Human Factors in Cybersecurity, Vol. 208, 2026, 21–30
https://doi.org/10.54941/ahfe1007409

Eye Tracking Study to Analyze Context Encoding During Phishing Decision Making

Tianhao Xu and Prashanth Rajivan

Industrial and Systems Engineering, Seattle, WA 98195, USA

ABSTRACT

Phishing and spear-phishing remain among the most persistent cybersecurity threats. This study examines end-user decision-making in spear-phishing contexts by modelling the relationship between visual attention and responses using eye-tracking measures. Forty-eight university students completed an email management task while eye movements were recorded with a Tobii Nano Pro tracker. Participants classified 50 emails (phishing, spear-phishing, promotional, legitimate) drawn from a corpus of 481 messages. Cognitive load was manipulated via concurrent puzzle solving and validated using NASA-TLX. To minimize bias, participants assumed fictional personas and made realistic decisions without explicit phishing instructions. Cognitive load and fatigue showed no significant effect on phishing susceptibility, though later trials exhibited a marginal increase in response bias. In contrast, eye-tracking metrics strongly predicted decisions. Logistic regression revealed that longer first fixation durations and higher saccade counts increased likelihood of responding, while larger pupil diameters were negatively associated with responses. Beyond results from the experiment, the study proposes attention-based representations integrating eye-tracking with natural language processing to improve cognitive models.

Keywords: Phishing, Attention, Decision making

INTRODUCTION

Phishing and spear-phishing attacks remain among the most persistent cybersecurity threats, despite continued advances in automated detection technologies. While spam filters successfully block the majority of mass phishing emails, personalized spear-phishing attacks frequently evade these systems, placing the burden of detection on end-users. Consequently, understanding how people perceive, process, and respond to email-based threats is critical for improving human-centered cybersecurity defences.

Prior research has largely attributed user susceptibility to phishing to inattention or failure to notice salient cues (Parsons et al., 2015; Harrison et al., 2016). However, this explanation offers limited insight into the underlying cognitive mechanisms driving decision-making in realistic email contexts. Human decisions are shaped not only by momentary attention but also by how information is encoded into memory and retrieved during subsequent judgments. Instance-Based Learning theory suggests that individuals rely on accumulated experiences when making decisions,

Received February 21, 2026; Revised April 4, 2026; Accepted April 19, 2026; Available online July 20, 2026

activating relevant memories based on contextual similarity (Gonzalez et al., 2003). Yet, empirical evidence linking visual attention during email processing to memory encoding and phishing-related decisions remains limited.

Recent eye-tracking studies have examined how users attend to non-textual elements of phishing emails, such as headers, links, or visual layout (McAlaney et al., 2020). Comparatively little work has focused on how attention to textual content influences cognitive processing and behavioural outcomes. Moreover, the interaction between cognitive workload, fatigue, and attention in phishing decision-making is not well understood, even though real-world email management frequently occurs under time pressure and mental strain (Boksem et al., 2005).

This paper addresses these gaps by investigating how visual attention and cognitive workload affect end-user decision-making in phishing contexts. We report findings from an eye-tracking experiment in which participants performed realistic email management tasks under varying levels of induced cognitive load. Eye movement data were collected to characterize attention patterns during email processing, and behavioral responses were analyzed using statistical models.

The contributions of this work are twofold. First, we empirically examine the effects of cognitive workload and fatigue on phishing discrimination and response bias. Second, we identify specific eye-tracking metrics—such as fixation duration, saccade counts, and pupil diameter—that significantly predict user responses to emails. Together, these findings advance understanding of the cognitive processes underlying phishing susceptibility and highlight opportunities for designing adaptive, human-centered cybersecurity interventions.

BACKGROUND

Phishing attacks exploit human cognitive processes through deceptive communication, impersonation, and contextual manipulation. While automated filtering systems effectively block most mass phishing emails, personalized spear-phishing attacks frequently bypass such defenses, leaving end-users as the primary detection mechanism. Consequently, understanding the cognitive foundations of human phishing susceptibility has become an important research focus.

Most prior human-centered phishing research has attributed susceptibility to failures of attention or insufficient processing of warning cues (Parsons et al., 2015). However, this perspective provides only a partial explanation of user behavior. Decision-making in phishing contexts is shaped not only by immediate perceptual cues but also by prior experiences stored in memory. Cognitive theories suggest that individuals rely on past encounters when evaluating novel situations, retrieving similar experiences to guide current judgments.

Cognitive models like Instance-Based Learning (IBL) theory formalizes this process by modeling decision-making as a function of accumulated experiences, or *instances,* each consisting of contextual features, actions, and outcomes (Gonzalez et al., 2003). Decisions emerge through a blending

mechanism that retrieves and weights past instances based on similarity, frequency, and recency. IBL has been successfully applied to a variety of dynamic decision-making domains and provides a principled framework for modelling intuitive, System-1–like behavior.

Recent phishing research has increasingly adopted cognitive models to explain and predict end-user responses to malicious emails (Cranford et al., 2019; Xu, Singh & Rajivan, 2022; Malloy & Gonzalez, 2024; Xu & Rajivan, 2026). Prior work demonstrated that cognitive models can capture human phishing decisions more effectively when email representations incorporate higher-order contextual meaning rather than surface-level semantic similarity alone (Xu, Singh & Rajivan, 2022; Xu & Rajivan, 2026). These findings suggest that users encode subjective interpretations of email content into memory, which subsequently influence future decisions. Moreover, cognitive models have been shown to perform particularly well for individuals exhibiting rapid, intuitive decision strategies, indicating alignment between IBL mechanisms and human heuristic processing.

Despite these advances, a critical gap remains in understanding how email content is encoded into memory during initial exposure. While IBL assumes that experiences are stored as instances, the perceptual and attentional processes governing instance formation are not directly observable. Prior eye-tracking studies in phishing contexts have primarily examined attention to non-textual cues such as headers, links, and visual layout (McAlaney et al., 2020; Pfeffel et al., 2019). These studies provide valuable insights into visual behavior but offer limited understanding of how users attend to linguistic content and how such attention shapes memory representations.

Furthermore, phishing decisions often occur under cognitive load and fatigue, conditions known to impair attention and working memory (Boksem et al., 2005). Yet, few studies have explicitly examined how workload interacts with attentional processes to influence phishing susceptibility from cognitive modeling lens.

The present work builds on previous work by using eye-tracking data to investigate how visual attention contributes to instance formation and decision-making. By combining gaze metrics with instance-based learning models, this study aims to bridge perceptual processes and memory-based cognition, providing a more comprehensive account of human phishing behavior.

RESEARCH QUESTIONS

This study focuses on understanding how cognitive workload and visual attention affect end-user decision-making in phishing contexts. We hypothesize that users encode information in emails that they pay attention to into memory and rely on these representations to guide subsequent responses when making decisions on whether to respond or not to a phishing email. Accordingly, this work addresses two primary research questions.

RQ1: What is the effect of fatigue and cognitive load on attention during phishing decision-making?

Decisions on phishing threats are not made in isolation. They occur as part of regular workflows which is likely to involve mental strain, time pressure, and accumulated fatigue. Cognitive workload is known to affect attention (Yang & Kim, 2019) and working memory, potentially altering how users process email content. This research question examines whether experimentally induced workload and fatigue influence visual attention and decision outcomes.

RQ2: How does eye movement influence human decision-making in email contexts?

Visual attention provides an observable window into how people process emails. Eye-tracking metrics such as fixation duration, saccade frequency, and pupil diameter reflect how users allocate attention to email. This question investigates the relationship between these eye-movement features and behavioral responses, with the goal of identifying attentional indicators predictive of phishing-related decisions.

Together, these questions aim to clarify how workload-related fatigue and visual attention interact to shape memory encoding and decision-making in phishing scenarios, providing empirical foundations for attention-informed cognitive models and human-centered cybersecurity interventions.

METHODS

Experimental Design

The objective of the experiment is to analyse parts of the emails that participants fixated upon prior to decision making and how these influence their final decision-making during email classification tasks. The study was conducted using a standard desktop computer equipped with a Tobii Nano pro eye tracker, which recorded participants' eye movement information as they made decisions about the presented email messages. Our hypothesis is that there is strong relationship between attention and memory recall processes on sentences and words that the participants find to be relevant to their decision-making.

Students from the University of Washington were recruited for the experiment. All the experiment sessions were held in person in the lab and were recruited via email. The median age of the participants was 22 (SD = 4.3). 29.1% were pursuing undergraduate, and those remaining were pursuing graduate-level degrees. 56.25% out of 48 participants' reported their native language as English. 56% of participants identified themselves as Female, 37% as male, and 4.1% as non-binary. Participants were asked to do the email management task. They were told that the goal of the study was to understand how people manage their emails because past research has shown that when participants know that the study is about detecting phishing emails, they tend to become cautious and biased towards reporting more emails as phishing (Parsons et al., 2015). For each email, they were asked to choose one of these options: 'Response immediately', 'Response and

follow up later', 'Leave in mailbox', 'Delete the email', and 'Delete and Block the Sender'. Each participant processed a total of 50 emails.

The context for decision-making is crucial in phishing and spear phishing. Therefore, the participants were given a fictional role and were asked to make decisions on behalf of the role, a.k.a persona. The profiles utilized in previous experiment were also retained for this current study (Xu, Singh, & Rajivan, 2023). They also had access to full information about their persona, including their name, personal details such as location and birth date, details about the persona's work and personal interests. The persona provided contains individual information, family information, professional information, and social information. For more information about the persona, please refer to our previous experiment (Xu, Singh, Rajivan, 2023).

Mental workload Manipulation: Workload have been suggested as important indicators of human susceptibility to phishing attacks (Zhuo et al., 2024). For example, it has been observed that participants who reported having been victims of phishing attacks also reported experiencing high levels of stress at the time of susceptibility (Vishwanath, 2015). Fatigue is also associated with detriment in human attention (Boksem et al., 2005). So, in this study, participants conducted a task that was expected to induce significant workload and fatigue prior to processing emails presented to them. To manipulate participants' fatigue levels, participants were asked to solve a series of puzzles (Haran et al., 2013) before making decisions on the emails. They were simple puzzles that required counting the mathematical symbols shown in an image. In the lower cognitive load condition, participants were required to solve only 2 such puzzles whereas participants in the high cognitive load group responded to 20 trials of puzzles. After solving the puzzles, participants in both condition were asked to filling a NASA TLX survey (Hart et al., 1998) to access their cognitive load.

RESULTS

We found that the mental demand, temporal demand, and effort, frustration as measured using NASA TLX scale were significantly different between the two conditions. The participants in the high-load condition generally report significantly higher values than the low-load condition group. In other words, they report a higher cognitive workload which suggests that we were able to vary workloads between the two condition using the puzzles. However, we did not find a significant effect of workload on participant eye movement metrics. For example, we did not find a significant difference in the number of fixation between the two conditions $t(69.41) = 1.82$, $p = 0.07$. We did observe that as trial number increased, participants were spending less fixation on emails $t(-6.35) = 2109.59$, $p < 0.0001$ which suggests that participants paid more attention during the initial trials compared to latter trials. Results are presented in Table 1.

Table 1: Fixation across trials and experiment condition.

	Estimate	Std. Error	df	t-Value	p-Value
(Intercept)	0.27	0.11	69.49	2.43	0.02
Trail number	–0.01	0	2109.59	–6.35	0
Condition (Low|High)	0.28	0.15	69.41	1.82	0.07

Next, we applied Signal Detection Theory measures such as d-prime and response bias were to analyze the effect of workload on participants' decisions to both the ham emails and phishing emails. The metric d-prime (d') serves as a measure of a participant's ability to discriminate between phishing and ham emails effectively. A larger value of d' indicates that the participant is effective in identifying phishing more often, whereas a lower d' value indicates that the participant could not discriminate phishing from ham emails. The metric response bias c reflects the overall bias of the participant to classify all emails as either phishing or ham. Both these measures are calculated using the hit rate and false alarm rate. The hit rate (HR) is a measure that quantifies the proportion of correctly identified phishing emails from the total number of phishing emails presented to participants (Signal). In other words, it measures how many of the phishing emails were correctly detected as phishing by the participants. On the other hand, the false alarm rate refers to the number of regular emails incorrectly classified as phishing emails. It is also known as false positives, in which regular emails are falsely identified as malicious.

$$HR = {Hits}/{Signal}$$

$$FAR = {False\ Alarms}/{Noise}$$

Using hit rate and false alarm rate, d-prime(d') and response bias c were calculated as defined in signal detection theory. When estimating d-prime and response bias for extreme proportions of stimulus, where either the hit rate (HR) or the false alarm rate (FAR) is 0 or 1, a correction is necessary to avoid division by zero. A common approach is to replace 0 with a small positive value(0.001) and 1 with a value slightly below(0.999) before calculating d-prime and response bias (Singh et al., 2023).

$$d' = z(HAR) - z(FAR)$$

The sensitivity index d prime refers to how easy or difficult it is to detect that a signal exists in the presence of noise. The d prime is the distance between the means of signal and noise distributions. If a participant has a low d-prime, it means that the participant has difficulty distinguishing phishing emails from benign emails. The response bias c represents the number of standard deviations from the midpoint between signal and noise distributions. d prime and c were calculated using the following equations.

$$c = -0.5 * \left[z(HAR) - z(FAR) \right]$$

A response bias greater than 0 indicates a bias toward classifying emails as ham and a response bias less than 0 indicates a bias toward classifying emails as phishing. To answer research question 1, we used a linear mixed effect model on both the d prime and response bias as the dependent variable. The independent variables were the experiment condition and whether the measurements were calculated in the first half or second half of the experiment.

To reduce noise from individual differences and emails, the random effects in the logistic regression model include participants' ID and email ID. Since each participant may receive a random set of emails, these are considered crossed random effect factors.

As shown in Table 2, for d', we found that participants demonstrated no difference between the first half and second half of the experiment trials s ($F(1,46) = 0.31$, $p = 0.58$), and no significant difference due to experiment condition (high and low cognitive load ($F(1,46) = 1.01$, $p = 1.01$)).

Table 2: Mixed effect model predicting d-prime.

	Sum Sq	Mean Sq	F-Value	P-Value
Experiment Phase (first half | second half)	0.18	0.18	0.31	0.5810
Condition (High | Low)	0.58	0.58	1.01	0.3191
Phase* Condition	0.84	0.84	1.46	0.2327

In terms of response bias, we found a marginal difference in participants' performance between the first half and second half of the experiment ($F(1,46) = 3.82$, $p = 0.0569$). See Table 3. The average was 0.79 in the first half of the experiment and 0.9 in the second half of the experiment. As participants experienced more fatigue towards the end of the experiment, they were much more likely to choose to respond to the emails presented to them. Again, the experiment condition did not show a significant effect on the response bias ($F(1,46) = 3.82$, $p = 0.2868$).

Table 3: Mixed effect model predicting response bias.

	Sum Sq	Mean Sq	F-Value	P-Value
Experiment Phase (First half | second half	0.39	0.39	3.82	0.0569*
Condition (High | Low)	0.12	0.12	1.16	0.2868
Phase*Condition	0.26	0.26	2.55	0.1168

From both tests, we did not observe effects from the experiment condition that would suggest that cognitive load and fatigue have a significant effect on susceptibility to phishing and spear-phishing emails. However, we found that the response bias has a marginal difference between the first half and

second half of the experiment, which may suggest that fatigue may have an effect on response bias.

Next, we analyzed the relationship between eye movement and decision-making. These analyses aims to answer research question 2: how would eye movement affect human decision-making? or what kind of eye metrics would result in decision-making in the email classification?

Participants' response to each email was encoded as response vs non-response. The response label includes the *response immediately*, and flag and *follow up later*. On the other hand, Non-response label includes the *leave in the mailbox*, *delete email*, *delete the email and block the senders*. To analyze the effect of eye-tracking metrics on decision making, we initially fit all the metrics to the model predicting the decision making and conducted feature selection to avoid correlating features. Table 4 shows the results from the final model. If the coefficient is positive, then participants were more willing to respond to the email, and vice versa.

Table 4: Mixed effect logistic model predicting decision-making.

	Estimate	Std-Error	Z-Value	P-Value
Intercept	1.31	0.22	5.85	0.00
Average duration of whole fixations	–0.18	0.11	–1.69	0.09
Duration of first whole fixations	0.23	0.10	2.38	0.02*
Average whole fixation pupil diameter	–0.33	0.16	–2.08	0.04*
Number of saccades	0.63	0.11	5.74	0.00***
Maximum peak velocity of saccades	0.11	0.09	1.28	0.20

We found that the duration of the first whole fixation and the number of saccades were the significant features that positively affected the decision. If there are interesting elements in the email, then participants generally spent more time on the first fixation. A similar pattern could be found in the number of saccades. However, the average duration of whole fixations is found was found to be insignificant. On the other hand, the average whole fixation pupil diameter is the factor that negatively significantly affects decision-making. In other words, the more average whole fixation pupil diameter, the more likelihood participants not responding to the email.

CONCLUSION

Using eye movement data, we investigated the relationship between decision-making and attention to textual features within emails. we found that factors such as the duration of the first whole fixation and the number of saccades were significantly associated with a higher likelihood of response. This would suggest that a person's first glance at an email strongly affects their decision-making. Our findings also reveal that the smaller the average whole fixation pupil diameter, the more likely participants were to respond. When a person's mental fatigue increases, their pupils constrict (Kaakinen & Hyona, 2014). This suggests that participants who were experiencing greater fatigue were

more likely to respond to emails, which means they were also more likely to fall victim to phishing attacks.

ACKNOWLEDGMENT

This work was supported by a grant from the National Science Foundation (NSF grant # 2142888). The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

REFERENCES

Boksem, M. A., Meijman, T. F., & Lorist, M. M. (2005). Effects of mental fatigue on attention: an ERP study. Cognitive brain research, 25(1), 107–116.

Cranford, E. A., Lebiere, C., Rajivan, P., Aggarwal, P., & Gonzalez, C. (2019). Modeling cognitive dynamics in end-user response to phishing emails. Proceedings of the 17th ICCM.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. Cognitive Science, 27(4), 591–635.

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. Judgment and Decision making, 8(3), 188–201.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139–183). North-holland.

Harrison, B., Svetieva, E., & Vishwanath, A. (2016). Individual processing of phishing emails: How attention and elaboration protect against phishing. Online Information Review, 40(2), 265–281.

Hopstaken, J. F., Van Der Linden, D., Bakker, A. B., & Kompier, M. A. (2015). The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics. Biological psychology, 110, 100–106.

Kaakinen, J. K., & Hyönä, J. (2014). Task relevance induces momentary changes in the functional visual field during reading. Psychological Science, 25(2), 626–632.

Malloy, T., & Gonzalez, C. (2024). Applying Generative Artificial Intelligence to cognitive models of decision making. Frontiers in psychology, 15, 1387948.

McAlaney, J., & Hills, P. J. (2020). Understanding phishing email processing and perceived trustworthiness through eye tracking. Frontiers in Psychology, 11, 1756.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. Computers & Security, 52, 194–206.

Pfeffel, K., Ulsamer, P., & Müller, N. H. (2019, June). Where the user does look when reading phishing mails–an eye-tracking study. In International Conference on Human-Computer Interaction (pp. 277–287). Cham: Springer International Publishing.

Singh, K., Aggarwal, P., Rajivan, P., & Gonzalez, C. (2023). Cognitive elements of learning and discriminability in anti-phishing training. Computers & Security, 127, 103105.

Vishwanath, A. (2015). Examining the distinct antecedents of e-mail habits and its influence on the outcomes of a phishing attack. Journal of Computer-Mediated Communication, 20(5), 570–584.

Xu, T., Singh, K., & Rajivan, P. (2022, January). Modeling Phishing Decision using Instance Based Learning and Natural Language Processing. In Hicss (pp. 1–10).

Xu, T., Singh, K., & Rajivan, P. (2021, September). Spearsim: Design and evaluation of synthetic task environment for studies on spear phishing attacks. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 65, No. 1, pp. 1500–1504). Sage CA: Los Angeles, CA: SAGE Publications.

Xu, T., Singh, K., & Rajivan, P. (2023). Personalized persuasion: Quantifying susceptibility to information exploitation in spear-phishing attacks. Applied Ergonomics, 108, 103908.

Xu, T., & Rajivan, P. (2026). Analyzing instance representation in cognitive models of phishing decision-making: T. Xu, P. Rajivan. User Modeling and User-Adapted Interaction, 36(1), 7.

Yang, X., & Kim, J. H. (2019). Measuring workload in a multitasking environment using fractal dimension of pupil dilation. International Journal of Human–Computer Interaction, 35(15), 1352–1361.

Zhuo, S., Biddle, R., Betts, L., Arachchilage, N. A. G., Koh, Y. S., Russello, G., & Lottridge, D. (2024). The Impact of Workload on Phishing Susceptibility: An Experiment. In Symposium on Usable Security and Privacy (USEC).