing emails in a naturalistic setting. Interviewing people who have expertise at identifying phishing messages will help to know what skills are needed to train non-experts to protect themselves. What is a reasonable level of skill that a human being can possess at identifying phishing emails? And how do people with that skill actually go about identifying phishing messages?

3.1 Participants

For this study, I wanted to identify and interview experts: people who have some high level of expertise at identifying phishing emails. Applying the cognitive/perceptual definition of expert from Klein and Hoffman [28], phishing experts should be people who understand what typical phishing emails look like (and, presumably, what non-phishing emails look like). They should be able to make fine perceptual distinctions about features of the emails, and be able to notice things that others might miss. And they should have a strong understanding of cause-and-effect; what might have caused this email to be sent, or what would happen if links were clicked or emails deleted?

To identify people who were likely to possess these skills, I sought out a professional IT organization with a security team which has to deal with many phishing messages every day. IT professionals have the technical skills and knowledge to understand cause-and-effect and to understand small technical details of emails that everyday people do not. A professional security team is also able to develop a high level of skill at detecting phishing messages through direct experience. Other members of the team and people who work adjacent to the security team also likely develop some of the expertise by vicariously hearing about phishing messages and having to deal with issues that arise out of phishing attacks.

I recruited 21 members of the professional IT staff at a large midwestern university. 11 of the 21 were members of the security team who are responsible for identifying phishing messages and dealing with them for the university. Participants were recruited via snowball sampling with the assistance of one of the members of the IT security team and with approval of the management. All of the participants reported having technical or IT training and years of experience working with technology and email. They all reported hearing about phishing emails regularly as part of their job and knowing members of the IT security staff and working closely with them. Table 1 summarizes these participants.

During the interview process, I verified that all 21 participants possessed the three cognitive/perceptual skills in my definition of phishing expertise: understand what typical phishing emails look like; making fine perceptual distinctions about details of emails; and understood cause-and-effect with phishing emails. By this definition, all 21 participants possessed expertise in phishing security.

Professional experts are a difficult population to sample; they have limited time. Two hours of time from an expert is a large amount to request. 21 participants is a reasonable number for a study

Gender			*Role*			*Email Received*		
Man	16	76%	IT Security	11	52%	Legitimate	4	20%
Woman	5	24%	General IT	8	38%	Phishing	15	75%
			IT Adjacent	2	10%	Training	1	5%

Table 1. Information about the participants and the emails they described. The last column indicates whether the participant ended up concluding that the email they received was legitimate, was a phishing message, or was a fake phishing message sent as training by their organization. "IT Adjacent" is a self-description used by participants without strong technical backgrounds who work in non-technical roles inside the IT organization.

like this because of the specialized knowledge and expertise of the participants, and was enough to achieve theoretical saturation [39]. The interview respected their time (I split the interview into two 1-hour sessions if needed) and resulted in a large amount of detailed data from each participant. While the number of participants is not large, the level of detail and the difficulty of access means that I was able to generate detailed findings from this group of participants. As with most qualitative research, findings generalize not because of statistical sampling or large numbers, but by clearly describing the participants (so the reader can "transfer" results to new situations) and grounding theory development in the details of the participants [39].

While all participants work in the same organization (and thus might be subject to the same organizational messaging about phishing), most of them have years-to-decades of experience in IT and security, often in more than one organization. It is unlikely that the organizational messaging overly influences these results; indeed, since the organization's phishing training is written by some of these participants, it is more likely that these experiences influence the training more than the training influences the experiences.

3.2 Interviews

The interviews followed the Critical Decision Method [11]. The interviewer first worked with the participant to identify a critical incident: an instance of an event in the past when the person's skill and expertise were challenged. For these interviews, I asked about incidents where the participant had received a "potentially dangerous email", and preferred discussing incidents that "particularly challenged" the participant's knowledge, skills, and expertise to identify whether it was dangerous or not. I specifically sought to identify challenging emails because identifying these emails goes beyond routine or procedural knowledge, and better allows us to see the components of expert decision-making [11]. All participants were able to identify multiple potential incidents (emails that they had received), including a number of messages that were initially thought to be potentially dangerous but ended up being legitimate emails.

Once a specific incident was identified, then the interview proceeded through three sweeps through the incident. In the first sweep, the participant and the interviewer walked through the incident together, and the interviewer drew a visual timeline of the incident on a shared whiteboard in the interview room. The timeline enabled the interviewer and participant to develop a shared understanding of the order that events unfolded, and what happened during the incident. It provided a shared display that helped ensure that the interviewer accurately understood the ordering of events and helped the participant be more complete in their recall of the incident. The interviewer typically spent about 15–60 minutes of the interview developing this shared timeline.

In the second sweep, the interviewer then asked numerous questions about each time point on the timeline with the goal of "deepening" the understanding of what happened. In particular, the interviewer asked at each point in time what cues (properties in the world) the participant was noticing at that point in time, what goal the participant was trying to accomplish at that point in time, properties of decision-making like time pressure and timing, options that the participant considered for what to do next, experience and background knowledge that the participant brought to bear to help with the decision-making, and whether at this point what they were seeing was typical of these kinds of incidents. This deepening sweep was usually the longest portion of the interview, and the interviewer usually spent about 30-60 minutes of the interview asking these questions about each event and decision in the timeline.

In the third sweep, the interviewer went through the incident and asked "what-if" questions. These questions probed for what would happen if things were different – if cues were or were not present, if they were earlier in their career and didn't have the same level of expertise, etc. These hypothetical what-if questions focused on possible alternative courses of action, and surfaced aspects of expertise that the participant possesses but the interviewer might not know to ask about.

In total, each interview lasted about 2 hours, and was capped at no longer than 2 hours. All interviews reached the third sweep, though not all interviews completed the third sweep within the 2 hour time limit. Participants were compensated with $40 USD for their time. After the interview was complete, the interviewer took photographs of the timeline on the whiteboard, had the interviews transcribed for analysis, and wrote up a short summary story of the incident. This protocol was approved by my university's IRB as expedited.

3.3 Initial Analysis

The data was analyzed in two phases. In the first phase, I combined analysis methods from the Critical Decision Method [11] with open coding and data matrices [34]. Analysis was initially driven by looking for elements of the Recognition Primed Decision Model[24, 26]. I began by building a structured description of the decision-making process for each participant separately [22]. This involved using the transcripts of the interview along with the timeline to identify major decision points that the participant faced. For each decision, I identified components of decision-making that the participant mentioned: cues that the participant noticed; background knowledge that they brought to bear on the situation; expectations about what would happen; goals that the participant was trying to achieve at that time; and assessments of the situation as a whole [22].

After identifying these, I placed these structures in chronological order based on the timeline diagrams, and compared across participants using data matrices [34] to identify patterns in the decision process. From this, I identified a three-stage process for phishing detection that was followed by 19 of the 21 participants. This process is described in Section 4, below. I also conducted open coding of the transcripts to serve as a validity check; this coding helped to verify that the descriptions of the decision process matched what the participants said during the interview, to check that quotes and descriptions represented multiple participants accurately, to check the meaning of outliers, and to search for negative cases and counterexamples that did not match my description [36]. I also conducted member checks, showing early draft of this paper and its analysis to participants to ensure that my interpretations are accurate.

In doing this analysis, I discovered that a specific part of this process – the transition from the first stage to the second stage – was a critical process for identification of phishing emails, and that this transition was poorly described in existing literature. To address this, I conducted a second analysis, described in Section 5, where I focused on the cues that participants noticed to better understand this transition.

	Stage 1: *Face Value*	*Stage 2:* *Suspicion*	*Stage 3:* *Decision*
Cognitive Action	Sensemaking: What is this email?	Hold 2 alternative frames in head simultaneously	Deal with it
Stated Goal	How does this email relate to other things in my life?	Is this a phishing message?	Take action to deal with the email
What is noticed	Information present in email, taken at face value	Explicitly seeking information to help make determination	Nothing
Expertise Used	Surface topic of email	Phishing	None?
Key information	Discrepancies with the emerging understanding	Technical details of email	N/A
Major decision	What other explanation might make sense?	Which frame makes more sense?	What should I do about it?

Table 2. The three stages of detecting a phishing email

4 THREE STAGES OF PHISHING DETECTION

Nineteen of the 21 participants described a process that proceeded through three distinct stages. Table 2 provides an overview of these stages. First, the participant would try to make sense of the email and understand what it was saying. Second, the participant would become suspicious that the email wasn't legitimate, and investigate further. And third, after investigation, the participant would make a final decision about the email. As an example, consider this email that Josh received[1]:

Case 1. Bank Hacked: *Josh saw an email from his primary bank with the subject line indicating he needed to verify his account information. Josh was alarmed; he was uncertain if his account had been compromised. He was also surprised that his bank would send an email stating that he needed to verify his account information rather than simply stating he had a notification on his account.*

Confused as to what was happening, Josh opened the email to read it more closely. The format of the email was similar to newsletters the bank periodically sends out. The email said he was being notified that his bank had been hacked and he needed to login to his account to verify his information. He also noticed a link at the bottom of the message directing him to his login page. Being an experienced web developer, he knew that if he hovered over the link he could see the url. Once he did so he saw the url was similar to what he remembers the bank url being, but he always forgets the details, in particular if the bank if a .com or .net.

[1] A *case* is a summary of an incident described by one of the participants. The interviewer wrote a case description immediately after completing each interview, which was then checked against the transcripts and timeline for accuracy. Cases are presented here as detailed examples and to provide context, but findings are not specific to a case; all findings were checked for representativeness against data from all participants [36]. All names were replaced with pseudonyms.

But having a link in an email is still suspicious; why would the bank ask him to log in rather than just inform him that he "has a notification on your account"? Uncertain, Josh decided to go to the bank login page by typing it into the address bar. Josh let autofill complete the url address.

When he logged into his account he saw a notification alert on his account summary, but before opening the notification he quickly skimmed his balance on his accounts. Everything looked normal. So, he read the bank notification and saw a notice that customers had been receiving an illegitimate email asking them to verify their account information. The bank instructed him not to engage with this email.

At this time, Josh knew that the email was a phishing attack. But just to be sure, he opened his transaction history and verified the content for the last few weeks, seeing no unusual transactions. He then deleted the email.

4.1 Stage 1: Sensemaking

All participants except one described a very similar process as the first thing they did when they received the email. They began with two goals. First, they try to understand *why* they received the email. Each email is not just a simple request; understanding the larger context of the email is important to be able to appropriately respond. Their goal is effectively to construct an initial frame of the email that situates the email in a larger story involving other situations in their life.

The second goal that most, but not all, participants explicitly described was trying to figure out what action they had to take to deal with this email. They saw email in general as a series of action items that had to be "dealt with", and dealing with email was a major time commitment in their work. To try to minimize that time commitment, they explicitly stated a goal of trying to quickly figure out what action they had to take do "deal with" the email.

Reading an email is a complex task that involves connecting words and images on the screen with many other aspects of one's life in order to understand what the email is about and make sense of the email. Weick [52] calls this process "sensemaking", and shows that it is one of the primary cognitive activities that people engage in. Sensemaking is usually initiated when people encounter new information (such as an email), or when some surprise indicates the inadequacy of their current understanding, and continues until people feel like they understand the situation they are in.

Klein et al. expanded on Weick's idea of sensemaking, and introduced what they call the 'data-frame' theory of sensemaking [29]. When a person is in a situation, they have a "frame" for understanding that situation [18]. This frame could be a story that explains the chronology of what is going on, a map that explains spatial relationships, a script explaining roles, or a plan for what should happen. Frames provide an explanatory structure that describes relationships between entities, such as between the email and other events in the person's life.

In order to deal with the email quickly, participants had a goal of identifying what action was needed. They built an initial frame of understanding for the email with this goal in mind. Some emails were just informative, and no action was needed; participants liked those emails because reading them was all that was necessary and they could get through those quickly. Some emails needed responses, or required more labor. So the sensemaking goal they had when reading an email was to identify what action was needed. Once they understood the needed action, then they could make intelligent decisions about priorities, etc.

4.1.1 Taking the Email at Face Value. During this sensemaking process, participants took the email at face value; they accepted information presented in the email as fact, including the displayed sender and information presented in the content of the email body. For example, P16 received a phishing email presenting itself as a "party planner" invitation from a friend for a potluck dinner.

During this first stage, he took the email at face value, looking at his calendar to see if he was available on the dates in the email and looking up recipes for dishes he could bring to the event. In Case 1, Josh received an email pretending to be from his bank, asking him to verify his account. Assuming the email was legitimate, he became concerned that his account had been compromised and was concerned that his money had been stolen.

4.1.2 Problem Detection. As participants noticed new aspects of the email (for example, as they read the body of the email), they added this new information to their understanding to enrich their current frame. However, some of the cues that they noticed did not cleanly fit into the current frame. These *discrepancies* are often noted, but do not necessarily change the current frame of the email.

Josh noticed that the email from his bank directly asked him to verify information. He found this odd; he remembers that emails from his bank normally just say that he has a notification on his account. This is a discrepancy: a fact that was noticed about the situation that does not cleanly fit into the current frame.

A single discrepancy is usually not enough to cause the participants to question their current frame for the email. However, as they make sense of the email, they may notice multiple discrepancies that do not fit into the current frame. Eventually, there is a point where the participant decides that their current framing — taking the email at face value — isn't sufficient, and he or she starts questioning the frame. That is, they detect a problem with their current framing of the email. *Problem detection* requires an active decision by a person that the current frame might not accurately reflect an appropriate understanding of the situation / email [25].

4.2 Stage 2: Is this phishing?

Once a person detects a problem with their current understanding of the email, they experience a cognitive shift. Participants changed their activity away from sensemaking, and instead tried to determine if the email was legitimate. This shift was usually evident in the timeline description from participants; participants would describe a clear change in the goal they were trying to accomplish. Rather than trying to figure out what the email was asking them to do, participants described their new goal as trying to determine *if* the email was legitimate. That is, they wanted to know if the email was what is said it was, or if it was a phishing email. Josh experienced this shift in the 3rd paragraph of Case 1.

This new goal captures the second stage of phishing detection. Rather than making sense of the email, participants begin searching for evidence that would help them determine if the email is legitimate. For example, participants would hover over the link the email (P12), click on the name to view the email address of the sender (P9), open up the headers of the email to see where it came from (P7), or run the attachment through a web-based antivirus detector (P4).

4.2.1 The role of training. When asked about how they knew what actions to take that would help them achieve this goal, almost all of the participants cited explicit security training that they had received. Looking at what information participants reported using this point, this information often matches with common security training around phishing — specifically hovering over links and looking at `from` addresses or senders (as Josh did in Case 1).

Most of the information that was sought during this stage is what I call *conclusive distinguishers*: information that can conclusively establish that an email is a phishing email. If a person hovers over a link and it doesn't link to the expected domain, then with high confidence the email can be declared to be phishing (or, at least, not what it says it is). Most current phishing training focuses on helping people identify conclusive distinguishers [23, 40], and it is exactly these distinguishers that the experts looked for: URLs, email headers, IP addresses, and other hard-to-fake technical information. Participants described some time pressure here and hoped to quickly identify if the email was phishing using these distinguishers.

4.2.2 *Difficulty Inversion.*

In this stage, I observed an inversion of difficulty. The vast majority of emails are not phishing, so participants initially assumed legitimacy. However, after experiencing this cognitive shift and becoming suspicious of the email, participants found it easy to conclude a message was phishing but difficult to conclude it was legitimate.

For example, P4 looked at an email from someone asking for a job that was sent to a high-level manager. She was suspicious that it was a phishing email due to the link in the message. She hovered over the link (looking for a conclusive distinguisher) and it looked like a legitimate link to a resume. She examined the from address and headers (another possible conclusive distinguisher) and they matched exactly with the stated identity in the email. She ran the attached PDF through an anti-virus (a third possible conclusive distinguisher), which reported that the file was safe. She looked up the domain names in the links and they appeared congruent with the rest of the email. Then she opened a disposable virtual machine and visited the link in this safe virtual machine and examined the resulting website for inconsistencies. Only after all of this work was she willing to conclude that the email was not a phishing message.

Four participants described a message that they thought might be phishing but actually turned out to be legitimate. For these participants, they followed their training and looked for conclusive distinguishers, but found that those distinguishers did not establish the email as phishing. They had to go through much more effort to then convince themselves that the email was actually a legitimate email. Conclusive distinguishers provide strong evidence of an email's phishing nature, but provide only weak evidence of its legitimacy.

4.3 Stage 3: Dealing with the email

After coming to a conclusion about whether the email was phishing or not, participants entered a third stage where they had to decide what to do with the email. At this point, participants stopped looking for new information or trying to change or enrich their framing of the email. Instead, they focused on action: finally accomplishing that first goal of dealing with the email. This goal change is evident in the final paragraph of Case 1.

The majority of participants had a very simple action for phishing messages: delete the email (9 participants). These participants often justified this action in very practical terms: their goal was to take whatever action they needed to to deal with the email, and no action was required (since it was a phishing message), so they were done with the email and deleted it. In Case 1, Josh deleted the email after verifying his bank accounts.

Instead of just deleting the email, 9 more participants felt some responsibility to the organization to help them protect against phishing attacks. For two of these, it was explicitly part of their job, but the remaining seven participants did so because they wanted to be helpful. All of these participants described additional actions, such as forwarding the email, along with a note with their findings, to the abuse@XXX.edu email address where phishing messages are supposed to be reported. Four participants undertook additional investigative steps (such as scanning attachments for viruses or opening the link in a virtual machine) before reporting the email to the organization responsible for it.

5 HOW EXPERTS BECOME SUSPICIOUS

The shift from stage 1 to stage 2 appears to be the critical event for IT experts in the detection of phishing emails. At this point, participants detect a problem with their current understanding of the email (the current cognitive *frame* [29]), and shift their goal from doing whatever the email asks to figuring out whether the email is really a phishing message. Since these experts found it easy to determine that an email was a phishing message after this shift, it appears that this shift is the critical event.

5.1 Research on Problem Detection

Klein defines problem detection as identifying that a current cognitive frame is inadequate to explain the current situation. Many complex real-world decisions involve problem detection. For example, Klein, Crandall, et al. examined how neo-natal nurses recognize when a newborn baby is in the process of developing sepsis — which is really identifying when the normal framing (the baby is healthy) is the incorrect frame for the baby [25].

People often identify problems by noticing discrepancies between the real world and their understanding of the world. Once enough discrepancies build up, the person *detects* that there is a problem. This buildup of discrepancies model is a very common model of *problem detection* [10].

Klein identifies two major critiques with this buildup-of-discrepancies model of problem detection: [25] First, it requires effort and expertise to *notice* a discrepancy. Two people can be in the same situation, but will notice different aspects of that situation. Second, problem detection does not occur simply by the accumulation of discrepancies. Klein instead proposes that problem detection is a frame shift from one frame to another. That is, before a person can detect a problem, they need to identify an alternative framing for the situation that they can shift their understanding to. Without this alternative explanation, discrepancies do not trigger problem detection.

5.2 Analyzing Cues to Understand Suspicion

Building on this conceptualization of problem detection, I re-analyzed the interview transcripts to focus on the *cues* that my participants noticed. Cues provide the primary input to decision making – they are the data about the world that a person uses when making decisions. They provide stimulus; noticing a cue in the world can cause a person to think more carefully and change aspects of the decision-making process.

The cues I identified have three major properties [24]. First, cues exist in the world; in theory, they could be independently factually verified by other people. Cues are things in the world, not inside a person's head. Cues are external to the decision-making process, but provide new information and input into that process. Second, not everything in the world is a cue; cues have to be explicitly *noticed* by a person for them to become an input into the decision-making process. What people notice and don't notice is complicated; people don't notice everything in front of them, and different people often notice different things in the same situation. And finally, noticing a cue is contemporaneous; remembering something that was seen or known in the past is not a cue. Even though remembering a fact that was first noticed in the past can serve as a new input in decision-making, that fact had to be noticed at some point in the past first.

Two research assistants and I carefully re-analyzed the transcripts for the 19 participants that described specific email incidents, and for each incident we made a *cue inventory* [11, 22]: a list of all of the cues that the person reported noticing during the incident. Because these were retrospective accounts, it is possible that the participant noticed something at the time that they did not remember or mention during the interview. However, most of the important cues – the cues that influenced and shaped the decision-making process of the participant about this email – were likely to be mentioned and included in the cue inventory [22]. Two members of the research team independently created a cue inventory for each transcript, and then met weekly to compare the inventories, discuss discrepancies, and decide on a final list of cues based on the definition of "cue" above.

After the cue inventory was identified, we ordered the cues in the chronological order that they were first noticed in the incident. Often, participants were not clear about the order in which cues were noticed; we grouped cues into clusters that were all noticed at approximately the same time. We cannot definitely identify ordering within a cluster, but participants were clear of the ordering across clusters. After clustering and ordering cues, we coded the cues for which of the three decision-making stages the participant was in when they first noticed the cue: face-value sensemaking, suspicion, or dealing with the email.

Finally, for each cue that we identified, we coded for properties of that cue in the decision-making process of the participant. We coded whether the participant identified the cue as a discrepancy with their current understanding, and if so, what type of discrepancy. We coded if the cue was explicitly scanned for, or simply noticed. We coded for negative cues — noticing when something is missing from the world (which are more difficult to notice and require expertise to know that it should be present). And we coded for reactions that the participant might have taken to each cue, such as explaining away discrepancies, or having a really strong reaction to the cue (what Klein et al. call an "antibody reaction" [25]).

This coding was also done by two members of the research team. The level of analysis was a "cue", and each cue was associated with a series of statements in the transcript. We used an excel spreadsheet with columns for the different codes and rows for each cue. We met after coding each 2-3 interviews to measure reliability and discuss any differences. Differences were discussed until a final agreement was reached, and the final agreement was used as the code for the data. This process is based on the one described by Hoffman et al. [22].

During the process, we developed written guidance for what is a cue, and a codebook defining the properties of cues. These documents are available in the supplementary materials for this paper.

5.3 Cues: Noticing Requires Expertise

A person has to *notice* a cue before he or she can use it to influence their decisions. Noticing is cognitive work. People don't notice everything, but instead only notice things in the world that they have a mental framework to understand that in some way contribute meaning to them. Two people can look at the same thing in the world but notice different things because those different things have different meaning to them.

Noticing requires *expertise*, and the expertise an individual possesses influences and shapes what they notice about an email. Possessing different expertise is why two people can look at the same thing and notice different things about it. Klein and Hoffman, in their paper "Seeing the Invisible", identify three ways that expertise influences noticing [28]: (1) experts can see what typically happens in a situation, and when elements are missing that are typically present; (2) experts can see fine distinctions in situations that are overlooked by non-experts; and (3) experts can see antecedents and consequents, visualize how a situation got into its current state, and how it will continue to develop.

Consider what Ashley notices in this example:

Case 2. Not My Paypal: *On her birthday, Ashley (P9) was working and checking her email. She noticed an email come in from PayPal with a subject like "VISA expired". She thought this was confusing, because she doesn't have a PayPal account. Her husband does, and it is linked with their bank account, but she doesn't. She was concerned by this email, and wanted to make sure she wasn't being thefted [sic]. She clicked on the email to read it. She noticed that it didn't have an attachment (good; she usually just deletes emails with attachments). She looked to see whether the email was directly asking for a new credit card number, or whether it was asking her to log in to her account. Directly asking is a common tactic she has seen in past phishing, but this email wasn't directly asking. She checked to see if the login button went to a real site by hovering over it, and it seemed to — it went to a* `paypal.com`

URL. She also noticed that the email looked "on brand" — it had appropriate colors, the images were of appropriate resolution, and it looked purposefully created and designed like it was coming from a CRM tool. She works with brand standards for her job, so she naturally recognizes things like this.

She still was skeptical and wanted to be sure there weren't any identify theft issues. She clicked reply and looked at who the reply defaulted to — it went to a `paypal.com` *email address and didn't fill in any other information. She deleted the reply; she just wanted to see what email address it would go to. She then went to the web for more information. On paypal.com, she verified that the branding in the email matched the webpage. She decided to forward the email to a paypal address she found to "cover her butt."*

At that point, she decided she wasn't freaking out. She decided "meh"; she didn't need to follow up any more because there was no danger, so she just moved on.

When she first opened and looked at the email, Ashley noticed things that most people would notice — she noticed what the email was about and what it was asking her to do. But, because of her background and experience in marketing and education, she also noticed things like the resolution of the images and the shades of colors that many of us would likely have overlooked.

In examining what the IT experts notice about emails, I found evidence that all three cognitive-perceptual skills influenced what was noticed. In Case 2, Ashley uses her expert knowledge of phishing emails, and what is typically present in phishing emails, to recognize things that were missing from her email. She noticed that the email was not directly asking for her credit card number. She specifically noticed that this request was missing because she applied her expertise to recognize that phishing emails (at least in her experience) often directly request the information they want, and this email wasn't.

Ashley's ability to recognize the resolution of the images is an example of the second cognitive-perceptual skill: seeing fine distinctions. Ashley was able to recognize the resolution of the images as not too high and not too low, but appropriate for the type of message being sent. Most people don't recognize the resolution of images, and even if they did, wouldn't know whether it was higher or lower than it should be. But Ashley's expertise in CRM software enabled her to apply meaning to the images and recognize fine distinctions in resolution that other people wouldn't.

Finally, Ashley's expertise in phishing enabled her to understand how the situation would develop in the future, which enabled her to not freak out and end up deciding "meh" about the email. She understood that, because she received this email, the attackers did not yet have the information they wanted from her, and her account was still safe.

The important point here is not what she noticed (image resolution and shades of color), but how she noticed it. Ashley used expertise that she had developed for other means to notice features of the email that had meaning to her. Other participants do not have her expertise in CRM systems, and thus did not recognize those exact cues. But all participants used their own knowledge and expertise, including non-technical and non-security related expertise, to notice features in emails that have meaning to them.

5.4 Cues: Noticing and Mindset

Expertise shapes what people notice in a situation because it gives meaning to the things that they see. Expertise is relatively static, but the same person might notice different things at different times. Consider this example:

Case 3. The Typo: *A year or two ago, Sarah (P11) had received an email that seemed to be from Sprint (but wasn't); all she remembers about the email was that it had a typo in it. When she clicked on it her computer went blank and stopped working. She lost the computer for a week and had to use* *a loaner computer while it got fixed. Since she was a manager, all of her co-workers knew what had happened.*

Sarah was at work in the afternoon one day checking her work email. She saw an email from "AT Order" with the subject "Ann Taylor order". She remembered that she had recently (last couple of days) ordered from Ann Taylor. She has a 2-second rule for deciding if an email can be deleted, or if it needs to be dealt with later. She decided to deal with it later because it could be something about her recent order, and could have financial implications.

Later in the day, she opened the email and read it to try to figure out if it was important and was something she needed to deal with. Why were they sending this email to her? She noticed that it looked similar to other (advertising) emails from Ann Taylor — it included the logo, the legalese at the bottom, and was directly addressed to her. It said there was an "issue with your order" and had a big box where she could click to "check status" of the order.

Near the end, she noticed a typo in the email. She didn't remember the specific typo, but she remembered there was one. This caused her "audit mind" to kick in, and she wondered if someone was trying to scam her (remembering the Sprint email). She read the email more closely and started to notice "lots of stuff" that was off about the email. She noticed more typos in the email. Lots of typos. She noticed the salutation was not normal for an American (she is Canadian) – something like "Pleasant Day". She noticed grammar issues. She noticed that the paragraph structure didn't seem like a normal American structure.

She thought of this email as "gray" – she wasn't sure if it was legitimate, but wasn't convinced. She checked the email address, and saw that the email address was a bunch of characters at a domain, rather than a name (like customer service) or noreply or something like that. This told her conclusively that this wasn't a legitimate email.

In this case, Sarah read through the same email twice. Sarah's initial frame associates the email she received with the clothing order she placed a few days prior. She noticed things like the logo, the main point of the communication (an issue with her order), and where she had to click to move on.

The second time through the email, though, she was suspicious of the email. She noticed typos and other issues that she didn't the first time through. This change to a suspicious mindset changed what had meaning to her, and thus allowed her to notice things that she didn't notice the first time.

According to Klein et al., the frame that a person has influences what data (cues) that person notices in the world, and the data that is noticed influences the frame that people use to understand the data [29]. Klein et al. argue that "the data identify the relevant frame, and the frame determines which data are noticed. *Neither of these comes first.* The data elicit and construct the frame; the frame defines, connects, and filters the data." Another way of saying this is that the mindset a person has when reading an email influences what he or she notices, and what is noticed simultaneously influences that person's mindset.

5.5 Discrepancies

As each participant read the email, he or she would try to integrate each new cue that they noticed into their existing frame for the email. Sometimes, though, they would notice cues that had meaning to them, but that they couldn't easily integrate into their frame. Each of these cues was a discrepancy. Discrepancies play an important part in the emerging sensemaking process as people try to understand the email. As Klein argues, they also play an important part in problem detection [25].

45%[2] of all cues that my participants noticed were discrepancies with their current frame. During the first stage, when participants were still taking the email at face value, 35% of the cues were discrepancies. The majority of cues noticed were not discrepancies; because these were reasonably good phishing emails (or legitimate emails), most of the cues made sense and could be integrated into the story for that frame.

Across the emails that participants described, I found three different types of discrepancies that played roles in participant decision-making processes: inconsistencies, typicality violations, and frame violations.

Inconsistencies. Inconsistencies are internal contradictions between cues. An individual cue cannot be an inconsistency, but two cues can be inconsistent with each other. Inconsistencies are internal to the a given frame; they are two things that both make sense as part of the story but cannot both be part of the same story. However, inconsistencies are external to the decision-maker; the information needed for the inconsistency comes entirely from cues noticed in the world, and not from background knowledge.

Inconsistencies played a relatively small role in detecting phishing. Only 27% of discrepancies (19% of all cues) were inconsistent with another cue that was noticed. For example, P7 noticed that an email appeared to come from someone with an email address in his same organization (same domain after the '@'), but when he looked at the headers of the email, it was not sent from a mail server in that organization. He described this as an inconsistency; either it came from inside the organization or not, but both can't be true. In this example, P7 didn't notice this inconsistency until after he was already suspicious, and explicitly sought out the information in the headers.

Typicality Violations. The second type of discrepancy I found were typicality violations. A typicality violation is an individual cue that violates the person's expectations for what is typically present in similar situations. The person has some expectation for the email and the cue that is noticed doesn't meet that expectation. 43% of all discrepancies were typicality violations.

In typicality violations, expectations come from patterns in prior, similar situations. The person notices that prior, similar situations typically have some feature present, and then expect the same feature in this situation. Typicality is not necessarily the result of cause-and-effect, and people don't need to have a strong reason to believe that the cue is important. It just needs to violate what the person has previously noticed as typically present. For example, P3 got an email from someone in his same organization who he had never previously communicated with. The email was written in a comic sans font. P3 found this to be odd (that is, a discrepancy) because he doesn't normally see professional, work emails written in that font (it violated what he typically sees).

Typicality violations can be either positive or negative. Positive typicality violations are when a person notices something present that is different than it typically is. The font discrepancy that P3 noticed was a positive typicality violation because he directly noticed the font.

Negative typicality violations are when the person notices something is not present that typically is. Negative violations are more difficult to notice because it requires more expertise to understand that something typically is present but is not in this situation. In Case 3, Sarah noticed that the email did not contain any details about her order. She recognized that this is unusual; in her experience emails from retailers usually state what order the email is about. This was a negative typicality violation because she noticed something was missing, which required her to think back to previous messages to recognize that it wasn't present. Negative typicality violations are less common than positive ones. Only 24% of typicality violations were negative; the other 76% were positive.

[2] Note that these percentages refer to the specific participants and the specific emails received by these participants. These numbers are descriptive, but are unlikely to generalize to other situations.

Frame Violations. Third, frame violations occur when a person thinks through the frame (the emerging story) that they are developing for the email, and deduce logically what should or should not be present. A cue is a frame violation if the person observing the cue cannot integrate the cue into his or her current frame of understanding for the email.

Frame violations were relatively rare in my data; they only represented 13% of the cues that were noticed (25% of discrepancies). Because we coded typicality violations separately, frame violations were only coded when the participant explicitly talked about some logical thought process that led them to conclude that the cue didn't fit into the frame.

5.6 The Need for an Alternative Explanation

Noticing discrepancies isn't enough to recognize the email as phishing. Instead, the role that discrepancies seem to play is that they create a need.

What I've called discrepancies, Klein calls "disturbances" [25]. He argues that cues (data elements) that don't fit into the current frame disturb the person who notices them and create a feeling of anxiety. That low level anxiety causes people to feel a need to redefine their understanding of the situation. As Klein says, "the positive role of anxiety in problem detection can be seen as helping to question whether the current assessment or current line of activity are still appropriate." [25]

In the participants, this anxiety manifested itself as a feeling of a need for an alternative explanation. They noticed discrepancies between what they were seeing in the email, and the story that was forming in their head. Participants felt uncomfortable about this, and felt like they needed to find some way to understand or explain these discrepancies.

To see this, consider the following case where the participant did not develop anxiety. The participant explained away the discrepancies he noticed, and as a result did not feel this need for an alternative explanation. This actually prevented him from becoming suspicious and considering alternative explanations until he was forced to by his friend's mother.

Case 4. The Party Invite: *Nick (P16) was checking his personal email on his phone while in the car with his wife driving. He noticed an email from Mary, the mother of his friend Zack, who he knew well. It was an invitation to a party from a party planning service; it was a form email about the event a couple weekends from now with a customized message like "sign up for a dish to pass". This seemed like something she would send; she frequently has such parties though she normally talks to him in person about them because he often makes the main meat for such parties.*

When they got home, his wife went to check the physical calendar they keep on the fridge to see if that weekend was open. While she was doing that, Nick went to his computer and started looking at recipes for things he might want to make for the party. His wife confirmed that they were available that weekend, and he started to get excited about a party — particularly about making some interesting food for the party (which is his favorite part of such get togethers). After he made some decisions about what to bring, he went to sign up through the website. Following his normal practice, he typed in the URL in the email rather than clicking on the link. When he did that, the website came back with an "event not found" error message. He guessed that she just fat-fingered the URL, so he went and hovered over the URL. He noticed that the link actually went to a site in the UK — it looked like it might be a CDN (content delivery network) — but it didn't seem like the same website for the party planner.

Still excited about the party, he decided to just call up Mary. He asked her what the event code for the party was. Mary responded confused; she wasn't planning a party. She said her son Zack had also gotten an email about a party from her. At this point, Nick knew that the email wasn't legitimate. Mary said that Zack thinks she was hacked. Nick, trying to be helpful, started making suggestions from his IT background about what to do. Mary responded that her son Zack has told her the same

Cue	# Participants
Action Link	8
Typo	3
Suggested by another person	3
Recipients in from line	2
Unusual subject line	2
Attachment	1
URL	1
Unusual information requested	1
Never became suspicious	2

Table 3. Triggers: the last cue noticed by participants before shifting to be suspicious. Counts are the number of participants who reported this cue last before shifting.

things. Nick concluded that "Zack was on this" and hung up. Disappointed that there wasn't a party, Nick deleted the email and informed his wife that the party wasn't real.

Nick noticed multiple discrepancies with this party invite email. He thought it was unusual that he got an email about it rather than hearing about it in person (a typicality violation). He got an "event not found" error (a frame violation). He noticed the URL didn't go where he expected it to (an inconsistency).

As he noticed each discrepancy, he explained it away. For example, he assumed he "fat-fingered" the URL by typing it in incorrectly, and that was why he was getting the event not found error message. By explaining the discrepancy away, he removed the anxiety that comes with discrepancies. He no longer felt a need for an alternative explanation, and thus when faced with evidence that normally would be conclusive (the bad URL), he wasn't in the frame of mind to recognize it.

5.7 Triggering an Alternative Framing: Something to shift TO

"Failures of problem detection are not so much failures to detect an indicator, but rather they are failures to reconceive or redefine the situation." – Klein et al. [25]

Above, I argued that becoming "suspicious" about an email is actually a cognitive shift; the person shifts from framing the email at face value, and instead becomes suspicious that the email might be a phishing email. Klein shows that a person can't just "detect a problem"; instead, the person needs a possible alternative explanation to shift into [25]. For phishing, "the email is phishing" is one such alternative explanation.

To identify potential triggers for this shift, I looked at the cue inventory for each participant, and identified the last cue that was noticed before shifting to suspicion. If this cue is strongly associated with phishing in the participant's head, then it is possible that noticing this cue could bring to mind phishing emails, and enable the person to realize that phishing is a potential alternative explanation.

For each of the possible triggers, Table 3 shows how many participants reported noticing that cue last before shifting to suspicion. The most common trigger was an action link. Action links are links in emails that the email is asking the person to click in order to do something. Action links are not URLs; in all cases, the participant had not yet seen the URL. Instead, action links are just the presence of a link requesting an action. Most participants described that they believed that this was a common feature of phishing emails.

Triggers are not necessarily discrepancies, but discrepancies can serve as triggers. Indeed, if the email is being taken at face value, often an action link makes sense to be present in the email, and therefore participants don't see it as a discrepancy. For one participant, typos made sense in the email (because it was written quickly) and he saw it as odd when the email didn't have any typos.

All of the triggers in Table 3 seem to be associated with phishing emails in the minds of the participants. The discrepancies noticed previously created a need for an alternative explanation, and then the trigger cue enabled the participant to identify phishing as a possible alternative explanation. That is, the trigger cue helped the participant identify an alternative framing of the email that they could shift to. Only once they had this possible alternative did they actually become suspicious.

6 DISCUSSION

Every email that a person receives could possibly be a phishing email. Many people receive tens or hundreds of emails per day, and only a small number of them are phishing messages. As Sasse [42] argues, "Usable security acknowledges that users are focused on their primary goals" and that "disrupting these primary tasks" with security concerns can create "a huge workload" for users. Each email message received is a decision (is this a phishing email?) that needs to be made, and if done poorly, making these decisions can be a huge workload.

In this paper, I describe a common process that IT experts follow to identify phishing messages. This process has been successfully integrated into the workflow of these office workers, scaled up to normal levels of email, and helps them restrict their time-intensive investigations to only those emails that warrant it.

All of these participants described applying their phishing training in stage 2 (and occasionally in stage 1), and that training helped them to know what to do. This suggests that existing training is helpful and effective, at least for these IT professionals. However, the participants rarely described any influence of phishing training during stage 1. It is not until *after* becoming suspicious and questioning the initial framing of the email that the phishing training is normally used. However, most email messages never make it out of stage 1. This suggests that current training is missing an important aspect of how these experts detect phishing messages.

Past research has identified a number of features of emails that can indicate phishing, such as typos and grammatical issues [6], URLs that do not link to the correct website [30, 45], social influence strategies [35], and appealing to people's curiousity [5]. Many of these features are already mentioned in end user training as things to look for when identifing phishing messages.

Many of the cues noticed by participants in this study are mentioned in prior research. However, it isn't enough just to recognize the presence of a given cue in an email; you have to know what to do with it. It is not enough to known *which* cues signal phishing to people; we need to understand the different *roles* that each cue plays in the complex decision-making process of identifying phishing emails. Some cues build anxiety (discrepancies); some help people remember alternative explanations (triggers), some support definitive conclusions (conclusive distinguishers), and some serve multiple roles at different points in the process. All three types of cues are needed for experts to identify a phishing email, and no single type of cue is enough.

6.1 How Expertise Matters

In order to detect an email as a phishing message, all of my participants used their expertise in important ways. To detect phishing messages, three distinct types of expertise were used.

First, expertise in the *face topic* of the email is needed. The face topic is whatever topic the email says it is about, assuming the email is truthful. For a large number of my participants, this was actually expertise in the organization they worked for. For example, participant P1 talked about

Failure	*Normal Process*	*How It Fails*
Automation failure	Making Sense of the Email	Not engaging in enough sensemaking; automatically doing what the email asks
Discrepancy failure	Knowing what is typical, and noticing discrepancies	Not noticing discrepancies because you don't know what to expect
Accumulation failure	Discrepancies create a need for an alternative explanation	Explaining away discrepancies
Alternative failure	Identifying phishing as an alternative explanation	Noticing discrepancies, but not identifying other possible explanations (not becoming suspicious)
Failure to Investigate	Examining email for phishing indicators	Become suspicious, but don't know what to look for
Evidence Failure	Examining email for phishing indicators	Be suspicious, investigate, but find bad information that leads to the wrong conclusion

Table 4. Ways that people can fail to detect a phishing email

how their organization doesn't usually send business emails outside of business hours (for an email asking to update benefits information). P8 complained that an email didn't have a full, long signature, which is normal for his organization. P17 got an email from a woman who worked down the hall from him, but he felt was weird because she never emails him, and the company culture was to personalize emails and this email wasn't personalized. Face value expertise is primarily used during the first stage, and helps people to recognize discrepancies.

Second, expertise in *phishing scams* is also needed. This expertise includes background knowledge that phishing emails exist, and some knowledge about what these emails look like. Phishing expertise is used is to recognize phishing as a possible alternative explanation for the email. Something needs to trigger this alternative explanation. For most participants, this expertise in phishing scams did not come from formal training. Rather, it came from experience seeing phishing messages, and from office gossip about prior phishing messages that people had seen.

Third, once suspicious, people need to know what information to look for to determine whether this email is actually phishing or not. Participants, which were all IT experts, turned to their *technical expertise* and explicit training in phishing. Each person uses that expertise to specifically look for things (like URLs, email senders, email headers, or IP address information) that can help them determine if the email is legitimate or phishing. This is important because becoming suspicious isn't enough to conclude an email is phishing; people need to investigate to be sure.

6.2 Ways to Fail in Detecting Phishing

By examining the process that IT experts use to identify phishing messages, I identified six places where this process can fail. These failures are described in Table 4.

In the normal successful process, the first thing that experts do is try to make sense of the email. This can fail when processing email automatically. I call this an *automation failure*. If a person doesn't try to make sense of the email, and instead just automatically does what the email says, then this bypasses the whole detection process and doesn't enable them to notice discrepancies, become suspicious, or further investigate the email.

A *discrepancy failure* occurs when a person is trying to make sense of an email, but doesn't notice discrepancies. Most discrepancies are typicality violations or frame violations, but a person can only notice these when he or she has enough expertise in the face topic of the email. This failure often comes from a lack of expertise; but it is not phishing or technical expertise that is missing, but rather expertise in the face topic of the email.

An *accumulation failure* occurs when a person explains those discrepancies away one by one without allowing them to build up into a need for an alternative. As Klein observes, explaining away discrepancies more commonly occurs among experts because they are able to come up with more plausible explanations that can be used to explain away discrepancies [24]. While discrepancy failures occur due to lack of expertise, accumulation failures occur from too much expertise in the face topic of the email. Nick, in Case 4, had an accumulation failure.

An *alternative failure* occurs when the person never identifies phishing as a possible alternative explanation. Almost all of the participants in this study seemed to be triggered to remember that phishing is a possible explanation by something they noticed in the email. Without such a triggering cue, the person never identifies an alternative explanation and leaves the discrepancies unexplained. We never fully understand emails we receive, so even in the face of discrepancies, a person might not become suspicious. Identifying phishing as an alternative relates to phishing expertise, so this failure is most likely to occur in people who don't have much experience or expertise in phishing emails.

Once a person is suspicious, normally they investigate the email by explicitly seeking out information that will help them determine whether the email is legitimate or not. Most of the participants indicated that they knew what to look for because of phishing training. However, people with less technical expertise might not know what to look for or what to do, and thus might *fail to investigate* the email.

Finally, even if a person does investigate the email, it is possible that the investigation produces erroneous or incorrect information. This is an *evidence failure*. This type of failure occurred during the 2016 US Presidential election, and played an important part in that election [32, 37].

6.3 Implications for Non-Expert Users

The primary contribution of this paper is a description of the *process* that IT security experts use to identify phishing messages. It is important to understand and improve the ability of experts to detect these messages; they are often targeted for phishing attacks.

However, the knowledge of the cognitive processes described in this paper also has implications for improving the ability of non-expert users to detect phishing. This paper describes a realistic upper bound for how well end users can do in phishing. Even security experts are not perfect and do not detect every phishing email. Security experts do not do full investigations of every email they receive (according to data presented above). We cannot and should not expect end users to be perfect in recognizing phishing emails. This paper describes a reasonable upper bound for expectations of end users – that is, it describes a reasonable, usable process that end users could use to identify a reasonable number of phishing emails if they were trained well.

Knowledge of the process described above can help improve end-user training. While this paper does not have data about how to deliver effective phishing training, it can help inform the content of that training. Most end-user training focuses on general awareness ("this is what phishing is") and on specific things to look for ("hover over links") [17]. Rarely does existing advice mention processes, and often the process advice is unrealistic (e.g. "hover over every link in every email"

[6]) and isn't even followed by experts. Instead, we can provide more actionable training to not investigate every email but only ones that are "suspicious", and focusing our efforts on helping users become appropriately suspicious.

It is also possible to improve "awareness" training about phishing. Much awareness training emphasizes what phishing is, and what is trying to do [17]. The detailed process described above shows the role that this awareness plays: it provides an alternative explanation. It also suggests a new goal for awareness training: to cognitively associate "triggers" present in many emails with the idea of phishing emails. For example, end-user training might usefully focus on associating *action links* with phishing (rather than trying to associate all links with phishing). Triggers should be something that people already naturally notice (like requesting an action) rather than things that people need to go out of their way to notice (like hovering over a link to see the URL). The details presented here emphasize the importance of triggers (even suspicious experts wouldn't investigate the right things to notice a phishing email without them) and important constraints on them (already notice rather than explicitly investigate).

Another finding of this paper is in the role that contextual knowledge and non-technical expertise can play in phishing detection. This suggests that phishing training can be usefully customized for different contexts and groups of users with shared expertise. Such customization should encourage users to use their knowledge of context and existing expertise in other domains to become suspicious of emails.

6.4 Negative Uses of this Theory

The goal of this work is to develop a theory of how people identify phishing emails, so that we can better train users to recognize phishing emails and prevent them from falling victim to phishing messages. However, this theory (and any similar theory) is dual-use, and may possibly be used by malicious actors to improve their phishing messages. In developing and describing this theory, I focused on how it can be used to improve defensive training, and I did not identify methods that can trigger detection failures.

6.5 Limitations

All of the expert participants in this study work for the same organization. They received formal training from a variety of schools, possess a variety of different types of degrees, and many worked for other organizations before this one. Still, organizational culture is strong, and could be a strong influence on the approaches and strategies used by these experts. It is possible that this study is biased by the fact that all of the participants work in the same organization.

7 CONCLUSIONS

Phishing is a difficult problem to protect against and current technical solutions do not completely solve the problem. In this paper, I describe the naturalistic process that IT experts use to recognize and identify phishing emails in their own inboxes. The crux of this process is noticing discrepancies between the email and what is typically expected to be in similar emails, and then identifying phishing as a potential alternative explanation for these discrepancies.

Many of my participants described a triage process (e.g. a "five second rule") for quickly determining if an email is worth reading at all. Emails that were not relevant were deleted without being read carefully, and this could include phishing emails. I did not study this, and the theory above does not cover this case. Participants used the process presented here for more difficult emails that are attractive or important to the recipient for some reason.

Much prior work has attempted to design ways to train users to identify phishing messages [30, 45, 51]. Most of this work focuses on training users to use conclusive distinguishers — features of the email, such as URLs, that can conclusively establish that the email is a phishing message. In this paper, I show that there is a very important role for non-conclusive distinguishers — features of emails that seem off, but cannot conclusively identify that the email is a phishing message. In particular, all but one of the experts I studied had to notice a number of non-conclusive discrepancies before they ever began looking for conclusive distinguishers.

I found that phishing expertise played two important roles — triggering the person to identify phishing as a possible alternative explanation, and knowing what information to look for once they become suspicious. Current training focuses on this second role – knowing what to look for. It would be helpful if phishing training could also help people to associate common features of phishing emails (like action links) with phishing generally, because that will enable those features to better trigger people to remember phishing as a possible alternative explanation.

In addition to phishing expertise, I also found a role for expertise in the face topic of emails. This expertise helps people notice discrepancies by recognizing what is typically present in situations. Organizations should try to be consistent in legitimate emails, and give people many opportunities to see and recognize those consistent features of legitimate emails. This will help people they interact with recognize discrepancies when messages do not follow those typical practices.

Interviewing IT experts, Theofanos et al. found that experts claim that they are distrusting of everything online [49]. As general comments, some of my participants also stated this (unprompted). The naturalistic decision making literature refers to this as a "stance" [24]; IT experts have a generally skeptical stance toward emails they receive. However, it is not clear how helpful this stance was or even if it is truly present in the cases I studied. All but one of the participants initially trusted the email; it was only after noticing discrepancies and being triggered that they became suspicious of emails. Some end user training encourages users to adopt a similarly skeptical stance; it isn't clear from this research whether such encouragement is beneficial.

One participant in this study did not follow the 3-stage process. This person encourages end users to hover over all links in emails (like most phishing advice does). She conscientiously follows this advice herself, and hovers over all links in every email she receives before she actually reads the email. This also appears to be an effective strategy for identifying phishing emails, though only one of the 21 participants followed this process.

In this paper, I show that human methods of detecting phishing use very different information than most technical methods of detecting phishing. Humans extensively use *contextual* knowledge of how the email is related to other parts of their life. They also use information about *typical* behaviors of people and organizations that they interact with. Thus, human detection complements technical detection as it is likely to have very different gaps and vulnerabilities, and using the two methods in tandem should result in greater security than either would alone.

I described a way that end users can integrate phishing detection into their normal, everyday email processes in a realistic and reasonable manner. Current phishing training, with its focus on conclusive distinguishers like URLs, could benefit from additionally training people to recognize discrepancies in emails and to more easily trigger to identify phishing as a potential explanation for those discrepancies. Future work should examine which types of failure mostly commonly lead to problems and compromise, to help focus training on the weakest parts of the process.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant No. 1714126. I greatly thank Molly Cooper for helping to recruit and coordinate the interviews. Connor O'Rourke and Caitlin Myers assisted with the coding of cues. I thank the MSU BITLab for feedback on early versions of this paper.

REFERENCES

[1] Ammar Almomani, B. B. Gupta, Samer Atawneh, A. Meulenberg, and Eman Almomani. 2013. A Survey of Phishing Email Filtering Techniques. *IEEE Communications Surveys & Tutorials* 15, 4 (2013), 2070–2090. https://doi.org/10.1109/SURV.2013.030713.00020

[2] Ankesh Anand, Kshitij Gorde, Joel Ruben Antony Moniz, Noseong Park, Tanmoy Chakraborty, and Bei-Tseng Chu. 2018. Phishing URL Detection with Oversampling based on Text Generative Adversarial Networks. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 1168–1177. https://doi.org/10.1109/BigData.2018.8622547

[3] Nalin Asanka Gamagedara Arachchilage and Steve Love. 2013. A game design framework for avoiding phishing attacks. *Computers in Human Behavior* 29, 3 (May 2013), 706–714. https://doi.org/10.1016/j.chb.2012.12.018

[4] Alejandro Correa Bahnsen, Eduardo Contreras Bohorquez, Sergio Villegas, Javier Vargas, and Fabio A. Gonzalez. 2017. Classifying phishing URLs using recurrent neural networks. In *2017 APWG Symposium on Electronic Crime Research (eCrime)*. IEEE, 1–8. https://doi.org/10.1109/ECRIME.2017.7945048

[5] Zinaida Benenson, Freya Gassmann, and Robert Landwirth. 2017. Unpacking Spear Phishing Susceptibility. In *FC 2017: Financial Cryptography and Data Security*. Vol. 10323 LNCS. 610–627. https://doi.org/10.1007/978-3-319-70278-0_39

[6] Mark Blythe, Helen Petrie, and John A Clark. 2011. F for Fake: Four Studies on How We Fall for Phish. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, New York, USA, 3469–3478. https://doi.org/10.1145/1978942.1979459

[7] Deanna D Caputo, Shari Lawrence Pfleeger, Joshua D Freeman, and M Eric Johnson. 2014. Going Spear Phishing: Exploring Embedded Training and Awareness. *Security & Privacy, IEEE* 12, 1 (Jan. 2014), 28–38.

[8] Robert Cialdini. 2009. *Influence: The Psychology of Persuasion* (revised ed.). HarperCollins.

[9] Vincent C. Conzola and Michael S. Wogalter. 2001. A Communication–Human Information Processing (C–HIP) approach to warning effectiveness in the workplace. *Journal of Risk Research* 4, 4 (2001), 309–322. https://doi.org/10.1080/13669870110062712 arXiv:https://doi.org/10.1080/13669870110062712

[10] David A. Cowan. 1986. Developing a Process Model of Problem Recognition. *Academy of Management Review* 11, 4 (Oct 1986), 763–776. https://doi.org/10.5465/amr.1986.4283930

[11] Beth Crandall, Gary Klein, and Robert Hoffman. 2006. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*. A Bradford Book. 332 pages.

[12] Lorrie Faith Cranor. 2008. A Framework for Reasoning About the Human in the Loop.. In *Usability, Psychology, and Security (UPSEC)*. https://www.usenix.org/legacy/event/upsec08/tech/full_papers/cranor/cranor.pdf

[13] Rachna Dhamija, J. D. Tygar, and Marti Hearst. 2006. Why Phishing Works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 581–590. https://doi.org/10.1145/1124772.1124861

[14] Serge Egelman, Lorrie Cranor, Jason Hong, and Yue Zhang. 2007. Phinding Phish : Evaluating Anti-Phishing Tools Phinding Phish : Evaluating Anti-Phishing Tools. In *Network and Distributed System Security*. San Diego, CA.

[15] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 2893–2902. https://doi.org/10.1145/2702123.2702442

[16] Gil Friedrich. 2018. URL Defense Link Rewrites: The Good, The Bad, and The Over-Promised. https://www.avanan.com/resources/url-defense-link-rewrites

[17] Bill Gardner and Valerie Thomas. 2014. *Building an Information Security Awareness Program: Defending Against Social Engineering and Technical Threats* (1st ed.). Syngress. 214 pages.

[18] Erving Goffman. 1974. *Frame Analysis: An Essay on the Organizatino of Experience*. Harper and Row.

[19] The Radicati Group. 2019. *Email Statistics Report 2019-2023 Executive Summary*. Technical Report. The Radicati Group.

[20] Ryan Heartfield, George Loukas, and Diane Gan. 2017. An eye for deception: A case study in utilizing the human-as-a-security-sensor paradigm to detect zero-day semantic social engineering attacks. In *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*. IEEE, 371–378. https://doi.org/10.1109/SERA.2017.7965754

[21] Robert R. Hoffman. 1998. How Can Expertise be Defined? Implications of Research from Cognitive Psychology. In *Exploring Expertise*. Palgrave Macmillan UK, London, 81–100. https://doi.org/10.1007/978-1-349-13693-3_4

[22] Robert R Hoffman, Beth Crandall, and Nigel Shadbolt. 1998. Use of the Critical Decision Method to Elicit Expert Knowledge: A Case Study in the Methodology of Cognitive Task Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 40, 2 (jun 1998), 254–276. https://doi.org/10.1518/001872098779480442

[23] Jason Hong. 2012. The state of phishing attacks. *Commun. ACM* 55, 1 (jan 2012), 74. https://doi.org/10.1145/2063176.2063197

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

[25] Gary Klein, Rebecca Pliske, Beth Crandall, and David D Woods. 2005. Problem detection. *Cognition, Technology & Work* 7, 1 (mar 2005), 14–28. https://doi.org/10.1007/s10111-004-0166-y

[26] Gary A Klein, Roberta Calderwood, and Anne Clinton-Cirocco. 1986. Rapid Decision Making on the Fire Ground. *Proceedings of the Human Factors Society Annual Meeting* 30, 6 (Sep 1986), 576–580. https://doi.org/10.1177/154193128603000616

[27] G A Klein, R Calderwood, and D MacGregor. 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics* 19, 3 (Jan 1989), 462–472. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=31053

[28] Gary A Klein and Robert R Hoffman. 1992. Seeing The Invisible: Perceptual–Cognitive Aspects of Expertise. In *Cognitive Science Foundations of Instruction*, M Rabinowitz (Ed.). Erlbaum, 203—-226. http://cmapspublic3.ihmc.us/rid=1G9NSY15K-N7MJMZ-LC5/SeeingTheInvisible.pdf

[29] Gary A Klein, Jennifer K Phillips, Erica L Rall, and Deborah A Peluso. 2007. A Data-Frame Theory of Sensemaking. In *Expertise Out of Context: The Sixth International Conference on Naturalistic Decision Making*, Robert R Hoffman (Ed.). Lawrence Erlbaum Associates, Inc., 13–155.

[30] Ponnurangam Kumaraguru, Steve Sheng, Alessandro Acquisti, Lorrie Faith Cranor, and Jason Hong. 2010. Teaching Johnny not to fall for phish. *ACM Transactions on Internet Technology* 10, 2 (May 2010), 1–31. https://doi.org/10.1145/1754393.1754396

[31] Tian Lin, Daniel E. Capecci, Donovan M. Ellis, Harold A. Rocha, Sandeep Dommaraju, Daniela S. Oliveira, and Natalie C. Ebner. 2019. Susceptibility to Spear-Phishing Emails: Effects of Internet User Demographics and Email Content. *ACM Trans. Comput.-Hum. Interact.* 26, 5, Article 32 (July 2019), 28 pages. https://doi.org/10.1145/3336141

[32] Eric Lipton, David E Sanger, and Scott Shane. 2016. The Perfect Weapon: How Russian Cyberpower Invaded the U.S. *The New York Times* (dec 2016). https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html

[33] MacEwan University. 2017. University Discovers Online Fraud. https://www.macewan.ca/wcm/MacEwanNews/PHISHING_ATTACK

[34] Matthew B Miles, A. Michael Huberman, and Johnny Saldaña. 2013. *Qualitative Data Analysis: A Methods Sourcebook* (third ed.). Sage Publications.

[35] Daniela Oliveira, Harold Rocha, Huizi Yang, Donovan Ellis, Sandeep Dommaraju, Melis Muradoglu, Devon Weir, Adam Soliman, Tian Lin, and Natalie Ebner. 2017. Dissecting Spear Phishing Emails for Older vs Young Adults: On the Interplay of Weapons of Influence and Life Domains in Predicting Susceptibility to Phishing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 6412–6424. https://doi.org/10.1145/3025453.3025831

[36] Anthony J Onwuegbuzie and Nancy L Leech. 2007. Validity and Qualitative Research: An Oxymoron? *Quality and Quantity* 41 (2007), 233–249.

[37] Will Oremus. 2016. "Is This Something That's Going to Haunt Me the Rest of My Life?". *Slate* (Dec 2016). https://slate.com/technology/2016/12/an-interview-with-charles-delavan-the-it-guy-whose-typo-led-to-the-podesta-email-hack.html

[38] Justin Petelka, Yixin Zou, and Florian Schaub. 2019. 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*. ACM Press, New York, New York, USA, 1–15. https://doi.org/10.1145/3290605.3300748

[39] Denise F. Polit and Cheryl Tatano Beck. 2010. Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies* 47, 11 (Nov 2010), 1451–1458. https://doi.org/10.1016/j.ijnurstu.2010.06.004

[40] Emilee Rader and Rick Wash. 2015. Identifying patterns in informal sources of security information. *Journal of Cybersecurity* 1 (Dec 2015), tyv008. https://doi.org/10.1093/cybsec/tyv008

[41] Karol G Ross, Gary A Klein, Peter Thunholm, John F Schmitt, and Holly C Baxter. 2004. The Recognition-Primed Decision Model. *Military Review* (Aug 2004). http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA521492

[42] Angela Sasse. 2015. Scaring and Bullying People into Security Won't Work. *IEEE Security & Privacy* 13, 3 (May 2015), 80–83. https://doi.org/10.1109/MSP.2015.65

[43] Bruce Schneier. 2000. Semantic Attacks: The Third Wave of Network Attacks. https://www.schneier.com/crypto-gram/archives/2000/1015.html{#}1

[44] Steve Sheng, Mandy Holbrook, Ponnurangam Kumaraguru, Lorrie Faith Cranor, and Julie Downs. 2010. 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* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 373–382. https://doi.org/10.1145/1753326.1753383

[45] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. 2007. Anti-Phishing Phil. In *Proceedings of the 3rd symposium on Usable privacy and security - SOUPS*

'07. ACM Press, New York, New York, USA, 88. https://doi.org/10.1145/1280680.1280692

[46] Sami Smadi, Nauman Aslam, and Li Zhang. 2018. Detection of online phishing email using dynamic evolving neural network based on reinforcement learning. *Decision Support Systems* 107 (2018), 88–102. https://doi.org/10.1016/j.dss.2018.01.001

[47] Rebecca Smith. 2016. How a U.S. Utility Got Hacked. *Wall Street Journal* (Dec 2016). https://www.wsj.com/articles/how-a-u-s-utility-got-hacked-1483120856

[48] Symantec. 2019. *Internet Security Threat Report*. Technical Report February. https://doi.org/10.1016/S1353-4858(05)00194-7

[49] Mary Theofanos, Brian Stanton, Susanne Furman, Sandra Spickard Prettyman, and Simson Garfinkel. 2017. Be Prepared: How US Government Experts Think About Cybersecurity. In *Workshop on Usable Security (USec)*. Internet Society.

[50] Verizon. 2019. *2019 Data Breach Investigations Report*. Technical Report. https://doi.org/10.1016/s1361-3723(19)30060-0

[51] Rick Wash and Molly M. Cooper. 2018. Who Provides Phishing Training?. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–12. https://doi.org/10.1145/3173574.3174066

[52] Karl E. Weick. 1995. *Sensemaking in Organizations*. Sage Publications. 248 pages.

[53] Josephine Wolff. 2018. *You'll See This Message When It Is Too Late: The Legal and Economic Aftermath of Cybersecurity Breaches*. MIT Press, Cambridge, MA, USA. https://mitpress.mit.edu/books/youll-see-message-when-it-too-late

[54] Weiwei Zhuang, Qingshan Jiang, and Tengke Xiong. 2012. An Intelligent Anti-phishing Strategy Model for Phishing Website Detection. In *2012 32nd International Conference on Distributed Computing Systems Workshops*. IEEE, 51–56. https://doi.org/10.1109/ICDCSW.2012.66

Received January 2020; revised June 2020; accepted July 2020