Errors, Irregularities, and Misdirection: Cue Utilisation and Cognitive Reflection in the Diagnosis of Phishing Emails

Mitchell Ackerley
Charles Sturt University, Bathurst, Australia

Ben W. Morrison
Macquarie University, Sydney, Australia
ben.morrison@mq.edu.au

Kate Ingrey
Charles Sturt University, Bathurst, Australia

Mark W. Wiggins
Macquarie University, Sydney, Australia

Piers Bayl-Smith
Macquarie University, Sydney, Australia

Natalie M. V. Morrison
Western Sydney University, Blacktown, Australia

Abstract

The study aimed to examine the role of, and potential interplay between, cue utilisation and cognitive reflection in email users' ability to accurately (and efficiently) differentiate between phishing and genuine emails. 145 participants completed the Cognitive Reflection Test (CRT), a phishing diagnostic task, and the Expert Intensive Skill Evaluation (EXPERTise 2.0) battery, which provided a gauge of users' cue utilisation in the domain. The results revealed an interaction between users' cognitive utilisation and cue reflection, whereby users low in both facets performed significantly worse in diagnosing phishing emails than all other groups. Further, those participants with both higher cue utilisation and cognitive reflection took significantly longer to make their diagnosis. It is concluded that a high level of cognitive reflection was able to compensate for a lower level of cue utilisation, and vice versa. Participants reported using seven types of cue during diagnosis, however, there was no significant relationship between the types of cues used and users' level of cue utilisation. Taken together, the findings have implications to the design of user-level interventions in relation to the identification of vulnerable users, as well as the need to consider training approaches that extend beyond the use of simple cue inventories.

Keywords: cyber security, phishing, cue utilisation, cognitive reflection, expertise.

1 Introduction

Within a technologically advancing world, organisations have become increasingly reliant on Information Technology (IT) to facilitate business services. However, this reliance is met with certain risks to an organisation's cyber security. The Australian Government suggests that cyber-crime results in losses of one billion dollars per year for Australian organisations (Department of Prime Minister and Cabinet, 2019). The most common means of accessing personal information by cyber criminals in Australia is via *phishing* emails (ACSC, 2021; Moore, 2021).

Phishing emails are socially engineered messages that are designed to appear legitimate communications in order to lure email users to 'bite' or engage with the email (Mayhorn & Nyeste, 2012). Engaging with these emails (e.g., clicking on links) invariably leads to breaches of sensitive information, malware infections, and in some cases the forceful request of direct payment (Abawajy, 2012; Fette et al., 2007). Consequently, understanding the factors that underlie users' misdiagnosis of phishing emails as trustworthy is a critical line of enquiry. While system design is undoubtedly vital to countering phishing scams, user-level differences play a significant role (Downs et al.; Fette et al., 2007; Jones et al., 2015). Email users' skilled use of visual cues (i.e., features present within an email that signal a problem) have been implicated in their ability to detect phishing emails (Bayl-Smith, et al., 2020; Nasser et al., 2020a, 2020b). For example, irregularities such as an unfamiliar email address or a typographical error will regularly cue the email user to be suspicious of the author's intentions.

In more sophisticated phishing attempts, scammers will tailor content to revolve around a message that prompts the user to act urgently. The message will often include an element of social proof (i.e., everyone is said to be doing the act, or have already done it). Further, the author will frequently mimic an authority figure, such as the email user's boss. Indeed, Hassandoust et al. (2020) recently found that email users are more vulnerable to phishing attacks when the messages they receive are specific to their context, thereby targeting their individual vulnerabilities. These phish authors aim to conceal their intentions by bypassing users' analytical or rational modes of thinking (often referred to as System 2 in dual process information processing models; Stanovich & West, 2000), and instead target users' heuristic system of mental shortcuts (i.e. System 1 reasoning).

System 1 is automatically favoured upon perception of information, as processing is associated with a miserly use of cognitive resources. This system is very efficient, and extremely useful, leveraging a powerful associative network in memory. However, it is also associated with an increased risk of bias-related error (Tversky & Kahneman, 1974;1975), which may lead to the misapplication of cues, particularly in conditions characterised by uncertainty and time pressure (Kahneman & Tversky, 1972). For instance, emotive messages may trigger an *affect* heuristic, which often lead us to make decisions based on how we feel, rather than think. In this sense, just as a magician uses their knowledge of the nature and limitations of human perception to draw the observer's attention away from key details of their tricks (e.g., distracting them with movement in one hand, while doing something with the other), the phish author can engineer their email to exploit the cognitive biases of email users. However, such an approach assumes that email users will be largely unable to utilise diagnostic cues, either as a result of misdirection tactics, or a lack of user knowledge and/or skill.

Cue utilisation has shown to differ as a function of domain expertise, with more experienced people leveraging their store of mental cues to yield more accurate and reliable responses to complex tasks (Brams et al., 2019; Shanteau, 1992). In addition to expertise, the acquisition and use of cue-based associations may be related to an individual's typical mode of cognitive reflection, with more intuitive users being more prone to simple heuristic reasoning, compared to more rational users (i.e., System 2 users) who may naturally devote more attentional resources to inspecting emails for suspicious features (Harrison et al., 2016; Luo et al., 2013; Vishwanath et al., 2016; Vishwanath et al., 2011); the latter group being better positioned to ultimately develop *skilled intuition* within a particular context, which will manifest in the skilful recognition of diagnostic cues (Kahneman & Klein, 2009).

The current study aimed to examine the role of, and potential interplay between, cue utilisation and cognitive reflection in email users' ability to accurately (and efficiently) differentiate between phishing and genuine emails. In doing so, this work directly extends upon existing phishing victimology frameworks that have established the role of cue utilisation (Nasser et al., 2020a, 2020b; Bayl-Smith et al., 2020) and factors relating to dual process information processing (Butavicius et al., 2016; Frauenstein & Flowerday, 2020; Harrison et al., 2016; Jones et al., 2019; Luo et al., 2013; Valecha et al., 2015; Vishwanath et al., 2011; Vishwanath et al., 2016; Vishwanath et al., 2018; Yan & Gozu, 2012; Zhang et al., 2012) in determining users' phishing susceptibility. Critically, clarifying the interplay between users' perceptual-cognitive skills and their existing preferences concerning information processing have implications to the future design of cyber security interventions such as education and training programs, and more broadly, to our understanding of the acquisition of skilled intuition and expertise in similar domains of judgement and decision-making.

2 Cue Utilisation

Relatively greater levels of skilled cue utilisation have been associated with effective decision-making in a range of work domains (Wiggins, 2021) such as medicine (Galanter & Patel, 2005), lie detection (Morrison et al., 2020; Walczyket al., 2009), disaster recovery (Gacasan et al., 2016), aviation (Wiggins, et al., 2014), drone piloting (Shekhet al., 2018), Rugby League (Johnston & Morrison, 2016), forensic investigation (Morrison et al., 2018; Morrison & Morrison, 2015) and electricity power control (Wiggins et al., 2014). This is partly because, in complex work settings, the frugal use of a limited number of valid cues will invariably reduce the demands placed on individuals' working memory during decision-making (Brams et al., 2019). Consequently, a cue-based strategy enables users to manage multiple tasks simultaneously (Klein et al., 1986), which is particularly relevant when considering a phishing diagnosis context wherein engagement with emails is often considered a secondary or 'background' task.

One model that explores the relationship between cue utilisation and decision-making in a phishing email context is the Instance-based Learning Theory (IBLT; Gonzalez, 2013). The IBLT extends on Anderson's (1996) Adaptive Control of Thought-Rational (ACT-R) architecture, proposing that dynamic decisions are made based on generalisations of previous experiences. In line with the ACT-R model, these instances reside as chunks within the users' declarative memory and are activated by cues (e.g., poor grammar) within the email. Once cued, the chunk with the highest activation level in memory is retrieved. This is similar to Klein et al.'s (1986) Recognition-Primed Decision (RPD) model, which posits that experienced decision makers will use cues to draw on a repertoire of stored patterns that accumulate as a result of domain-specific experiences. In matching these patterns, the operator will non-consciously generate and prioritise a series of plausible responses, which they can test via a process of mental simulation (a notion consistent with deGroot's concept of 'progressive deepening'; 1978). Thus, cues are thought to be central to both intuitive (System 1) and deliberate (System 2) aspects of the RPD model. Upon first attending to email content, it is expected that users form an impression of validity based on the unconscious activation of cue-based associations (the make-up and quality of this network is largely contingent on their domain knowledge). If this impression exceeds a 'suspicion threshold', the user will engage a more deliberate review of the email's features and content in forming a final judgement of its validity.

Shonman et al. (2018) investigated the application of IBLT on users' suspicion thresholds, maximum cues processed, and strength of prior cue associations (i.e., chunks) in a phishing diagnosis task. Results showed that users who exhibited higher levels of prior activation in memory chunks more accurately identified phishing emails. Moreover, users who identified up to four deceptive cues, and therefore reached a suspicion threshold, showed greatest accuracy (Shonman et al., 2018). This suggested that users' capacity for cue utilisation directly shaped their susceptibility to being deceived by phishing emails.

In a similar study of cue utilisation and phishing diagnosis, Nasser et al. (2020a) employed a Cyber Security edition of EXPERTise 2.0 (Wiggins, 2016; Wiggins, et al., 2014), which is a software-based assessment battery designed to objectively assess users' recognition of cues. The program comprised five tasks, asking individuals to identify diagnostic features (Feature Identification Task; FIT), recognise key features (Feature Recognition Task; FRT), associate features and/or events in memory (Feature Association Task; FAT), discriminate relevant from less relevant features (Feature Discrimination Task; FDT), and prioritise information during problem solving (Feature Prioritisation Task; FPT). Participants who showed greater levels of phishing cue utilisation in the program were found to perform better in a separate performance measure of phishing email diagnosis (Nasser et al., 2020a). Despite differentiating levels of performance based on phishing cue utilisation, the authors found no differences in the types of cues used across 'high' and 'low' cue users. However, this may be attributed to the fact that participants were asked to select the cues that they used from a list, rather than recall them from memory, which may have led to a rationalisation effect. A further limitation of the work was that it did not address the impacts of users' individual differences, many of which have been implicated previously in phishing susceptibility. In particular, email users' cognitive reflection tendencies have been a focal point of recent studies relating to phishing email detection (Butavicius et al, 2016; Frauenstein & Flowerday, 2020; Jones et al, 2015; Jones et al., 2019; Yan & Gozu, 2012). Critically, such differences may moderate the cue utilisation-performance relationship.

3 Cognitive Reflection

Dual-process models (see Evans, 2008; Stanovich & West, 2000) of information processing have been widely adopted by cyber security researchers as a theoretical framework for understanding phishing victimology (Bayl-Smith et al., 2020; Frauenstein & Flowerday, 2020; Harrison et al., 2016; Luo et al., 2013; Valecha et al., 2015; Vishwanath et al.,2016; Vishwanath et al., 2018; Zhang et al., 2012). In such models skilled cue utilisation is presumed to occur as part of the associative mechanisms of System 1, which is characterised by rapid, automatic, unconscious, and relatively effortless responses. However, relatively unskilled cue utilisation and the use of simple 'rules of thumb' are also theorised to reside in System 1; the difference between the two types of user behaviour will largely be a function of one's accumulated expertise in a domain (i.e., whether they possess the *right* cue-based associations for the job). In contrast, mental simulation, and 'deepening' processes (Kahneman & Klein, 2009) are better aligned to System 2, which is characterised by slow, systematic, deliberate, and relatively effortful progressing.

The Elaboration Likelihood Model (ELM) also presents a dual-cognitive model (Petty & Cacioppo, 1986). The model suggests that individuals using the central information processing pathway activate two sub-processes: attention (i.e., mental focus) and elaboration, described as the process forming connections between the present experience and past knowledge (Harrison et al., 2016). According to proponents of this model, a proposed benefit of the System 2 pathway is that individuals engage deeply with contextual features such as those found in phishing emails. Here, upon receipt of an email, activation of the System 2 pathway may prompt a user to continue their information search and carefully examine the email for suspicious features (Harrison, et al., 2016; Luo et al., 2013; Vishwanath et al., 2016; Vishwanath, et al., 2011).

A body of research has implicated *cognitive reflection* as a significant activator of the systematic thinking pathway (Isler et al., 2020; Kahneman & Klein, 2009; Pennycook & Rand, 2019). Cognitive reflection is defined as an information processing technique and is measured by behavioural outcomes characterised along a spectrum of rapid, impulsive decision-making or slower, rational, and stepwise approaches (Frederick, 2005). Using the Cognitive Reflection Test (CRT; Frederick, 2005), Butavicius et al. (2016) tested the relationship between participants' pre-existing cognitive reflection tendencies and their ability to detect phishing email threats. Their results revealed that participants who were less cognitively impulsive (i.e., scored higher on the CRT) were more likely to judge fraudulent emails as dangerous. Here, the authors highlighted the importance of heuristic use, suggesting that participants engaging in less reflective and more intuitive processes (i.e., greater System 1 and less System 2 thinking) may be more prone to increased errors in judgement when assessing phishing emails (Butavicius et al., 2016). Jones et al. (2019) revealed a similar relationship between participants' CRT results in predicting user susceptibility to phishing emails.

In line with these findings, Yan and Gozu (2012) investigated differences in intuitive or rational decision-making strategies of 171 email users. Participants were instructed to give either a rapid (i.e., intuitive) response or assess the email carefully (i.e., rationally) before submitting their response. The results revealed that those in the rational decision-making group correctly identified more emails as suspicious than those who provided an intuitive response (Yan & Gozu, 2012). Comparable effects on performance have been found when implementing a time pressure, which was presumed to induce intuitive responses (Jones et al., 2019). Similarly, Bayl-Smith et al., (2020) found that those participants who deliberated longer per email, seemingly demonstrating a greater degree of cognitive reflection, demonstrated an increased ability to correctly identify phishing features.

4 Study Aims

While the heuristic-systematic framework accounts for an unskilled perspective of phishing susceptibility akin to Kahneman and Tversky's heuristics and biases paradigm (Tversky & Kahneman, 1974; 1975), it fails to account for instances of skilled cue utilisation, which are typically studied within a Naturalistic Decision-Making framework (Klein, 2008). Indeed, with the accumulation of operational experience, we posit that email users will possess numerous cue-based associations that reside in long-term memory, which may be triggered upon reviewing an email. Here, if a suspicion threshold is exceeded (presumably based on activation strength; see Anderson, 1996), users will engage a more deliberate (and conscious) process of feature scrutiny and verification; presumably via System 2 reasoning. Further, we contend that reaching suspicion threshold may be moderated by a multitude of individual differences, for instance, a user's natural inclination of engaging a process of cognitive reflection appears highly relevant. Therefore, we propose that email users who demonstrate higher cue utilisation in a phishing context and/or a general tendency to engage in cognitive reflection, are likely to be best positioned to detect phishing threats. Conversely, we would expect those users with relatively low cue utilisation and cognitive reflection to be most vulnerable to phishing scams.

The study aimed to examine the role of, and potential interplay between, cue utilisation and cognitive reflection in email users' ability to accurately (and efficiently) differentiate between phishing and genuine emails. As in Bayl-Smith et al. (2020), and Nasser et al. (2020a; 2020b), EXPERTise 2.0 (Wiggins et al., 2014) was used to gauge participants' level of phishing cue-utilisation. Consistent with their findings we expected that those participants with higher levels of cue utilisation would demonstrate superior performance on a phishing diagnosis task compared to those with lower levels of cue utilisation, but that, based on the theorised relationship between greater cognitive reflection and the development of skilled cue utilisation (Kahneman & Klein, 2009), this relationship would also be moderated by users' cognitive reflection tendencies. Specifically, we hypothesised (H1) that those participants found to be high in both cue utilisation and cognitive reflection would perform significantly better (in relation to phishing diagnosis) than all other participant groups. Conversely, we hypothesised (H2) that those participants found to be low in both cue utilisation and cognitive reflection would perform significantly worse than all other participant groups.

As higher scores on the Cognitive Reflection Test (CRT) have been associated with more 'patient' decision-making, presumed to be more informed by careful deliberation (Frederick, 2005), it was expected that such deeper processing will manifest in significantly longer response latencies for formulating diagnoses about email suspiciousness. Consequently, we hypothesised (H3) that those participants found to possess relatively lower cognitive reflection would demonstrate significantly shorter response latencies when diagnosing an email, compared to those with high cognitive reflection. Further, based on the theorised role of cognitive reflection tendencies in the development of skilled cue utilisation (Kahneman & Klein, 2009), we expected that cue utilisation may moderate the relationship between cognitive reflection and efficiency, in that those individuals with both higher cognitive reflection and cue utilisation are most likely to resemble expert operators in this domain and demonstrate both accurate and efficient visual diagnosis (Bram et al., 2019). In this sense, we expect those with high cognitive reflection tendencies to be efficient in domains in which they have accumulated a repertoire of domain-specific associations. Therefore, we hypothesised (H4) that those participants who demonstrated a high degree of cognitive reflection will take significantly less time to diagnose when they also possess a high level of cue utilisation, compared to a low level of cue utilisation.

Finally, despite previous work differentiating levels of performance based on phishing cue utilisation (Nasser et al., 2020b), and others eliciting a set of 'correct' phishing cues (Parsons et al., 2016), differences in the types of cues used across high and low cue users are yet to be captured. This has largely been attributed to limitations in previous designs (e.g., asking participants to recognise the features they use from a pre-defined list; Nasser et al., 2020b). As such, a second aim of the study was to, using an open-response format, elicit the cues used by email users when deciding whether an email is genuine or not, and test whether the types of cues reportedly used were associated with the degree of cue utilisation demonstrated by participants (i.e., whether high and low cue utilisers were found to use different cues when making their decisions).

5 Method

5.1 Participants

A convenience sample of 145 (79 female, 66 male) participants comprised first-year undergraduate psychology students enrolled at Charles Sturt University (CSU; Australia) and members of the public. Female ages ranged from 18 to 76 years ($M_{age} = 38$, $SD_{age} = 14$), and male ages ranged from 18 to 80 years ($M_{age} = 34$, $SD_{age} = 14$).

5.2 Materials

5.2.1 Cognitive reflection

The Cognitive Reflection Test (CRT; Frederick, 2005) was used to measure participants' cognitive reflection tendencies. The CRT is a three-item behavioural-based questionnaire that measures participants' tendency for intuitive or rational judgment. It is designed to be consistent with Stanovich and West's (2000) conception of a System 1 and System 2 cognitive framework. When completing the measure, participants were asked to respond to three items in open text. A total score of three was obtainable, with each correct answer equalling one point, consistent with Frederick's (2005) scoring method. A higher score on the CRT indicated higher levels of cognitive reflection.

5.2.2 Phishing diagnosis performance

The phishing diagnosis task was hosted and completed on Qualtrics (Qualtrics, 2019). The task was developed in-house by the research team and measured participants' accuracy in differentiating between genuine and phishing emails. The task comprised images of five legitimate and five phishing emails, which were presented to participants one at a time, in a random order. Legitimate emails were compiled from emails the research team had received, with identifiable information removed. Phishing emails were confirmed phishing attempts extracted from the research team's spam folders over a 6-month period. Upon presentation of each email image, participants were asked to click one of two on-screen labels to designate whether they thought the email was "trustworthy" or "suspicious".

5.2.3 Cue utilisation

A cyber-security edition of EXPERTise 2.0 was used to evaluate behaviour that was indicative of cue utilisation in the phishing domain. The predictive validity for EXPERTise has been assessed in several domains such as drone search and rescue tasks (Shekh et al., 2018), audiology (Watkinson et al., 2018) aviation (Wiggins, 2014), and most recently in cyber security (Nasser et al., 2020b).

5.2.3.1 Feature Identification Task (FIT)

The FIT was used to assess participants' ability to identify key visual features from 15 individual emails (presented individually) by using their cursor to click on a box appearing on the screen labelled "trustworthy" or clicking on suspicious features of the email itself. The FIT measures a participant's ability to recognise and recall diagnostic cues from the environment (Loveday et al., 2013). Experts are primed to use diagnostic cues and should reflect lower response latencies (Wickens et al., 2013). Participants with higher cue utilisation show lower mean response times across scenarios when identifying key features (e.g., broken images) that were suspicious.

5.2.3.2 Feature Recognition Task (FRT)

The FRT measured participants' ability to accurately recognise diagnostic features of phishing emails (Wiggins et al., 2018). The FRT generates a score rather than measuring response latencies. Participants were shown 20 emails for a short period of time (1 second) and were then instructed to classify each one as "trustworthy", "untrustworthy", or "impossible to tell". Research suggests that experts possess a larger and more refined library of cues in memory, which allows for more rapid responses when identifying prominent cues (Loveday et al., 2013; Wiggins et al., 2018).

5.2.3.3 Feature Association Task (FAT)

The FAT measured participants' capacity to recognise associations between specific diagnostic features of a phishing email. Evidence has shown that associated concepts are activated within a shorter time frame due to pre-existing associations residing within memory (Morrison et al., 2013). Two words were presented at the same time, for a short period, with participants having to indicate the relatedness of the words. A total of 15 pairs were shown, with participants indicating relatedness by using a 7-point Likert-type scale (1 = *Extremely Unrelated* to 7 = *Extremely Related*). The variance of responses across individual participants were calculated, with higher levels of cue utilisation found to be associated with greater levels of variance in responses (Morrison et al., 2013).

5.2.3.4 Feature Discrimination Task (FDT).

The FDT is based on the Cochran-Weiss-Shanteau index (Shanteau, et al., 2002), which is based on the premise that expertise can be measured by a ratio of discrimination divided by consistency in responses (Weiss & Shanteau, 2003). From this formula, the FDT only utilises the discrimination component of this ratio, with experts predicted to display greater variance in responses (Loveday, et al., 2013). The FDT measured participants' ability to discriminate between relevant and less relevant features of a suspected phishing email. Participants were asked to read two scenarios about a phishing email they encountered with an associated picture of the email. Participants were then given several choices (e.g., pay the duty as requested) and were instructed to choose one. Once a decision had been selected, the participants were asked to rate, using a 10-point Likert-type scale (1 = *Not Important at All* to 10 = *Extremely Important*) the extent to which the 10 features of the scenario had impacted their decision. Greater levels of variance between responses tend to be associated with greater discrimination, and higher levels of cue utilisation (Loveday, et al., 2013; Weiss & Shanteau, 2003).

5.2.3.5 Feature Prioritisation Task (FPT)

The FPT measured participants' capacity to prioritise cue-based information. Participants were given 30 seconds to click on separate menus related to different features of the email (e.g., time the email was sent) across two different scenarios. Clicking on separate menus revealed information about the email. At the end of the 30 seconds, participants were asked to formulate a decision about how they would respond to the email, similar to the FDT. The FPT examines the ratio of sequential transitions to total transitions and is based on the finding that superior information acquisition strategies involve discrimination between features based on their utility, while relatively unskilled acquisition, typically seen in novices, tends to be largely sequential in nature (Wiggins et al., 2018).

5.3 Cue type survey

After completing the 10 email diagnosis task questions, participants were asked one open-text question: "What feature(s) of the emails helped you make your decision?".

5.4 Procedure

Upon commencing the study, participants were asked to generate a unique six-digit identification code. Participants then completed the CRT (Frederick, 2005), the phishing diagnostic task, and the cue type survey on the Qualtrics platform (Qualtrics, 2019). Upon completion, participants were re-directed to the EXPERTise 2.0 platform, where they would again enter their six-digit code, which was used to link data between the Qualtrics and EXPERTise 2.0 platforms. After clicking through, participants recorded their demographic information and completed the five tasks.

6 Results

6.1 Data Reduction

Participants' responses on the phishing diagnosis task were scored in relation to their precision in diagnosing a suspicious email among both suspicious (5) and trustworthy (5) emails. This score was calculated by summing the number of correct diagnosis divided by the sum of correct and incorrect diagnoses. Participants' diagnosis efficiency (i.e., time taken to diagnose an email as trustworthy or suspicious) was recorded in seconds (s) and an average response time was calculated for the ten emails.

Participants' responses on the CRT were scored as either correct or incorrect and tallied to yield a total score out of 3. To identify two relatively distinct groups of participants based on their cognitive reflection tendencies, participants who scored either '0' (low cognitive reflection; n = 57) or '3' (high cognitive reflection; n = 32) were retained (N = 89). All other participants were excluded from analyses involving cognitive reflection.

In line with the conventional method for analysing EXPERTise 2.0 data (Bayl-Smith, et al., 2020; Brouwers et al., 2016; Sturman et al., 2019), each of the five tasks of EXPERTise 2.0 were standardised into z-scores (see Table 1) prior to conducting a *k*-means cluster analysis. Scores on the FAT, FDT, and FPT yielded statistically significant mean differences across a two-group model intended to reflect relatively high and low levels of cue utilisation. Participants displaying positive standardised mean for the FAT and FDT, and a negative standardised mean for the FPT, as observed in the current findings, are generally considered above average in relation to cue utilisation performance (Brouwers et al., 2016; Loveday et al., 2013; Loveday et al., 2014). Scores on the FIT and FRT failed to reveal a statistically significant difference between the two groups and were excluded from further analysis. Overall, thirty-nine participants comprised a group whose behaviour was consistent with a relatively 'high' degree of cue utilisation, while 50 participants comprised a 'low' cue utilisation group. See Table 1 for Standardised means from the EXPERTise 2.0 tasks by derived cluster centroids for the three retained tasks.

EXPERTise 2.0 Task	**High cue utilisation cluster (*n* = 39)**	**Low cue utilisation cluster (*n* = 50)**
Feature Association Task (variance/time)	.483	-.416
Feature Detection Task (variance)	.888	-.671
Feature Prioritisation Task (ratio)	-.244	.170

Table 1. Standardised means from EXPERTise 2.0 tasks by derived cluster centroids (high and low cue utilisation) for the three retained tasks

Note: The F test differences between clusters were statistically significant ($p < .05$).

6.2 Cue utilisation, cognitive reflection, and phishing diagnostic performance

6.2.1 Diagnosis Precision

Participants' phishing diagnosis precision was examined using a 2 x 2 factorial between-groups analysis of variance (ANOVA). Results revealed a statistically significant interaction between cognitive reflection and cue utilisation for precision scores, *F* (1,85) = 5.06, *p* = .027, partial η^2 = .056. A main effect for cognitive reflection was also found, *F* (1,85) = 4.57, *p* = .035, partial η^2 = .051, but not for cue utilisation, *F* (1,85) = 1.04, *p* = .311, partial η2 = .012. Descriptive statistics are provided in Table 2, and the interaction is shown in Figure 1.

	High cognitive reflection		**Low cognitive reflection**		**Total**	
Cue utilisation	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	.79	.11	.80	.12	.79	.11
Low	.83	.11	.70	.15	.75	.15
Total	.81	.11	.74	.14	.77	.14

Table 2. Mean Precision for Cue Utilisation by Cognitive Reflection

To further investigate the interaction, four simple effects tests were performed for cue utilisation and cognitive reflection groupings. A Bonferroni adjusted alpha of .0125 was used to control the familywise error rate at .05 (Field, 2013). The simple effect for cue utilisation for those in the low cognitive reflection group was statistically significant, *F* (1,85) = 7.43, *p* = .008, partial η^2 = .075, with those in the high cue utilisation group (*M* = .796, *SE* = .03) demonstrating greater precision than those in the low cue utilisation group (*M* = .703, *SE* = .02). The simple effect for those in the high cognitive reflection group was not statistically significant, *F* (1,85) = .589, *p* = .445, partial η^2 = .007, with those in the high cue utilisation group (*M* = .793, *SE* = .03) scoring no differently to those in the low cue utilisation group (*M* = .828, *SE* = .03).

The simple effect for cognitive reflection for those in the low cue utilisation group was statistically significant, *F* (1,85) = 10.988, *p* = .001, partial η^2 = .114, with those in the high cognitive reflection group (*M* = .828, *SE* = .03) demonstrating greater precision than those in the low cognitive reflection group (*M* = .703, *SE* = .02). The simple effect for those in the high cue utilisation group was not statistically significant, *F* (1,85) = .005, *p* = .941, partial η^2 = .000, with those in the high cognitive reflection group (*M* = .793, *SE* = .03) scoring no differently to those in the low cognitive reflection group (*M* = .796, *SE* = .03).

In other words, those with either high levels of cue utilisation or cognitive reflection (or both) performed to a similar level. This finding does not provide support for the clear advantages predicted among the high cue utilisation and cognitive reflection group (H1), as it appears that a high level of cue utilisation is able to compensate for a low level of cognitive reflection, and vice versa. However, participants with both low levels of cue utilisation and cognitive reflection were found to perform significantly worse on the phishing diagnosis task than all other groups, supporting H2.

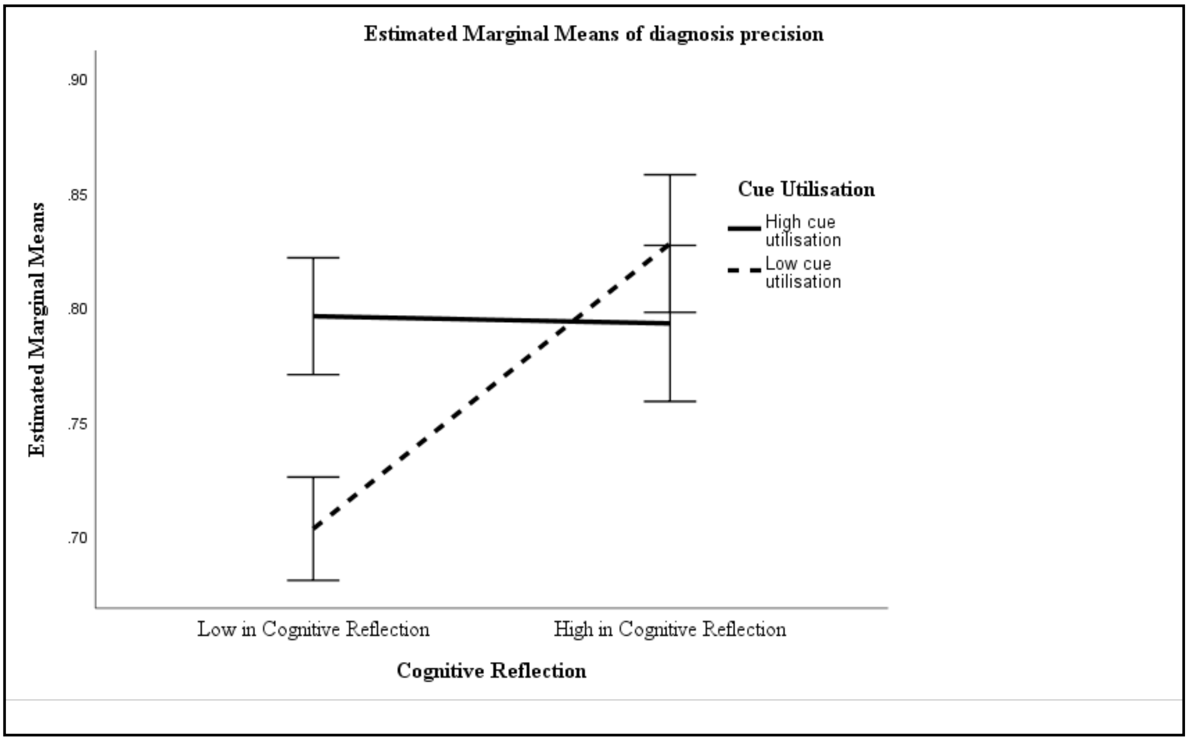

Figure 1. Interaction between cue utilisation and cognitive reflection for diagnosis precision. Error bars represent standard errors (±1 SE)

In other words, those with either high levels of cue utilisation or cognitive reflection (or both) performed to a similar level. This finding does not provide support for the clear advantages predicted among the high cue utilisation and cognitive reflection group (H1), as it appears that a high level of cue utilisation is able to compensate for a low level of cognitive reflection, and vice versa. However, participants with both low levels of cue utilisation and cognitive reflection were found to perform significantly worse on the phishing decision task than all other groups, supporting H2.

6.2.2 Diagnosis efficiency

Participants' phishing diagnosis efficiency was examined using a 2 x 2 factorial between-groups analysis of variance (ANOVA). Results revealed a statistically significant interaction between cognitive reflection and cue utilisation for average diagnostic time, F (1,85) = 6.61, p = .012, partial η^2 = .072. A main effect for cognitive reflection was also found, F (1,85) = 6.49, p = .013, partial η^2 = .071, but not for cue utilisation, F (1,85) = 2.85, p = .095, partial η^2 = .032. Descriptive statistics are provided in Table 3, and the interaction is shown in Figure 2.

	High cognitive reflection		**Low cognitive reflection**		**Total**	
Cue utilisation	*M*	*SD*	*M*	*SD*	*M*	*SD*
High	17.54	5.55	12.14	4.98	14.08	5.75
Low	13.27	4.16	13.55	5.52	13.45	5.03
Total	15.14	5.20	12.93	5.29	13.72	5.34

Table 3. Mean diagnosis time for Cue Utilisation by Cognitive Reflection

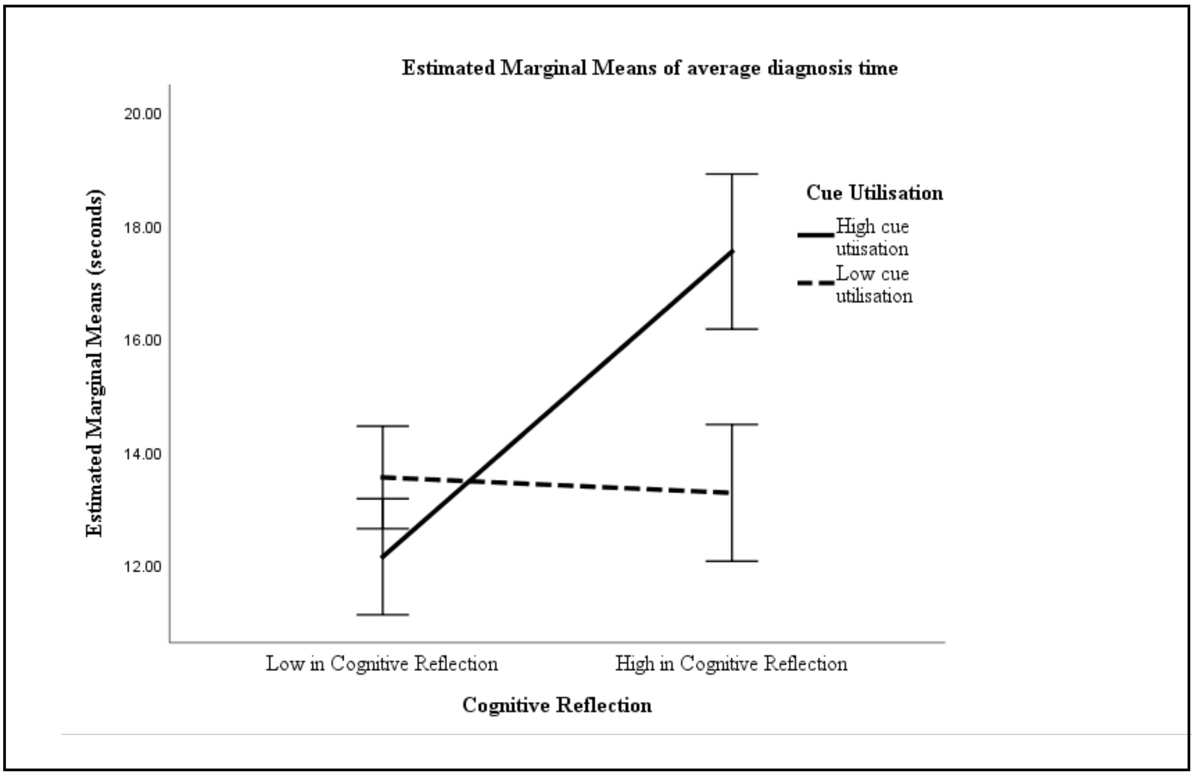

Figure 2. Interaction between cue utilisation and cognitive reflection for average diagnosis time. Error bars represent standard errors (±1 SE)

To further investigate the interaction, four simple effects tests were performed for cue utilisation and cognitive reflection groupings. A Bonferroni adjusted alpha of .0125 was used to control the familywise error rate at .05 (Field, 2013). The simple effect for those in the low cognitive reflection group was not statistically significant, *F* (1,85) = .542, *p* = .464, partial η^2 = .006, with those in the high cue utilisation group (*M* = 11.95, *SE* = .88) taking a similar amount of time to those in the low cue utilisation group (*M* = 12.82, *SE* = .78). The simple effect for cue utilisation for those in the high cognitive reflection group was statistically significant, *F* (1,85) = 7.08, *p* = .009, partial η^2 = .077, with those in the high cue utilisation group (*M* = 16.98, *SE* = 1.18) demonstrating longer average response latencies than those in the low cue utilisation group (*M* = 12.97, *SE* = 1.04).

The simple effect for those in the low cue utilisation group was not statistically significant, *F* (1,85) = .000, *p* = .986, partial η^2 = .000, with those in the high cognitive reflection group (*M* = 12.79, *SE* = 1.04) taking a similar amount of time to those in the low cognitive reflection group (*M* = 12.82, *SE* = .78). The simple effect for cognitive reflection for those in the high cue utilisation group was statistically significant, *F* (1,85) = 11.65, *p* = .001, partial η^2 = .121, with those in the high cognitive reflection group (*M* = 16.98, *SE* = 1.18) demonstrating longer average response latencies than those in the low cognitive reflection group (*M* = 11.95, *SE* = .88).

While participants' cognitive reflection did appear to impact their efficiency in the predicted direction (H3), cue utilisation moderated this relationship, revealing that those with high levels of cue utilisation took significantly longer than those with low cue utilisation, as well as all other participant groups. This finding does not provide support for advantages predicted among the high cue utilisation group in reducing the time taken to diagnose emails (H4) and suggests that this group were in fact the least efficient.

6.3 Types of cues used by participants

Open-text responses for each participant from the phishing diagnosis task were reviewed by two raters who established thematic labels that represented different cue types (Bengtsson, 2016; Johnston & Morrison, 2016). In line with expectations that participants would use similar cues based on previous literature (Parsons et al., 2016; Williams et al., 2018), seven main cue types were identified from the 145 text responses (see Table 4 for seven cue types).

Cue Type	Example Features
Look of the email	Spelling, formatting, punctuation
Company brand and information	If it was a recognised brand, contact information
Email sender	Whether the address matched the sender, or the address was surprising
Relevance of email content	Similar transactions in the past, or offers too good to be true
Urgent prompt for action	Link requesting prompt payment, personal information, asking to reset passwords
Perceived authenticity of embedded URLs	Whether the link(s) appeared authentic or suspicious, the rationale for clicking the links
Look of email images	Image quality of email images, graphic professionalism, quality of images fit with the brand

Table 4. Identified Cue Types and Example Features

6.3.1 Cue type frequencies and inter-rater reliability

In preparation for subsequent tests of association relating to cue type, two raters independently assigned all reported features to one of the established cue types. Cohen's κ was used to assess the level of agreement between the two raters' judgement whilst controlling for chance agreement, revealing excellent agreement, κ = 0.93 (Fleiss et al., 2003). All agreed assignments (287) were retained for the subsequent analyses (Gwet, 2001).

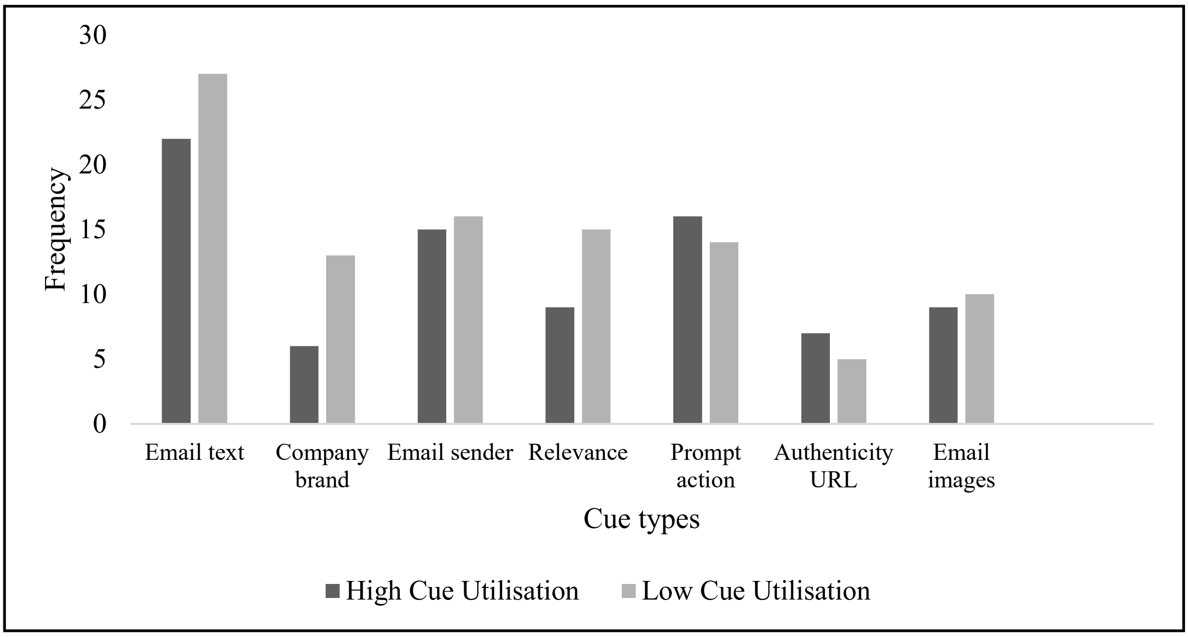

Figure 3. Frequency of type of cue used by cue utilisation group (high and low)

6.3.2 Cue type and cue utilisation

A Chi-square test of independence was used to investigate the frequencies of cues used between higher (*n* = 39) and lower cue utilisation groups (*n* = 50). Results revealed no statistically significant association between type of cue used and cue utilisation, $\chi^2(6) = 3.78$, *p* = .707. This meant that the proportion of each cue type observed in the higher and lower cue utilisation groups was not significantly different. These results can be seen in Figure 3.

7 Discussion

The study aimed to examine the role of, and potential interplay between, cue utilisation and cognitive reflection in email users' ability to accurately (and efficiently) differentiate between phishing and genuine emails.

A second aim of the study involved using an open-response format to elicit the cues used by email users when deciding whether an email is genuine or not, and test whether differences in the types of cues used existed as a function of differences in cue utilisation (i.e., whether high and low cue utilisers were found to use different cues when making their decisions).

7.1 Key Findings

7.1.1 Cue utilisation and cognitive reflection in diagnosing phishing emails

Participants high in both cue utilisation and cognitive reflection were found to perform at a similar level to those with high cue utilisation but low cognitive reflection, as well as those with low cue utilisation but high cognitive reflection (H1). However, participants with both low levels of cue utilisation and cognitive reflection performed significantly worse in diagnosing phishing emails than all other groups (H2).

Contrary to expectations, higher levels of cue utilisation did not translate to increased accuracy on the phishing diagnosis task (H1). While these results may indicate that the relationship between cue utilisation and task performance proposed here may not translate in a cyber security context, we speculate that the findings may be attributed to the poverty of cues within a phishing email compared to other 'cue-rich' domains (e.g., aviation, driving, criminal investigation). Indeed, the limited range of visual markers may restrict the development of perceptual-cognitive skills, limiting the performance advantage gained from valid cue-based associations (Brams et al., 2019; French & Nevett, 1993; Harré et al. 2012). Therefore, cue utilisation should be considered alongside other potent predictors of phishing vulnerability, in this case individual differences relating to cognitive reflection.

These results imply that the cue-performance relationship, at least in this context, appears to be somewhat contingent on individual differences relating to cognitive reflection tendencies. Indeed, taken together with the finding that a high degree of cue utilisation was associated with an *increase* in the time taken to diagnose emails among high cognitive reflectors (H3 and H4), it may be posited that the most crucial element in predicting performance is an individual's willingness to 'engage' in the diagnostic process, whether by attending to the cues that arouse suspicion, or simply by employing an analytical approach to assessment.

While there is precedent for reduced efficiency among higher cue users in other domains (Brouwers et al., 2016), we can also speculate that in the current context, EXPERTise 2.0 may not have discerned genuine differences in expertise that are attributed to varying degrees of skilled intuition. When considering the phishing context, we note Kahneman and Klein's

(2009) conception of the development of expertise and the pre-requisite of extended exposure to cue-based associations in a 'high validity' environment. Indeed, we speculate that the causal relationship between suspicious indicators and fraudulent emails may be less stable than in other contexts where skilled intuition is likely to develop and thrive. Relatedly, users' opportunities for feedback in this context are presumably limited, with most emails being valid and trustworthy communications. As such, in this context, EXPERTise 2.0 may be less sensitive to domain-specific cue utilisation differences, and more of an indication of general engagement with diagnostic tasks.

Overall, this notion of enhanced engagement remains consistent with previous work that has adopted a heuristic-systematic information processing framework in explaining phishing victimisation (Bayl-Smith et al., 2020; Frauenstein & Flowerday, 2020; Harrison et al., 2016; Luo et al., 2013; Valecha et al., 2015; Vishwanath et al., 2018; Zhang et al., 2012). Indeed, a vital factor in reducing phishing susceptibility appears to be an individual's inclination to diligently examine email features when considering its authenticity (Harrison et al., 2016; Luo et al., 2013; Vishwanath et al., 2016; Vishwanath et al., 2011). The missing piece of the puzzle may be an accurate record of users' operational experience in diagnosing phishing emails, which is expected to be both difficult to gauge and limited in variation. For instance, identifying a group possessing extensive deliberate practice in phishing diagnosis is extremely unlikely. Nevertheless, future studies should consider strategies to identify genuine experts in this field to enable an investigation of operational experience as an additional moderator in the cue utilisation, cognitive reflection, and phishing performance relationship.

In terms of practical implications for organisations, identifying individuals who are less likely to engage in an assessment of an email's validity, either via the CRT or EXPERTise 2.0, may be useful in the implementation of targeted cyber-security training and education interventions. Importantly, such programs would need to be intensive and ongoing, with some findings showing the transient nature of behavioural changes among emails users who habitually rely on heuristics (Canova et al., 2014; Vishwanath, 2015). Further, workplaces may benefit from work design interventions that recognise the importance of workers' conscious attention during email processing (e.g., dedicated email times without concurrent tasks).

7.1.2 Cue types and cue utilisation

In line with expectations, participants reported using cue types largely consistent with previous literature (Parsons et al., 2016; Williams et al., 2018). Parsons et al. (2016) identified cues such as consistent logos (e.g., company brand), links, personalisation, and spelling cues. The current study found cues relating to inconsistencies within the text of the email (e.g., spelling, formatting, and punctuation), perceived authenticity of embedded URLs (e.g., if the links appeared authentic) and company brand. Interestingly, an "urgent prompt for action" cue was identified by participants, which was conceptually similar to the "urgency" cue/influence technique found in Williams et al. (2018).

The current study explored cue types further, by analysing if they differed significantly among different levels of cue user (i.e., high and low). However, cue type did not significantly differ between different levels of cue utilisation, suggesting that participants were not using significantly different cues during decision making. This finding is noteworthy, as previous research has advocated for the use of critical cue inventories as the basis for training interventions (Morrison et al., 2013; Wiggins & O'Hare, 2003). Similarly, many cyber-experts have promoted a rule-based approach to phishing education (Basnet et al., 2012; Moghimi &

Varjani, 2016). Critical cue inventories and rule-based approaches place an emphasis on identifying cues as the primary factor of good decision-making. However, the disparity in the current study's findings suggests that simply having an awareness for said cues does not appear to determine performance on its own. Instead, it appears that knowledge may be a pre-requisite alongside a willingness to engage in a process of assessment. Thus, while presumably useful, the effectiveness of interventions centred on the learning of critical cue inventories may be largely moderated by other attributes of individual learners.

It must be noted that the choice to survey respondents on the cues used across all emails makes it difficult to identify specific predictors of performance. It was decided that surveying of cue usage after each email, although ideal, would be too onerous on participants considering the time commitment associated with participation in the experiment. Future research could consider more 'online' measures of cue utilisation, such as the use of eye-tracking.

7.2 Limitations

The primary limitation of the research approach was the artificial nature of the phishing diagnostic performance measure, which may have yielded several experimental artefacts. For instance, the fact that participants were advised that the study involved detecting phishing emails prior to participation may have elicited expectation effects not present during real-world operation (see Levine's Truth-Default Theory; 2014). At worst, such a process may inadvertently prime some participants to engage more analytical, System 2 processes during decision-making, which may confound the true effects of users' cognitive reflection tendencies. Work is underway to investigate the impact of differences in the ratio of trustworthy to suspicious emails in controlled phishing experiments. It is possible that participants' will anticipate an inflated number of phishing emails in experimental settings, which would potentially impact their precision in detecting phishing threats (see Canfield et al., 2016). We also intend to test the generalisability of our results beyond controlled designs via the use of naturalistic study techniques, such as the use of simulated phishing emails delivered to participants' actual inboxes at infrequent times.

8 Conclusion

The current study presents several novel findings. First, results revealed that participants with relatively low cue utilisation and cognitive reflection were found to be significantly more vulnerable to deceptive phishing emails, and that performance could be greatly improved with an elevation in either quality. Next, a number of phishing email cues were identified by participants, however, the types of cues utilised did not differentiate between high and low cue users. These findings extend on a raft of others purporting the significance of cue utilisation in diagnostic performance, this time in a cyber security context, but also uncover an important moderating effect based on individual differences in email users' cognitive reflection tendencies, which has not been established previously. Taken together, the findings have implications to the design of user-level interventions in relation to the identification of vulnerable users, as well as the need to consider training approaches that extend beyond the use of simple cue inventories.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American psychologist, 51*(4), 355-365. https://doi:10.1037/0003-066X.51.4.355

Basnet R.B., Sung A.H., & Liu Q. (2012) Feature Selection for Improved Phishing Detection. In: Jiang H., Ding W., Ali M., & Wu X. (eds) *Advanced Research in Applied Artificial Intelligence. IEA/AIE 2012*. Lecture Notes in Computer Science, vol 7345. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-310

Bayl-Smith, P., Sturman, D., & Wiggins, M. (2020). Cue utilization, phishing feature and phishing email detection. In M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne, & M. Sala (Eds.), Financial Cryptography and Data Security: FC 2020 International Workshops, AsiaUSEC, CoDeFi, VOTING, and WTSC, Revised Selected Papers (pp. 56-70). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12063). Springer. https://doi.org/10.1007/978-3-030-54455-3_5

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *Nursing Plus Open, 2*, 8-14. https/doi:10.1016/j.npls.2016.01.001

Brams, S., Ziv, G., Levin, O., Spitz, J., Wagemans, J., Williams, A. M., & Helsen, W. F. (2019). The relationship between gaze behavior, expertise, and performance: A systematic review. *Psychological Bulletin, 145*(10), 980–1027. https://doi.org/10.1037/bul0000207

Brouwers, S., Wiggins, M., & Griffin, B. (2018). Operators who readily acquire patterns and cues, risk being miscued in routinized settings. *Journal of Experimental Psychology: Applied, 24*(2), 261-274. https://doi:10.1037/xap0000151

Butavicius, M., Parsons, K., Pattison M., & McCormac, A. (2016). *Breaching the Human Firewall: Social engineering inPphishing and Spear-Phishing Emails*. ArXiv, abs/1606.00887.

Canfield, C. I., Fischhoff, B., & Davis, A. (2016). Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 58*(8), 1158–1172. https://doi.org/10.1177/0018720816665025

deGroot, A. D. (1978). *Thought and choice in chess*. The Hague: Mouton

Downs, J. S., Holbrook, M. B., & Cranor, L. F. (2006). Decision strategies and susceptibility to phishing. In *Proceedings of the second symposium on Usable privacy and security* (pp. 79-90). ACM. https://doi:10.1145/1143120.1143131

Fette, I., Sadeh, N., & Tomasic, A. (2007). Learning to detect phishing emails. In *WWW '07 Proceedings of the 16th international conference on World Wide Web* (pp. 649-656). New York: ACM. https://doi:10.21236/ada456046

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London, England: Sage.

Fleiss, J. L., Levin, B., & Paik, M. C. (2013). *Statistical methods for rates and proportions.* Hoboken: John Wiley & Sons.

Frauenstein, E. D., & Flowerday, S. (2020). Susceptibility to phishing on social network sites: A personality information processing model. *Computers & Security, 94*, 101862. https://doi.org/10.1016/j.cose.2020.101862

Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives, 19*(4), 25-42. https://doi.org/10.1257/089533005775196732

French, K. E., & Nevett, M. E. (1993). The Development of Expertise in Youth Sport. In J. L. Starkes & F. Allard (Eds.), *Advances in Psychology* (pp. 255-270). North-Holland. https://doi.org/https://doi.org/10.1016/S0166-4115(08)61475-2

Gacasan, E. M. P., Wiggins, M. W., & Searle, B. J. (2016). The role of cues in expert project manager sensemaking. *Construction Management and Economics, 34*(7-8), 492-507. https://doi:10.1080/01446193.2016.1177190

Galanter, C. A., & Patel, V. L. (2005) Medical decision making: A selective review for child psychiatrists and psychologists. *Journal of Child Psychology and Psychiatry, 46*(7), 675-689. https://doi:10.1111/j.1469-7610.2005.01452.x

Gonzalez, C. (2013). The boundaries of instance-based learning theory for explaining decisions from experience. In V. S. C. Pammi & N. Srinivasan (Eds.), *Progress in brain research* (Vol. 202, pp. 73-98). Oxford, UK: Elsevier.

Gwet, K. (2001). *Handbook of inter-rater reliability.* Gaithersburg: STATAXIS Publishing Company.

Harrison, B., Svetieva, E., & Vishwanath, A. (2016). Individual processing of phishing emails. *Online Information Review, 40*(2), 265–281. https://doi.org/10.1108/oir-04-2015-0106

Harré, M., Bossomaier, T., & Snyder, A. (2012). The perceptual cues that reshape expert reasoning. *Scientific Reports, 2*(1), 502–502. https://doi.org/10.1038/srep00502

Hassandoust, F., Singh, H., & Williams, J. (2020). The Role of Contextualization in Individuals' Vulnerability to Phishing Attempts. *Australasian Journal of Information Systems, 24*. DOI: https://doi.org/10.3127/ajis.v24i0.2693

Johnston, D., & Morrison, B. W. (2016). The application of naturalistic decision-making techniques to explore cue use in rugby league playmakers. *Journal of Cognitive Engineering and Decision Making, 10*(4), 391-410. https://doi:10.1177/1555343416662181

Jones, H.S., Towse, J. N., & Race, N. (2015) Susceptibility to email fraud: A review of psychological perspectives, data-collection methods, and ethical considerations. *International Journal of Cyber Behaviour, Psychology and Learning, 5*(3). 13-29. https://doi:10.4018/IJCBPL.2015070102

Jones, H.S., Towse, J. N., Race, N., & Harrison, T. (2019). Email fraud: The search for psychological predictors of susceptibility. *PloS ONE 14*(1), e0209684. https://doi:10.1371/journal.pone.0209684

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American psychologist, 64*(6), 515-526. https://doi:10.1037/a0016755

Klein, G. A. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid Decision making on the Fire Ground. *Proceedings of the Human Factors Society Annual Meeting, 30*(6), 576-580.

Kobus, D. A., Proctor, S., & Bank, T. E. (2000). *Decision-making in a dynamic environment: the effects of experience and information uncertainty*. Technical Report 1832. San Diego, CA: Spawar Systems Center.

Levine, T. R. (2014). Truth-Default Theory (TDT). *Journal of Language and Social Psychology, 33*(4), 378–392. https://doi:10.1177/0261927x14535916

Loveday, T., Wiggins, M., Festa, M., Schell D., & Twigg, D. (2013). Pattern recognition as an indicator of diagnostic expertise. In C. P. Latorre & F. A. Sanchez (Eds.), *Pattern recognition – Applications and methods* (pp. 1-11). Berlin: Springer.

Luo, X., Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating phishing victimization with the Heuristic-systematic model: a theoretical framework and an exploration. *Computer Security, 38*, 28–38. https://doi.org/10.1016/j.cose.2012.12.003

Moghimi, M., & Varjani, A. Y. (2016). New rule-based phishing detection method. *Expert Systems with Applications, 53*, 231–242. https://doi:10.1016/j.eswa.2016.01.028

Morrison, B. W., Johnston, D., Naylor, M., Morrison, N. M. V., & Forrest, D. (2020). "You can't hide your lyin' eyes": investigating the relationship between associative learning, cue awareness, and decision performance in detecting lies. *Journal of Cognitive Engineering and Decision Making, 14*(2), 99-111. https://doi.org/10.1177/1555343420918084

Morrison, B., & Morrison, N. (2015). Diagnostic cues in major crime investigation. In M. W. Wiggins, & T. Loveday (Eds.), *Diagnostic expertise in organizational environments* (pp. 91-98). Ashgate Publishing

Morrison, B. M., Wiggins, M. W., Bond N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making, 7*(2), 211-226. https://doi:1177/1555343412459192

Morrison, B. W., Wiggins , M. W., & Morrison, N. V. (2018). Utility of expert cue exposure as a mechanism to improve decision-making performance among novice criminal investigators. *Journal of Cognitive Engineering and Decision Making, 12*(2), 99-111. https://doi:10.1177/1555343417746570

Nasser, G., Morrison, B. W., Bayl-Smith, P., Taib, R., Gayed, M., & Wiggins, M. W. (2020a). The effects of cue utilization and cognitive load in the detection of phishing emails. In AsiaUSEC'20: proceedings of the Workshop on Usable Security (pp. 1-10). Malaysia: Springer.

Nasser, G., Morrison, B. W., Bayl-Smith, P., Taib, R., Gayed, M., & Wiggins, M. W. (2020b). The Role of Cue Utilization and Cognitive Load in the Recognition of Phishing Emails. *Frontiers in big data, 3*, 546860. https://doi.org/10.3389/fdata.2020.546860

Parsons, K., Butavicius, M., Pattinson, M., McCormac, A., Calic, D., & Jerram, C. (2016). *Do users focus on the correct cues to differentiate between phishing and genuine emails?* arXiv preprint arXiv:1605.04717.

Qualtrics (Version 2019). [Web-based software]. Provo, UT: Qualtrics. Available from http://www.qualtrics.com.

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research, 136*(2), 253-263. https://doi:10.1016/S0377-2217(01)00113-8

Shekh, S., Auton, J. C., & Wiggins, M. W. (2018). The effects of cue utilization and target-related information on target detection during a simulated drone search and rescue task. *Proceedings of the Human Factor and Ergonomics Society Annual Meeting, 62*(1), 227-231. https://doi:10.1177/1541931218621053

Shonman, M., Li, X., Zhang, H., & Dahbura, A. (2018). Simulating phishing email processing with instance-based learning and cognitive chunk activation. In S. Wang, V. Yamamoto, J. Su, Y. Yang, E. Jones, L Iasemidis & T. Mitchell (Eds.), *Lecture Notes in Computer Science: Vol 11309. Brain Informatics* (pp. 468-478). Cham: Springer. https://doi:10.1007/978-3-030-05587-5_44

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645-665. https://doi:10.1017/S0140525X00003435

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*(3), 430-454. https://doi.org/10.1016/0010-0285(72)90016-3.

Tversky, A., & Kahneman, D. (1974). Heuristics and biases: Judgement under uncertainty. *Science, 185*(1974), 1124-1130. https://doi:10.1126/science.185.4157.1124

Tversky A., & Kahneman D. (1975) Judgment under Uncertainty: Heuristics and Biases. In: Wendt D., Vlek C. (eds) Utility, Probability, and Human Decision Making. Theory and Decision Library (An International Series in the Philosophy and Methodology of the

Social and Behavioral Sciences), vol 11. Springer, Dordrecht. https://doi.org/10.1007/978-94-010-1834-0_8

Vishwanath, A. (2015). Habitual Facebook Use and its Impact on Getting Deceived on Social Media. *Journal of Computer-Mediated Communication, 20*(1), 83-98

Vishwanath, A., Harrison, B., & Ng, Y.J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Commun. Res.* 1–21. https://doi.org/10.1177/0093650215627483

Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, Cognition, and Automaticity Model of Phishing Susceptibility. *Communication Research, 45*(8), 1146–1166. https://doi.org/10.1177/0093650215627483

Vishwanath, A., Herath, T., Chen, R., Wang, J., & Rao, H.R. (2011), "Why do people get phished? Testing individual differences in phishing vulnerability within an integrated information processing model". *Decision Support Systems, 51*(3), 576-586.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology*, 24, 33-49. https://doi:10.1007/s10869-009-9090-8

Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology, 57*(9), 681-687. https://doi:10.1080/14992027.2018.1476782

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 45*(1), 104-116. https://doi:10.1518/hfes.45.1.104.27233

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance.* New York: Psychology Press. https://doi:10.4324/9781315665177

Wiggins, M. W. (2021). A behaviour-based approach to the assessment of cue utilisation: implications for situation assessment and performance. *Theoretical Issues in Ergonomics Science, 22*(1), 46-62. https://doi.org/10.1080/1463922X.2020.1758828

Wiggins, M. W. (2016). Expertise and cognitive skills development for ab-initio pilots. In R. A. Telfer & P. J. Moore (Eds.), *Aviation training: Learners, instruction and organization* (pp 54-66). Abington, Oxon: Routledge.

Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014). Trait-based cue utilization and initial skill acquisition: implications for models of the progression to expertise. *Frontiers in Psychology, 5,* 541. https://doi:10.3389/fpsyg.2014.00541

Wiggins, M. W., Crane, M., & Loveday, T. (2018). Cue utilization, perceptions, and experience in the interpretation of weather radar returns. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62*(1), 721-725. https://doi:10.1177/1541931218621164

Wiggins, M, W., Griffins, B., & Brouwers, S. (2019). The potential role of context-related exposure in explaining differences in water safety cue utilization. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 61*(5), 825-838. https://doi:10.1177/0018720818814299

Wiggins, M., Loveday, T., & Lyons, L. (2014). Cues and cue-based processing: Implications for system safety. *Procedia Engineering, 84*, 55-61. https://doi:10.1016/j.proeng.2014.10.409

Wiggins, M. W., Whincup, E., & Auton, J. C. (2018). Cue utilisation reduces effort but increases arousal during a process control task. *Applied Ergonomics, 69*, 120-127. https://doi:10.1016/j.apergo.2018.01.012

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring susceptibility to phishing in the workplace. *International Journal of Human-Computer Studies, 120*, 1-13. https://doi:10.1016/j.ijhcs.2018.06.004

Yan, Z., & Gozu, H. Y. (2012). Online decision-making in receiving spam emails among college students. *International Journal of Cyber Behavior, Psychology and Learning, 2*(1), 1-12. https://doi:10.4018/ijcbpl.2012010101

doi: https://doi.org/10.3127/ajis.v26i0.3615