VICTORIA UNIVERSITY OF WELLINGTON

Te Whare Wānanga o te Ūpoko o te Ika a Māui

A Modern Perspective on Phishing:

**An investigation into susceptibility to phishing
attacks between mobile and desktop email clients.**

By

Jayden Nowitz

A thesis

submitted to the Victoria University of Wellington in fulfilment of the
requirements for the degree of Master of Science in Computer Science.

Faculty of Engineering and Computer Science

Victoria University of Wellington

2018

Abstract

Research on how to counter phishing from a user behavior perspective has been explored for over a decade, yet the prevalence of such threats is increasing. This thesis aims to provide a modern perspective by considering if there is a difference in how susceptible an individual is on a mobile device versus a desktop email client. Currently very few studies consider phishing on mobile devices and the research is unclear as to the potential difference in susceptibility rates between the two device types. Initially a review of 60 phishing emails received by the university that had passed mail filtering were used to assist in the design of the messages to be used in the second stage of the study. Following this a simulated phishing attack on two groups in one unit of professional administrative staff in the university (141 in total with 71 in Group A and 70 in Group B) within the university was undertaken. The defining characteristic between the groups was how they responded to a message with a 'loss versus gain' appeal. This area has received limited exploration in the research and findings remain unclear. This study found that people were statistically far more susceptible to the 'gain' message of *a free coffee* at 28.2% than the 'loss' message of *Office365 account suspension* 7.1% . For device type there appears to be no statistically significant difference, even between the groups. This study highlights the complexities of device usage around phishing, that have not been clearly highlighted in previous studies, such as people viewing emails with one device and falling victim on another device.

Acknowledgements

I would like to thank my supervisors A/Prof Ian Welch and A/Prof Val Hooper. Without their support, guidance, insights, and wisdom this thesis would not have been possible. This research also would not have been possible without the enthusiasm and dedication of a few individuals (that won't be named for confidentiality) in the university and outside who assisted me in making data collection possible and tirelessly helped me to get the permission to conduct this study. Without all of your support, I would have never been able to do this masters thesis. To my friends and family you have kept my morale up, a massive thank you for all your encouragement throughout this journey.

Table of Contents (Chapters)

1.Introduction	1
2. Literature Review.....	3
2.1 Previous phishing Research.....	4
2.2 Mobile Research.....	10
2.3 Common theories in phishing research.....	13
2.3.1 Elaboration Likelihood Model (ELM)	13
2.3.1.1 ELM in Security.....	14
2.3.2 Heuristic Systematic Model (HSM).....	15
2.4 Message Characteristics	17
2.5 Methods in behavioural Phishing research	19
2.6 Research issues	22
2.6.1 Mobile research	22
2.6.2 Phishing attacks and messages	23
2.6.3 Generalisability of sample population	24
3. Methodology and Study One design	24
3.1 Coding scheme	25
3.2 Results	25
3.3 Analysis and discussion of Study One	28
4. Methodology round 2: Study two	29
4.1 Study tool selection	30
4.2 Study two method	32
4.2.1 Message design	33
4.2.2 Human Ethics.....	35
4.3 Hypotheis development.....	35

4.3.1 'Loss versus gain'	36
4.3.2 The effect of training.....	36
4.3.3 Device Type	36
4.4 Results of Study two	37
4.5 Analysis of study two.....	41
4.5.1 Hypothesis One.....	42
4.5.3 Hypothesis 3	44
5. Discussion.....	47
6. Limitations, implications for practice and future research.....	54
6.1 Limitations	54
6.2 Implications for practice	56
6.3 Future Research.....	57
7 Conclusion	58
<u>Appendices</u>	
Appendix A: Prior Phishing research and methods.....	i
Appendix B: Previous constructs used in phishing studies	vii
Appendix C Coding schema	ix
Appendix D (part one): Technical considerations	xi
Appendix D (part two): Tool Selection	xiii
Appendix E: Message design	xvi
Appendix F: Human ethics.....	xix
Appendix G: Summary of raw data study two	xx
References.....	80

List of Tables and figures:

Figures

Diagram A	Help desk scam	26
Diagram B	Phishing email imitating Xero	27
Diagram C	Emails sent in experiment	34

Tables

Table 1	Tool selection scoring	31
Table 2	Experiment design	33
Table 3	Study two results (aggregate)	38
Table 4	Contingency and Chi-square test Hypothesis 1	42
Table 5	Chi-square Hypothesis 1 (by action)	44
Table 6	Contingency and chi-square test Hypothesis 3	45

1. Introduction

Phishing attacks are a current Information Security threat to both personal use and corporate/enterprise environments alike, that has recently gained attention despite this threat having been prevalent for more than 10 years. The Anti-Phishing Working Group observed an increase of 5753% in phishing email campaigns over 12 years (APWG, 2016). The total number of phishing attacks in 2016, was also 65% greater than in 2015, at 1,220,523 attacks (APWG, 2016). Demonstrating that despite the attack vector having been measured by the Anti-Phishing Working Group since 2004 and despite extensive research efforts it remains a present threat. This research is focussed upon users' susceptibility to phishing attacks between mobile and desktop email clients. It also aims to establish the residual effect of awareness materials on user susceptibility between both device types.

User susceptibility is an important aspect of phishing research as awareness and security behaviour are still considered important aspects of Information Security defence, albeit a 'last line'. Furthermore, phishing emails are one of the most common methods to spread 'malware' such as ransomware. Awareness is considered by the Internet Crime Complaint Center (2015) as the primary way to mitigate ransomware because it targets the end-users in enterprise environments. The SANS institute (2016) in its technical guide on ransomware also highlights the importance of awareness in that one of the most important prevention methods is for users not to fall victim to the phishing emails. As illustrated with zero-day malware spread through phishing emails that the rate of infection was decreased by organisations having Information Security awareness training (Tran, Campos-Nanez, Fomin & Wasek, 2016). Indicating the current need for increased awareness and a focus on user's susceptibility to phishing attacks.

The majority of the current phishing research has been heavily focussed upon users of desktop email clients. It has been suggested by recent academic research that users are more susceptible on mobile devices due to the size of the devices and other factors around their usage and called for further research in this area (Harrison, Vishwanath, Ng & Rao, 2015; Iuga, Nurse & Erola, 2016). The Anti-Phishing Working Group (2013) in its Mobile Fraud white paper back in 2013 highlighted the range of hardware and software issues make them potentially

more vulnerable to phishing or Smishing (SMS based phishing attacks), the susceptibilities around the architecture of the Android ecosystem making them more susceptible to malware. Despite this issue having been identified for a while in Information Security literature little of the previous research has considered testing this assumption or how much more susceptible people are on mobile versus desktop email clients. Furthermore, research into mobile phishing is limited especially from a user perspective (Iuga et al, 2016; Ngoqo & Flowerday, 2015; Bottazi, Cassalicchio, Cingolani, Marturana & Piu, 2015 and Dang-pham & Pittayachawn, 2015). The above illustrates the need for a study that compares susceptibility between the two devices for both Information Security professionals and academic research communities. This area as it relates to mobile phishing is of more importance, since as a lot of Information Security resources on phishing have been intended for desktop environments and as such contain little information or even transferable information to mobile devices (Armin, 2013). This research provides two distinct advantages, in that it aims to understand susceptibility between mobile and desktop in greater detail and how well phishing training translates across modalities in reducing this susceptibility.

The following research questions were proposed to guide this study:

- 1) What is the difference in susceptibility to phishing attacks between a mobile versus a desktop email client?
- 2) What is the residual impact of traditional Information Security workshops on the level of susceptibility

The aim behind this thesis is to provide actionable research that is relevant to operational Information Security professionals, and provides guidance. The aim is to consider the use of simulated phishing attacks as a method to build immunity to phishing amongst employees in enterprise environments. The approach to this research is informed by Robey and Markus (1998, 7) in that it aims to provide research that balances rigour with relevance to a practitioner audience, which they term “consumable research”. Arguably the ability to profile user’s behaviour, resulting from this research would be less rigorous than forming a theory explaining why people fall victim but more relevant to a practitioner audience. It is hoped that this can potentially guide medium to large enterprise environments in identifying how to

establish or improve robust ways to measure the risk of employees to phishing attacks by taking into account awareness training.

The study aims to utilise a field experiment to explore and understand the issues around these two questions. It aims to strike a balance between the amount of control researchers have in a traditional usability experiment versus capturing people's natural behaviour. This research will consider the relevant security literature in the formulation of its methodology. To achieve this it reviews the relevant previous streams in mobile security and trust as well as phishing research. This informs the first round of the methodology which is to determine the representative phishing messages of those generally received and the difference in how they appear between mobile and desktop email clients. This evaluation will then be used to craft phishing emails that are typical of actual phishing emails in how they try to deceive users.

2. Literature Review

The phishing research can be classified into two streams: (a) technological research, that looks at the application of technology to detect phishing emails and (b) behavioural research. Phishing as a security threat consists of both sociological and technological aspects (Mohammad, Thabtah & McCluskey, 2015). This research focusses on the behavioural phishing research and tries to uncover why people fall victim to phishing emails and then how to reduce their susceptibility. The aim is to understand how people react and process the messages containing deceptive cues which are aimed at encouraging victims to carry out the action the attacker desires. Examples of recent technical based phishing research is evidenced by the following studies Bottazzi et al (2015) and Wu, Du and Wu (2016). For instance, Wu et al (2016) developed a tool using Optical Character Recognition (OCR) software to detect phishing URLs on mobile devices. This study does not consider technological approaches; as despite advances in this area of research the threat remains present furthermore as both professional bodies have highlighted with types of malware spread by phishing such as Zero-days and ransomware awareness is a major mechanism of defence against this (SANS Institute, 2016 & Tran et al, 2016).

Within this area there have been a broad range of studies and approaches. Research ranges from, Human Computer Interaction based research, like Iuga, Nurse & Erola (2016) using software designed to help assess usability such as mouse cursor heat maps or eye-tracking as

in Alsharnouby, Alaca & Chiason (2015), to studies predicting susceptibility such as Luo, Zhang, Burd and Seazzu (2013). Due to the variety of perspectives utilised to try predict and explain phishing susceptibility this literature review aims to synthesise the past literature across the variety of streams taking a behavioural approach. Some past researchers have approached this by the methodology used (Pattinson et al, 2012) and similarly by considering the different approaches within a given method (Jansson & Von Solms, 2013). Other studies have applied models from other disciplines (Vishwanath Hearth, Chen Wang & Rao, 2011; Harrison, et al, 2015). Since this paper is trying to explore the difference between device types it does not directly apply a theory. The varying approaches applied in the wider body of literature can be seen in Appendix A. This thesis aims to present a modern account of the research in this area.

Due to the lack of research around the difference between mobile and desktop devices or on mobile phishing. The available literature on mobile has generally been based on assumptions or mobile has not been considered (Iuga et al, 2016; Vishwanath, 2016 & Harrison et al, 2015). Due to the wide range of perspectives into addressing this attack vector the literature as briefly outlined above is intended to provide an overview of where there is scope for new research especially around device type as a variable. The literature review is organised as follows: (a) To considers what is known around phishing and the mobile privacy and security literature (to understand potential differences or perceptions between device types). (b) To consider the theories around the processing of phishing emails and awareness materials and to understand the literature on deception. (c) To review the major research methods used in the phishing research. Finally to consider the gaps and limitations in the current research.

2.1 Previous phishing Research

Prior to understanding the role device differences can play in phishing susceptibility, it is important to consider what is known from previous behavioural research. Phishing attacks are designed to bypass technical defences and exploit the user's cognitive biases and get them to rely on their heuristics, being a form of social engineering, the attackers utilise concepts from marketing to persuade the user to perform an action by creating senses of urgency, fear of loss, or importance (Mohammad et al, 2015; Workman, 2008; Luo et al, 2013; and Davinson & Sillence, 2010). Vishwanath et al (2011) argue that people do not fall victim due to a lack of

ability but instead a failure of cognitive processing of the email and other environmental factors. For instance, those who receive a lot of emails tend to engage in less cognitive processing and those who ritually check emails tend to engage in less consideration of the message (Vishwanath et al, 2011). Phishing emails are designed to deceive a person based upon factors that would at face value establish credibility, because if someone methodically analysed the message they would be likely to correctly identify the phishing. Luo et al (2013) found from a simulated spear-phishing attack in a university that 36% clicked the link and 15% gave credentials. They consider that the phishing email they employed mimicked what victims would expect as well as appearing to come from a legitimate authority figure, and presented what would appear to be a genuine scenario for its purported purpose (Luo et al, 2013). This also seemingly lends support to the notion that the cues within the phishing message are able to fool a person into incorrectly thinking the message is legitimate.

Other methods for deception and establishing credibility are the framing of the message, the perception of presence and targeting the message. Such factors have been elaborated on in recent studies where they have manipulated the messages sent to the test subjects (Harrison et al, 2015; Vishwanath, 2016 & Goel, Williams & Dincelli, 2017). Harrison et al (2015), by applying the Heuristic Systematic Model (HSM) (this will be discussed later), considered the use of visual cues in phishing emails; where the 'rich' message (used images and contact numbers) had higher social presence (the perception of distance from the sender). They found participants tended to heuristically process the emails, and people were twice as likely to fall victim to the 'rich' email message (Harrison et al, 2015). One would assume based upon the above that if systematically processed then images and phone numbers would not affect a person's likelihood to regard the message as legitimate, since they would be able to identify the deception techniques used by the sender. Goel et al (2017) found people are more susceptible to messages framed around loss or gain or with greater contextualisation (more targeted emails). They suggest that contextualised loss based messages are the most effective technique; however, across the different messages people are more motivated by the prospect of winning rather than losing (Goel et al, 2017). This research highlights that the techniques used in the message are a significant factor in the success of a phishing campaign.

The credibility and appearance of authority was supported in earlier research around personality and susceptibility. The role of personality has been considered in a range of studies

to predict people's susceptibility to phishing (Workman, 2008; Vishwanath, 2015; Alseadoon, Othman, Foo & Chan, 2013; Halevi, Lewis & Memon, 2013; Moody, Galletta, Walker & Dunn, 2011). Workman (2008) found that in both self-reported and the observed behaviour that people who were more trusting or who comply more willingly to authority are more susceptible to falling victim to phishing. Moody et al (2011) considered the following personality factors: degree of trust, curiosity, risk, boredom, entertainment and ability to focus, on phishing susceptibility and found overall that few of these variables are good predictors. They argue instead that researchers should consider other personal characteristics such as self-efficacy or security orientation (Moody et al, 2011). Suggesting that a person's personality is not necessarily a reliable predictor of susceptibility. There has been found to be a link between having a suspicious personality and self-efficacy around detecting phishing scams. Harrison, Vishwanath and Rao (2016a) found, that people who have higher levels of suspicion around communications, also have better self-efficacy, meaning they are more likely to perceive deficiencies in material conveyed for them to assess. They are more likely to systematically process emails resulting in less trust of the phishing message (Harrison et al, 2016a). They found though not originally hypothesised that the higher self-efficacy around identifying deception was significant in predicting the perception around messages containing limited cues (Harrison et al, 2016a). While this does highlight a personality predictor it is related to a behavioural characteristic. Vishwanath (2015) consider that personality types do not predict if a person will systematically (cognitive) or heuristically process the phishing message, but interestingly perception of a lack of information encourages greater systematic processing. Overall, personality is a poor predictive factor of susceptibility to phishing and instead dual process theories offer a better framework to predict this.

Behavioural characteristics, consistent with what Moody et al (2011) encouraged, have been a continued area of research development in explaining why people fall victim. Common behavioural characteristics that have been considered quantitatively range from knowledge and self-efficacy through to an individual's belief's and habits, refer to Appendix B (Vishwanath et al, 2011; Vishwanath, Harrison & Ng, 2016; Arachchilage & Love, 2014; Harrison, Svetiva & Vishwanath (2016b); Sun, Yu, Lin & Tseng, 2016). Research considering knowledge has been inconsistent in its relationship to measures relating to the process (will be elaborated on later). Field related knowledge as a variable was not supported by Vishwanath et al (2011) but in Harrison et al (2016b) phishing and email knowledge was

supported in predicting attention and elaboration. Thus, emphasising that with the knowledge construct there has been a range of conceptualisations in the research to predict one's behaviour or intentions toward the threat. The complexity around this construct with phishing can be evidenced by Arachilage and Love (2014), in a survey they considered the effect of procedural and conceptual knowledge on a person's self-efficacy. Procedural knowledge is the person's ability to know how to carry out an action whereas conceptual is the ability to explain the why, they found that only the intersection of the two explains the individual's self-efficacy (Arachilage & Love, 2014). Considering the role of the different dimensions of knowledge on understanding phishing, behaviour appears to be an area that needs further consideration.

Self-efficacy is an important variable in predicting behaviour around susceptibility. Higher self-efficacy has been shown to increase the ability to avoid succumbing to phishing attacks (Arachilage & Love, 2014). It has been found that increased self-efficacy around internet usage is associated with greater self-efficacy in coping with phishing, i.e. the ability to more successfully handle this threat safely (Sun et al, 2016). Though it seems probable from the literature that self-efficacy impacts behaviour or intention rather than the processing, but as Vishwanath et al (2011) highlighted it does not explain the extent of elaboration of the message. Studies of self-efficacy have inconsistent results around predictive capability. Other studies have illustrated that previous experience can be a predictor of being able to better handle these threats (Moody et al, 2011 & Benenson, Gassmann & Landwirth, 2017). This would support the rationale behind simulated phishing research and attacks as an awareness measure (to be considered later).

One stream of research that has gained more attention from a usability perspective has been the websites used by the phishers rather than the original messages. Alsharnouby et al highlight (2015) where the participants stated how confident they were that the websites presented to them were genuine or illegitimate, only 53% of the participants could correctly identify the phishing websites. This underscores that while people find it hard to identify phishing emails, the websites themselves are also carefully designed to deceive the potential victim. Elements used in the design of phishing websites has been highlighted by Iuga et al (2016), Alsharnouby et al (2015) and Downs, Holbrook and Cranor (2006). Iuga et al (2016) examined how the participants determine if a website is legitimate, using Facebook, as an

example and found that many participants could not safely distinguish between pop-up based phishing sites especially if coupled with other indicators such as an https address, and argue that many do not consider URLs in detecting phishing websites. This study supports the previous literature, with respect to emails as to what informs credibility to users. A major limitation of this study is that it only focussed on Facebook which, is a flaw in that it has been found many user's trust emails that appear to come from trusted sources such as banks or government institutions (Parsons et al, 2015a). The finding around security indicators such as https was supported in a much earlier study. Downs et al (2006) who found that participants could identify fake sender names but could not apply this properly, such as not noticing aspects of the URL correctly and mistaking locks in the webpage as equivalent to the one in the browser. They further found that in their sample the best predictor was familiarity with phishing techniques, since many focussed on the message content (Downs et al, 2006). This supports the findings of Iuga et al and Alsharnouby et al around phishing websites. Alsharnouby et al (2015), had the users pretend they arrived at the website from an email. They found, by utilising eye-tracking, that the detection rate for phishing websites in their experiment was 53% compared to 79% for the legitimate websites (Alsharnouby et al, 2015). Eye tracking showed many participants are inconsistent with the strategies they use to identify fraudulent websites but the most used strategies were in order: the content of the page, looking for bugs in the website, studying the URL, cross referencing using a search engine and the presence of a SSL symbol (Alsharnouby et al, 2015). These studies emphasise that people have difficulty in determining the legitimacy of URLs, while the bug strategy would not work in email due to its static nature. Arguably it is more important for research to look at how to lower susceptibility and encourage better management of emails, rather than to worry about the strategies used.

The managing of genuine versus illegitimate emails rather than the websites used by phishers has been considered in a few studies. Emails potentially provide phishers with more cues to exploit, while websites are usually the second stage in phishing attacks (Alsharnouby et al, 2015). Parsons et al (2015a) confirm the importance of visual cues in that participants rejected genuine emails due to a lack of company logo in the message. Furthermore, they also suggest that authority plays an important role with emails that appear to be from banks and government institutions as more likely to require 'follow-up' and that people are more susceptible to emails that present a gain or loss, which they posit may encourage people not

to consider security in these situations (Parsons et al, 2015a). While the visual component suggests conformation of the findings of Harrison et al and Goel et al around visual cues and the potential to lose or gain and reinforces Luo et al findings on authority. Another study Pattinson et al (2012) emphasise that familiarity around computers was strong predictor being less likely to incorrectly manage the phishing emails. Downs et al (2006) also emphasised that one of the best predictors of being able to identify phishing emails is prior experience with the scams. One would expect this emphasis on behavioural characteristics such as experience and knowledge to underscore the efficacy of awareness and training around phishing. Emphasising that the message presented to the participants has a significant impact upon whether people consider it legitimate, however the person's knowledge and experience may yet be the best protection from a behavioural perspective.

Awareness training is often emphasised as a solution for combatting the general susceptibility of people to phishing campaigns. Much of the current literature has emphasised how their findings can inform awareness training around phishing (Halevi et al, 2013; Moody et al, 2011 & Pattinson et al, 2012). The impact of regular awareness training around zero-day malware spread by phishing has been emphasised by Tran et al (2016), who highlight that while there is no significant difference between monthly and yearly training it has a clear impact as compared to no training. They found in their simulations of the spread of the malware that the incident rate where there is no awareness training was 132,368 machines per hour compared to almost 20,000 for yearly training and 15,474 with monthly training. Calls for regular awareness training around phishing and emphasising its risks, has been regularly made in numerous studies (Workman, 2008; Pattinson et al, 2012; Alseadoon et al, 2013 & Greitzer et al, 2014). Jansson and VonSolms (2013) emphasise the efficacy of embedded training in simulated phishing attacks as an awareness training measure amongst participants with no prior informed consent.

These findings can be contrasted with those of Caputo, Pfleeger, Freeman & Johnson (2014) who found that very few people spent enough time on the training pages to read them properly or that many more skim read it and 11% continued to click regardless of training. Participants who ignored company alerts, infact suggested company alerts as a method for the company to communicate this information (Caputo et al, 2014). Caputo et al (2014) used participants who were informed that they were going to be assessed on their security

behaviours ; the susceptibility rates decreased significantly between trial one and two in their study, 60% to 34%. Though there is debate around the effectiveness of training arguably simulated attacks do foster an improvement in decreasing susceptibility rates. Ngoqo and Flowerday (2015) emphasise with respect to SMS based phishing that over time awareness levels increase with exposure to more training but the person's intention towards behaving securely around phishing increases to a lesser extent. Underscoring that awareness is a slow and iterative process whereby immediate effects cannot be expected after three cycles, as in Caputo et al.

Hesitancy around awareness has been emphasised in wider behavioural research suggesting it is an area of debate. Downs et al (2006) also emphasise that understanding certain indicators such as SSL and https does not necessarily translate into correct results as the participants that were aware of this were just as likely to fall victim to the phishing email that contained this cue. They argue that awareness cannot necessarily be linked to greater perceptions of vulnerability or necessarily better strategies around identifying phishing messages (Downs et al, 2006). However some authors have also been hesitant over awareness as a solution. Vishwanath et al (2011) states that while it should increase knowledge this is only useful if the person shows deliberation in processing of the message, while Benenson et al (2017), consider that phishing tests can foster a culture of suspicion within an organisation if the security team runs simulated phishing campaigns. Awareness training as has been emphasised in the awareness literature is not a 'silver bullet' to behavioural Information Security issues (of which phishing is only one example). Pfleeger, Sasse and Furnham (2014) note that while the clear majority of organisations have implemented awareness training, few are realising increased levels of awareness or the better security behaviour they aimed for. While its efficacy of awareness can be demonstrated, there are issues surrounding its implementation and achieving the desired results.

2.2 Mobile Research

It is important to consider the impact of mobile device type separate from the wider literature around phishing due to the lack of empirical consideration of the potential impact of device type on susceptibility. Some studies have suggested that there is a difference in susceptibility, between desktop and mobile environments. More recent research has considered that mobile

is an avenue that is missing from the current phishing literature (Jansson & Von Solms, 2013 & Iuga et al, 2016). Iuga et al (2016) call for research into susceptibility on mobile devices, due to a number of their participants initially opening the web-based experiment on these devices, suggesting that the screen size might be a factor in any difference in susceptibility (Iuga). Whereas, Harrison et al (2015) consider that due to the shorter periods of use and less concentration that people use on mobile devices will heuristically process, regardless of presence. Vishwanath et al (2016) consider their model which extends the HSM could be used to explain SMS based mobile attacks that get the user to download a mobile app or click on a malicious link. The current literature provides very little indication as to whether susceptibility differs between devices, let alone how many respond on mobile.

The closest studies to this thesis that have considered the role of mobile devices in phishing, are the behavioural studies are Vishwanath (2016) and Ngoqo & Flowerday (2015). Ngoqo and Flowerday (2015) examining smishing and the correlation between awareness and the behavioural intent (considered to be the number with the 'naïve mistake' behavioural profile). Mobile phones are not just prone to phishing through message platforms. Vidas et al (2013), found that with QR phishing that 75% visited out of curiosity to the page, furthermore most of these hits came from just the sticker, followed by the one with instructions and the Social Network study advertisement. For email we should know if they are more susceptible on mobile Vishwanath (2016) considered both device types, though it was self-reported they found that mobile devices reinforced habits of frequently monitoring email, where victimisation is explained by habit and heuristic processing resulting from this. They considered that the role of device was not significant in relation to the processing of cues (Vishwanath, 2016). This confirms the model of Vishwanath et al (2016) around self-regulation impacting habits in predicting suspicion. While it explains the behavioural factors that lead to falling victim, what remains uncertain is if there is a difference between the proportions who fall victim on mobile relative to desktop email clients.

One needs to consider the wider literature on perceptions around mobile device security to more fully understand how susceptibility may differ between device type. Foundational to an understanding of the differences is the consideration of the earlier mobile research, into devices that could handle email and other productivity applications. Botha, Furnell & Clarke (2009) presented a comparison of how security indicators are shown on a Windows PDA

relative to the desktop version. They found that with mobile, they present less security indicators and functionality, than with the desktop version, noting that the issues around usability and security are in how they are implemented (Botha et al, 2009). This means that for security in a mobile environment there is a trade-off between the familiarity with the desktop version, the ease of use of the mobile and the detail provided to the user. There are some differences in perceptions around the security of mobile technology features. This was highlighted by Chin, Felt, Sekar and Wagner (2012) who found that with tasks involving financial transactions there was more security concern with mobile than desktop, with banking 13% said they did not perform this on their phone for security concerns compared to 2% for desktop. Though, they found there was no difference between the willingness or past actions around handling work emails on their phone as compared to their laptop (0% for both device types) (Chin et al, 2012). This study indicates that unlike banking there should not be a difference between opening emails on the two device types.

It has been highlighted that there is a level of awareness around some mobile threats but people may not necessarily carry out the necessary protective behaviour. Alsaleh, Alomar and Alarifi (2017) found that behaviours in different aspects of mobile security among Android users are associated with each other: protecting Wi-Fi networks is associated with performing backups and installing smartphone protection applications, and the vice-verse was also found. They found that users generally would still perform potentially risky behaviour such as connecting to public Wi-Fi networks or location services, despite being aware of the potential risks. Presumably if a user perceives the cost of the risk as being great and has the knowledge to mitigate the risk then they will perform the appropriate actions to protect themselves. This potentially echoes the findings of Ngoqo and Flowerday (2015) in that increased awareness levels do not result in an equivalent increase in intention towards carrying out the protective behaviours. It has also been highlighted; that people have a good understanding of malware, device authentication, data-theft and almost 60% read application permissions before downloading (Parker, Ophoff, Van Belle & Karia, 2015). However, there was less adoption and understanding around data encryption and a lack of understanding around App permissions risks, which the authors suggest there is a trade-off that the users do not necessarily know how to resolve between functionality and privacy (Parker et al, 2015). Indicating that while it is important to build awareness and knowledge, the challenge is in making less common protections seem simple. Potentially people are familiar enough with mobile phones that they

are roughly equivalent in their ability to manage malicious messages between the device types, so no difference is seen as in Vishwanath (2016). Alternatively if there is a difference it may potentially be due to habits or knowledge of where to get the more detailed message cues in the interfaces of the email clients. In which case it would suggest the need for regular awareness training like what was performed in the Ngoqo and Flowerday or Janssson & Von Solms studies.

2.3 Common theories in phishing research

An important avenue of the desktop based phishing literature has been the use of dual process theories to explain why people fall victim to phishing attacks. These theories are the Heuristic Systematic Model (HSM) or the Elaboration Likelihood Model (ELM). The models assume that either the information is processed on heuristics or peripherally (using mental shortcuts people develop) or alternatively systematically (cognitively) through an elaborated judgemental decision based on the persuasiveness of the message (Chaiken & Maheswaran, 1994). Both models have been applied to a range of studies on communication where they aim to explain persuasion. One could consider the phishing research using these models as a subsection of both the behavioural phishing and online persuasion literature. The Elaboration Likelihood Model, which will be considered first, is often used in conjunction with other theories to understand victimization or the likelihood of (Workman, 2008 & Vishwanath, et al 2011).

2.3.1 Elaboration Likelihood Model (ELM)

The Elaboration Likelihood Model is a dual process theory that consider heuristic and cognitive processing of information that have been applied in the social engineering and information security fields. The ELM considers that when someone processes the argument contained in a message this happens through one of two routes; centrally or peripherally depending on the perceived merit. If the message is considered to be of high merit then the detailed central analytical route is chosen while if lower value then the peripheral heuristic route is used. (Petty & Cacioppo, 1986). This theory of the ability of a message to persuade the recipient via, it's two processing routes is largely similar to the Heuristic Systematic Model (as will be explored later).

2.3.1.1 ELM in Security

The Elaboration Likelihood Model has been applied in the behavioural phishing literature to emphasise how material can persuade someone into taking the desired action of the sender. The ELM was applied by Workman (2008) to understand how the victim is persuaded into carrying out the action that the social engineer wants them to (compliance), they found that phishing campaigns try to encourage peripheral processing by using techniques that promote a perception of trust from the victim. The ELM has been considered together with other factors in the behavioural phishing research field. Vishwanath et al (2011) hypothesised that both a person's consideration of message factors such as an email's source or the presence of grammatical errors and behavioural factors such as knowledge predicts the level of elaboration (the extent the message is processed). They found that only the urgency cues within a message are elaborated upon; making it more likely to cause a person to respond and make quick generalisations around a message (Vishwanath et al, 2011). This finding underscores the advantage of ELM and the HSM, in that these models predict how a person processes the characteristics of a potential phishing message. Harrison et al (2016b) hypothesised in their research model that fear-based appeals in phishing emails would lead to less attention and elaboration of the message, whereas leakage cues, such as obvious grammatical errors, would seemingly increase attention and elaboration. They however could not support both hypotheses on message factors. While specific factors in the message may be uncertain, similar findings in the processing routes as emphasised in the HSM studies have been demonstrated.

ELM has been applied more widely than the phishing body of research within the behavioural security research field. Komatsu, Takagi and Takemura (2013) applied it with Protection Motivation Theory to understand the intention of users into removing malware from their PC after receiving an alert email that their machine was infected, they found that for low-knowledge users relied on their perception of trust in the sender of the message, but to change their attitude the receiving user needs to be able to comprehend the content of the warning. While this was applied to a legitimate security context this reveals how users can be 'persuaded' into installing malicious content, such as ransomware, on to their computer through phishing emails. The Elaboration Likelihood Model has been used to inform the

design of awareness materials in an intervention for field experiments based on physical social engineering (Bullée, Montoya, Pieters, Junger and Hartel, 2015). Bullée et al (2015) tested the difference between a group exposed to an awareness intervention, that appealed to both the central and peripheral processing routes through a detailed poster for the former and a lanyard for the latter; to see the difference in the proportion of the of the 'aware' group handed over the digital keys to someone who appeared to be trusted compared with those that did not. They found that 37% of the 'aware' group fell victim, compared to 62.5% of the group exposed to no interventions (Bullée et al, 2015). This highlights the efficacy of the ELM in the design of training materials to help reduce vulnerability to social engineering attacks. ELM has clear applicability for the analysis of the persuasive elements of a message, albeit applied to an intervention to change behaviour.

2.3.2 Heuristic Systematic Model (HSM)

The Heuristic Systematic Model or HSM has been applied more widely in the phishing literature to understand how people process the phishing messages. This can be seen by Luo et al (2013) who applied constructs informed by the HSM, in predicting why people fall victim to spear-phishing attacks, being the 'argument quality' of the email, perception of 'source credibility', 'genre conformity' (extent it replicates real messages), the extent of 'time pressure' and finally the individuals need to cognitively (systematically) process it. However, since they did not complete the survey as planned in the research they could only support the relationship of the 'argument quality', 'source credibility' and the 'genre conformity'. Thus, providing an indication as to how the HSM has been applied in the study design, and what can be measured from simulated attacks alone.

This can also be evidenced in its application to predict victimisation or the likelihood of falling victim, in some of the more recent phishing literature, across a variety of types of messages. The HSM in its recent applications to understand how people process phishing emails, has consistently underscored that they are heuristically processed and the greater the amount of cues in the emails strengthens this (Harrison et al, 2015; Vishwanath, 2016; Vishwanath et al, 2016). The findings on more cues in the messages increasing heuristic processing in

Vishwanath (2016) lends itself to further confirming the study of Harrison et al (2015)'s results on information richness conveying social presence.

One such extension of the HSM can be illustrated by Vishwanath, Harrison and Ng (2016) who tried to predict with a model the impact of heuristic or systematic processing and habit upon an individual's suspicion of a phishing email. They found that when heuristically processing people tended to not view the phishing message as suspicious whereas with systematic processing they were more likely to with both the link and the attachment based phishing emails (Vishwanath et al, 2016). Furthermore, they found that a person's beliefs on cyber risks predicted if they would systematically or heuristically process (Vishwanath et al, 2016). This reinforces the importance of the HSM in explaining the behaviours of what makes people susceptible and highlights the importance of behavioural factors in explaining when this is more likely to occur. Another such explanation of characteristics in predicting the likelihood to identify phishing attacks, considered more personality related factors than behavioural characteristics in predicting when the processing routes of the HSM applied. Harrison et al (2016b) associated suspicion with explaining a behavioural characteristic being the perception of having enough information, in predicting the processing route. Using a personality based construct in their model, this provides a clear link to the arguments of Moody et al around relying more on behavioural factors as predictors in phishing research. While this explains the processing route this is still fundamentally answering the 'why' related to victimisation to phishing emails.

Further manipulations of the phishing emails are highlighted in a later study of the HSM. Goel et al (2017) applied the HSM in the considering how different types of phishing messages affect the rate of victimisation and found that contextualised messages tend to result in more heuristic processing, which can direct people to ignore other factors that would reveal it is a phishing message. Furthermore, they consider based upon their findings in the context of the HSM literature that appeals for gain make it more likely a person will heuristically process the message (Goel et al, 2017). What is clear from the HSM literature is not only how the message can affect victimisation but also in trying to find an explanation for why people heuristically process emails, which as mentioned before is associated with being more likely to fall victim. Ultimately whether there is a difference in the predictive capabilities of the ELM or the HSM though highly similar models, the HSM in phishing arguably has had more direct application

to the processing of the message. This does not mean one can ignore the importance of the message characteristics as Goel et al illustrate, as this can account for wide variety of results in terms of the success of phishing attacks.

2.4 Message Characteristics

The objective of the studies considering the message characteristics is to emphasise how the language and presentation of the messages try to deceive the victims into complying. Rather than manipulating the type of message shown to participants based upon their appeals (Goel et al, 2017; Harrison et al, 2016b). This research emphasises qualitative analysis of a sample of phishing emails (Kim & Hyun Kim, 2013). Research in this field is more limited than research on explaining the victimisation to phishing messages.

The characteristics and language used in phishing emails has been a starting point for research into understanding the cues used within messages. Blythe, Petrie & Clark (2011), emphasises the use of logos, and grammatical errors that are often contained in phishing messages, 38% of their sample did not contain any grammar errors and 68% used company logos to appear legitimate. They consider that grammar alone cannot predict phishing messages from genuine emails as 38% did not contain these errors, furthermore in this sample 50% contained one or two typographic errors (Blythe et al, 2011). Given what was highlighted before from heuristic processing if someone is making quick inferences then unless the error is obvious, then a person could potentially not notice it. Furthermore, common techniques used by phishing websites and emails do not just rely on a psychological explanation, rather simple deceptive tactics are employed to fool users who are less knowledgeable around the structure of the internet. This is highlighted in a study of phishing websites by Dahmija, Tygar & Hearst (2006); showing common techniques used by phishers are to manipulate domain names by registering domains that appear close to the target domain, the use of SSL certificates, as well as using images to mask text of the phishing site. These highlight aspects in the actual message rather than how a person psychologically processes the characteristics of the message. Also, looking at phishing websites Jakobsson (2007) highlights, with URLs and security indicators, that the presence of Verisign and safe site symbols are trusted, even though these can be mimicked by phishers, as well as incorrect domains by replacing the .org belonging to the legitimate site with one run by the attackers at the .com domain. While these common message tricks can be applied by phishers to fool their victims into 'trusting' their website or email, it emphasises

phishing emails also use the actual elements in the message to lure the victims into believing its supposed origin.

Apart from logos other cues the language used is designed to encourage compliance with phishing emails. This is evidenced by cues in messages containing psychological prompts in the language. Parsons, Butavicius, Pattinson, Calic & McCormac (2015b) categorised the cues contained in phishing emails. They emphasise in this categorisation of message cues that certain prompts emphasise requests based on emotional factors by trying to emphasise consequences, though often illegitimate, or encourage urgency in the response. The prevalence and range of appeals that researchers have observed has been further elaborated by Kim and Hyun Kim (2013) who carried out a semantic analysis of phishing messages from the APWG website to analyse their persuasion methods. Interestingly, only 40% of the sample included a time-pressure cue for a response from the potential victim in the message and 70% included reasoning explanation as to why the message was sent to deceive the potential victim (Kim & Hyun Kim, 2013). Another aspect they highlight is that almost 95% of the sampled emails contain the argument of the message first, rather than later (Kim & Hyun Kim, 2013). Thus the structure of the messages themselves is used to both deceive the receiver of the email and also to emphasise the action they want the user to take. Further there was a high prevalence of credibility appeal with just over half the sample using company logos, to deceive the user into thinking the phishing message comes from a legitimate source. This further supports the findings of Parsons et al (2015a) as to their prevalence and potentially it encourages peripheral processing, suggesting that phishing emails focus on simple elements to make the user believe their legitimacy, such as the address bar as found by Vishwanath et al, (2011).

In phishing messages a range of methods is used by the phisher to garner and encourage compliance dependent upon the aim of the attacker. It has been highlighted that to encourage compliance with the messages most of the phishing emails that aim to steal data focus on providing detailed false information, to encourage people into thinking they are doing what people they know want or like them to do and appear to come from a person or organisation in authority (Ferreira, Coventry & Lenzini, 2015). Ferreira et al (2015), found that phishing emails that aimed to spread malware applied the same liking principle as in the data theft but instead of authority tended to distract the user by focussing the user on the actions they want

them to perform. Finally, they found for emails where the aim of the attacker is to commit fraud most commonly there are elements of commitment and reciprocity (the element of using personal connection for performing a favour) and the distraction technique (Ferreira et al, 2015).

Emotional drivers are used to encourage compliance with the message, through psychological factors. Ferreira et al (2015) emphasise how reward and loss and can be used to this effect through the principle of distraction technique. This was one of the more common techniques found for emails classified as data theft (such as help desk scams), which comprised 30 of the 52 emails analysed (Ferreira et al, 2015). This is best illustrated by the phishing emails that emphasise the loss of access to something of value to the victim. The prevalence of such emails is illustrated by Kim and Hyun Kim (2013) who found, over 75% of the sample of phishing emails, appealed to needs of safety (being protection from an undesirable event such as account suspension) and over 60% of the sample used fear based appeals. This finding relates closely to factors that have been considered in susceptibility studies around the role of the message. Harrison et al (2016b) emphasise this aspect in their study with the fear versus reward (other studies refer to this as loss/gain, see Goel et al, 2017) aspect highlighted in the message factors. Fear is the element that by not complying with the phisher's message the person will lose something they value, such as account access while reward is the premise behind the 'Nigerian' email scams (Goel et al, 2017 & Harrison et al, 2016b). Reinforcing the ideas of the Heuristic Systematic Model in that the message aims to encourage heuristic processing of the emails, through the appeals made to the victim directly or by mimicking the credibility of legitimate sources.

2.5 Methods in behavioural Phishing research

Methodology distinguishes the different research avenues in behavioural phishing research, depending on what the studies aimed to explore. Pattinson et al (2012, 20-21) consider there to be three broad types of behavioural or human focussed phishing research namely: "Real phishing experiments", "Phishing IQ tests" and "Scenario based experiments". This section of the literature review extends this classification of behavioural phishing research of Pattinson et al by not only considering phishing game research a subsection of the phishing IQ tests but also by recognising the phishing studies that code out a sample of phishing emails. Phishing

game based research broadly helps to improve the the games player's knowledge on how to better identify phishing emails by focussing on teaching strategies such as identifying fraudulent URL strings (Arachchilage et al, 2016 & Davinson & Silence, 2010). Coding based research can be considered an important avenue for exploration of this genre, as it highlights the tactics used by phishers to gain compliance. Ferreira et al (2015) note that understanding the strategies around social engineering that are used by those carrying out such attacks can lead to further identifying and developing ways to mitigate these techniques.

A further avenue of research sets out to use large sampling sizes to predict factors behind susceptibility to phishing attacks. This is performed by simulating actual phishing campaigns, usually spear phishing, in a university/corporate environment. Pattinson et al (2012) consider with this method (apart from difficulty in conducting it due to ethical concerns) does not show how people deal with a range of different phishing emails and they argue it is not easy to accurately measure for biases or other factors about each person. Though that being stated a range of studies using this technique this have aimed to uncover some of these factors through using surveys as well (Workman, 2008; Ngoqo & Flowerday, 2015 & Vishwanath et al, 2011). This research is quantitative and arguably aims to achieve more generalizable results by forming models explaining victimisation.

Other aspects of phishing research are less concerned with trying to replicate real-world conditions and instead unpack in greater detail the psychology behind falling victim to this type of attack. One area considers studies that try to test users' knowledge about phishing or look at ways to improve this knowledge such as phishing games. They tend to be concerned with testing a single or set of behavioural factors. This approach covers generally two methodologies being survey and experiment based research. These involve directly informing participants and asking how participants would react in the situation presented to them by the researchers; this research is criticised as having a subject expectancy effect type of cognitive bias present (Pattinson et al, 2012). Since the research on this is focussed on understanding how people react naturally one could extend this to cover research on how to retrain user's process such emails. The goal of phishing game research is to develop and evaluate, using experiments, solutions that help educate people to be less susceptible, through providing them with the knowledge, such as identifying phishing URLs, to successfully handle such emails if they encounter them in real life (Davinson & Silence, 2010 &

Arachchilage, Love & Beznosov, 2016). The objective is to either test perceptions or to reshape them so they can better identify phishing emails.

This research methodology for phishing tries to get elements of 'real world' phishing in highly controlled lab environments to gain rich insights or to isolate out an individual variable/factor in understanding susceptibility. Therefore, it only uses experiment methodologies to understand these factors. It involves not only controlling for variables but focussing on behaviour in the defined circumstances that the researchers wish to explore. Pattinson et al (2012) researchers use qualitative interviews during the experiment with participants performing role-play. Furthermore other tools are used for data collection during the experiment such as eye tracking as in Alsharnouby et al (Downs et al, 2006 & Alsharnouby et al, 2015). An issue with these studies is when to inform the participants. Downs et al (2006) did not inform participants when recruiting them, whereas Alsharnouby et al (2015) in studying phishing websites gave the participants a brief explanation of phishing and to imagine they visited the websites presented to them from a link in an email. Parsons, McCormac, Pattinson, Butavicius and Jerram (2015a) found in comparing the susceptibility between the informed and uninformed group that the informed group were better able to identify the genuine and phishing emails than the uninformed group, but the difference was not statistically significant. They argue that this confirms the expectancy effect/bias that previous studies believed but it is not as strong as previously thought to be (Parsons, et al, 2015). It reiterated the importance of visual presentation such as logos in making an email seem legitimate to participants and that also if the email targeted loss or promoted the chance to gain financially, such as survey prizes (Parsons, et al 2015; Harrison et al, 2015 & Goel et al, 2017). This suggests the importance of message cues and contextual factors as well as emphasising that more weight should be given to studies that have not informed their participants.

The coding based phishing studies, however observe real events or samples and code the responses out. This coding of a sample of phishing messages can be used either as the sole method or the first stage in the study methodology (Kim & Hyun Kim, 2013; Ferreira, 2015 & Blythe et al, 2011). This research focuses on analysing the messages or the behaviour of people around handling the messages, however unlike simulated phishing attacks or experiments the researchers are only observing and not controlling events. Valecha, Chen,

Hearth, Vishwanath, & Wang (2015) observed the behaviour of people while Kim and Hyun Kim (2013) used this method for the analysis of phishing messages. Both studies rely on the researchers coding observed 'real world' data rather than field-experiment/experimental design. Kim and Hyun Kim (2013) carried out a semantic analysis of phishing messages from the APWG website to analyse the persuasion methods used. This is a field of behavioural phishing research that should be considered for its ability to inform what are the more prevalent forms of phishing emails/messages, as has been well emphasised in the literature there are a wide variety of message factors that can be manipulated by phishers, as emphasised by Goel et al and Harrison et al.

2.6 Research issues

After having surveyed a range of behavioural phishing research and the limited amount of mobile research available. The following section outlines some of the potential limitations in the current body of literature and highlights areas for future research, of which some which will be addressed in the methodology section of this thesis.

2.6.1 Mobile research

Mobile in the phishing research has only been considered with regards to susceptibility with only one study considering the role of the device type (Vidas et al, 2013; Ngoqo & Flowerday, 2015 & Vishwanath 2016). This is a limitation and gap in the current body of literature. The available study illustrates that device type is not a statistically significant factor and that email habits explain any difference in susceptibility (Vishwanath, 2016). The limitation of this, which is arguably of greater concern to a practice audience is the proportion of users using mobile devices and then those that fall victim on this device type. Another limitation is that the mobile studies of Vidas et al (2013) and Ngoqo and Flowerday (2015), were field studies rather than scenario-based experiments. Established scenario based experiments allow for greater detail as to how people interact with phishing messages and the strategies they employ to determine the legitimacy of a message. As Iuga et al (2016) emphasise, many participants tried to open the landing page of the usability experiment on a mobile device and call for more research into mobile phishing. Replicating this study for mobile or that of Alsharnouby et al (2015) would allow for comparison of people's strategies around identifying phishing emails or websites with those utilised for desktop devices. This could provide additional inferences

to be made as to how people evaluate how message cues are presented due to the difference in form of the device (Harrison et al, 2015 & Iuga et al, 2016).

2.6.2 Phishing attacks and messages

Another concern with the current body of literature is around the messages used in the simulated phishing attacks and if they are representative of those sent by actual phishers. This section also concerns the manipulation message factors such as loss versus gain in phishing research. Firstly, there appears to be a gap in linking the coding based studies of phishing messages with the messages used in the simulated phishing attack research. This is arguably evidenced by the studies of Blythe et al (2011), Pattinson et al (2012) and Parsons et al (2015a) using actual phishing and genuine emails in the experiments or surveys, however Pattinson et al and Parsons et al only presented screenshots in the experiment not actual emails. Blythe et al (2011) used a sample of phishing emails that informed the survey data collection. Therefore, the potential issue underlying researcher collected and written emails, is that while they mimic what a researcher thinks are the techniques a phisher uses they may not be representative of typical phishing emails. There is variation in the success rates observed in simulated phishing attack studies, as can be seen by the results of Goel et al (2017) in attacks that were both contextualised but differed in the loss versus gain manipulation such as the bank card and the firewall, received 4.4% and less than 1% click rates on the link respectively. As compared to the most successful three messages which had click rates ranging from 20.6% to 37.3% (Goel et al, 2017).

A final concern and gap in the current literature is its focus on the user clicking the link as being equivalent to being successfully phished, this ignores the variety of exploit mechanisms. As Ferreira et al (2015) emphasise in that what they classified as data theft phishing emails (this category included help desk scams), comprised the bulk of the sample 30 out of 52 phishing emails. These phishing emails are focussed at trying to gather the credentials from vulnerable users. Luo et al (2013) results with spear phishing that of the 105 users targeted 35 clicked on the link to the phishing website, yet only 16 credentials were harvested. Assuming the majority of attacks are aimed at harvesting credentials, while this is risky behaviour that they clicked the link does not necessarily mean they were successfully victimised. Yet from a literature perspective how the user manages the phishing email is only

the preceding step to how they handle the phishing website. This was emphasised in the briefing given to participants in Alsharnouby et al (2015) experiment. While simulated attacks in research have emphasised links, few have considered attachments which could either redirect users to a site or be used to install malware (Vishwanath et al, 2016 & Vishwanath, 2016). While these are more limitations as compared to gaps, such as with mobile, in the current body of knowledge, they emphasises the relatively underexplored approaches in the current literature.

2.6.3 Generalisability of sample population

Another limitation of behavioural phishing research is its lack of emphasis on the generalisability of the sample population. This is potentially a concern as university students have been used by researchers who are noted to have higher levels of vulnerability to phishing attacks (Goel et al, 2017; Vishwanath et al, 2016 & Harrison et al, 2015). This therefore has the potential risk skewing the results since the sample is potentially not representative of susceptibility in the general adult population. As Coopamootoo and Groß (2016) note generalisability and representativeness of sample populations have been one of the major methodological concerns expressed by behavioural security researchers.

3. Methodology and Study One design

In considering what is susceptibility to phishing attacks in the target organisation used in this study two, it was deemed necessary to collect a sample of phishing messages that have passed through the organisation's email filters. The target organisation chosen (for its size and ease of access in getting permission to perform a simulated phishing exercise) is a large New Zealand university. To get this sample from the university's exchange server, spam emails were exported into a dummy account because this allowed for ease of viewing and running in a virtualised environment to avoid accidentally triggering any incidents. The aim behind this step was twofold: firstly to get an idea of the most common types of phishing emails people in the university receive and secondly the techniques used by the phishers. This step was guided by the studies of Ferreira et al (2015) and Blythe et al (2011) that looked for the most common types of phishing emails sent and analysed the data collected to determine the methods used by the phishers to deceive end users.

3.1 Coding scheme

A deductive coding schema was considered appropriate for this research as there has been a wide range of phishing studies conducted previously. Secondly this was considered appropriate this method is not highly original. Furthermore, the coding schema was adapted from that used for the rankings by the participants in Parsons et al (2015b). This coding schema was adapted by utilising the deception techniques looked at in Kim & Hyun Kim (2013). It did not use the deception techniques used in Ferreira et al (2015) as it was aiming at an overview of the techniques, so the emails sent were more representative rather than providing a detailed guide as to how to craft phishing emails that were most effective. This coding schema in Appendix C was adapted from the elements in Parsons et al (2015b) but also includes the Message appeals factors from Kim & Hyun Kim (2013). This adapted coding schema was considered appropriate for use as the literature provides an overview of the range of techniques used within phishing messages and how they can be used to garner compliance from victims. Finally, the adapted coding schema before being used for analysis was, validated by academics from other New Zealand universities and IT professionals in a different city in New Zealand, where it was well received and no changes were considered necessary.

3.2 Results

The sample of emails came from a 5-month period of spam emails for part of the 2017 year, this resulted in 160 spam emails in the folder of the dummy account, alan.turing@vuw.ac.nz. This sample of 160 spam emails contained 60 unique individual phishing emails, the other emails were marketing emails or 'Nigerian mail scams'. Only unique phishing emails were coded based upon the sender/ subject header and content of the email (For an example of how the coding schema was applied refer to Appendix C). Only unique emails were analysed because some of the emails appeared multiple times and were identical in every respect apart from the aspects specific to each targeted user, being the time-stamp on the email or the addresses it was sent to.

The clear majority of the emails in the phishing email sample were spear-phishing emails with only 11 containing attachments and the rest using hyperlinks or links embedded in images.

The bulk of these were broadly help desk scams, that threatened the user with loss of access to their university accounts unless they clicked the link and gave their credentials to either re-validate the account or receive an update, these accounted for 21 out of the 60 emails. From this sample of 60 emails 12 tried to target a gain appeal, in 10 this factor was not clear and 38 were loss emails. This was considered the principal differentiating factor as it is a more consistent measure in the results, than others such as contextualisation. In terms of the appeals used almost half used fear, 29 messages attempted to persuade the victim and just over 30 appealed to victim's need to protect themselves. Though while half of these messages tried to frighten the user into taking an action supposedly designed to protect them, it still emphasises that overwhelmingly this sample focussed on messages to encourage the user to comply, to avoid the threatened or inferred consequence. Only 10 messages in the 60 sample messages used any pressures for time or urgent responses.

16 of the 60 unique phishing emails examined were contextualised either to the organisation or the service targeted these were most likely spear phishing or credential harvesting focussed attacks and had a loss focus. 15 of the 60 unique emails used image buttons or embedded the link in an image the URL can be revealed by hovering over the image. Another strategy was using a hyperlink rather than a plain text link, of the eight unique emails targeting the university's Outlook/Office365 accounts all hid the URLs as either hyperlinks and one did this in an image (refer to Diagram A below). The other help desk scams were more generalised just targeting university credentials. Emphasising that most of the attacks show varying degrees of contextualisation, with the ones targeting the Outlook enterprise applications arguably the most contextualised.

Diagram A: Help desk scam

This content is unavailable

Please consult print version for access

Another interesting finding was observed in the top-level domain where the attacker tried to imitate the domain of the targeted service's or something related to that. This was more prevalent in the service phishing attacks, seeking payment for invoices (see diagram B). For instance the following attack targeting Spark and Xero where the sending email addresses are both to .nz top level domains that contain the company name in them. While it appears like it is using a plain text URL the actual company domain redirects to an entirely different website. This therefore takes advantage of the unmoderated domain space for .net.nz and .co.nz, it also is only easily detectable on a desktop version, where addresses are displayed as well as the from header.

Diagram B: Phishing email imitating Xero

This content is unavailable

Please consult print version for access

Interestingly the gain emails that appeared out of the sample of unique emails were ones promising Bitcoin's, though this only appeared 4 times in the 160 spam emails, yet only 3 emails promised receiving funds in actual currency. While the phishing links and attachments were not examined due to potential security concerns, only 10 of the 60 phishing messages had https in the links on the websites. The most popular companies targeted were Netflix, Apple and SAP with these emails all being loss focussed: with 6 emails targeting Netflix, 4 targeting SAP (claiming bills were overdue and had to be paid) and 4 for Apple IDs. Only one directly targeted banking being ASB bank (also a loss focussed email). What is interesting that though targeting work email addresses, the mail was likely targeting personal services.

3.3 Analysis and discussion of Study One

This part of the study aimed to find the major variables as defined by the literature that are exploitable in the phishing attacks in this institution. The aim behind this was that the emails and the variables of the messages in study two are more representative of those commonly found in the inboxes of University staff.

As highlighted above the overwhelming majority of this sample were 'loss' focussed and likely aimed at harvesting credentials for accounts, as they sought users to validate their details. These findings are largely consistent with the findings in Ferreira et al (2015) where around 57% of their sample were classified as trying to steal data, as in help desk scams. In this study the most common technique for this was to appear to be from an authoritative source, and either to give information in the emails (though likely false to appear more trustworthy) or to emphasise the costs of not performing the action (Ferreira et al, 2015). In this study, it can be seen with the helpdesk scams that they try to imitate the university's IT support service or Microsoft and threaten suspension of the account if the user does not comply, by providing their credentials. Compared to the message factors in Goel et al (2017) where the messages were contextualised, loss or gain focussed and the motive presented to the user. Our sample emphasised the loss aspect with 38 of 60 targeting this and some were contextualised to varying degrees to either the target organisation or by directly using the name of university's IT support unit or services). This may be explained by the fact that this is an enterprise environment then therefore potentially being skewed by the high prevalence of help desk scams targeting university email/ credentials. Whereas, Ferreira et al (2015) used their mailboxes plus online, a potentially wider sample.

The results of this study can also be framed around the emotions targeted in the messages that aim to elicit the desired response. Phishers emphasise credibility aspects and threaten loss. Therefore the language used in the above sample the prominence of fear as an emotional appeal being made and encouraging a user to protect themselves is consistent with the literature. Kim and Hyun Kim (2013) found that safety was used in around 78% of the sample and fear in almost 65%, whereas in our sample they were nowhere near as prevalent at face value. Considering the help-desk scams which are heavily present in this study's sample the language used tends to encourage the user to protect themselves out of the fear of the

negative consequence of not complying. The results differ with regards to the use of urgency which was far more prevalent in Kim and Hyun Kim (2013), where there were over double the proportion found in this study. This is even more interesting when compared to other studies since as Vishwanath et al (2011) highlight that urgency cues increase the likelihood of a person complying with the phisher's message. It therefore seems strange that in this study so few messages exploit urgency though as noted by Parsons et al (2015a, 2) messages that use this or "forceful" language tends to attract suspicion from end users.

An area that showed more consistency was in the use of visual cues in the messages. As emphasised in the literature by Harrison et al (2015) the presence of logos and other visual cues increase the susceptibility of individuals to phishing attacks. One would expect to see more visual cues and logos in the help-desk scams in this sample, as it purportedly has an impact upon perceptions. Interestingly there are some clear differences between aspects in the messages in the literature and actual phishing messages. Such as the links that were in plain text at the bottom of the email in Goel et al (2017)'s messages. In our sample they were all hyperlinks behind images or text. In this study just over a quarter had logos used in the design, this may have been biased to an extent by the over representation of the help desk scams in this study. This seems contradictory to the literature as Kim & Hyun Kim (2013)'s sample had almost 54% and Blythe et al (2011) 64%. Therefore the design of phishing emails from in sample is different to other samples in the literature. Therefore, as phishing emails sent in later rounds need to mimic legitimate emails, they would need to be adjusted to address visual presentation used in emails to mimic commonly used techniques.

4. Methodology round two: Study two

For the behavioural research component, this research carried out a simulated phishing attack as a field experiment. This because susceptibility to phishing has been validated and considered in past field-experiment studies which have informed the design of this stage of the study (Luo et al, 2013; Jansson & Von Solms, 2013 & Goel et al, 2017). It was also considered the appropriate method based upon the literature as the mobile based studies all included a field-experiment in their methodology, though some with other methodology as in the Vidas et al study on Qrishing (Vidas et al, 2013; Ngoqo & Flowerday, 2015 & Vishwanath, 2016). Arguably the study could be considered in a laboratory experiment but it would have

been logistically harder to get multiple rounds of data collection, which would have made it more difficult to determine how exposure to phishing impacts susceptibility.

Also, in the university there was the opportunity to perform a simulated attack across a valid sample by carrying out a test on a set group of professional staff, rather than academics or students. This allowed for a more generalizable result as well as ensuring greater homogeneity across the sample population but was done at the expense of the detailed insights that could have been obtained from a smaller sample in a lab experiment. The methodology used has two distinct benefits: it allows for testing on a sample population that is not university students, which as shown in Appendix A which is different to the majority of the current research, and secondly it allows for testing on a group where all had in prior months attended a general information security workshop conducted by an external security company. It was further considered that a lab experiment would have most likely have had mostly student participation, emphasising why a simulated attack across the target group of professionals in the university was thought to be a better sample group.

4.1 Study tool selection

The tool selection criteria were informed by the literature review and the technical concerns listed above primarily. This section sets out each criterion with why it was included:

- 1) Landing page: This section sets out that each tool needed to be able to at least mimic the Office365 login page or the University's Microsoft account login page for the spear phishing; similar to Luo et al (2013).
- 2) User-agent: This is an essential criterion as it needed to store the device-type and IP address of the devices visiting according to the research question and primary objective of this study. Secondly, we wanted to avoid having to use self-reported data as in (Vishwanath, 2016) on the device used, because the objective was to study spear-phishing.
- 3) URLs: For this study, the URLs need to be to an external landing page, but must appear as if it is an actual website. Customisability of the domain was a desirable but non-essential factor but it was essential it did not appear as an IP address, as this is easily identified as being illegitimate web address (Jakobsson, 2007).
- 4) Sending email address: As highlighted above as it is common knowledge in the university among staff and students and due to this address being displayed prominently in the header of the desktop email client, that it could not use an address from a common free mail service (such as Gmail, Outlook or iCloud mail).
- 5) Simplicity: This was considered a desirable factor since this is meant to replicate conducting phishing tests in an operational security environment. This is

considered desirable as phishing tests would only be conducted periodically so one would want to avoid configuring software on servers. Though this also included favouring solutions that are cheaper to implement in time and cost when run repeatedly.

To find the best selection against the proposed criteria the following scoring was used for each section. Each option considered was scored against the criteria making it easy to select the appropriate tool (for further details on each option see Appendix D). It was proposed that each criterion would be scored out of 10 to make computation and comparison simple. For aspects such as the landing page it was scored for its effectiveness in making a solution that would be easily repeatable and meet the above requirements. The following indicative scoring for each option was the result of this (for a discussion refer to part two of Appendix D):

Table 1: Tool selection scores

Option:	Landing page score	User-agent Score	URLs score	Sending Email Address score	Simplicity	Totals /50
Marketing tools	6	6	5 (through registering domains)	5 (through registering domains)	5	27
GoPhish	8	8	4 (criteria partially met as landing page was an IP Address)	6	6	32
Phish5	8	8	9	9	8	42
PhishingBox	8	9	9	9	9	44

It was considered that the open source and Marketing option were both let down in their simplicity, while they are financially cheaper options this was offset by the time that would be needed to configuring them for each test. Therefore a commercial tool was to be used. Zhou and Aleroud (2017) note that there are a variety of commercial tools designed that allow companies to launch regular simulated campaigns marketed as being for training employees. It was decided that this would allow for the most comprehensive option that would provide

more 'professional' looking results making the emails harder to detect using simple visual cues alone.

4.2 Study two method

This study set out to not only consider susceptibility between devices but to also highlight the role of the message and the impact of a repeated test. This used a two-stage field experiment as in Jansson & Von Solms (2013), whereby each group is served one phishing message in Stage One and then a different message in Stage Two of the experiment. Also it allowed for an easy way to test between a gain and loss targeted messages. This characteristic was chosen as emphasised in study one as it seemed to be the main distinguishing factor in our messages, plus this has a foundation in the literature from Goel et al (2017); Harrison et al (2016b) and Parsons et al (2015b). This is interesting due to the variety of results observed in the literature. Goel et al (2017) found that generally the 'gain' messages generally achieved high click through rates of around 20%, whereas the most successful message was a highly contextualised 'loss' email which achieved just over 37%. Harrison et al (2016b), like Goel et al (2017) who hypothesised the 'loss' or 'fear' message would be more likely to result in victimisation, found that even though the 'fear versus reward' factor did not affect the susceptibility, found the gain messages resulted in lower elaboration.

The 141 staff from the target/reference group were sorted into groups A or B, where the first group would receive the 'gain' focussed message and the second the 'loss' message. The list of staff in the targeted group had 193 listed email addresses. Two were eliminated as, one was a general email address and the other a test address. On review a further 50 had to be removed as they did not directly belong to the proposed group. The remaining 141 staff from the .txt were imported into a .csv file, where each person was assigned systematically to group A or B. Since this list was already alphabetical this minimised the likelihood of an entire team within the group receiving the same phishing email. Another precaution was for the launch of the messages to be staggered over a 48-hour period prior to the campaign to minimise the chance of arousing suspicion if co-workers simultaneously received identical messages. The field-experiment consisted of two stages, where each campaign lasted one week with a week in between the first and the second campaigns. This was to allow any heightened senses of awareness created in stage 1 to subside before stage 2. The message each group received was

assigned by the external cyber security professional strengthening the appearance of a random allocation of recipients to the message they received.

The study was set out as follows, with each campaign called stages:

Table 2: Experiment design

Stage/Message	Stage 1	Break	Stage 2
Gain Message	Group A 'VicBooks'		Group B 'VicBooks'
Loss Message	Group B 'Office365'		Group A 'Office365'

The messages were switched since the group assignments were in effect random allowing for comparison of how the message was handled at each stage. One can assume there were no extraneous variables between the groups to affect the comparison. Since the same principles in detecting phishing attacks apply regardless of the gain versus loss characteristic in the messages comparison could be made of each groups Stage 1 and Stage 2 performance. The study design was informed by the Jansson & von Solms study, and the message characteristic factor by Goel et al (2017) but this study only involved a single stage. Due to the limited sample population we could only use two messages rather than 8 and still be able to keep an acceptable number per group. Even though a previous study has the minimum acceptable sample size, of 90 according to Ngoqo & Flowerday (2015). The 71 in Group A and 70 in Group B were considered to be sufficient to make a valid inference since the ethics approval only allowed for the test upon the specific unit in the university that had 141 staff members. This research extends the Goel et al (2017) study by more than just considering mobile devices, in that it isolates the effect of the loss versus gain, where the message design was informed by the messages analysed in study one to ensure the messages sent to the reference group were close to what has been encountered by university staff in the past.

4.2.1 Message design

As highlighted the study proposed to consider and design messages that were both highly contextualised with a gain versus loss perspective from the Goel et al (2017) study while being consistent with what was seen in study one. To isolate the loss versus gain effect each message was highly contextualised being specific to the university and personalised to each receiver. For the loss message this was done by targeting Office365 which the university uses and having the vuw.ac.nz domain in the body of the message, whereas the gain message used

the lure of a free item from the university coffee shop. The messages and the message from study one that they were adapted from are shown in Appendix E. Each message was validated by an external cybersecurity professional before sending.

Diagram C: Emails sent in experiment (larger images refer to Appendix E)

<p>This content is unavailable</p> <p>Please consult print version for access</p>	
<u>'Gain' email free VicBooks coffee</u>	<u>'Loss' email Office365 suspension</u>

The decision for the loss was to adapt a help-desk style or issue with a popular service such as Netflix style phishing message, as seen in study one, being the most common ‘loss framed’ messages. Similar messages have been commonly used in the phishing messages in the literature being the loss of access or suspension of something important to the target, so in these studies it has been the ability to access to student systems or records (Goel et al, 2017; Harrison et al, 2015; Harrison et al, 2016b & Vishwanath et al 2016). Therefore, in keeping with study one the Help Desk scam was more appropriate than the Netflix account as this is an enterprise environment, it could be safely assumed not many would use a work email address to sign up for an entertainment service. One of the account suspension emails from study one was adapted, by correcting the grammar and adding a logo, this was then externally validated with the wording changed slightly in validation.

The gain based was aimed at a more realistic offering than had been shown in the literature and in study one. Firstly, the gain emails in study one offered ranging from a bitcoin to \$75 dollars on Kogan. Goel et al’s messages (2017) promised similarly highly valuable items such as an iPad resulting in a 19.7% click through, and a \$50 gift card for completing a survey that

achieved 21.4%. Originally, the plan was to replicate the Kogan message for study two, however in validation it was recommended to use a more realistic promise, similar to ones that the expert had used with corporate clients. The message was altered to promising a coffee form a university coffee shop.

4.2.2 Human Ethics

In conducting this study, the ethics was an important aspect due to the strong initial element of deception to avoid the effects of an expectancy bias. Parsons et al (2015a) confirm the presence in their experiment of an expectancy effect, where the informed group performed better than the uninformed group in the lab experiment. So therefore in this study there was no informed consent and it did not provide people with the opportunity to withdraw from the study. Vishwnath et al (2011); Goel et al (2017) and Pattinson et al (2012) emphasised the ethical concerns around the lack of informed consent in simulated phishing attack research.

To mitigate this in the Human Ethics Approval (Ethics id: 25045), everyone would be sent a debriefing email at the end of the exercise, which explained its purpose (Appendix F). Furthermore, this study had clearance and support from high levels in the University's IT service. The data would be kept confidential in line with past studies only aggregated data would be presented and the results would not be used to discipline any employees based upon the results. To minimise privacy and security concerns the tool selected did not capture user submitted data, and the login page in the Office365 phishing test the "sign-in" button went blank if clicked. This study follows the procedure in Jansson & Von Solms (2013), Goel et al (2017) and Bullée et al (2015), making it consistent with prior studies of this nature.

4.3 Hypothesis development

The following hypotheses were made from the method above and the literature, for testing the variables that may affect the dependent variable being the susceptibility. The dependent variable was measured as the percentage of click throughs and the proportion that fell victim to each attack. Jansson & Von Solms (2013) did not test any hypotheses in their field experiment and Ngoqo and Flowerday also did not as it was interpretivist research. This research tests a hypothesis for each section that is also falsifiable as set out in Coopamotoo & Groß (2016).

4.3.1 'Loss versus gain'

Since the effect of a 'loss' versus 'gain' appeals in messages remains unclear as shown above. This study considered based upon the findings in Goel et al (2017) and Harrison et al (2016b) that the gain message will be more effective than the loss one. Though, these studies both hypothesised the opposite and Parsons et al (2015a) found that loss focussed messages were more likely to succeed at victimising people.

Therefore, the hypotheses are:

H1: People are more susceptible to phishing messages that promise a 'gain' to the potential victim (end user) than ones that pose a 'loss' if the person does not comply.

H1₀: There will be no difference in the susceptibility of end-users to both the 'loss' and 'gain' focussed phishing messages

4.3.2 The effect of training

This hypothesis aims to test whether exposure to training positively impacts upon the susceptibility rates of the end-users. Whereby as people are exposed to messages more often they become less susceptible to phishing messages. So, we expect more people to fall victim in Stage 1 than Stage 2.

H2: The overall susceptibility of the victims for each group will decrease in stage 2 as compared to Stage 1

H2₀: There will be no difference in the susceptibility between each group between Stage 1 and Stage 2

4.3.3 Device Type

This hypothesis explores the primary motivation of the study being to test the effect of the using a mobile device versus the desktop client on susceptibility. Due to the lack of literature

in this area it will test the assumptions of Iuga et al (2016) and Jansson & Von Solms (2013) that the mobile user are more susceptible than the desktop users. The null hypothesis for this will be that of Vishwanath (2016) where there is no effect of the device on the susceptibility.

H3: The proportion of end users that fall victim on mobile devices will be greater than for desktop across the two groups

H3₀: There will be no difference in the susceptibility rates between the mobile device users and the desktop device users.

4.4 Results of Study two

The 'gain' focussed phishing email triggered a Cyber Incident during the 48 hours of staggered sending yet managed to achieve a significantly higher response rate than the loss manipulation. The exercise was live for 72 hours until both emails were then blacklisted ending the experiment. The decision to call off the data collection during the campaign was because both messages had been reported to the University as suspicious during this time, therefore the reference group were now suspicious of both messages and so this unit inside the university would have an expectancy bias in any subsequent rounds of data collection, meaning that there would likely be significantly lower rates of victimisation. If the 'gain' message (free VicBooks coffee) had not been reported as malicious, it would have been suspicious that the job request reporting it had not been closed and it was still in the inboxes of those who received it. Furthermore, as Group B was in that same reference group used they would have been suspicious of the message had it arrived again in stage 2. Despite this both messages achieved reasonable response rates in this shortened period of data collection, with 44 out of the 71 having opened the email in Group A (62%) and 32 out of 70 for Group B (46%), as shown in Appendix G (for a sample of the raw data). Though IP Addresses (version 4) were collected by the tool, human ethics restricted us by not allowing us to use GeoIP databases meaning we could not analyse where in the city the traffic came from, as in were people at or near the university when they fell victim. GeoIP databases were not used in case it revealed the areas in which people live, though it is not usually highly accurate.

The susceptibility rates also varied significantly between the groups as can be expected with the 'gain' focussed mail having achieved a better response rate. For both tests a 'fail' was defined as clicking the link and a further fail was having preformed the action on that page. The failure rates were as follows : Group A 28.2% fail and Group B 7.1%. For the two groups in this test the following pass/fail and susceptibility rates were observed:

Table 3: Study two results (aggregate)

	VicBooks "Gain appeal"	Office365 "Loss Appeal"
Number targeted	71	70
Number Open	44	32
Open Proportion	62%	46%
Number click link	16	3
Link Proportion	22.5%	4.3%
Number action Preformed	4	2
Action Proportion	5.6%	2.9%

For Group A, there was an embedded training exercise for those who fell victim. If the exercise was not stopped due to the cyber incident, in a second round this would have allowed for testing the awareness as it would have taken place a week after the end of the first campaign. In the design of the experiment it was not known if there was an embedded training exercise in the configuration. Where Group A would have been exposed to embedded training unlike Group B, especially as 17 of the 20 who failed viewed the training module. While this is a good uptake in the training, we only know the number of times and device the person viewed this page on not if they read it. The highest risk users in traditional spear phishing, are those that carry out the action on the page. This was less than 6% for both groups. While it is a small percentage, it does however mean that the attack would have been successful several times in the short time that it ran, potentially other cyber-attacks could be launched before there appropriate responses could be carried out. Clicking on malicious pages could be risky from a vulnerability perspective emphasising concern over the failure rates for the gain message (VicBooks) in particular.

The results highlight that the notion of mobile versus desktop susceptibility, in seems to be a simplification of the issue as device interaction is more mixed. A range of Operating Systems were observed being in use with iPhones and iPads for mobile (both iOS) and for desktop users

Windows and Linux were the most popular followed by Mac OSX, therefore the device type was separated by operating system. The iPad was not treated differently to the iPhone not only due to the operating system but the limited numbers in the sample, it was regarded as an iPhone with a larger screen. This complexity is seen in Group A with 2 viewing on an iPhone and then a windows device, another viewing on Windows before an iPad and 2 viewing on the iPhone but 'handing-off' to a desktop device (either Mac or Windows). Interestingly Group B was less complex, 4 of the 33 who opened the email viewed it on both a mobile and then a desktop, with no one handing off the link to another device. Suggesting that in an enterprise environment there is a complex interplay between devices, especially with those in Group A who viewed on the one device but opened the link from a desktop device. All of those in Group A who used both devices for viewing clicked the link on a Windows desktop machine, whereas in Group B only one failed who used both device types to view. Seemingly with these dual device users there is no distinguishable correlation as to their susceptibility, evidenced by the difference and limited sample size of both groups. A reasonable minority of participants used both device types, 13.6% (of the 44 who opened the email) in Group A and 12.1% in Group B (32 opened the emails).

This complexity in the devices used to manage these emails also translates across into the susceptibility rates between mobile and desktop within these groups. In Group A 11 of the 44 (25%) who opened the email initially on a mobile device and 15 of the 32 (47%) in Group B. Unexpectedly there is a difference in the device type used to initially view the email between the groups given both campaigns were launched simultaneously. In terms of susceptibility of those who originally opened on a mobile device 7 failed in Group A versus 1 in Group B, though no one preformed the extra action. Of those who originally viewed on mobile in Group A (63.6%) failed whereas in Group B it was (6.7%). At face value the results of those who opened the email on a mobile device in each group and the proportion of those who then fell victim clearly differs. The wide variability is likely due to the limited sample size as the original size of each group in the group targeted within the university was small. This still indicates the above point of complexity in understanding the role of device type in phishing. When comparing the desktop susceptibility rates as determined by the Operating System due to the limited sample Mac OSX, Linux and Windows users were classified as desktop, like the iPhone and iPad were classified as mobile. We can assume as this is an enterprise environment there is a reasonable amount of similarity between the desktop devices. Of those who used the

desktop as their initial viewing device 33 (80.4% opened on a desktop Operating system) in Group A, of which 13 failed (39.3% of desktop failed) compared to 17 (53.3%) in Group B where only 4 failed (23.5% of desktop users). The susceptibility was filtered by the first Operating system of the five potentially used for viewing the email, this was due to the complexity of interactions, where people opened the email on multiple devices or even operating systems. This assumes in their behaviour that if they thought it was illegitimate at first viewing they would mark the email as spam or delete it on the initial viewing device. Therefore, it was on the initial device that they decided whether to action the email. In these statistics we again observe a variability between the groups which is most likely due to the limited sample size where $n=32$ for Group B, which is the smallest. We can assume the viewing device difference between the messages is essentially random, as the targets unwittingly classified themselves.

There was greater consistency in the results between groups for the number of times the email was opened and average number of times people clicked the link. This is shown by the number times the message was viewed on each operating system, with only five in each group viewing it on a second operating system. For Group A the average number of times the message was viewed on the first device was 2 with a median of 2, for the five people who viewed on a second device this was 2.4 and respectively. However since only five people in this group viewed on a second device this confirmed the use of the initial viewing device for calculating the devices used for the susceptibility calculations. For Group B the median was 2 with a mean of 3.875 and for the five who viewed on a second device this was 11 and 9 respectively. This indicates that people normally view around the same number of times on the first device, however for both groups this was skewed by a few people who opened the email numerous times, which introduced the variability. For clicking the link a similar pattern can be seen. In Group A people clicked the link with a median of 1 time but the mean across the group was 3.5 yet in group B the median was 2 with an average of 3. This can be assumed to be roughly similar for both but skewed by the presence of one or two extreme values. Group B is more difficult to draw an inference from for failure due to the smaller rate of approximately 7% as compared to approximately 28%. Since Group A had training this was more predictably viewed only once, with a median of 1 and an average of 1.5. Potentially emphasising for this group more consistency with in the times people clicked the link in comparison to the median.

Though surprisingly similar results can be seen in how many times people in the reference group reported the message as spam. Over the period of the one working day that the test spanned 11 people reported the Office365 message as a spam message to the University compared to 15 for VicBooks, in fact one person contacted the café in question and in approximately 90 minutes from the first VicBooks email being reported a job was established in the Systems queue of the university to permanently block the message. Furthermore, by 5 hours after the first messages on that day both messages had been traced back. Providing an insight into the reaction times of the staff in responding to a perceived phishing attack. With the VicBooks spam message 15 minutes after the job had been opened, it was reported as a potential phishing-attack. After this all parties involved in this research had decided that no second test should be carried out on this group as the reference group were aware of both messages. An unintended consequence of the simulated phishing test is that it allowed for testing of the responsiveness of the reference group as a sample of the professional staff to phishing attacks. Interestingly 15.7% of people in Group B (70 people) and just over 21% had in Group A reported the message as spam. For the loss message (Group B) a greater proportion of people reported it as spam than who received a fail result and interestingly this is roughly a comparable proportion, to the amount victimised in the gain message (Group A). However, no cross referencing, to check if people who fell victim had reported, was carried out in this analysis because the focus of this study is on susceptibility not how it was managed and we do not know how many in their individual mailboxes marked it as spam.

4.5 Analysis of study two

In testing the hypotheses above that formed the basis for the experiments for this section only the hypotheses related to Gain versus loss message and device type could be tested. Meaning that Hypothesis two above testing the effect of exposure to simulated attacks could not be assessed. This is due to there being no data from a second stage to make the comparison for exposure. The, hypotheses were tested based upon the criteria in Coopamootoo and Groß (2016) around behavioural security research whereby there must be clearly falsifiable hypotheses, through considering the p-value. In determining the threshold for the significance level, this study applied a 95% confidence interval as per the examples in Coopamootoo & Groß (2016). For testing both hypotheses it was considered to use a chi-

squared test as the data was categorical and the groups were independent (Kraska-Miller, 2013). The test allows for testing of the significance between such groups and is also used for testing within a sample population (Kraska-Miller, 2013 & Sauro & Lewis, 2012). It was considered that a Chi-Square test with a large sample size was adequate for this rather than an n-1 Chi Square as recommended by Sauro and Lewis (2012) due to each group having a sample size greater than the n=30 threshold, with Group B being the smallest at 32. Though, as Sauro & Lewis (2012) highlight as compared to an n-1 two-proportion or n-1 chi square (which are mathematically alike tests) this does not underestimate probability in smaller sample sizes. It was deemed for this that the test samples were sufficiently large even in applying the cell check method for the first test (Sauro & Lewis, 2012). The two tests were conducted as follows using the chi-square test is JASP Statistics (version 0.8.5.1).

4.5.1 Hypothesis One

The first test sought to compare the differences in susceptibility between the two messages. This involved comparing how the message was handled by the participants in each group. As mentioned before Group A was assigned the gain message and Group B the loss message by the external cybersecurity professional. Therefore the chi-square tests allows for comparison of the overall susceptibility between the groups. For testing this the full samples for each group was used and initially compared against the pass versus fail rate between these groups. Allowing for comparison of those who clicked the link against those who did not in the two groups. The following results were observed for this chi-square test:

Table 4: Contingency and Chi-square test Hypothesis 1

Contingency Tables

Group	Pass/Fail		Total
	Failed	Pass	
a	20	51	71
b	5	65	70
Total	25	116	141

Contingency Tables			
Group	Pass/Fail		Total
	Failed	Pass	
Chi-Squared Tests			
	Value	df	p
X ²	10.68	1	0.001
N	141		
Nominal			
	Value		
Phi-coefficient	0.275		
Cramer's V	0.275		

With respect to our null hypothesis that there would be no statistical difference between the susceptibility of the gain group as compared to the loss group, our threshold would be that the P value is greater than 0.05 (95% confidence level). As shown above we fail to accept the null hypothesis thereby lending support to the hypothesis that gain messages are more effective in terms of susceptibility than loss ones, as our P value of 0.001 is less than the 0.05 threshold. This supports H1 in that there is higher susceptibility to gain focussed messages. However the Cramer's V indicates that there is a slight degree of dependence between these two variables. As compared to Cramer's V interpretation guide in (Kraska-Miller, 2013) values of 0.3 shows a weak association between the variables. Thus the value of 0.275 shows a reasonably weak dependency between the two variables in Hypothesis 1.

It was decided to compare the handling of the message to see if this still means we fail to accept the null hypothesis. For this test due the very small numbers that 'performed the action' and the difference in these actions was consider to differ significantly between the messages. For Group A this was refreshing the page and in Group B clicking the 'sign-in' button. This results in a 2x3 chi-square test:

Table 5: Chi-square Hypothesis 1 (by action)

Contingency Tables

		Email status			Total
		Clicked Link in Email	Email Opened	No Status	
a	Count	20.00	24.00	27.00	71.00
	Expected count	12.59	25.68	32.73	71.00
b	Count	5.00	27.00	38.00	70.00
	Expected count	12.41	25.32	32.27	70.00
Total	Count	25.00	51.00	65.00	141.00
	Expected count	25.00	51.00	65.00	141.00

Chi-Squared Tests

	Value	df	p
χ^2	11.03	2	0.004
Likelihood ratio	11.68	2	0.003
N	141		

Nominal

	Value
Phi-coefficient	NaN ^a
Cramer's V	0.280

^a Value could not be calculated - At least one row or column contains all zeros

Comparing how the gain message in Group A and the loss message in Group B was handled, there is a significant difference between the groups. This is evidenced by a P value of 0.004 which like the test above is below the 0.05 threshold. This provides evidence that the participants in Group A who received the gain message were more likely to click the link and viewed the email than that in Group B (loss message). Though the Cramer's V value for this test was slightly higher than in the result before, it still shows the same result of a very weak dependency between the how the email was managed and the message type.

4.5.3 Hypothesis 3

To test hypothesis three, concerning the impact of the device type, the devices were categorised based upon the operating system. This was done as in the section above where iPhone and iPad were classified as mobile and Mac, Linux and Windows as desktops. Furthermore it was considered that due to the constrained size of each group in this study this was easier than carrying out a test to see if users of a particular operating system were more susceptible, especially given the very limited number of Mac users as well. As mentioned in the results section we considered the initial viewing device, making the assumption that it was

on this device that the user made the evaluation as to the email's legitimacy for the message they were exposed to.

For conducting comparisons with this group we had to know the operating system so those responses that were recorded as 'No Status', which means the participants that did not open the email were eliminated from each group. There is no knowledge of operating systems for people in the 'No Status' responses and null values would skew the results in the test. Overall the test considers susceptibility between mobile versus desktop against the pass/fail metric. Since using 'Email Status' would be equivalent to the pass/fail totals as the 'No Status' results were removed. The chi-square test between mobile versus desktop, was sorted between each group as well as the impact of the message type, which gave the following results:

Table 6: Contingency and chi-square test Hypothesis 3

Contingency Tables				
Group	Pass/Fail	Viewing Device		Total
		Desktop	Mobile	
a	Failed	13	7	20
	Pass	20	4	24
	Total	33	11	44
b	Failed	4	1	5
	Pass	13	14	27
	Total	17	15	32
Total	Failed	17	8	25
	Pass	33	18	51
	Total	50	26	76

Chi-Squared Tests				
Group		Value	df	p
a	X ²	1.956	1	0.162
	X ²			
	continuity correction	1.100	1	0.294
	Likelihood ratio	1.961	1	0.161
	N	44		
b	X ²	1.719	1	0.190
	X ²			
	continuity correction	0.678	1	0.410
	Likelihood ratio	1.839	1	0.175
	N	32		
Total	X ²	0.081	1	0.776
	X ²			
	continuity correction	7.336e - 4	1	0.978
	Likelihood ratio	0.081	1	0.775
	N	76		

Nominal		
Group		Value
a	Phi-coefficient	0.211
	Cramer's V	0.211
b	Phi-coefficient	0.232
	Cramer's V	0.232
Total	Phi-coefficient	0.033
	Cramer's V	0.033

The results emphasise that there is no distinguishable effect between the two variables. As evidenced by the p value from the Chi-Square tests for each group and overall. The null hypothesis being tested was that there was no difference between the mobile and desktop users on susceptibility regardless of message. Therefore we fail to reject the null hypothesis for each group and overall as none of the p values are below the 0.05 threshold (95% confidence interval). Meaning that there is no significant statistical difference between the rates of mobile and desktop susceptibility not only overall but also within each group, being the gain and the loss message. As highlighted by Kraska-Miller (2013) that if the results mean that the null hypothesis is rejected, meaning no statistically significant difference, then the Cramer's V result can be disregarded. Though, this would have happened anyway, even if the null hypothesis was rejected as the overall Cramer's V is close to zero at 0.33. It can be considered that there is no clear statistical effect for the device type between mobile and desktop, and the susceptibility to the phishing attacks.

5. Discussion

This research primarily offers two major implications from its findings, it considers the effect of the message and the complexity of devices types on susceptibility to phishing attacks. Since this study employed a simulated phishing attack, it fits according to the taxonomy of responses to phishing as presented by Aleroud & Zhou (2017, 171) as a study around awareness and within "Training and Educating users". This is because this study focussed on the susceptibility of users to phishing and the use of simulated attacks to understand users' responses. However, the educating component could not be empirically tested as the second round of data collection in Study two was not able to be completed.

This study instead emphasises the behaviour of professional administrative staff who have recently received training in information security (including phishing). This is different from the majority of past studies that used university students as the test group (Jansson & Von Solms, 2013; Goel et al, 2017; Harrison et al, 2016b). It is more comparable in its method and generalizability to Luo et al (2013) who used one message across staff and faculty in a United States university and to a lesser extent Caputo et al, 2014. To the best knowledge of the author only one previous study for mobile devices being Vishwanath (2016) considered both mobile and desktop devices. With respect to this we considered susceptibility but provided more detailed understanding of people's behaviour around this factor. We also further aimed to

consider the impact of the message factors on susceptibility by considering the 'gain versus loss' of Goel et al (2017) and 'fear versus reward' aspect from Harrison et al (2016b), labelled as the former in this study.

The results of this study should be considered in the context of the other simulated phishing studies that explored susceptibility or victimisation to phishing, before putting it in context of our limited insights into mobile susceptibility. It is considered by the author that the results for susceptibility we achieved in terms of proportions are similar, at face value, to other studies. In comparison to Goel et al (2017) which is most similar to this research in terms of messages used. They achieved with two of their gain messages promising a \$50 gift card or an iPad both achieved, click through rates on the link of 21.4% and 19.7% respectively (Goel et al, 2017). These are perhaps the most similar to Group A in this study that had 28.2%, which is roughly similar with potentially the slightly higher rate in this being explained by the offer seeming more likely as it promised a free coffee or alternatively that our message was contextualised unlike those gain emails in Goel et al (2017). The loss based email was also equivalent to previous studies, though not equivalent to Goel et al (2017) where the loss emails ranged from 37.3% for the course registration to around less than 1% for the credit card attack. In comparison to similar messages in Jansson & Von Solms (2013), being the 'virus scan' message and 'Database crash' messages are the most similar to our 'loss' one trying to get the victim's university Office365 credentials, by asking them to revalidate their account. We managed to achieve a click through rate on the link of 7.1% on the link compared to 9.07% and 8.78% respectively for the above similar phishing messages in Jansson and VonSolms (2013). Suggesting that our loss results while not comparable to Goel et al (2017) are comparable to those in Jansson and Von Solms (2013), both of which were instrumental in guiding the design of this study. Our results while consistent with the those in some of these studies are less than for others, such as 60% in the first test in Caputo et al (2014) and 47% in Harrison et al (2016b). The results from this research still provide enough evidence that our susceptibility findings can be seen as reasonably consistent with prior research for both message manipulations.

While not being able to perform a second round is a clear limitation. The study yielded interesting results around mobile and desktop as well as the message type. It also highlighted other factors to consider in the design of simulated phishing experiments especially in

professional environments, as discussed later. The concern over this limitation was mitigated by that Luo et al (2013) also did not use quantitative post measures (due to time-frame in their case), and achieved 38 out of 105 faculty and staff in a single test having submitted credentials. Therefore indicating the value in a simulated exercise and also the short life of phishing campaigns can have in research environments. This thesis is also more consistent with this in that they conducted it across a single uniform sample group. In this thesis rather than staff and faculty it was 141 professional administrative staff, though a potentially interesting factor in this study when compared to the studies considered above was that our group had received prior awareness training that included phishing, including the technique to successfully mitigate both messages. Despite this and our consistency with prior research, this does not explain our variance between the 'gain' group and 'loss' group to contextualised phishing emails.

The comparison of our two groups being Group A with the 'gain' and Group B the 'loss' reveals interesting results compared to other studies. Though, it does reinforce the validity of the above study compared to others, it helps demonstrate that the above findings are reasonably consistent with the prior literature. In this it is worth considering the consistency of our gain email (VicBooks coffee) being more effective than the loss one (Office365). This is in part consistent with some of the findings from Goel et al (2017) and Harrison et al (2016b) studies. It is worth noting that the 'gain versus loss' or 'fear versus reward' message appears to be the best knowledge of the author a limited research stream with unclear results in these few studies.

While 'gain versus loss' and contextualisation can be considered important factors when comparing study results, potentially it is also influenced by how the former characteristic is presented/interpreted by the potential victim. Despite Harrison et al (2016b) finding overall there was no statistically significant difference between the 'reward' and the 'fear' message in predicting the processing or susceptibility to phishing attacks. However, they found that there is evidence to suggest that where potential victims consider there is a chance to incur a financial benefit they showed less elaboration on the message (Harrison et al, 2016). Confirming Goel et al (2017) being evidence suggesting that the 'gain' messages are more effective than 'loss' based ones. Goel et al (2017) also had unclear results founding that highly contextualised loss emails were the most effective, whereas the others interpreted in this way

were not and received low rates for clicking on the link. They note that the 'gain' messages used had very little contextualisation yet achieved reasonably high rates of clicking the link at approximately 20% (Goel et al, 2017). Suggesting validation of the 'gain' message in this study that differs from their study. At face value the VicBooks message received higher results at closer to 30% and in the reward as compared to Goel et al (2017)'s contextualised gain message. Though the contextualised gain message did not promise free goods but free firewall/virus software (Goel et al, 2017). The results from this study are in clear contrast to the findings around loss focussed email messages. Furthermore, despite the 'loss' email in this study being contextualised it did not receive a high susceptibility rate like Goel et al's (2017) 'course registration' email that had 37.3%, whereas ours had 7.1%. Potentially suggesting that 'gain' messages regardless of contextualisation will generally but not always be more successful than loss messages. The relevance of the outcome to the individual may yield more explanatory power in that it depends on the type of offer. From these results it seems that the voucher, iPad and in this study the coffee may have had a more relevant value, meaning if this assumption above is true then it would confirm the finding around 'rewards'/gains of Harrison et al (2016b).

Another interesting insight of this study was around the genre and familiarity to the two messages sent. This highlights an alternative explanation to the message factor, it was found to be a another potential explanation for the variance between the gain and loss groups (Group A and B respectively) based upon considering the timings of when the messages were reported as spam in the university's service desk tools. This explanation however does not have any definitive way for comparison as no follow up measures were conducted after the field-experiment, such as a survey as in Goel et al (2017) to check how the message was interpreted. It could be considered that the 'gain' message (VicBooks coffee voucher) was a variation upon a classic phishing email, as it was not directly adapted from one of the common messages in Study One, whereas the 'loss' one (Office365 validation) was built directly from a typical service desk phishing message. Thus, potentially familiarity with the message would have been an indicator to many that it was a phishing attack not a legitimate email. Thereby implying the VicBooks would have been seen by those who initially saw it as a marketing rewards email, enhancing the belief of earning a free coffee or that it appeared to be from a university café was enough to encourage them to click the link.

This explanation potentially relates to the familiarity of the messages in this study to the participants. So for the 'gain' focussed email the hyperlink was to work-rewards.com, with the reply address being to VicBooks@ the above domain; whereas the Office365 was potentially more obvious with msupdate.net and the reply to field on that domain as msonlineserviceteam. Assuming they did not remember or try to apply the training, this would appear doubtful as per the findings of Alsharnouby et al (2015) and Downs et al (2006) on the elements explored in determining phishing sites and emails. If the participants checked the domain of the link, then in the case of the Microsoft message, MS not Microsoft could be assumed to be an obvious imitation and thus a phishing message whereas with the VicBooks message, one might have assumed that the café in question was using a marketing service to send the voucher. Both emails contained logos, so this factor can be safely ignored as a driver. Meaning either familiarity with the message or the interpretation of the domain would be the driver of their interpretation relating it back to familiarity considerations in the prior research rather than the message factors, placing the emphasis on human predictors of susceptibility rather than the message.

The literature underscores that potential victims who are less experienced or familiar with the techniques used by phishers are more susceptible (Harrison et al, 2016b & Downs et al, 2006). Suggesting that since the VicBooks email was not a traditional message and the plausibility of its domain it could have potentially fooled even more savvy users. It also seems unlikely that these decisions would have been based entirely upon familiarity with the source. Since as Moody et al (2011) found perceptions of familiarity with the sender is associated with greater susceptibility to the phishing attack. Meaning that when considering the source, one can assume it would have been based upon other factors in the message above as both organisations the phishing messages would be well-known to the users. This provides an explanation, albeit like the others an unverified one, as to the clear difference in susceptibility between the two groups. While there are multiple explanations behind this large difference in susceptibility of the groups, potentially no theoretical explanation should exclude both the human behavioural element and the message factors in explaining the susceptibility. Explaining susceptibility was not the primary focus of this study as it was originally concerned with the overall rates of this between mobile and desktop email clients.

Against the current mobile literature this study reveals extra insights due to the tool used collecting the device agent each time the email was opened, the link clicked or action was preformed but it is difficult to generalise from this result. Our results appear consistent with the closest study around mobile versus desktop as a factor in susceptibility to phishing. Vishwanath (2016) found that the type of device had no impact upon susceptibility to phishing and instead consider that a user's habits is a better predictor of understanding susceptibility. Instead it is considered in this study that smartphone and heavy mobile usage increase habituation, that is considered by the author to be a better predictor of susceptibility (Vishwanath, 2016). This study also suggest a similar result while providing extra insights into people's viewing behaviours on different devices. This study confirms that neither overall or for the sub groups did the mobile phone show greater susceptibility, since none of the values in the Chi-Square tests were below the threshold p-value of less than 0.05 (as per the testing in hypothesis 3). This confirms the findings on device type, though where our study provides further interest for future exploration is in considering those who viewed the emails multiple times and especially across multiple devices, as seen in the raw data (Appendix G). This may not necessarily mean that it will predict susceptibility but highlights how people interact with an email, The data shows that it is not as simple as engaging or deleting the message on a single device since a proportion of users as viewed or clicked the link on different devices. This shows how people interact with emails between devices and also relates back to the research on perceptions of security on devices. Exploring this would potentially mean linking these perceptions back to the findings of Chin et al (2012) being no difference in perceptions around security in conducting emails on mobile versus desktops computers and the belief measures around the riskiness of actions in Vishwanath et al (2016). Highlighting that for studies that want to go beyond the scope of this research in explaining how device type relates to phishing, should consider the perceptions of devices in assessing why people would interact with the same phishing email across multiple devices.

Unfortunately we could not evaluate in this context the impact of awareness training upon behaviour, as in Jansson and Von Solms (2013) or Ngoqo and Flowerday (2015) due to the limited time frame that the exercise was live for and no second round of data collection was possible. The campaign period of one week might have been too long given the lifespan of phishing websites. Aleroud and Zhou (2017) note as the average uptime for phishing attacks between 2008 and 2013, ranges from just under 23 to 72 hours. The initial time frame above

was used as it is consistent with the literature, however because of this a second round could not be conducted with those emails as they had achieved a potential 'expectancy effect', as emphasised in Parsons et al (2015a). This research emphasises how proximity can potentially have a beneficial impact upon awareness training in organisations, which the studies considered in the literature such as Jansson and Von Solms (2013) could not do due to the broad and large sample populations. Therefore instead of organisations trying to reduce susceptibility they should emphasise resiliency when conducting and evaluating simulated phishing exercises.

Finally considering awareness literature, this study can be seen as providing support for it, as a part of building and understanding resiliency rather than immunity to phishing attacks. This is evidenced by the results from the loss message where 7.1% 'failed', in that they at least clicked the link. The terminology used in the tool may not be the most research appropriate. The result still demonstrates that even with a potentially very familiar phishing attack, there will still be those that fall victim, thereby exposing an enterprise environment to further malicious risks. Even in the Goel et al (2017) study where the least successful phishing emails never had a zero percent rate of susceptibility, though 2 emails did have a less than 1% rate. Awareness implies that users with the adequate training and knowledge would never fall victim. As Vishwanath (2011) emphasise that awareness may have limited effect as phishing messages often rely on cues that encourage processing based on heuristics, which is increased by ones information consumption habits, whereas education encourages elaboration. As indicated in the literature it remains a major strategy to combat phishing as a security threat (Aleroud & Zhou, 2017).

Therefore, one should focus on this a part of resiliency, whereby an organisation understands its points of weakness and puts in place processes to decrease the recovery time from an incident. Furthermore, testing for phishing could also be emphasised as a part of resiliency, as demonstrated by Caralli, Allen and White (2010) in the CERT-RMM model. The model has a guide around 'insider' threats to security where they advocate for: ascertaining where 'insiders' pose a risk, including unintentionally; and investigate this to help actively mitigate these risk areas (Caralli et al, 2010). This does not mean that organisations can reduce or stop awareness efforts but instead they should be framed as a part of a response to phishing as a threat not the solution, even if technical controls like DMARC and advanced filtering is applied

to corporate/enterprise email inboxes. The results provide an indication of the proactive value as emphasised above and also looking at the reporting times of the incidents of spam in this study, meaning its short life span also provides support for awareness activities as it provides indication of a 'watercooler effect'. This term describes where people located in close proximity or have working ties with each other informally share their own knowledge - termed after work related conversations around an office watercooler (Sarker, Sarker, Nicholson & Joshi, 2003). With the VicBooks phishing message in particular staff members targeted even checked with the café about the deal and made them aware of this message. Meaning a distinct advantage of the proximity in our sample population, which is not clear in the other studies, illustrates how the 'watercooler effect' if present may help others in determining the nature of phishing messages. Indicating that awareness can potentially have a group level effect resulting from an individual level in an enterprise environment.

6. Limitations, implications for practice and future research

This section presents the limitations of the research, the implications for practitioners and finally the avenues for future research for an academic audience of the study.

6.1 Limitations

It is considered that this study had three major limitations affecting the design of the second round of data collection and therefore the final results. The main limitation concerns the limited sample sizes, second is that the research was called off part way through data collection in Stage One, and finally the descriptive rather than explanative nature of this study.

Limited sample size was constrained by access in this study. It was considered that the opportunity to conduct the study on a set unit of professional administrative and support staff was of greater value than that of using the more common sample population such as students. Since the size of this professional unit in the organisation was 141 employees it meant that it especially limited the results around mobile versus desktop susceptibility. This is why results were classified between these two and not the five device Operating Systems. This limitation is perhaps best seen in the highly variable results between each group around the proportions that used the mobile operating systems versus desktop ones, as evidenced by only 15 opening the email initially on mobile in Group B versus 11 Group A, meaning that for comparisons between this subset of the population this was still a small sample being far less

than $n=30$. If the study was conducted across a larger sample population and assuming it had a similar rate of opening, then one would expect less variability amongst the susceptibility rates on mobile.

The second limitation was that the proposed research in round two was called off with no replication carried out to make comparisons between study two and a potential follow-up study. The reason why there was not a third study - being a replication of study two (even without any initial awareness workshops), was that the time-frame for a second human ethics approval (due to the complex ethical nature of this study around informed consent). However, potentially for the small sample population and the proximity of members in the reference group the campaign time was already too long. This meant that the study could not test any hypotheses around exposure to simulated campaigns or even the impact of residual training, because of the two factors described above. There was an adequate amount of data collected for considering the impact of the message and gaining an insight into the complexities around device type and phishing susceptibility on it. Finally, the study highlights additional aspects that would be of value for future research, especially around the complexities in device usage behaviour, where it is not a choice of one or the other.

The final limitation was that this study does not predict or model the 'why' of the observed susceptibility results, so it cannot provide further explanatory insights into what was collected from the simulated experiment or directly validate the observed results. This would have highlighted the 'why' behind the variance between the two messages or the complexity around device interaction. This is a limitation in the scoping of the whole study, evidenced by the initial research questions, as it focussed on the 'what' rather than explaining 'why' observed phenomena happened. As Gregor (2006) emphasises not all theory is necessarily predictive or explanatory in predicting why phenomena occurs, also highlighting the role of 'Explanatory theories' amongst others. Explanatory theories are ones that do explicitly define the boundary or even modal relationships (Gregor, 2006). Therefore taking into account the nature of the discussion and of the original scope of this study it was intended to be a descriptive study around an area that has received little attention in the literature. This means that factors around knowledge, familiarity or even memory of the prior awareness training session could not be evaluated against the observed results from the simulated campaign. The study did highlight more detailed insights into not only the role of the device but also the

characteristics of the message content, such as 'gain versus loss' in affecting susceptibility that provides an insight into the phenomena for further research to explore in greater detail.

6.2 Implications for practice

This study sought to provide insights that would be useful for both a practitioner audience and academic audiences. This was guided by the principle of 'consumable research' by Robey and Markus (1998). The implications for the practitioner audience is identified below with guidance and the areas of future research are in the next section. The primary contribution of this research to the practitioner audience is around the complexity of device usage as it relates to phishing as an Information security threat. It can be seen that considering susceptibility as simply between mobile and desktop oversimplifies this important factor, which is demonstrated by the number of participants using multiple devices in the above field experiment. This is a more complex interaction happening between device types in both this research and in Vishwanath (2016), where in that study the participants were asked what device they were viewing on. There by suggesting that enterprise environments need to be aware of the potential role of device in how they address phishing, especially as some people in the field-experiment (study two) 'handed off' the phishing page from the mobile device to a desktop computer.

A contribution of this study to operational IT departments and general Information Security professionals is around the message type. This is shown by the differences in not only the open rates but also the susceptibility between the gain focussed message and the loss one. Emphasising the susceptibility of people to gain focussed phishing messages and or variations upon classic social engineering attacks. The results of this exercise emphasise the use of simulated phishing exercises to test staff reaction times. Therefore rather than aiming for a 100% pass rate in simulated phishing exercises the target should be around how quickly the operational IT departments can react to the simulated attack. This is a shift in focus to one of building resilience rather than immunity, especially as emphasised in the discussion around the Cyber Resilience framework for insider threats (Caralli et al, 2010).

A final area of relevance for practice relates to the message types used. It was considered that both phishing messages could have been stopped by the users if they followed the advice on phishing in the general security training received. It was intended that by looking at the URLs

they would have seen that the links and sending addresses did not belong to the actual domains the target organisations use. The domain used in the 'gain' message may have looked more legitimate to the participants, suggesting that the target group had not remembered this part of the training that they received around two months before this exercise. The assumption is that awareness training has an even effect upon decreasing the susceptibility to 'gain' versus 'loss' targeted phishing attacks, as the principles behind successfully identifying them remain the same.

6.3 Future Research

The implications for an academic audience is its benefits in extending this body of research for further academic inquiry into understanding the nature of susceptibility to phishing, and why this still remains an elusive goal. It is considered the primary areas for future research from this thesis are around the complexity of device type, which was over simplified by considering mobile versus desktop, and further understanding the effects of awareness training. As emphasised by Aleroud and Zhou (2017) present technical measures on mobile remain limited and they call for more work to be done on increasing user awareness.

An area for future research and consideration is the residual effect, if any, of awareness training for phishing and the willingness of people to engage in this. The role of awareness has been emphasised in the behavioural security literature, as well as the few studies that have taken a longitudinal approach (Crossler et al, 2013 & Warkentin, Johnston, Shropshire & Barnett, 2016). In considering awareness around social engineering Bullée et al (2015) call for studies around social engineering awareness to consider how long the impacts of an awareness campaign last for so as to determine, when to cycle training. This points to a clear need for longitudinal studies. In the phishing literature above this has consisted of two or three cycles of testing (Jansson & Vonsolms, 2013; Ngoqo & Flowerday, 2015 and Caputo et al, 2014). The closest study assessing for the residual impact of awareness training is Caputo et al (2014) who tested in three stages (February, May and September) the first involved all participants, then the second containing only those who clicked the link, and the last was across all participants again. They note that regular repeats must be done to see how long the impact of training lasts (Caputo et al, 2014). However this study does not show how awareness messages relate to different message appeals, though the principles behind awareness training are the same regardless of message. Though embedded training is simpler to

implement its results may differ from those where participants attend a workshop. An opportunity exists to show the difference between embedded and general security training, that encompasses social engineering (to avoid an expectancy effect) and establish what residual effect there is after a regular time interval. Secondly the Caputo et al (2014) study used informed participants that were debriefed after each round which means it can be assumed to be affected by the expectancy bias demonstrated by Pattinson et al (2012).

The awareness and the device type aspects of this study emphasises a potentially novel area for further research, being around the perceptions of security and susceptibility on different devices. Currently in the security literature studies have started to consider more the role of mobile security behaviour (Crossler & Bélanger, 2017 & Thompson, McGill & Wang, 2017). The models used in these studies could be used to inform research into social engineering perspectives on different devices, especially for mobile variations of phishing such as Smsishing or attempts to get the user to install a malicious Android application. The model presented around mobile privacy and security by Crossler and Bélanger (2017) considers both knowledge and awareness, while the privacy side of this model is potentially less relevant to social engineering research. This model would be important in this context as it would consider both awareness and knowledge in a single model unlike Arachchilage and Love (2014) who consider knowledge and self-efficacy in their model. Including awareness in a model for social engineering is important as it has long been considered a remedy as illustrated above and even in earlier behavioural phishing research, such as Workman (2008). Furthermore, an issue relating to device that should be examined in greater detail is emphasised in the results of this study which showed that many participants viewed on multiple devices. The use of multiple devices and even 'handing off' of the links in the emails could be further investigated by behavioural security researchers or usability researchers alike as it spans across the two disciplines. This would provide insights into when desktop versus mobile security becomes a factor, however it maybe more suited for the richer insights gained in a traditional experiment, such as in Parsons et al (2015a).

7 Conclusion

In summary this research set out to consider the use of simulated phishing attacks as a training measure, in determining if there is a difference in susceptibility based upon device type.

Initially from the literature it was uncertain as to the impact the device type would have on susceptibility to phishing attacks and also the effect of the persuasive language used by phishers in messages, around a 'gain' versus a 'loss' message. Comparing a sample of phishing emails received by the university in study one testing for differences in the message factor was seen as one of the key variables from the coding of the phishing emails.

This study provides a clear distinction from the past literature in comparing susceptibility in the sample population used to evaluate the susceptibility based on device type and message factors, being a simulated experiment across professional and administrative staff who had received external information security training. While it was planned that a second round was to be conducted to evaluate the impact of the simulated training both emails were reported to the university's service desk as phishing emails and had to be blocked on university networks. While it appears from the results that the 'gain message' was more effective than the 'loss' one, there was no clear impact of device type between the groups on susceptibility. Our findings highlight considerations for further research especially around device usage around phishing emails as users opened the emails on both a desktop and mobile device. This study's results emphasise the complexity of understanding device usage and security behaviour around a common threat to any enterprise environment, which to the best knowledge of the author has not been clearly illustrated in the prior literature on device type and phishing susceptibility. Furthermore, this research was original in considering both the message and the device type in the susceptibility of professional administrative staff to phishing. It is hoped further research will consider the impact of awareness training and how this relates to both message factors and device usage around phishing in modern enterprise environments.

Appendix A: Prior Phishing research and methods

Authors (date)	Theory	Synopsis	Methodology	Sample population	Measures
Alsharnouby, Alaca & Chiasson (2015)	N/A	Modern browser interfaces and greater awareness around phishing has improved user's ability to determine phishing websites	Experiment with eye tracking and interview during experiment	Carleton university through posters and e-mails	Browser interface elements looked at & correctly identifying phishing websites
Arachchilage & Love (2014)	Technology Threat Avoidance Theory	Impact of Procedural or conceptual knowledge affects a person's self-efficacy in being able to identify phishing links.	Survey	Undergraduate students at two UK universities	Procedural knowledge, self-efficacy, avoidance motivation and avoidance behaviour
Ngoqo & Flowerday (2015)	Two factor taxonomy of Security Behaviour and Theory of Planned Behaviour	Mobile SMS phishing and the impact of awareness campaigns in susceptibility.	Simulated phishing attack and survey (each group gets exposed to two awareness campaigns.	Undergraduate Information Systems students in South Africa	Awareness and behavioural intent (from TPB). Participants separated into two groups based upon those who responded to the first SMS
Caputo, Pfleeger, Freeman & Johnson (2014)	N/A	Replication study assessing the effectiveness of embedded training to spear phishing attacks	Simulated phishing attacks (Four treatments according to different focus of awareness materials) and interviews	Organisation in Washington DC	Number who clicked the link, amount of time a person spent on the page.

Pattinson, Jerram, Parsons, McCormac & Butavicius (2012)	N/A	Explored why some participants managed genuine and phishing emails better than others.	Experiment (Two Groups One was informed it was a phishing study the other was not)	Students at the University of Adelaide	Scoring system based on response to genuine or phishing email
Jansson & von Solms (2013)	N/A	Exploring the impact of simulated phishing attacks with embedded training	Simulated phishing attacks (Each faculty received a different type of phishing message)	Students in each faculty at a South African University	Number of users who logged into email, number who clicked on the link and number who used the awareness material
Parsons, McCormac, Pattinson, Butavicius & Jerram (2015)	Signal Detection Theory	Explored difference between informing participants about phishing experiment versus not.	Experiment	Students at University of Adelaide	Time taken, Response taken to each email.
Valecha, Chen, Herath, Vishwanth & Wang (2015)	Heuristic-systematic Model	What information will people share in social-media phishing attacks	Coded tweets on Twitter, then statistical analysis from codes	Web based research (netnographic)	Coding schema on phishing and victimisation
Luo, Zhang, Burd & Seazzu (2013)	Heuristic-Systematic Model (HSM) and	Exploring susceptibility of users to spear-phishing spread by email mimicking the university's portal	Simulated phishing attack and survey	University Staff and faculty within a school	Hit-rate, amount of credentials gathered, Survey was never completed
Vishwanath, Herath, Chen, Wang & Rao (2011)	Interpersonal Deception Theory & ELM	The research shows a model to explain susceptibility to email phishing attacks by predicting the person's likelihood to respond to the threat	Survey	University students following two major phishing incidents	Email Cues (Interpersonal Deception Theory), Involvement (ELM), knowledge, self-efficacy, Elaboration (ELM)

Iuga, Nurse & Erola (2016)	N/A	Research explores factors that help users identify legitimate webpages from phishing websites using Facebook as an example	Web-based user study	Primarily university students or recent graduates	Heat maps from mouse cursor, success of detecting phishing site, time
Kumaraguru et al (2009)	Deception Theory	Looks at retention of embedded phishing training messages	Simulated attack (3 groups: Control, single treatment and multiple)	CMU Students and staff but 56% of sample students	Survey, Percentage who clicked and percentage who gave credentials
Workman (2008)	ELM/ Health Belief Model	Explores the behaviours associated with susceptibility to falling victim to phishing attacks	Field study (Simulated phishing attack and survey)	Company employees in USA	Normative commitment, Continuance commitment, Affective commitment, Trust Fear and Reactance
Harrison et al (2015)	HSM	Explores the impact of the degree of social presence and information richness on victimization	Factorial experiment (Simulated attack)	125 undergraduate students in North East USA	Heuristic Processing, Systematic Processing, victimisation
Harrison et al (2016a)	HSM	Looks at personalities that are more suspicious perception of trust	Simulated attack and survey	University students USA	Efficacy, suspicion, Information Insufficiency, Heuristic processing, systematic processing, Trust
Harrison et al (2016b)	ELM	Looks at message factors and user's knowledge and experiences in predicting victimisation	Field experiment (Simulated phishing attack)	University students North East USA	Fear versus reward message, leakage cues, subjective knowledge, objective knowledge, attention, elaboration, victimisation
Vishwanath (2015)	HSM, "Big - five personality traits"	Looks at the role of habit and processing on phishing victimisation.	Field experiment (Simulated phishing attack)	Undergraduates (University of Buffalo)	Email habit, Heuristic processing, Systematic processing, "Big-five personality", Information Insufficiency, victimisation

Vishwanath et al (2016)	Extension of HSM to form Suspicion, Cognition, and Automaticity model (SCAM) (Pg3)	Considers how beliefs and habits inform suspicion towards phishing emails, comparing hyperlink and attachment	Experiment	Undergraduates (University of Buffalo)	Risk-beliefs, self-regulation, Heuristic processing, Systematic Processing, Email habits, suspicion
Vishwanath (2016)	HSM	Looks at the role of device affordance on heuristic processing and habits on victimisation	Experiment	Students (University of Buffalo- Singapore)	Device affordance, heuristic processing, email load, habit, co-presence, self-regulation, victimisation
Goel et al (2017)	N/A	Considers the role of contextualisation, gain/loss and motive on susceptibility to different types of phishing emails	Field experiment	Students (University of Albany)	Opened email, links clicked, past experiences in survey
Bennenson et al (2017)	N/A	Difference in susceptibility to phishing attacks between Facebook and email	Field experiment with follow up questionnaire	University students	The number clicked, data on friend request etc on Facebook, and qualitative coding of reason for clicking
Jagactic et al (2007)	N/A	Explores susceptibility to phishing attacks where the phisher is spoofing a social media friend of the victim	Field experiment	Students aged 18-24	Number clicking through

Moody et al (2011)	N/A	Considers the effect of personality type and internet experience on susceptibility to phishing attacks	Online experiment (simulated attack and survey)	Psychology and Information Systems students	Personality, internet experience and number of clicks on the link
Downs et al (2006)	N/A	Explores the strategies used by people with little security knowledge to identify phishing emails	Scenario Experiment	Students and professionals	Suspicion, interview data and strategies reported by participants
Sun et al (2016)	N/A	Explores the effect of self-efficacy with regards to phishing as a mediating variable between internet self-efficacy and behaviour around phishing	Survey	College and university students	Self-efficacy (internet), self-efficacy (phishing), anti-virus adoption, action confirmation, learning for protection
Alseadoon et al (2013)	Model of Detecting Deception	Explores the different types of victims to phishing attacks	Survey, simulated attack and interviews	Students aged 18-25	Personality and interview data
Halevi et al (2007)	"Big-five personality types"	Explores personality types and disclosure of information on Facebook with susceptibility to phishing	Survey and simulated phishing attack	Engineering students	Facebook privacy settings, personality, and
Dhamija et al (2006)	N/A	Usability studying considering how phishing websites fool users	Usability experiment	University staff and students	Interview data and participant behaviour
Blythe et al (2011)	N/A	Exploring what aspects of phishing emails and the factors behind victimisation	Coding of phishing emails, survey, interview with blind users	People aged 18-65	Qualitative analysis, participant's recognition of genuine vs phishing emails
Vidas et al (2013)	N/A	Explores the susceptibility of people to Qrishing (phishing by QR codes)	Field experiments	People over 18 in Pittsburgh	Surveillance data, number of scans, survey data on QR code knowledge and behaviour

Appendix B: Overview of the constructs used in phishing studies

Study	Behavioural construct	Intermediary variable	Dependent variable	Hypothesis Supported
Arachilage & Love (2014)	Knowledge	Self-efficacy	Avoidance motivation	Interaction between procedural and conceptual knowledge and self-efficacy
Vishwanath et al (2011)	Knowledge	Elaboration	Response likelihood	Partial support
	Computer self-efficacy	Elaboration	Response likelihood	Not supported
Harrison et al (2016b)	Fear vs Reward	Attention and elaboration	victimization	Not supported
	Leakage Cues	Attention and elaboration	victimization	Not supported
	Email knowledge and phishing knowledge	Attention and elaboration	Victimization	Both Supported
Vishwanath et al (2016)	Risk beliefs	Heuristic and systematic processing	Suspicion	Supported for decrease heuristic, supported only in link for systematic and directly
	Self-regulation	Email habits	Suspicion	Supported
Sun et al (2016)	Internet self-efficacy	Anti-phishing self-efficacy	Anti-phishing behaviour	Supported
Harrison et al (2016a)	Efficacy (email)	Suspicion		Supported
	Suspicion	Information Insufficiency and processing	Trust	Supported for insufficiency, partially supported for the processing routes and both processing routes supported on trust
Vishwanath (2016)	Richness	Device	Heuristic processing	Richness directly increases heuristic processing
	Email habit Heuristic processing	Device		Email habits precedent to victimisation
Vishwanath (2015)	Personal traits (conscientiousness, etc)	Email Habit		Supported for conscientiousness and emotional stability but

				not for information insufficiency
	Personality, information insufficiency	Processing (Heuristic and systematic)		Only supported for information insufficiency on systematic processing and direction for heuristic
		Email habit, systematic processing & heuristic processing	victimisation	All supported
Goel et al (2017)	contextualisation		susceptibility	Partial support
	Loss/gain		susceptibility	Loss effect for contextualised otherwise gain had high click through
	motive		susceptibility	Supported
Moody et al (2011)	Source		susceptibility	Supported
	Text link		Susceptibility	Significant for full model
	Personality			Not supported
	Internet experience		susceptibility	Supported
Luo et al (2013)	Argument quality		Victimisation	Supported
	Genre and source		Victimisation	Supported
	Pretext and damage potential		Victimisation	Supported
Benenson et al (2017)	Medium (Facebook)		Attack success rate	Supported
	Sender profile		Attack success rate	Not supported
	Gender		Attack success rate	Not Supported
Workman (2008)	Commitment traits		Susceptibility	Supported
	Trust and obedience		Susceptibility	Supported
	Reactance		Susceptibility	Positive though not supported

Appendix C: Coding schema and application

Coding table adapted from Parsons et al (2015b) and Kim and Hyun Kim (2013)

Cue/element type	Cue/element	Aspect of cue	Publications
Message cues	Links	HTTPS, URL	Parsons et al (2015b); Alsharnouby et al (2015); Arachchilage & Love (2014)
	Visual presentation	Logo, banners company logo, design	Parsons, et al (2015b), Kim & Hyun Kim (2013); Harrison et al (2015), Blythe et al (2011)
	Contextualisation	Personalised, organisation specific	Goel et al (2017)
	Spelling and grammar	Errors, consistency	Downs et al (2006); Luo et al (2013); Parsons et al (2015b); Blythe et al (2011); Harrison et al (2016a)
	Sender	Sending address, contact details, and reply address	Parsons et al (2015b); Kim & Hyun Kim (2013); Jagactic et al (2007); Benenson et al (2017)
	Credibility	Source credibility, Trusted organisation	Parsons et al (2015b); Parsons et al (2015a), Kim & Hyun Kim (2013); Jansson & von Solms (2013)
Message appeals	Urgency	Time pressure, Forceful language	Kim & Hyun Kim, (2013); Luo et al (2013); Vishwanath et al (2011); Parsons et al, (2015b)
	Reasoning (argument quality)	Cause, Analogy, sign/associated event	Kim & Hyun Kim (2013); Luo et al (2013);
	Emotional appeal	Fear (loss), Affection, Happiness (gain)	Kim & Hyun Kim (2013), Harrison et al, (2016a); Goel et al (2017); Workman (2008); Parsons et al (2015a)
	Motivation	Belonging, self-esteem, safety	Kim & Hyun Kim, (2013) & Ferreira et al, (2015)

Coding schema application (sample)

Note: n/a means could not be determined or not enough information so not applicable

This content is unavailable

Please consult print version for access

Appendix D (part one): Technical considerations

The following technical considerations around spam filtering, whitelisting and the technical safeguards put into enterprise mail systems to prevent phishing were considered in designing study two (the simulated phishing exercise). This was a concern as the university mail systems use Microsoft Exchange Server with DMARC. Meaning that simple techniques to send out test phishing emails to our test @vuw.ac.nz account would result in them potentially being filtered as spam or greylisted. This technical consideration gives an overview, but not an exhaustive discussion on these technologies, because the research questions are not around evaluating phishing technologies and secondly this research does not focus on technical safeguards. The literature review placed technical research on phishing outside of its boundary conditions. These factors were considered so as to get around the filtering, for study 2.

Spam filtering

Spam filtering on email was considered the primary technical counter-measure in place, as the aim of any phishing attack is to maximise the quantity of messages received in order to get higher click through rates. In the testing preceding study two we wanted to avoid greylisting. Greylisting delays the arrival of the messages classified as such (Kucherawy & Crocker, 2012). It works where by the receiving SMTP server, has a cache of incoming IP addresses and if it is not in this or has not met the minimum age period set since the initial connection then the receiving server will return a temporary failure (Kucherway & Crocker, 2012). This was a concern for research as delaying when the message appeared in a target's inbox would shorten the length of the campaign. A noted limitation of this technology is that if it returns a false positive then it delays legitimate emails (Kucherway & Crocker, 2012). While sending IP addresses could have been whitelisted even during testing it was initially regarded that this would ideally be avoided, so as to replicate the challenges faced in performing phishing attacks. Though whitelisting would have likely needed to be avoided in some of the studies (Vishwanath et al, 2011; Luo et al, 2013, Vishwanath et al, 2016 & Harrison et al, 2015) as they used Gmail accounts, and therefore would likely have passed greylisting. Using an @outlook.com account was not deemed to be an option in this study since it is common knowledge amongst university staff and students that the university uses @vuw.ac.nz email addresses. As illustrated above in the screen shots from the Outlook desktop client, the email

address appears in the header for emails. Email whitelisting even during testing would have avoided this concern and was an option for this research as it was conducted in collaboration with the systems team in the University's Information Technology Service.

Spam filtering employs a range of techniques to determine if a message is what it considers legitimate email or 'spam'. For this study this was not assessed, as the thesis was not aimed at evaluating the efficacy of spam filter techniques. Spam filtering involves machine learning techniques that classify emails based on keywords (Guzella & Caminhas, 2009). Furthermore, filters often use two or more techniques, where validation of open source tools such as SpamAssassin have achieved 99% accuracy rates, also bundled into filters is Optical Character Recognition to find text embedded behind and in images (Guzella & Caminhas, 2009). Therefore, indicating that the use of images to obscure aspects in phishing emails may work from a visual deception to a typical end-user but is not guaranteed to bypass spam filters. Despite the sophistication of defences such as spam filtering, social engineers are still getting through.

DMARC and DKIM

DMARC and DKIM are an important consideration in running tests for the sending of the email messages. This is a technique used to try and ensure that addresses are not being faked. For the study to succeed any tool used would need to pass controls such as DMARC, DKIM and SPF. This would not have been a problem for some of the simulated attacks in the literature review as they used Gmail accounts, or like Goel et al (2017) who used actual email accounts on the university domain. DMARC was built on top of the existing DKIM and SPF frameworks with the explicit aim of reducing phishing and spammers ability to fake legitimate trusted domains (DMARC, 2018). This aims to stop the ability of spammers imitating the exact domain of the legitimate sources and verifying the DKIM signatures and SPF records attached to the sending domain, meaning that swapping the top-level domain and sending an email address from that would be able to pass DMARC. As is evidenced in the study 1 results where the social engineers for the Xero invoice phishing attack, sent from a xero.net.nz domain. It meant that for sending phishing emails for this study tools either needed to be built to pass DMARC, DKIM & SPF or

to configure mail servers coming from similar domains to the target domain (Kucherawy & Zwicky, 2015).

Appendix D (part two): Tool Selection

This appendix provides greater detail of the options that were evaluated in the selection of the tool used for the simulated phishing exercise in study 2. It aims to provide a broad overview as to why each option was scored as it was.

Marketing tools

This tool selection was based on the literature and involved using tools that are meant for email- marketing. It was considered that for launching and sending campaigns and tracking clicks this was a viable option, though less simple for a spear phishing landing page. This solution would have required the use of a G-Suite account registered to a domain such as the vlctoria.ac.nz, which was purchased for this study, marketing software to track the opening rates and Google analytics on our replicated landing page to track where the click came from and the device type used. Replicating the landing page would have meant hosting a webserver with effectively a copy of the page source code, and server-side code for redirecting users to another webpage hosted on that server if characters submitted were greater than one character in length for the username and password. It would also mean tracking for click through rates from our simulated login page for tracking how many submitted credentials. Since this solution would combine data from multiple sources, it was deemed a viable solution, though more time-consuming to implement and validate.

GoPhish

An open source project on GitHub for creating, launching and tracking phishing campaigns was considered to avoid commercial costs. This project is programmed in the Go Language and has a Go installer package for installing the program (GoPhish, 2017). This tool was validated by installing and running it on an Ubuntu 14.04 server running on Microsoft Azure. It was run on an external server because if it was going to be selected then this would be required for the landing page to always be available. This tool was rejected because when

creating the campaign, the email is linked to a landing page created in the service. Therefore, this was rejected because the link in the email address would likely appear to be from an IP address and port number.

This service uses port 3333 for the GoPhish service and then uses MailHog (a separate GitHub project) for creating test SMTP servers, using port 1025 to send and 8080 for the mailbox, this is run on the same SMTP server (GoPhish, 2017). Another issue encountered in validation of this service, was that it uses self-signed https certificates. For device tracking it captures the full device agent as follows, in this case: "Mozilla/5.0 (Macintosh; Intel Mac OS X 10.12; rv:57.0) Gecko/20100101 Firefox/57.0". This gives the full details like Google Analytics of the browser and device used. The advantages of this tool are it being open source and landing pages can replicate the page of a given target. However, there was budget in the project for a commercial tool.

Phish5

The first commercial tool evaluated for this research was Phish5, commercial tools would likely be used by enterprise environments as they make it simpler to launch and monitor campaigns. This tool was rejected on pricing when compared to the other commercial offering. The tool complies with human ethics in that it does not collect credentials in the campaigns, users can remove any of their data from the site and provides SMTP logs (Phish5, 2017). These were all considered necessary features to meet the research requirements. The tool offers customisation around the landing page, which was considered a benefit with respect to landing pages and URLs. However, in validating this tool using pre-set templates for Microsoft Office the following issues were found when testing; on one university email service that uses Linux systems, the email was grey listed. In subsequent testing on the test account on the staff domain which uses, Microsoft Exchange this was allowed through initially, but when testing the https option for the landing page it used self-signed certificates. Finally, on a repeat of the initial test to the staff domain the message failed to pass Microsoft Outlook's spam filters and a quarantine report was sent to the targeted email address. In conducting this test it was later realised by the researcher, that the service needs to be whitelisted.

Overall the tool met all the criteria, including desired aspects such as simplicity, but it was expensive with billing on a per campaign basis.

PhishingBox- through local reseller

The PhishingBox tool offered a similar product to the above commercial product, but had more favourable pricing and so it was selected for this research. This service added more assurances as it integrates with Microsoft services, and the landing pages are optimised for mobile devices, including ones that mimic Microsoft landing pages which was determined to be important for study 1 (PhishingBox, 2017). It was accepted from this point that all tools regardless of approach would need to be whitelisted to ensure emails ended up in each target's inbox. Ultimately the choice between the two commercial tools came down to price. From a research perspective since the thesis was not aiming to evaluate efficacy of different phishing tools, or review this but to explore any differences in susceptibility between users of mobile or desktop devices.

Appendix E: Message design

'Gain' Message and landing page

This content is unavailable

Please consult print version for access

'Gain' Message promising free VicBooks Coffee

This content is unavailable

Please consult print version for access

Landing page for VicBooks message

This content is unavailable

Please consult print version for access

Sample of 'gain' message from Study 1

Loss Message and landing page

This content is unavailable

Please consult print version for access

'Loss' Message targeting Office365

This content is unavailable

Please consult print version for access

'Loss' Landing page for Office365

This content is unavailable

Please consult print version for access

Sample of Loss message from Study 1

Appendix F: Human ethics

Copy of debriefing email in Ethics:

Dear VUW *[Blank]* staff

Recently you would have received emails offering either a free coffee or the other wanting you to provide your login credentials, to get access back to your university Office365 account. Both were phishing emails sent as a part of a phishing test. You were randomly assigned as to which email you would receive. Both were variations of reasonably common phishing emails that you or anyone else in the university could receive. Phishing messages are often intentionally designed to take advantage of well documented human responses and reactions in how we process information to achieve a higher response rate.

These campaigns were granted Human Ethics approval from the university's central human ethics committee (id: 25045) and approval from the relevant senior managers, including the CIO. These messages were interested in testing your susceptibility to two common strategies: being promised something or fooling you into thinking you will lose something.

This experiment was not conducted to identify those who fell victim but only to profile the University's risk to phishing as a cyber-attack vector. This is a part of a larger area of research that aims to better understand and build the University's resilience to this type of threat. You were not informed about this exercise as it may have created an expectancy effect that would bias the results. All data will be kept confidential and information collected from this will be aggregated so you can not be identified. Furthermore, no user submitted data was captured in this exercise.

Your data will not be used to personally identify you. Furthermore, these emails were benign unlike real ones, so no harm was caused to University systems. If you have any concerns or any questions please do not hesitate to contact me or my supervisor A/Prof Ian Welch (email: ian.welch@vuw.ac.nz)

Yours Sincerely,

Jayden Nowitz

(email: jayden.nowitz@vuw.ac.nz)

(Engineering and Computer Science, VUW)

Appendix G: Summary of raw data study two

Group A: VicBooks

This content is unavailable

Please consult print version for access

Group B: Office365

This content is unavailable

Please consult print version for access

References

- Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. *Computers & Security*, 68, 160-196.
- Alsaleh, M., Alomar, N., & Alarifi, A. (2017). Smartphone users: Understanding how security mechanisms are perceived and new persuasive methods. *PloS one*, 12(3): e0173284, 1-35.
- Alseadoon, I. M., Othman, M. F. I., Foo, E., & Chan, T. (2013). Typology of phishing email victims based on their behavioural response. In *Proceedings of the 19th Americas Conference on Information Systems*, 19. AIS
- Alsharnouby, M., Alaca, F., & Chiasson, S. (2015). Why phishing still works: user strategies for combating phishing attacks. *International Journal of Human-Computer Studies*, 82, 69-82.
- APWG. (2016). Phishing activity trends report: 4th Quarter, 2016. Anti-Phishing Working Group. Retrieved from: http://docs.apwg.org/reports/apwg_trends_report_q4_2016.pdf
- Arachchilage, N. A. G., & Love, S. (2014). Security awareness of computer users: A phishing threat avoidance perspective. *Computers in Human Behavior*, 38, 304-312.
- Arachchilage, N. A. G., Love, S., & Beznosov, K. (2016). Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior*, 60, 185-197.
- Armin, J. (2013). Mobile Threats and the Underground Marketplace. APWG White Paper: Mobile. Retrieved from: http://docs.apwg.org/reports/mobile/apwg_mobile_fraud_report_april_2013.pdf
- Benenson, Z., Gassmann, F., & Landwirth, R. (2017). Unpacking Spear Phishing Susceptibility. In *International Conference on Financial Cryptography and Data Security* (pp. 610-627). Springer, Cham.
- Blythe, M., Petrie, H., & Clark, J. A. (2011). F for fake: four studies on how we fall for phish. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3469-3478). ACM.
- Botha, R. A., Furnell, S. M., & Clarke, N. L. (2009). From desktop to mobile: Examining the security experience. *Computers & Security*, 28(3-4), 130-137.
- Bottazzi, G., Casalicchio, E., Cingolani, D., Marturana, F., & Piu, M. (2015). MP-Shield: a framework for phishing detection in mobile devices. In *IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*, 2015 (pp. 1977-1983). IEEE.

Bullée, J. W. H., Montoya, L., Pieters, W., Junger, M., & Hartel, P. H. (2015). The persuasion and security awareness experiment: reducing the success of social engineering attacks. *Journal of experimental criminology*, 11(1), 97-115.

Caputo, D. D., Pfleeger, S. L., Freeman, J. D., & Johnson, M. E. (2014). Going spear phishing: Exploring embedded training and awareness. *IEEE Security & Privacy*, 12(1), 28-38.

Caralli, R. A., Allen, J. H., & White, D. W. (2010). *CERT resilience management model (CERT-RMM): a maturity model for managing operational resilience*. Addison-Wesley Professional.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of personality and social psychology*, 66(3), 460.

Chin, E., Felt, A. P., Sekar, V., & Wagner, D. (2012). Measuring user confidence in smartphone security and privacy. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, (p. 1). ACM.

Coopamootoo, K. P., & Groß, T. (2016). Evidence-based methods for privacy and identity management. In *IFIP International Summer School on Privacy and Identity Management*, (pp. 105-121). Springer, Cham.

Crossler, R. E., & Bélanger, F. (2017). The Mobile Privacy-Security Knowledge Gap Model: Understanding Behaviors. In *Proceedings of the 50th Hawaii International Conference on System Sciences*. DOI: 10.24251/HICSS.2017.491

Crossler, R. E., Johnston, A. C., Lowry, P. B., Hu, Q., Warkentin, M., & Baskerville, R. (2013). Future directions for behavioral information security research. *Computers & Security*, 32, 90-101.

Dang-Pham, D., & Pittayachawan, S. (2015). Comparing intention to avoid malware across contexts in a BYOD-enabled Australian university: A Protection Motivation Theory approach. *Computers & Security*, 48, 281-297.

Davinson, N., & Sillence, E. (2010). It won't happen to me: Promoting secure behaviour among internet users. *Computers in Human Behavior*, 26(6), 1739-1747.

Dhamija, R., Tygar, J. D., & Hearst, M. (2006, April). Why phishing works. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 581-590). ACM.

DMARC (2018). DMARC Overview. DMARC.org. Retrieved from: <https://dmarc.org/overview/>

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

Ferreira, A., Coventry, L., & Lenzini, G. (2015, August). Principles of persuasion in social engineering and their use in phishing. In *International Conference on Human Aspects of Information Security, Privacy, and Trust* (pp. 36-47). Springer.

Greitzer, F. L., Strozer, J. R., Cohen, S., Moore, A. P., Mundie, D., & Cowley, J. (2014, May). Analysis of unintentional insider threats deriving from social engineering exploits. In *2014 IEEE Security and Privacy Workshops (SPW)*, (pp. 236-250). IEEE.

Goel, S., Williams, K., & Dincelli, E. (2017). Got Phished? Internet Security and Human Vulnerability. *Journal of the Association for Information Systems*, 18(1), 22.

GoPhish. (2017). Gophish User Guide. GoPhish Version 0.3-dev. Retrieved from: <https://www.gitbook.com/book/gophish/user-guide/details>

Gregor, S. (2006). The nature of theory in information systems. *MIS quarterly*, 30(3), 611-642.

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, 36(7), 10206-10222.

Harrison, B., Vishwanath, A., Ng, Y. J., & Rao, R. (2015). Examining the impact of presence on individual phishing victimization. In *48th Hawaii International Conference on System Sciences (HICSS)*, 2015 (pp. 3483-3489). IEEE.

Harrison, B., Vishwanath, A., & Rao, R. (2016a). A user-centered approach to phishing susceptibility: The role of a suspicious personality in protecting against phishing. In *49th Hawaii International Conference on System Sciences (HICSS)*, 2016 (pp. 5628-5634). IEEE.

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

Halevi, T., Lewis, J., & Memon, N. (2013). Phishing, personality traits and facebook. *arXiv preprint arXiv:1301.7643*.

Internet Crime Complaint Center (IC3). 2015 Internet crime report (2015 annual report). FBI and DoJ. Washington, D.C. Retrieved from https://www.ic3.gov/media/annualreport/2015_IC3Report.pdf

Iuga, C., Nurse, J. R., & Erola, A. (2016). Baiting the hook: factors impacting susceptibility to phishing attacks. *Human-centric Computing and Information Sciences*, 6(1), 1-20.

Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94-100.

Jakobsson, M. (2007). The human factor in phishing. *Privacy & Security of Consumer Information*, 7(1), 1-19.

Jansson, K., & Von Solms, R. (2013). Phishing for phishing awareness. *Behaviour & Information Technology*, 32(6), 584-593

Kim, D., & Hyun Kim, J. (2013). Understanding persuasive elements in phishing e-mails: A categorical content and semantic network analysis. *Online Information Review*, 37(6), 835-850.

Komatsu, A., Takagi, D., & Takemura, T. (2013). Human aspects of information security: An empirical study of intentional versus actual behavior. *Information Management & Computer Security*, 21(1), 5-15.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (p. 3). ACM.

Kraska-Miller, M. (2013). *Nonparametric statistics for social and behavioral sciences*. CRC Press. ProQuest E-book central. Retrieved from: <http://ebookcentral.proquest.com/lib/vuw/detail.action?docID=1511153>.

Kucherawy, M., & Crocker, D. (2012). Email Greylisting: An Applicability Statement for SMTP. Internet Engineering Task Force (IETF). Request for Comments: 6647 ISSN: 2070-1721. Retrieved from <https://tools.ietf.org/html/rfc6647>

Kucherawy, M., & Zwicky, E. (2015). Domain-based message authentication, reporting, and conformance (DMARC). Internet Engineering Task Force (IETF). Independent Submission Request for Comments: 7489 ISSN: 2070-1721. Retrieved from: <http://buildbot.tools.ietf.org/html/rfc7489>

Luo, X. R., Zhang, W., Burd, S., & Seazzu, A. (2013). Investigating phishing victimization with the Heuristic–Systematic Model: A theoretical framework and an exploration. *Computers & Security*, 38, 28-38.

Mohammad, R. M., Thabtah, F., & McCluskey, L. (2015). Tutorial and critical analysis of phishing websites methods. *Computer Science Review*, 17, 1-24.

Moody, G., Galletta, D., Walker, J., & Dunn, B. (2011). Which phish get caught? An exploratory study of individual susceptibility to phishing. In *32nd International Conference on Information Systems*. Association for Information Systems, Shanghai

Ngoqo, B., & Flowerday, S. V. (2015). Information Security Behaviour Profiling Framework (ISBPF) for student mobile phone users. *Computers & Security*, 53, 132-142.

Parker, F., Ophoff, J., Van Belle, J. P., & Karia, R. (2015). Security awareness and adoption of security controls by smartphone users. In *Second International Conference on Information Security and Cyber Forensics (InfoSec)*, 2015 (pp. 99-104). IEEE.

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

Parsons, K., Butavicius, M., Pattinson, M., Calic, D., McCormac, A., & Jerram, C. (2015b). Do Users Focus on the Correct Cues to Differentiate Between Phishing and Genuine Emails?. *Australasian Conference on Information Systems*, Adelaide, p.57

Pattinson, M., Jerram, C., Parsons, K., McCormac, A., & Butavicius, M. (2012). Why do some people manage phishing e-mails better than others?. *Information Management & Computer Security*, 20(1), 18-28.

Petty, R. E., & Cacioppo, J. T. (1986). *The elaboration likelihood model of persuasion*. In *Communication and persuasion* (pp. 1-24). Springer New York.

Pfleeger, S. L., Sasse, M. A., & Furnham, A. (2014). From weakest link to security hero: Transforming staff security behavior. *Journal of Homeland Security and Emergency Management*, 11(4), 489-510.

Phish5. (2017). Phish5 Tour: See how Phish5 Works. Retrieved from: <https://phish5.com/tour>

PhishingBox. (2017). Phishing Simulator: Phishing Tests & Training for Security Awareness. Retrieved from: <https://www.phishingbox.com/products-services/phishing-simulator>

Robey, D., & Markus, M. L. (1998). Beyond rigor and relevance: producing consumable research about information systems. *Information Resources Management Journal*, 11(1),

SANS Institute (2016). Enterprise Survival Guide for Ransomware Attacks. SANS Institute. Retrieved from: <https://www.sans.org/reading-room/whitepapers/incident/enterprise-survival-guide-ransomware-attacks-36962>

Sarker, S., Sarker, S., Nicholson, D., & Joshi, K. (2003, January). Knowledge transfer in virtual information systems development teams: an empirical examination of key enablers. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, 2003. (pp. 10-pp). IEEE.

Sauro, J., & Lewis, J. R. (2016). Quantifying the user experience: Practical statistics for user research. Elsevier Inc: <http://dx.doi.org/10.1016/B978-0-12-802308-2.00005-9>

Sun, J. C. Y., Yu, S. J., Lin, S. S., & Tseng, S. S. (2016). The mediating effect of anti-phishing self-efficacy between college students' internet self-efficacy and anti-phishing behavior and gender difference. *Computers in Human Behavior*, 59, 249-257.

Thompson, N., McGill, T. J., & Wang, X. (2017). "Security begins at home": Determinants of home computer and mobile device security behavior. *Computers & Security*, 70, 376

Tran, H., Campos-Nanez, E., Fomin, P., & Wasek, J. (2016). Cyber resilience recovery model to combat zero-day malware attacks. *Computers & Security*, 61, 19-31.

Valecha, R., Chen, R., Herath, T., Vishwanath, A., Wang, J., & Rao, R. (2015). An exploration of phishing information sharing: A heuristic-systematic approach. In Proceedings of the 10th Pre-ICIS Workshop on Information Security and Privacy- Proceedings 2, Fort Worth.

Vidas, T., Owusu, E., Wang, S., Zeng, C., Cranor, L. F., & Christin, N. (2013, April). QRishing: The susceptibility of smartphone users to QR code phishing attacks. In *International Conference on Financial Cryptography and Data Security* (pp. 52-69). Springer, Berlin, Heidelberg

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

Vishwanath, A. (2016). Mobile device affordance: Explicating how smartphones influence the outcome of phishing attacks. *Computers in Human Behavior*, 63, 198-207.

Vishwanath, A., Harrison, B., & Ng, Y. J. (2016). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research*, 1-21, DOI: 10.1177/0093650215627483

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

Warkentin, M., Johnston, A. C., Shropshire, J., & Barnett, W. D. (2016). Continuance of protective security behavior: A longitudinal study. *Decision Support Systems*, 92, 25-35.

Workman, M. (2008). Wisecrackers: A theory-grounded investigation of phishing and pretext social engineering threats to information security. *Journal of the Association for Information Science and Technology*, 59(4), 662-674.

Wu, L., Du, X., & Wu, J. (2016). Effective defense schemes for phishing attacks on mobile computing platforms. *IEEE Transactions on Vehicular Technology*, 65(8), 6678-6691.