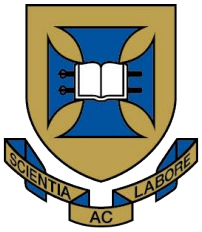


THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Characterisation of Human Behaviours Against Cyber Attacks

by

Chanon Kachornvuthidej

School of Information Technology and Electrical Engineering

A thesis submitted for the Bachelor of Science Honours degree at The University of Queensland in Year 2020

Chanon Kachornvuthidej

c.kachornvuthidej@uqconnect.edu.au

6 November 2020

Prof Amin Abbosh
Acting Head of School
School of Information Technology and Electrical Engineering
The University of Queensland
St Lucia QLD 4072

Dear Professor Abbosh,

In accordance with the requirements of the Degree of Bachelor of Science (Honours) in the School of Information Technology and Electrical Engineering, I submit the following thesis entitled

"Characterisation of Human Behaviours Against Cyber Attacks"

The thesis was performed under the supervision of Associate Professor Dan Kim. I declare that the work submitted in the thesis is my own, except as acknowledged in the text and footnotes, and that it has not previously been submitted for a degree at the University of Queensland or any other institution.

Yours sincerely

Chanon Kachornvuthidej

Abstract

Modern cybersecurity researches are concentrated on the development of sophisticated algorithms and security tools. Many research has shown despite effective automated tools, users are still falling for these online scams as a result of poor usability. Limited research is focusing on factors that contribute to human users to fall for online scams by applying psychological principles to investigate their thought processes. The presence of many pitfalls and challenges researchers faced such as reproducibility of results and complexity when designing the experiments are the root causes of the scarcity of human-centric cybersecurity researches. The current project aims to address these issues by constructing a generic framework that can aid the design of future research projects of a similar kind. A phishing study will be conducted to demonstrate as an example of how the framework could be implemented. Participants in the study will engage in a computer-based email classification task with the use of an eye tracker to capture their behaviour (gaze movement) followed by a short questionnaire. Results indicate participants exercise common online protection techniques such as not clicking on unknown links, verifying the sender's address, and not giving out personal information. Findings also uncover numerous techniques scammers used exploiting the psychological vulnerability of many online users such as visual deception, emotional manipulation, and cognitive overloading. Discussion on applications of the framework and suggestions on future human-centric research project ideas are also examined throughout the current report.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text.

The content of my thesis is the result of work I have carried out since the commencement of my Honours degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Keywords

cybersecurity, human-centric, eye-tracking, phishing, scams, psychology, human-computer interaction, cognition

This thesis is dedicated to my parents and sister, Adisak, Siri, and Thanathorn Kachornvuthidej for always supporting me and help me get to where I am today. I may not be the best of all but I am glad to have made you proud. I love you.

I would also like to take this opportunity to express my sincere gratitude to the following list of individuals who have taught me many valuable life lessons and unforgettable memories:

Dr Dan Kim - my supervisor, This project could not have been possible without your guidance. Thank you for trusting and believing in me, and be my inspiration to pursue the academia. I hope one day I can return as a PhD candidate under your supervision again.

Dr Jacki Liddle, Mr Peter Worthy and Dr Kelly Matthews for your mentorship during my Summer/Winter Research Program where my passion for research began;

All my professors and tutors at The University of Queensland;

All my teachers at Amnuay Silpa School;

And all my friends from around the world.

ขอคุณทุกคนทีทำให้ผมมีวันนี้ได้

Contents

Chapter 1

Introduction

1.1 Motivation

Corporations and governments around the world are responding to the growing cyber-attacks over recent years. For many, the attack seems distant however your data and private information are potentially among those targeted or even stolen. Over 577 million accounts of user information including name, date of birth, and social security number were leaked between 2017-2018. Shockingly this figure only includes mere four major corporations [1]. There are many vulnerabilities in modern computer systems such as the use of a weak password combination, obsolete software, or communicating through a non-secure channel compromising two or more components of the CIA triad (Confidentiality, Integrity, and Availability) [2]. There is a consensus notion that humans are the weakest link in the cyberspace, however, limited investments are made towards user-training, security policies, and user-friendly interface to combat cyber attacks [3], [4], [5]. This is in line with reported statistics that 90% of successful data breaches originated from a phishing attack deceiving human users [6]. Although automated threat detection software and browser extensions are widely available in the market, many usability issues are preventing full usage of these tools.

On top of the sparse investments, majority of academic researches and industry resources are concentrated on developing sophisticated algorithms and software that could detect and warn users of potential threats. Besides, majority of existing human-centric cybersecurity research only concerns the outcome of user action (e.g. whether the user has correctly classified between phishing vs genuine email or not) and not the *why/how* of such behaviour (e.g. why did the user click open on certain email and not the other, why did the user choose to mark this email as spam). Limited attention has been given to the users themselves to understand the thought processes and behaviours of online users at the instance they encountered a potential cyber threat. Such knowledge not only will help developers program smart and useful automated tools enhancing user experience, but also develop user training programmes that could help shape online behaviour for user to become more aware when encountering such threats. There are some evidence outlining why many research projects are avoiding the focus on human users. [7], [8] and [9] mentioned issues with generalisability and reproducibility of results of

phishing studies involving human participants, in addition to various challenges and considerations involved that researchers faced when conducting such projects. These issues are as a result of lacking a generic guide or framework which can be utilised consistently throughout all human-centric research projects worldwide. This lack of framework and investigation of the thought processes will be the focus of the current research.

In addition, with the recent setup of UQ Cyber Squad in the School of ITEE, the proposed framework in the current project is crucial as it will pave the ground for future human-centric cybersecurity research within the School.

1.2 Topic Definition

This current project aims to address the issues faced by researchers associated with the human-centric cybersecurity domain by constructing a generic framework to aid future projects of a similar kind. To begin we will review existing literature relating to three aspects: 1) usability of automated anti-phishing tools; 2) properties of human users; and 3) components of the phishing materials. The framework will be derived from an example phishing study conducted in the second part of this project with the focus on generic email phishing which is the most common form of cyberattack. The example phishing experiment will demonstrate a series of tasks to be completed when conducting human-centric research. The activities include collating email samples to be used in the experiment which will replicate a real phishing attack as much as possible. The use of eye-tracking technology in cybersecurity research will then be investigated. In this study, participants will be required to view the email samples and classify whether it is genuine or a phishing attempt. To investigate the thought processes and behaviours during such an encounter with a potential phishing attack, an eye-tracker will capture what components of the email participants look at. Data analysis using generated visual heat-map will be conducted to extract components of the email that is of interest to the user. Extracted components are effective at informing users of a potential threat if that particular email sample is correctly classified, or deceiving if incorrect otherwise. Based on the captured data, suggestions on how to address the deceiving cues making it more salient will be proposed. Evaluations on the effectiveness of the framework will be discussed as well as it's applications in other cyber-security, or usability research contexts.

1.3 Existing Research

Currently, there are many known exploitation methods that attackers used ranging from Botnet, delivering a distributed denial-of-service attack; ransomware, attackers encrypt victim's files and demanding payment in exchange; or water holing, setting up a fake website in an attempt to steal information of visiting users [10]. Amongst those are commonly practised phishing techniques which is the focus of the current research: disguising as a trustworthy entity with intentions to gain sensitive information or encourage certain behaviours commonly for financial gain [11]. Many exploitations used by hackers often rely on phishing and social engineering as an entry point to bypass security

detection and protection systems. This is because phishing works similarly to a reverse payload attack method.

As shown to the right, Fig. 1.1 demonstrates the common two exploitation vectors of a typical cyberattacks. Bind payloads are defined as an instance where hackers try to gain access to a system or information by initiating a connection or attack from externally through the existing system vulnerabilities. These includes denial of service attacks, SQL injection, or ransomware. Where reverse payload the hackers deliver their exploitation materials (e.g. a phishing email) and wait for someone (or the internal system) to reconnect/reply and give out information. The reverse payload style phishing is very beneficial to scammers since it is easier to deceive human judgement than a technical machine or firewall which can react and mitigate the risk faster than a human user.

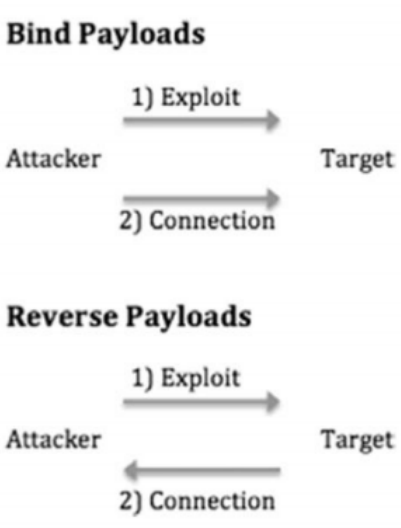


Figure 1.1: Reverse vs Bind Payload Source: [12]

Phishing itself also has many types and delivery vectors to the target including emails, SMS, voice messages/phone calls, or fake websites. The most common is generic email phishing, where attackers will register a fake domain that mimics an organisation and sends an email to the target tricking users to perform a certain action. This type of phishing will be the focus of current research. Emails are the preferred method of phishing delivery since it is low cost (only need to pay for the custom email domain), can be sent to international recipients (broaden the market of users to exploit), and easy to send in bulk (increase the chance of success). [13] reported statistically, when attackers send out 10 emails, there is a 90% chance that at least 1 person will fall for the phishing scam. This sounds like a small chance but when the scale of the attack is enlarged to hundreds and thousands of emails sent by attackers, the significance becomes more observable. Other notable types of attacks include spear phishing, similarly to generic email but includes specific personal information of the victim to the message sent; whaling, an even narrower delivery targeting only senior executives of companies; and clone phishing, attackers may use a legitimate email that was sent before but replace the attachment and links to a fake content instead [13], [14]. These techniques are not individually utilised but are often employed in combination with one another to increase the chance of success.

Chapter 2

Usability of Anti-Phishing Tools

Tremendous efforts are made to develop sophisticated security tools in the form of browser extensions and plugins that can automatically detect malicious activities. Google Safe Browsing, SiteAdvisor, SmartScreen Filter are a few examples of the tools trusted by banking institutions and major organisations globally [15], [16]. Functionalities of the tools range from IP filtering and packet inspection, email attachment scanning, to detecting inconsistencies in contents and name spellings [17], [18]. A study demonstrated these automated tools and modern browsers can weed out up to 80% of harmful phishing sites. The tool displays security warning that the site may be fake prompting users to re-evaluate their action before continuing to proceed at own's risk or abort [19]. It is also important to note that the majority of such tools are only developed and examined in the context of English language contents. Other natural languages like Chinese and Arabic could be the next target for major security breaches in line with expert's forecast that there will be reports of attack spikes in South-East Asia, China, and East African region in the near future [3], [20]. The advancement of detection tools over the years, in addition to such a high detection rate, should reduce the statistics of successful attacks. This logic contradicts the exponential trend of data breaches with a staggering 1,300% increase in reported cases from 2006 to 2015 [1]. There are no doubt these cases are the tip of the iceberg not accounting for other unreported data breaches that occurred during the period. [16] further added these automated security tools are only as effective as a Scam Block as humans are still the most vulnerable subject in the equation. It is first important to examine why the rate of data breaches are growing despite having these advanced tools. Proposed rationales behind this contradicting theory are: 1) the security warnings are being ignored, 2) users failed to utilise security indicators to protect themselves, and 3) the role of human psychology in cybersecurity. These arguments shall be explored further in the following sections to understand what issues users faced when using these automated tools.

2.1 Ignoring the Warnings

Although anti-phishing tools have a high detection rate, users still ignore the warnings and proceed to the potentially harmful contents anyway. In this research context, we define security warnings as those messages that *actively* interrupt the access flow of users when visiting a site or opening a message content forcing a manual user input selection of the available options.

Figure 2.1: Warning Message 1
Source: [19]

Figure 2.2: Warning Message 2
Source: [19]

Fig. 2.1 and 2.2 present examples of the warning messages when click opening a link which may be harmful. [21] examine the effectiveness of these anti-phishing tools and found the warning messages did make participants become more cautious and selective of sites they are visiting, although, the average visitation rate of sites with warning messages is still up to 83%. A simple cost-benefit analysis may reveal the reason. Personal information is not a physical object that users can see when it is lost, users may not fully grasp the value of the information they are giving out and may not realise the harm right away [22]. For example, users giving out credit card numbers may not notice until weeks later when their bank statement is issued with suspicious payments. In addition to [23] and [24] findings where convenience and desire to access the site and use the services suppressed the fear for

security concerns. This is more evident when sites are offering very promising returns for visiting such as promotional code or even free product samples. These benefits outweigh the cost that users perceived to be at risk hence proceeding and ignoring the warnings while the true risk of such action remains unrealised and causes further harm in the long run. So there exists a threshold to which users will choose to proceed or exit from the site. More research is needed to thoroughly examine factors that contribute to this cognitive cost-benefit analysis that influences the threshold beyond what was examined here.

Another factor for ignoring the warning messages is the difficult technical languages used. [25] found participants disregard the security warnings and chose to proceed to the site because they could not fully understand the messages with technical jargons. The lack of understanding then affects users' evaluation, in most cases, underestimating the significance of the risks. Interestingly even those with technical backgrounds are still ignoring the messages hinting their understanding of the high false-positive nature of the tool [23], [26]. This is supported by a study that found some anti-phishing tools possess as high as 88.5% of false-positive warnings [19]. So these studies have demonstrated security warnings are populated with technical languages that the average users could not understand, and in addition to that, the warnings are overly produced. These combinations of issues triggered the users to ignore these warnings altogether.

2.2 Fail to Utilise Security Indicators

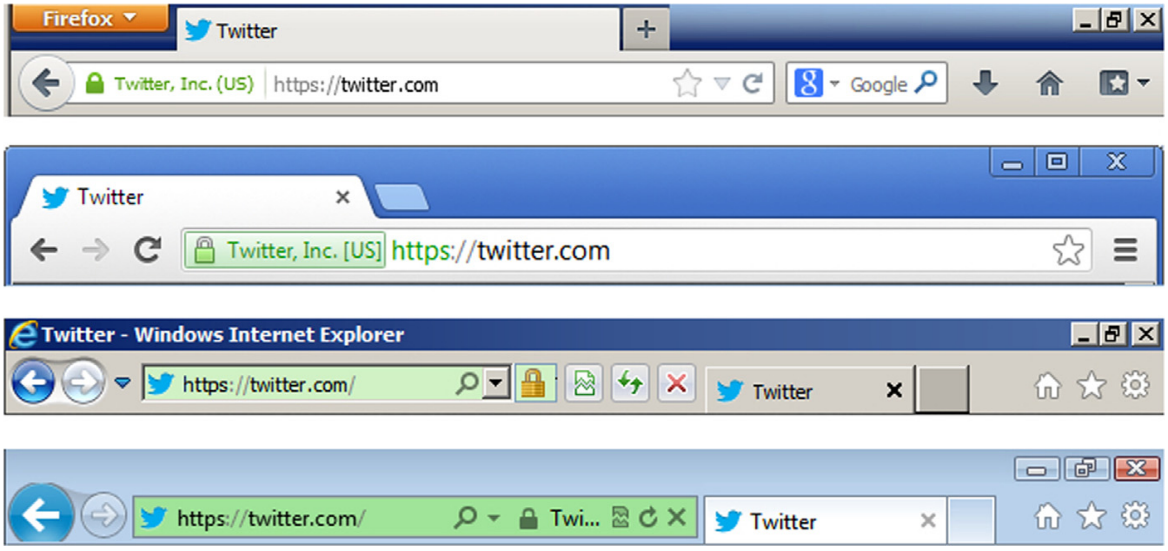


Figure 2.3: Security Indicators
Source: [15]

Modern browsers and anti-phishing extensions are providing indicators to alert users of potential threats. Fig. 2.3 are some examples of security indicators to help users gauge the safety of the website. These are different from security warnings where indicators are *passive* icons or text that may change shape/colour however it does not actively interrupt the user access flow with the manual options. [15] found using eye-tracker in an experiment and requiring participants to view and make a decision whether the website is safe/unsafe to proceed, the majority of novice computer users did not utilise security indicators to their advantage. The total length of time users glanced at indicators are insignificant compared to the total duration that users spent looking at the site before making a decision. It is reasonable to argue that time may not be the most appropriate unit of measure in this context. A non-text based indicator does not need the user to stare at it for a significant duration for it to be understood hence not marked as an area of interest [15]. Use of time as a metric for cognitive processes operates on the theory that **the point that people are looking at for a long period is what they are concentrating their cognitive processes on** [27] (this theory will be crucial for the experiment conducted in the second part of this project). This is one commonly practised approach in cognitive research as no technology to date can confidently measure human cognitive processes. Another evidence suggests that users are unaware of the usefulness of security information these indicators give, more importantly in some cases, they don't even know where to find such information (failed to notice the indicators presented on screen) [26]. [18] reported the need for modern security indicator designs to be more salient. One example is the use of colour coding and ordering of wireless options from safest (green) to least (red) when selecting a WiFi connection like common traffic lights. Use of identifiers that the majority of users are familiar with (traffic light colours) help users in making a quick and effortless informed judgement which has been shown to better assist security decisions [28].

It could be argued that users took fewer precautions for day to day web browsing because of the perceived low-risk information handling. For example just visiting a site to read an article, browsing through an online catalogue where a certain level of anonymity can be maintained. As a result, users are more conservative in spending unnecessary effort to verify the authenticity of websites and contents with the security indicators in line with previous findings that convenience outweighs security in this case [10], [24]. However, this understanding contradicts with [29]'s finding that even when users are engaged in formal high-risk information handling (consciously giving out personal information) with an example of an online e-voting system, only 10% of the users glanced at security indicators with the attempt to verify the authenticity of the site before proceeding. These are explained with the inadequate useful information indicators offered in helping the user to make the right informed decision [29]. Users do not understand the message that the indicators are conveying, causing ambiguity and impacting the judgement on the extent of the risks and appropriate action to perform hence wrong decisions are made posing further compromisation [30].

2.3 Human Psychology

[4] claims humans are the 'weakest link' in the cyber world and complex security systems failed at preventing social engineering techniques hackers exploit. One evident case is the very recent phishing attack on the notable Shark Tank host Barbara who was scammed over $400,000 but was later able to recoup the loss [31]. Targeting attacks towards humans rather than computer systems is also less expensive to do. An attempt to bypass a firewall and gain access to a computer system to obtain information would require an abundance of computational power and hardware equipment to successfully carry out such stunts [4]. The social engineering methods commonly used by scammers have three common psychological goals: 1) "provoke the subject to perform certain actions", 2) "block rational thinking zones" and 3) "influence the emotional and affective sphere". The more goal combinations attackers can employ, the higher the likelihood they can obtain the target's information [32]. For example, email triggering emotion of fear regarding law enforcement (goal 3) should you not follow the link to pay the fine (goal 1) within the next 72 hours (sense of urgency + seriousness, goal 2). This will be explored further in the third section of this paper on components of the phishing emails. In another study, participants read a series of emails and decide whether to upload/download attachments or report the email as spam. Participants encounter fake emails at a rate of 1%, 5% or 20% replicating different sensitivity levels of anti-phishing tools. Results indicate high prevalence effects on those in the 1% condition with significant misses of harmful emails. While the tools can filter out some harmful contents, should it fail to detect some and users encounter them, their ability to effectively take notice and react to those emails was hindered. This is accounted to the lowered frequency of encountering harmful emails in general causing lack of exposure and too high reliance on such tools [33].

Maintaining situational awareness is also another key to ensure safe online activities. Situational awareness helps people easily obtain necessary information for better decision making when encountering problems and have to choose a course of action [34]. Evidence suggests anti-phishing tools cause lowered situational awareness from the poor computer-human task handover. When security warning messages are given, commonly the warning does not provide users with sufficient information about the nature of the attacks, what prevention or scan has the tool done, and recommended course of action [19]. Furthermore, these messages are often displayed when users are focusing on other tasks and least expect a warning. This is because the tool's processes are often done in the background away from what users see on screen so they are unaware of the initiation of these scans [22]. As a result, users are likely to choose any default option that allows them to disregard the warnings and continue with their work indicating the divided attention principle at play [30]; to prevent cognitive overloading from performing multiple tasks at once, users will eliminate or postponed tasks that are deemed less priority which in this case the security warnings [35]. The combination of lack of awareness and divided attention principle influence users to choose the wrong action and fall victim of these scams.

Another fascinating pattern of human psychology is risk aversion and gain maximisation. This is similar to the cost-benefit analysis of ignoring the warnings raised previously that the perceived risk is less than the benefit hence users choose to proceed to dangerous sites. When users, or humans in general, are presented with options, we tend to choose the option that would lead to us maximising the gain while minimising the loss factoring in the probability of such occurrence happening [36]. For example, if you have to choose between receiving $500 cash or flip a coin, if it lands on head you will get $600, if it lands on a tail you will get $0. Most people would choose to grab the $500 since the perceived risk of losing the money is too high (50% chance of landing on tail and losing everything just for an extra $100 is not worth it). Consider a second case where you can receive $500 cash or flip a coin, if it lands on head you will get $600, while if it lands on a tail you will get $450. With this case, some people might change their mind and gamble with the coin because the potential loss here becomes less than that of gaining even though the probability of both outcomes are still the same (instead of 50% chance of landing on tail and losing everything, now I am just losing $50 but I also have 50% chance of landing on head and gaining that extra $100. Gain 100 >lose 50 hence gamble). Consider a third case where you can receive $500 cash or flip a biased coin. If it lands on head you will get $10,000 while if it lands on tails you will get $0. The coin is biased such that there is an 95% chance of landing on head. In this third case, most people would still gamble with the coin because the perceived reward or gain is larger than that of loss (gaining 20 fold of original value justified losing everything with mere 5% probability). It can be derived from these scenarios that if the perceived risk of loss is lower than or equal to that of gaining extra then people would accept the risk and gamble in hope to maximise the amount gained. While if the perceived risk of losing is higher, then people would switch into risk aversion mode minimising the loss by taking the default option to save whatever they can [36]. These principles are what drives lottery systems and casinos into making you gamble more and more in hope to hit the jackpot. Applying this to cybersecurity, the loss that users perceive is their personal information while the gain that they believe outweighs this loss could be access to information, entertainment, or discount codes leading to users accepting the risk and gambling with these conditions where you will never win. Some users do not understand the significance of such disclosure that even knowing your name can lead to not only compromising your physical safety but also a more damaging loss to finance and other sensitive information as well. This issue builds upon [22] findings previously that personal information is not a physical object that users could see when it is lost hence the perceived loss is much lower than reality. One clear exercise to demonstrate this is for the reader to try searching your name online and see what results come up on the search. Some may find articles or images of yourself dated back years ago, this demonstrates how much information about you can be extracted by someone who has harmful intentions. This is in line with findings that reports on cyberstalking, child predator, harassment are on the rise. The majority of these criminal cases originated online especially towards younger users because they may not fully grasp the full-scale danger of disclosing their personal information publicly [37].

2.4 Chapter Summary

The first two proposed rationales indicate issues with the usability of security measures to warn users of potential threats with difficult jargons and users failing to notice the indicators. The second and third rationale hinted at the overlooked vulnerabilities in human cognitive psychology, namely the prevalence effects and divided attention, that influence the ability to effectively utilise available security indicators to their advantage. The effects cascade as a result of poor interface designs and computer-human task handover undermining situational awareness. As presented, many research has pointed out that these automated anti-phishing tools are populated with combinations of multiple design issues and pitfalls affecting the overall usability, hence leaving users exposed to cyber-attacks. These findings and arguments are the grounds for the current study to shift attention towards human users to further investigate *why* such actions and decisions are made and what can be done to tackle these issues.

Chapter 3

Properties of Human Users

As introduced in the above section, we will now shift our focus towards aspects of human users. This section will address *what* contributes to certain groups of people becoming more susceptible to online scamming than others which will entail *why* people are still falling for these phishing scams.

3.1 Signal Detection Theory

An important concept to explore first is using signal detection theory to model the detection performance of harmful contents by both human users and automated tools. Although very similar, this theory is not to be confused with the previous example on the cost-benefit criterion which focuses on the user's ability to *choose an appropriate action when encountering a threat*. The current theory concerns the user's ability to *detect and classify threats*. Signal Detection Theory is a model for analysing the ability to differentiate between the target (signal) and distractions (noise). Possible outcomes are grouped into 4 categories

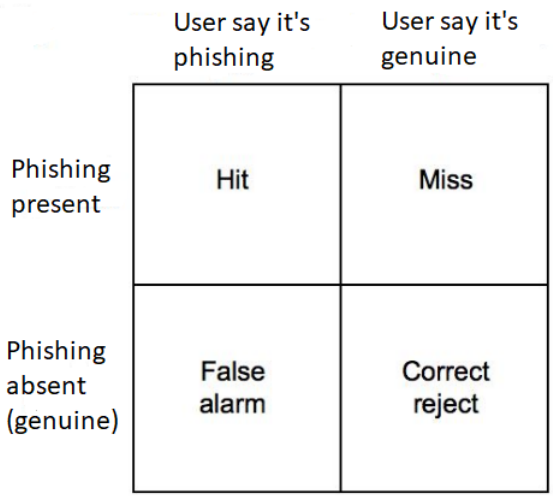


Figure 3.1: Phishing and Signal Detection Theory

Fig 3.1 demonstrates the four possible decision outcomes of Signal Detection Theory with application in cybersecurity. What influences the accuracy is a criterion, it is a benchmark to which respondent will report the target is present or not.

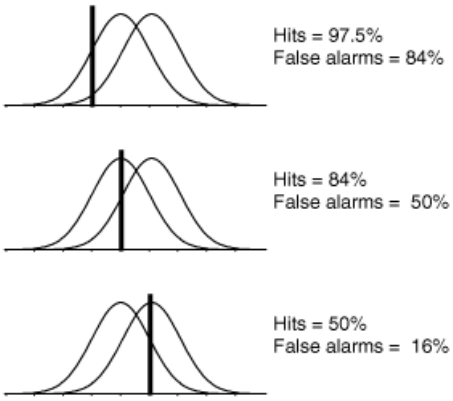


Figure 3.2: Effect of shifting the criterion
Source:[38]

Fig. 3.2 demonstrates how the criterion will affect each of the 4 possible decision outcomes. Should the criterion be raised high (the vertical bar is moved towards the right), there will be higher cases of misses and correct rejection while lower hit and false alarm. While if the criterion were lowered (vertical bar is moved towards the left), there will be higher cases of hit and false alarm but lower misses and correct rejection: A trade-off between security and efficiency. (Note the percentage numbers reported in the figure above is given as an example, it does not reflect the real amount changed) The criterion is dynamic such that it may be raised or lowered depending on numerous factors such as age, experiences, or education, and varies from person to person. These factors will be explored in further details.

Applying this to the current project, an ideal performance for both users and the automated tool is maximising hit and correct rejection while minimising miss and false alarm. Meaning when a phishing content is present, users and automated tools should be reporting that harmful content is present. And when safe content is present users and automated tools should not be reporting any warnings otherwise. The previously reported average of 88.5% false-positive warnings reveal the majority of automated tools in the market have too high phishing sensitivity level (criterion bar is too much to the left) [19]. This trade-off between safe and effective automated tools while not overly producing false-positive alarms by setting appropriate criterion is a very challenging task faced by developers which is one of the major reasons contributing to the poor usability of these tools. Reports on the performance of human users in classifying phishing emails are also alarmingly low. [39] reported corporate employees detect on average seven out of 10 generic phishing emails in a classification simulation task while four out of 10 spear-phishing emails were detected. Average phishing detection abilities throughout a number of phishing types were found to be mere 40% accuracy. Although performance on certain types are higher, the average phishing detection skills of users are deemed poor [39].

3.2 Consequences of High Misses and False Alarms

There is a greater cost associated with missing a threat than reporting a false alarm. [40] reported each cybersecurity breach cost on average $412,000 which could be as a result of failing to recognise just one phishing email. Other costs associated with a successful infiltration could involve loss in the confidentiality of data, organisation's reputation, data tampering and many more. False alarm on the other hand mainly concerns the loss of productivity and time. Employees spending time dealing with false alarm notifications or alerts produced by automated detection systems could be an opportunity cost for the business [39]. A study reported majority of security operation centres received on average 10-20 alerts per day with some receiving up to 50 or more taking up between 2.5 and 5 hours each day just dealing with the alerts alone. Shockingly out of these alerts received more than half are false alarms and does not require any major intervention to the system's operation [41]. With such high figures, operators are prone to alarm fatigue: the human sensory system is overloaded with excessive numbers of alarms resulting in desensitisation causing delayed response time and a higher chance of missing actual important alarms. The effects of alarm fatigue are very well established and are under active research in the healthcare industry where nurses and doctors missed important alerts produced by patient's life support systems and monitors resulting in death and serious injuries [42], [43]. This links to the previous section discussing why warnings are being ignored, as a result of alarm fatigue and desensitisation in addition to the limited cognitive processing capacity [22], [19]. The same study of [41] further detailed the most common response of operators to high false alarm rate is to adjust the sensitivity of the alerting feature to reduce the alert volume or ignore certain categories of alerts altogether (making the system more vulnerable as attacks may go unnoticed). Despite this adjustment, it seems that such approaches are ineffective as many security centres are still reporting a high volume of alerts [41]. The study did not outline the effects of adjusting the threshold on the security level and any changes in cyberattack trend however referring back to Fig. 3.2 on Signal Detection Theory, it could be predicted that decreasing alert sensitivity (moving the vertical bar to the right so it produce less alerts) would result in lower false-positive and hit with increasing correct rejection and miss. Meaning the high volume of alerts still experienced could be from actual cyber breaches that were missed as a result of a more vulnerable system. Further investigations are required to verify this reasoning.

3.3 Factors Influencing the Criterion

With the presence of a criterion that users subconsciously use when deciding whether the content is safe or unsafe, it is crucial to understand what factors contribute to setting such a criterion.

3.3.1 Training

In a pre and post-test training program, [44] report having employee training can increase phishing email detection effectiveness. Employees are presented with a series of emails and must classify these as fake or genuine before answers are revealed followed by training on strategies to detect email scam in workplace. Results showed a significant increase in the post-training detection simulation performance with different phishing email samples. [45] and [46] supports the claims that training does significantly improve phishing detection (hit rate) over the long term. [45] further adds that false positive rate also decreases indicating training not only increases detection hit rate but also better fine-tuning classification between phishing and genuine email. This contradicts with the Signal Detection Theory where higher hit rate would always be coupled with higher false-positive rates revealing the theory's limitation with complex modelling such as human learning abilities [38]. In fact, to the writer's understanding, no statistical or mathematical model has come close in replicating human cognitive abilities as noted in the field of AI and machine learning.

A skill once learned can be unlearned or forgotten if not practised regularly, cybersecurity skill is no different. It is recommended that consistent training are put in place to ensure retention of skills for maximum online protection [47]. Interestingly [48] investigated cybersecurity training among primary school students and found those with training performed 14% better at recognising phishing email than classmates that did not receive training. However, after four weeks the effects of training wears off and performance reclined to the pre-training level which hinted, in addition to [47] claim, a retraining every four weeks could be the optimal frequency. [48] also noted age was a major factor where older students outperformed those younger possibly due to experience with technologies and learning capacity. Age factor will be explored in a later section.

3.3.2 Prior Experience

[49] found participants who had prior experience being victimised by phishing scams have increased awareness and are better at detecting phishing emails. Interestingly those who had financial loss performed better than those with loss of credentials however with small sample size, the author note, the findings are inconclusive without further evidence. This is in support of [22] finding that credentials are not physical objects or entities that users could realise the harm right away when lost, unlike finances. [50] conducted an experiment presenting phishing banking URLs to participants and measure click and interaction behaviour of these sites. Those suffering highest prior financial losses to online scams and those using online banking regularly are the least to click visit and elicit the most cautious behaviours when browsing the sites. Younger participant groups that lack online banking experience

and phishing awareness, and those with lower prior financial losses are the worst performing in terms of visitation to these sites. Interestingly those who self-reported high in phishing awareness performed similarly to those without experience in terms of high visitation rate to phishing sites. These findings infer that overconfidence in cybersecurity skills can result in poorer performance due to lowered alertness and underestimating the risks presented [50]. Lastly, participants that often ignore scam warning messages and those reported high trust in the bank is the worst-performing at reporting the site as legitimate and intend to transact with the site giving out personal credentials [50]. The finding is supportive of [19] associating warning messages as having a high false alarm record and difficult technical jargons. Evidence indicates this participant group lacks understanding of cybersecurity and the techniques used by scammers to deceive their target, hence, they uphold a wrong concept of trust believing that they will always interact with the real bank and not a scammer. With these findings, it could be concluded that users at both ends of the spectrum (those lacking online experience and those overconfident of their skills) are susceptible to phishing and that prior experience exposed to scams can contribute to more cautious behaviour online.

These findings can be further explained with [51] well-known psychological theory on Operant Conditioning; a theory on learning by encouraging/discouraging certain behaviours through reward and punishment.

	Reinforcement	**Punishment**
Positive	Something is *added* to *increase* the likelihood of a behavior.	Something is *added* to *decrease* the likelihood of a behavior.
Negative	Something is *removed* to *increase* the likelihood of a behavior.	Something is *removed* to *decrease* the likelihood of a behavior.

Figure 3.3: Operant Conditioning
Source:[52]

Fig. 3.3 above outlines the components of the operant conditioning [52]. The theory is the key that underlies many daily activities. One example is the education system, a student who completed their homework and study the materials will be *rewarded* with high marks while students who do not study or do their homework will be *punished* with a lower mark and may have to repeat the class. This model helps shape the majority of students encouraging them to elicit desirable behaviour which is to study and complete their homework. Other applications of this theory include animal training, law and punishment, parenting, marketing, and even psychological therapy [51]. [52] and [53] further claims that the intensity of punishment or reinforcement positively correlates to how well the subject retain and learns the training. In [50] findings, those with high prior losses experience an intense negative punishment (losing money) hence they are the least likely to elicit the behaviour of visiting the site again to avoid this repeated loss. While those who experience a less intense punishment will be more likely to repeat such behaviour again following the gain maximisation concept perceiving benefits to be higher than losses gambling between convenience and safety [36]. The most effective form of behavioural shaping is positive reinforcement, although is challenging to implement in the given

context as success/safety is not a physical object user can see as previously mentioned [22]. A user acting appropriately and safely online does not actively experience or feel that they have been rewarded (although it could be viewed that by being safe and not scammed is the reward). The perceived absence of reward, in this case, could overtime prompt users to seek their own reward and potentially putting themselves at risks such as accessing dodgy websites for promotional code or entertainment [53]. Positive punishment on the other hand is the least effective when trying to reduce a certain behaviour because it can induce side effects such as fear, stress, and anxiety [51]. These conditions are the root cause of technophobia; a mental condition common among the older population afraid of interacting with any form of digital technologies. Patients diagnosed with technophobia often suffer from a poorer quality of life and lower social inclusion fueling further problems such as depression, loneliness, and self-harm [54]. Even though such effects could prompt users to become more cautious when browsing online contents however the risk of inducing technophobia is greater than the benefits hence deemed an ineffective and unethical method.

3.3.3 Age

The lack of online experience introduced previously could be attributed to age. Younger generations although are more familiar and confident with technologies in general, the purpose and interactions for using them is mainly for browsing news, social media, or educational where they are protected to some extent (i.e. kids friendly contents or safe browsing mode). This differs from adults where the purpose of interactions with technologies extends to conducting online bankings, managing documents, and working, where handling of sensitive information is involved and is often the target of scamming [55]. This lack of exposure in the younger generation makes them more vulnerable to phishing attacks when they do encounter them since they do not fully grasp the full danger and potential harm associated with such attacks as reported with growing criminal cases on cyberstalking [37]. Remarkably, the current findings further demonstrate that the prevalence effects [33] previously outlined span across a much wider age range than originally found with only adults. This argument does not assert that online protection currently put in place for kids should be removed, rather, they should concurrently receive adequate training and skill development since a young age. Such development will create awareness of the issues and equipped them to be able to act appropriately when facing cyberattacks that the filtering tools may have missed similarly to the issue on the effectiveness of automated anti-phishing tools discussed.

When talking about age, another vulnerable group to cyber scamming is seniors. Seniors are the most common demographic to fall for any types of online or telephone scamming, and with an ageing population in many countries, cyber fraud targeting this population group is expected to grow [56]. Contrary to popular belief, the elderly population is also another major internet user group with 79% of seniors aged 57-65, 67% of 66-74 years old, and over 40% of those aged 75 or more reporting they use the internet regularly [57]. Although seniors are familiar with using and understand the internet concept, this does not infer they are familiar with applications and platforms on the internet such as

online shopping sites, banking, or social media. [55] also found a mere 26% of internet users aged 65 and over reported they feel confident and prepared to deal with potential cyberattacks or technical errors when using the internet. This complements arguments in the above section that unfamiliarity and lack of experience with online platforms can lead to a poorer judgement of harmful contents [58], [50]. Findings further demonstrate seniors are aware that they lack knowledge of online platforms from the low confidence rating making them vulnerable, however for unspecified reasons, they choose to remain an active user of these internet platforms. Potential justification for such contradicting behaviour is the benefits that online platforms bring to daily activities such as groceries shopping, scheduling services, or finding information. It is also an effective platform to maintain a social connection with friends or family avoiding psychological issues associated with those who don't similarly to technophobia patients [54]. Utilising online platforms for social connection although can also bring harm. The issue of loneliness among seniors is growing, with this, scammers often exploit this sensitive emotional state of seniors, gaining trust, building rapport and (romantic) relationship with the target hoping to gain personal information [56], [58]. There are a number of reasons seniors are often a target of online scamming, this includes: 1) wealth. [59] and [60] reported those aged 50 years old or more own up to 70% of the total US nation's wealth. It would be more likely for scammers to obtain a large sum of money from seniors compared to those who are less financially equipped. 2) Politeness and trustworthiness. These are common traits among the senior population hence scammers exploit this with often novel scenarios that they are urgently in need of finance or in some cases, pretend to be someone in close contact with the target to increase the chance of success [61]. 3) not only elderlies but most people when falling for scams don't know where to report the crime or the feeling of shame overwhelmed any subsequent rational thinking. A survey found seniors are afraid of their carer's opinion of them if they do report falling for a scam. Relatives might felt that you can no longer take care of your personal finances anymore hence stricter measures will be put in place to control your activities online. Scammers realised this getaway potential without being reported hence direct their effort towards seniors [58], [60]. 4) poor memory and cognitive capabilities. Over 40% of people aged 65 or older are diagnosed with memory impairment in the US [62]. Seniors may have trouble accurately recalling the event and the delayed realisation that they have been scammed further made it difficult to trace back the source. 5) health promises. Unsurprisingly reports of health and insurance scams among seniors are on the rise. Scammers often target common health issues or symptoms that relate to ageing and offer insurance packages claiming to cover the cost of medication and life insurance sum to loved ones when they passed. Elderlies, in addition to vulnerabilities above, believe in the advertisement and fall for these scams.

3.4 Chapter Summary

In summary, this current section has explored the effects of training and prior experiences on aiding detection of phishing scams as classified with the Signal Detection Theory. The theory also noted challenges faced by developers in setting an appropriate criterion for automated tools which contributes to the usability issues outlined in the previous section on high false-positive rates causing alarm fatigue. Evidence also suggests human users are generally poor at detecting phishing attempts with those lacking online and prior experience performing the worse. Age contributed to the lack of experience among younger generations and personality traits of that of seniors significantly contributed to these population groups becoming vulnerable to scams.

Chapter 4

Anatomy of Phishing Contents (Email)

The first section outlines usability issues with advanced phishing detection tools and an introduction to human psychology which is the reason *why* rates of data breaches and scams are growing. Expanding on human psychology, the second section explored the human aspects of *what* contributes to certain groups of the population becoming more susceptible to phishing and online scamming than others. In order to protect ourselves from these scams, we must first reverse engineer these phishing materials to identify the thoughts and techniques used by scammers. This last section will break down phishing emails and explore *how* scammers design emails such that it is convincing to the user.

4.1 Emotional Manipulation

Recall that there are three common psychological goals as discussed in the earlier section that scammers are aiming to achieve (provoke the subject to perform certain actions, block rational thinking zones, and influence the emotional and affective sphere) [32]. Emotions of fear when paired with punishment were found effective at provoking compliance [64]. One example would be the imposing of fines and imprisonment for those breaking COVID-19 lockdowns. The fear of law enforcement and punishment in the form of imprisonment would encourage people to comply with government policies. In addition, this is in line with risk aversion and gain maximisation theory where people will act in a way that maximises their own benefits while minimising loss [65]. Having to pay fines, in this case, would be an undesirable action to do. [66] further adds the attempt to exert excessively strong fear in the context of cybersecurity can result in individuals being alarmed and starts to question the reliability of the information being presented. This is different from the technophobia condition previously outlined that is triggered as a result of intense punishment. The attempt by scammers to exert excessive fear is often coupled with irrational claims. Receivers that do not exercise common sense assessing the situation could fall victim to these scams [66]. One famous example is the Hitman email. It will claim that money has been paid to kill or cause serious harm to the receiver of the email and they have 12 hours to match the payment in exchange for protection. An average person receiving such an email should ignore or report a potential scam to local law enforcement as the likelihood that it is a scam

attempt is higher than that of the claimed assassination happening [67].

Another effective emotion hackers often exploit is a sense of urgency. The benefits of urgency, when paired with fear and punishment is it ensures users have as little time to think rationally as possible before acting, which often results in compliance to the demands to minimise loss [68]. In the same Hitman email example, the constraints of 12 hours are to rush the receiver into panicking that they only have a very limited amount of time with their life at stake hence an action needs to be taken right at that instance without further thinking. Which in cases that belief in such a claim will comply with the payment demands. A generic guide when exercising common-sense is to compare the expected professional behaviours of genuine companies with the current scenario. For example, it would be appropriate, in a professional context, that any urgent communications regarding attempted financial fraud or unusual activities in credit card usage be done via a phone call or SMS by an authorised person of the bank instead of email communication asking you to visit a webpage. Such activity should be deemed suspicious and necessary actions should be taken to verify the message prior to giving out any informations to the sender of these emails [69].

In addition to fear and urgency, authority is another well-known factor that can provoke obedience. One renowned psychological study of authoritative power is the Milgram shock experiment. Participants were assigned the role of a teacher while a student (a confederate) sits in another room, both can communicate with each other but cannot see each other during the experiment. The teacher will ask the student questions and if the student got the answer wrong, the teacher will be instructed by an experimenter dressed in a lab coat (to represent power and authority) sitting in the same room with the teacher to press a button on a machine which will deliver an electric shock on the student. The shock will start from mild to life-threatening lethal power as more answers are wrong throughout the experiment session (the shock is never actually delivered but the student sitting in another room will scream when the button is pressed as if he/she is in pain). Results found over 65% of participants although showed anxiousness to comply, did continued to deliver maximum lethal shock power following the order of those in power, in this case, the experimenter dressed in a lab coat as authoritative figure [70]. Scammers often exploit this human weakness, in the earlier COVID-19 example, if the message were sent by a person you perceived as having authority (e.g. from a police department, defence force, the government) the receiver would be more likely to comply with the instructions given to avoid any conflict with those figures.

4.2 Visual Deception, Familiarity, and Trust

Another feature scammers used when prompting users to click certain links on email is link manipulation. In programming languages like HTML, the displayed text and the hyperlinks can be different. Links to malicious sites can be hidden for example click here or www.facebook.com. If you click on either, both of these links will open www.google.com however it demonstrates potential malicious sites could be hidden under these texts [71]. Other similar techniques include using a URL shortener, redirects, DNS spoofing and many more. Another form of link manipulation worth discussing is misspelling the link. For example, a real government site www.centrelink.com.au scammers could register a domain name with very similar spellings **www.centerlink.com.au** (r and e swapped places) which is used to phish user's credential [72]. This also applies to email addresses such as **something@gov.in** vs **something@govs.in** where there is an extra 's' after the word gov. A quick glance of these texts can very easily miss the manipulation and fall for these scams. The science behind this can be explained with the well-known internet memes circulated around 2013 that pepole can sitll raed wrods and sentecnes evne thoguh the lettres are jubmled up (people can still read words and sentences even though the letters are jumbled up). Humans don't read every letter of a word; rather, we rely on specific character positions of the word (mostly beginning and ending) to quickly judge what the word is [73], [74]. This is in line with the previous findings presented that human cognitive capacity is a limited resource and with a lot of information to be process from all human sensory inputs at any given time, reading letter by letter will be very time and resource-consuming [30], [35]. Scammers exploit this ability to associate manipulated links with a genuine one to lure users into clicking it. A simple but effective strategy to counter these phishing attempts is to hover your cursor above the hyperlink before clicking open. Hovering the cursor above the hyperlink will reveal the actual link to the website that is about to open when the link is clicked [71]. A better practice, instead of clicking on the given links in the email or attachments, open the organisation's webpage and navigate to the login page yourself. A search on a reputable search engine would generally display a genuine link to the queried organisation [73]. These methods will ensure that you are always transacting with the genuine server.

Making phishing contents looks like a genuine one will increase the chance of success. It is a simple statement however the science behind such phishing content design utilises many psychological aspects. A key feature determining the success of phishing email is how much the target *trusts* what they are reading. And to create a sense of trust, one precondition that must be met is *familiarity*. [75] defines the two terms as familiarity deals with "an **understanding** of the current actions of other people or of objects" while trust deals with "**beliefs** about the future actions of other people". Consider an example of an online shopping website. My *familiarity* with the online shopping site will increase as I visit the site more often. This is because I have a better **understanding** of features and layout of the site and basically how the site works (e.g. where to click on to submit an order, where I can find information about a certain product, who to contact if I have an issue). As I become more familiar, my level of *trust* towards the site also increases. If I provide them with my credit card information to submit my order, I expect (or predicting the near-future actions) that they will only charge me

the amount for the product that I want to purchase and not use the information for anything else. In this case, I rationalise and reason, based on previous exposure and familiarity of the site to trust that they will also behave rationally. This example also utilises the core concept of learning: operant conditioning [51]. My behaviour and attitude towards the site become more positive each time I interact with it as I am positively rewarded (successfully purchase something on the site, the emotion of joy is the reward). It only requires one instance of positive punishment (I got scammed or charged extra without me knowing) to ruined the trust that I have with the site [51]. This infers that building trust and familiarity is much more effortful than ruining such a relationship with clients. [76] applies the concept of trust and familiarity with e-commerce websites and found both factors fluence intention for users to inquire about products with the site and intentions to purchase the product with them. They also briefly note the concept of a minimum threshold of trust and familiarity required in order for the above actions to be elicited. Whether this threshold varies depends on our intended action with the object or person are potential research topics to explore further (e.g. more trust level is needed when I am giving out my credit card information compared to when I am giving out my phone number contact details in the same context or not for example). Applying this to phishing context, building familiarity and trust from the ground up as mentioned is time-consuming and effortful for scammers. Instead, they chose to impersonate well-known corporations and organisations where preexisting trust and familiarity with the general population are already established which will save them time, effort, and increase the likelihood of success. This is also the reason why corporations around the world are promoting cybersecurity awareness and consultations for their clients in hope to maintain the trust and ensure company image is not falsely jeopardised by groups of scammers [76].

Interestingly familiarity with a certain organisation could also assist users in detecting a phishing attempt when the behaviours deviate from normal [75]. For example, if the preferred communication channel between you and a company is selected as SMS but an email is received from the company it could be suspected a phishing attempt. Users similarly should exercise common-sense to attempt to verify the source of the message before acting [69]. It is important to note insufficient evidence was discussed in the original study. A more thorough investigation is needed to support this claim that familiarity does assist in better detection skills [75].

There are also trends in the industry and corporations scammers often impersonate. Contrary to popular beliefs that most phishing emails are of banking institutions or online shopping, a study reported cloud services came first in domains that are often impersonated, followed closely by financial services, social media, e-commerce logistics, internet telecommunication and government respectively [77]. This is in line with expert's commentary that brands like Microsoft, Dropbox and DocuSign offering cloud services with large pre-established users trusting the organisation are frequently impersonated in phishing scams. In addition to the link manipulation and misspelling techniques scammers used, they also utilise familiar visual cues such as a logo or copyrighted symbols to further gain the trust of users. [78] found participants in a phishing study reported trusting emails with the presence of recognisable logo or copyright symbol inside an email more compared to just plain-text mention of the company name or none at all. This not only presents professionalism and genuineness but is also in line with the

concept of familiarity and trust above that these visual cues help trigger memory of prior interaction with these organisations [75]. Participants of the study also mentioned factors with the format and wording of the email that helped them decide whether to trust an email or not. Phishing emails are often populated with spelling mistakes, grammatical errors, images that are low quality, or layouts that are not perfectly aligned [78], [79]. Recall that scammers do not want to put in a lot of effort into preparing generic emails. They have a better chance of succeeding sending many average looking emails but to a wide audience than spending time preparing a few good emails [80]. These design mistakes became good indicators to users that such email may be designed by a non-professional who does not actually represent the organisation.

4.3 Marketing Strategies

In addition to techniques used by scammers, it is also relevant to briefly review techniques used by marketing materials as both have similar goals in capturing user attention and increasing click rate to their websites based on contents and links sent in the email.

When scammers send phishing email, making the email pass through an automatic detection system and not marking it as 'spam' is the first of many challenges. Even if the email is successfully delivered, this does not guarantee the target will even open and read the email. We are currently living in a face-paced era where professional workers are receiving overwhelming number of new emails each day. With many emails, limited time and cognitive capacity to process all this information, the average response rate to emails is only 25% with people spending on average four to eight seconds glancing through each email [81]. The subject line is then the most important part of the email to grab user attention into clicking opening it. [82] recommended subject line should be five to seven words in length summarising the purpose and main message of the email. These words must be carefully chosen such that it achieves the three psychological goals previously mentioned in order to maximise success potential whether for marketing or malicious intent [32]. The choice of words is highly dependent on the industry and purpose of the message, for example, the use of emoji in marketing emails along with words such as 'big sale', 'introducing', 'exciting' would better stimulate the interest of the reader into finding out more about this sale program while if these are used in informative or corporate emails could be seen as unprofessional and less trustworthy to open [79]. [83] reported a list of words that are commonly used across many industries in their email subject line including 'upgrade', 'just', 'content','go', 'wonderful'. According to the researcher, these words are effective to use in the email subject line partly attributed to the suitability in both formal and informal emails [83]. [84] further adds the use of the receiver's name in the email subject line prompted a 14.6% increase in open rates compared to the ones without. The use of personalisation techniques not only indicate the seriousness of the message but also the source of the email from someone who is of close contact [108]. This argument will be discussed further below.

4.4 Personalised emails

The act of personalisation is not uncommon, especially in the retail and hospitality sector. One prominent example is when visiting the famous cafe chain, the barista will ask for your name to write down on the coffee cup and calls you to the pickup window when the order is complete. This is to add value to the product or services offered and create memorable experiences for revisitation. Personalisation in email operates on the same psychology. Emails personalised to you may contain your personal information such as name or preferences to customise the content of the email best suited to you in hope to increase open rate. The same principles are implemented in social media and video streaming services where contents of the site will differ according to your preferences and pages that you follow. Many research [82], [83], [84], [79], [85] agree that personalised email does indeed prompt for higher open rates however this is also a double edge sword. Overdoing or inappropriate use of the recipient's personal information can risk offending the receiver and loss of trust.

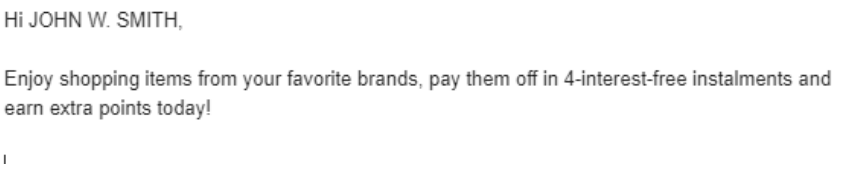


Figure 4.1: Sample Marketing Email

Fig. 4.1 above for example, although the email has attempted to deliver personalised contents, it became awkward when the name of the recipient is written in all capitals and with the use of a middle name initial which is very unnatural. In this case, it could lower the reputation and opinions of the recipient towards the sender given the inappropriateness of the email [85]. Another similar pitfall of a too personalised email is not only inappropriate, but a risk to an invasion of privacy and potential abuse of the information for further social engineering should the information were mishandled.

4.5 Chapter Summary

This final section discussed components and considerations that goes into designing a phishing email. The use of psychological manipulators such as fear, urgency and authority inhibits victims to exercise common sense and think rationally. Technical methods such as link manipulation, misspelling domains reveal further limitations in support of previous findings on human cognitive capacity. The factor of familiarity was mentioned which contributes to users willingness to interact and trust potential web pages. Techniques to identify potential phishing email relying on design mistakes were briefly outlined. Finally, techniques used in marketing and personalised emails were explained how it could be applied in phishing email designs.

Chapter 5

The Framework

Following relevant literature reviews, section two of the current research project is to develop a generic framework to aid future human-centric cybersecurity research particularly projects that involve experiments with human participants. The following chapters document the first-hand processes and considerations the researcher of this project has experienced when conducting an email phishing study with eye-tracking technology. The study will investigate components of emails that participants look at that prompted them to decide whether the email is a phishing attempt or a genuine email. These components will reveal the thought processes of online users. An eye tracker will be used to capture the eye gaze movement while participants are tasked to classify the presented emails. This experiment will serve as a demonstration on how the proposed framework could be applied in the real world research, which consists of 4 related tasks in the following order: 1) Stimuli design; 2) Research and selection of tools; 3) Data gathering and analysis; 4) Discussion and improvement suggestions. These components will be expanded in this current section.

5.1 Stimuli Design

The aim of this first task is to create sets of stimuli to be used in the experiment based on concepts outlined in previous literature. In this research, the stimuli are email samples with some being a phishing email and some are genuine. Recall that these email samples must be designed such that it is effective with the three common psychological goals to represent the real attack as much as possible (provoke the subject to perform certain actions, block rational thinking zones, and influence the emotional and affective sphere) [32]. The first consideration when designing is the domain or industry of the email samples. Financial, e-commerce logistics, and government are selected as the target domain in this experiment to be consistent with [77]'s claim that the respective industries are most likely to be impersonated. The second consideration is what companies or organisations will represent the chosen domains and should be included in the samples. [76] suggests using large well-known organisations meaning there is a well established pre-existing level of trust and familiarity between participants and the companies. The current set of email samples will include organisations

that are well known to the local residents of Brisbane, Australia where the current research is being conducted. These include primarily Commonwealth bank (financial), National Australian Bank (financial), Australia Post (e-commerce logistics), Queensland Health (government), Australian Tax Office (government) and many more. To ensure the phishing samples are representative of the real phishing attacks, the samples are collected from the website of the selected organisations. Most well-known organisations nowadays will have a cyber-aware page outlining basic cybersecurity information and ways to protect themselves for their customers. [86], [87], [88], [89] are a few examples of the awareness page outlining techniques to protect ourselves online. These page often featured sample screenshot images of phishing emails, techniques scammers used to impersonate their organisation in the past, as well as samples of genuine email customers can expect to receive to help educate their users. Fig. 5.1 below is one example of such awareness page by Australia Post.

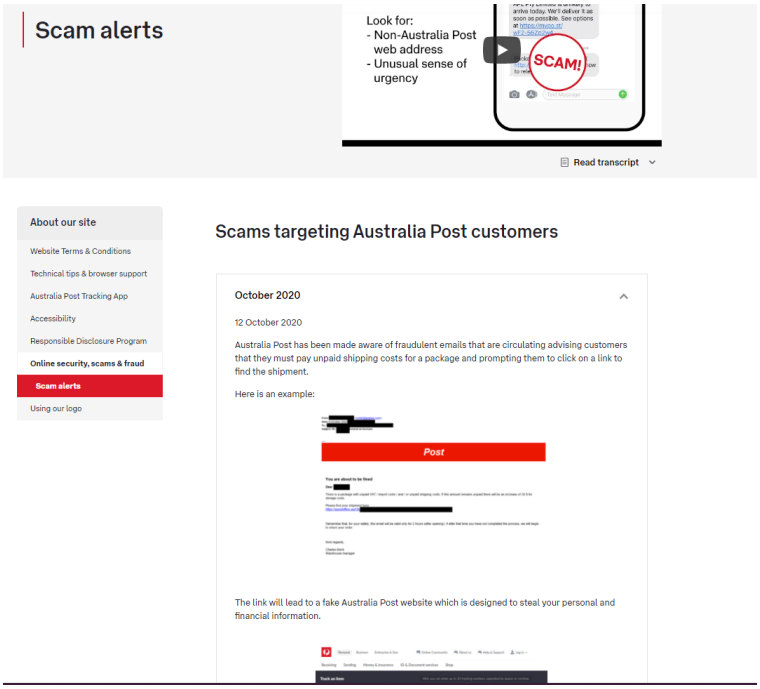


Figure 5.1: Australia Post Scam Alerts page. Source: [86]

The current study will utilise such information and samples to construct the sets of email stimuli. Full list of sources to the sample screenshots obtained can be found in this project folder [90]. After obtaining the sample set, the images gathered are then resized such that it matches with the screen size of the computer that the experiment will be conducted on. This is to ensure the images are not displayed in the low-quality out of ratio form and all text and components can be clearly seen. Duration of the current task took up four weeks of the project time to research, gather, and modify/resize the set of email stimuli to use.

Alternatively, free and paid phishing simulators could be utilised for the experiment. These simulators will automatically generate phishing emails to be used as well as setting up the experiment environment (detecting keyboard responses, mouse clicks and movements, and information entered). These simulators, however, are often restricted to the trial version with a limited number of companies available to select to generate the phishing emails in additional to the basic data analysis options.

Moreover, these simulators are not designed to be compatible with eye tracker equipment hence would be inappropriate for the current project investigating gaze data. [91] and [92] outlines considerations on how such simulators could be integrated into a phishing research project and further limitations associated with such tools, the researcher deemed this a good resource for future projects to consider.

5.2 Research and Selection of the Apparatus

After designing the stimuli set to be used, it is necessary to explore what equipment will be required to collect and analyse data from human participants. There are many human-centric tools for research depending on the nature and goal of the project. These include electroencephalogram (EEG), wearable devices for heart rate/blood pressure, pupil detection, head motion, and many more. In this project, we are interested in understanding what makes the user believe an email is genuine or a phishing, and one parameter to measure is what the user looks at when reading these email content. Data of this parameter can be collected using an eye tracker which will detect and record the movement of your eye gaze over time. The rationale behind using eye gaze movement is based on [27] principles that what you are looking at is what you are thinking about hence by inspecting elements that participants look at should reveal their thought processes while performing the task.

5.2.1 Applications of Eye-Tracking

The use of eye-tracking technology in research is not uncommon, especially in the field of psychology, gaming, marketing, and human-computer interaction designs. Eye-tracking in information security specifically is generally used to capture user behaviour when interacting with a computer in the form of the fixation point (user's eye gaze on screen), scan-path (gaze movement), heat maps (indicating area of interest), and salient picture component [93]. Typical research projects will utilise many, if not all, of the data forms which the present research will also follow. Other areas of research such as psychology may be interested in additional parameters such as pupil dilation or head motion as appropriate.

Some benefits of using an eye tracker to capture data over other methods like self-report includes the ability to explore subconscious natural behaviour. In the current study if the participants were explicitly asked to recall and describe what components of the email did they read or look at there may be bias or inaccuracy towards the reported answers [94]. The answers may not be comprehensive, skipping minor details or components but maybe the key to the decision making. These minor variations in answers are part of what causes inconsistencies and generalisability issues faced in traditional human-centric phishing studies [7], [8], [9]. The eye tracker will be able to overcome that limitation by capturing user behaviour which is a true reflection of their thought processes [94], [27].

Like many products, there are different purpose of eye-tracker devices: consumer and research. Consumer-grade eye trackers are those designed for personal use in gaming, entertainment, or as assistive technology. While research-grade trackers often come with higher price tags, those are

specifically designed for professional use in both commercial and academic research such as usability, marketing, or simulation projects. One main difference is that some consumer-grade trackers are subject to regular firmware updates to keep up with the constantly changing gaming industry, and the use of commonly available hardware which can influence consistency and repeatability of the data gathered. Research grade trackers are different from that which aims for consistency of data by utilising specifically made hardware components which are often of higher quality [94].

5.2.2 Types of Eye-Tracker

There are three main types of eye-tracker commonly used in research: 1) screen mounted, the tracker is attached to the computer screen; 2) wearable glasses, user wear a specialised glasses which is fitted with the eye-tracking component; 3) eye-tracking in VR headset, specialised VR headset fitted with eye-tracking component and head movement detection [95]. Trackers that require physical contact with the user (glasses and headset) are generally more accurate than screen mounted and enable capturing of data in real-world interactions with constant physical movement. It is however not recommended for prolonged use as it could induce eye fatigue or motion sickness to the operator [96], [97]. Screen mounted, on the other hand, is most effective for observing screen-based stimuli and it requires participants to remain stationary hence can be limited if participants are required to interact with a real 3D physical object [98]. Screen mounted generally cost the least, followed by wearable glasses and VR headset as the most expensive. Note however that having expensive equipment does not guarantee the suitability of the device to the nature of the research being conducted. In the current project we are investigating screen-based stimuli (presenting emails on screen) hence the screen mounted model is selected. Another form of eye tracker worth noting is web-cam based tracker. These are often browser-based plugins and extensions that take advantage of the built-in webcam in computers and laptops. Initial development of the current project also explored the use of web-cam based eye tracker and found the quality of data obtained is very poor (50% reported accuracy rate at best) hence it was ruled out as impractical for use in this project. Although the quality does not meet academic research standards however these web-cam based trackers is a perfect tool and widely used for mass-market research projects as users could participate without additional tool setup or travelling to the study lab.

5.2.3 Choosing the Right Eye-Tracker

Over the course of the current project, the researcher noticed there are a limited number of vendors that manufacture research-grade eye-tracking devices, and even more so for consumer-grade. The writer of this article believes this is due to the expensive price and very niche market of only companies with well-funded R&D departments or academic institutions that would be interested in purchasing one.

When purchasing a consumer-grade eye tracker, it is often easily done through the vendor's website with secure payment like a common online shopping process. Once paid, the product will be shipped to your nominated address within standard delivery time and return policies. Research grade however will require consultations with a sales representative who will aim to understand the nature of your research

project and conduct a live product demonstration before recommending product options. The sales representative will be the main contact person throughout the project or the lifespan of the product for any repairs, updates or maintenance of the tracker. This is indicative of quality over quantity business model similar to purchasing an item of value such as a car or a house which could not be done purely just browning product information online. The researcher believes this is a point worth noting for future research projects to keep in mind.

Consultation sessions with a Sydney based sales representative who proposed a list of eye tracker models that would fit within the current research design were carried out. An information flyer from the consultation including recommended models can be found here [90]. The consultation and email corresponding with the sales representative in the current project took up over 12 weeks, which resulting in the conclusion that purchasing a research-grade eye tracker for the current project was deemed infeasible due to restricted budget and the short-term nature of the honours year.

In addition to the sales consultation sessions, when choosing the right model for research a few factors will have to be taken into consideration. [99] outlined three metrics that can be used to evaluate the quality of an eye tracker model and its suitability to the research project:

Accuracy: the difference between the captured and the true gaze direction and position. Best possible accuracy as reported by many manufacturers is 0.5-degree difference between captured and true gaze position which is considered very accurate.

Precision: how consistent the captured gaze position is when the true gaze is constant and fixed at one location.

Effect of data loss: how does the tracker behave when it captures invalid data during blinking, participants wearing contact lenses or glasses.

Other factors also include familiarity of the eye-tracker operator during the experiment, the environment that is being conducted in, operational costs, demographics of participants, screen size and position of the eye-tracker installed.

5.2.4 Note on Price

There are many tiers of eye-tracker depending on the functionalities and the factors considered above. From our discussion with two eye-tracker suppliers, research-grade eye-tracker can range from $1,100 up to $51,000 with an average tracker costing approximately $5,500. This does not include software packages ($4,700), licensing ($3,400) and training ($800) which can be an add-on at additional cost. Some suppliers offer research packages which include all of the above components and ongoing support throughout the research project while some also offer a one-off sale of individual components. Alternative options for accessing research-grade tracker is opting to a rental program of this equipment. Rental rates can cost between $4700-$5400 per month with three months access to complementary analysis software. An important point to note is scientific hardware rentals are classed as temporary imports by most Customs and Border protection agencies which may charge additional import duty and fees. All of these figures are reported in AUD as of 2020 with 10% academic discount [90].

5.2.5 Tobii 4C

Cost of research-grade eye-tracker was a major limiting factor in the current project hence a consumer-grade eye-tracker was used. One leading vendor in the consumer market is Tobii and the only eye tracker model it offers is the Tobii 4C at the time of this project (now Tobii 5 as of late 2020) [100]. It is specifically designed for gaming and eSport training with a price of 169 Euro ($300). Key features of this eye tracker model are the ability to track head motion, eye gaze, and is compatible with 97% of demographics. The product comes with a free core software and SDK that expose interaction APIs and functions for developers or for recreational use [101], [102]. There are no reports on the accuracy of the current tracker model since it was not designed for scientific use, however, communities and user reviews are on a positive side with reported high robustness to data loss, and excellent accuracy and precision [101], [102]. This is certainly one limitation to note, the inability to formally assess the accuracy of the tracker to evaluate its performance.

Figure 5.2: Tobii eye tracker 4C used

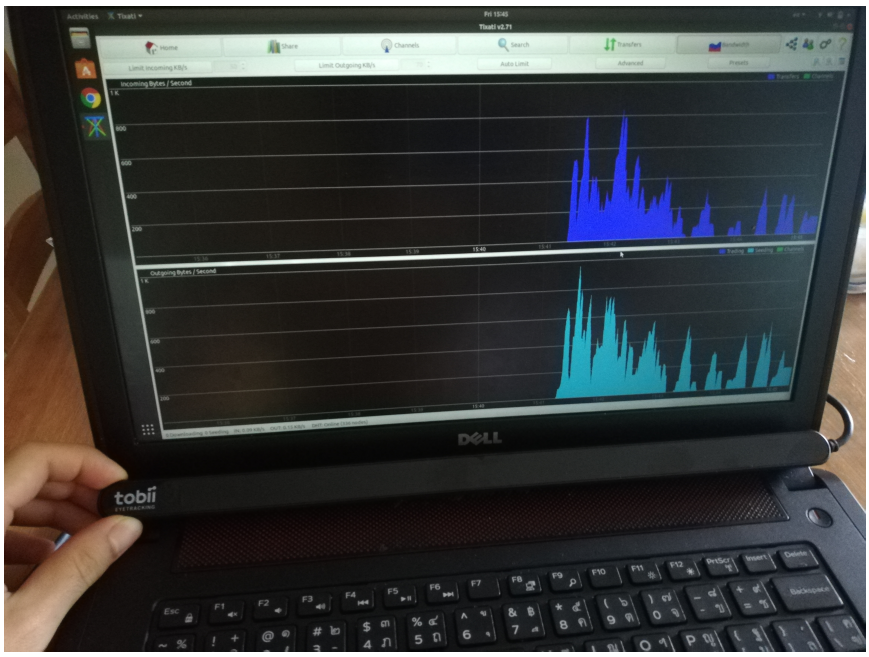


Figure 5.3: Tobii eye tracker 4C mounted onto laptop screen

5.2.6 Heat map Analysis and Simulation Program

The free software package offered by Tobii does not cover any data analysis feature rather, only allows data of head motion and eye gaze to be read and exported as x,y coordinate fixation point. The analysis is offered as part of a pro-SDK which cost an additional 2,160 Euro ($3,500) to access. Instead, the researcher of this project developed an application which takes the captured data and produces a heat map from it utilising python programming and the open-source Matplotlib library. The rationale for visualising collected data as a heat map as it can clearly outline areas of interest (hotspots) where the participant looks at most frequently and longest. This will help address the core investigation of the study to understand the thought processes of participants while encountering potential phishing attempts.

In addition to the analysis program, the experimenter also developed a second program, which is a simulation program used to display the sets of email samples in randomised order while detecting and recording user keyboard responses and time taken. The samples are presented as a basic slideshow in full-screen mode where stimuli will be displayed one after the other once a keyboard response is detected. Additional information on these applications will be mentioned below and can be found at [90].

Chapter 6

Methodology

Prior to conducting the experiment, approval will have to be sought especially when human participants are involved in the study. The following study has been assessed in accordance with the UQ Risk Assessment Guideline [103] which deemed the current project as a low-risk category. Operations of the study are also designed and consulted with relevant parties and found to be in accordance with the UQ Ethics Guideline [104].

This section outlines the phishing experiment that was conducted reflecting the step-three of the framework to gather data from participants. Seeking approval of the project and consultation on ethical policies is a progressive task which spans across the initial 15 weeks of the research although, the actual running of experiment with participants only requires four weeks.

6.1 Experiment Design

The current experiment design will be following an inductive approach where no theory or hypothesis will be proven, rather focusing on the generation of new theories and ideas to aid future research designs through discussions. Inheriting similar experimental designs from relevant studies of [33], [105], [15], [32] outlined in the previous sections, the independent variable in this experiment is whether the email samples are genuine or a phishing email. While the dependent variable is participant eye gaze on-screen as a fixation point, time that is taken before making a decision, and the key pressed on the keyboard. The controlled variable is the set of email samples that will be kept the same for all participants but will be presented in randomised order. Location of the experiment, eye tracker, laptop displays will also be kept the same to ensure consistencies in results.

6.2 Participant

In total 12 participants partake in the current research experiment with an average age of 24.5 years old. All are current undergraduate or masters students. To maximise representativeness, the participants' sample came from diverse academic backgrounds including Marketing, Psychology, Computer Science, Nutrition, and Education. Nine females and three males represent the current sample with various level of experience and knowledge of cybersecurity. Call for participants were advertised and recruited through friends of the experimenter and on a student social media page.

6.3 Apparatus and Stimuli

Informed consent and participation agreement forms are used summarising the details, aims, and rights of the participants in the current experiment. It outlines what data will be captured, how it will be used, and disposal of the data upon completion of the project including the rights to withdraw at any time during the experiment with results excluded from the analysis. Participant instruction sheet is also used which outlined step-by-step the two tasks that participants will have to complete (phishing email response task and cybersecurity survey). The Tobii 4C eye tracker is also used to capture participant eye gaze which is connected to a Dell Inspiron 7447 laptop running Ubuntu 18.04.4 with 14.1 inches display at 1920 x 1080 resolution. The experiment is conducted at the experimenter's own home in a quiet undisturbed and well-lit room to ensure maximum performance of the eye tracker as recommended by the manufacturer [106]. Tobii eye tracker software is downloaded from [90] which provides an installer for Linux based OS (Windows-based machine can download the same software from the vendor's website at [107]). This program is needed to operate the tracker and capture eye gaze. The researcher developed python experiment program is used to present 40 email samples made from step one of the framework in full-screen mode (three were later removed from the analysis) and detect participant keyboard responses. The samples consists of 20 genuine and 20 phishing emails to the user to give equal representation and exposure to both as well as covering domains including finance (15), e-commerce & logistics (8), social media (6), government (7) and others (4).

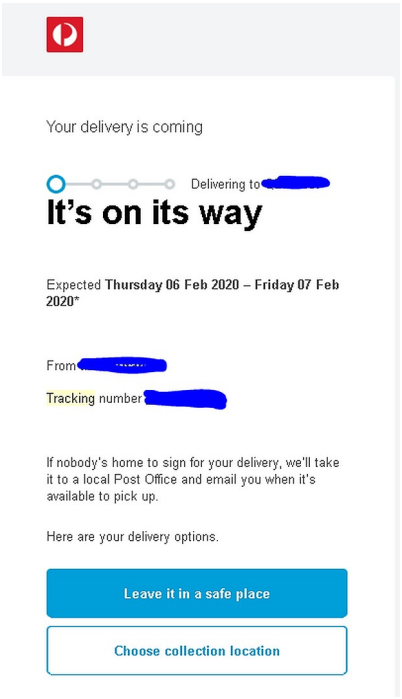


Figure 6.1: A sample of genuine email
Source:[86]

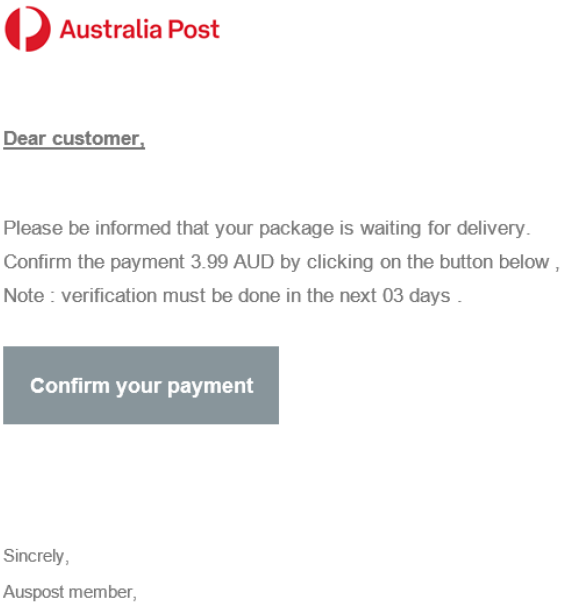


Figure 6.2: A sample of phishing email
Source:[86]

Fig. 6.1 and 6.2 above demonstrates sample of stimuli used in the experiment as produced by the first Stimuli Design Task of the framework. All apparatus and stimuli used can be found here [90] and the appendix of this report.

Finally, a survey consisting of three demographic questions (age, gender, and study majors) and six cybersecurity awareness questions were also used. The questionnaire will examine participants opinions on the following topics relevant to the concepts previously reviewed:

1) Self-rating of cybersecurity awareness (overconfidence in cybersecurity [50]),

2) Strategies used to protect themselves from phishing scams (familiarity with online services [55]),

3) Past encounters with email or online scams (prior experience being victimised and Skinner operant conditioning [49], [51]),

4) Prior experiences with cybersecurity training (prior training experience [44]),

5) Perceived usefulness of cybersecurity training (the overlooked human psychology and human properties dealing with cyber-attacks [36], [41]),

6) And suggestions or plans to improve one's cyber safety (effectiveness of anti-phishing tools compared to human performance [39], [22]),

The questions will employ quantitative Likert scale of one (I know very little) to five (I know a lot) and open ended answers allowing participants to further discuss their thought processes and self-reflect on their skills and performance.

6.4 Procedure

Upon arrival to the premise, participants will be greeted and led to the experiment room. All participants will read and complete the informed consent form and information sheet which clearly outlines their rights to withdraw from the experiment and that full anonymity is ensured. Participants will then be given an instruction sheet which outlined the two tasks that the participant will have to complete and give the opportunity for any questions or concerns to be addressed.

The experimenter then launches the Tobii EyeTracker program on the laptop which initiates a calibration process as shown in Fig. 6.3 and 6.4 below.

Series of phishing emails will then be presented to the participant in full-screen mode indicating the start of the experiment. Participants will have to classify whether the presented email is a phishing email or a legitimate email by pressing 'r' key for real and 'p' key for phishing on the keyboard. Any other key pressed will be discarded. Participants were instructed there are no time limits for each of the images presented and they should behave as if they were reading emails normally. After an 'r' or 'p' is pressed, the next image will be presented. The eye tracker will be running during the whole experiment in the background hence it is encouraged that participants sit as still as possible. The set of emails is broken down to four chunks each containing 10 emails. A three-minute break will be given in between the chunks to counter the effects of fatigue and discomfort staring at a computer screen and not moving for a prolonged period. During the break, participants may move around however it is

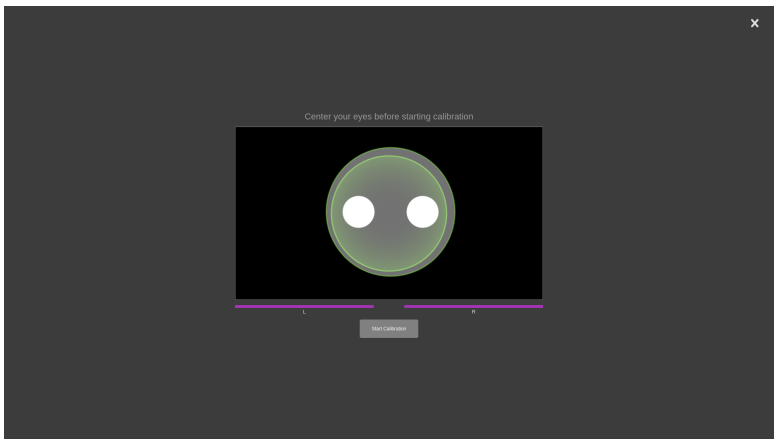


Figure 6.3: Positioning process guiding participant to sit aligned with the tracker

Figure 6.4: Calibration process displaying series of dots participant must look at

requested that they do not leave their seat and sit facing the direction of the laptop (but not necessary to always look at the laptop) to avoid having to repeat the calibration process again. Upon completing the first task, the eye tracker system will be shut down and participants will be given the survey to complete on the same computer. Participants are then thanked for their time and are free to leave the experiment venue. The whole experiment last on average 60 minutes and all participants will be reimbursed with candy/snacks of their choice for their time. All experiment sessions conducted are in line with Queensland Government's social distancing rules and that all equipment is cleaned prior to, and after conducting each session.

6.5 Data Analysis

Data that are captured during the experiment sessions are 1) user eye gaze as (x,y) coordinates fixation point on-screen in milliseconds time. Each image presented will assume the origin point (0,0) is top left corner; 2) user keyboard responses 'r' or 'p' and time taken in milliseconds for the user to respond to each of the images presented; 3) written survey results. User eye gaze and keyboard responses are saved in two different CSV file formats for further analysis. The manufacturer's Tobii EyeTracker program was designed such that it will not operate with a separate third-party program well unless low-level configurations of the program are set which may result in unexpected behaviour and/or damage to the tracker. To mitigate this, both researcher's developed program and Tobii program will be running consecutively but independently of each other hence the result of two separate files made. The Tobii program will handle capturing eye gaze while the researcher's developed program will capture user responses.

Since data is saved in two separate files and that the two programs are running independently of each other, there is no way to signal the Tobii program that a new email stimulus has been presented. The eye tracker data captured hence are continuous without any dividers to indicate which rows of the data belongs to which image stimuli. The first step of data analysis is to combine the data based on the timestamp in order to map the eye tracker data with the user responses. For instance the first row of Fig. 6.6 below the timetaken column is 21.988196 seconds for the imageFileID 1. Converting this to milliseconds will give 21988.196. We find the row in Fig. 6.5 under the Time column that has the same milliseconds value, let's suppose we found the 1000th row to have the same value. The row 0 up to 1000 of the eye gaze data file will be marked as for imageFileID of 1. Then the whole process repeats for the next row imageFileID of 2 finding the row in gaze data from row 1001 that match with the timeTaken values in milliseconds.

Reading	Time (miliseconds)	X Coord	Y Coord
0	392996	0.295854	0.632006
1	398495	0.295854	0.632006
2	408150	0.295854	0.632006
3	509040	0.286782	0.682679
4	520082	0.320808	0.717654
5	531801	0.331772	0.719584
6	543307	0.339986	0.713302
7	553783	0.346771	0.70109
8	565351	0.35116	0.683393
9	576564	0.356081	0.668189
10	587184	0.36045	0.652242
11	598692	0.364293	0.638954

Figure 6.5: A sample of eye gaze reading data

Then using the researcher's developed analysis program, plot the eye tracker data onto the sample email screenshot images which will display the eye movement paths. Since there are over 37 email stimuli and 12 participants that means there are 37x12 = 444 screenshot images to be analysed. Instead, all participants' eye gaze is combined and plotted onto each of the 37 email stimuli. Fig. 6.7 below

imageFileID	imageFileName	domain	userResponse	timeTaken (Sec)	correctResponse?
1	Scam4	finance	p	21.988196	TRUE
2	Scam17	logistics	p	31.112552	TRUE
3	Scam9	logistics	p	21.464975	TRUE
4	Real3	others	r	13.780226	TRUE
5	Real13	finance	r	9.934671	TRUE
6	Scam20/fig6	finance	r	25.681062	FALSE
7	Scam15/fig7	finance	r	8.396392	FALSE
8	Scam6	finance	p	28.681705	TRUE
9	Scam18	social media	p	27.790168	TRUE
10	Real1	social media	r	14.773639	TRUE
11	Real11	government	r	18.086204	TRUE
12	Real7	finance	r	23.135371	TRUE
13	Scam13/fig13	finance	r	23.002121	FALSE

Figure 6.6: A sample of user keyboard responses

displays an example raw plotting data of all participants' eye gaze of this particular email stimuli. Then with this raw plotting, filters out and only highlights the area of interest. The area of interest will follows the same definition as [93] where this is defined as the clusters that contain significantly more gaze (x,y) fixation points than other areas of the image. Fig. 6.8 blow displays an example of the filtered data with area of interest included. The rationale behind combining all participants data together onto one single screenshot for each sample is to obtain an average eye movement pattern for ease of data analysis. More information on the developed analysis program including documentation of the code can be found here [90].

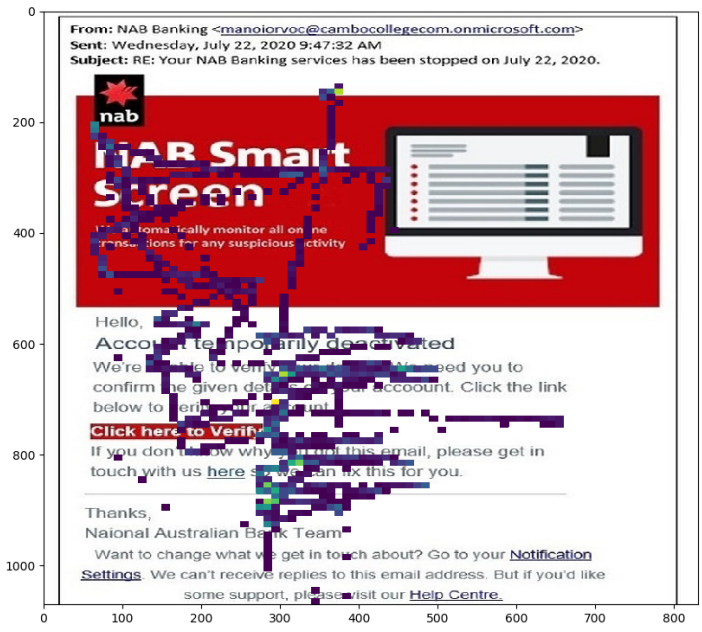


Figure 6.7: Raw plotting of eye gaze data

The user response data are manually processed using Excel spreadsheet to extract information on the rate of correct/incorrect responses as will be outlined below. This entire process is then repeated

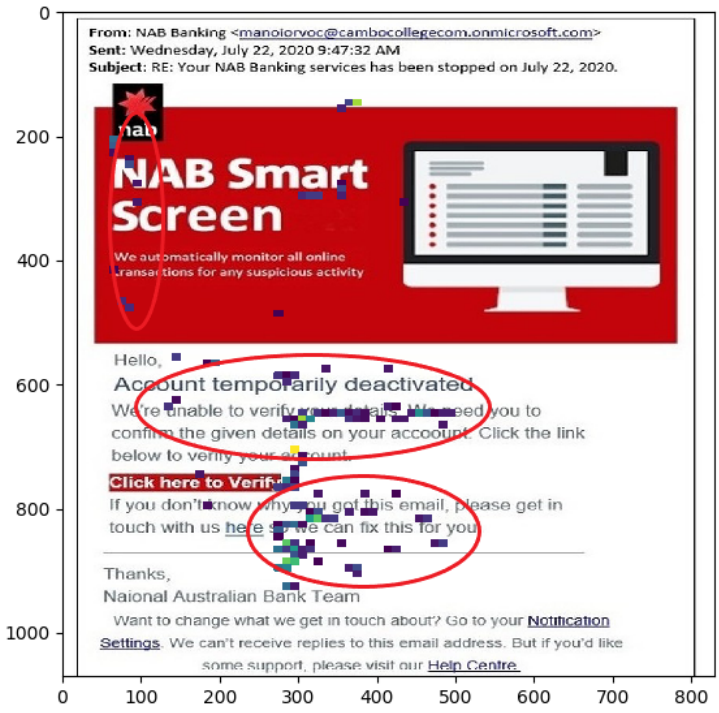


Figure 6.8: Filtered plotting of eye gaze data

for each of the participants' data before deriving the average performance on correct and incorrect responses of this participant group.

Chapter 7

Results

The final task of the framework is to report the results of the study by applying models and discuss the implications of such results according to previous literature. Since the collected data concerns behaviour of human participants, the Signal Detection Theory [38] will be applied to classify participant results.

7.1 Eye-tracking and User Responses

The emails included in the analysis consist of 17 phishing and 20 genuine samples made up of finance (14), e-commerce logistics (8), social media (5), government (7) and others (3). Results of three emails were discarded due to the eye tracker stopping prematurely as a result of a bug in the researcher's developed software. The average time spent reading each email is 25.87 seconds and the overall correct response rate is at 70.27% or 26 emails.

	User responded phishing	User responded genuine
Phishing present	64.71%	35.29%
Phishing absent	25%	75%

Figure 7.1: Responses classified by Signal Detection Theory

Fig. 7.1 above displayed the results using Signal Detection Theory. There are six misses (phishing email but user responded as genuine), five false rejection (genuine email but user responded phishing), 11 hits (phishing email and user responded phishing) and 15 correct rejection (genuine email and user responded as genuine) corresponding to the percentage values above.

Fig. 7.2 below outlines the performance according to each email domains.

	Number of stimuli	Correct response	%	Incorrect response	%
Finance	14	9	64.2857	5	35.7143
E-commerce/Logistics	8	6	75	2	25
Social Media	5	5	100	0	0
Government	7	4	57.1429	3	42.8571
Others (Uber and Translink)	3	2	66.6667	1	33.3333

Figure 7.2: Responses classified by domains

Due to over 37 stimuli sets included in the current experiment it is impractical to insert all screenshot results of the eye tracker data in the current report. The following series of pages will only outline the key findings of the data gathered as classified by Signal Detection Theory. Each figure includes the original screenshot that participants see during the experiment and the result of the average eye gaze pattern of all participants combined. Area of interests is highlighted in red which is the area where participants spend the most significant amount of time looking at. More complete results of the email stimuli can be found in the Appendix and all data gathered can be found at [90].

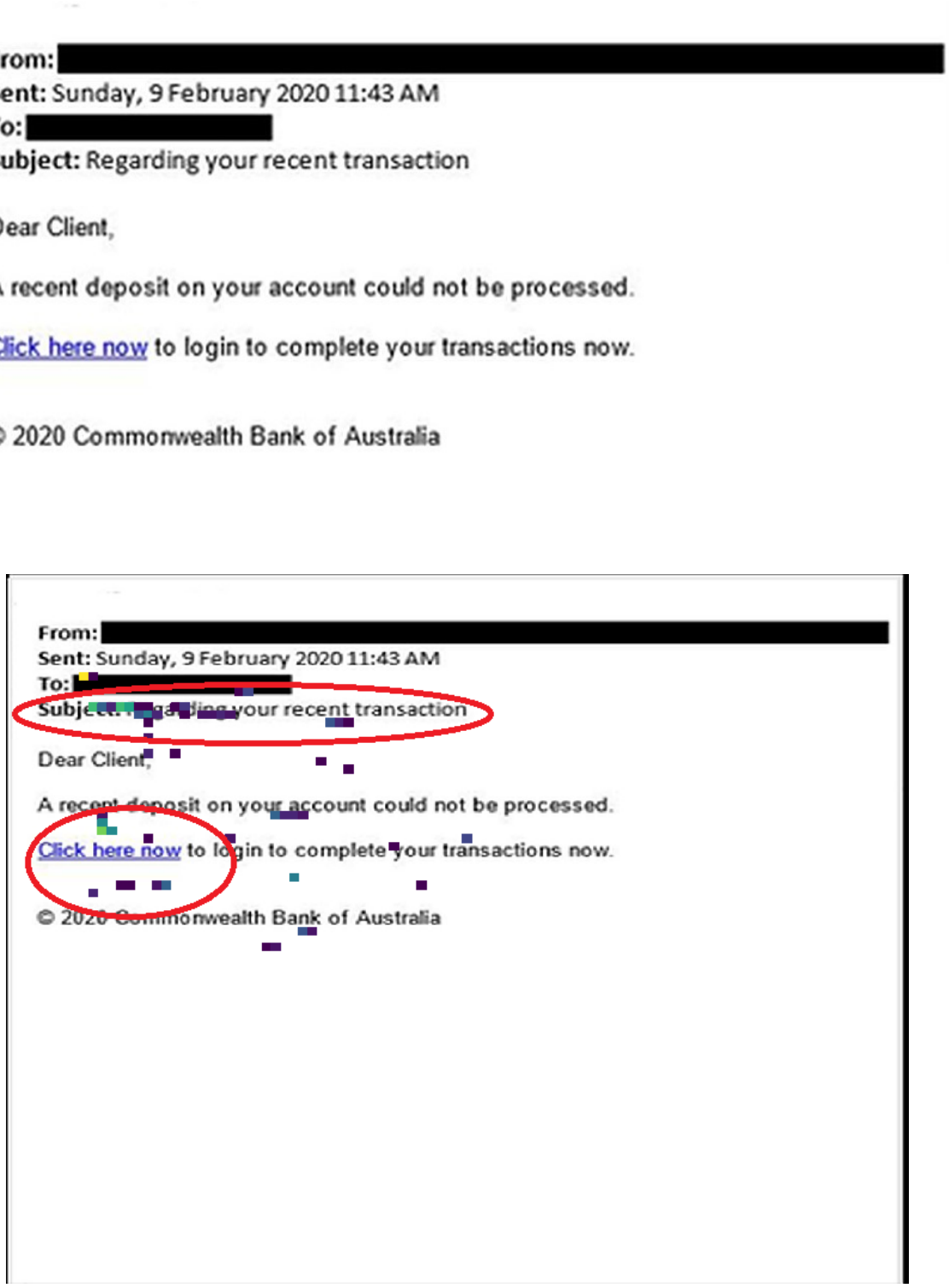
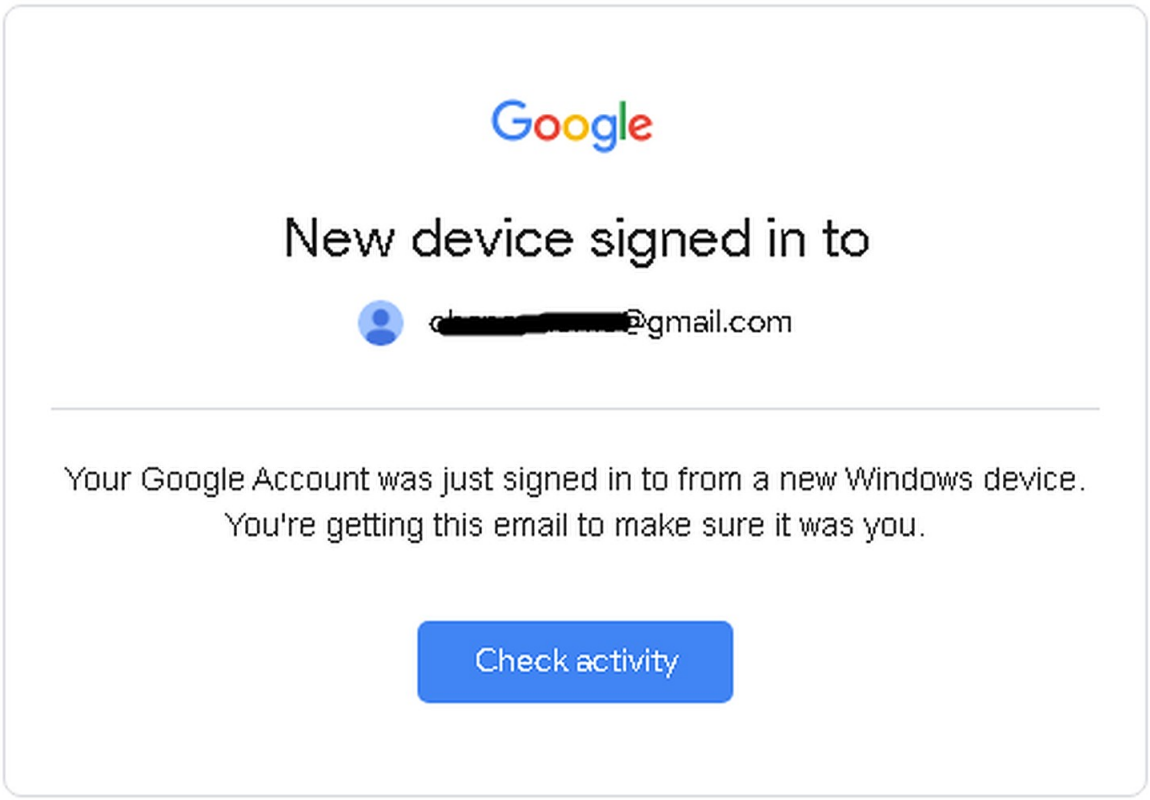


Figure 7.3: A sample of phishing email user responded as phishing (hit)



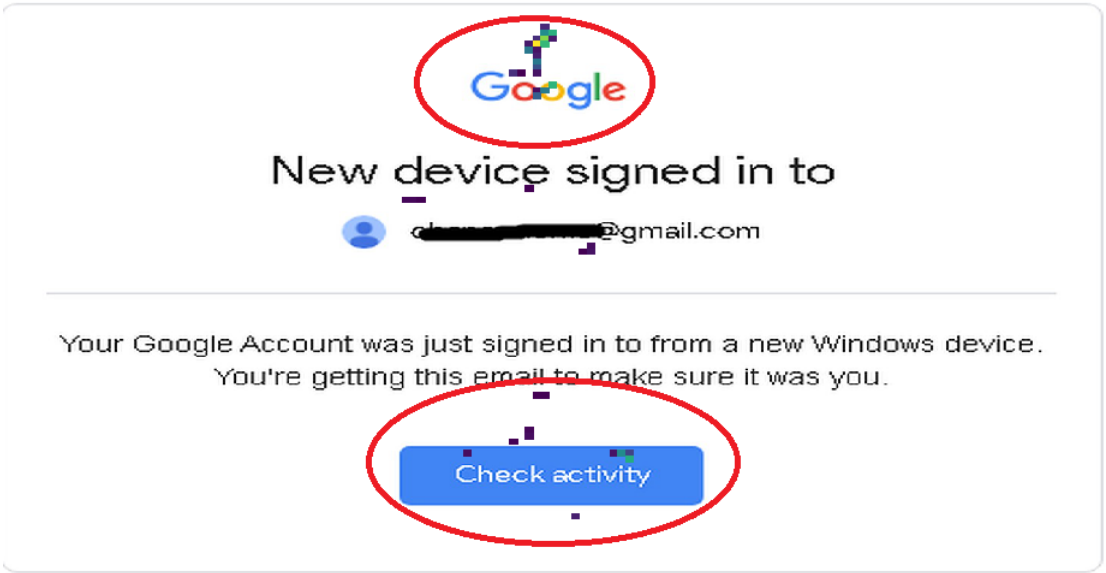


Figure 7.4: A sample of genuine email user responded as genuine (correct-rejection)

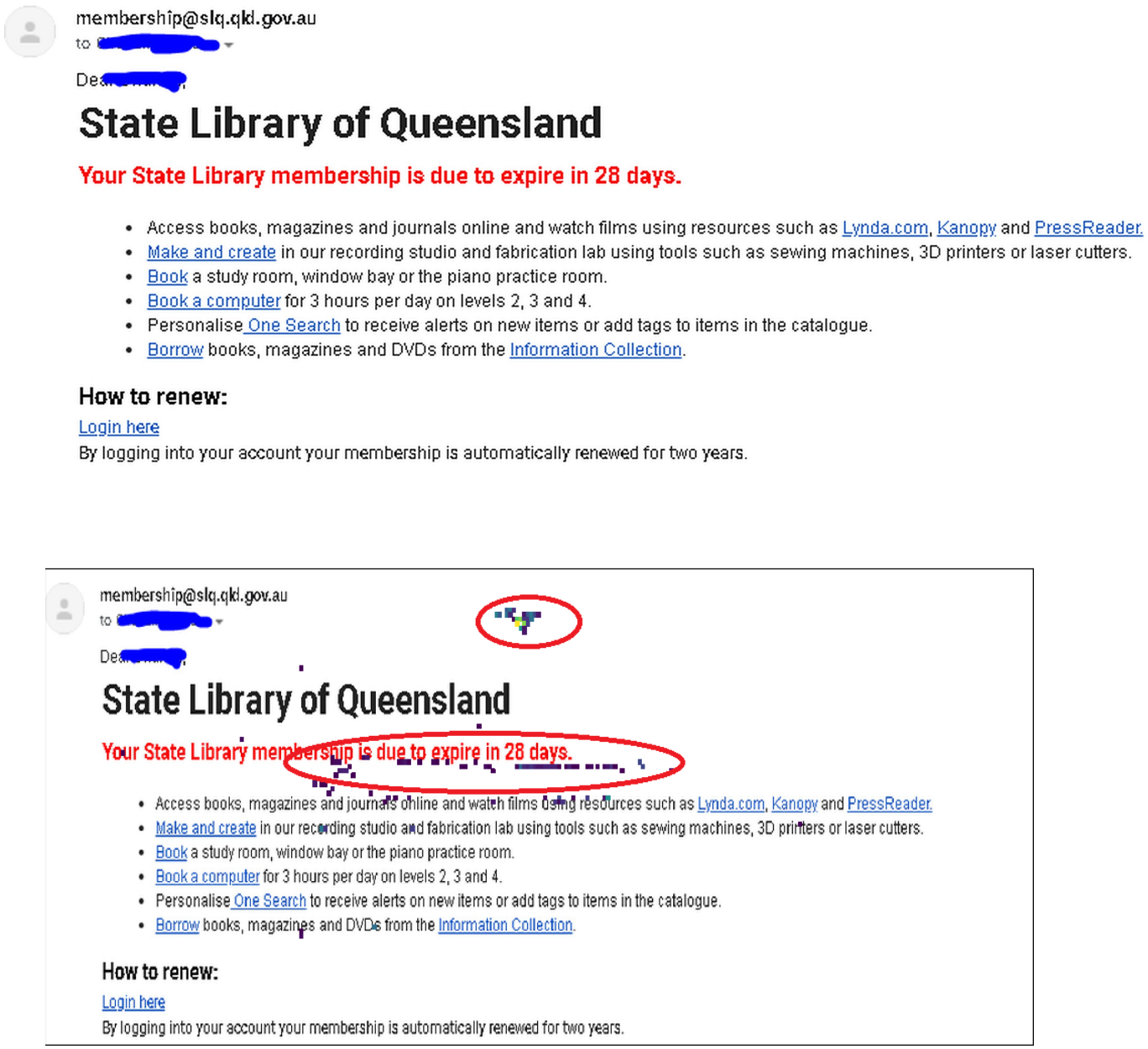


Figure 7.5: A sample of genuine email user responded as genuine (correct-rejection)

In emails that user responded correctly (both hit and correct rejection), there are interests toward logo of the organisation and graphics used in the email as demonstrated in Fig. 7.4 and Appendix A.4. Where a time-frame is mentioned in the email, participants focus their attention towards those key information as shown in Fig. 7.5. Appendix A.5 and A.6 that users responded correctly also noted participants are utilising unique information such as tracking number, ID number to indicate genuineness. Interestingly the area of interest of those email responded correctly are often clustered around the beginning and the ending portion of the email. Very few samples contain an area of interest that is clustered around the body of the email. Fig. 7.3 supports this argument where the area of interest is shown in the header section of the email.

In emails that the participant responded incorrectly (misses and false alarm), the area of interest in the emails includes the email subject line and email address of the sender as noted in Fig. 7.6 and 7.7 consistently to those correct responses above. Banking information and instructions outlined in the email are also viewed at most frequently as depicted in the two samples mentioned. These include the scammer's bank account number the money should be transferred to and links to follow. In many samples similarly to those samples responded correctly, participant spends a considerable amount of time glancing at the closing paragraph of the email. Fig. 7.7 and Appendix A.7 depicted this behaviour where these areas commonly include further instructions or constraints for example 'do not reply to this email', 'find more information here'. Conversely to correct responses samples, Appendix A.2 indicate participant did look at the security indicator (SSL certificate and lock symbol) however they still responded incorrectly to the sample.

The common area of interest that participants look at most frequently in both correct and incorrect response are hyperlinks (embedded under text such as 'click me' and plain links) and clickable buttons as depicted in Fig 7.8 and 7.3. The use of bold and bright red colour to highlight certain information or phrases attracts user attention indicating it as an area of interest in many samples Appendix A.1, A.2 and Fig. 7.5.

From: CommBiz Notifications [mailto:alerts@commdebit.com]
Sent: Thursday, 31 October 2019 9:10 AM
To: [redacted]
Subject: CommBiz: Direct Debit initiated by Australian Taxation Office.

Dear customer,

Our records indicate that Direct Debit requested from your account by this creditor:

Creditor name: Australian Taxation Office
Creditor bank: Reserve Bank of Australia
BSB: 093-003
Account number: xxx385

Please confirm or amend Direct Debit requests in Direct Debit management system.

Thank you for using CommBiz.

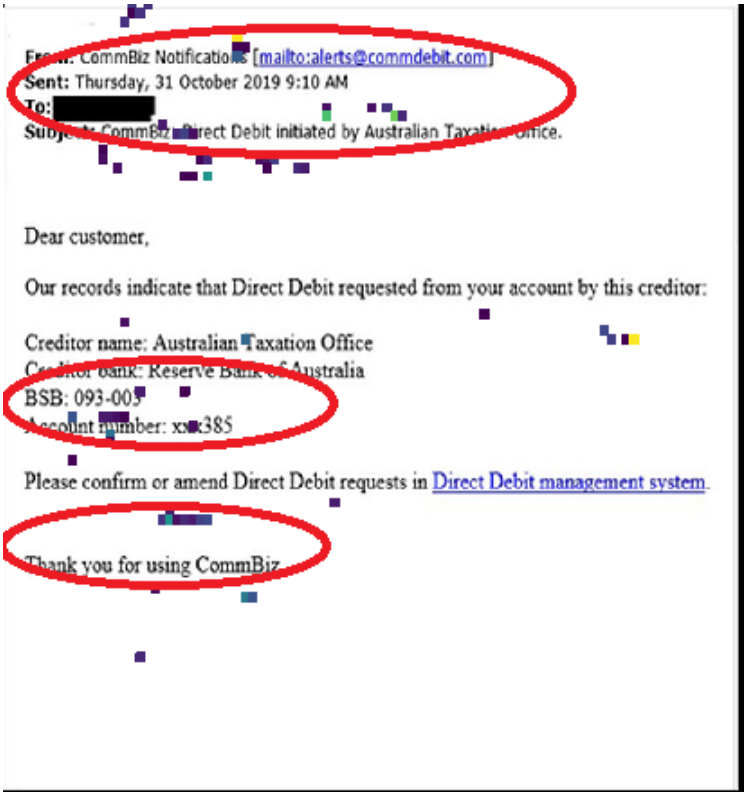


Figure 7.6: A sample of phishing email user responded as genuine (miss)

Dear [redacted],

Your Term Deposit for account ending in [redacted] has matured.

Next Steps
To check the details of your Term Deposit:

1. Log on to NetBank.com.au
2. Go to View Accounts then Information.

Or for more detail see your reinvestment notice by going to View Accounts and select Statements.

Yours sincerely,
CommBank

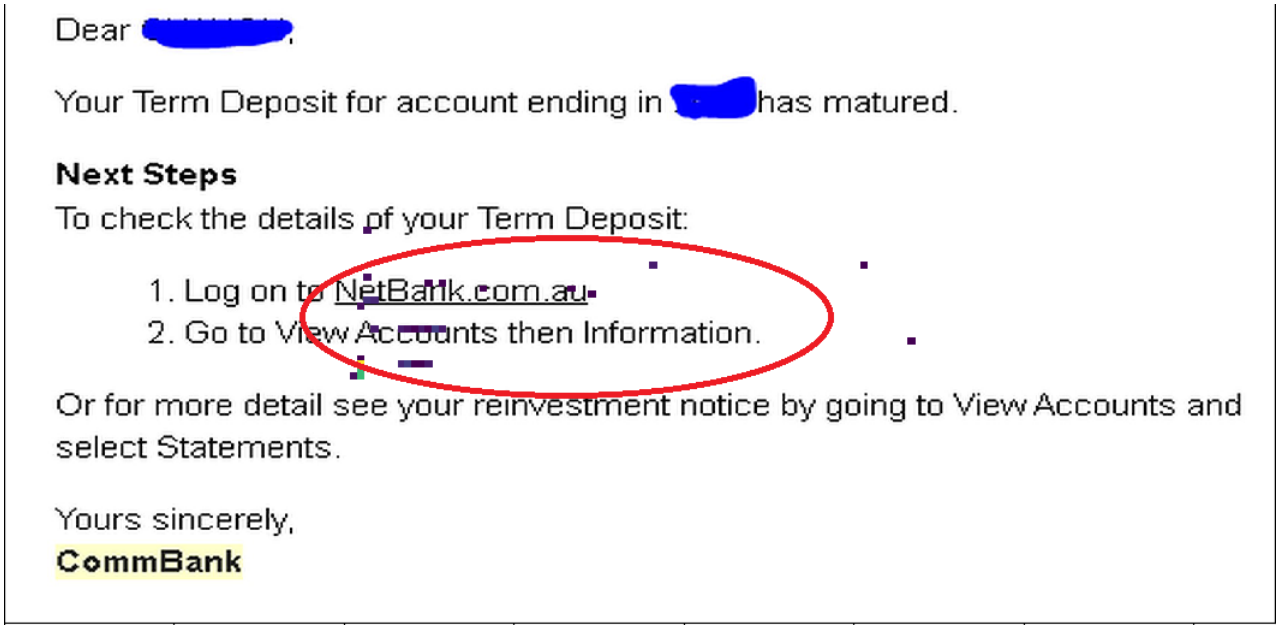


Figure 7.7: A sample of genuine email user responded phishing (false alarm)

7.2 Survey Results

Despite having high correct responses rate, the majority of participants self-rated their cybersecurity awareness as low-average with a mean score of 2.9 out of 5 on the Likert scale.

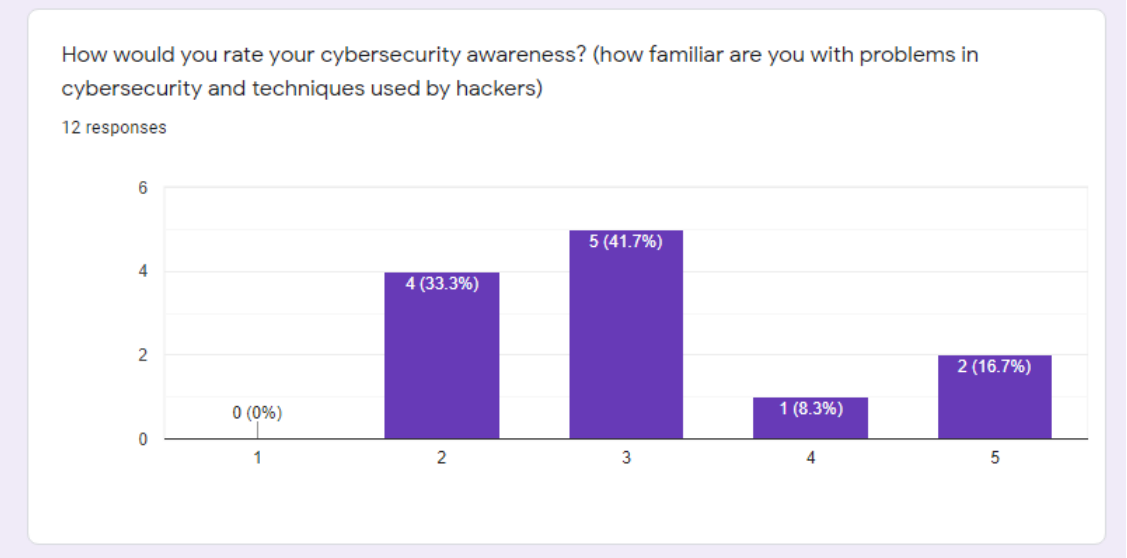


Figure 7.8: Self rating of cybersecurity awareness

When asked about the effectiveness and importance of cybersecurity awareness campaigns and training all participants unanimously agree that such strategies can increase the level of protection for users and is a vital key to online safety. However mere two participants have had experience in cybersecurity training and induction themselves as part of their study requirement.

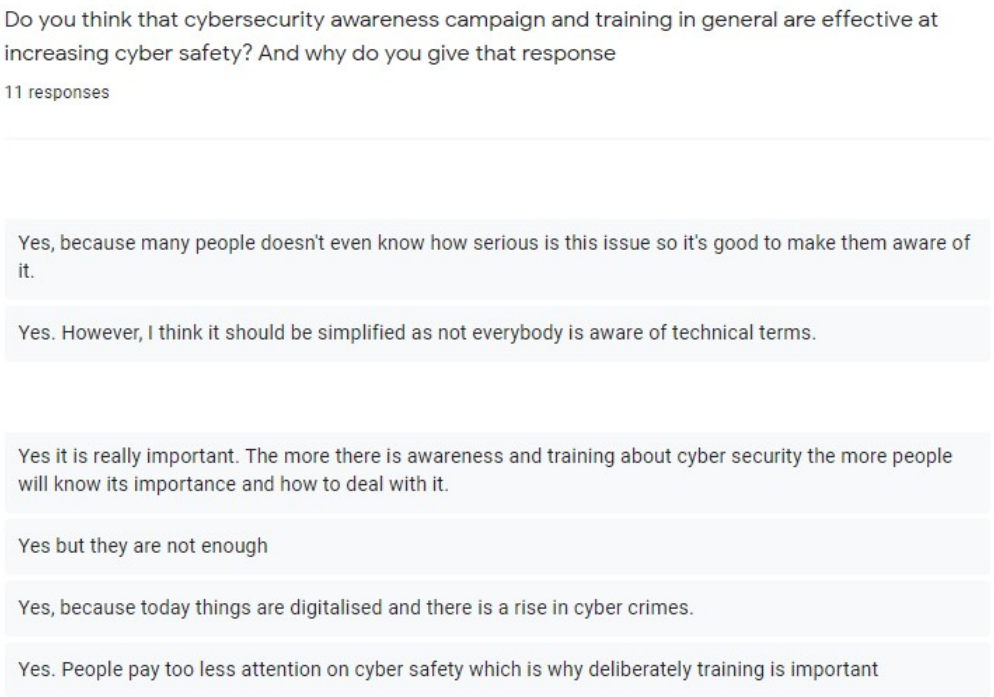


Figure 7.9: Opinion towards cybersecurity training

Fig. 7.9 outlines some key responses and reasoning participants supplied in regards to the effectiveness of awareness campaign and training.

When asked to comment on their personal plan to improve their cybersecurity skills one participant expressed interest in participating in cybersecurity training in the immediate future to improve their awareness. Two participants mentioned that although cybersecurity training is effective and important, relying on training alone is insufficient and will be looking for alternative methods to improve their skills instead. Other participants showed no opinions or immediate plan to improve their cybersecurity awareness despite commenting on the importance of training in the previous question. The figure above also demonstrates these opinions.

Eight out of twelve participants claimed they have had direct experience with phishing scams before. All the scams were related to finance and were delivered to them via email asking for personal information such as credit card number, bank account number, and password. This also includes email containing links to a phishing website also asking for personal information. Basic comparison on correct user response between those with phishing scam experience and those without reveal no significant difference in performance between the two groups. The average correct user response for those with phishing scam experience is 71% while those without are at 69%. Fig. 7.10 below further demonstrates key responses and experiences of participants with scamming.

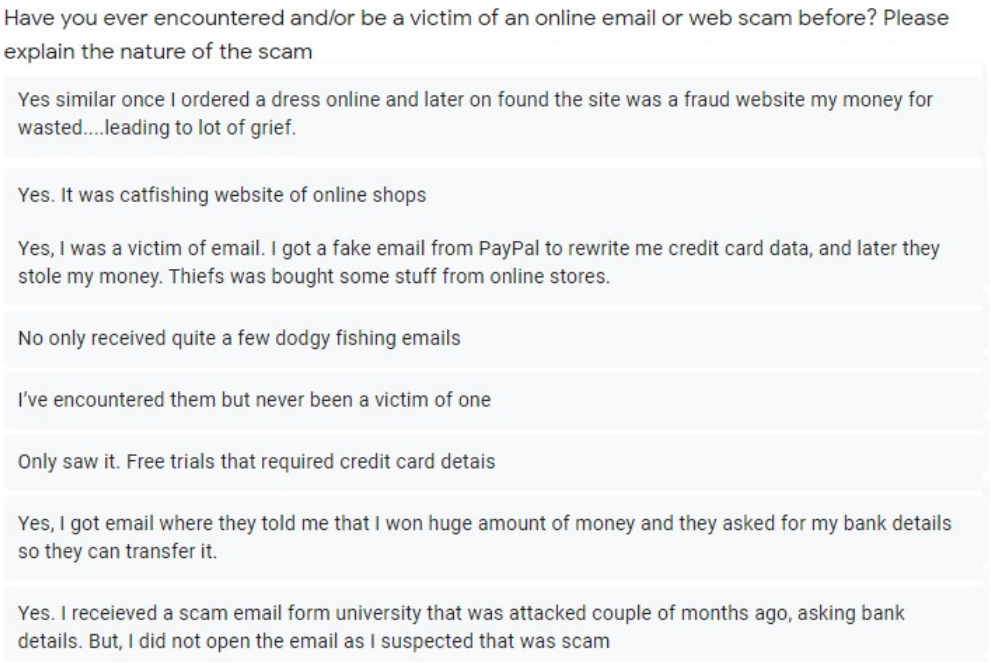


Figure 7.10: Previous experience with scamming

Strategies that all participants reported performing to protect themselves online on a daily basis include not clicking on unknown links, checking the sender address, not accepting cookies, frequently changing their password and not saving personal information online, deleting communications from an unknown sender, and checking reviews of sites before purchasing or conducting online transactions. Two participants reported using automated protection tools such as AdBlock and DuckDuckGo browser with built-in protection features. One participant commented judging the tone and language of the

email received whether it is over-exaggerating or not which would indicate a potential scam. Fig. 7.11 below outline key responses on strategies participants used to protect themselves online.

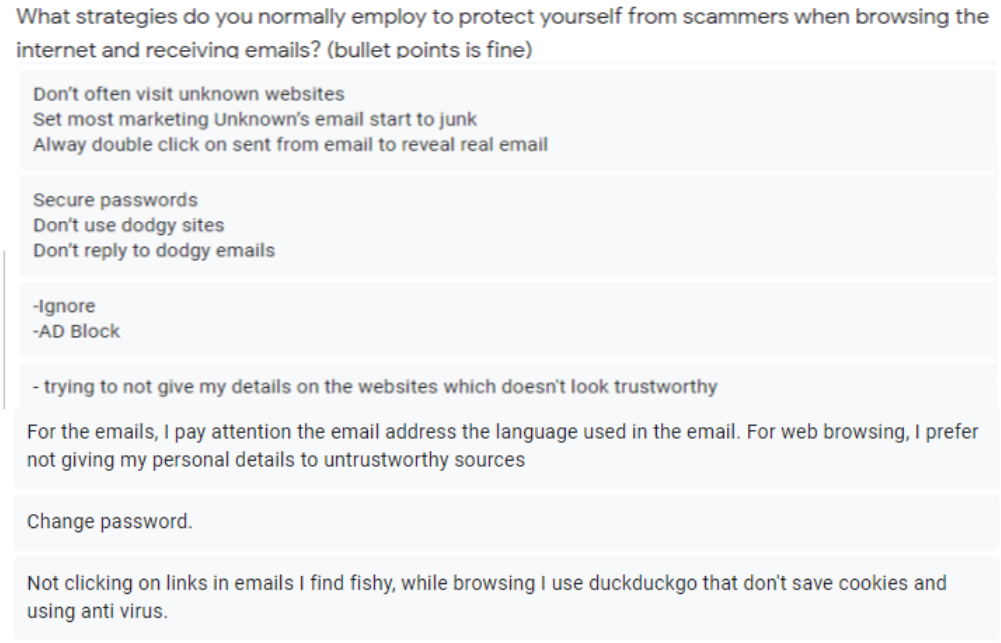


Figure 7.11: Strategy to protect themselves

When asked about strategies to promote cyber safety among the general population all 12 participants agreed that training with the use of simulation and both offline in-person and online courses will help increase the awareness. Four out of 12 participants discussed the idea for such training to be mandatory within education institutions. Two participants suggested using social media and the use of online influencers to target specific user groups delivering messages about online safety. Other proposals involve the training to be free, basic and easy to understand; localise the training so people within the same communities can share ideas; set up an information booth in high foot traffic areas like a shopping mall to promote the awareness to the general population more. Fig. 7.12 below outline key responses participants discussed.

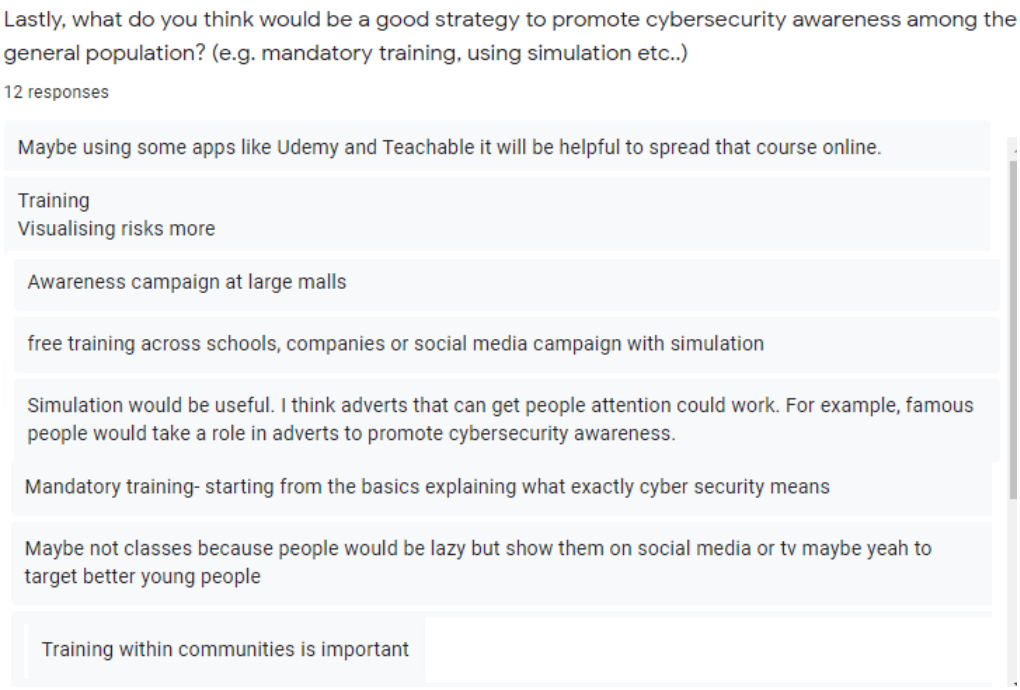


Figure 7.12: Recommendations to improve awareness

Chapter 8

Discussion

One key driver that the current study operates on is [27] theory that what you are looking at is what you are thinking about. From this, it could be inferred that components in the phishing emails that participants look at but responded incorrectly (misses) are effective at deceiving, while components in emails responded correctly (hit) are effective at assisting users to make an informed decision. Similarly, components in genuine emails that participants look at but responded incorrectly (false alarm) are ineffective at informing users of trust, while components in emails responded correctly (correct rejection) is effective at informing users of trust. Results will be interpreted following those classifications.

Firstly, the low-average self-reported cybersecurity awareness score from the survey clearly indicates participants are realising the danger of cyberattacks however they believe their knowledge of how to protect themselves is lacking hence the low score. This is proven as an underestimation of their own skills and understanding since the correct response rate in eye-tracking task was up to over 70.27%, similarly to the average figure [39] reported on generic phishing email detection performance of novice online users. This supports [50] findings that overconfidence in cybersecurity skills correlate with poorer performance as users take fewer precautions trusting in their own skills more. Underestimation of skills on the other hand promote participants to engage in more precautionary action when interacting online. Since they perceive their skills to be lower and more vulnerable, hence increasing the need to protect themselves and be more vigilant [50]. By protecting themselves, participants could be lowering the criterion of the Signal Detection Theory and reporting more emails they encountered as phishing [38]. This phenomenon was not observed in the current participant group where the false positive rate is relatively low, confirming [45] findings that originally uncover limitations of the Signal Detection Theory in modelling complex human learning and cognitive processing abilities beyond simple movement of the criterion bar.

It is clear also from the survey results that all participants realised the importance and the increasing needs to develop personal cybersecurity skills however when it comes to action on these comments, very few have actually taken steps such as receiving training. Shockingly only a handful of participants have immediate plans to improve their cyber skills despite realising its importance and rating their

cybersecurity understanding as low-average. This lack of previous training directly correlates to the fact that the majority of the participants also have fallen for phishing scams before which is consistent to [45] and [46] findings on vulnerability as a result of lack of training. Participants' responses contradict itself where if one perceives cyber skills to be the key in staying safe online, and support that training is effective at exercising such skills, why do so limited number of participants have prior training experience in the first place? Two participants raised this concern that relying on training alone is insufficient. This could perhaps indicate the ineffective design of such training courses offered in the market, such as lack of retraining program after four weeks as [48] claims to be the most optimum frequency, out of date materials being taught, difficult to understand/not tailored for beginners, inaccessibility (too pricey, not suitable for certain population groups, in foreign languages), or even poor marketing of such courses. These factors on training courses were not investigated in the current project however is an important question to be answered by future research.

Those with a prior encounter with phishing does not perform better in decision making (correct response rate) compared to those without for current participant groups which contradict with [49] findings that prior experience helps increase awareness and detection skills. As noted by the experimenter of that study, the small sample size in both current and [49] study could influence these contradicting findings. Although, the findings in present study could be explained with the Pavlovian Operant Conditioning theory. The repeated behaviour falling for these phishing scams in the current study despite prior phishing encounters indicates the intensity of prior punishment (losing money or personal information) are low [51], [53]. When the intensity of punishment is low, learning from experience would be unlikely to happen. Furthermore, as [22] claims, personal information is not a physical object that the harm can be realised right away hence the perceived loss and intensity of prior punishment is lowered even more. All scam attempts participants reported encountering are of the finance domain supporting [77] and [11] findings that this is one of the most commonly phished email types with the ultimate goal of obtaining personal information for financial gain. Despite the indifference in performance due to experience, it is surprising to see how common it is for the general population to encounter some form of cyberattacks with over eight out of 12 participants reported prior experience being scammed. Over the course of this project, the researcher has also received on average one email phishing attempts per week mostly in foreign languages such as Spanish, Russian or French. These emails are asking to renew a subscription to online services or clear a bank transfer which the researcher has no knowledge of these languages and has never associate with any of the claimed services and financial institutions. This experience is consistent to [3] and [20] findings that non-English speaking online users will be the next target for scammers in the near future due to the majority of research and development focus at present are done in English language context. Appendix B.1 and B.2 provides an example of a real phishing email that the researcher of the current project recently received.

Reported strategies that participants perform to protect themselves online on a daily basis indicate some understanding of methods used by hackers and the countermeasures against them. Since the majority of the participants have had no training before, it could be inferred that these countermeasures

and understanding participants reported were derived from their own direct experience with phishing or public awareness campaign they came across. These are consistent with research recommending that the claimed countermeasures and the exercise of common sense can indeed combat and help online users detect potential phishing scams [72], [71], [66]. How much of these reported strategies are practised in real life is unknown relying on self-report method alone which is the main rationale for the use of an eye-tracker to investigate natural behaviours revealing the true understanding. Evidence from eye-tracker data confirms that participants do exercise these techniques naturally as previously reported in Fig. 6.6 and 7.3 checking hyperlinks; 7.4 corporate logo and clickable buttons; and time frames in Fig 7.5. In addition, mere two participants mentioned utilising automated protection tools to protect themselves online. This is in line with arguments presented in the earlier sections of this report that these tools are populated with usability issues hence participants resort to self-reliance and own skills to detect these potential attacks. Logically it could be assumed that if participants make use of these automated tools frequently and/or see these as being useful, then more participants in this study should have recommended or mentioned the use of these tools but this was not the case.

Many great ideas were discussed when participants were asked on strategies to promote cybersafety. One compelling idea is mandatory training in school as part of the curriculum. Such a strategy could definitely promote more awareness not only among the general population over the long-term but also younger generations that are equally vulnerable but are often overlooked with the reported increase in cyberstalking and harassment among younger generations [37]. The use of social media influencers is another interesting idea to target a specific population group. A point worth noting with these suggestions is that such messages or training courses implemented should be done with consideration of differences in learning abilities attributed to age. Older children or teens may be able to remember and comprehend more complex topics, while younger ones may need more visual-based prompters as found in the study of cybersecurity training among primary schoolers by [48]. Making people realising the harm that could be done through online attacks “visualising the risk more” as reported by one participant could definitely make people become more vigilant. This is in line with participants of [28] where it was suggested that security indicators need to be more silent in guiding the user to act appropriately and informed user to the extent of the risks posed to them. This argument is also supportive of the earlier discussion on the issues with users ignoring the warnings and not utilising indicators to their advantage.

Furthermore, Results contradict with [81]’s findings that the average user spent on average four-eight seconds reading emails while findings in the current experiment indicate staggering 25.87 seconds. This could occur as a result of participants spend time reading the whole email since it is in a research context compared to when casually reading emails in their own time which may mean just skimming through. All participants were informed of the purpose and aim of the research prior to commencing the experiment, this could have contributed to users spending more time reading than usual. An uninformed blinded study could replicate more natural behaviour of the participants in future studies. Very interestingly social media email samples have 100% correct response rate. This is definitely attributed to social media having the smallest sample size of all email domain stimuli (only

six samples). However this is in line with [55] since current participants group are young adults (mean age 24.5 years old) this population group are very familiar with social media, it would be easier for them to detect when the email presented deviates largely from those that they're familiar with causing an alarm. Moreover, results indicate participants is clearly influenced by the perceived authority in government emails which performed the poorest out of all sample domains. The perceived authority puts pressure on participants to comply with the demands of the email in hope to avoid conflicts with law enforcement [70].

Moving to eye-tracking data, consistently with [82] that subject line is the most important part of the email grabbing user attention. Indicated area of interest where participants spent a considerable amount of time glancing at includes the subject line and the email address of the sender. Such action complements the previous argument on participant knowledge of techniques used by scammers and reported strategies used to protect themselves online, in line with experts recommendations on self-protection [72], [71], [66].

In email samples with security indicators (SSL certificate) as depicted in Appendix A.1 and A.2, the indicators were marked as an area of interest conversely with participants in [15] and [27] studies that found no area of interest on security indicators. As noted in those two studies, since the indicators do not require users to stare at it for a considerable amount of time in order to understand what it meant, it was not flagged as an area of interest in those studies however it does not imply users did not take notice. Despite these conflicting findings the figures mentioned above clearly indicate participants did take notice of the indicators as shown with fixation points populated near the symbols. This action further strengthens the findings that participants are aware of basic strategies to protect themselves online by verifying the contents they are viewing. This finding, however, contradicts the arguments initially discussed in the current paper that users failed to utilise security indicators [15], [27]. With mere two email samples containing the security indicators in the current study, the result regarding utilisation of security indicators is inconclusive.

Focusing the attention on samples that participants responded incorrectly, banking information and instructions outlined in the email is the key area of interest. This is clear evidence of the three common psychological goals in effect (provoke the subject to perform certain actions, block rational thinking zones, and influence the emotional and affective sphere) [32]. The threats outlined in the email such as cease of membership (Fig. 7.5), suspension of account (Appendix A.2), or refusal of payment (Fig. 7.7) provoked an emotion of fear blocking the rational thinking zones (goal 3). The loss of finance, in this case, is an undesirable outcome in addition to the limited time to action outlined in the email further fueling the fear and urgency (goal 2). As a result, users look for strategies to minimise this loss which they perceive the best option is to comply with the demands of the scammer (goal 1) as [68] outlined in the concept of maximising gain and minimising loss. Participants' behaviour of focusing their attention on the banking information and instructions is the evidence of compliance with those demands to mitigate the loss perceived. In addition, an interesting behavioural pattern of participants among samples that were responded incorrectly is the focus towards the closure paragraph of the email, Fig. 7.7 and Appendix A.1 demonstrates this phenomenon. As noted, the closure paragraph often

includes links or further instructions/constraints to the situation such as ‘do not reply to this email’. These additional cues support the argument above that it further contributes to goal 2 and goal 3 of the three psychological goals further hindering the support and options that users perceived to be on offer and force them to comply with the demands of the scammers believing this is the best option [32].

Taking on board results from the current experiment, one sample suggestion on what these anti-phishing tools that is more ’human-friendly’ could look like considering the human factors discussed is depicted in Fig. 8.1 below. Firstly, upon receiving a potential phishing email, the overall rating will be displayed informing the user how likely this email is a phishing attempt and the risk involved. Secondly, sets of instructions for the user to follow will be immediately displayed to guide users step-by-step in verifying the email received. This includes checking the sender’s address, hovering on links to detect link manipulation techniques, and tips on contacting host organisations to verify this email. The techniques employed follow those recommended by experts [78], [79], [80]. In addition to the instructions, key information is also highlighted which is the date or time frame mentioned in the email to help with decision making. All anti-phishing tool generated texts and highlights will be done in bold and bright colour to attract user attention towards those recommendations.

Figure 8.1: Ani-phishing tool recommendation

Chapter 9

Summary of The Framework

The proposed framework consists of 4 related tasks derived from the phishing study conducted as outlined above. It can be summarise as follows:

1) Stimuli Design: the creation of sets of stimuli which to be used in the experiment

2) Research and selection of tools: for human-centric research, there is the need to utilise appropriate apparatus to collect data from human participants which often requires the involvement of a specialist

3) Data gathering and analysis: conducting of the experiment involving human participants and ensure compliance with relevant ethical and safety policies

4) Discussion and improvement suggestions: Extract meaningful and interpretation of results applying models and theories in accordance with the previous literature findings

Specific considerations of each of these steps of the framework are as outlined above which will help future research project designs.

9.1 Limitations

There are definitely many limitations to the current phishing experiment. One major factor was the restricted financial budget which limits the research to be conducted with a consumer-grade eye tracker. Given the current project is a one-year honours research, investing in a research-grade eye tracker is not a viable option since there are currently no concrete plans or research projects that would continue to utilise the equipment. When operating the eye-tracker there are noticeable 0.5 to one-second delay between actual gaze and the system detecting the movement. This definitely has a major impact on the accuracy and reproducibility of the data gathered especially when the experiment is operated on a timescale of milliseconds. Not only the hardware that was limited, but the analysis software also needs to be developed by the researcher. A self-developed analysis software poses great risks for undetected bugs or errors in code to persist which could lead to inaccurate results obtained when compared to more professionally tested and developed software that comes with a research-grade tracker. This is evident where the results of three of the email samples were discarded due to a bug in

the software. The self-developed software also limits to only basic analysis that can be performed on the data due to limited time for development as well as the availability of open-source libraries to use free of charge. Results discussed above although are in line with previous literature, the small sample size prevents a concrete assertion of the findings, which unfortunately puts this in the pitfall that many other human-centric research projects are also suffering with reproducibility and generalisability of results [7],[8], [9]. This, however, is justified similarly to the above point with a restricted budget and short-term nature of the project itself. The main aim of the current project also is to derive a generic framework from the phishing study that would help future research designs hence more focus is given towards documenting the processes and considerations involved instead.

9.2 Implications and Suggestions for Future Research Design

The proposed framework could be applied to future human-centric cybersecurity research involving human participants. The framework can help guide researchers on planning, acquiring equipment, setting up the environment, running the experiment and analysis of the data. The framework also extends beyond cybersecurity research projects but can also be applied to human-centric research in other domains such as marketing or psychology with some variations as appropriate. Suggestions for future research design is to purchase a professional well-developed analysis software and/or a research-grade eye tracker (both of these components often come together as a bundle). Although the cost will be significantly higher, the benefits that such investment will bring to the quality of data obtained, reproducibility of results, ease of setting up the test environment and reputation in publishing the research papers will be justified. With such investment involved, it is recommended that similar human-centric research be done as part of a longer-term research project i.e. postdoctoral research or PhD programme. This will help ensure continuity in equipment usage as well as having sufficient time and resources to conduct quality research.

Many future research potentials were discussed throughout this report. The notable proposal includes exploring the design of cybersecurity training courses to ensure maximal effectiveness in skill development to protect against modern cyberattacks; factors influencing cost-benefit criterion of users when making a decision whether an email is a phishing attempt or genuine; or a more thorough investigation of the current exemplar phishing study to include larger participant groups, complex result analysis, and use of research-grade apparatus. These potential topics will build up upon findings of the current study, utilising the proposed framework, and to contribute to tackling human-centric issues faced in the cybersecurity domain.

Chapter 10

Conclusion

The current project produced a generic framework to aid future human-centric cybersecurity researches which consists of four related tasks: stimuli design, eye tracker tools, data gathering and analysis, and discussion and improvement suggestions. These tasks were obtained from conducting a phishing study with the use of eye-tracking technology. Data from the study are analysed with the Signal Detection Theory and Operant Conditioning which uncovers issues relating to usability of anti-phishing tools and limitations of the human cognitive processes namely the prevalence effects, cognitive overloading, gain maximisation, and familiarity/trust, similar to those outlined in previous literature. The eye tracker captured user behaviour of checking links, sender's address, and banking information which confirms their understanding of basic cybersecurity protection strategies in line with the high performance result in the email classification task conducted. The survey prompted interesting discussion regarding effectiveness of training program, previous scamming experiences, and strategies to promote cybersecurity among the general population. A display of potential improvement of anti-phishing tool to be more user friendly based on the gathered data were also presented. Limitations of the project mainly concern the restricted finance which impacted the utilisation of a research-grade eye tracker. Despite this, implications of the proposed framework and suggestions for future research projects were discussed which also includes its applications beyond cybersecurity research projects.

Chapter 11

Bibliography

[1] C. Evans and C. Smith, ”Beyond Obfuscation: The Defence Industry’s Position within Federal Cybersecurity Policy,” The National Defense Industrial Association 2019.

[2] C. Raiu, ”Cyber-threat evolution: the past year,” Computer Fraud Security, vol. 2012, pp. 5-8, 2012.

[3] A. Gostev, ”Cyber-threat evolution: the year ahead,” Computer Fraud Security, vol. 2012, pp. 9-12, 2012.

[4] B. K. Wiederhold, ”The Role of Psychology in Enhancing Cybersecurity,” Cyberpsychology, Behavior, and Social Networking, vol. 17, pp. 131-132, 2014.

[5] S. Kraemer, P. Carayon, and J. Clem, ”Human and organizational factors in computer and information security: Pathways to vulnerabilities,” Computers Security, vol. 28, pp. 509-520, 2009.

[6] Deloitte, ”Understanding Phishing Techniques,” 2019.
Available: https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/sea-risk-cyber-101-part10.pdf.

[7] K. Parsons, A. McCormac, M. Pattinson, M. Butavicius, and C. Jerram, ”The design of phishing studies: Challenges for researchers,” Computers security, vol. 52, pp. 194-206, 2015.

[8] S. Das, A. Kim, Z. Tingle, and C. Nippert-Eng, ”All About Phishing: Exploring User Research through a Systematic Literature Review,” 2019.

[9] R. Zhao, S. John, S. Karas, C. Bussell, J. Roberts, D. Six, B. Gavett, and C. Yue, ”Design and evaluation of the highly insidious extreme phishing attacks,” Computers security, vol. 70, pp. 634-647, 2017.

[10] G. O. f. Science, ”Using behavioural insights to improve the public’s use of cyber security best practices,” 2014.
Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads
/attachment_data/file/309652/14-835-cyber-security-behavioural-insights.pdf.

[11] N. C. S. Centre, ”Common cyber attacks: reducing the impact ” 2016.
Available: https://www.ncsc.gov.uk/collection/10-steps-to-cyber-security/
introduction-to-cyber-security/common-cyber-attacks-reducing-the-impact.

[12] G. Bai, ”COMP3320 Week 7: Software Vulnerability Scanning, Exploitation and Identification,” The University of Queensland, 2020.

[13] F. Y. Rashid. (2017). Types of phishing attacks and how to identify them. [Online]. Available: https://www.csoonline.com/article/3234716/types-of-phishing-attacks-and-how-to-identify-them.html.

[14] L. Irwin. (2020). The 5 most common types of phishing attack. [Online].
Available: https://www.itgovernance.eu/blog/en/the-5-most-common-types-of-phishing-attack.

[15] M. Alsharnouby, F. Alaca, and S. Chiasson, ”Why phishing still works: User strategies for combating phishing attacks,” International Journal of Human - Computer Studies, vol. 82, pp. 69-82, 2015.

[16] S. Thompson, ”ScamBlocker,” The Journal of the American Taxation Association, vol. 27, pp. 113-115, 2005.

[17] Infosec. (2020). Anti-Phishing Hardware Software. [Online]. pp. 5-10.
Available: https://resources.infosecinstitute.com/category/enterprise/
phishing/phishing-countermeasures/anti-phishing-hardware-software/gref.

[18] PhishProtection. (2020). What You Need to Know About Anti-Phishing Software. [Online].
Available: https://www.phishprotection.com/content/anti-phishing-solution/anti-phishing-software

[19] N. Mazher, I. Ashraf, and A. Altaf, ”Which web browser work best for detecting phishing,” in Information Communication Technologies 5th International Conference, At Karachi, Pakistan, 2013.

[20] H. Zuhair, A. Selamat, and M. Salleh, ”Survey of anti-phishing tools with detection capabilities,” in Biometrics and Security Technologies (ISBAST) 2014 International Symposium, Kuala Lumpur, 2014.

[21] A. Abbasi, F. Zahedi, and Y. Chen, ”Impact of anti-phishing tool performance on attack success rates,” in 2012 IEEE International Conference on Intelligence and Security Informatics (ISI), 2012.

[22] A. Jones, ”How do you make information security user friendly?,” Information Security Technical Report, vol. 14, pp. 213-216, 2009.

[23] S. Sharma, ”Using Contextual Information to Improve Phishing Warning Effectiveness,” ProQuest Dissertations Publishing, 2015.

[24] H. Tuttle, ”The Risks of Mobile Payment Technology,” Risk Management, vol. 62, p. 36, 2015.

[25] N. J. Daniel, ”Predatory Personalities as Behavioral Mimics and Parasites: Mimicry-Deception Theory,” Perspect Psychol Sci, vol. 9, pp. 445-451, 2014.

[26] L. Li, E. Berki, M. Helenius, and S. Ovaska, ”Towards a contingency approach with whitelist- and blacklist-based anti-phishing applications: what do usability tests indicate?,” Behaviour Information Technology, vol. 33, pp. 1136-1147, 2014.

[27] M. A. Just and P. A. Carpenter, ”Eye fixations and cognitive processes,” Cognitive Psychology, vol. 8, pp. 441-480, 1976.

[28] D. Jeske, P. Briggs, and L. Coventry, ”Exploring the relationship between impulsivity and decision-making on mobile devices,” Personal and Ubiquitous Computing, vol. 20, pp. 545-557, 2016.

[29] P. Realpe-Muñoz, C. A. Collazos, J. Hurtado, T. Granollers, J. Muñoz-Arteaga, and J. Velasco-Medina, ”Eye tracking-based behavioral study of users using e-voting systems,” Computer Standards Interfaces, vol. 55, pp. 182-195, 2018.

[30] S. Furnell and N. Clarke, ”Power to the people? The evolving recognition of human aspects of security,” Computers Security, vol. 31, pp. 983-988, 2012.

[31] J. Valinsky. (2020). ’Shark Tank’ judge Barbara Corcoran gets her $400,000 back from scammers. [Online]. Available: https://edition.cnn.com/2020/03/02/business/barbara-cocoran-email-hack-money-returned/index.html.

[32] K. R. Yulia, ”Psycholinguistic Aspects of Humanitarian Component of Cybersecurity,” Psycholinguistics, vol. 26, pp. 199-215, 2019.

[33] B. D. Sawyer and P. A. Hancock, ”Hacking the Human: The Prevalence Paradox in Cybersecurity,” Human Factors: The Journal of Human Factors and Ergonomics Society, vol. 60, pp. 597-609, 2018.

[34] A. Horneman. (2019). Situational Awareness for Cybersecurity: An Introduction. [Online]. Available: https://insights.sei.cmu.edu/sei_blog/2019/09/situational-awareness-for-cybersecurity-an-introduction.html.

[35] C. D. Middlebrooks, T. Kerr, and A. D. Castel, ”Selectively Distracted: Divided Attention and Memory for Important Information,” Psychological Science, vol. 28, pp. 1103-1115, 2017.

[36] A. Harley. (2016). Prospect Theory and Loss Aversion: How Users Make Decisions. [Online]. Available: https://www.nngroup.com/articles/prospect-theory/.

[37] BBC, ”Cyber-stalking,” 2020. Available: https://www.bbc.com/news/topics/c5elz932pjyt/cyber-stalking.

[38] D. Heeger. (2006). Signal Detection Theory. [Online]. 1(1).
Available: https://www.cns.nyu.edu/ david/courses/perception/lecturenotes/sdt/sdt.html.

[39] J. Martin, C. Dubé, and M. D. Coovert, ”Signal Detection Theory (SDT) Is Effective for Modeling User Behavior Toward Phishing and Spear-Phishing Attacks,” Hum Factors, vol. 60, pp. 1179-1191, 2018.

[40] M. Jaclyn, ”Something Looks Phishy Here: Applications of Signal Detection Theory to Cyber-Security Behaviors in the Workplace,” University of South Florida: Dissertations Theses, pp. 1-71, 2017.

[41] CriticalStart, ”The Impact of Security Alert Overload,” CriticalStart 2019.
Available: https://www.criticalstart.com/wp-content/uploads/CS_MDR_Survey_Report.pdf.

[42] S. Sendelbach and M. Funk, ”Alarm fatigue: a patient safety concern,” AACN Adv Crit Care, vol. 24, pp. 378-386, 2013.

[43] K. Gaines. (2019). Alarm Fatigue is Way Too Real (and Scary) For Nurses. [Online]. Available: https://nurse.org/articles/alarm-fatigue-statistics-patient-safety/.

[44] H. R. Division and J. Martin, ”Training Report: Effectiveness of Cybersecurity Training,” Taxas Department of Transportation 2019. Available: http://ftp.dot.state.tx.us/pub/txdot-info/hrd/training/report-phishing-dark-waters.pdf.

[45] K. F. Tschakert and S. Ngamsuriyaroj, ”Effectiveness of and user preferences for security awareness training methodologies,” Heliyon, vol. 5, p. e02010, 2019.

[46] R. Dodge, K. Coronges, and E. Rovira, ”Empirical Benefits of Training to Phishing Susceptibility,” Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 457-464.

[47] M. Petock, ”Anti-Phishing Training: IS IT Working? Is It Worth It?,” Carnegie Mellon University, 2020. Available: https://insights.sei.cmu.edu/insider-threat/2020/01/anti-phishing-training-is-it-working-is-it-worth-it.html

[48] E. Lastdrager, I. Gallardo, P. Hartel, and M. Junger, ”How Effective is Anti-Phishing Training for Children?,” Industrial Engineering Business Information Systems, vol. 1, pp. 229-239, 2017.

[49] A. Vance, B. Anderson, C. Kirwan, and D. Eargle, "Using Measures of Risk Perception to Predict Information Security Behavior: Insights from Electroencephalography (EEG)," Journal of the Association for Information Systems, vol. 15, pp. 679-722, 2014.

[50] A. Abbasi, M. Zahedi, and Y. Chen, "Phishing Susceptibility: The Good, the Bad, and the Ugly," IEEE International Conference on Intelligence and Security Informatics, vol. 1, pp. 1-6, 2016.

[51] B. F. Skinner, "Two Types of Conditioned Reflex: A Reply to Konorski and Miller," The Journal of general psychology, vol. 16, pp. 272-279, 1937.

[52] L. Learning. (2020). Reinforcement and Punishment. [Online].
vailable: https://courses.lumenlearning.com/waymaker-psychology/chapter/operant-conditioning

[53] U. o. S. Coast. (1998). The Basic Findings In Instrumental/Operant Conditioning. [Online]. Available: https://userweb.ucs.louisiana.edu/ cgc2646/LRN/Chap4.html.

[54] D. Di Giacomo, J. Ranieri, M. D'Amico, F. Guerra, and D. Passafiume, "Psychological Barriers to Digital Living in Older Adults: Computer Anxiety as Predictive Mechanism for Technophobia," Behav Sci (Basel), vol. 9, p. 96, 2019.

[55] M. Anderson and A. Perrin, "Barriers to adoption and attitudes towards technology," Pew Research Center 2017. Available: https://www.pewresearch.org/internet/2017/05/17/barriers-to-adoption-and-attitudes-towards-technology/.

[56] A. o. C. F. Examiners. (2019). Elderly fraud scams: How they're being targeted and how to prevent it. [Online]. Available: https://www.acfe.com/fraud-examiner.aspx?id=4294997223.

[57] L. Raine, "Senior Citizens and Digital Technology," Pew Research Center 2012. Available: https://www.pewresearch.org/internet/2012/09/15/senior-citizens-and-digital-technology/.

[58] A. C. C. Commission. (2020). Advice for older Australians. [Online].
Available: https://www.scamwatch.gov.au/get-help/advice-for-older-australians.

[59] D. Lormel, "Fraud Against the Elderly," Federal Bureau of Investigation 2001. Available: https://archives.fbi.gov/archives/news/testimony/fraud-against-the-elderly.

[60] E. L. Carlson, "Phishing for elderly victims: as the elderly migrate to the Internet fraudulent schemes targeting them follow," The Elder law journal, vol. 14, p. 423, 2006.

[61] M. Wong, E. Gardiner, W. Lang, and L. Coulon, "Generational differences in personality and motivation: Do they exist and what are the implications for the workplace?," Journal of managerial psychology, vol. 23, pp. 878-890, 2008.

[62] Dementia. (2019). Memory Loss with Aging: What's normal, what's not? [Online]. Available: https://www.dementia.com/memory-loss.html

[63] C. Iuga, J. R. C. Nurse, and A. Erola, "Baiting the hook: factors impacting susceptibility to phishing attacks," Human-centric computing and information sciences, vol. 6, pp. 1-20, 2016.

[64] Z. Leron, The Psychology of Information Security. Ely: IT Governance Ltd, 2016.

[65] C. A. Harper, L. P. Satchell, D. Fido, and R. D. Latzman, "Functional Fear Predicts Public Health Compliance in the COVID-19 Pandemic," Int J Ment Health Addict, p. 1, 2020.

[66] D. Dariusz, Techniques of Social Influence. London: Taylor and Francis. Taylor Francis Group Routledge, 2015.

[67] I. Rijnetu. (2019). Here are the Top Online Scams You Need to Avoid Today [Updated 2019]. [Online]. Available: https://heimdalsecurity.com/blog/top-online-scams/hitmanscam.

[68] K. Sheridan. (2017). Phishing Emails that Invoke Fear, Urgency, Get the Most Clicks. [Online]. Available: https://www.darkreading.com/endpoint/phishing-emails-that-invoke-fear-urgency-get-the-most-clicks/d/d-id/1330100.

[69] L. Karisny. (2013). Common Sense Cybersecurity. [Online].
Available: https://www.govtech.com/dc/articles/Common-Sense-Cybersecurity.html.

[70] S. Milgram, "Behavioral Study of obedience," Journal of abnormal and social psychology, vol. 67, pp. 371-378, 1963.

[71] A. Gendre, "Anatomy of a Phishing Email," VAD Secure 2013.
Available: https://www.vadesecure.com/en/anatomy-of-a-phishing-email/.

[72] P. Organisation. Phishing Techniques. [Online]. Available: https://www.phishing.org/phishing-techniques

[73] U. o. Cambridge. (2003). Psycholinguistic evidence on scrambled letters in reading. [Online]. Available: https://www.mrc-cbu.cam.ac.uk/people/matt.davis/cmabridge/.

[74] R. Keith, J. W. Sarah, L. J. Rebecca, and P. L. Simon, "Raeding Wrods with Jubmled Lettres: There Is a Cost," Psychol Sci, vol. 17, pp. 192-193, 2006.

[75] N. Luhmann, H. Davis, M. King, C. Morgner, J. Raffan, and K. Rooney, Trust and Power. Newark: Polity Press, 2017.

[76] D. Gefen, "E-commerce: the role of familiarity and trust," Omega (Oxford), vol. 28, pp. 725-737, 2000.

[77] VadeSecure, ”Phishers’ Favorites Top 25,” 2019. Available: https://www.vadesecure.com/en/phishers-favorites-q2-2019/.

[78] E. J. Williams and D. Polage, ”How persuasive is phishing email? The role of authentic design, influence and current events in email judgements,” Behaviour Information Technology, vol. 38, pp. 184-197, 2019.

[79] M. Bill, ”Improving response with personalised emails,” Direct Response, p. 29, 2006.

[80] ”Effective email communication,” Administrative Assistant’s Update, p. 7, 2015.

[81] M. Plummer. (2019). How to Spend Way Less Time on Email Every Day. [Online]. Available: https://hbr.org/2019/01/how-to-spend-way-less-time-on-email-every-day

[82] R. Sobers. (2020). The Anatomy of a Phishing Email. [Online].
Available: https://www.varonis.com/blog/spot-phishing-scam/.

[83] N. D. Smith, ”The best (and worst) words to use in an email subject line,” DM news, vol. 37, p. 12, 2015.

[84] C. Monitor. (2019). The 15 Most Powerful Words in Subject Lines. [Online]. Available: https://www.campaignmonitor.com/blog/email-marketing/2019/02/power-words-in-email-subject-lines/.

[85] T.-N. Heidi, ”The psychology of personalization,” Graphic Arts Monthly, p. 13, 2000.

[86] A. Post. (2020). Scam Alerts. [Online]. Available: https://auspost.com.au/about-us/about-our-site/online-security-scams-fraud/scam-alerts.

[87] A. C. C. Commission. (2020). Current COVID-19 (coronavirus) scams. [Online]. Available: https://www.scamwatch.gov.au/types-of-scams/current-covid-19-coronavirus-scams.

[88] A. C. C. Commission. (2020). Spot the scam signs. [Online].
Available: https://www.scamwatch.gov.au/about-scamwatch/tools-resources/online-resources/spot-the-scam-signs.

[89] C. Bank. (2020). Latest Email and SMS Scams Mimicking CommBank. [Online]. Available: https://www.commbank.com.au/support/security/sms-phishing-scams.html.

[90] C. Kachornvuthidej, ”Honours Eyetracker,” 2020.
Available: https://github.com/CNK-THA/Honours_EyeTracker.

[91] Infosec. (2020). Top 9 Phishing Simulators [Updated 2020]. [Online].
Available: https://resources.infosecinstitute.com/top-9-free-phishing-simulators/gref.

[92] S. Govardhan. (2017). How to design an effective phishing simulation? [Online]. Available: https://www.cisoplatform.com/profiles/blogs/how-to-design-an-effective-phishing-simulation.

[93] Z. Zorz. (2019). Phishing attacks are a complex problem that requires layered solutions. [Online]. Available: https://www.helpnetsecurity.com/2019/10/24/phishing-attacks-solutions/.

[94] T. Pro. (2020). What are the benefits of eye tracking in research? [Online]. Available: https://www.tobiipro.com/blog/what-is-eye-tracking

[95] B. Farnsworth. (2018). Eye Tracking: The Complete Pocket Guide. [Online]. Available: https://imotions.com/blog/eye-tracking/.

[96] M. Cognolato, M. Atzori, and H. Müller, "Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances," J Rehabil Assist Technol Eng, vol. 5, p. 205566831877399, 2018.

[97] Y. Wang, G. Zhai, S. Chen, X. Min, Z. Gao, and X. Song, "Assessment of eye fatigue caused by head-mounted displays using eye-tracking," Biomed Eng Online, vol. 18, pp. 111-119, 2019.

[98] iMotions. (2015). Screen-Based Eye Tracker vs Eye Tracking Glasses - What's the Difference? [Online]. Available: https://imotions.com/blog/screen-based-eye-tracker-vs-eye-tracking-glasses/.

[99] K. Holmqvist, M. Nyström, and F. Mulvey, "Eye tracker data quality: what it is and how to measure it," ACM, 2012, pp. 45-52.

[100] Tobii. (2020). Tobii Eye Tracker 5. [Online]. Available: https://gaming.tobii.com/.

[101] Tobii. (2020). Specifications for the Tobii Eye Tracker 4C. [Online].
Available: https://help.tobii.com/hc/en-us/articles/213414285-Specifications-for-the-Tobii-Eye-Tracker-4C.

[102] Tobii, "Core SDK Link," 2019. Available: https://developer.tobii.com/community/forums/topic/core-sdk-link/.

[103] J. Carmichael. (2020). Occupational Health and Safety Risk Management. [Online]. Available: https://ppl.app.uq.edu.au/content/2.30.01-occupational-health-and-safety-risk-management.

[104] T. U. o. Queensland. (2020). Ethics, integrity and compliance. [Online].
Available: https://research.uq.edu.au/research-support/ethics-integrity-and-compliance.

[105] R. Adrian and B. Lorenzo, "Ten inbox secrets: What eye tracking reveals about designing better emails," Journal of Direct, Data and Digital Marketing Practice, vol. 14, p. 46, 2012.

[106] Tobii. (2016). Frequently asked questions. [Online]. Available: https://help.tobii.com/hc/en-us/articles/212824889-Frequently-asked-questions.

[107] Eitol, ”Tobii eye tracker linux installer,” 2018.
Available: https://github.com/Eitol/tobii_eye_tracker_linux_installer.

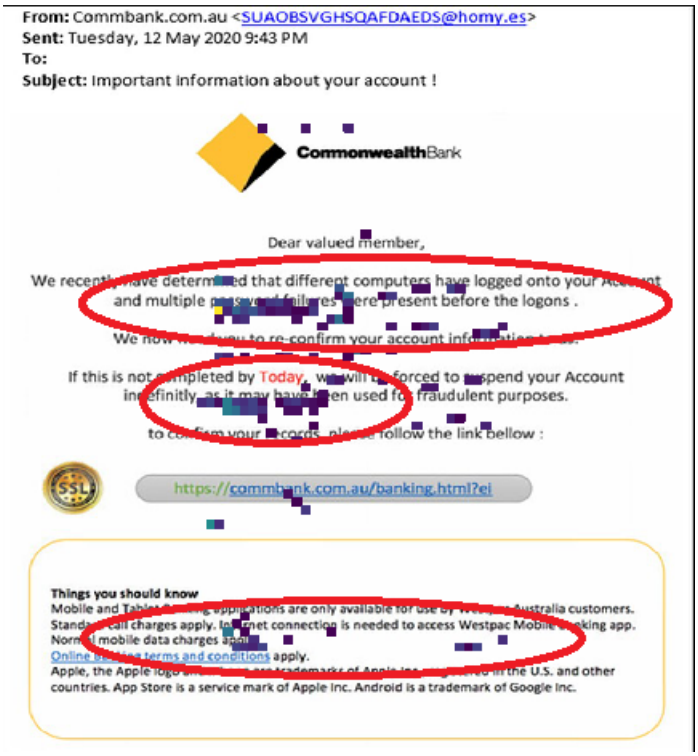
[108] S. Clements, ”5 Tips for an Effective Email,” Arkansas business, vol. 35, pp. 19-19, 2018.

Appendix A

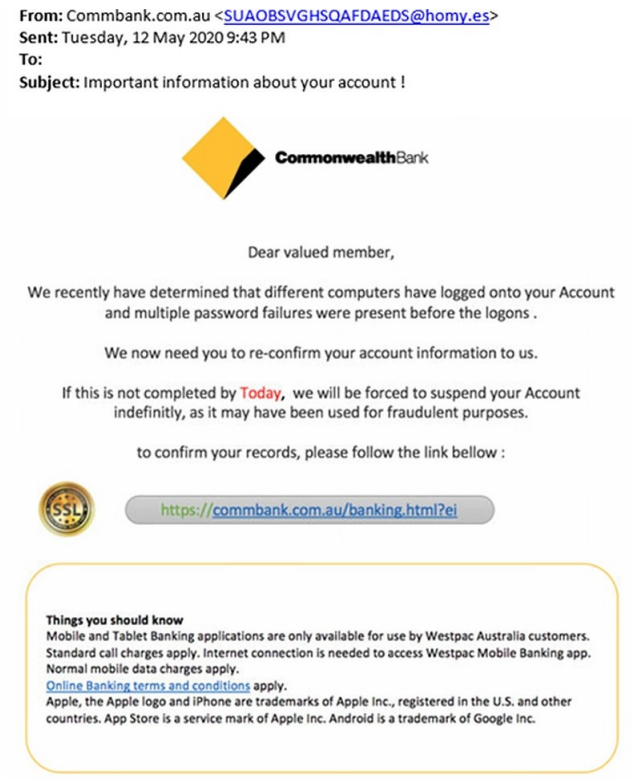
Eye Tracking Results

The following figures are results of the eye-tracking experiment as mentioned in the Result and Discussion sections of this report. The full set of images produced from the experiment including raw data can be found here [90].

A.1 Appendix-1

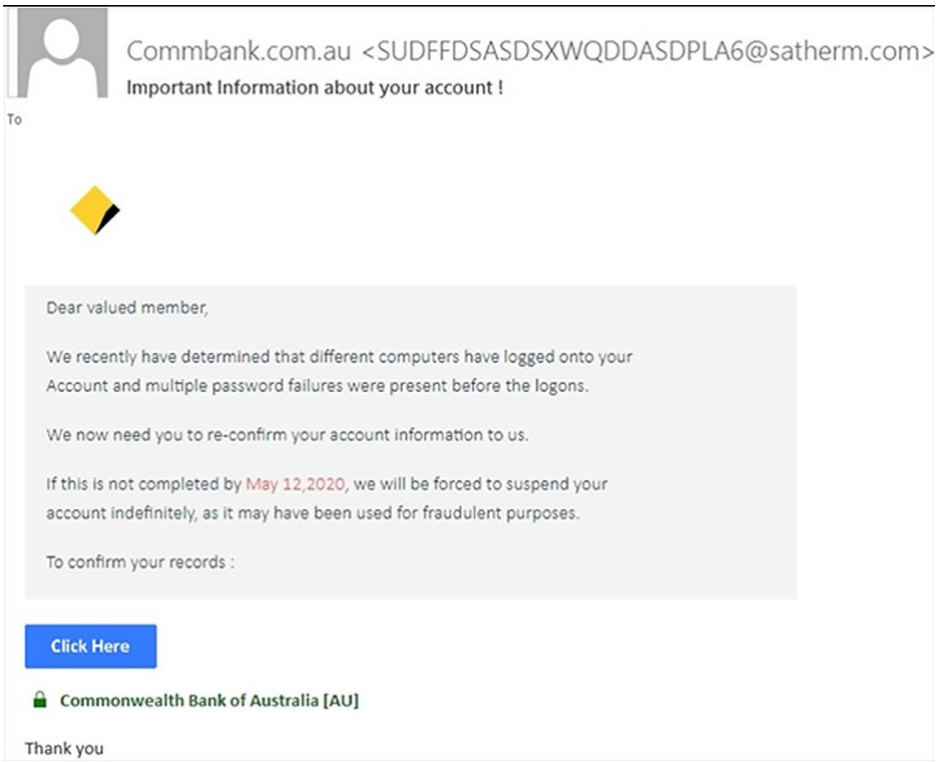


-

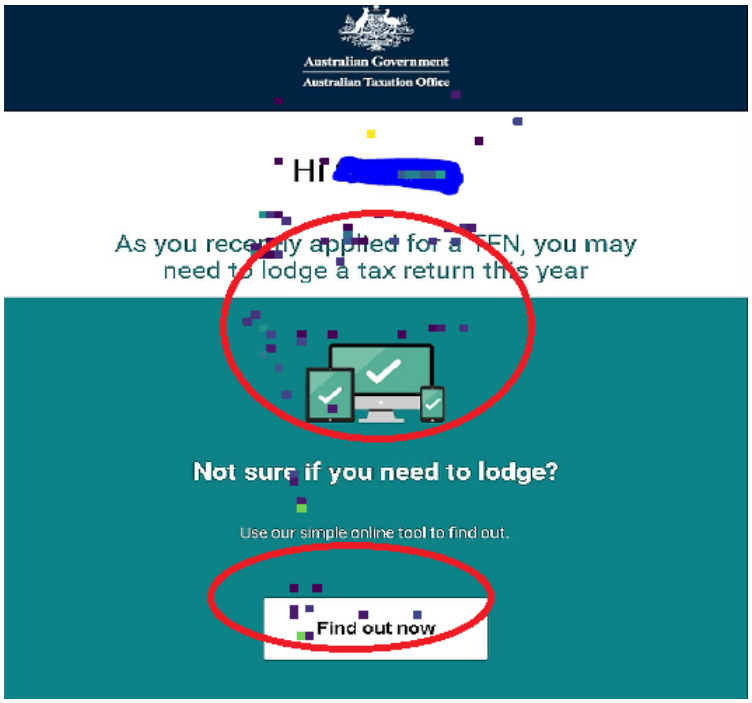


A.2 Appendix-2

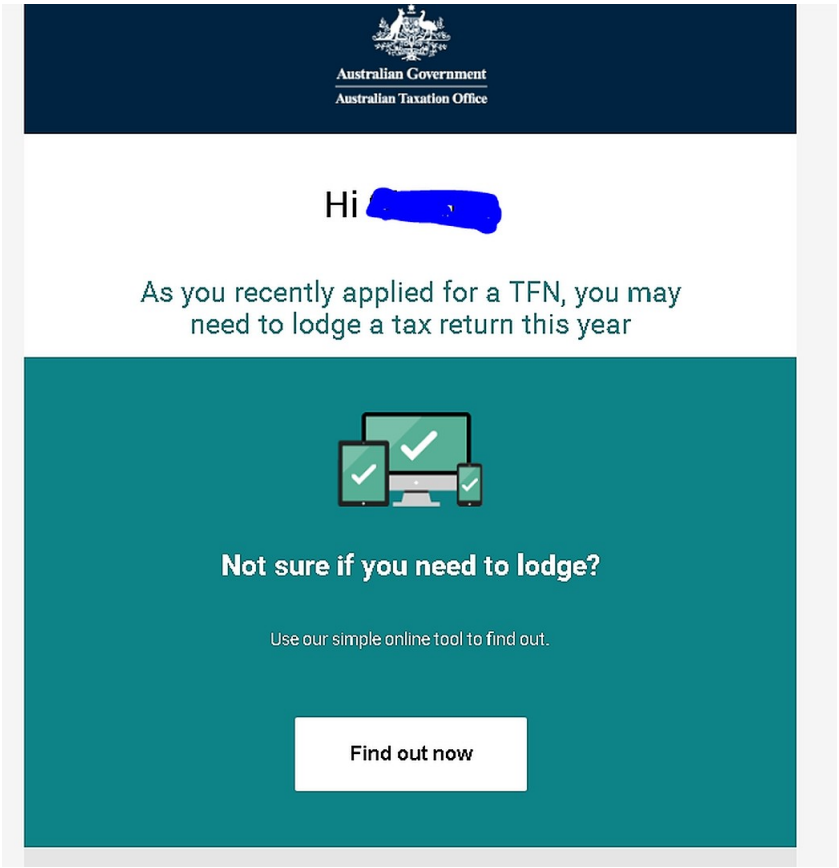
-



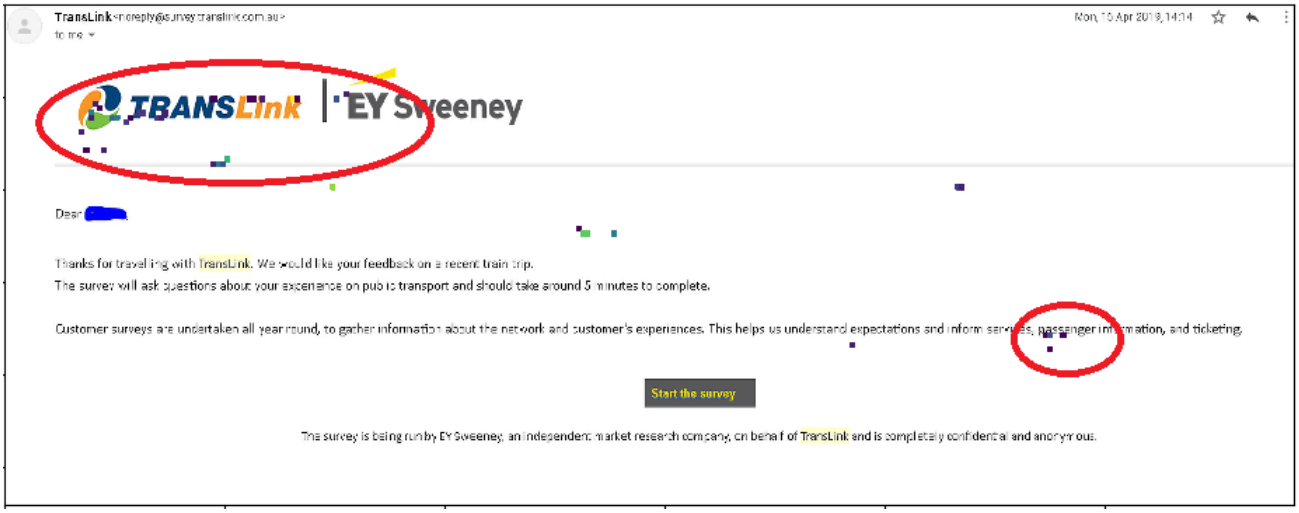
A.3 Appendix-3

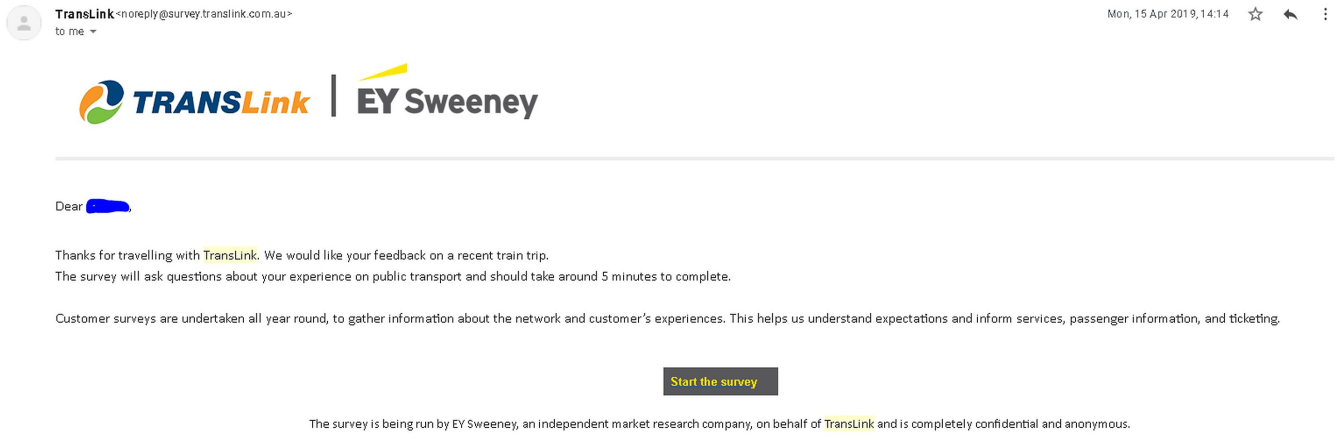


-

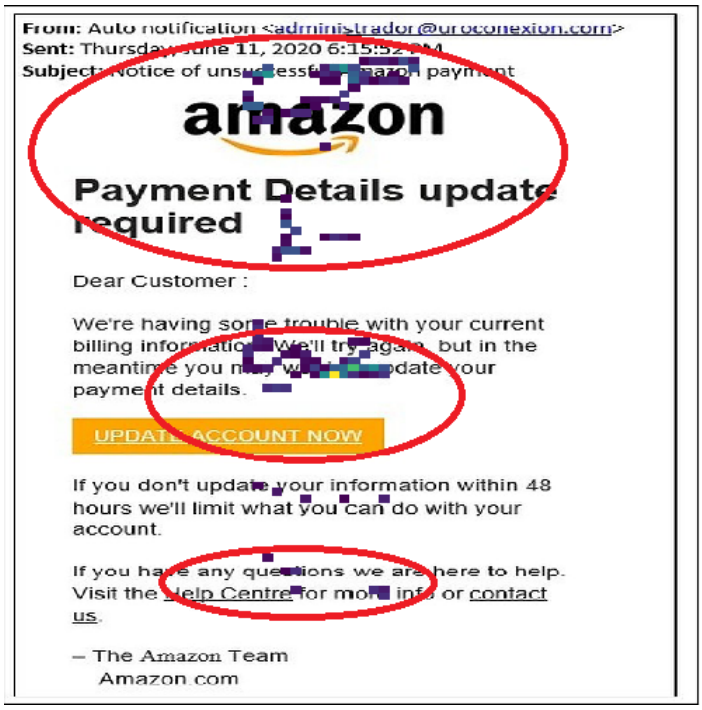


A.4 Appendix-4

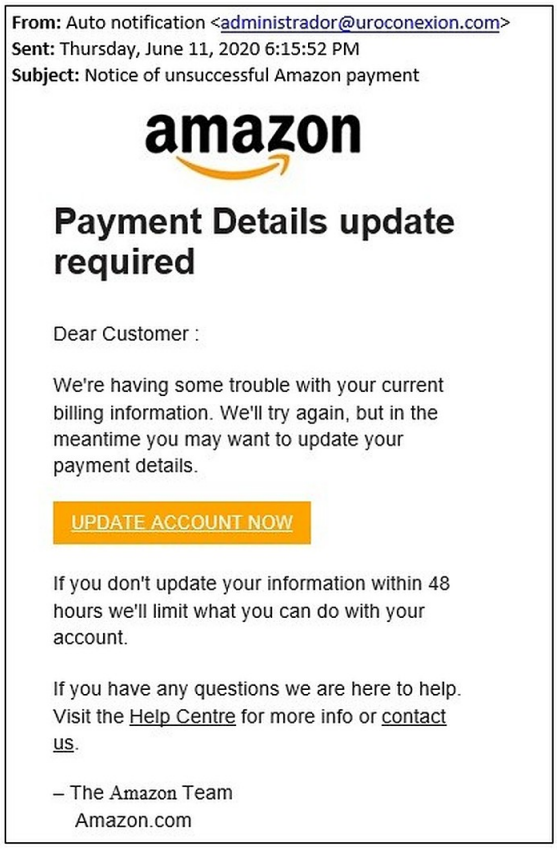




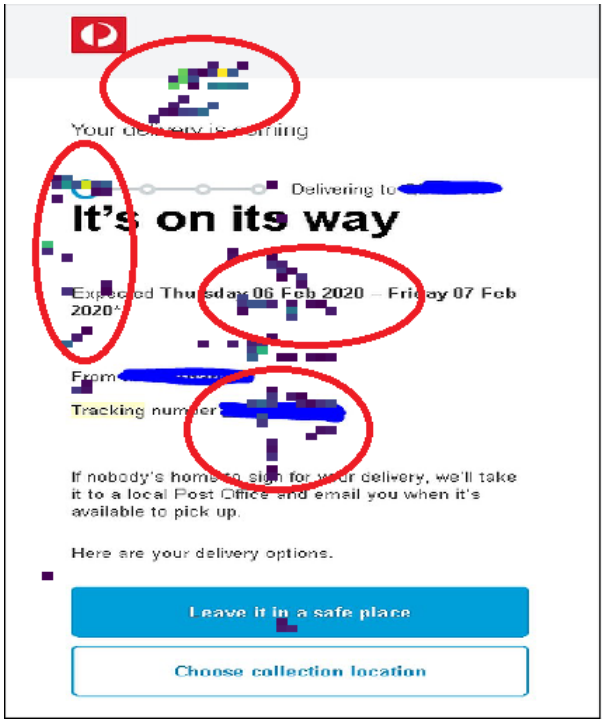
A.5 Appendix-5



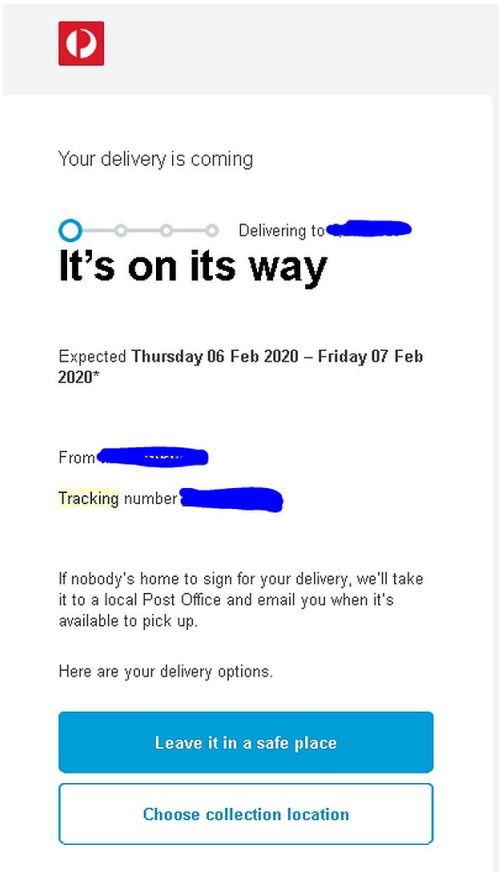
-



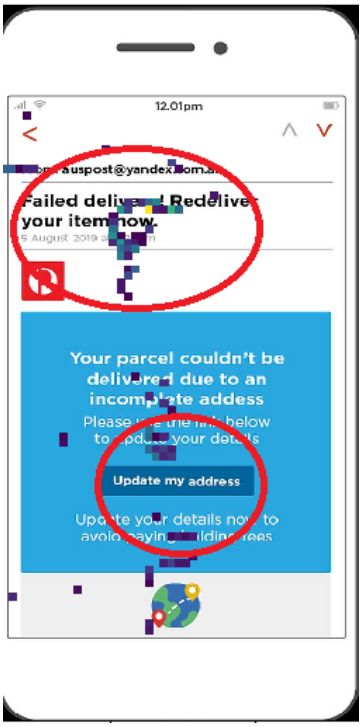
A.6 Appendix-6



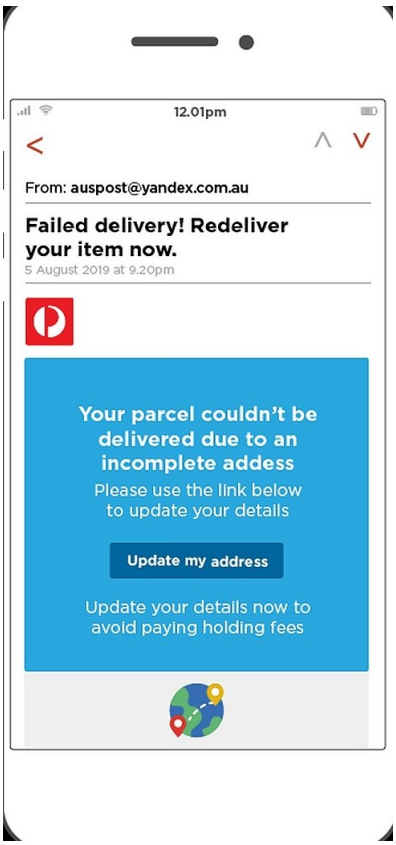
-



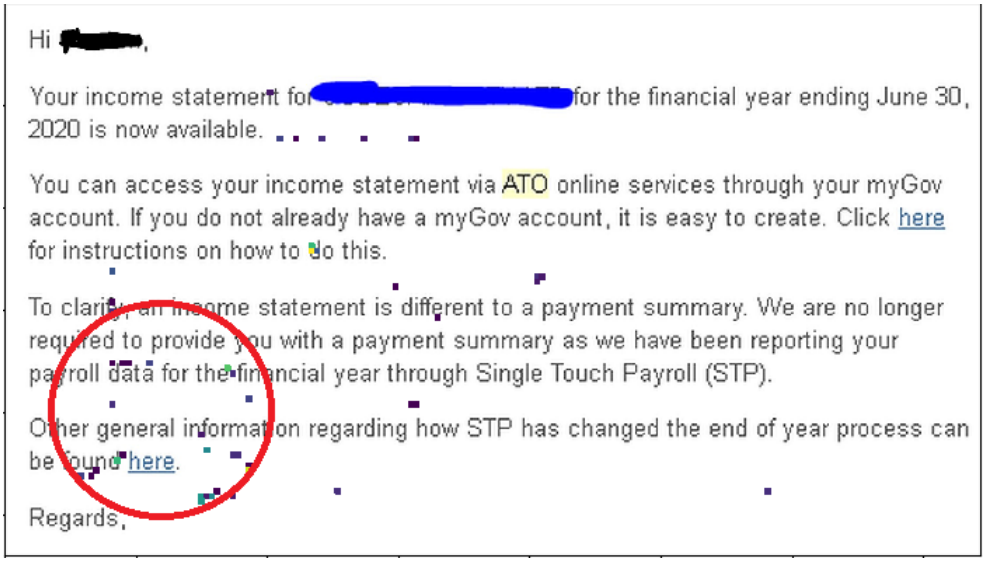
A.7 Appendix-7



-



A.8 Appendix-8



Hi [redacted],

Your income statement for [redacted] for the financial year ending June 30, 2020 is now available.

You can access your income statement via ATO online services through your myGov account. If you do not already have a myGov account, it is easy to create. Click here for instructions on how to do this.

To clarify, an income statement is different to a payment summary. We are no longer required to provide you with a payment summary as we have been reporting your payroll data for the financial year through Single Touch Payroll (STP).

Other general information regarding how STP has changed the end of year process can be found here.

Regards,

-

Hi [redacted],

Your income statement for [redacted] for the financial year ending June 30, 2020 is now available.

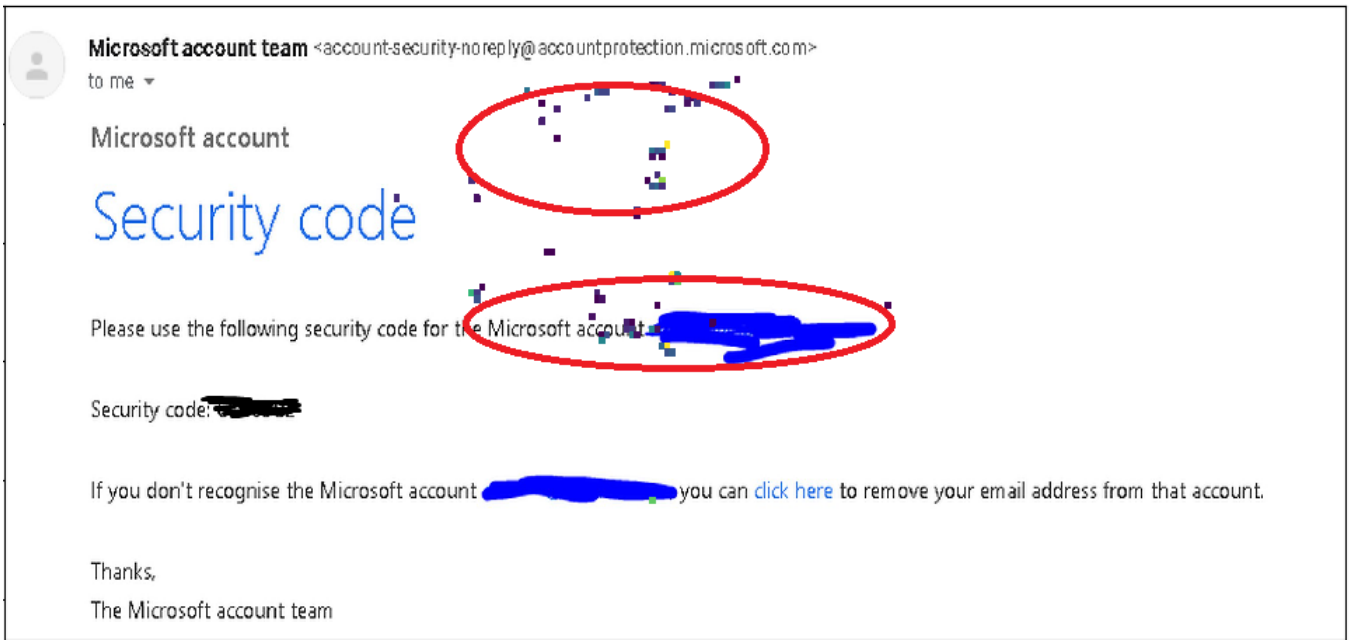
You can access your income statement via ATO online services through your myGov account. If you do not already have a myGov account, it is easy to create. Click here for instructions on how to do this.

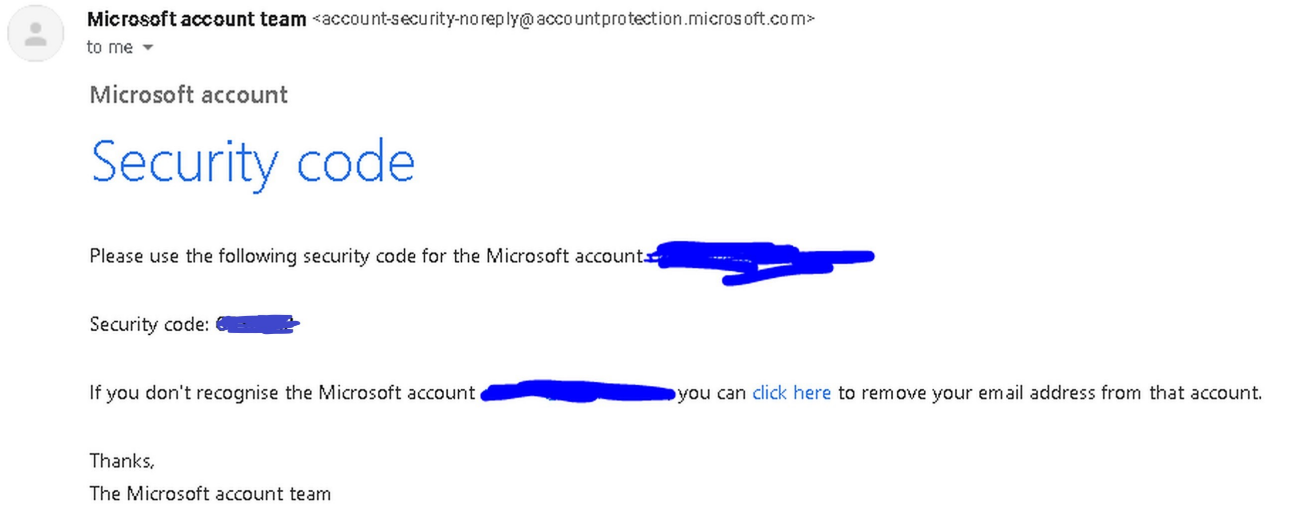
To clarify, an income statement is different to a payment summary. We are no longer required to provide you with a payment summary as we have been reporting your payroll data for the financial year through Single Touch Payroll (STP).

Other general information regarding how STP has changed the end of year process can be found here.

Regards,

A.9 Appendix-9

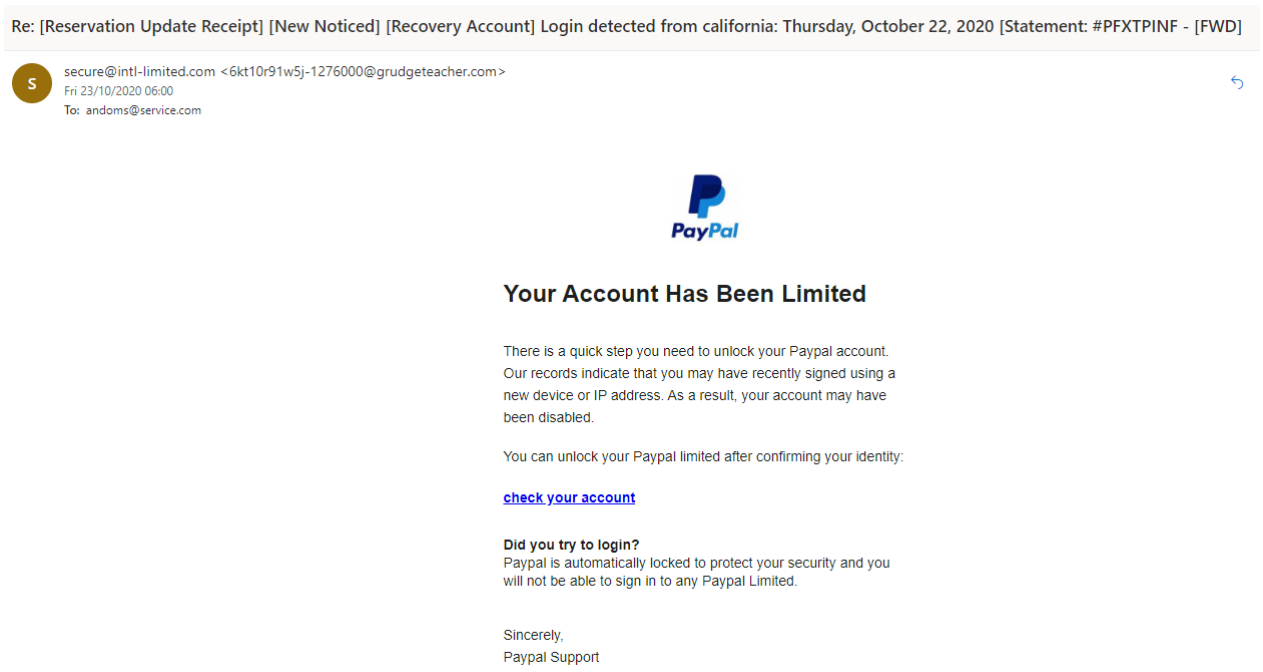




Appendix B

Additional Real Phishing Email Samples

B.1 Appendix-1



Re: [Reservation Update Receipt] [New Noticed] [Recovery Account] Login detected from california: Thursday, October 22, 2020 [Statement: #PFXTPINF - [FWD]

S secure@intl-limited.com <6kt10r91w5j-1276000@grudgeteacher.com>
Fri 23/10/2020 06:00
To: andoms@service.com

PayPal

Your Account Has Been Limited

There is a quick step you need to unlock your Paypal account. Our records indicate that you may have recently signed using a new device or IP address. As a result, your account may have been disabled.

You can unlock your Paypal limited after confirming your identity:

check your account

Did you try to login?
Paypal is automatically locked to protect your security and you will not be able to sign in to any Paypal Limited.

Sincerely,
Paypal Support

B.2 Appendix-2