Cue Utilization, Phishing Feature and Phishing Email Detection

Piers Bayl-Smith[0000-0001-8014-0633], Daniel Sturman[0000-0002-5025-598X], and Mark Wiggins[0000-0002-6422-9475]

Macquarie University, New South Wales, Australia

Abstract. Cognitive processes are broadly considered to be of vital importance to understanding phishing email feature detection or misidentification. This research extends the current literature by introducing the concept of cue utilization as a unique predictor of phishing feature detection. First year psychology students (*n=127)* undertook three tasks measuring cue utilization, phishing feature detection and phishing email detection. A multiple linear regression model provided evidence that those in a higher cue utilization typology (*n*=55) performed better at identifying phishing features than those in a lower cue utilization typology (*n*=72). Furthermore, as predicted by the Elaboration Likelihood Model (ELM) and Heuristic-Systematic Model (HSM), those who deliberated longer per email demonstrated an increased ability to correctly identify phishing features. However, these results did not translate into improved performance in the phishing email detection task. Possible explanations for these results are discussed, including possible limitations and areas of future research.

Keywords. Phishing; Cue utilization; Feature identification; Elaboration Likelihood Model; Heuristic-Systematic Model

1 Introduction

1.1 Study Aims

Despite significant investment in cyber security solutions, employees remain the most significant risk to maintaining a protected information environment. Specifically, phishing emails are a major attack vector through which an organization's information security can be compromised. Recent research has suggested that for businesses, 74% of all cyber threats originate via email sources [1], whereas in Australia, phishing was the top registered scam category reported to the Australian Competition and Consumer Commission [2]. Costs to businesses and individuals have steadily been on the rise at a global level, occasioning in business disruption, information and intellectual property loss, and revenue loss, with damages reported in the hundreds of millions of dollars [2, 3].

Given the importance of human factors and phishing, this study investigates what cognitive factors may influence phishing detection. In particular, whether

cognitive processing impact an individual's ability to detect suspicious features characteristic of a phishing email, as well as their ability to correctly distinguish between legitimate and non-legitimate email communications. Unique to this study, we examine the role of cue-based processing when assessing phishing emails.

1.2 Phishing Features and Cognitive Processing

Phishing emails are deceptive forms of communication that endeavor to appear legitimate, but are in fact attempts to obtain personal credentials or sensitive information [4]. By promising some form of false utility, they aim to convince targets to either reply, click on an embedded URL link, or download an attachment. Researchers have identified common features inherent in phishing emails that may be used by recipients to identify that an email is malicious [5–7]. These can include poor visual presentation, spelling and grammatical errors, a sender's address that does not match the expected domain name, and questionable URL links. By identifying such features, individuals can increase their likelihood of recognizing an email as a phishing attempt, and can take appropriate protective actions [4, 8]. In contrast, when such features are either misidentified or neglected, the recipient is at increased risk of complying to the phishing email. In support of this contention, several studies, both qualitative and quantitative, have demonstrated a link between the identification of suspicious features and phishing detection performance [e.g., 4, 9, 10].

Avoiding phishing scams however does not merely require an ability to identify the relevant features of an email, but also relies on applying appropriate levels of cognitive resources to assess the entirety of an email message [4]. To encourage limited processing, phishing emails will often resort to social engineering techniques, such as appeals to authority or urgency [6, 11]. By utilizing these forms of persuasion, an email recipient may be inspired to respond quickly and without deliberation – only later realizing that they have become a victim to a phishing scam [12]. Researchers have also posited that individual factors may lead to less cognitive processing of emails including personality [13], habitual use of emails [14], threat perceptions [15] and self-efficacy [16].

From a theoretical perspective, phishing research has utilized dual-process cognitive models that differentiate between easy and rapid processes from those that are more effortful, time intensive and deliberate [17–20]. Utilizing theoretical frameworks such as the Heuristic-Systematic Model (HSM)[21] and Elaboration Likelihood Model (ELM)[22], researchers have proposed that recipients fail to identify phishing emails due to quick-fire heuristic processes rather than being systematically attentive to the content of the message [14]. That is, rather than examining the broad range of features within an email, such as sender's address, URL hyperlinks and formatting issues, people respond to emails only with a cursory appraisal. Social engineering strategies, individual factors and context can play an important role in whether people are motivated to engage in more systematic or elaborative processes when analyzing an email. However, utilization of more deliberative processes, whilst effortful and taking longer, will improve detection of the salient features of an email that indicate it to be suspicious [14, 23].

Previous phishing research has typically measured systematic or elaborative processing via survey, asked after exposure to a phishing email [4, 17, 19]. Results have generally been supportive of the notion that deeper processing of emails leads to more accurate levels of phishing detection. However, given the debate on whether cognitive processes can be explicitly understood and reported [24], survey items may not provide an accurate gauge of systematic or elaborative cognitive processing. Harrison et al. [23] utilized an alternative method to assess elaboration, where participants were asked an open ended question on why they did or did not respond to a phishing email. Here, word count was used as an indicator of cognitive processing. Although this method overcomes the potential shortcomings of survey questions, this measure is indirect, open to possible confounds and occurs post-exposure to the phishing email. In contrast to these approaches, by using evaluation time, this study provides a more direct indication of cognitive processing whilst the participant is assessing an email. Therefore, in line with these considerations, we hypothesize the following:

Hypothesis 1: The more time taken to assess an email will be associated with an increased correct identification of suspicious features characteristic of a phishing email.

Hypothesis 2: The more time taken to assess an email will be associated with an increased detection of phishing emails (hit rate) and lower incidence of incorrectly identifying a genuine email (false alarm).

Although there is much to commend using the existing theoretical paradigm, we will also investigate an alternative theoretical framework drawing from research on cue utilization and expertise [25].

1.3 Cue Utilization

Cues comprise associations between situation specific environmental features and task-related events or objects. Through repeated exposure to feature-event relationships, cues are acquired and retained in long-term memory. These cues can later be activated rapidly and non-consciously when key features are identified, enabling fast and accurate responses [26–28]. As cues are acquired through repeated exposure, cue-based processing is often associated with expert performance [29, 30]

Experts appear to have the facility of being able to innately identify an appropriate response to a problem based on limited information [31, 32]. They are also faster at generating situation assessments with greater accuracy than novices [33, 34]. According to the Recognition-Primed Decision (RPD) model [29], experienced decision makers base their judgements upon satisficing rather than a deliberate analytical assessment of all available options. Through the acquisition of cues based in memory, experts recognize patterns in the situation, what to expect, what further cues need to be attended to, and what goals need to be realized [35]. These patterns in turn actuate scripts that are then implemented into action. In contrast, non-experts or those who are inexperienced, are unlikely to know which cues to attend to, and do not know how to utilize cues in a meaningful way.

Cue utilization refers to an individual difference in the capacity to acquire, recognize and apply cues [36, 37]. Effective cue utilization allows individuals to attend to features of greater relevance, reducing the overall number of features to which they attend, and thereby increasing speed and performance [38–41]. To measure cue utilization, this study uses the online assessment tool, EXPERT Intensive Skills Evaluation (EXPERTise 2.0) [42]. This tool is designed to assess behaviors that are consistent with the utilization of cues, distinguishing relative participant performance in the operation of cues. EXPERTise 2.0 has demonstrated construct validity [43, 44], predictive validity [45], and test-retest reliability [46]. In the context of phishing, individuals with relatively higher cue utilization, as measured by EXPERTise 2.0, should be able to more rapidly identify features which are indicative of phishing emails, in turn enabling more accurate classifications of phishing emails. Therefore we hypothesize:

Hypothesis 3: Participants in the higher cue utilization typology, as determined by performance in EXPERTise 2.0, will be associated with an increased identification of suspicious features characteristic of a phishing email, compared to participants in the lower cue utilization typology.

Hypothesis 4: Participants in the higher cue utilization typology, as determined by performance in EXPERTise 2.0, will be associated with an increased detection of phishing emails (hit rate) and lower incidence of incorrectly identifying a genuine email (false alarm), compared to participants in the lower cue utilization typology.

2 Method

2.1 Participants

Students enrolled in a first-year psychology program at Macquarie University, Australia, were invited to participate in an online study investigating the impact of cue utilization on phishing detection tasks. In total, 127 students participated in the study. Of these, 65.4% were female with an average age of 22.7 years (SD = 8.3 years), ranging from 18 to 54 years. The majority (81.9%) were in some form of paid employment, with 12.5% having managerial responsibilities. Only 14 respondents (11.0%) had received formal cyber security training by their organization. Students who completed the study were provided with course credit. The ethical considerations of this study were reviewed and approved by the Macquarie University Human Research Ethics Committee.

2.2 Materials

Expertise 2.0 – Phishing edition
The present study employed the phishing edition of EXPERTise 2.0, which comprises a battery of four tasks: The Feature Identification Task (FIT); the Feature Recognition

Task (FRT); the Feature Association Task (FAT), and; the Feature Discrimination Task (FDT).

During the FIT, participants are required to identify, as quickly as possible, key features of concern in a series of domain related stimuli. In the phishing edition of EXPERTise, participants were presented with 10 scenarios, each consisting of a single phishing email. For each email that was presented, participants were required to click on the area of the email that aroused the most suspicion, or to click on an icon titled "Trustworthy Email". For this task, response latency for each scenario was recorded. Higher cue utilization is generally associated with a lower mean response latency [25, 47].

In the FRT, participants are presented with domain related stimuli for short periods, and then required to categorize the stimuli. The phishing edition of EXPERTise consists of 20 email stimuli, 10 which contain a genuine email and 10 which contain a phishing email. Each email is presented for only 1000ms, after which participants are asked to classify the email as "Trustworthy", "Untrustworthy", or "Impossible to tell". The FRT assesses the capacity to rapidly extract key information, therefore the short display time was chosen to reflect the nature of this task. Higher cue utilization is typically associated with a greater number of correct classifications [48].

For the FAT, participants are presented with two phrases used in a given domain and are required to rate the perceived relatedness of each phrase. In the phishing edition of EXPERTise participants are presented with 14 pairs of phrases which are related to the computing environment and phishing (e.g., 'Email', 'Malware'). Each pair of stimuli are presented side by side for 2000ms, after which participants were asked to rate the perceived relatedness of the words on a scale ranging from 1 (Extremely unrelated) to 6 (Extremely related). For the FAT, higher cue utilization is typically associated with a greater mean variance in ratings, being selected within a shorter period of time [49].

In the FDT, participants are presented with the details of a problem-oriented scenario and are required to select an appropriate response. Participants are then provided with a list of features from the scenario and are asked to rate the perceived importance of each feature in determining their chosen response, ranging from 1 (Not important at all) to 10 (Extremely important). The phishing edition of EXPERTise consists of a spear phishing email, claiming that an unpaid invoice is going to result in legal and financial costs. The 11 features being rated contain factors related to the work environment and email (e.g., "your bosses anger", "the senders email address"). Higher cue utilization is typically associated with a greater variance in feature ratings [41, 50].

Phishing Feature Task

The phishing feature task was setup on Qualtrics [51], an online survey platform. This task involved participants viewing a genuine email that had one of four features manipulated by the researchers; the senders email address with an incongruent domain name, introduction of poor spelling, insertion of a URL with an incongruent domain name, or changes to the look of the email to make it appear more rudimentary (see Appendix for an example stimuli). Participants were informed that each image was legitimate but had one feature changed. Each email image was displayed for a maximum of 20 seconds,

but they could proceed forward at any time. After the image was displayed, respondents were asked which feature most aroused their suspicion. Nine features of an email were then listed as options, including a tenth option of "I don't know". The time spent evaluating each email and the feature selected was collected for each participant.

Phishing Detection Task

The phishing detection task was also setup within Qualtrics [51]. In this task, an image of an email was displayed for a maximum of 20 seconds, after which participants were asked to judge whether the email was trustworthy or suspicious. Respondents were able to move forward to the question of trustworthiness before the 20 seconds had elapsed. All emails were either genuine or examples of real phishing attempts that had been collected overtime by the researchers (see Appendix for an example stimuli). In total, there were ten emails that needed to be assessed by each respondent; five genuine and five phish. The time each email was attended to and trustworthiness responses were collected for each participant.

Demographic and Cyber-security expertise

To control for possible confounds, we have included age, gender and self-reported cyber-security expertise. Previous research has suggested younger adults (18-25 years) are more susceptible to phishing attacks [52]. This may be associated with less exposure to phishing emails, lower internet use across one's lifetime, lack of cyber education, or the use of specific types of attack strategies within a phishing email [52–54]. Gender has also been identified as being an important consideration when examining phishing susceptibility, with females potentially at more risk of responding to phishing emails [52, 55]. This may be explained by differences in personality, self-efficacy and lack of technical training [16, 52, 55]. However, it should be noted that the effects for gender have not been consistently found significant across all research [56]. Cyber-security proficiency was assessed by a single item, "What is your proficiency in cyber security." For this study, the five-point Likert-type response has been converted into a categorical variable, where the options "Very Good" and "Expert" have been labelled "High proficiency", and "None", "Poor" and "Average" are categorized as "Low proficiency". Overall, 27.6% of participants considered cyber security proficiency to be high. Self-efficacy and cyber security knowledge has been implicated as a protective factor against phishing attacks [16].

2.3 Procedure

Participants were sent a link to the Qualtrics survey platform where they were first asked a series of demographic questions along with items pertaining to their cyber security history and knowledge. Respondents then completed the phishing detection task and phishing feature task before being automatically forwarded to the EXPERTise 2.0 platform. Within EXPERTise 2.0, participants completed the four tasks associated with cue-utilization (FIT, FRT, FAT, FDT), with detailed instructions being provided for each task. The total time to complete the online study was approximately 30 minutes.

3 Results

3.1 Cue utilization typologies

Using a standard approach for classifying participants into cue utilization typologies [57], scores for each EXPERTISE task were converted to standardized *z*-scores, and a cluster analysis was performed to identify two cue utilization typologies. Fifty-five participants were classified as having relatively higher cue utilization and 72 participants were classified as having relatively lower cue utilization. The higher cue utilization typology consisted of participants with shorter mean response latencies on the FIT, greater mean accuracy on the FRT, a higher mean ratio of variance to reaction time on the FAT, and a greater mean variance in ratings on the FDT. There were significant differences in FIT, FAT, FRT, and FDT mean scores between the higher and lower cue utilization typologies (see Table 1). Additional clustering solutions were examined post-hoc but were not found to be suitable due to low participant numbers in additional clusters (n<5).

Table 1. Raw and standardized EXPERTise task scores by cue utilization typology

	Higher cue utilization (n=55)			Lower cue utilization (n=72)			
	Mean	*SD*	*z*-score	Mean	*SD*	*z*-score	*t*
FIT	3749	2708	-0.35	9423	5895	0.27	-3.58**
FRT	11.0	2.62	0.44	10.25	2.94	-0.33	4.64**
FAT	1.60	0.88	0.50	0.87	0.60	-0.39	5.53**
FDT	10.0	3.65	0.77	3.93	2.59	-0.59	10.33**

* Significant at the 0.05 level (two-tailed); **Significant at the 0.01 level (two-tailed)

3.2 Performance on the Phishing Feature Task

Across the 16 emails used in the phishing feature task, participants were on average able to detect the suspicious feature 6.5 times. A multiple linear regression was conducted to determine the effects of cue utilization typology and average email deliberation time upon phishing feature task performance. Age, gender and subjective cyber security proficiency were included as control variables. A summary of results is displayed in Table 2. Overall, the combined predictors had a significant effect in explaining the variance of feature detection performance (r^2 = .21, F(5,126) = 6.40, p < .01). In support of Hypothesis 1, the mean review time for each email was associated with a significant positive effect, such that on average for every 4.4 seconds of additional viewing time (1 *SD* increase in time), an additional 1.12 features were correctly detected. Hypothesis 3 was also supported, where those in the high cue utilization typology were significantly more likely to detect an additional 1.42 features than those grouped in the low cue utilization typology. No significant effects for age, gender or subjective cyber security proficiency were found.

Table 2. Multiple Linear Regression for the Phsihing Feature Task

	DV = Phishing Feature Detection Performance			
Predictor	*b*	*SE*	β	*t*
Age	-0.04	0.03	-0.11	-1.25
Gender	-0.25	0.45	-0.46	-0.56
Cyber proficiency	0.56	0.48	0.10	1.16
Phishing feature review time	0.25	0.52	0.43	4.58**
Cue utilization typology	1.42	0.45	0.27	3.17**

** Significant at the 0.01 level (two-tailed)

3.3 Performance on the Phishing Detection Task

Performance on the phishing detection task indicated that on average, 3.2 emails were correctly identified as phish (hit rate), ranging from 0 correct to 5 correct. For emails that were genuine, on average 1.1 emails were incorrectly identified as being suspicious (false alarm), ranging from 0 false alarms to 4 false alarms. Calculating for potential bias [58], no significant differences were found contingent upon cue utilization typology.

To examine the effects of cue utilization typology and average email deliberation time upon phishing detection performance, two separate multiple linear regression models were tested. The first model included phishing detection hit rate as the dependent variable, with average email review time, cue utilization, age, gender and subjective cyber security proficiency included as predictor variables. Results indicated that the predictor variables accounted for 10.3% of total variance ($F(5,126) = 2.79$, $p = .02$), with only gender significantly related to phishing detection hit-rate ($\beta = 0.19$, $p = .03$), where on average, males were more likely to correctly identify phishing emails when compared to females. The second model utilized the same independent variables, but included phishing detection false alarms as the variable of interest. Overall, this model was not significant ($F(5,126) = 1.94$, $p = .09$), with no predictors demonstrating a significant relationship with the dependent variable. Therefore, both models lack evidence to support Hypotheses 2 and 4.

4 Discussion

This study examined the influence of processing time and cue utilization upon the identification of suspicious phishing features and phishing detection. Overall, the results suggest that both the time processing emails and a high cue utilization typology have a positive impact upon being able to perceive features that may indicate that an email is suspicious. However, these factors did not translate into an enhanced proficiency to discriminate phishing emails nor a lower incidence of incorrectly identifying a phishing email.

According to dual-process cognitive theories such as HSM [21] and ELM [22], increased systematic processing or elaboration of an incoming communication should improve the detection of suspicious features that may identify the message as fraudulent [14]. The current study provides additional support for this contention. Participants who on average assessed email images for longer periods of time, demonstrated a greater ability to identify the suspicious feature that had been changed in an otherwise genuine email. A similar result was found for cue utilization. That is, those in the high cue utilization typology exhibited an improved ability in detecting suspicious features within an email. This supports the notion that those with higher cue utilization are more able to generate accurate situational assessments, thereby identifying features that do not fit with the expected patterns of the situation. Practically, these results suggest that users do not approach emails with similar cognitive processes or capabilities. According to Downs et al. [8], phishing succeeds when attackers are able to manipulate recipients into generating inaccurate mental models. Therefore, it is incumbent upon organizations to adequately train users on phishing feature identification to minimize differences in cognitive processing and cue-utilization. Furthermore, email clients should allow individuals to easily identify features that indicate that an email may be suspicious, thereby maximize the opportunity to create accurate metal models.

Although longer deliberation time and higher cue utilization was associated with increased ability to identify suspicious features in an email, this did not translate into improved phishing detection or lower rates of genuine email misidentification. This supports the contention made by Vishwanath et al. [4] that phishing is more than just the ability to detect phishing features. Research has indicated that phishing detection can be influenced by a large variety of factors, including personality, threat perceptions and self-efficacy [13–16]. As an alternative explanation, there may be some methodological considerations that may account for the null results. First, with only five phishing emails and five genuine emails, there may not have been enough variability in the task. Even if participants had chosen randomly, on average they would have correctly identified 2.5 emails correctly. Future research should include a larger section of phishing emails, including the possibility of changing the base-rate of phishing to genuine emails. Second, the task may have been too easy. The phishing examples used in this study were not overly sophisticated, nor were they personally addressed to the participant. That is, they may contain multiple features that indicate their fraudulent nature and therefore be too easy to detect. Furthermore, any persuasion strategies used by the phishing emails will be mollified by not being of personal import to the participant (e.g., they were under no threat or open to the utility being offered). Future research should then try to increase the fidelity of the experiment by simulating the work environment or use actual phishing simulations upon employees naïve to the study.

Of some interest, phishing detection was significantly related to gender. This supports previous research that suggests females may be more vulnerable to phishing than males [52, 55]. However, no such effect was not found with the phishing feature detection task. Explanations for our results then must be explained by factors not examined in this study, including self-efficacy, differences in personality or less online experience or knowledge [13, 52].

4.1 Limitations and Future Research

Apart from the limitations noted in the previous section as they relate to the phishing detection task, as an introductory study examining the role of cue utilization in phishing detection, this study has several limitations which also provide future avenues for further research. First, cue utilization as measured by EXPERTise 2.0 only gauges individual differences in the ability to detect and discriminate domain relevant features. What aspects of an email communication are actually being examined when discriminating between genuine from phishing emails was not able to be determined using EXPERTise 2.0. Future research should engage eye-tracking technologies to determine which features are attended to when reviewing an email message. This, in conjunction with cue utilization performance, should provide a more comprehensive understanding of what security features of an email are important and which features may need to be highlighted to ensure a secure email environment.

Second, this study had participants assess images of phishing emails rather than respond to actual emails. Therefore, the results may not be applicable in actual work or personal settings. The images used were displayed within an email application shell, but they were not able to be operated as a real email message would (e.g., hovering over hyperlinks to reveal additional information). Future studies would benefit from the use of more sophisticated simulations allowing researchers to draw more meaningful real-world inferences.

Third, drawing from a sample of first year psychology students, the sample for this study was broadly homogenous. In organizations and personal contexts, a wide range of people of different ages, experiences, knowledge and network privileges have access to email. Therefore, future research needs to continue to investigate additional individual and contextual factors to understand why individuals may fall for phishing scams. This study suggests that cue utilization may be a key feature, although more research is needed with a broader demographic sample.

4.2 Conclusion

Phishing scams are on the rise globally, costing millions in damages. This study again reinforced the notion that more deliberative or systematic processing of incoming communications reduce the risk posed by phishing scams. Furthermore, this was the first study to investigate the potential role of cue utilization in phishing feature and email detection. It was found that a more deliberative processing of emails and higher cue utilization resulted in an improved ability to detect suspicious features in an otherwise genuine email but did not necessarily improve overall phishing detection. Linking these cognitive processes to phishing detection may provide additional capacities to understanding the threat posed by phishing, and thereby improve possible protective interventions, usability initiatives and training programs.

5 References

1. Neely, L.: 2017 Threat Landscape Survey: Users on the Front Line. (2017).
2. Australian Competition & Consumer Commission: Targeting scams: Report of the ACCC on scams activity 2018. Canberra (2019).
3. Bissell, K., LaSalle, R., Dal Cin, P.: The cost of cybercrime. (2019).
4. Vishwanath, A., Herath, T., Chen, R., Wang, J., Rao, H.R.: Why do people get phished? Testing individual differences in phishing vulnerability within an integrated, information processing model. Decis. Support Syst. 51, 576–586 (2011). https://doi.org/10.1016/j.dss.2011.03.002.
5. Parsons, K., Butavicius, M., Pattinson, M., Calic, D., Mccormac, A., Jerram, C.: Do Users Focus on the Correct Cues to Differentiate Between Phishing and Genuine Emails? In: Australasian Conference on Information Systems. pp. 1–10. , Adelaide (2015).
6. Zielinska, O.A., Welk, A.K., Mayhorn, C.B., Murphy-Hill, E.: A temporal analysis of persuasion principles in phishing emails. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 60, 765–769 (2016). https://doi.org/10.1177/1541931213601175.
7. Furnell, S.: Phishing: can we spot the signs? Comput. Fraud Secur. 2007, 10–15 (2007). https://doi.org/10.1016/S1361-3723(07)70035-0.
8. Downs, J.S., Holbrook, M.B., Cranor, L.F.: Decision strategies and susceptibility to phishing. In: Proceedings of the second symposium on Usable privacy and security - SOUPS '06. p. 79. ACM Press, New York, New York, USA (2006). https://doi.org/10.1145/1143120.1143131.
9. Molinaro, K.A., Bolton, M.L.: Evaluating the applicability of the double system lens model to the analysis of phishing email judgments. Comput. Secur. 77, 128–137 (2018). https://doi.org/10.1016/j.cose.2018.03.012.
10. Williams, E.J., Hinds, J., Joinson, A.N.: Exploring susceptibility to phishing in the workplace. Int. J. Hum. Comput. Stud. 120, 1–13 (2018). https://doi.org/10.1016/j.ijhcs.2018.06.004.
11. Parsons, K., Butavicius, M., Delfabbro, P., Lillie, M.: Predicting susceptibility to social influence in phishing emails. Int. J. Hum. Comput. Stud. 128, 17–26 (2019). https://doi.org/10.1016/j.ijhcs.2019.02.007.
12. Hadnagy, C., Fincher, M.: Phishing dark waters : The offensive and defensive sides of malicious e-mails. Willey, Indianapolis (2015).
13. Halevi, T., Memon, N., Nov, O.: Spear-Phishing in the Wild: A Real-World Study of Personality, Phishing Self-Efficacy and Vulnerability to Spear-Phishing Attacks. SSRN Electron. J. (2015). https://doi.org/10.2139/ssrn.2544742.
14. Vishwanath, A., Harrison, B., Ng, Y.J.: Suspicion, Cognition, and Automaticity Model of phishing susceptibility. Communic. Res. 1–21 (2016). https://doi.org/10.1177/0093650215627483.
15. Jansen, J., van Schaik, P.: Persuading end users to act cautiously online: a fear appeals study on phishing. Inf. Comput. Secur. 26, 264–276 (2018). https://doi.org/10.1108/ICS-03-2018-0038.
16. Sun, J.C.-Y., Yu, S.-J., Lin, S.S.J., Tseng, S.-S.: The mediating effect of anti-phishing self-efficacy between college students' internet self-efficacy and anti-

phishing behavior and gender difference. Comput. Human Behav. 59, 249–257 (2016). https://doi.org/10.1016/j.chb.2016.02.004.
17. Musuva, P.M.W., Getao, K.W., Chepken, C.K.: A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. Comput. Human Behav. 94, 154–175 (2019). https://doi.org/10.1016/j.chb.2018.12.036.
18. Workman, M.: Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. J. Am. Soc. Inf. Sci. Technol. 59, 662–674 (2008). https://doi.org/10.1002/asi.20779.
19. Vishwanath, A.: Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. J. Comput. Commun. 20, 570–584 (2015). https://doi.org/10.1111/jcc4.12126.
20. Luo, X., Zhang, W., Burd, S., Seazzu, A.: Investigating phishing victimization with the Heuristic–Systematic Model: A theoretical framework and an exploration. Comput. Secur. 38, 28–38 (2013). https://doi.org/10.1016/j.cose.2012.12.003.
21. Chaiken, S.: Heuristic versus systematic information processing and the use of source versus message cues in persuasion. J. Pers. Soc. Psychol. 39, 752–766 (1980). https://doi.org/10.1037/0022-3514.39.5.752.
22. Petty, R.E., Cacioppo, J.T.: The Elaboration Likelihood Model of persuasion. Commun. Persuas. 19, 1–24 (1986). https://doi.org/10.1007/978-1-4612-4964-1_1.
23. Harrison, B., Svetieva, E., Vishwanath, A.: Individual processing of phishing emails. Online Inf. Rev. 40, 265–281 (2016). https://doi.org/10.1108/OIR-04-2015-0106.
24. Nisbett, R.E., Wilson, T.D.: Telling more than we can know: Verbal reports on mental processes. Psychol. Rev. 84, 231–259 (1977). https://doi.org/10.1037/0033-295X.84.3.231.
25. Loveday, T., Wiggins, M.W., Searle, B.J.: Cue utilization and broad indicators of workplace expertise. J. Cogn. Eng. Decis. Mak. 8, 98–113 (2014). https://doi.org/10.1177/1555343413497019.
26. Brunswik, E.: Representative design and probabilistic theory in a functional psychology. Psychol. Rev. 62, 193–217 (1955). https://doi.org/10.1037/h0047470.
27. Ericsson, K.A., Lehmann, A.C.: Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints. Annu. Rev. Psychol. 47, 273–305 (1996). https://doi.org/10.1146/annurev.psych.47.1.273.
28. Salthouse, T.: Expertise as the circumvention of human processing limitations. In: Toward a general theory of expertise: Prospects and limits. pp. 286–300. Cambridge University Press, Cambridge, NY (1991). https://doi.org/10.1037/e578082012-006.
29. Klein, G.A.: A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In: Decision making in action: Models and methods. pp. 139–147 (1993). https://doi.org/10.1002/bdm.3960080307.
30. Anderson, J.R.: Rules of the mind. Lawrence Erlbaum, Hillsdale, NJ (1993).

31. Abernethy, B.: Anticipation in squash: Differences in advance cue utilization between expert and novice players. J. Sports Sci. 8, 17–34 (1990). https://doi.org/10.1080/02640419008732128.
32. De Groot, A.D.: Thought and choice in chess. Mouton, The Hague.
33. Calderwood, R., Klein, G.A., Crandall, B.W.: Time Pressure, Skill, and Move Quality in Chess. Am. J. Psychol. 101, 481 (1988). https://doi.org/10.2307/1423226.
34. Müller, S., Abernethy, B., Farrow, D.: How do World-Class Cricket Batsmen Anticipate a Bowler's Intention? Q. J. Exp. Psychol. 59, 2162–2186 (2006). https://doi.org/10.1080/02643290600576595.
35. Klein, G.A.: The Recognition-Primed Decision (RPD) model: Looking back, looking forward. In: Zsambok, C.E. and Klein, G.A. (eds.) Naturalistic Decision Making. pp. 285–292. Lawrence Erlbaum Associates, Mahwah, NJ (1997).
36. Wiggins, M.W., Loveday, T., Auton, J.C.: EXPERT Intensive Skills Evaluation (EXPERTise) Test. Macquarie University, Sydney (2015).
37. Lansdale, M., Underwood, G., Davies, C.: Something Overlooked? How experts in change detection use visual saliency. Appl. Cogn. Psychol. 24, 213–225 (2010). https://doi.org/10.1002/acp.1552.
38. Brouwers, S., Wiggins, M.W., Helton, W., O'Hare, D., Griffin, B.: Cue utilization and cognitive load in novel task performance. Front. Psychol. 7, 1–12 (2016). https://doi.org/10.3389/fpsyg.2016.00435.
39. Sturman, D., Wiggins, M.W., Auton, J.C., Loft, S.: Cue utilization differentiates resource allocation during sustained attention simulated rail control tasks. J. Exp. Psychol. Appl. (2019). https://doi.org/10.1037/xap0000204.
40. Williams, A.M., Ward, P., Knowles, J.M., Smeeton, N.J.: Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. J. Exp. Psychol. Appl. 8, 259–270 (2002). https://doi.org/10.1037/1076-898X.8.4.259.
41. Weiss, D.J., Shanteau, J.: Empirical Assessment of Expertise. Hum. Factors J. Hum. Factors Ergon. Soc. 45, 104–116 (2003). https://doi.org/10.1518/hfes.45.1.104.27233.
42. EXPERTise 2.0 [Computer Software], https://expertise.mq.edu.au/, (2019).
43. Wiggins, M.W., Azar, D., Hawken, J., Loveday, T., Newman, D.: Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. Saf. Sci. 65, 118–124 (2014). https://doi.org/10.1016/j.ssci.2014.01.006.
44. Small, A.J., Wiggins, M.W., Loveday, T.: Cue-Based Processing Capacity, Cognitive Load and the Completion of Simulated Short-Duration Vigilance Tasks in Power Transmission Control. Appl. Cogn. Psychol. 28, 481–487 (2014). https://doi.org/10.1002/acp.3016.
45. Watkinson, J., Bristow, G., Auton, J., McMahon, C.M., Wiggins, M.W.: Post-graduate training in audiology improves clinicians' audiology-related cue utilisation. Int. J. Audiol. 57, 681–687 (2018). https://doi.org/10.1080/14992027.2018.1476782.
46. Loveday, T., Wiggins, M.W., Festa, M., Schell, D., Twigg, D.: Pattern recognition as an indicator of diagnostic expertise. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36530-0.

47. Schriver, A.T., Morrow, D.G., Wickens, C.D., Talleur, D.A.: Expertise Differences in Attentional Strategies Related to Pilot Decision Making. Hum. Factors J. Hum. Factors Ergon. Soc. 50, 864–878 (2008). https://doi.org/10.1518/001872008X374974.
48. Wiggins, M.W., O'Hare, D.: Expert and Novice Pilot Perceptions of Static In-Flight Images of Weather. Int. J. Aviat. Psychol. 13, 173–187 (2003). https://doi.org/10.1207/S15327108IJAP1302_05.
49. Morrison, B.W., Wiggins, M.W., Bond, N.W., Tyler, M.D.: Measuring Relative Cue Strength as a Means of Validating an Inventory of Expert Offender Profiling Cues. J. Cogn. Eng. Decis. Mak. 7, 211–226 (2013). https://doi.org/10.1177/1555343412459192.
50. Pauley, K., O'Hare, D., Wiggins, M.: Measuring Expertise in Weather-Related Aeronautical Risk Perception: The Validity of the Cochran–Weiss–Shanteau (CWS) Index. Int. J. Aviat. Psychol. 19, 201–216 (2009). https://doi.org/10.1080/10508410902979993.
51. Qualtrics core-XM [Computer Software], https://www.qualtrics.com/au/core-xm/survey-software/, (2019).
52. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L.F., Downs, J.: Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In: Proceedings of the 28th international conference on Human factors in computing systems - CHI '10. p. 373. ACM Press, New York, New York, USA (2010). https://doi.org/10.1145/1753326.1753383.
53. Gavett, B.E., Zhao, R., John, S.E., Bussell, C.A., Roberts, J.R., Yue, C.: Phishing suspiciousness in older and younger adults: The role of executive functioning. PLoS One. 12, (2017). https://doi.org/10.1371/journal.pone.0171620.
54. Oliveira, D., Ebner, N.C., Rocha, H., Yang, H., Ellis, D., Dommaraju, S., Muradoglu, M., Weir, D., Soliman, A., Lin, T.: Dissecting Spear Phishing Emails for Older vs Young Adults. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17. pp. 6412–6424. ACM Press, New York, New York, USA (2017). https://doi.org/10.1145/3025453.3025831.
55. Halevi, T., Lewis, J., Memon, N.: A pilot study of cyber security and privacy related behavior and personality traits. WWW 2013 Companion - Proc. 22nd Int. Conf. World Wide Web. 737–744 (2013). https://doi.org/10.1145/2487788.2488034.
56. Bullee, J., Montoya, L., Junger, M., Hartel, P.: Spear phishing in organisations explained. Inf. Comput. Secur. 25, 593–613 (2017). https://doi.org/10.1108/ICS-03-2017-0009.
57. Wiggins, M.W., Brouwers, S., Davies, J., Loveday, T.: Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. Front. Psychol. 5, 1–8 (2014). https://doi.org/10.3389/fpsyg.2014.00541.
58. Stanislaw, H., Todorov, N.: Calculation of signal detection theory measures. Behav. Res. Methods, Instruments, Comput. 31, 137–149 (1999). https://doi.org/10.3758/BF03207704.

6 Appendix

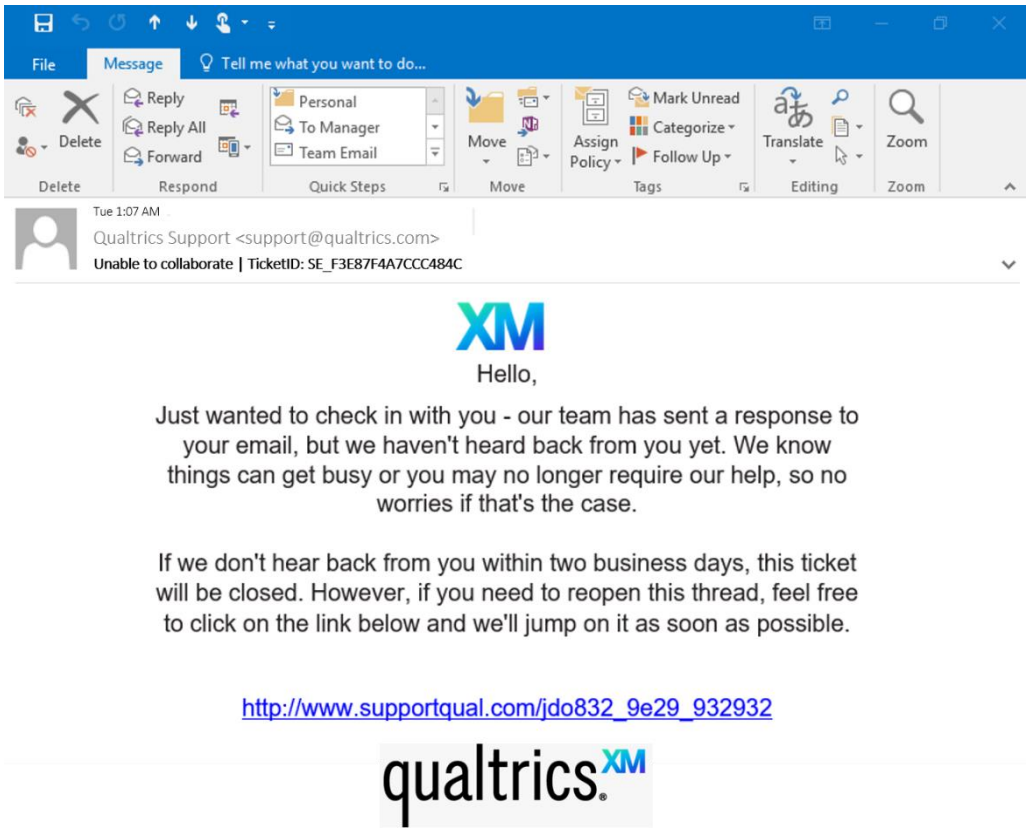

Figure 1: Example stimuli for Phishing Feature Task – URL was changed

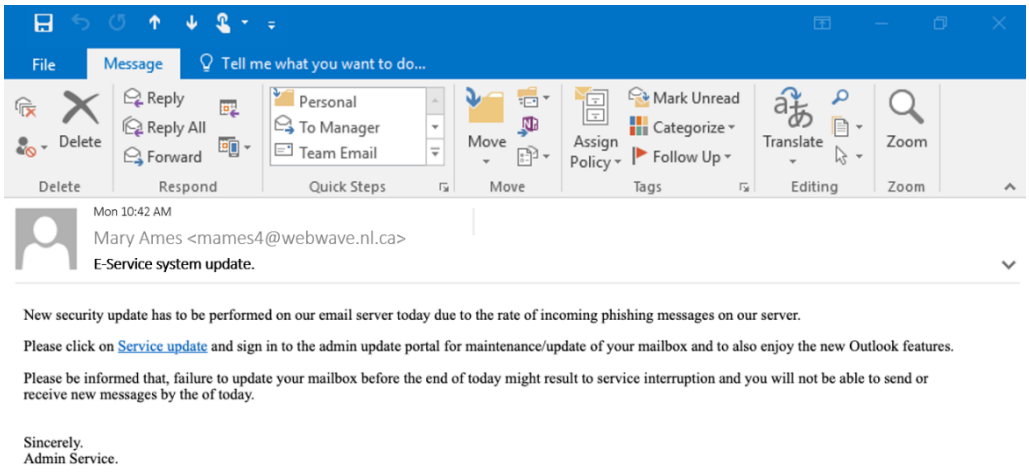

Figure 2: Example stimuli for Phishing Detection Task