The Role of Cue Utilisation and Background Office Noise on Phishing Email Detection

[redacted]

This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of Psychological Science (Honours)

School of Psychology

The University of Adelaide

September 2023

Wordcount: 6934

Table of Contents

List of Figures

List of Tables

Abstract

Phishing email attacks have increased significantly in recent years, resulting in losses over $200 million for both Australian individuals and organisations in 2022. In order to reduce this risk, individuals need to be able to effectively identify phishing emails. While previous research has shown that cue utilisation improves phishing email detection, none have examined cue utilisation within noisy work environments, where email communication is highly prevalent. Understanding how situational factors like office noise impact phishing detection is crucial for the protection of an organisation's cybersecurity. The aim of the present study was to explore the relationship between cue utilisation and background office noise on an individual's ability to differentiate phishing from genuine emails. In a lab-based study, a sample of Australian residents (N = 49) completed two tasks in which they were required to sort a series of emails (50% phishing) into one of ten categories. Participants completed one task under simulated background office noise and completed the other task with no noise. Cue utilisation was measured using the cybersecurity edition of EXPERTise 2.0 situational judgement test. Consistent with past findings, this study found that higher levels of cue utilisation were associated with a greater ability to differentiate phishing from genuine emails. Participants performance was not affected by background office noise, however, background noise resulted in greater perceived cognitive load. Overall, the findings in this study may aid in the development of education and training programs to enhance phishing detection.

Keywords: Cue Utilisation, Phishing email detection, cybersecurity, office noise, EXPERTise

Declaration

"This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time."

[redacted]

September 2023

Contributor Roles Table

ROLE	ROLE DESCRIPTION	STUDENT	STUDENT	SUPERVISOR 1	SUPERVISOR 2
CONCEPTUALIZATION	Ideas; formulation or evolution of overarching research goals and aims.	x	x	x	x
METHODOLOGY	Development or design of methodology; creation of models.	x	x	x	x
PROJECT ADMINISTRATION	Management and coordination responsibility for the research activity planning and execution.	x	x	x	x
SUPERVISION	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.			x	x
RESOURCES	Provision of study materials, laboratory samples, instrumentation, computing resources, or other analysis tools.			x	x
SOFTWARE	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code.	x	x	x	x
INVESTIGATION	Conducting research - specifically performing experiments, or data/evidence collection.	x	x		
VALIDATION	Verification of the overall replication/reproducibility of results/experiments.			x	x
DATA CURATION	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for	x		x	x

The Role of Cue Utilisation and Background Office Noise on Phishing Email Detection

Phishing is a fraudulent practice in which cybercriminals disguise themselves as trustworthy entities or institutions to obtain sensitive information from online users (Thapa et al., 2023; Williams et al., 2018). Most commonly used via email, phishers trick victims into clicking on malicious URL links or attachments that request sensitive information (e.g. credentials, passwords), demand direct payments and/or illicitly install malware on their devices (Ackerley et al., 2022; Butavicius et al., 2016). Phishing emails are still one of the most effective forms of cybercrime, exploiting the vulnerabilities of individuals, governments and organisations (Ayaburi et al., 2019; Sturman et al., 2023; Zhuo et al., 2022).

Globally, the annual cost of cybercrime is rising, and is predicted to cost organisations US$8 trillion in 2023 and grow to US$10.5 trillion by 2025 (Morgan, 2020). In Australia, the financial damage of successful phishing email attacks has increased, costing organisations over $98 million in 2022 compared to $81.45 million in 2021 (Australian Cyber Security Centre [ACSC], 2022, 2021). In addition to financial losses, phishing attacks cause reputation damage, decrease productivity, and revenue loss to organisations (Alkhalil et al., 2021; Vishwanath et al., 2011).

Many organisations have invested in technological solutions to block and filter phishing emails; however, these solutions do not guarantee 100 percent protection (Buckley et al., 2023; Slifkin & Neider, 2022). Therefore, individuals are the last line of defence once phishing emails reach their inboxes, and their ability to identify phishing emails is detrimental for the protection of organisations (Weaver et al., 2021; Williams et al., 2023). Almost all cybersecurity breaches are caused by human error, with a report from IBM Cybersecurity Intelligence Index (2014) finding that 95% of corporate breaches were caused by employees clicking on malicious links and attachments.

Since human error constitutes the majority of weaknesses in the security chain, phishing research has begun examining psychological factors that can be targeted for improvement, such as how an email is cognitively processed (Musuva et al., 2019; Vishwanath et al., 2011). Understanding how individuals make judgements to determine email legitimacy can help the development and implementation of effective phishing prevention programs (Alkhailel et al., 2021; Zhou et al., 2022).

Cue-based Processing and Phishing Email Detection

Dual-system theories of reasoning propose that individuals' engage in one of two processes when determining an email's legitimacy (Kahneman, 2003). System 1 processing is characterised as rapid, automatic, and intuitive, while System 2 processing involves more slow, deliberate and analytical decision-making. In situations where individuals are under time pressure and high cognitive demands, they often rely on System 1 processing, utilising rules-of-thumb and environmental cues to make fast decisions (Kahneman, 2003; Tversky & Kahneman, 1974).

Cues comprise learnt associations between environmental features, events or objects (Brunswik, 1955; Klein, 2008). Cues are formed by the interaction between an individual and their environment, where the individual identifies a relationship of a causal factor (feature) with a situation (object/event) in memory (Brunswik, 1955; Klein, 2008). For example, when appraising an email, individuals might associate a 'Suspicious looking URL' (feature) with a 'phishing email'(event). Through repeated exposure of feature-event relationships, cues are established, strengthened, and retained in long-term memory (Ericsson & Lehmann, 1996; Wiggins et al., 2014). Once established in long-term memory, cues can become rapidly and unconsciously activated and subsequently influence decision-making (Wiggins et al., 2021).

Phishing emails are designed to encourage System 1 processing, provoking individuals to make quick decisions rather than deliberating (Goel et al., 2017). For example,

some phishing emails create a sense of urgency, claiming that an individual's 'account will be deactivated' or urging them to 'pay immediately' (Musuva et al., 2019). This sense of urgency leads individuals to act quickly, resulting in them failing to recognise a phishing email, even when the email contains cues that indicate phishing (Vishwanath et al., 2018). Additionally, phishing emails have advanced in sophistication and personalisation, posing a greater challenge for individuals to identify them while using System 1 modes of thinking (Rajivan & Gonzalez, 2018). This may be relevant in workplace settings, as employees are often busy juggling multiple tasks and urgent deadlines, potentially increasing their reliance on System 1 processing (Buckley et al., 2023; Kahneman, 2003).

Although System 1 processing has been associated with phishing susceptibility (Musuva et al., 2019, Vishwanath et al., 2011; 2018), the ability to identify diagnostic cues can aid in phishing email detection (Nasser et al., 2020). Theoretically, relying on cues in an environment is a fundamental component for effective decision-making, and skilled performance (Wiggins et al., 2021). Expert-novice paradigms propose that expertise and skilled performance is established by individuals utilising cues in their environment to make rapid and accurate conclusions (Brunswik, 1955; Klein, 2008). Through experience and extensive exposure, cues trigger mental representations in long-term memory, thereby reducing the amount of information that needs processing in the environment (Brouwers et al., 2017; Wiggins, 2021). This allows for more residual resources in working memory to be allocated for higher level processing and optimal decision-making (Ericsson & Lehmann, 1996; Wiggins et al., 2014). As the rapid activation of cues imposes fewer demands on working memory, this enables experts to perform tasks with high levels of accuracy (Wiggins, 2014; 2021).

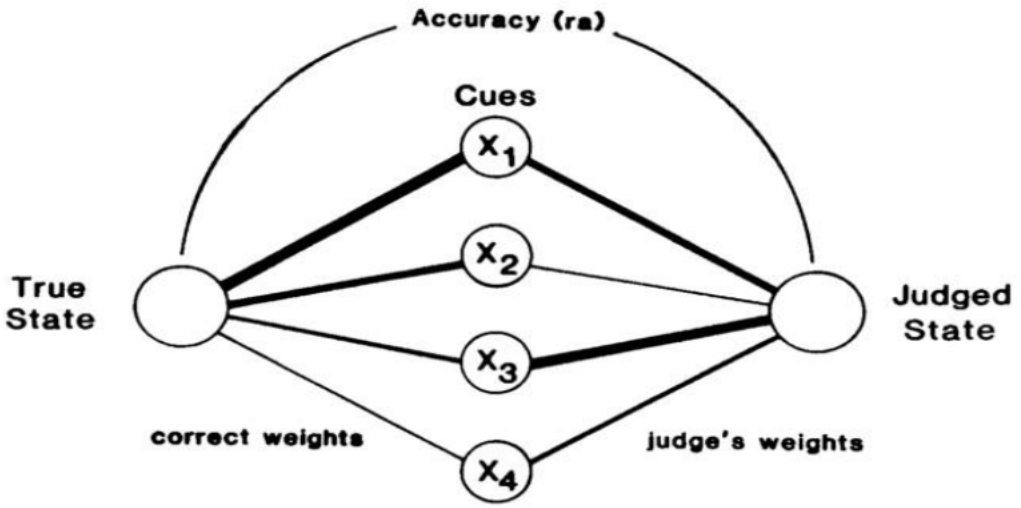
The Brunswik (1955) lens model offers a conceptual framework to understand how individuals can utilise cues to judge the legitimacy of emails. According to Brunswik (1955)

individuals do not have direct access to the 'true state' of an email, therefore individuals must rely on a ray of cues in the email (e.g., senders address, URL link) that are available to them to determine the email's legitimacy. As shown in figure 1, the thickness of the rays connecting the observer's judged state to the cues represents how much the individual weighs and interprets the cues to be indicative of the 'true state' of the email. For example, if an individual evaluates a cue, such as a suspicious looking URL as a strong indicator of a phishing email, an individual may give it substantial weight when judging the email's legitimacy.

Each cue provides a piece of information of the 'true state' of an email, with some cues providing more valid predictions of whether the email is a phishing or genuine (Koehler & Harvey, 2008). This is represented by the thickness of the rays connecting the cues to the 'true state' of the email (see figure 1). The thicker rays indicate the validity and reliability of the cues in predicting the 'true state' of the email (phishing or genuine). As cues are probabilistic in nature, an individual's ability to combine and rely on more valid cues and less on ambiguous cues leads to more accurate judgements on an email's legitimacy (Mosier & Kirlik, 2004).

Figure 1

A schematic illustration of Brunswik's Lens Model that conceptualises the process of intuitive judgements (Wigton, Hoellerich, & Patil, 1986).



Individual differences in the capacity to attain, recognise and apply cues is referred to as cue utilisation (Wiggins et al., 2014). In comparison to their less experienced counterparts, experts have been shown to have greater capacity of cue utilisation (Wiggins, 2014). This ability is demonstrated through the rapid identification and recognition of meaningful information from cues in a situation, the capability to differentiate relevant from less relevant cues, and the ability to apply and prioritise critical cues to make high-order evaluations of a situation (Brouwers et al., 2017; Wiggins, 2014; 2021). To measure an individuals' capacity for cue utilisation, researchers have utilised an online situation judgement test: EXPERT Intensive Skills Evaluation (EXPERTise 2.0), which has been validated in many domains such as cybersecurity (Sturman et al., 2023) driving (Yuris et al., 2019), air traffic control (Falkland & Wiggins, 2019), and medicine (Crane et al., 2018). Within these domains, an individual's capacity for higher cue utilisation has been associated with greater objective

performance (Crane et al., 2018; Falkland & Wiggins, 2019; Sturman et al., 2023; Yuris et al., 2019).

Some existing literature suggests an individual's capacity for higher cue utilisation is associated with greater accuracy in phishing detection (Ackerley et al., 2022; Nasser et al., 2020; Williams et al., 2023), and an increased ability to identify cues that are indicative of phishing emails (Baly-Smith et al., 2020). These findings support the theoretical notion that individuals with higher cue utilisation abilities can identify key features of relevance within their environment to make accurate decisions (Ackerley et al., 2022; Bayl-Smith et al., 2020).

Although these studies have provided valuable insights to the significance of cue-based processing, none have examined the impact of the environment on phishing detection. Most of these studies have been conducted in quiet and controlled lab-based settings, which may not reflect real-world conditions (Slifkin & Neider, 2022). Since individuals do not process information in perfect environments, it is important to consider situational factors that influence decision-making (Buckley et al., 2023). Despite the increasingly vulnerability of phishing attacks within organisations, less is known about the influence of specific aspects of the work environment on phishing susceptibility (Buckley et al., 2023; Williams et al., 2018). While some studies have examined the workplace factors on phishing detection (Butavicius et al., 2022; Vishwanath et al., 2018), none to date have explored the aspects of the physical environment, such as workplace office noise on phishing susceptibility.

Cue Utilisation and Office Noise

Problems experienced in offices has been widely documented, indicating that employees often find themselves distracted from the background environment while performing demanding tasks (Jahncke & Hallman, 2020). Distractions such as noise can manifest in many forms, however, printers, typing of keyboards, phones ringing and colleagues talking in the background are considered highly disruptive in office environments

(Banbury & Berry, 2005; Brennan et al., 2002; Brocolini et al., 2016). Emerging evidence has found office noise can exert different effects on decision-making and negatively impact task performance (Bronolini et al 2016; Hygge & Knez, 2001). For example, Hygge and Knez (2001) found that employees tended to work faster in noisy environments, resulting in reduced accuracy on memory tasks. Yadav et al. (2017) and Jahncke and Hallman (2020) also found office noise was associated with decrements in performance on cognitive and short-term memory tasks.

The effects of workplace noise on employees decline in performance can be attributed to the additional cognitive demands placed on working memory resources (Biondi et al., 2022; Sweller, 1998). As working memory has a limited capacity, background noise competes for the same working memory resources as those required for the primary task (e.g. checking an email's legitimacy). As attention is moved from the primary task to the distraction (e.g., noise), there are fewer residual resources allocated to the task at hand, therefore reducing an individual's capacity for higher-order information processing (Ericsson & Lehmann, 1996; Falkland et al., 2020; Yadav et al., 2017). As higher cognitive demands has been found to be a contributing factor associated with errors in performance, it is plausible that background office noise may increase phishing susceptibility (Buckley et al., 2023; Williams et al., 2018).

The Current Study

The aim of the current study was to examine the relationship of cue utilisation and background office noise on an individual's ability to differentiate phishing from genuine emails. Using a lab-based design, participants were required to complete two Email Sorting Tasks containing a series of phishing and genuine emails. During one of the Email Sorting Tasks, participants were required to listen to simulated background office noise. Cue utilisation was operationalised using an online software package (EXPERTise 2.0; Wiggins et

al., 2014), which assessed participants behaviours that are indicative with the utilisation of cues within the domain of cybersecurity.

H1. In comparison to participants with lower cue utilisation, participants with higher cue utilisation would be better able to differentiate phishing from genuine emails.

H2. Compared to the background office noise condition, participants will be better able to differentiate phishing from genuine emails in the no background office noise condition.

H3. The background office noise condition would result in an increase in the number of errors in differentiating phishing from genuine emails, although an interaction is hypothesised with lower cue utilisation associated with a greater loss of performance compared to those with higher cue utilisation.

Method

Participants

Participants comprised 49 Australians (30 females, 19 males) ranging in age from 17 to 51 years (M =20.78, SD =5.75). Thirty-eight first-year students were recruited from the University of Adelaide Psychology subject pool, via the university's SONA research participation system. Students received a one hour of course credit in exchange for their participation. Additionally, 11 participants from the general public were recruited using snowballing and convenience sampling. These participants received a $20 Coles Myer giftcard in exchange for their participation in the study. Participants reported receiving on average 10.84 emails per day (SD =8.5), spending 1.13 hours (SD =1.84) reading and responding to emails per day and reported spending 5.9 hours per day using a computer (SD =2.64). In order to participate, participants were required to be fluent in English.

Materials

The current study employed a lab-based design conducted on a computer using Qualtrics. Participants were required to complete a series of tasks which included demographic questions, two Email Sorting Tasks and two manipulation checks. After, participants completed the EXPERTise 2.0 - cybersecurity edition on a separate platform.

Demographics Questionnaire

Participants completed a series of demographic questions including their age, gender, confidence with computer use, how many hours spent using a computer a day, the devices used to access their emails, the number of hours spent reading and responding to emails per day and the number of emails received per day.

Email Sorting Tasks

Two Email Sorting Tasks were designed to measure participants performance ability in differentiating between phishing from genuine emails. Participants were instructed to roleplay as a research assistant for a fictitious character name "Professor Alex Jones". In this role, participants were asked to examine 40 of Prof. Alex Jones emails, which were presented to them individually for period of 10s. After 10s, the page automatically progressed and participants were required to sort the email into one of ten predefined categories ('Urgent', 'Teaching', 'Research', 'Banking', 'Online purchases', 'Social media accounts', 'Official', 'Spam', 'Phishing', 'Miscellaneous'). Prior to beginning the task, participants were notified that 50% of the emails were phishing emails, with phishing emails defined as "fraudulent, fake or otherwise deceptive message designed to trick a person into revealing sensitive information". Participants then were provided with task instructions and a practice trial.

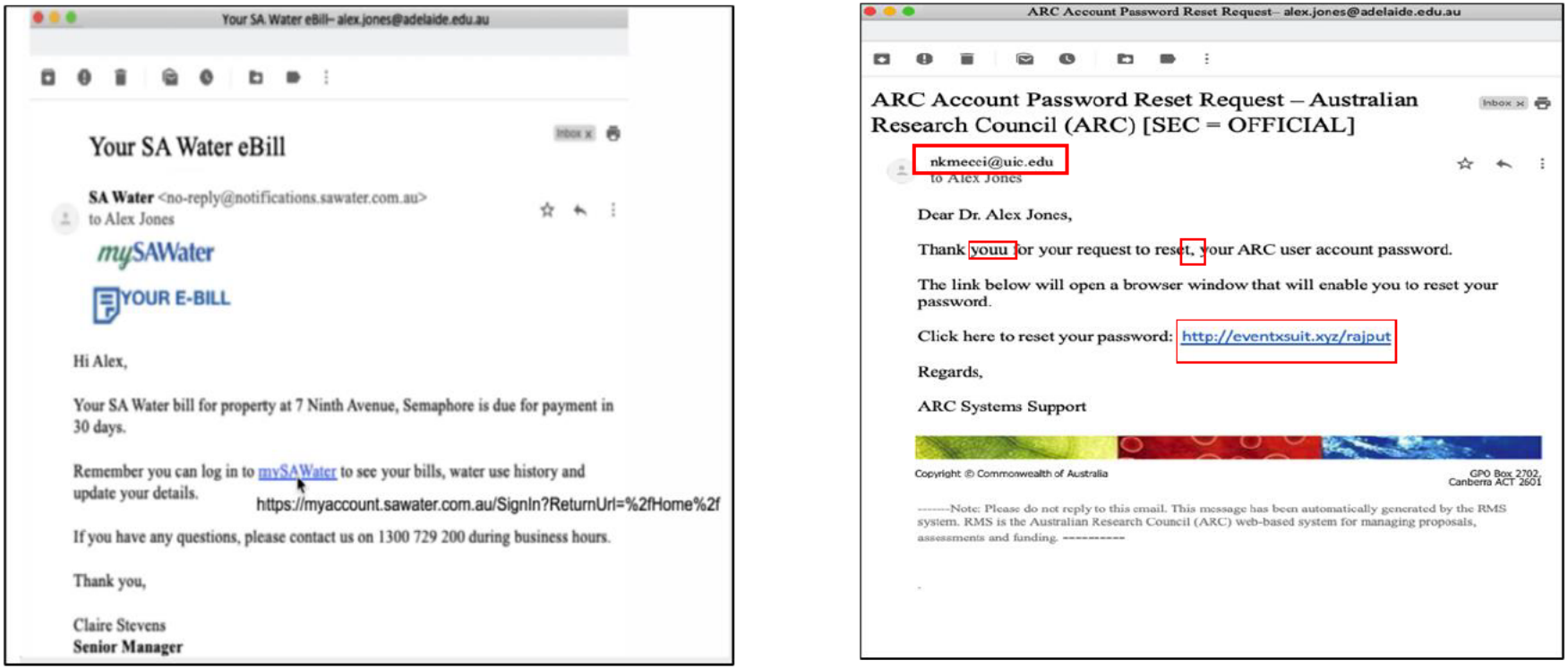
The email stimuli was developed from genuine emails that were received by the researchers of this study. All original emails utilised in the study were genuine emails (e.g. not phishing emails), included a URL link or a 'click' button and comprised 100 words or less

to ensure participants had adequate time to assess and appraise each email. Half of the 40 emails were modified systematically to create stimulated phishing emails that contained three features commonly associated with phishing emails: two spelling/ grammatical errors in the first line (e.g. “Thank you for enrolling too vote or updating your, details”), an illegitimate email senders address (e.g., nkmecci@uci.edu; see figure 2), and the legitimate URL link contained in the original email was replaced with a suspicious URL link obtained from a previously known phishing email. The URL links were either displayed in the text of the email or embedded in a prompt button (e.g., ‘Verify here’). The participants were told that the URL links could be viewed by hovering the mouse over the link or prompt button.

All emails were addressed to Alex Jones, containing both personal and work-related emails (e.g., see Figure 2 for an example of a genuine (personal) email and a phishing (work-related) email). Participants completed two Email Sorting Tasks, each containing 20 emails: 10 phishing and 10 genuine emails. Further, within each Email Sorting Task, the order of the emails were randomised.

Figure 2

An example of two of the email stimuli within the Email Sorting Task. A genuine (personal) email is presented on the left, displaying the hover feature. A phishing (work) email is presented on the right.



Note. The Phishing email (right) includes three phishing features highlighted in the red boxes: illegitimate sender address, two spelling and grammatical errors within the first line of text, and a suspicious URL link. The red boxes were not included in the study but were presented in the figure for clarity.

Office Noise Recording

To examine the impact of background office noise on participant's ability to differentiate genuine from phishing emails, an office noise recording was employed to simulate a typical office environment. The office noise recording comprised office sounds encountered in real office workplaces, including typing of keyboards, printer and fax machines, telephones ringing, and conversations between workers. The office noise was selected based on prior research, which have reported these types of office sounds to be most disruptive on participants when performing tasks (Banbury & Berry, 2005; Brennan et al., 2002; Brocolini et al., 2016; Jahncke & Hallman, 2020).

To create the noise recording, a sound clip obtained from https://www.youtube.com/watch?v=9y0wmSgRNRk was used, consisting of office noises mentioned above. Further, four voices were recorded over the sound clip to mimic colleagues talking to each other. Topics that were discussed between the four voices included work meetings, fire drills, broken work machinery and coffee orders. The noise recording was played on a loudspeaker, with the sound ranging from 60 to 65 dB (A), typical sound level of an open-plan office (Witterseh et al., 2004). The loudspeaker was placed on the desk next to the computer where participants performed the experiment.

In this study, the office noise recording was utilised during one of the two Email Sorting Tasks. To minimise order, practice and fatigue effects, the order of the two Email Sorting Tasks and the office noise recording were randomly counterbalanced across participants. Before starting the Email Sorting Task using the background office noise recording, participants were informed by the researcher that their environment would stimulate an office work environment. When completing the other Email Sorting Task, the background office noise recording was not used, and participants were informed that their office environment simulated a private office.

Manipulation Checks

To measure whether participants were more distracted and experienced greater cognitive load when completing the Email Sorting Task under the background office noise condition, two self-report measures were administered to measure subjective distraction and cognitive load. Immediately after completing each of the Email Sorting Tasks, participants were required to respond on a distraction scale, consisting of a single item question asking, "I was distracted during this Email Sorting Task" and were required to rate how distracted they were on a 6-point Likert Scale from 1 (*Strongly disagree)* to 6 (*Strongly agree*).

The participants then also completed the NASA-Task Load Index (NASA-TLX: Hart & Staveland, 1988) after completing each Email Sorting Task. NASA-TLX has been conducted in previous studies assessing participants cognitive load during cognitive demanding tasks (Biondi et al., 2022; Smith-Jackson & Klien, 2008). Prior research has also reported that participants rated greater levels of cognitive load when performing tasks within an office noise environment (Sheng et al., 2022; Smith-Jackson & Klein, 2008).

NASA- TLX is a multidimensional rating scale that measures perceived mental load after completing a task. NASA-TLX consists of six subscales of mental workload: mental demand; physical demand; temporal demand; performance; effort; frustration. In this study, an adapted version called Raw Task Load Index (RTLX) was used, consisting of one item from each subscale. Items are scored on a Likert scale from 1 (*Low*) to 7 (*High*) ($\alpha = .77$). Sample items include 'How high were the mental demands during this task?' and 'How high was your level of frustration during this task?'. The NASA TLX has demonstrated high convergent validity (Rubio et al., 2004) and good re-test reliability (Xiao et al., 2005) (see Appendix B).

EXPERT Intensive Skills Evaluation 2.0 (EXPERTise 2.0)

EXPERTise 2.0 is a shell software platform that can be customised to assess behaviours that reflect the utilisation of cues in specific domains (Wiggins et al., 2014). EXPERTise 2.0 comprises multiple situational-based tasks, which assess different components of cue utilisation. Since cue utilisation is domain-specific, the stimuli incorporated in the software is dependent on the domain of interest. EXPERTise 2.0 has been utilised in serval domains including power transmission control (Small et al., 2014), audiology (Watkinson et al., 2018), driving (Yuris et al., 2019), air traffic control (Falkland & Wiggins, 2019) and medicine (Crane et al., 2018). EXPERTise 2.0 has demonstrated construct validity (Loveday et al., 2014; Small et al., 2014), predictive validity (Watkinson et al., 2018) and test-retest reliability (Loveday et al., 2013). The cybersecurity edition of EXPERTise 2.0 was utilised in this study and comprises four tasks: Feature Identification Task (FIT), Feature Recognition Task (FRT), Feature Association Task (FAT) and Feature Discrimination Task (FDT).

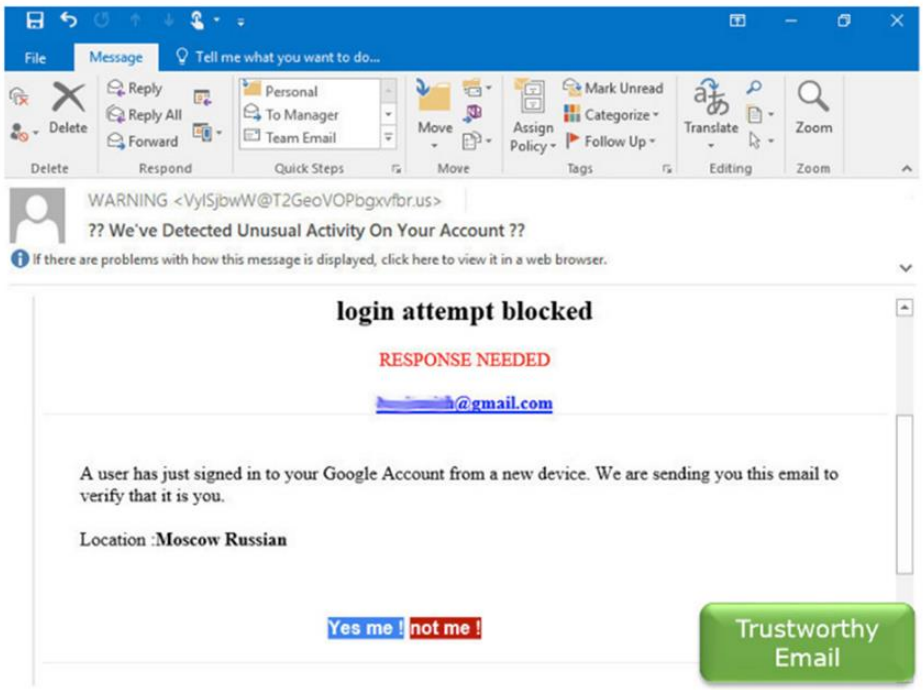
During the FIT, participants were required to identity relevant key features of concern presented in a series of 16 scenarios. Each scenario consisted of a phishing email, which included features such as a senders address, a greeting, spelling/grammar, a corporate logo and URL Link. Participants were asked to click, as quickly as possible, on the feature that aroused the most suspicion or had an option to click on a ‘Trustworthy Email’ icon (see figure 3). The response time from the initial display of each email and the moment participants selected a suspicious feature or the ‘Trustworthy Email’ icon were recorded. Higher cue utilisation is associated with faster response times when identifying key diagnostic features (Loveday et al., 2014).

In the FRT, participants were assessed on their ability to rapidly recognise predictive features of phishing emails to form an accurate judgement. Participants were presented with a

series of 22 emails (11 phishing, 11 genuine). Each email was presented for 1000 ms, and then prompted to categorise the email as either ‘Trustworthy’, ‘Untrustworthy’ or ‘Impossible to tell’. The number of correct classifications were recorded for each participant. Higher cue utilisation is associated with greater accuracy in email classifications (Bayl-Smith et al., 2020; Brouwers et al., 2017).

Figure 3

Example of a Phishing Email Presented in Feature Identification Task



During the FAT, participants were presented with a series of 15 pairs of words that were related to cybersecurity (e.g., ‘Email’, ‘Malware’), with each pair presented simultaneously on the screen for 2000 ms. Participants were then asked to rate the extent the two words are related on a Likert-scale ranging from 1 (Extremely unrelated) to 7 (Extremely related). The mean variance of each response was recorded for each participant.

Higher cue utilisation is associated with greater variance in the perceived relatedness of cybersecurity terms (Morrison et al., 2013).

In the FDT, participants were required to rate the importance of different features in an email when evaluating an emails legitimacy. Participants were presented with two problem-oriented phishing email scenarios. In both scenarios, participants were presented with a phishing email claiming an invoice was overdue and legal action will be taken if not paid. Participants were asked to read the scenario and then required to select an appropriate response from a range of options, in regard to the emails legitimacy (e.g., 'Ignore the email' , 'Contact company to confirm request'). After, participants were asked to rate the importance of 10 features presented in the scenario (e.g., 'Date of email', 'Email address') on a Likert-scale ranging from 1 (*Not important at all*) to 10 (*Extremely important*). The mean variance of responses were reported, with higher cue utilisation associated with greater variance in ratings (Pauley et al., 2009).

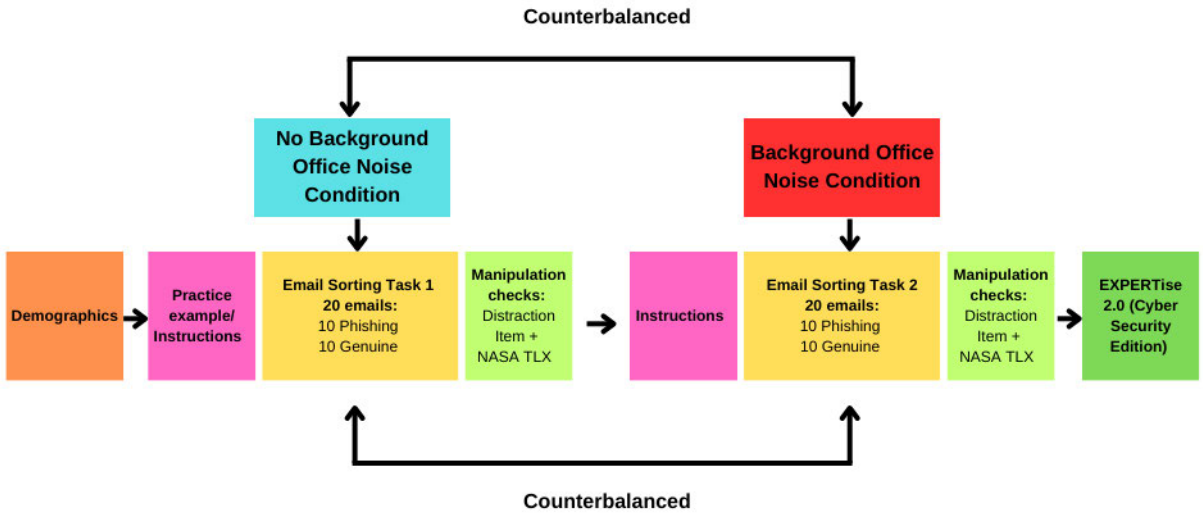
Procedure

The study was approved by the University of Adelaide Human Research Ethics Committee (H-2023-25) with voluntary informed consent obtained by all participants. Upon arrival, participants were presented with an information sheet on Qualtrics and notified the study was interested in users behaviour and the management of emails. Once participants provided informed consent, participants completed a demographics questionnaire. Before commencing the Email Sorting Tasks, participants were informed to wait for further instructions from the researcher. Once the researcher notified the participant which condition (background office noise, no background office noise) they were in, participants completed the Email Sorting Task. Participants completed two manipulation checks. After, participants

were directed to a separate platform to complete the EXPERTise 2.0 (Cybersecurity edition). The total time to complete the study was approximately 60 minutes.

Figure 4

Experimental workflow of the study.



Results

Overview of Analysis

The objective of this study was to investigate whether participants cue utilisation capacity (higher, lower) and the presence of a noise condition (background office noise, no background office noise) affected participants' ability to differentiate phishing from genuine emails. The data analysis was conducted in two stages, including a preliminary data analysis involving two manipulation checks using the IBM Statistical Package for Social Sciences (SPSS) Version 27. The first stage involved using a k-mean cluster analysis to determine two typologies of cue utilisation (higher, lower) based on participants performance scores on four EXPERTise 2.0 tasks. The second stage involved the examination of the hypothesis using analysis of variance (ANOVA).

Preliminary Data Analysis

Manipulation checks

Data from the distraction scale, which consisted of a single item, was analysed for each participant in both the no background office noise and background office noise condition. Higher scores on the distraction item indicated of higher levels of perceived distraction experienced by the participants.

Participant responses across the RTLX six items were reduced to a single RTLX score for each participant in each noise condition (no background office noise, background office noise). The item measuring performance (P), which used a scale from 1 (*not successful*) to 7 (*successful*), was reserve coded. A single RTLX score was represented by the sum score of the six responses, where higher scores reflecting greater perceived workload for that condition.

Data Reduction

The dataset from EXPERTise 2.0 and the Email Sorting Task was screened for missing data and outliers. One case with missing EXPERTise 2.0 data were removed. Forty-eight were retained for the initial analysis.

The data from the Email Sorting Task, and EXPERTise 2.0 (Cybersecurity edition) went through data reduction. Consistent with the standard approach to data reduction of EXPERTise 2.0 data (Bayl-Smith et al., 2020; Brouwers et al., 2016; Sturman et al., 2023), the FIT mean response time taken to identify the most suspicious feature within an email was calculated across 16 scenarios. For the FRT, participants accuracy in classifying emails was summed across a set of 22 emails. For the FAT, participants perceived association ratings between feature-event pairs across 15 scenarios were calculated to generate a mean variance score. For the FDT, participants' mean variance in rating of the importance of 10 email features across 2 scenarios were calculated as a mean sum.

Participants' capacity to discriminate between genuine and phishing emails in the email sorting task was calculated using Signal Detection Theory (Stanislaw and Todorov, 1999). Hits and false alarms for each participant were calculated separately in both noise conditions (background office noise, no background office noise). Hits were operationalised as the number of emails that were correctly classified as phishing emails. False alarms were operationalised as the number of genuine emails that were falsely classified as a phishing email. After hits and false alarms were converted to z scores. Sensitivity scores (d') were calculated by subtracting z(hits) from z(false alarms). Larger d' scores indicate a participant has a better ability to differentiate between genuine and phishing emails (higher ratio of hits and less false alarms).

Data Analysis

Stage 1: Establishing Typologies

To establish cue utilisation typologies, a k-means cluster analysis was conducted to categorise participants into two cue utilisation typologies (higher, lower) based on performance across the four EXPERTise tasks (Sturman et al., 2019; Wiggins et al., 2014). With the standardised scores, the cluster analysis yielded statistically significant mean differences between two cue utilisation clusters, reflecting participants with relatively higher (n= 32) or lower (n= 16) levels of cue utilisation. The higher cue utilisation cluster consisted of participants with higher variance ratings in the FAT and FDT tasks, greater accuracy on the FRT and lower response latency in FIT, in comparison to the lower cue utilisation cluster (see Table 1). Independent sample t-test demonstrated statistically significantly differences between the two cue utilisation clusters across all EXPERTise tasks.

Table 1

Standardised z scores from EXPERTise 2.0 tasks by cue utilisation typology.

	Higher cue utilisation (n= 32)	Lower cue utilisation (n= 16)
Feature Identification Task	-.34**	.67**
Feature Recognition Task	.35**	-.71**
Feature Association Task	.29*	-.57*
Feature Discrimination Task	.27*	-.54*

Note. *p <.05 (two tailed) **p <.01 (two tailed)

Descriptive Statistics

Paired sample t-tests was run for both manipulation checks. For the distraction item, participants reported significantly higher levels of distraction in the background office noise condition (M= 3.57, SD= 1.35) compared to the no background noise condition (M= 1.92, SD= 1.10), t (1,48) = 7.47, $p.$ < .001.

In the R-TLX, there was a significant difference between the noise conditions, with participants in the background office noise condition (M = 4.21, SD = .96) reported a higher perceived workload compared to the no background office noise condition (M =3.53, SD = .84), t (1,48) = 6.57, $p.$ < .001.

Table 2

Descriptive statistics in both noise conditions.

Variable	No background office noise				Background office noise			
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
Hit Rate (z score)	4.33	2.86	0	10	4.18	2.88	0	9
False Alarms (z scores)	1.04	1.17	0	5	1.12	1.33	0	5
Sensitivity (z scores)	1.32	1.28	-1.04	4.65	1.15	1.30	-1.48	3.61

Stage 2: Hypothesis testing

To test the hypothesis, a 2 x 2 mixed method ANOVA was conducted. For each independent variable a 2 x 2 (Cue utilisation [higher, lower] x noise condition [background office noise, no background office noise]) mixed ANOVA was run, with cue utilisation as the between subjects variable and the noise condition as the with-in subjects variable. The dependent variable was sensitivity (d').

The examination of histograms revealed the dependent variable was normally distributed for each condition. Further, Shapiro-Wilk tests were non-significant, indicating that the assumption of normality was met. Levene's Test of Equal Variance revealed there was no significant differences in each group's variance, indicating the assumption of homogeneity of variances was not violated across groups. For the repeated measure variable, Mauchly's test revealed that the assumption of was sphericity was met for each condition ($p > .05$).

H1: There was a significant main effect of cue utilisation on sensitivity scores F (1, 46) = 4.756, p = .034, $\eta 2$ = .094. Participants with higher cue utilisation (M= 1.5, SD= .196)

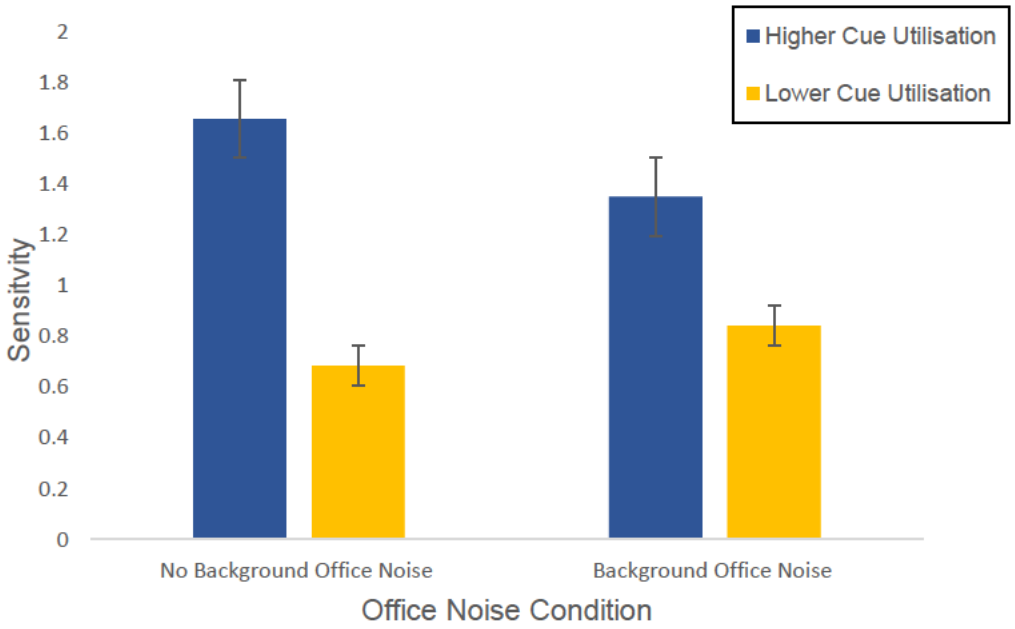
were better able to differentiate between genuine and phishing emails compared to participants with lower cue utilisation (*M*= .76, *SD*= .28). Therefore, H1 was supported.

H2: There was a statistically non-significant main effect of noise conditions on sensitivity scores F (1, 46) = .174, p = .679, η2= .004. This indicates there was no difference in participants ability to differentiate between phishing from genuine emails in both noise conditions. Therefore, participants sensitivity scores in the background office noise condition (*M*=1.15, *SD*= 1.30) compared to the no background office noise condition (*M*=1.32, *SD*= 1.28) exhibited similar levels of accuracy when differentiating phishing from genuine emails. This result does not support H2.

H3: The ANOVA revealed non-significant interaction between cue utilisation typology and noise conditions on participants ability to differentiate phishing from genuine emails F (1,46) = 1.66, p = .204, η2 = .35. This result indicates that differences in ability to differentiate genuine from phishing emails between participants with relatively higher and lower cue utilisation was not affected by the noise conditions (see figure 5). This result provides a non-support for H3, suggesting that the background office noise condition did not significantly increase the numbers of errors when differentiating genuine from phishing emails among participants with higher and lower cue utilisation.

Figure 5

The mean sensitivity scores by Cue Utilisation and Nose Condition



Note. Error bars represent standard errors.

Discussion

The aim of this study was to examine the relationship between cue utilisation and the presence of office noise on participants ability to differentiate between genuine and phishing emails. Overall, H1 was supported as participants with higher levels of cue utilisation were better able to differentiate phishing from genuine emails, compared to participants with lower cue utilisation. However, H2 was not supported, revealing that the presence of background noise made no difference in participants ability to differentiate phishing from genuine emails. Further, participants with relatively lower cue utilisation levels did not experience a significant decrease in performance when background office noise was present, therefore H3 was not supported.

In support for H1, the current study showed that participants with higher cue utilisation were better able to differentiate genuine from phishing emails compared to

participants with lower cue utilisation. This is consistent with empirical studies which have found that higher cue utilisation is associated with increased performance in identifying phishing cues and detecting phishing emails (Ackerley et al., 2022; Bayl-Smith et al., 2020; Nasser et al., 2020; Sturman et al., 2023).

Additionally, this finding provides support for the theoretical notion of Brunswik lens model (1955), suggesting participants with higher cue utilisation are better able to identify cue-based associations and weigh them appropriately to make more accurate judgements regarding email legitimacy.

Contrary to H2, participants performance in differentiating genuine from phishing emails was not negatively impacted by background office noise. Instead, the results indicated that participants performance stayed consistent when exposed to office noise. Overall, the findings in this study do not support previous research which found office noise to significantly worsen performance on a range of cognitive measures (Brocolini et al., 2016; Jahncke and Hallman, 2020; Yadav et al., 2017).

Findings from this study indicate that office noise did not negatively impact participants ability to differentiate genuine from phishing emails, however participants subjective ratings of cognitive workload and levels of distraction were significantly greater when completing the Email Sorting Task with the presence of background office noise. According to Resource Allocation theories, when additional demands are placed on working memory, individuals allocate a greater proportion of available cognitive resources to meet the demands of the task (Kahneman, 1973). Therefore, participants may have exerted more attention, effort and cognitive resources when exposed to office noise, allowing them to maintain performance on the Email Sorting Task (Small et al., 2014).

Additionally, the lack of effect may be linked to the duration (e.g., 10 mins) participants were required to complete Email Sorting Task. Similar studies examining noise

and cognitive performance found participants were able to maintain performance levels when exposed to noise during short time periods (e.g., under 15 minutes) (see Haka et al., 2009; Kjellberg, 1997; Landström et al., 2002). Further, these studies also found participants reported higher perceived levels of effort (Kjellberg, 1997; Landström et al., 2002) and disturbance (Haka et al., 2009) when exposed to noise.

Further, examining background office noise in lab-based environment may have lacked ecological validity. The results of this study might not reflect how individuals act in real-world workplaces when appraising emails. For example, in this study the primary task was to classify emails into categories. The findings in this study may be underestimated, as in many real-world settings receiving emails is usually a secondary task to a person's work (Zhou et al., 2022). Therefore, in the real-world when employees are faced with additional cognitive demands such as noise, employees may not allocate additional cognitive resources when appraising emails and may be more likely to fall for a phishing attack.

H3 was not supported, finding that participants with lower cue utilisation did not experience a significant loss in performance in the presence of office noise compared to participants with higher cue utilisation. Theoretically, one of the advantages of higher cue utilisation, is that cues are retained in long-term memory, imposing fewer demands on working memory (Wiggins, 2021). Therefore when participants with lower cue utilisation encountered additional cognitive demands imposed by office noise, it was anticipated that they would experience a greater loss in performance compared to participants with higher cue utilisation.

Cognitive Load Theory (CLT; Sweller, 1998) proposes that working memory has a limited capacity, and when the total cognitive load associated with the task at hand exceeds memory capacity, it can lead to a decline in performance (Haji et al., 2015). Therefore it is

plausible that the background office noise did not impose a significant amount cognitive demands on working memory to negatively impact performance.

Additionally, according to Dual-system theories of reasoning, the deployment of System 1 processing depends on the amount of cognitive strain placed on individual in a given situation (Jones et al., 2019). Therefore, it is also plausible that the background office noise did not force or require participants to use System 1 processing. Therefore the office noise may not have placed enough cognitive demands on working memory to result in participants not needing to rely on cues when differentiating genuine from phishing emails.

Strengths

This study demonstrates several methodological strengths, one of which is attributed to the experimental design, enhancing internal validity and ecological validity. By employing a lab-based design, background office noise was able to be measured while other variables were kept constant, therefore increasing the likelihood to rule out other alternative explanations on participants performance on the Email Sorting Task. Secondly, a limitation commonly discussed in phishing detection studies (Buckley et al., 2023; Hanel & Vione, 2016; Williams et al., 2023), is that many lab-based designs are performed in quiet and controlled conditions, lacking ecological validity. By utilising an office noise recording, the study in enhanced ecological validity to gain a better understanding of individuals decision-making on phishing detection in more a realistic setting. By enhancing ecological validity, this study can inform practical implications and future research (Williams et al., 2023).

Second, another notable strength was the email stimuli used for this study. Firstly, past phishing research has utilised real phishing emails (not stimulated) therefore the cues in the emails were not standardised (Bayl-Smith et al, 2020). As phishing emails can vary in sophistication, context and cues, past studies have found it challenging to determine which

cues in the phishing emails influenced accuracy or inaccuracy in detecting phishing emails (Buckley et al., 2023; Williams et al., 2018; Williams et al., 2023).

By stimulating the phishing emails, the phishing cues were standardised across all the emails used in the study, enhancing internal validity. As participants were exposed to the same cues in the simulated phishing emails, the study could draw more accurate and reliable conclusions on the relationship between cue utilisation and background office noise on phishing detection. Subsequently, future studies can use this approach to investigate what phishing cues result in the most or least accuracy on participants ability to differentiate between genuine and phishing emails (Bayl-Smith et al., 2020).

Limitations and Future Directions

This research, however, is subject to potential limitations. One limitation in this study pertains to the relatively small sample size (*n= 49*). The small number of participants can limit the generalisability of findings and the sufficient statistical power to detect significant results and potential main effects. Due to the smaller sample size the generalisability of these results are not certain, however we did find that higher cue utilisation is associated with greater ability to differentiate genuine and phishing emails, consistent with past research with larger samples sizes (Ackerley et al., 2022; Nasser et al., 2022; Sturman et al., 2023). Future research should replicate this study with a larger sample size with individuals from the broader population.

Additionally, this study solely focused on the influence of office noise on phishing detection. However, in real-world work settings, employees are often subjected to or encounter a range of other workplace factors that may potentially contribute to phishing susceptibility. Phishing research have found workplace factors such as high email load (Vishwanath et al., 2018), busy work periods (Conway et al., 2017), time constraints (Butavicius et al., 2022) to negatively impact an individual's ability to detect phishing emails

(Butavicius et al., 2022; Conway et al., 2017; Vishwanath et al., 2018). Such studies have concluded that when individuals are busy, receive a high load of emails and/ or have less time to sufficiently appraise each email, individuals rely on their intuition (System 1 processing), which, in turn, decreases phishing detection performance (Butavicius et al, 2022; Conway et al., 2017; Vishwanath et al., 2018). This study provides insight about the influence of office noise on phishing detection, but it only informs a small aspect of the complex and dynamic nature of workplaces. Although these workplace factors were beyond the scope of this study, future research could possibly take a more holistic approach examining workplaces factors and their interplay on phishing detection (Buckley et al., 2023; Williams et al., 2023).

Also, future research may benefit using the background office noise stimuli for a longer duration and in present it in different variations to replicate real-world settings. In the workplace, employees often contend with prolonged exposure to various noise levels that fluctuate throughout the workday. For example, Banbury and Berry (1997) found participants were able to habituate to continuous office noise, but after short periods of quiet, participants dishabituated, leading to errors in performance. While this study used continuous office noise stimuli, examining the effects of noise for longer durations and the impact of intermediate noise may provide greater insight about office noise and performance in real-world settings.

Practical Applications and Implications

At an applied level, the findings of the present study have implications for future phishing detection training programs in organisational settings. As higher cue utilisation participants were found to be better at differentiating genuine from phishing emails, cue-based training may be an effective intervention. Cue-based training facilitates the acquisition of cues associated with phishing emails via simulation training (Sturman et al., 2023). Further simulation training allows the employee to learn, practice and reinforce cue-base associations without posing risk to the organisation’s digital infrastructure (Falkland et al., 2019). Also

cue-based training programs have been found to improve phishing detection and lower the likelihood of classifying a genuine email as phishing (Weaver et al., 2021). As employees are receiving a higher volume of emails, and a small proportion of them are phishing emails, training programs need to focus on reducing the number of decisions that need to be made to make sure the employees time is utilised sufficiently. As cue-base associations alleviate the cognitive demands on working memory, cue-based training may be an attractive avenue for organisations.

Conclusion

This current study was designed to examine the relationship of cue utilisation and background office noise on participants ability to differentiate between genuine and phishing emails. The outcomes of this study extended on previous findings that participants with higher cue utilisation was associated with a greater ability to detect phishing emails (Ackerley et al., 2022; Bayl-Smith et al., 2020; Nasser et al., 2020; Sturman et al., 2023; Williams et al., 2023). However, the presence of noise did not adversely affect participants ability to differentiate genuine from phishing emails, nor was there a statistically significant interaction between cue utilisation typology and the presence of noise on phishing detection performance. Overall, the outcomes of this study suggest that cue-based processing remains to be an effective strategy for detecting phishing emails. Therefore, cue-based training may be beneficial for the protection of organisations cybersecurity.

References

Ackerley, M., Morrison, B. W., Ingrey, K., Wiggins, M. W., Bayl-Smith, P., & Morrison, N. (2022). Errors, Irregularities, and Misdirection: Cue Utilisation and Cognitive Reflection in the Diagnosis of Phishing Emails. In *Australasian Journal of Information Systems.*

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. In *Frontiers in Computer Science* (Vol. 3). Frontiers Media S.A. https://doi.org/10.3389/fcomp.2021.563060

Australian Cyber Security Centre. (2021). *ACSC Annual Cyber Threat Report.* https://www.cyber.gov.au/sites/default/files/2023-03/ACSC%20Annual%20Cyber%20Threat%20Report%20-%202020-2021.pdf

Australian Cyber Security Centre. (2022). *Annual Cyber Threat Report.* https://www.cyber.gov.au/sites/default/files/2023-03/ACSC-Annual-Cyber-Threat-Report-2022_0.pdf

Ayaburi, E., Kofi Andoh-Baidoo, F., & Ayaburi, E. W. (2019). Association for Information Systems Association for Information Systems Understanding Phishing Susceptibility: An Integrated Model of Understanding Phishing Susceptibility. *International Conference on Information Systems (ICIS).* https://aisel.aisnet.org/icis2019/cyber_security_privacy_ethics_IS/cyber_security_privacy/43

Banbury, S., & Berry, D. C. (1997). Habituation and dishabituation to speech and office noise. *Journal of Experimental Psychology: Applied, 3*(3), 181–195. https://doi.org/10.1037/1076-898X.3.3.181

Banbury, S. P., & Berry, D. C. (2005). Office noise and employee concentration: Identifying causes of disruption and potential improvements. *Ergonomics*, *48*(1), 25–37. https://doi.org/10.1080/00140130412331311390

Bayl-Smith, P., Sturman, D., & Wiggins, M. (2020). Cue utilization, phishing feature and phishing email detection. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12063 LNCS*, 56–70. https://doi.org/10.1007/978-3-030-54455-3_5

Biondi, F. N., Saberi, B., Graf, F., Cort, J., Pillai, P., & Balasingam, B. (2023). Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied Ergonomics*, *106*. https://doi.org/10.1016/j.apergo.2022.103867

Brocolini, L., Parizet, E., & Chevret, P. (2016). Effect of masking noise on cognitive performance and annoyance in open plan offices. *Applied Acoustics*, *114*, 44–55. https://doi.org/10.1016/j.apacoust.2016.07.012

Buckley, J., Lottridge, D., Murphy, J. G., & Corballis, P. M. (2023). Indicators of employee phishing email behaviours: Intuition, elaboration, attention, and email typology. *International Journal of Human Computer Studies*, *172*. https://doi.org/10.1016/j.ijhcs.2023.102996

Brouwers, S., Wiggins, M. W., Helton, W., O'Hare, D., & Griffin, B. (2016). Cue utilization and cognitive load in novel task performance. *Frontiers in Psychology*, *7*(MAR). https://doi.org/10.3389/fpsyg.2016.00435

Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics*, *60*(11), 1500–1515. https://doi.org/10.1080/00140139.2017.1330494

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. Psychological review, 62(3), 193.

Butavicius, M., Parsons, K., Pattinson, M., & McCormac, A. (2016). Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *arXiv preprint arXiv:1606.00887*.

Butavicius, M., Taib, R., & Han, S. J. (2022). Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails. *Computers and Security*, *123*. https://doi.org/10.1016/j.cose.2022.102937

Chen, W., Guo, X., Chen, Z., Zheng, Z., & Lu, Y. (2020). *Phishing Scam Detection on Ethereum: Towards Financial Security for Blockchain Ecosystem*. https://bitcoin.org/bitcoin.pdf

Conway, D., Taib, R., Harris, M., Yu, K., Berkovsky, S., & Chen, F. (2017). A Qualitative Investigation of Bank Employee Experiences of Information Security and Phishing. *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017),* 115-129.

Crane, M. F., Brouwers, S., Wiggins, M. W., Loveday, T., Forrest, K., Tan, S. G. M., & Cyna, A. M. (2018). "Experience Isn't Everything": How Emotion Affects the Relationship Between Experience and Cue Utilization. *Human Factors*, *60*(5), 685–698. https://doi.org/10.1177/0018720818765800

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual review of psychology*, *47*(1), 273-305.

Falkland, E. C., & Wiggins, M. W. (2019). Cross-task cue utilisation and situational awareness in simulated air traffic control. *Applied Ergonomics*, *74*, 24–30. https://doi.org/10.1016/j.apergo.2018.07.015

Falkland, E. C., Wiggins, M. W., & Westbrook, J. I. (2020). Cue Utilization Differentiates Performance in the Management of Interruptions. *Human Factors*, *62*(5), 751–769. https://doi.org/10.1177/0018720819855281

Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems*, *18*(1), 22–44. https://doi.org/10.17705/1jais.00447

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*, *52*, 139-183. https://doi.org/10.1016/S0166-4115(08)62386-9

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E., & Hongisto, V. (2009). Performance effects and subjective disturbance of speech in acoustically different office types–a laboratory experiment. *Indoor air*, *19*(6), 454-467.

Hanel, P. H. P., & Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0168354

Hanus, B., Wu, Y. A., & Parrish, J. (2022). Phish Me, Phish Me Not. *Journal of Computer Information Systems*, *62*(3), 516–526. https://doi.org/10.1080/08874417.2020.1858730

Hygge, S., & Knez, I. (2001). Effects of noise, heat and indoor lighting on cognitive performance and self-reported affect. *Journal of Environmental Psychology*, *21*(3), 291–299. https://doi.org/10.1006/jevp.2001.0222

IBM Cybersecurity Intelligence Index. (2014). *IBM Security Services 2014 Cyber Security Intelligence Index.* https://i.crn.com/sites/default/files/ckfinderimages/userfiles/images/crn/custom/IBMSecurityServices2014.PDF

Jafari, M. J., Khosrowabadi, R., Khodakarim, S., & Mohammadian, F. (2019). The effect of noise exposure on cognitive performance and brain activity patterns. *Open Access Macedonian Journal of Medical Sciences*, *7*(17), 2924–2931. https://doi.org/10.3889/oamjms.2019.742

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

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58*(9), 697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kalaharsha, P., & Mehtre, B. M. (2021). Detecting Phishing Sites – An Overview. *arXiv:2103.12739*. http://arxiv.org/abs/2103.12739

Kjellberg, A. (1997). Noise. In H. A. Waldron, & C. Edling (Eds.), Occupational health practice (4th ed., pp. 241–256). Oxford: Butterworth Heinemann.

Klein, G. 1993. "A Recognition-Primed Decision (RPD) Model of Rapid Decision Making." In Decision Making in Action: Models and Methods, edited by G. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, 138–147.

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

Koehler, D. J., & Harvey, N. (2008). Blackwell Handbook of Judgment and Decision Making.

Landström, U., Söderberg, L., Kjellberg, A., & Nordström, B. (2002). Annoyance and performance effects of nearby speech. *Acta acustica united with acustica*, *88*(4), 549-553.

Lin, T., Capecci, D. E., Ellis, D. M., Rocha, H. A., Dommaraju, S., Oliveira, D. S., & Ebner, N. C. (2019). Susceptibility to spear-phishing emails: Effects of internet user demographics and email content. *ACM Transactions on Computer-Human Interaction*, *26*(5). https://doi.org/10.1145/3336141

Loveday, T., Wiggins, M. W., & Searle, B. J. (2014). Cue utilization and broad indicators of workplace expertise. *Journal of Cognitive Engineering and Decision Making*, *8*(1), 98–113. https://doi.org/10.1177/1555343413497019

Martin, J., Dubé, C., & Coovert, M. D. (2018). Signal Detection Theory (SDT) Is Effective for Modeling User Behavior Toward Phishing and Spear-Phishing Attacks. *Human Factors*, *60*(8), 1179–1191. https://doi.org/10.1177/0018720818789818

Morgan, S. (2020). Cybercrime to cost the world $10.5 trillion annually by 2025. *Cybercrime Magazine*, *13*(11). https://cybersecurityventures.com/ cybercrime-damage-costs-10-trillion-by-2025/.

Morrison, L. J., Neumar, R. W., Zimmerman, J. L., Link, M. S., Newby, L. K., McMullan, P. W., Hoek, T. Vanden, Halverson, C. C., Doering, L., Peberdy, M. A., & Edelson, D. P. (2013). Strategies for improving survival after in-hospital cardiac arrest in the United States: 2013 consensus recommendations: A consensus statement from the American heart association. *Circulation*, *127*(14), 1538–1563. https://doi.org/10.1161/CIR.0b013e31828b2770

Mosier, K. L., & Kirlik, A. (2004). Brunswik's lens model in human factors research: Modern applications of a classic theory. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 48*(3), 350-354.

Musuva, P. M. W., Getao, K. W., & Chepken, C. K. (2019). A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility. *Computers in Human Behavior*, *94*, 154–175. https://doi.org/10.1016/j.chb.2018.12.036

Mynatt, E. D., Hudson, S. E., Fitzpatrick, Geraldine., Association for Computing Machinery., & SIGCHI (Group : U.S.). (2010). *CHI 2010 : we are HCI : conference proceedings, Atlanta, Ga, USA, April 10-15, 2010*. Association for Computing Machinery.

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

Parsons, K., Butavicius, M., Delfabbro, P., & Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human Computer Studies*, *128*, 17–26. https://doi.org/10.1016/j.ijhcs.2019.02.007

Parsons, K., Butavicius, M., Pattinson, M., Calic, D., Mccormac, A., & Jerram, C. (2016). Do users focus on the correct cues to differentiate between phishing and genuine emails? *Australasian Conference on Information Systems*. https://doi.org/10.48550/arXiv.1605.04717

Parsons, K., Mccormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Phishing for the Truth: A Scenario-Based Experiment of Users' Behavioural Response to Emails. In *Security and Privacy Protection in Information Processing Systems: 28th IFIP TC 11 International Conference, SEC 2013, Auckland, New Zealand, 405* (28), 366-378. https://doi.org/10.1007/978-3-642-39218-4_27

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2015). The design of phishing studies: Challenges for researchers. *Computers and Security*, *52*, 194–206. https://doi.org/10.1016/j.cose.2015.02.008

Pauley, K., O'Hare, D., & Wiggins, M. (2009). Measuring expertise in weather-related aeronautical risk perception: The validity of the Cochran-Weiss-Shanteau (CWS) index. *International Journal of Aviation Psychology*, *19*(3), 201–216. https://doi.org/10.1080/10508410902979993

Rajivan, P., & Gonzalez, C. (2018). Creative persuasion: A study on adversarial behaviors and strategies in phishing attacks. *Frontiers in Psychology*, https://doi.org/10.3389/fpsyg.2018.00135

Small, A. J., Wiggins, M. W., & Loveday, T. (2014). Cue-based processing capacity, cognitive load and the completion of simulated short-duration vigilance tasks in power transmission control. *Applied Cognitive Psychology*, *28*(4), 481–487. https://doi.org/10.1002/acp.3016

Slifkin, E. J. D., & Neider, M. B. (2023). Phishing interrupted: The impact of task interruptions on phishing email classification. *International Journal of Human Computer Studies*, *174*. https://doi.org/10.1016/j.ijhcs.2023.103017

Smith-Jackson, T. L., & Klein, K. W. (2009). Open-plan offices: Task performance and mental workload. *Journal of Environmental Psychology*, *29*(2), 279–289. https://doi.org/10.1016/j.jenvp.2008.09.002

Stanislaw, H. (1999). Calculation of signal detection theory measures. In *Behavior Research Methods, Instruments, & Computers*, *3*(1), 37-149.

Sturman, D., Wiggins, M. W., Auton, J. C., Loft, S., Helton, W. S., Westbrook, J. I., & Braithwaite, J. (2019). Control room operators' cue utilization predicts cognitive resource consumption during regular operational tasks. *Frontiers in Psychology*, *10*(AUG). https://doi.org/10.3389/fpsyg.2019.01967

Sturman, D., Valenzuela, C., Plate, O., Tanvir, T., Auton, J. C., Bayl-Smith, P., & Wiggins, M. W. (2023). The role of cue utilization in the detection of phishing emails. *Applied Ergonomics*, *106*. https://doi.org/10.1016/j.apergo.2022.103887

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, *10*, 251-296.

Thapa, C., Tang, J. W., Abuadbba, A., Gao, Y., Camtepe, S., Nepal, S., Almashor, M., & Zheng, Y. (2020). *Evaluation of Federated Learning in Phishing Email Detection*. https://doi.org/10.3390/s23094346

The Corner of Ambient Sounds & ASMR. (2019, September 25). *ASMR: Office Ambience 8 HOURS of OFFICE SOUNDS, Working | Study | Writing | ASMR | White Noise* [Video]. Youtube. https://www.youtube.com/watch?v=9y0wmSgRNRk

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. In *New Series*, *185*(41).

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. https://doi.org/10.1111/jcc4.12126

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. https://doi.org/10.1016/j.dss.2011.03.002

Wang, J., Herath, T., Chen, R., Vishwanath, A., & Rao, H. R. (2012). Research article phishing susceptibility: An investigation into the processing of a targeted spear phishing email. In *IEEE Transactions on Professional Communication*, 55(4), 345–362. https://doi.org/10.1109/TPC.2012.2208392

Wash, R. (2020). How Experts Detect Phishing Scam Emails. *Proceedings of the ACM on Human-Computer Interaction*, *4*. https://doi.org/10.1145/3415231

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.org/10.1080/14992027.2018.1476782

Weaver, B. W., Braly, A. M., & Lane, D. M. (2021). Training Users to Identify Phishing Emails. *Journal of Educational Computing Research*, *59*(6), 1169–1183. https://doi.org/10.1177/0735633121992516

Wiggins, M. W. (2014). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, *15*(3), 283–292. https://doi.org/10.1080/1463922X.2012.724725

Wiggins, M. W. (2020). Cue Utilization as an Objective Metric in Naturalistic Decision-Making. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 209–213. https://doi.org/10.1177/1071181320641051

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., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118–124. https://doi.org/10.1016/j.ssci.2014.01.006

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.org/10.1016/j.ijhcs.2018.06.004

Williams, R., Morrison, B. W., Wiggins, M. W., & Bayl-Smith, P. (2023). The role of conscientiousness and cue utilisation in the detection of phishing emails in controlled and naturalistic settings. *Behaviour & Information Technology*, 1–17. https://doi.org/10.1080/0144929X.2023.2230307

Wigton, R. S., Patil, K. D., Hoellerich, V. L. (1986). The effect of feedback in learning clinical diagnosis. *Journal of Medical Education,* 61(10), 816-22.

Witterseh, T., Wyon, D. P., & Clausen, G. (2004). The Effects of Moderate Heat Stress and Open-Plan Office Noise Distraction on SBS Symptoms and on the Performance of Office Work. *Indoor Air, 14*(8), 30-40.

Yadav, M., Kim, J., Cabrera, D., & de Dear, R. (2017). Auditory distraction in open-plan office environments: The effect of multi-talker acoustics. *Applied Acoustics*, *126*, 68–80. https://doi.org/10.1016/j.apacoust.2017.05.011

Yuris, N. C., Wiggins, M. W., Auton, J. C., Gaicon, L., & Sturman, D. (2019). Higher cue utilization in driving supports improved driving performance and more effective visual search behaviors. *Journal of Safety Research*, *71*, 59–66. https://doi.org/10.1016/j.jsr.2019.09.008

Zhuo, S., Biddle, R., Koh, Y. S., Lottridge, D., & Russello, G. (2022). SoK: Human-Centered Phishing Susceptibility. *ACM Transactions on Privacy and Security*. https://doi.org/10.1145/3575797

Zuhayr, N. A., Girinoto, Qomariasih, N., & Setiawan, H. (2023). Detection Model for URL Phishing with Comparison Between Shallow Machine Learning and Deep Learning Models. *Springer Nature,* 146-156. https://doi.org/10.2991/978-94-6463-174-6_13

Appendix A

The Email Sorting Task Categorisation

Which category would you sort this email into?

- ◯ **Urgent** (emails, personal or work-related, that Alex needs to respond to within the next 24-48 hours)
- ◯ **Teaching** (emails from colleagues regarding the coordination of university courses)
- ◯ **Research** (emails regarding Alex's research and research opportunities)
- ◯ **Banking** (Alex's personal banking)
- ◯ **Online purchases** (receipts from purchases Alex has made)
- ◯ **Social Media accounts** (notifications from Alex's social media accounts)
- ◯ **Official** (personal emails from official agencies e.g. Medicare, ATO, AFP)
- ◯ **Spam** (advertisement emails of no consequence)
- ◯ **Phishing** (emails that seem fraudulent, fake or otherwise deceptive)
- ◯ **Miscellaneous** (emails that don't fit into any other category)

Appendix B

The NASA-Task Load Index Scale items

1. How high were the mental demands during the task.
2. How high were the physical demands during the task.
3. How high was the time pressure during the task.
4. How hard did you have to work to accomplish your level of performance.
5. How high was your level of frustration during the task.

Reversed scored

6. How successful do you think you were in terms of your performance.

(Hart & Staveland, 1988)