The Employee Experience in Cybersecurity and How to Mitigate Risk

Laura M. Bishop

School of Psychology, Cardiff University

Supervisors:

Prof. Phillip L Morgan, Dr Phoebe Asquith, Prof. Dylan M Jones, Prof. William J Macken.

April 2023

Acknowledgements

I would like to begin by offering my sincere thanks to my lead supervisor, Professor Phillip Morgan, my professional constant and dear friend. Someone without whom - this thesis and the phenomenal opportunities experienced would not have been possible. I must also thank my two late supervisors, Professor Dylan Jones and Professor Bill Macken, who sadly passed away during the time of my PhD. I am honoured to be one of their final PhD students and am thankful for the advice and wisdom learned, over many pints of beer. Also, huge appreciation goes to Dr Phoebe Asquith, a close friend and colleague who stepped in as supervisor after the loss of Dylan and Bill, offering guidance at times of complete and utter bewilderment.

I was fortunate to work with many organisations during my PhD most significantly Airbus, who I collaborated with throughout and am now honoured to lead their Human-centric Cybersecurity Research Programme within the wonderful Innovations and Scouting Team. Thanks also to ThinkCyber, a UK startup who apply behavioural science to security awareness training via their nudging software product Redflags®, for their funding and support during Study 6, and OutThink where I worked as Director of Human Risk Science whilst writing the Playbook portion of this thesis. Special thanks goes to Ceri Jones at Lego for her guidance, friendship and many hours putting the cybersecurity world to rights.

During the process of my PhD I had to largely work full time, become teacher to three children during Covid-19 and suffer a severe personal trauma. Eternal thanks goes to my parents, sister and brother-in-law for their unremitting support. My beautiful nieces and nephews and unbelievable friends including Helen and Linn, not forgetting my Grandmother Gloria who early in my life demonstrated to me that it is never too late to return to study.

This PhD is dedicated to my three children Jake, Aurora and Felicity for their unbelievable patience and kindness. Mum did it kids!!!!

Abstract

With society now heavily invested in computer systems and internet connectivity, it has never been more vital to identify ways to safeguard cyberspace (Asquith & Morgan, 2019). In 2021, over 23,896 cyber security incidents were reported to have taken place across the globe, with a data breach confirmed in over 5,212 of these incidents (Verizon, 2022). Despite many organisations now applying time and budget to cybersecurity awareness training, 82% of security breaches are still found to involve a human element (Verizon, 2022).

The aim of this PhD was to better understand the human experience in cybersecurity, internal individual differences that can result in decision-making vulnerabilities, but also the impact of additional external pressures such as offender persuasion attempting to leverage on human susceptibility, to the impact of persuasive interventions generated to promote secure behaviour. The result – a Cybersecurity Awareness Framework (CAF) that can guide organisations on how to better measure and manage human-centric cybersecurity moving forward. In addition, an improved understanding around the persuasion techniques most likely to increase human vulnerability, as well as findings around the impact of several interventions currently being utilised to persuade end-users to behave in ways that counter that vulnerability. Together, these outputs provide a more holistic understanding around the employee experience in cybersecurity, the challenges they face, and recommendations for future intervention.

Following a critical literature review chapter (including multiple models and theories) concerning the background and context to this PhD thesis, two empirical chapters follow presenting a total of eight studies (N = 2055):

1. The first empirical block identifies a Cybersecurity Awareness Framework that combines *six key vulnerabilities* that appear to indicate higher susceptibility to risk

including threat appraisal, information security self-efficacy, information security awareness, information security attitude, information security operation policy and cybersecurity experience and involvement.

2. The second empirical block improves understanding around *additional external persuasive factors* that are being used to further influence human decision-making including key cybercriminal social engineering techniques and the usefulness of number of current protective interventions.

Within the final chapter, a general discussion is presented consolidating the findings of both empirical blocks, to provide a novel and concise overview of the human experience in cybersecurity. As well as recommendations around a number of key interventions that can be actioned and tested as a result of the Cybersecurity Awareness Framework (CAF), findings will be discussed in relation to both offender and defender persuasive communications and how to move forward. The key objective being, to furnish organisations with a more holistic picture around the human experience in cybersecurity and its associated vulnerabilities, a framework to assist in measuring susceptibility and recommendations for future intervention. Helping support employees towards becoming a stronger human line of defence for their organisations from today.

Keywords: Decision-making, cybersecurity, behaviour change, infosec, cybersecurity awareness

Contents

Tables and Figures

Tables

Figures

General Introduction

Since the creation of the first electronic computer by Alan Turing in 1946, and the inception of the world wide web by Tim Berners-Lee in 1993, technology has become increasingly more integrated into everyday human life. Across the last 30 years, developments in connectivity have witnessed society becoming increasingly more comfortable shopping online e.g., Amazon, (established c.1994), searching for and exploring information e.g., Google (est. c.1998) and Wikipedia (est. c.2001), communicating with friends through social media e.g., Myspace (est. c.2003) and being able to do all of this from one small handheld mobile device (Versus, 2020). Online connectivity has matured into the internet of things (IoT) whereby even transportation, security, health and home appliances can be monitored and managed remotely.

Across many parts of the world, humans are now able to conduct most activities online and from the safety of their own home, witnessed most markedly within the Covid-19 pandemic. During this time, many employed adults were able to conduct work remotely and people of all ages were able to maintain contact with loved ones and order online shopping, whilst remaining as safe as possible in isolation. Many organisations were able to continue service through a work-from-home programme where employees could access company systems and data remotely. However, despite technology providing humans with the ability to work, shop, exercise and socialise from almost anywhere around the world, it has also provided the growing opportunity for largely anonymous and remote cybercriminals to gain access to end-user finances, technology, and personal data (Li & Liu, 2021). Such attacks have resulted in the need for people and organisations to find ways to better protect their information and systems online, with society therefore focusing now more than ever on the enhancement of cybersecurity.

Cybersecurity is largely defined as the processes used to protect the theft or damage of sensitive information online, with a cyber-attack described by the National Institute of Standards and Technology at the U.S. Department of Commerce as attempts to destroy, corrupt or deny access to information or information systems (NIST, 2023; Taherdoost, 2022). Cybersecurity is deemed a way to set standards that indicate whether an information system is compliant with company, national or international standards and is therefore secure (Taherdoost, 2022). Much complexity has been witnessed in the public domain as well as across industry and academia in reference to the meaning of cybersecurity. This is largely due to the term being used interchangeably with other similar expressions, such as computer security, cyber security, digital security, information technology security, and information security (NIST, 2023; Wikipedia, 2023; Von Solms & Van Niekurk, 2013). Even further confusion is added when considering its multiple forms of spelling – Cybersecurity / Cyber security / Cyber-security. Of all of the above variations, particular confusion is found in relation to the terms cybersecurity and information security, with both often referred to as the general protection of information - however one related to specifically to information online.

Information security was a concept originally created to describe the protection of information and the information systems in which this data is held, be it contained physically in an office filing cabinet or within a floppy disk sat inside a company computer. However, as technology has progressed, a paradigm shift has occurred whereby information has become increasingly held within some form of electronic systems such as a laptop, mobile phone or in the cloud, rather than existing in tangible form, for example on paper (Althonayan & Andronache, 2018). The digitalisation of a large portion of societal information has now led to the line between the two terms, information security and cybersecurity, becoming increasingly blurred and therefore utilised interchangeably, everywhere from social media hashtags, dictionaries, online information sources and academic papers. Confusion is also present within most organisations who have a head of cybersecurity termed CISO (Chief Information Security Officer) as well as cybersecurity awareness programmes that also educate on the more physical protection of information.

A number of academic papers have addressed the challenges faced by the incidental merging of the two terms (information security and cybersecurity) acknowledging that subtle differences do indeed still exist. For example, as previously discussed, information security focuses on the protection of information held digitally or physically, however cybersecurity references not only the protection of digital information, but with the added aspect of protecting the human (Von Solms & Van Niekurk, 2013). An example of human protection that cybersecurity considers is that of cyberbullying, often taking place within social media platforms. Despite the addition of human protection in the concept of cybersecurity, organisations within their awareness programmes rarely touch on this aspect, focusing largely on the protection of information only. Adding further confusion by often including some features of physical information security into training e.g., tailgating.

Even as of 2023, the terms cybersecurity and information security are still all too often used within the same streams in academia to describe the protection of digital information, with any psychometric instruments used within such studies referring to either term in place of the other (e.g., Akkad et al., 2023; Rohan et al., 2023; Shaikh & Siponen, 2023). This does generate concern around whether information security or cybersecurity is being measured – or indeed both. Antunes et al. (2021) put forward the view that the term information security be used to describe the protection of information, and cybersecurity the policies and procedures put in place to ensure its security. However, whilst the field awaits a full review of the most suitable terminology and how best to utilise it moving forward, what is clear is that the generalised digitalisation of information in society has blurred the line between the two concepts. This thesis therefore takes the view of Antunes et al. (2021) where cybersecurity is used as an overarching term to describe the processes and behaviours required to ensure information security *online*. As the work within this thesis is focused on cybersecurity within industry, cyberbullying receives no emphasis, with the priority being to investigate human behaviour in protecting company, employee and customer information albeit it all now held in cyberspace. Future work must be conducted to determine which term should now be utilised to describe the protection of human information online, in a far more digitalised world, and commitment to future academic publications adhering to the agreed wording in order to streamline research.

As society have become more reliant on the cyber world and conscious to the consequences of attack, governments, associations and organisations have become more cognizant to the important role employees play in mitigation. Each year, more time and budget is being applied to human-related interventions in the hope to better protect companies from the potential devastation cyber-attacks can cause. In 2023, it is predicted that the total cost of cybercrime will hit US$8 trillion, rising to a predicted US $10.5 trillion by 2025 (Techtarget, 2023). This is in unison with no significant reductions in cyber-attacks involving the human element taking place over the past 10-20 years, despite ongoing intervention attempts (Verizon, 2021, 2022). There are a number of indications as to why this may be the case, such as awareness training often taking a one-size-fits-all approach, with a scope not tailored to the needs of individuals (Aldawood et al., 2019 ; Alshaikh et al., 2018; Bada et al., 2019; Scholl et al., 2018; Skinner at al., 2018) leaving many organisations unclear on what they need to do to improve this alarming statistic.

As society becomes increasingly alert to the need to protect and secure information within cyberspace, government bodies around the world are generating control mechanisms to ensure organisations take ownership around the safety of their business and its staff. The Data Protection Act (2018) is the UK application of the European General Data Protection

Regulation (GDPR) that states that those responsible for utilising personal data are also held accountable for its protection. This includes the mandatory need to provide GDPR training to all employees that handle such sensitive personal data. However, despite such training, cybersecurity challenges continue to persist at the same level, with a lack of significant changes to mitigation despite numerous attempted interventions (Techtarget, 2023; Verizon 2022).

It has to be questioned, whether it is fair to place the onus of employee behaviour on organisations, when they have little understanding around how their employees actually experience cybersecurity. To date, a number of technical solutions have found some success, such as the filtering of phishing emails (Dada, 2019). However, such solutions are currently limited by the need to be continuously programmed and reprogrammed using human expertise, with a current focus on machine learning algorithms to reduce this requirement (Karim, 2019). Complexity also exists due to the everchanging landscape of cyber-attacks, with continual developments in hardware, software and artificial intelligence heavily impacting the realisation of such technical accomplishments. A recent example being the use of The Chat Generative Pre-training Transformer (ChatGPT) to generate human-like communications, such as emails, for malicious ends (Alawida et al., Renaud et al., 2023). The result – humans remain ultimately responsible for the success or failure of most cyber-breaches with ~82% of attacks reported to be a result of intentional or unintentional human action (Verizon, 2022). It has never been more vital to get to the root of why current security education, training and awareness programmes are not very effective within organisations, in order to generate interventions that can result in reduction of attack success. In order to do so, it is first vital to gain understanding around the holistic experience of the human in cyberspace and their interaction with cyber-risk.

Aims

The main aims of this PhD are twofold:

1. To investigate how humans experience cybersecurity and those aspects of a human (more specifically employees) that might leave them vulnerable to falling victim to attack attempts (Chapter 2);
2. To explore how factors external to humans may further influence their experience. First by exploring the most potent social engineering techniques utilised by cybercriminals in phishing emails to manipulate vulnerability, as well as the impact several interventions can have on steering behaviour away from susceptibility.

Key outcomes from this research include a cybersecurity employee vulnerability tool that organisations can use to measure how humans are experiencing cybersecurity within their organisation, allowing them to apply more tailored interventions – e.g., moving away from those that are one-size-fits-all – that are known to be ineffective. In addition, a set of experiments are presented that detail the manipulative influence of employee decision-making biases, both for the good and bad. Together, these outputs should provide organisations, government bodies, academics and others, with a greater understanding of the human experience in cybersecurity and whether it is fair or indeed shrewd, to continue to lay fault or culpability at the level of individuals – e.g., employees.

Chapter One: Research Context and Key Questions

Chapter 1 provides a comprehensive analysis of the systematic and idiosyncratic risks humans face in relation to cybersecurity. It highlights the importance of providing organisations with a holistic overview of the intentional and unintentional human behaviours that can result in a cyber-attack. Also presented are the key questions to be answered in the subsequent chapters and overall empirical aims.

Chapter Two: Human Cybersecurity Vulnerabilities

This marks the start of the empirical portion of this thesis, by investigating what makes humans generally, and more specifically, susceptible to succumbing to a cyber-attack attempt(s). It explores the potential underlying decision-making strategies that may result in these vulnerabilities and how it may be possible for organisations to better support their employees moving forward. This investigation brings together the broad quantity of psychological, sociological and behavioural economics research currently available in relation to behaviour change models, decision-making theory, acceptance of technology theory as well as individual factors that have previously been suggested as relating to cybersecurity behaviour. The purpose of this novel and thorough theoretical review, is to provide a sound overview of the employee experience in cyberspace and how humans may succumb to attack. The output being, a human-centric cybersecurity measure of vulnerability that organisations can use to help better understand, support and manage workforce susceptibility.

The breadth of research currently available that can be used to inform organisations on human-centric cybersecurity appears complex, with the number of potential factors believed to inform susceptibility perhaps a result of overfitting. This likely renders employers unable to comprehend how to identify risk in their particular organisation, with therefore little hope of alleviating the problem. The objective of the three studies in Chapter 2 are to therefore bring together a repository of models and factors pertinent to human behaviour, to create a vulnerability assessment tool and subsequently - a parsimonious set of metrics that organisations can use to assess, potentially control and then reassess cybersecurity risk to humans moving forward.

Chapter Three: Human Vulnerability Exploitation and Mitigation

The aim of this second empirical chapter, is to explore how the employee experience can be further influenced by external factors (including cybercriminal manipulation and persuasive debiasing interventions), that have the purposeful aim of impacting cybersecurity vulnerability, both for the good and the bad. In conjunction with the previous chapter, a more holistic overview can be provided to organisations in relation to the employee experience, and perhaps why challenges within human factors, in relation to cybersecurity, may not be as simple as previously assumed.

Human Cybersecurity Vulnerability Exploitation

Chapter 3 begins this shift in perspective, by investigating how cybercriminals look to further manipulate human decision-making heuristics, to increase the likelihood of a successful security breach. The studies within this chapter explore the top social engineering strategies being utilised by cybercriminals in particular relation to phishing emails (emails sent with the distinct intention to manipulate humans into revealing sensitive information or installing malware), as well as the impact they may have on the employee. Many offenders are aware of how humans make decisions, associated cognitive biases and how they can leverage on such error to increase the chances of end-users following their command. It is not enough for organisations to focus only on the human risks inherent in their employees, but also how offenders then further manipulate known vulnerabilities for personal gain. By investigating both defender vulnerabilities and offender strategies a holistic picture can be painted with organisations better guided on the support they need to offer their employees to encourage reductions in risk.

Interventions to Mitigate Human Cybersecurity Vulnerability

Chapter 3 also considers a number of interventions that are being used within the literature and organisations, to help positively influence the employee experience by targeting human decision-making vulnerabilities at their root cause. This set of experiments investigates how useful real-time soft-paternalistic nudging (the subtle guiding of choice), motivation education and adapting cognitive strategies (the development of new habitual behaviours) are at supporting employees whilst interacting with emails in a corporate mailbox. The aim being, to understand whether such interventions are useful in positively influencing the employee cybersecurity experience, by supporting the reductions in reducing attack success, or they are in fact serving as an additional burden. Should this be the case, more innovative solutions are now required to mitigate the number of damaging cyber-attacks being experienced but with the human experience at the centre.

Chapter Four: General Discussion

Drawing upon the themes and findings from Chapters 1 to 3, Chapter 4 presents a comprehensive overview of the investigations undertaken and findings and what they indicate about the future of the human and cybersecurity with employee vulnerabilities, offender strategies and proactive interventions in mind. A critical and novel summary is provided around how the human experiences risk in cyberspace with a forward focus on how the strengths and vulnerabilities of both the human and technology now need to be brought together if true strides in attack mitigation are to be witnessed. The overall objective of this thesis being, to furnish organisations and academics with a holistic look into the world of the human and cybersecurity and whether there is hope for improving the current risk to humans (specifically employees within organisations) both systematically and dynamically towards mitigation.

Chapter One: Research Context and Key Questions

Since the inception of the world wide web in 1993, the integration of technology into every aspect of human life has progressed at a galloping pace. The number of appliances and systems at home, work and otherwise (e.g., connected transport, wearable health monitors, connected home heating, self-scanning in shops, security biometrics) that are connected to cyberspace grow in number each year. Despite the benefits offered by technology and connectivity, for example improvements in healthcare, communication, speed of production and more (Laplante & Laplante, 2016), cybercriminals are left able to access the assets of others remotely and often anonymously in an online world. Over 23,000 security incidents were reported in 2021 alone, with a human element believed to be a key driver in around 82% of successful breaches (Verizon, 2022). During this same year, phishing remained one of cybersecurity's top threat actions with over 80% of organisations experiencing an attack, with phishing still regarded a top threat type in 2023 (Techtarget, 2023). A number of technical solutions have been applied to support reductions in the risk this level of connectivity has generated, yet cybercriminals (and new technologies made available to them) continue to find increasingly sophisticated ways to evade detection, leaving humans persistently responsible for the success or failure of a large number of attacks (Verizon, 2022).

Over recent years, organisations have become more sympathetic to the burden employees face in relation to cyber-attacks, with many investing time and budget into at least some form of compliance training. However, cybersecurity education and awareness schemes often provided within organisations are reported as technical, off-the-shelf programs that many employees find difficult to transfer or find unrelatable to their working day (Scholl et al., 2018). The lack of lustre in respect of cybersecurity awareness training is perhaps the result of perplexity around how organisations even begin to manage human cybersecurity risk from a psychological and/or human factors perspective. It is therefore important to provide stakeholders and decision-makers with guidance around how to reduce cybersecurity human risk, in a way that is accessible and directly usable. In order to do this, it is important to first recognise the internal challenges humans face in cybersecurity, to ensure intervention is focused around supporting those particular vulnerabilities both correctly and sympathetically. This includes investigating the potential influence of a number of socio-psychological factors (e.g., personality) perceptual factors (e.g., threat appraisal), habitual factors (e.g., experience and involvement) and socio-emotional factors (e.g., intrinsic maladaptive rewards) that previous literature suggests being of most concern.

Individual Differences and Decision-making

Whilst at least some current cybersecurity awareness interventions may take into account a small number of human individual differences that impact how they experience cybersecurity, none are able to simultaneously account for the multitudes of potential characteristics and perceptions humans possess that are believed to correlate with their decision-making during cyber tasks (such as found in e.g., Egelman & Peer, 2015; Gratian et al., 2018; McGill & Thompson, 2017; Posey et al., 2015; Safa et al., 2015). Examples of relevant individual differences include - a person's belief on the probability of a cyber-attack, their level of impulsivity, and possibly their level of commitment to an organisation (Egelman & Peer, 2015; Meyer & Allen, 1991; McGill & Thompson, 2017). The array of factors potentially involved, presents a challenge for organisations having to choose the most cost-effective risk areas to target (i.e., it is unlikely that all factors can be considered). It is therefore important to improve understanding around the key human-centred cybersecurity metrics and underlying measures organisations can use to drilldown to cybersecurity human risk within their business and how it is experienced by employees.

Coinciding with the above, is the absence of an effective approach to measuring cybersecurity human vulnerabilities, and therefore recommended interventions to deploy (see

Alshaikh et al., 2018). It is important that organisations better understand the human-centric cybersecurity experience and resulting metrics they might need to generate, in order to actively inform the correct choice of control mechanism. These metrics need to measure those individual differences and perceptions most likely to influence behaviour, so that organisations can more effectively apply time and budget, tailored to the needs of an individual or group of individuals. A key aim of the current research is to therefore better understand what makes humans vulnerable to a cyber-attack, as well as how cybercriminals may attempt to prey on these vulnerabilities and what can potentially be done to support end-users with such cognitive constraints in mind. As well as the systematic and more idiosyncratic risks employees may pose, offenders also dynamically perpetuate this risk by manipulating employees using social engineering techniques. This can include methods of persuasion and other ancillary tactics to further encourage intuitive thought and the decision-making errors that can occur as a result (Luo et al. 2013; Williams et al. 2018). An example could be the offender presenting themselves as an expert, perhaps the CEO of an organisation, preying on the human heuristic to obey those in authority. Therefore, a subset of the research within the current thesis will investigate not just intrinsic human vulnerability, but also the ways in which cybercriminals target end-users and how they further leverage on human decision-making constraints. This will be addressed in the third chapter of this thesis, across two studies.

Current Challenges in Relation to Cybersecurity Mitigation

Several factors have been suggested as accountable for the lack of success in cybersecurity interventions, including the use of content that is too technical, not easily transferable and either too wide or too narrow in scope (Alshaikh et al., 2018; Bada et al., 2019; Scholl et al., 2018; Skinner at al., 2018). Further challenges that have been ascribed include constitutional barriers such as insufficiency of government support, business related obstacles such as budgetary constraints, and personal challenges for example employee individual differences and work-based pressures (Aldawood et al. 2019).

Many current interventions also rely heavily on educational activities, in the hope that employees will be able to consciously apply learnings during their often busy working day. However, humans are believed to process the majority of decisions unconsciously, and therefore it may be unrealistic to expect employees to recall large amounts of training knowledge to inform or assess each action they take (Bargh et al. 2001). In addition to this concern is the reality that employees are often multi-tasking and working under pressure, making it unreasonable to expect constant recollection and application of educational information when cognitive load is high.

During unconscious thought, humans apply a number of pre-determined cognitive strategies to help them arrive at a quick decision that whilst not always optimal, are often sufficient (Ceric & Holland, 2019; Kahneman, 2011). For example, intuitively choosing an unhealthy dinner option at a restaurant because an attractive photo of it appears on the menu. Cybersecurity decisions are largely made utilising the same sub-optimal, but productive decision-making processes. Therefore, to achieve some level of mitigation to human-related cyber breaches, interventions need to take into account how humans process decisions and the limitations this may bring to the interventions being applied. The thesis will therefore consider ways in which organisations can better support employees when in a more quick and intuitive mode of thought. This will also initially be investigated in the third chapter focused on external influencers, across three experiments.

Summary: Key Research Questions

With modern advancements in technology and a heavier reliance on interconnectivity, never has it been more vital to explore how humans experience the cyber world and its security, in order to find improved solutions for protecting organisations from cyber-threat. With humans, either intentionally or unintentionally, at the centre of most cyber breaches, focus needs to be shifted towards understanding their holistic experience in cyber space, and what makes them so vulnerable to attack.

It is important to further investigate the specific individual differences and perceptions most likely to lead to employee susceptibility, to allow organisations the opportunity to gain learnings around how their employees may be experiencing the situation. By better understanding the vulnerabilities of concern, such constructs can be more easily measured and managed by organisations with a clearer roadmap towards intervention. It is also imperative to explore the tactics cybercriminals are currently using to further limit employee ability to detect a cyber-attack, to ensure any intervention applied considers, holistically, what is actually within employee's capabilities. Finally, researchers must explore additional ways to support employees in relation to cyber-attacks outside of the current, largely educational, intervention practices that are perhaps not particularly effective during more intuitive decision-making. As discussed in later chapters, at the heart of change is active application of knowledge with this PhD focused on utilising the work of academic forefathers and contemporaries to inform current and future research, to encourage more effective change (Cox et al., 2006; Markey & Townsend, 2013).

The aim of the two empirical chapters within this thesis, are to build upon current research in relation to human decision-making vulnerabilities in cybersecurity, assembling findings into a novel and comprehensive format that details the all-embracing relationship between humans and cybersecurity, interventions that may aid success, and thoughts around future innovations that can be investigated now to make a difference to the cyber-attacks of tomorrow. Key aims therefore include:

1. Investigations into *why* employees (and other population samples) might be vulnerable to cyber-attacks;
2. Exploration into *how* cybercriminals further leverage on these vulnerabilities;
3. Explore *what* can be done (interventions) to better support employee vulnerabilities (and others) moving forward.

The resultant recommendations, include a human-centric cybersecurity framework for organisations to use, to understand the vulnerabilities employees are experiencing within their organisation, a benchmark for understanding cybercriminal exploitation of human vulnerabilities, and what (if any) interventions can be useful in reducing human decision-making error and therefore the probability of cyber-attack success. With all findings helping inform the question – *How does the human experience cybersecurity*.

Chapter Two: Human Cybersecurity Vulnerabilities

Chapter Summary

With society now heavily invested in cyber-technology and most cyber-attacks believed to be due to human error, it is imperative to now focus research on human-centric cybersecurity vulnerabilities in order to better tailor future intervention based on the most important human risk factors. Whilst a number of studies have investigated cybersecurity behaviour in relation to end-user individual differences, e.g., gender and personality (Egelman & Peer, 2015; Gratian et al., 2018) and factors within behaviour change theory such as threat appraisal and self-efficacy (Thomas., 2018; Yang et al., 2020), the following three studies are the first to bring together such a comprehensive inventory of human factors, in the hope to generate a parsimonious human-centric cybersecurity framework, useful for not only industry, but individuals and academics alike. This chapter improves understanding around why humans are vulnerable to cyber-attacks but also what can be done to reduce the number of security breaches experienced within this domain, as well as potential innovations for the future.

Across three studies, five-hundred and fifty-three participants completed a battery of questionnaires within a number of themes including socio-psychological factors (e.g., personality) perceptual factors (e.g., threat appraisal), habitual factors (e.g., experience and involvement) and socio-emotional factors (e.g., intrinsic maladaptive rewards) to understand which can be useful in predicting cybersecurity behaviour. Exploratory correlation analyses from Study 1, exploratory factor analysis and regression analyses within Study 2, and a further regression analysis in Study 3 help refine a large set of human related metrics into a manageable and predictive framework that can be used to measure human strengths and vulnerabilities in relation to cyber-attacks across organisations and provide guidance for intervention.

Introduction

Over the last 30-40 years or so, organisations have become increasingly reliant on the benefits that computer systems bring to business and its processes. The internet now affords seamless communication, increased productivity, and even more efficient information sourcing. However, alongside these benefits comes a cost – in 2021 alone, around 23,896 security incidents were reported to have taken place within organisations across the globe (Verizon, 2022). In the UK alone, the National Cybersecurity Centre (NCSC) provides support to around 15 cyber-attacks on UK organisations each week, including critical national infrastructure where a breach of security has the potential to cause severe and widespread disruption across the UK and beyond (NCSC, 2021; NCSC, 2022). Despite the great benefits this growth in technology has offered organisations, attacks towards online data and system integrity are continuing to evolve in both number and level of intelligence. This is despite the efforts of many organisations and academics working hard to create hardware and software that can protect end-users and security teams around the world. Perhaps due to this, Gartner (2023) have predicted high attrition in cybersecurity roles over the next few years, resulting in a skills shortage by 2025 due to burnout and low morale within the industry. It has therefore become time to place even more focus on aspects of the human that make them susceptible to attack, and what can be done to help mitigate them.

To date, the field of cybersecurity has largely focused its research on more technically orientated interventions, such as the analysis of aggregated logs and system monitoring. However, more attention is required around addressing end-user vulnerabilities and developing preventative human-centred security solutions (Verizon, 2022). Cybercriminals are finding increasingly sophisticated ways of bypassing technical efforts to reach end-users, allowing them to then take full advantage of human decision-making and processing constraints. Indeed, analyses suggest that ~83% of security breaches involve a human (entry point) element, with a common example social engineering where an attacker employs psychological manipulation to encourage people to e.g., click on malicious links, download malevolent attachments or reveal personal information (Ghafir et al., 2018, Verizon, 2022). Social engineering techniques can be deployed by cybercriminals physically, or by utilising various forms of synthetic media, such as email (phishing), telephone (vishing), text message (smishing) and more recently video technology (e.g., deep fakes). Phishing was believed to be the threat action that resulted in the majority of cybersecurity breaches involving the human in 2021 (Verizon, 2022). Despite the significant role the human plays in the potential success or failure of a cyber-attack, there remains a dearth of research relating to human-centred aspects, with better understanding required around what drives human cybersecurity behaviour and how to better protect them. By identifying both the systematic and individual differences at the root of human cybersecurity vulnerabilities, better tailored interventions can be devised to compliment the technical tools currently available and those being developed for the future.

Over recent years, largely as a result of the deployment of GDPR, the number of UK companies offering some form of security education, training and awareness (SETA) has increased. However, despite these efforts, human related security breaches have not abated, making it imperative to understand why current awareness interventions are failing to make significant improvements in security behaviour (Verizon 2022). Research can then become focused on the underlying human attributes and key drivers currently perpetuating risk within cybersecurity, in order to deliver more targeted and transferable interventions to those who require the most support, in the most needed areas at the right time.

A number of factors appear to be responsible for SETA’s current lack of success, including the use of highly technical content that employees find difficult to transfer into their working day (Alshaikh et al., 2018; Bada et al., 2019; Scholl et al., 2018; Skinner et al., 2018). However, of all the associated barriers, the 2021 SANS Security Awareness Report found shortages in time and resource applied within business to be the biggest challenge security awareness professionals face when trying to improve cyber interventions (SANS, 2021). This finding again suggesting focus must be applied to identifying the human vulnerabilities of most concern, in order to more effectively apply such limited budgets and time constraints.

The individual differences each human may experience in cybersecurity, provides a particularly robust barrier to the success of SETA programmes, due to the multitudes of characteristics and perceptions humans possess. This leaves organisations with the challenge of having to prioritise a certain number of individual differences over others, perhaps those most cost-effective, rather than the factors that will result in the largest impact. When examining the literature for guidance around the factors believed to influence behaviour, in particular cybersecurity behaviour, there are an overwhelming number of models organisations need to consider (Michie et al., 2014), despite many being descendants of the same original theorem or containing similar factors. Research in cybersecurity has therefore branched out in pockets, focused on each of the different models causing confusion and lack of consolidation. Safa et al. (2015) take some strides towards adjoining a number of these theories – Protection Motivation Theory (Rogers, 1975) and the Theory of Planned Behaviour (Ajzen, 1991) to investigate their predictive power in the cybersecurity domain, finding a number of key factors from both theories useful in understanding cybersecurity behaviour. Alongside this challenge is the large number of individual differences or socio-psychological factors not featured in behaviour change theory that are also suggested as related to or predictive of cybersecurity behaviour such as personality sub-types, risk-taking preferences and decision-making styles (e.g., Egelman and Peer, 2015; Gratain et al., 2018).

A key aim of the research presented in this chapter, is to therefore bring together key constructs from a number of leading behaviour change theories in the cybersecurity domain as well as several individual differences that have previously been indicated as related to or predictive of human cyber risk. The principle objective being, to provide organisations and awareness leads with a more simplistic view of the key underlying factors that help explain human experience, and therefore behaviour, in relation to cybersecurity (Jeong et al., 2019). By doing so, organisations will have available a more manageable set of metrics with which to focus intervention and measure success, a task not previously undertaken.

In order to begin exploration into the possible individual differences influencing cybersecurity behaviour, it is important to obtain an in depth understanding of the theory underpinning human behaviour and what may be required to influence and support positive change. The next section will review a number of key theories in relation to behaviour change, as well as other factors that have been identified in the literature as perhaps predictive of how end-users experience and interact within cybersecurity.

Demographics and Individual Differences

A number of user demographics and key individual differences should first be discussed in relation to cybersecurity behaviour, to provide an initial backdrop for the more innate aspects of the human that can influence how they may act. Whilst it is not always possible to manipulate the very basis of these factors e.g., someone’s age or how they naturally respond to decisions, knowledge of differences or correlations can help identify interventions that may reduce the impact these aspects can have and help support more secure behaviours moving forward.

In relation to user demographics, factors such as age and gender have been notably investigated in regard to cybersecurity behaviour with largely confirmatory findings. In 2009, Parrish, Bailey, and Courtney conducted correlation analyses between participant age and susceptibility to phishing activity, finding those aged 18-25 more at risk than other age groups. In 2010, Sheng et al. conducted a demographic analysis of phishing susceptibility and effectiveness of interventions, with outcomes also suggesting those aged 18-25 are more susceptible to phishing attacks than other age groups, as well women more susceptible than men. More ‘cyber-risky’ behaviours were also reported in those younger in studies by Whitty, Doodson, Creese, and Hodges (2015). These findings were confirmed again in Gratian et al. (2018) when conducting correlational investigations across 369 university staff and students, exploring relationships between user behaviour and age and gender, utilising the Security Behaviour Intentions Scale (SeBIS). The SeBIS consists of four security behaviours including device securement, password generation, proactive awareness, and updating. This study revealed that whilst age did not have a significant unique effect on the regression model, those aged 18 - 25 were found to generate weaker passwords than other age groups. Gender to be a unique predictor of more secure cybersecurity behaviours across all four measures, with women more susceptible to cyber-attacks than men.

The gender differences between women and men in cybersecurity are perhaps due to men, in general, perceiving themselves as having higher computer self–efficacy and general resilience than women, and therefore higher perceptions of ability increasing confidence (Anwar et al., 2017; Branley-Bell et al., 2022; Gratian et al, 2018). This is perhaps in some part due to the under-representation of women in IT and other STEM subjects, resulting in less exposure to topics relating to technology and gender asymmetry (Kshetri & Chhetri., 2022). However, a study by Fatokun, Hamid, Norman, and Fatokun (2019) whilst also finding a gender divide, found men to be particularly more susceptible to phishing attacks in the banking domain. Some variations in findings in relation to age, have also been uncovered in a more recent study by Branley-Bell et al (2022) in information and communication technology (ICT) cybersecurity behaviour across 579 participants. Age was found to be a significant negative predictor, with older users again found to be more secure than those younger, in respect of creating strong and secure passwords, however, those younger were predicted to be more likely to secure their devices. Additional studies have also found older adults to feel neither motivated nor capable in relation to cybersecurity (Morrison et al., 2021; Whitty et al., 2015). Despite some alternative findings to demographic differences in cybersecurity behaviour across age and gender, previous literature does suggest the idea that both women and those younger may require more support to ensure they remain secure in relation to cyber risk.

Egelman and Peer (2015) investigated how a number of additional human individual differences in relation to cybersecurity, such as risk-taking attitude, decision-making strategy and level of impulsivity, influenced cybersecurity behaviours in computer users. Less desirable cyber behaviours were found in those participants who were rated as more impulsive, more likely to take health/safety risks (e.g., drive under the influence of alcohol) and procrastinate or rely upon others when making a decision. The negative relationship between impulsivity and cybersecurity behaviour particularly, has been found within several studies (Aivazpour & Rao. 2018; Hadlington 2017; Parsons et al., 2013). This is perhaps due to a lower ability to process the contextual features required to detect cyber threat when reacting more rapidly (Jeske et al., 2016).

Gratian et al. (2018) built on the findings of Egelman and Peer (2015), by again investigating risk-taking attitude and decision-making style within an educational setting, as well as how gender and personality relate to cybersecurity behaviours. Regarding decision-making, a more rational processing style was linked to positive cybersecurity behaviours and a spontaneous style more negative. This differs from the study by Egelman and Peer (2015) where only avoidant decision-making was found to relate to behaviour. Gratian et al. (2018)

also found risk-taking attitude to be a good predictor of cybersecurity behaviour, with those taking financial risk generating stronger passwords and those more likely to take health/safety risks creating weaker passwords. Egelman and Peer (2015) also found that those more likely to take health/safety risks, reported fewer desirable behaviours. A number of contrasts within these findings do suggest further research is required to understand, particularly within a work-based setting, the constructs of most concern.

Models of Behaviour Change

There are a wide number of behaviour change models and theories – many of which are still under empirical investigation in relation to human-centric cybersecurity, creating a challenge for organisations eager to effectively educate and support their employees against cyber-risky decision-making. Organisations, as well as academic researchers, require a more clear and concise view of what factors they should focus on, to limit breach susceptibility in humans whilst both technical and psychological research continues to evolve. Examination of current literature provides a substantial number of individual differences and factors from behaviour change theory, and beyond, that therefore require further investigation in order to inform e.g., the development of a huma-centric cybersecurity assessment framework.

Protection Motivation Theory (PMT) is a particularly influential behaviour change model, originating within the health domain, that suggests two appraisal systems take place when assessing threat: (1) a threat appraisal, whereby the probability and severity of the threat is considered, and (2) a coping appraisal (McGill & Thompson, 2017). Coping appraisals are whereby judgements are made on how effective a response will be (‘response efficacy’), how effective end-users believe they will be in applying the response (‘self-efficacy’) and the associated costs to its application (‘response costs’). The outcome of these appraisals are suggested to influence the intention to adopt the behaviour required. For example, if a human perceives risk of threat to be low, and chance of response success to be low, then motivation to complete the behaviour will likely deplete (Rogers, 1975). Many studies have been conducted using and in relation to PMT in the cybersecurity field, largely around the deployment of fear appeals and/or coping messages (further analysed in Chapter 3) when considering the influence of external messaging on the employee cybersecurity experience. Fear appeals are messages focused on communicating the probability and severity of threat with the aim of increasing threat appraisal and coping messages providing information on how to remain secure and therefore improving coping appraisals. Research to date has found both increasing threat appraisal and coping appraisals to be effective in improving cybersecurity behaviour, however it appears perhaps coping messages more so than fear appeals (Shillair & Dutton, 2016; van Bavel et al., 2019). The application of both messages together does however appear to provide the most potent form of influence as well as believed to be more ethical (Dupuis & Renaud, 2021; Witt & Allen, 2000). The following paragraphs provide a deeper elucidation of the factors found within PMT and a number of psychological constructs identified as related.

Threat appraisal. Threat appraisal, defined as both the perceived probability of threat by the end user, as well as their estimation of harm should a cyber breach take place (McGill & Thompson, 2017), is the first factor within PMT to be discussed. If an employee perceives threat to themselves or their company to be high, they are more likely to be motivated to perform behaviours that will actively maintain protection.

When appraising cyber-threat, end-users often perceive risk to be lower than actual threat. Explanations around this appraisal include individuals believing themselves not to be of importance to hackers and not caring if their privacy is violated as they feel they have nothing to hide (Jones et al., 2021). Several human decision-making biases are presented within the psychological research that may help to explain why humans appraise threat as lower than actual risk, the first being the availability bias. *The availability bias* manifests as an inaccurate perception of the probability of an event occurring, determined by how readily relevant instances can be brought to mind (Taylor-Gooby & Zinn, 2006; Tversky & Kahneman, 1973). Should employees be shielded from security breaches befalling their organisation, they will not have such examples available to recall from memory, ultimately left assuming such occurrences are rare. As an example, humans often believe plane crashes or shark attacks to be more common than they are due to their coverage within media. Interventions must focus on ways in which to increase perceptions of breach probability ensuring threat is readily observable, possibly through regular updates delivered via posters, newsletters or computer pop-ups.

Often coinciding with the availability effect is that of saliency, whereby people are more likely to focus on prominent information than information that is more subtle, for example news stories that contain violence and suffering (Schenk, 2011). Further research is required around how to render cyber threat information as more available and salient. However, limited findings to date do suggest that employees will more readily recall breach details that they have verbally spoken rather than silently read (Tversky & Kahneman, 1973), imagined with clear and concrete detail (Carroll, 1978), and instances where more vivid information has been provided (Fontenelle & Howell, 1984).

Choice architecture utilises psychological principles to design choice for humans guiding them towards more positive decisions. Libertarian (or soft) paternalistic nudging, can possibly help increase threat appraisal by delivering contextual and salient breach examples. Nudging in relation to threat appraisal and how this is experienced by the human is further investigated in Chapter 3 of this thesis.

Another possible way to increase the saliency and availability of breach examples to improve threat appraisal is via the affect bias (Kahneman, 2011). The affect bias is whereby a decision is made based on emotion as opposed to rational thought, even when the emotions felt are not relative to the decision being made (Loewenstein & Lerner, 2003; Pfleeger & Caputo, 2012). For some time, decision-making was viewed as a purely cognitive process based on utility, however more recently the impact that affect can have on decision-making has moved front and centre. Human affect can impact a decision in one of two ways; the anticipated emotion should an action be chosen, and the immediate emotions experienced in relation to the decision including any irrelevant characteristic or environmental feelings e.g., happiness when the sun is shining (Loewenstein & Lerner, 2003). Previous research has highlighted the use of emotion in increasing human perception of risk, particularly in relation to fear (Keller, et al., 2006; Loewenstein et al., 2001, Peters et al., 2004; Pfleeger & Caputo, 2012; Slovic et al., 2002). Fear appeals are often used to reduce unrealistic optimism and increase appraisal of threat, resulting in humans displaying less risky and more conscious behaviours. How humans may experience this in cybersecurity is an idea investigated in Chapter 3 of this thesis.

Fear is an emotion characterised by negative valence and high arousal that results in the cognition of threat, motivating people to either escape or avoid a potentially harmful situation (Rogers, 1975; Witte & Allen, 2000). Research to date around the use of fear appeals to increase perceptions of risk have been mixed, however a number of meta-analyses undertaken across the years do show support for their application in increasing perceptions of susceptibility and severity (Lowry et al., 2023; Tannenbaum et al., 2015; Witte & Allen). Recent research has found fear appeals work better when using concrete examples, that include coping strategies, and when psychological ownership exists around the object or event. However, this proves a challenge when cybersecurity is a secondary task and attention limited (Briggs et al., 2017; Dupuis et al., 2021; Dupuis & Renaud, 2021; Schuetz et al, 2020). Some ethical concerns have been raised around the inducement of fear in relation to risk, with the provision of coping strategies an addition that may alleviate some of these concerns (Dupuis et al., 2021; Dupuis & Renaud, 2021). Immersive interventions such as virtual reality and perhaps augmented reality, can potentially provide powerful and affective examples employees can later easily reimagine, providing clear and concrete steps to avoiding such a breach in the future (Krupić, et al., 2021; Rosén et al., 2019). However, ethical concerns exist when looking to evoke sizeable affect.

Another potentially important aspect of threat appraisal is *the optimism bias*, whereby humans regularly overestimate personal positive outcomes and underestimate personal negative outcomes, on average in relation to others, impacting how they forecast risk (Pfleeger & Caputo, 2012; Warkentin et al., 2013). Therefore, whilst employees can be made aware of risk, it is believed that they will still underestimate it in relation to themselves and their organisation against the human average (Warkentin et al., 2013). The optimism bias is thought to have developed as an evolutionary factor to help reduce anxieties experienced during instances outside of human control, and therefore excessive reductions in optimism can potentially result in increased depression (Sharot, 2011; Weinstein & Klein, 1995). However, a small decline in domain specific optimism can support increases in the availability bias, resulting in threat appraisals that are more realistic and valuable (Arkes, 1991; Chen, et al., 2021; Weinstrin, 1980).

Across the literature, unrealistic optimism has been linked to poor threat appraisals in IT risk assessments, e-waste, and perception of risk to the coronavirus (Bottemanne et al., 2020; Chen et al., 2021; Loske et al., 2013; Warkentin et al., 2013; Shalev et al., 2014). Reducing the optimism bias is not an easy task, it is believed to be so robust that increasing knowledge around its existence will still result in people heuristically believing themselves as less susceptible to the bias itself (Croskerry et al., 2013; Jolls & Sunstein, 2006). There are however thought to be three main interventions with the potential to reduce the effect, including an accountability intervention, an insight intervention and an unambiguous intervention (Cutello et al., 2021; White et al., 2011). The accountability activity suggests end-users are made aware that they are being evaluated on their actions (even if they are not), however it is possible that this will simply reduce self-efficacy as opposed to increase perceptions of risk. An insight intervention is whereby end-users are asked to reflect on a more difficult task, reducing their optimism estimates in future activities. For example, reviewing their performance on a cybersecurity hazard perception task. Finally, an unambiguous definition intervention focuses on making the underlying factor they are judging clearer. For example, making explicit what constitutes good cybersecurity behaviour, allowing for more realistic evaluations.

Further intervention examples within the literature include - use of a similarity statement where end-users can draw comparisons to similar people, and the reframing of a security question whereby people state examples of why they might fail as opposed to why they might succeed (Arkes, 1991; Jolls & Sunstein, 2006; Soll et al., 2014). The addition of the optimism bias into a soft-paternalistic nudge (e.g., a computer pop-up suggesting the user to be at equal risk to a cyber-attack as others), looking to increase threat appraisal, is investigated within Chapter 3 of this thesis, when exploring interventions that may improve the employee experience and vulnerabilities within security. Threat appraisal and its relationship with cybersecurity is therefore investigated within this first empirical block, to better understand whether the creation of interventions that will help increase appraisals of threat within organisations will influence employee motivation to want to protect its data and systems.

Coping Appraisal. As well as PMT involving an appraisal of the probability and severity of potential threat, evaluation also takes place, around the perceived success of deploying a suggested response and the mechanisms involved in this process - *self-efficacy, response efficacy and response costs*. Self-efficacy is described as an expectancy or judgement around the skills and capabilities a person believes are required to bring about a certain course of action and whether they feel ability in relation to the response is something they possess (Maddux & Gosselin, 2012). Self-efficacy has taken on many terms over the years that in essence are all suggested to be measuring the same construct when informing within the context of cybersecurity – computer self-efficacy, information security self-efficacy, internet self-efficacy, privacy self-efficacy, coping self-efficacy, perceived behavioural control and so on across a number of research domains (see for example, Conetta, 2019; Raineri & Resig, 2020; Safa et al., 2015). Available models and frameworks to measure self-efficacy do therefore differ in how they present the construct, making it a challenge where no succinct operationalisation of the term has been found. This does lead to the potential for the jingle-jangle fallacy, whereby constructs within similarly named instruments are believed to represent the same thing, or that constructs within different sounding instruments do not (Zainal et al., 2022). For example, it assumed that all psychological instruments used to measure cybersecurity self-efficacy are measuring the same construct, but computer self-efficacy and cybersecurity self-efficacy are not. In order to eliminate the jingle jangle effect, perceived behavioural control, discussed later in this chapter, has been omitted to avoid further confusion.

Self-efficacy is believed to be determined by a biological and emotional want to master a task, as well as perceptions that the task is valuable and in itself can be effectively mastered (Maddux & Gosselin, 2012). Self-efficacy differs from ability and competency, due to its task specific focus, without consideration for factors such as cost and effort (Agha et al., 2019; van den Broeck et al., 2010). Self-efficacy is believed to be influenced by a number of human experiences, including previous practise and achievement of the behaviour, commendation of achievement by peers, and witnessing others mastering the behaviour (Maddux & Gosselin, 2012). When such perceptions around self-efficacy change, behaviour change is believed to follow. The influence self-efficacy has on a behaviour is also believed to be intertwined with response efficacy - *perception on the likeliness that a response will achieve a desired goal* (Cismaru et al., 2009). Response efficacy can be influenced by a number of factors that include social and cultural norms, for example believing response success to be less likely if conducted by a specific gender as often seem within STEM subjects (Keller, 2006). Bandura (1982) speaks of the balance between self-efficacy and response efficacy and how both must be aligned to achieve response success. For example, a human will likely not conduct a behaviour when the necessary environmental requirements are not in place, e.g., reliable firewall software, even if they feel fully confident in implementing the software. Alike self-efficacy, response efficacy is suggested to be impacted by perceptions of threat severity, with response success deemed more likely should an attack be perceived as severe, perhaps providing a larger opportunity for response success (Lewis, Watson & White, 2010).

Response efficacy is also believed to hold a multiplicative relationship with the final aspect of coping appraisal – response costs. Response costs include those factors that are aside from the skills and ability of the person and behaviour, such as the time, money and the effort it will take to make the response a success (Cismaru et al., 2009). Should reliable firewall software be available and an individual capable of installing, they may still refuse to do so if time and financial costs are not manageable. Response efficacy and response costs can be viewed as opposite ends of a continuum with response efficacy decreasing the more costs are required to conduct the behaviour (Cismaru et al., 2009). Whilst response efficacy and response costs are not as well research as threat appraisal and self-efficacy, their place within behaviour change models and relationships with other factors make it important to include them both within this set of explorative studies.

Similar to PMT, the Health Belief Model (HMB) focuses on the expectancy-value principle, whereby perceived expectation of risk and the costs (or benefits) of not taking action influence motivation to act (Anwar, 2017; Rosenstock, 1974). Despite differences in research application and model arrangement e.g., HBM offers a more hierarchical theory of behaviour change, and PMT focuses on behavioural continuums, both models have clear similarities, such as a threat appraisal factor, a self-efficacy factor, benefits to conducting the response and potential cost factors (Prentice-Dunn & Rogers, 1986). Avoidance Theory (AT), and its younger sibling, Technology Threat Avoidance Theory (TTAT) also present similar features - fear of threat - as the motivational driver for avoidance of a task, in conjunction with the perceived effectiveness of an alternative coping behaviour (Carpenter et al., 2019; Liang & Xue, 2009; Mowrer, 1939; Rachman, 1976). Whilst both the HBM and TTAT are still used within cybersecurity behaviour research, they do command less attention within the literature than the PMT. However, due to the continued use of all three measures, it is therefore important to synthesise research that focus on a combined theory, so that research can work in unison towards identifying the constructs most important in evoking positive behaviour change in cybersecurity.

An additional model often used to explain human behaviour that does provide additional constructs of interest is the Theory of Planned Behaviour (TPB; Ajzen, 1991). With intention, the motivational element that drives behaviour, central to this model, it presents three determining factors. First, that people consider their actions based on their overall evaluation of the impending behaviour (attitude). Second, their access to the relevant internal and external resources to perform the behaviour (perceived behavioural control) and finally whether significant others believe they should perform it (subjective norms; Ajzen, 1991; Burns & Roberts, 2013; Connor & Armitage, 1998; Safa et al., 2015). Perceived behavioural control is described by Ajzen (1991) as similar to Bandura’s (1982) self-efficacy construct, focused on the resources available to conduct the required behaviour. Whilst perceived behavioural control does place some additional focus on the availability of external tools and knowledge, in the pursuit of parsimony, its close similarity to the self-efficacy factor included in other models renders perceived behavioural control removed in favour of the more researched self-efficacy for the remainder of this study.

Attitude, an important aspect of the TPB, is defined as a person's general evaluation of something, such as an object or event, that in turn influences their behaviour towards it (Azjen, 1991; Conner & Armitage, 1998). Attitudes can be hidden from the world within a person's thoughts or feelings, or overtly expressed via their behaviour (Pickens, 2005), created as a result of a person's e.g., personality traits, motivations, beliefs, values (Pickens, 2005). In Fishbein and Ajzen's (1975) expectancy-value model (linked to TPB), attitudes are formulated around things, people, places or events either positive or negative in nature. With positive attitudes more likely to exist towards those things associated with better outcomes. The Elaboration Likelihood Model (Petty & Cacioppo, 1986), a theory based on the persuasion of attitude, describes how enduring these positive or negative attitudes become as a result of how high a degree of thought (elaboration) a person has given to a piece of information in relation to the object in question. Quantity of the elaboration ascribed can depend on a number of things, such as social contagion where people adopt the attitudes of those in their social group, often without awareness (Scherer & Cho, 2003). Humans strive for their behaviour to remain consistent with the attitudes they possess, or else they experience a mindful feeling of discomfort known as cognitive dissonance. For example, smoking under peer pressure despite a negative attitude towards it. People will work towards reducing this conflict in a number of ways, such as changing their behaviour or attempting to rationalise it by telling themselves that smoking is better than the use of a number of alternative illegal drugs. This vulnerability and how it can be manipulated, is further discussed within the 'commitment and consistency' portion of Chapter 3.

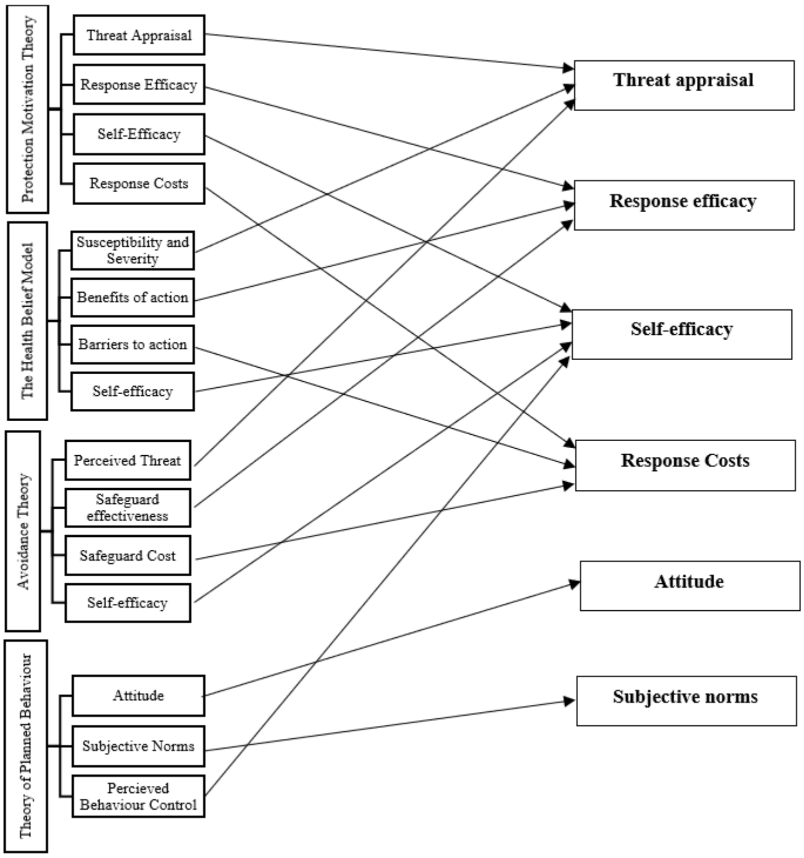
Scholl et al. (2018) in their Knowledge, Attitude and Behaviour Model (KAB) also acknowledge the importance of the relationship between attitude and behaviour and the need to separate attitude from knowledge alone. Whilst people may have the knowledge they need to protect themselves and their organisation from a cyber-attack, if they do not have a positive attitude towards the behaviours being suggested, they are unlikely to accept them or indeed adopt them over the longer term (e.g., following awareness training). Previous research in the field of cybersecurity has found some support for a link between cybersecurity attitude and behaviour suggesting, that a more positive attitude towards cybersecurity will result in less cyber-risky acts (Haddlington, 2018, 2017). The fear appeals mentioned earlier within this section have also been found to adapt general attitudes towards cybersecurity as well as threat and coping appraisals, through the form of persuasion (Olinas-Kukkonen & Harjumaa, 2008). This provides further support for exploration into the use of fear and coping appeal nudges Chapter 3 of this thesis.

The final factor included within the TPB to be discussed, is that of subjective norms – a person's perception around the likelihood that significant others will perform a suggested behaviour and that that person will therefore follow suit (Conner & Armitage, 1998; McGill & Thompson, 2017). The influence social and cultural relationships can have on the actions of others is a well-researched topic within the psychology domain (discussed further under social proof exploitation by cybercriminals in Chapter 3). It is well referenced that humans will behave like those around them, often as a source for intuitive heuristics (van Bavel et al., 2019; Raafat et al., 2009; Scherer & Cho, 2003). However, research in relation to subjective norms, around its influence on behaviour in the cybersecurity field, is often unfounded or mixed with it suggested that any relationship that may exist is mitigated by increasing self-efficacy (Ajzen, 1991; McGill and Thompson, 2017). The potentially higher the self-efficacy, the less likely people will look to the beliefs of others as an indicator of how to behave (Wang et al., 2015). Despite the mixed findings in relation to subjective norms and its impact on cybersecurity behaviour, it is important to include it within the aggregated framework of behaviour within this thesis, due to the known influence social norms can have, particularly in relation to persuasion (Cialdini, 2001).

Taken together, the four largely comparable theories of behaviour change (PMT, HBM, AT, TPB) suggest threat appraisal, response efficacy, self-efficacy, response costs, attitude and subjective norms to be at the centre of human behaviour change (see Figure 1). This suggestion is supported in a systemic review by Sulaimen et al. (2022), that finds previous research to support the integration of the TPB with PMT models, helping improve understanding around human cybersecurity behaviours.

Figure 3

The Aggregated Factors of Psychological Behaviour Change Theory



Note. Overlap of factors across models (left of figure) can allow for a simplified and encompassing set of behavioural influencers (right of figure).

Whilst there are some differences in the approach, application and importance of these concepts, it is imperative to highlight the clear overlap over many of the key themes that result in confusion within academia and organisations around what theory, and what factors with which to try and influence cybersecurity behaviour change that therefore explain the human experience.

Whilst four key behaviour change theories are discussed (PMT, HBM, AT, TPB), the constructs that each suggest as paramount in new behaviour adoption have similar concepts, with a need to merge the theories involved to generate a master theory more useful for application in industry and perhaps offering a more transparent understanding around how employees experience cybersecurity. Figure 1 highlights, how the fifteen factors contained within the four key theories, can be condensed to only six constructs of interest to create a more manageable research platform. For example, all four leading theories contain a self-efficacy element, and three of the four theories contain a factor related to how people appraise threat. By undertaking a commonsensical task of reduction, a more concise behaviour change model has been generated that is easier to investigate within future studies and for organisations to understand the employee experience moving forward.

Acceptance of Technology Model

As well as the psychological and social aspects deemed important in human behaviour change, research within the fields of Human-Computer Interaction (HCI) and Human-Robot Interaction (HRI) also investigate behaviour in relation to how the acceptance or adoption of the technology itself influences intention to behave a certain way (Sun et al., 2013). Whilst technical research is expanding its acceptance focus to include aspects of psychological behaviour change models, more is required to understand how technology acceptance may fit within a behaviour change model within the psychological domain (Chenoweth, 2007; Fei, et al., 2022). Integrated behaviour change and technology acceptance models have so far been investigated largely in the health industry, exploring behaviour towards the use of everything from electronic patient records to mobile health services and medical wearables (Hsieh et al., 2017; Mamra et al., 2017; Rahi et al., 2021; Singh et al., 2022). This will however be one of the first set of studies to investigate a possible converged behaviour change and acceptance of technology model within cybersecurity.

The Unified Theory of Acceptance and Use of Technology model (UTAUT; Venkatesh et al., 2003) is an additional theory of relevance with the aim of assessing users' acceptance of technology, said to influence intention to use/perform the aspect of technology in question, in this case cybersecurity. The UTAUT was developed by Venkatesh et al. in 2003 as a suggested upgrade and adaptation from the Technology Acceptance Model (TAM), of which was on its second version (Venkatesh et al., 2012). TAM focused largely on two main factors: *Performance expectancy* – the usefulness of a technology, and *effort expectancy* – its ease of use, to determine whether the technology in question was likely to be adopted by users, with positive results (Marangunić & Granić, 2015). The UTAUT added to TAM two additional predictive factors including: *Social influence* – potential peer impact (not dissimilar to social norms in TPB and social proof discussed in Chapter 3) and *facilitating conditions* – the knowledge and resources required for the technology to be successful, and therefore whether intentions are present that suggest future use. In 2012, Venkatesh et al. extended the UTAUT to formulate a UTAUT2, that included additional constructs for use within the consumer market such as *hedonic motivation* - does the technology provide fun or experiential benefits, *price value* - is it value for money, *habits* - what are the routines in relation to the technology, with even further extensions over the years including constructs such as *trust*. More recently talks have begun around the generation of a UTAUT3 by

UTAUT's forefather, Venkatesh, and others within the field, this time in specific relation to the acceptance of artificially intelligent (AI) systems e.g., smart technology (Kessler & Martin; 2017; Venkatesh, 2022; Wanner et al., 2022). This new version will likely include concepts such as *transparency of the system*, *trust propensity* and *attitudes towards AI*. The continual development of UTAUT within the technology domain, as well as very good reliability (α = .7 to .9) across studies in relation to other technological interventions, such as mobile banking, mobile internet and internet services (Oh & Yoon, 2013; Zhou et al., 2010), guiding the inclusion of the four factors found within UTAUT in Study 1. Whilst specific research around the use of UTAUT in the domain of cybersecurity has not yet been fully established, Alhalafi and Veeraraghavan (2023) have begun to conceptualise a cybersecurity based UTAUT model to also include the concepts of safety, resiliency, availability, confidentiality and integrity, with positive results. The growing utilisation of UTAUT as a key model to explain the acceptance of technology, dictated the need to include it within the initial assessment framework, in the hope that the factors included would provide further understanding around why end-users choose to perform or not perform secure behaviours, providing guidance around their experience. A number of additional factors found in the literature to be potentially influential in end-user cybersecurity behaviour will now be discussed.

Additional Factors of Potential Influence

As well as the inclusion of the key concepts found within behaviour change theory and the UTAUT, a number of additional factors that are often posited as related to cybersecurity behaviour within the literature, and antecedents to a number of previous constructs discussed, were also included within the current framework. These include overall cybersecurity awareness, level of experience and involvement in cybersecurity, value seen in its policy, commitment to a person's organisation, attachment to their organisation's technology, and potential maladaptive rewards (both intrinsic and extrinsic).

First in relation to the suggested antecedents of the factors found within the TPB, research by Safa et al. (2015) present three precursors to cybersecurity attitude, cybersecurity self-efficacy and subjective norms that are potentially of interest (see Figure 2):

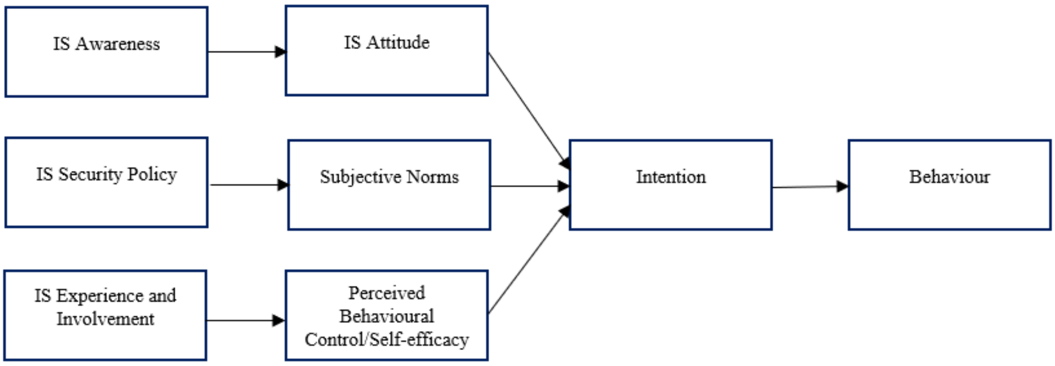
(a) Cybersecurity (or information security) awareness (ISA): The suggested antecedent to attitude, this relates to the need to maintain updated accurate knowledge in relation to cybersecurity risk and its effective coping behaviour;

(b) Cybersecurity experience and involvement (ISEI): The precursor to perceived behavioural control or self-efficacy, this relates to the time and energy required to increase experience and improve behaviour;

(c) Cybersecurity organisation policy (ISOP): This is the determinant of subjective norms and relates to the perception of organisational guidance and its effectiveness.

Figure 4

Supported Antecedents of the Theory of Planned Behaviour (Safa et al., 2015, permission granted)



Information Security Awareness. First in relation to ISA, the cybersecurity breach environment can be dynamic in nature, with a need for end-users to adapt to continual changes in how threat modifies and matures across time. This makes it critical that employees maintain a state of awareness whereby their knowledge around cyber-threats remains up to date in relation to current risk and the behaviours required to minimise its potential. It is suggested that there are three critical aspects related to maintaining employee awareness and these include (Safa et al., 2015; Zwilling et al., 2022):

a) Current and consistent awareness and training programmes completed;
b) A knowledge sharing culture;
c) Motivation for collaboration.

Together, these aspects encapsulate a culture whereby knowledge is learned, shared and used in an open and supportive environment. Knowledge can be defined as the understanding or awareness of objects, facts or skills that can exist either implicitly or explicitly. Knowledge that is implicit endures inside the mind of a human, whilst explicit knowledge is more outwardly communicated and perhaps documented within books, papers or manuals such as policy (Nickols, 2000). Whilst implicit knowledge is self-contained it can be described if requested. On the contrary, tacit knowledge whilst also held within the human mind, is learned through experience and not easily explained (Nickols, 2000). An example of tacit knowledge could be riding a bike, where explicit instructions started the journey, but the actual activity is so complex it cannot be actioned without gaining personal experience. As well as the implicit and explicit classification, knowledge can be described as declarative or procedural, with the first focused on knowledge that can be articulated (how to do something) and the latter similar to tacit knowledge, related more to the experience of doing (Nickols, 2000). Tacit and procedural knowledge are thought to be processed unconsciously and therefore offer benefits to cybersecurity in that they build habits helping reduce potential impacts in relation to productivity and social engineering, something essential when a task is secondary.

For employee awareness to be current and sustained, knowledge needs to not only be dispensed during formal induction and retention training, but also shared naturally across people and teams within the organisation. Knowledge sharing can be defined as the decision to make information openly available, rather than either intentionally or unintentionally hoarding it within one’s mind (Shaari et al., 2014). For knowledge sharing to be successful it requires two trading actions, the donation of information to others as well as the harvesting of required information others may possess (Shaari et al., 2014). Knowledge sharing is therefore not about the creation of subject matter experts or champions, but about providing all employees with an equal voice that helps to evolve universal wisdom. By supporting a knowledge sharing culture, less information leakage is experienced as part of natural employee attrition, helping to reduce occurrences where such loss of tacit knowledge can result in increased risk, as well as financial consequences (Bion, 2021).

There are a number of ways in which such information can be shared; the most common example occurring during the creation of training materials when the procedural knowledge of experts is translated into declarative knowledge, later imparted via e.g., presentations ad handouts. During training, this shared knowledge then transfers into declarative knowledge for the trainee, for future experience to convert this into their own procedural information. Outside of training, a knowledge sharing culture should also be encouraged with employees communicating useful information held in the mind, either verbally or tacitly through observation. Knowledge sharing should therefore be an unremitting process with information that is current, directly influencing the behaviours of employees and the usability of the processes they are being asked to undertake.

There are however a number of barriers that may encourage knowledge hoarding amongst employees, such as competition for promotion, bonuses and other forms of rivalry that may stifle motivation to share. That said, the benefits competition can bring to communication will also be discussed later on within this thesis. Employees may also be unaware of the wealth of knowledge they possess, as well as any deficiencies in knowledge that they need to rectify and that others may boast (Shaari & Rajab, 2014). Knowledge sharing can be intentionally encouraged through collaborative meetings and online portals, as well as fostered unintentionally through herding effects.

Herding effects have been investigated within the psychological literature through terms such as social contagion, group think, the bandwagon effect and social priming all focused on how views and behaviours become harmonious within groups, without being centrally steered (Raafat, Chater & Frith, 2009). Herding effects can result in humans making decisions based on what they believe to be the shared views and behaviours of others, even when personal knowledge may suggest an alternative behaviour. The root of herding effects (similar to social proof in the next chapter) is whereby humans make decisions based on how they believe others think, feel and behave (Hodas and Lerman, 2014). Although its use in social engineering has appeared to reduce in recent years (see Chapter 3), social proof is still a weapon of influence cybercriminals are aware of and use to persuade recipients to click on malicious links in phishing emails (Cialdini, 2001; Butavicius et al., 2016; Ferreira et al., 2015; Zielinska et al., 2016; Parsons et al., 2019). As herding effects can result in the distribution of both desirable and undesirable knowledge, attention needs to be given to how the most up to date and constructive information is allowed to infiltrate and influence organisational security awareness.

The bandwagon effect, where herding behaviours are based purely on belief popularity, can be used to promote positive messaging when delivering qualitative (positive comments) as well as quantitative (hits and likes) support (Lee et al., 2020; Waddell & Sundar, 2020). Group think, where the desire to maintain group harmony inhibits members delivering conflicting opinions, can be better supported during face-to-face interaction by providing impartial leadership and increased employee self-efficacy, encouraging social risk-taking. An additional way to encourage the infiltration of advantageous behaviours through the business is by utilising Behavioural Threshold Analysis, measuring the number of employees needed to conduct the behaviours before herding effects can have an impact (Snyman & Kruger, 2021).

Information Security Experience and Involvement. Information security experience and involvement (ISEI), noted as the antecedent to self-efficacy, is defined as the time and energy exerted on an object or event, with involvement increasing experience and resulting in improved behavioural intention and cybersecurity capabilities (Safa et al., 2015). The experiential journey from novice to expert permits the more skilled to recognise features and patterns in an object or event that can help formulate central principles from which more controlled future decisions can be based (Bion, 2021). Continual experience of a condition across time can result in far greater learning than declarative instructions, even providing the basis for interpreting information yet to be experienced e.g., the ability to comprehend sight read music (Lewis, 1988). To learn is to take the experience currently occurring and adapt any perceptions of the world within the mind, created by previously acquired knowledge (Bion, 2021). Through this systematic adaptation, tacit knowledge is incrementally built through learned experiences, providing capabilities that can be actioned but not readily communicated.

Whilst tacit knowledge is difficult to measure and explicitly communicate, it has been found to be the most valuable class of knowledge when problem solving or making predictions (Bion, 2021). This suggests more operational advocation of cybersecurity behaviours may be required to mitigate risk. Despite this, a study on small businesses by Patterson (2017) found 70% of the owners analysed do not involve their staff in any aspects of cybersecurity, with 40% believing that their employees came to their company sufficiently equipped with the required skills. It is therefore important that employee involvement is maximised so that experiences can be gained, tacit knowledge built, and this tacit knowledge observed by peers, allowing experience to permeate through the organisation and not be lost during natural attrition. This may be more challenging in companies who have large IT or separate cybersecurity departments solely responsible for making these decisions due to the bystander effect, whereby the knowledge that others will intervene is enough to demotivate action (Garcia et al., 2002).

Employee involvement in cybersecurity not only develops experience but can also increase motivation through empowerment (Amah & Ahiauzu, 2013). Allowing employees to set their own goals, make decisions and problem solve can improve workplace relations, innovation, self-esteem, organisational trust as well as generate more creative problem-solving (Freeman et al., 2000; Naqshbandi et al., 2019; Obiekwe et al., 2019). Involvement does however need to be influential not just passive, with employees inspiring actual decision-making rather than simply providing a voice (Cox et al., 2006; Markey & Townsend, 2013). In relation to cybersecurity, increased employee participation in its policies and strategies can improve their practical effectiveness, adoption and employee psychological ownership (Hedstrom et al., 2011; Lin & Wittmer, 2017).

A decision-making heuristic at the root of successful cybersecurity involvement is the IKEA effect, a decision-making bias, whereby humans place higher value on the objects or ideas that they have themselves helped create (Franke et al. 2010; Norton et al., 2012). Should a customer of IKEA (or similar store) self-build an item of flatpack furniture, they will value this product more highly than an identical item of furniture built by someone else. Norton et al. (2012) suggest that this is due to feelings and expressions of competence that ascend from being involved in successful creativity. This has links with psychological ownership, whereby investing more time in an object increases its perceived value and higher aversion to its loss (Baxter et al., 2015; Lee & Chen, 2011). The influence of employee involvement and experience on cybersecurity behaviour therefore suggests the inclusion of employees in the creation of policy and strategy in order to develop feelings of empowerment and build efficient expertise. This is being offered in some organisations through online platforms that collate employee feedback on the usability of policy during and after training, with sentiment analysis used to uncover where security workarounds are most likely to be occurring. Feedback can then support the creation of more usable policy with these changes actively fedback to employees, allowing them to feel more included in the cybersecurity process (Patterson, 2017; Reegård et al., 2019). The importance of this, is that with employee feedback, policy can become more usable resulting in less opportunity for problematic shadow security workarounds.

Employee engagement is a term heavily related to involvement with a focus on developing employees that are physically, mentally and emotionally connected to an organisation resulting in more effective workers (Osborne & Hammoud, 2017). Whilst strategies to improve employee involvement may increase participation, it does not guarantee engagement with this motivational aspect also requiring consideration (Nicholas & Erakovich, 2013). Several key requirements have been outlined as fundamental in achieving employee engagement that may assist commitment to cybersecurity, including an awareness culture, respectful and authentic leadership, employee empowerment and opportunities for personal growth (Gupta, 2015; Osborne & Hammoud, 2017).

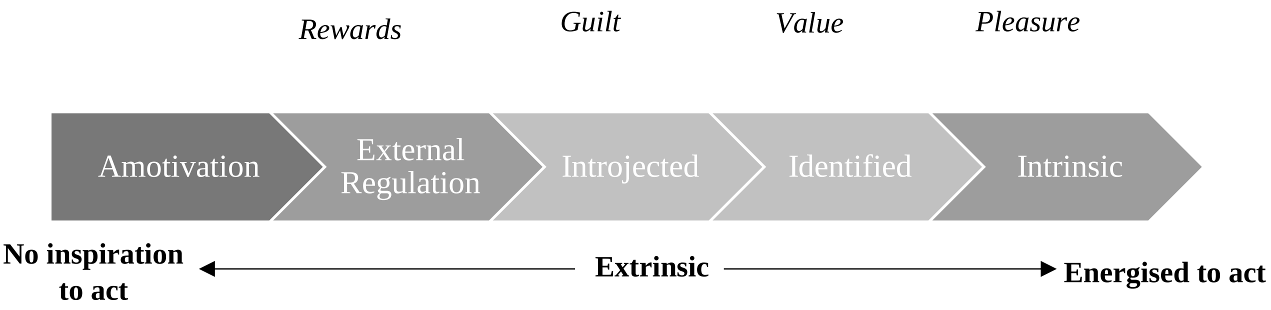
Organisations must promote open communication and a knowledge-sharing culture where employees are encouraged to share their thoughts and ideas around cybersecurity policy without criticism. Organisations should also communicate back to their employees with updates on cybersecurity status, keeping the communication hopeful whilst not being afraid to share bad news, resulting in a positive cybersecurity awareness culture. Employees need to feel trust towards leadership, both trust towards their principles in relation to cybersecurity but also that they respect employee contribution. Leaders must appear to be adhering to cybersecurity policy, practising what they preach (Osbourne & Hammoud, 2017). They must also respect the additional work risk-aware behaviours cost employees and appreciate their dedication to it (Gupta, 2015). Employees need to feel empowered, with organisations involving them in key cybersecurity company initiatives, allowing them to get involved in the bigger picture. Encouraging employee innovation and creativity in relation to cybersecurity, the more involved employees feel, the more value they perceive (Gupta, 2015; Nicholas & Erakovich, 2013; Norton et a., 2012). Even encouraging personal growth in other areas will have an indirect effect, by helping retain the most talented people in the organisation as well as those who have become experienced in company cybersecurity practices.

Should employees become incrementally successful at parts of a cybersecurity task, their self-efficacy for the task as a whole will increase. Therefore, providing cybersecurity training that starts with a simple breakdown of the behaviours required, with a gradual increase in level of difficulty would be extremely beneficial. An example could be used within current organisation phishing simulation tools, whereby the use of easy phishing simulation emails delivered can then gradually increase in level of difficulty, across time, providing feedback at each point and supporting employees towards enactive self-mastery (Elliot & McGregor, 2001; Nicholls, 1984; Ryan and Deci, 2020).

Employee engagement can also be understood in relation to self-determination theory (SDT; Deci & Ryan, 1985, 2012), a model that posits humans to be motivated into action along a continuum from amotivation, through extrinsic motivation to intrinsic motivation (see Figure 3; Ryan & Deci, 2000). Whilst intrinsic motivation is driven by pleasure in the action itself, extrinsic motivation is not automatically self-determined but can become more self-determined as activities are perceived as valuable. As employees are unlikely to ever become motivated to conduct cybersecurity tasks due to the pleasure they bring, it is important that their extrinsic motivation is based on the actual value of the activities rather than purely rewards and punishments so that it becomes as self-determined as possible.

Figure 3

The Continuum of Motivation



Antecedents of motivation within self-determination theory include competence, autonomy and relatedness; feeling capable, free from pressure and with a sense of social belonging (Guay et al., 2000; Ryan & Deci, 2020). Competence within SDT relates to the need for personal growth within employees, autonomy being trusted to reach goals and support company initiatives, relatedness the need for open social connection. Therefore, in order for employees to be engaged as well as involved in the cybersecurity chain, they must feel they have the ability, freedom and social support as enablers (interventions generated to directly target these concepts are investigate within Chapter 3 of this thesis).

ISEI requires organisations to support employee involvement in protecting the company from cybersecurity risk, as well as motivating and empowering them to convert this involvement into the experience required. This can be achieved via the inclusion of employees in the creation of policy and strategy in order to develop feelings of empowerment as well as perceptions of competence, autonomy and relatedness amongst peers. In return, organisations can better understand the employee experience and reasons why shadow security exists in the first instance, in order to generate policy that results in high employee usability. More of this to be discussed next, in relation to information security organisation policy.

Information Security Operation Policy. The construct of ISOP considers employee perceptions around the policies and procedures that governments, compliance agencies and organisations create to inform employees of the behaviours that are required of them to protect against cyber-attacks. The desired perceptions are that the policy is highly valued by employees, and that they feel a sense of appreciation of those procedures by others within their organisation. The importance of perceptions around cybersecurity policy is often overlooked, with the focus often only on the ticking of compliance boxes. However, even with the correct internal policies in place, employees often fail to follow them, resulting in unintentional insider threat (Gheyas & Abdallah, 2016). A 2017 qualitative investigation by Patterson explored the relationship between employees and policy within small businesses, highlighting a lack of employee involvement in its creation, resulting in ill-fit. The outcome can often be a “them versus us” culture, rather than agreed policy that is designed with employees, to be used by employees (Ashenden and Sasse, 2013; Hedstrom et al., 2011; Lin and Wittmer, 2017). Similar interviews took place but within a large multinational organisation by Kirlappos et al. (2014), with similar themes uncovered. Employee feedback around policy going unnoticed, security policy placing a burden on employees increasing cognitive load and impacting primary tasks.

There does appear some cross-over between the ISEI and information security operation policy, where some importance in both constructs is placed on the employee experience and higher exposure resulting in an improved perception of its value. Several studies also mention links between ISOP and social norms within TPB, whereby importance is placed on the behaviour of others (Briggs et al., 2017; Kirlappos et al., 2014). Whilst the literature may be currently lacking in direct relation to ISOP, some research focus has been placed the potential result of poor understanding in relation to operational policy. *Shadow security* - the consequence of policy that is not perceived as useable by employees, resulting in non-compliance and therefore risk. Employee policy compliance is not dichotomous, instead it runs along a continuum from compliance to non-compliance. Shadow security is firmly placed between these extremes, whereby a degree of usability is actually present in the policy, but it requires some form of adaptation in order to be functional (Kirlappos et al., 2014). How policy is adapted by employees can differ, with micro-cultures existing where groups of employees are all adapting policy in the same format. The reason these adaptations are termed ‘shadow’ behaviours is their propensity to run ‘underground’, invisible to those measuring cybersecurity risk and therefore providing a false sense of security for an organisation. Reductions in such risk can perhaps be achieved by providing policy in a clear, bite-sized format, where feedback is not ignored or listened to passively, but actively applied to provide instructions that ensure a ‘best-fit’ rather than ‘ill-fit’. Whilst all three of the factors discussed above are suggested as predictors of key TPB constructs, they are yet to be fully explored in their own right, despite evidence that they may be of direct predictive value to cybersecurity behaviour.

Organisational Commitment. Organisational commitment is an additional factor that has previously been found to relate to cybersecurity, defined as an employee's ability to identify with their organisation and align with their goals (Karim and Noor, 2017). The higher the sense of attachment an employee feels towards their workplace, the better their suggested level of productivity and lower their potential risk (Reeve et al., 2020). Meyer and Allen's (1991) tool, to measure organisational commitment, propose three key reasons why an employee may remain within an organisation - *because they want to* (have an emotional attachment), *because they have to* (need the money) or *because they feel they ought to* (feel obliged). Employee organisational commitment based more on an emotional attachment is said to result in the highest level of performance, also leaving them more likely to adhere to organisational policy (Karim and Noor, 2017; Scholl & Scholl, 2018).

In addition to connections between organisational commitment and ISOP, this particular factor has also been found to be related to threat appraisal, with higher organisational commitment also resulting in higher perceptions of severity of attack should one occur (Posey, Roberts and Lowry, 2015). Finally, organisational commitment has also been linked to improved employee engagement, an aspect discussed within above in relation to ISEI (Cox et al., 2006; Osborne & Hammoud, 2017). The presence of such overlaps with other concepts contained within the framework suggests it important to investigate the influence organisational commitment can have on reported cyber-security behaviours as well as how it interacts with the other cyber-security concepts discussed.

Psychological Ownership. An additional factor previously found to predict reported cybersecurity behaviour is psychological ownership: the feeling of mental claim or possession of an object driving the need to control it (Baxter et al., 2015). Research to date has found psychological ownership to be an internal motivator of cybersecurity behavioural intention, with those more attached to their organisation also more likely to protect their data (Raddatz et al., 2020). Psychological ownership has also been found to be closely associated to self-efficacy, whereby its impact on behaviour becomes more powerful the higher the perceptions of psychological attachment (Verkijika, 2020). Psychological ownership has also been linked to the adoption of digital technologies, such as increased physical connection to phones and computers via touch screens that increase attachment, and social media usage increased through customer/company co-creation of avatars and emoji stickers within apps (Brasel and Gips 2014; Kirk & Swain 2018; Zhao et al. 2016). When music became something to stream rather than something to physically buy, there was concern that adoption of new music would reduce due to lower levels of attachment. However, music streaming services rose to the occasion by drawing on psychological ownership theory to maintain feelings of ownership through customer creation of playlists and an increased sense of self present in the streaming application the more effort is added to generating them (Sinclair and Tinson 2017). Psychological ownership can therefore occur even when legal ownership is not present, increasing perceptions of responsibility (Peck et al., 2021).

When an object is perceived as psychologically owned, the holder will view it more favourably as it becomes an extension of themselves (Dyne & Pierce, 2004). Feelings of attachment will occur towards the object increasing its perceived value, and therefore a need to guard it in order to avoid its loss (Baxter et al., 2015). Psychological ownership is centred around a human decision-making heuristic, known as the endowment effect, whereby users place higher value on possessions they own, than something they do not (Pfleeger & Caputo, 2012). With its foundations in loss aversion, psychological ownership results in an unwillingness to swap an endowed item with one of similar value irrespective of which item was originally provided as well perceiving its sell value as higher than buyers are willing to pay. A study by Renaud et al., (2019) found psychological ownership can also be present for cybersecurity tasks, such as the creation of computer passwords. Participants reported attachment to their password routines resulting in them over-valuing their own personal strategy and remaining attached to the process and less willing to change. Increasing employee psychological ownership for company technology, data and policy will likely result in increased attachment and therefore higher perceptions of value and the need to avoid loss.

The literature suggests that a number of antecedent factors are present in relation to psychological ownership including - increasing control of the item, time and effort invested in it and coming to intimately know it (Baxter, Aurisicchio & Childs, 2015; Peck et al., 2021). In relation to a work computer, employees may control the technology by adjusting screen brightness, investing time downloading software and coming to intimately know its marks of wear and tear. First in relation to control, the more an end-user is able to manipulate technology for their personal comfort, the more possessive they will become over it with a need to therefore protect it (Lee & Chen, 2011). Baxter et al. (2015) discuss five ways in which an item can be controlled; spatially, through configuration, temporally, rate control and transformational control. Spatially, an item can be manipulated by physically moving it for personal comfort for example angling a computer away from the light or raising a laptop to eye level. Technology can also be controlled through its configurations, by personalising its sounds (e.g. alert tones) or setting a personal photo as a screen saver. Temporal control involves the ability to use the item when desired, and rate control, as desired with all features of work-based technology requiring constant availability for employees. Transformational control relates to leaving personal ‘marks’ on the technology item. For example, the addition of desktop icons, apps and taskbar links. The personal adaptation of objects allows employees to instantly recognise their technology just by viewing it or switching it on. Control therefore centres around the freedom to personalise the hardware, software and settings of the work phones, computers and tablets as desired, in order to increase perception of ownership and, in turn, improve security behaviour.

Self-investment is another way in which psychological ownership can be manipulated, whereby increasing the time, energy and effort exerted, results in perceiving an object as an extension of the self (Baxter et al., 2015). There are five ways in which employees can self-invest in work technology; through creation, repair and maintenance, using it as a repository, the use of emblems, and preference recall (Baxter et al., 2015). Whilst it would be challenging for employees to be involved in the creation of their work technology, allowing personalisation of its settings at the point of set up, as well as options around e.g., protective casing and software, can help evoke this contributing factor. Affording employees responsibility around the repair and maintenance of their technology, will also increase psychological ownership for example scheduling regular services. Employees using the technology as a repository of information has been challenged by modern cloud-based solutions, however allowing even a small amount of personal files (without breaching company security policies) to be stored on work mobile phones or laptops can increase perceptions of ownership. Emblems relate to how the technology confirms an employee’s identity, for example those in higher positions expecting higher quality technology, with those in the same team anticipating the same level of technology as each other. Finally, should hot desking or the regular sharing of technology be unavoidable, automatic recall of a person’s setting such as a computer profile can help to alleviate reductions in psychological ownership caused by not having a personal working space. Therefore, by employees investing time and effort in creating, personalising, maintaining, and storing information within work technology, psychological ownership can be increased even where legal ownership does not exist.

The third antecedent of psychological ownership is suggested to be intimate knowledge, where over time, people come to more intimately know an object, making it more special than similar items (Baxter et al., 2015; Lee & Chen, 2011). This factor has six contributing variables including ageing, disclosure, periodic signalling, enabling, simplification and proximity. Maturing alongside technology will result in employee ability to identify the items through the bumps and scratches received across time. Therefore, the longer the technology remains with the employee, the more attached they will tend to become, particularly if previous owner (if applicable) contamination (e.g., personalisation elements) is erased. Disclosure references the memories or experiences linked with the technology, such as the view from the office when using it or being passed the technology of a friend that has passed away. Even those more bothersome aspects of technology that periodically signal ownership can grow feelings of attachment, for example the app that automatically closes after certain actions - resulting in increased intimate knowledge.

Baxter et al. (2015) also state that the technology should be perceived as enabling the employee to achieve a desired experience, for example being able to efficiently complete their work, read emails remotely or remain in contact with colleagues. The technology should also simplify these experiences through features such as electronic reminders and saved contact numbers, with this simplification hindered by poor internet connection or limited memory. Finally, proximity will influence psychological ownership with a mobile phone held to one’s face evoking more psychological ownership than a technology that is voice activated. Psychological ownership can be enhanced by employees finding comfort in the aging, memories and experiences bound around the technology will increased familiarity resulting in more attachment and higher aversion of potential loss (Baxter et al., 2015).

Therefore, psychological ownership can be experienced without legal ownership and is enhanced through perceptions of control, self-investment and intimate knowledge. Within the workplace, technology can be controlled by allowing employees the freedom to personalise hardware, software and settings as desired such as screen height adjustment, the use of personal screensavers and unlimited mobile call hours. The time and effort invested in the technology can also increase psychological ownership by involving employees in the initial customisation of the technology, responsibility for scheduled maintenance and ongoing storage of information where loss of the item will become greater. Lastly, across time, employees will relate the technology to positive past experiences and memories so long as its features help simplify and enable positive experiences rather than become linked with negativity and frustration. The impact of past experiences on current behaviours draws similarities to the experience and involvement factor, whereby the more a human interacts with something, the more a connection they find to it.

Maladaptive Rewards. This concept relates to the intrinsic and extrinsic rewards a person may experience by not protecting themselves or their organisation from a cyber-attack. Intrinsic maladaptive rewards (IMR) relate to those benefits that may exist internally to a human, such as feeling personal gratification for not protecting their organisation. Extrinsic maladaptive rewards (EMR) however differ, relating to external benefits offered to an employee for not protecting their organisation, such as being financially rewarded. Should the maladaptive benefits experienced by an employee, outweigh their threat perception, they may opt for such internal and external benefits (Hassandoust & Techatassanasoontorn, 2020). Maladaptive rewards can result in behaviours that are unintentional, through neglect or lack of attention resulting in security ‘slip-ups’, or intentional such as helping provide system access to a cybercriminal due to the low organisational commitment discussed previously (Gheyas & Abdallah, 2016). Both the intentional and unintentional ‘risky’ behaviours, driven by maladaptive rewards, can result in huge financial loss for an organisation and are therefore important factors to consider.

A number of papers have focussed on extending behaviour change models by including intrinsic and extrinsic maladaptive threat behaviours (Hassandoust & Techatassanasoontorn., 2020; Safa et al., 2015). The literature around maladaptive rewards is small, however this is perhaps due to the difficulty in subjectively asking an employee whether they feel inclined to not protect their organisation for pleasure or monetary benefits (Liang et al., 2016). Research is available in relation to insider threat, a potentially similar concept. Defined as a current or former employee who exceeds, misuses or grants access to others in order to negatively impact an organisation's security (Greitzer et al., 2016). Similar to maladaptive rewards, insider threat can be deliberate or simply due to lack of care (Bradley et al 2017), motivated by aspects such as personal frustration, financial difficulties or reduced loyalty to a company. Insider threat employees do not suddenly become a risk, they are driven by a slow build-up of personal and work-based struggles, such as the emotions around a marriage breakup becoming more challenging due to an unsupportive manager at work. A number of psychological concerns have been identified as perhaps predisposing someone to this concern, such as an anti-social personality (Bradley et al., 2017), with some tools existing to try and extract current employee state of mind, for example personality mapping through the use of psycholinguistics, text analysis of life events and sentiment analysis to detect emotional state exhibited within their social media (Bradley et al., 2017). However, there are clear ethical questions that must be addressed before these techniques are utilised. It is also important to state that those scoring high in IMR and EMR are not solely end-users, but also the security operations staff and those sitting on an organisation's executive board (e.g., Directors).

There is still much to learn about the human characteristics and environmental situations that can lead an employee towards insider threat behaviour (Greitzer et al., 2016). A number of interventions have been described that may help reduce opportunities for insider threat to exist, with increasing organisational commitment one of those interventions (Bradley et al. 2017). However, more work is required in relation to how internal and external rewards impact employee security behaviours at work, providing reason for why both IMR and EMR were included within current framework.

The number of factors amassed from the above literature review, make it clear that organisations require more guidance around where to 0
intervention with the limited time and budget they have available. The aim of this research is to evaluate the multiplicity of discrete factors discussed above, in the hope to parsimoniously explain human cybersecurity behaviour in a way accessible to organisations. By streamlining these factors into a shared exploratory framework, interventions can be created to improve vulnerabilities, from today, whilst future research continues to investigate and interchange the factors underlying the model.

Aims of Studies 1-3

Following a comprehensive literature review, the main aim of this set of studies was to unearth and test factors believed to be influential in human decision-making, helping organisations predict risk in relation to the human in cybersecurity situations. By better understanding the constructs influencing behaviour, organisations can also design more targeted interventions and provide more tailored support. Based on this review, the following studies were designed to:

- Collectively explore a number of psychological models and individual differences that have been noted in the literature as potentially influential to cybersecurity behaviour but have never been brought together in a single study (Study 1);
- Examine the underlying structure of the large number of constructs under analysis and their potential relationships, to identify the existence of any latent factors (Study 2);

- Strengthen the validity of the framework by investigating how these latent factors significantly relate to reported cybersecurity behaviour, generally and within a more targeted (organisation) sample (Studies 2-3);

The scope of Study 1 is largely explorative, collating and investigating the numerous human individual differences underlying behaviour change theory as well as a number of other human characteristics previously found to relate to risky cybersecurity behaviours. Study 2 will seek to confirm the findings from Study 1, more specifically within an industry sample. It will also extend the research by investigating which of the correlating factors are found to significantly explain the variance in reported cybersecurity behaviour, with item reduction taking place prior to this through an exploratory factor analysis. Study 3 was designed to further verify the findings of the previous two studies by employing a larger sample with a key focus on the metrics highlighted as most influential in relation to reported cybersecurity behaviour. The main objective of this third study was to confirm the model created in Study 2 and generate a framework with which to further investigate employee cybersecurity vulnerability, to improve understanding around interventions that will better support at risk employees, something never attempted before in research. Therefore, the objective of Studies 1 to 3, were to work towards understanding human cognitive vulnerabilities in relation to cybersecurity and provide organisations with a parsimonious set of metrics with which they can benchmark, intervene and reassess risk.

Study 1

Recap of Study 1 Aims

Study 1 was exploratory by design, with the findings post analysis used to inform studies 2 and 3. Due to the large number of factors involved in the investigation, and without clear guidance on how they relate or may fit into a predictive model of behaviour, exploratory factor and regression analyses were saved for studies 2 and 3, respectively.

Study 1 addressed the following aims:

- Exploration of relationships (using correlation analyses) between a large number of factors previously identified as significantly relating to cybersecurity behaviour, albeit together in one study for the first time;
- Based on the findings of the above, the development of a second refined iteration of the framework used to identify human susceptibility to cyber-attacks would inform Study 2.

In order to better visualise the large number of variables included within this study, a human-centric framework identified by Albladi and Weir (2018) within cybersecurity, but in specific relation to social engineering attacks in social networks, is used to combine factors and aid orientation. Their model validates the use of a single framework encapsulating four overarching key themes that are believed to influence judgement - *perceptual attributes (e.g., threat appraisal), socio-psychological attributes (e.g., personality), socio-emotional attributes (e.g., trust) and an habitual theme (e.g., level of involvement; see Figure 4).* This useful distinction has been applied to initial interactions of the framework to aid categorisation and simplify understanding but however does not suggest the identification of latent variables.

Figure 4

User-centric Framework (created for this thesis from work by Albladi & Weir, 2018)

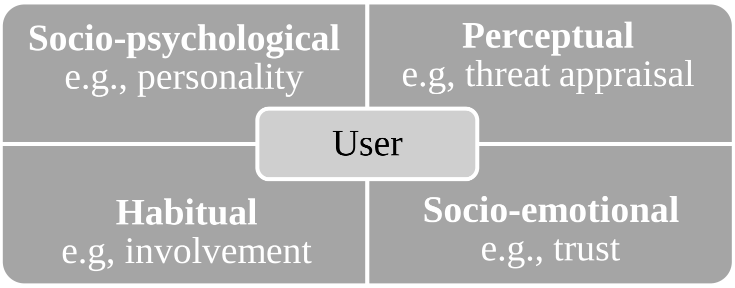
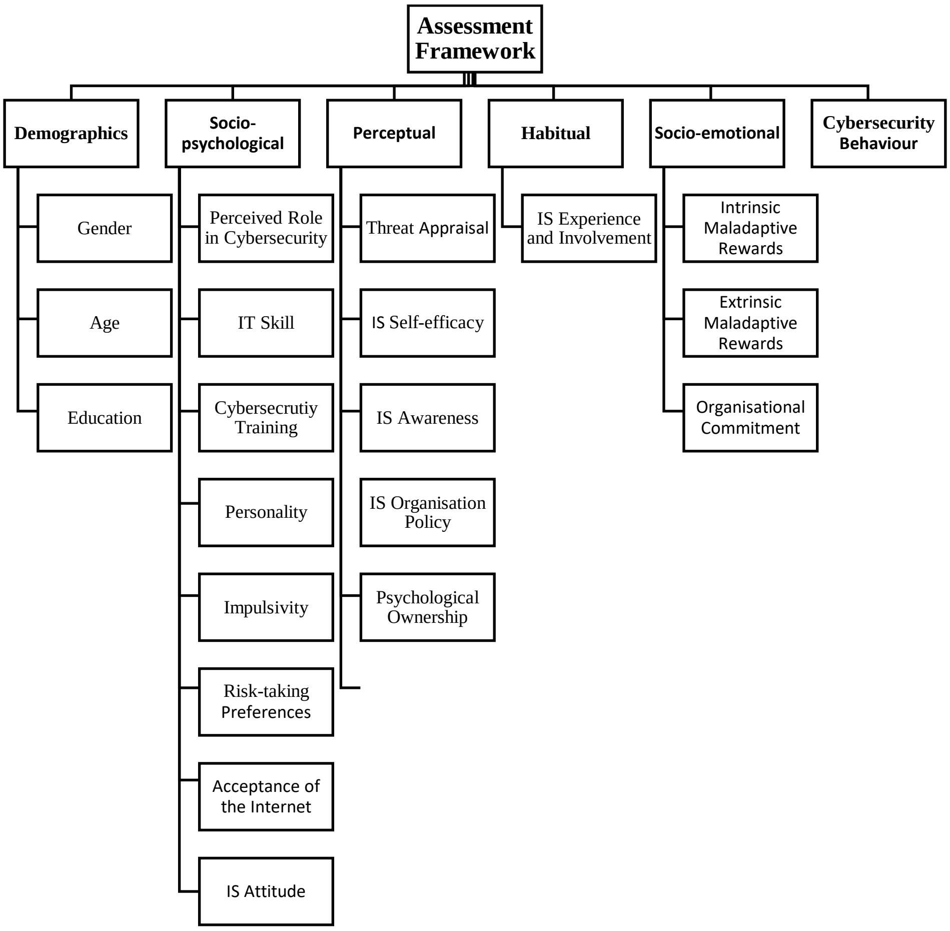


Figure 4 provides a graphical representation of the factors within the first iteration of the human vulnerability in cybersecurity tool (with the classifications of Figure 3 in mind) that will be examined within Study 1. Post-analysis within Study 1, this framework will be refined to include the factors that are considered significantly related to reported cybersecurity behaviour in preparation to further verify these relationships within Studies 2 and 3.

Figure 5

Iteration 1 of the Human-centric Cybersecurity Assessment Framework



Research Hypotheses

Study 1 was exploratory in nature. However, a number of key hypotheses were determined from the breadth of literature available within cybersecurity research and beyond (noting those highlighted in grey are the key factors present in important theories of behaviour change and likely to hold the most significant relationships):

Demographics. Study 1 (S1) H1 Reported cybersecurity behaviour was anticipated to significantly differ across a number of participant demographics (Gratian et al., 2018; Whitty et al., 2018):

S1 H1a It was expected that older adults will report significantly more secure behaviour than younger adults.

S1 H1b It was predicted that men will report significantly more secure behaviour than women.

Socio-psychological Factors. S1 H2 Several socio-psychological factors were also expected to significantly relate to

reported cybersecurity behaviour (Egelman and Peer, 2015; Gratian et al., 2018)**.**

S1 H2a In relation to personality, more secure behaviour were anticipated in both those more extrovert and those more conscientious.

S1 H2b It was predicted that less desirable behaviour will be reported by those more impulsive.

S1 H2c It was also anticipated that less desirable behaviour will be reported in those more likely to take health/safety risk, but less likely to take a financial risk.

S1 H2d In respect to decision-making styles, it was anticipated that those more likely to procrastinate, rely upon others to make decisions or be spontaneous will report less desirable behaviour. However, those with a more rational processing style will report more desirable behaviour.

Perceptual Factors. S1 H3 A number of perceptual factors found within validated models of behaviour change.

were expected to significantly correlate with cybersecurity behaviours (Burns and Roberts, 2013; Carpenter et al., 2019; Liang & Xue, 2009; Pickering et al., 2021; van Bavel, et al., 2019):

S1 H3a Threat appraisal would positively correlate with reported behaviour.

S1 H3b Response efficacy would positively correlate with reported behaviour.

S1 H3c Self-efficacy would positively correlate with cybersecurity behaviour.

S1 H3d Response costs would negatively correlate with behaviour.

S1 H3e Attitude would positively correlate with cybersecurity behaviour.

S1 H3f Subjective norms would positively correlate with behaviour.

S1 H3g IS awareness would positively correlate with reported cybersecurity behaviour.

S1 H3h IS organisation policy would positively correlate with reported behaviour.

S1 H3i Psychological ownership would positively correlate with cybersecurity behaviour.

S1 H3j A higher acceptance of cybersecurity would positively correlate with behaviour.

Habitual Factors. S1 H4 It is anticipated that the more experience and involvement participants feel they are in cybersecurity, the more secure they will report their behaviour (Safa et al., 2015).

Socio-emotional Factors. S1 H5 A number of socio-emotional attributes were also anticipated to significantly relate to reported cybersecurity behaviours (Posey, Roberts & Lowry, 2015):

S1 H5a Intrinsic maladaptive rewards would negatively correlate with reported cybersecurity behaviour.

S1 H4b Extrinsic maladaptive rewards would negatively correlate with reported cybersecurity behaviour.

S1 H4c Organisational commitment would positively correlate with reported behaviour.

Method

Participants. Seventy participants were recruited from the Cardiff University staff and PhD student pool (48% of the sample) as well as via the Prolific online marketing tool (52% of the sample). All participants were required to be in full or part-time employment in order to take part in the study. Of these participants, 31% were male, 68% female and 1% of a different identity, with an average age of 34.92 years (*SD* 10.67). Prolific pays an hourly rate (£12.05) to its users (as of circa. 2020 for completion of questionnaires, therefore participants were paid ~£8.00 for taking part. However Cardiff University staff and PhD students were not financially rewarded – e.g., students were awarded credits that count towards the research training element of their degree programmes. Whilst it is possible that differences may arise due to the payment offered to Prolific participants compared to the rest of the sample, analyses indicate no significant difference in reported cybersecurity behaviour between participants who were being paid to take part and those who received no such monetary reward *t* (69) 1.829, *p* = .095. Samples were also similar in so far as age and education. However, whilst 50% of participants within the Cardiff University sample were female, 84% of participants identified as female in the Prolific sample.

Design. A Kruskal-Wallis statistical test was employed to analyse differences in cybersecurity behaviour across a number of participant demographics (gender, age, education). A within participant correlational design was also used to investigate how reported cybersecurity behaviour related to several socio-psychological factors (level of IT skill, level of cybersecurity training, perception of importance of role in cybersecurity, personality, risk-taking preferences, decision-making styles, impulsivity, acceptance of the internet), perceptual attributes (threat appraisal, attitude, self-efficacy, subjective norms, perceived behavioural control, response efficacy, response costs, awareness, organisation policy), a habitual factor (experience and involvement) and socio-emotional factors (intrinsic and extrinsic maladaptive rewards, organisational commitment, psychological ownership). All questionnaires were randomised to reduce potential confounds of order effects.

Materials and Procedure. Each participant took part in one study only, accessing it via *Qualtrics*©, an online survey platform, on PCs and tablets (calibrated such that information was presented in a comparable manner despite the device being used). Participants were required to be in active employment to take part, due to the nature of the questions focused on their behaviour and perceptions, within the workplace. Before completing the battery of measures, participants were presented with an introduction sheet (Appendix A) and a request for informed consent to take part (Appendix B). Each participant also completed an optional demographics form collating information on their age, gender and level of education. Participants were asked to categorise their highest level of education across six categories including GCSEs, A-levels, undergraduate degree, master’s degree, PhD/Doctorate, or other. Also rated by participants were work-role importance in cybersecurity, from 1 (extremely important) to 5 (not at all important), level of IT skill rated from 1 (poor) to 5 (excellent) and level of training in cybersecurity from 1 (none) to 5 (expert). All of the following questionnaires and questions within those questionnaires were randomised.

Participants were asked to complete the IPIP personality traits measure (Appendix C; Goldberg et al., 2006) where they were presented with 50 statements (10 questions for each subscale including extroversion, openness to experience, neuroticism, conscientiousness, agreeableness) such as *'I make friends easily'* and asked to which extent each statement applied to themselves, rated from 1 (very inaccurate) to 5 (very accurate).

Risk-taking preferences were measured using the DOSPERT risk-taking preferences questionnaire (Appendix D; Blais & Weber, 2006) whereby participants were asked to rate how likely they were to engage in 30 risky behaviours from 1 (extremely unlikely) to 7 (extremely likely). The 30 questions, such as *'revealing a friends secret to someone else'*, covered five different forms of risk taking (social, recreational, financial, health/safety, ethical) with six questions per factor.

Participants completed the General Decision-making Styles Scale (GDMS; Appendix E; Scott & Bruce, 1995) indicating to which extent participants agree or disagree with 25 statements, such as *'I generally make snap decisions'*, with five overarching decision-making styles (intuitive, dependent, avoidant, rational, spontaneous) ranging from 1 (strongly disagree) to 5 (strongly agree). The Barratt Impulsiveness Scale was also employed (BIS-11; Appendix F; Patton et al., 1995) indicating how regularly they had experienced a list of 30 statements such as *'I don't pay attention'* ranging from 1 (rarely/never) to 5 (always). In addition, the UTAUT acceptance of the internet questionnaire was used (Appendix G; Venkatesh et al., 2003) containing 30 statements such as *'I find cybersecurity tasks useful in my daily life'* (with 9 subscales including performance expectancy, effort expectancy, social influence, trust, facilitating conditions, hedonic motivation, price value, habit and behavioural intention) rated from 1 (strongly disagree) to 7 (strongly agree). The IPIP, DOSPERT, GDMS and BIS-11 questionnaires were utilised within these set of studies to, where possible, replicate the methods utilised within both the Egelman and Peer (2015) and Gratian et al (2018) studies.

The combined TPB and PMT questionnaire (Appendix H; Safa et al., 2015) was rated using 42 statements such as *'I am aware of potential security threat'* from nine sub-scales e.g. threat appraisal, from 1 (strongly disagree) to 7 (strongly agree) and additional questions used within McGill & Thompson (2017) and Posey et al. (2015; Appendix I) rating 33 statements e.g., intrinsic and extrinsic maladaptive rewards such as *'I feel a high degree of ownership for my work computer and its contents'*, from four sub-scales e.g. organisational commitment from 1 (strongly disagree) to 7 (strongly agree).

Cybersecurity behaviour, was measured by the behaviour construct within the PMT and TPB questionnaire, rated from 1 (strongly disagree) to 7 (strongly agree) with five statements such as *'I consider security experts recommendations in my information security manner'*. After completion of all measures participants were provided with a study debrief (Appendix J) and thanked for taking part.

Results

The primary aim of Study 1 was to better understand the human experience in cybersecurity, and those aspects resulting in cyber-attack vulnerability. The objective being, to assemble several measures of individual differences previously found to relate to or predict human vulnerability in cybersecurity and determine their relationship to reported cybersecurity behaviour within this study. The main aim – to generate the first iteration of a novel human-centric cybersecurity framework, that organisations can use to measure and manage human vulnerability across a validated measure in order to periodically test human vulnerability in relation to cyber risk in their organisation.

Reliability of Measures. Initially, a test of internal consistency was applied to all measures employed within Study 1. Cronbach's Alpha tests revealed good to excellent reliability for the Barratt Impulsivity questionnaire (α = .87), GDMS decision-making style questionnaire subscales (α = .78 - .90), DOSPERT risk-taking preferences questionnaire subscales (α = .64 - .86), IPIP Personality Traits questionnaire subscales (α = .75 - .91), the combined TPB and PMT questionnaire subscales (α = .77 - .89), additional constructs included from the protection motivation questionnaire subscales (α = .69 - .88) and for the UTAUT subscales Cronbach's alpha tests reached acceptable to excellent reliability (α = .69 - .95). The key assumptions for parametric analysis were not met due to the use of ordinal data, and therefore non-parametric statistical tests were applied. Assumptions for the following statistical tests were analysed and met. Any missing observations within the dataset were replaced with the grand mean for each question and any outliers, determined as 3 interquartile range (IQR) from the mean, were windsorized to the next available value not considered extreme.

Cybersecurity Behaviour. Cybersecurity behaviour was operationalised within the study through the cybersecurity conscious care behaviour measure, found within the combined TPB and PMT questionnaire (Safa et al., 2014). Descriptive statistics for cybersecurity behaviour reveal a sample median score of (Mdn = 6, IQR = 1), indicating that on average the participants moderately agree that their cybersecurity behaviour is conscious and favourable.

Participant Demographics. A non-parametric Kruskal-Wallis test was conducted to determine if there were significant differences between participant demographics (age **S1 H1a**; gender **S1 H1b**, and level of education) and reported cybersecurity behaviour. Analyses revealed no significant differences between age (*H* = 11.562, *p* = .997), gender (*H* = 2.166, *p* = .339) and cybersecurity behaviour, nor education (*H* = 4.027, *p* = .402).

Individual Differences. Non-parametric Spearman's Rho correlation analyses were then applied to identify any significant relationships between reported cyber behaviour and participant ratings of IT skill (Mdn = 4, IQR = 1) with no significant differences found (*r* =

.07, n = 71, *p* = .579). Level of cybersecurity education was also reported by participants with a median score of 2 (Mdn = 2, IQR = 1), suggesting that participants, on average, rated their level of cybersecurity training as beginner with again no significant differences found (*r* = .20, n = 71, *p* = .093). Participants were then asked to rate how they perceive the importance of their role in the protection of their organisation's systems and data with a median score of 4 (Mdn = 4, IQR = 1), suggesting that, on average, participants rated their role significance as very important, however, no significant correlation with reported cybersecurity behaviour was unearthed (*r* = .17, n = 71, *p* = .169).

Next, the relationships between reported cybersecurity behaviour and a number of socio-psychological factors (**S1 H2;** personality, impulsivity, risk-taking preferences, decision-making styles) were explored. First analysed were possible relationships between personality styles and reported behaviour (**S1 Ha**). Those more conscientious (Mdn = 4, IQR = 1) were found to report significantly more conscious cybersecurity behaviour (*r* = .34, n = 71, *p* = .004) with a medium effect size. However, significant correlations were not identified between reported cybersecurity behaviour and levels of extraversion (Mdn = 3.5, IQR = 1; *r* = .20, n = 71, *p* = .100), agreeableness (Mdn = 4, IQR = .5; *r* = .01, n = 71, *p* = .924), neuroticism (Mdn = 2.5, IQR = 1.5; *r* = -.18, n = 71, *p* = .127) and openness to experience (Mdn = 4, IQR = 1; *r* = .20, n = 71, *p* = .103).

It was also hypothesised that significant relationships would be found between participant risk-taking preferences and reported cybersecurity behaviours. A significant positive relationship was found between social risk-taking (Mdn = 5.5, IQR = 1) and reported behaviour (*r* = .33, n = 71, *p* = .004; **E1 H2c**) with a medium effect size. However, no significant relationships were identified between reported cybersecurity behaviour and recreational risk-taking (Mdn = 2.5, IQR = 3; *r* = .13, n = 71, *p* = .276), financial risk-taking (Mdn = 2, IQR = 1.5; *r* = .16, n = 71, *p* = .198), health/safety risk-taking (Mdn = 2, IQR = 3;

r = .06, n = 71, *p* = .599) and ethical risk-taking (Mdn = 5.5, IQR = 1.5; *r* = -.01, n = 71, *p* = .927).

Decision-making styles were analysed to determine whether significant correlations were present between cybersecurity behaviour and the style participants choose when making decisions – e.g., intuitive (Mdn = 3, IQR = 1), dependent (Mdn = 4, IQR = 1), rational (Mdn = 4, IQR = 0), avoidant (Mdn = 2, IQR = 2), spontaneous (Mdn = 2, IQR = 1; **S1 H2d**). Reported behaviour was not found to significantly relate to any decision-making styles including intuitive decision-making (*r* = .04, n = 71, *p* = .766), that more dependant (*r* = .01, n = 71, *p* = .993), rational (*r* = -18, n = 71, *p* = .129), avoidant (*r* = -.13, n = 71, *p* = .287) or spontaneous (*r* = -.17, n = 71, *p* = .150). However, in relation to impulsivity (Mdn = 2, IQR = .5), a significant negative relationship was found (*r* = -.30, n = 71, *p* = .011; **E1 H2b**).

In addition, participant acceptance of cybersecurity measures was analysed through the factors contained within the UTAUT instrument. The average response for perceived effort expectancy was measured with participants, on average, moderately – strongly agreeing that cybersecurity tasks are easy to undertake (Mdn = 6.5, IQR = 1). Performance expectancy (Mdn = 6, IQR = 1.5), social influence (Mdn = 5, IQR = 2), facilitating conditions (Mdn = 6, IQR = 1.5) and trust (Mdn = 3, IQR = 3) were also analysed with participants interestingly, on average, slightly disagreeing that cybersecurity measures can be trusted. Of the five UTAUT factors included within the study, only effort expectancy (*r* = .30, n = 71, *p* = .012; **E1 H3j**) was found to significantly relate to reported behaviour both positively and with a medium-low effect size. Performance expectancy (*r* = -.21, n = 71, *p* = .074), social influence (*r* = .10, n = 71, *p* = .426), facilitating conditions (*r* = .19, n = 71, *p* = .122) and trust (*r* = -.14, n = 71, *p* = .230) were not found to significantly relate.

Again, utilising Spearman’s Rho correlation analyses, several perceptual factors from behaviour change theory were explored to determine whether a significant relationship could be determined between these factors and reported cybersecurity behaviour (**S1 H3)**. The following constructs were found to positively relate to perceived cybersecurity behaviour with a significance level of $p < 0.05$. Threat appraisal (Mdn = 6, IQR = 1; **S1 H3a**) with a medium effect size (*r* = .36, n = 71, *p* = .002), security self-efficacy (Mdn = 5.5, IQR = 1; **S1 H3c**) with a large effect size (*r* = .66, n = 71, *p* = < .001) and information security attitude (Mdn = 6, IQR = 1; **S1 H3e**) with a medium effect size (*r* = .43, n = 71, *p* = < .001). However, response efficacy (Mdn 5, IQR = 1; *r* = .17, n = 71, *p* = .163; **S1 H3b**), response costs (Mdn = 4, IQR = 2; *r* = -.205, n = 71, *p* = .087; **S1 H3d**) and subjective norms (Mdn = 5, IQR = 2; *r* = .12, n = 71, *p* = .333; **S1 H3f**) did not evidence a significant relationship.

The three antecedents of the TPB were analysed to determine potential correlations with reported behaviour including information security experience and involvement (Mdn = 5, IQR = 2; **S1 H4)**, information security awareness (Mdn = 5, IQR = 2; **E1 H3g)** and information security organisation policy (Mdn = 5.5, IQR = 1.5; **E1 H3h).** performance expectancy of cybersecurity tasks was 6 (Mdn = 6, IQR = 1.5), with participants, on average, moderately agreeing that cybersecurity measures are easy to undertake All three factors were found to hold significant positive relationships with reported behaviour, with large effect sizes (*r* = .64, n = 71, *p* = < .001; *r* = .63, n = 71, *p* = < .001; *r* = .54, n = 71, *p* = < .001 respectfully).

Table 1

Relationships Between Individual Differences and Reported Cybersecurity Behaviour

Construct	**Correlation**
Large Effect Size (>.50)	
Security self-efficacy	*r* = .66, n = 71, *p* = < .001
Information security experience and involvement	*r* = .64, n = 71, *p* = < .001
Information security awareness	*r* = .63, n = 71, *p* = < .001
Information security organisational policy	*r* = .54, n = 71, *p* = < .001
Medium Effect Size (>.30 and <.49)	
Information security attitude	*r* = .43, n = 71, *p* = < .001
Threat appraisal	*r* = .36, n = 71, *p* = .002
Conscientiousness	*r* = .34, n = 71, *p* = .004
Social risk-taking	*r* = .33, n = 71, *p* = .004
Impulsivity	*r* = -.30, n = 71, *p* = .011
Effort expectancy	*r* = .30, n = 71, *p* = .012
Small Effect Size (>.10 and <.29)	
Psychological ownership	*r* = .27, n = 71, *p* = .021

Finally four additional perceptual and socio-emotional (**S1 H5)** constructs were analysed within the framework to identify their relationship with reported cybersecurity behaviour including organisational commitment (Mdn = 5, IQR = 3; **S1 H4c**), psychological ownership (Mdn = 5, IQR = 2; **S1 H3i**) intrinsic maladaptive rewards (Mdn = 1, IQR = .5; **S1 H5a**) and extrinsic maladaptive rewards (Mdn = 1, IQR = 2; **S1 H4b**). Interestingly both styles of maladaptive rewards were rated on average by participants rating themselves as being very unlikely to wish to gain from loss to their organisation despite their profile being anonymous to the organisation they work with. Of these factors only psychological ownership significantly related to reported behaviour with a small effect size (*r* = .27, n = 71, *p* = .021), yet organisational commitment (*r* = .19, n = 71, *p* = .109), intrinsic maladaptive rewards (*r* = -.22, n = 71, *p* = .068) and extrinsic maladaptive rewards did not (*r* = .06, n = 71, *p* = .625). Table 1 provides a visual representation of the factors found to significantly relate to reported cybersecurity behaviour, as well as the level of effect these relationships possess.

Discussion

Study 1 provided an exploratory investigation into how several previously reported end-user demographics and individual differences – brought together within the same tool – significantly relate (or not) to reported cybersecurity behaviour. The factors included in this study had, in previous research, either been identified as correlating with/predictive of cybersecurity behaviour across a number of papers, theories and contexts or were included for exploratory purposes.

The key motivation behind Study 1, was to begin exploration into the generation of a powerful cybersecurity behaviour measurement tool, that – with further modification – based on findings – can be used periodically within organisations to health-check human cybersecurity vulnerability and therefore explain their perceived experience. The framework and associated metrics that will result from all three studies can inform more targeted intervention that can help benchmark and reassess status – post intervention to gauge control mechanism success. Factors included in iteration 1 of the framework can be found in Figure 5.

It was first hypothesised that reported cybersecurity behaviour would significantly differ between age groups and participant gender, with those both younger and female anticipated to be more at risk. Despite findings within previous literature, no significant differences were found within the current study between age and gender types and reported cybersecurity behaviour. Prior research did however largely focus on very specific cybersecurity tasks e.g., device securement or password management found within the SeBIS (Egelman & Peer 2015), rather than the more global perception of personal cybersecurity behaviour – as measured within the current study. Whilst no distinct hypothesis to date can be drawn around the impact of level of education on reported cybersecurity behaviour, this factor was included within the current study for exploratory purposes. No significant relationships were identified, and therefore there is no new evidence relating to this factor. However, approximately 50% of the sample were well educated e.g., studying or working within academia (with e.g. UG degree or higher qualifications) and therefore this finding cannot be generalised across the whole population.

Several additional participant individual differences were also investigated to determine whether there were significant relationships with reported cybersecurity behaviour, providing organisations with potential focus for future intervention. In respect of participant personality traits, it was predicted that more secure behaviour would be reported in both those more extroverted and conscientious. Results from Study 1 did find conscientiousness to be a personality trait significantly related to cyber behaviour, however extroversion was not. Those more conscientious are believed to be more self-controlled and orderly, with a higher chance of achievement in education, leadership and even marital stability (Roberts et al., 0). Higher conscientiousness is also seen in those more likely to complete work tasks both thoroughly and diligently and it therefore logical that conscientious users would also be more risk-aware in their cyber decisions. It is possible that the lack of relationship found between behaviour and extroversion is again due to the previous focus on a small set of very specific cybersecurity tasks, e.g., device securement rather than cybersecurity behaviour *globally and across behaviours.*

In reference to risk-taking preferences, previous research highlighted health and safety, ethical and financial risk-taking as significantly related to cybersecurity behaviour, with it therefore hypothesised that these factors would relate to cybersecurity behaviour within this current study (Egelman & Peer, 2015; Gratian et al., 2018). In contrast, it was found in the current study that security behaviours were most related to social risk-taking, with those more likely to disagree with the views of peers and significant others. It is perhaps those more comfortable in disagreeing with those around them, who would also be more likely to act against the shadow security workarounds that are often undertaken within the workplace, for example refusing to share a personal password with a colleague so that they can access their email inbox during annual leave (Kirlappos, 2016; Kirlappos et al., 2015; Kirlappos et al., 2014).

Impulsivity, defined as premature action or action imposed prior to conscious thought, was also predicted to be related to reported behaviour within this study. This suggests that those acting impulsively do so prior to the application of logic (Egelman & Peer 2015; Parsons et al., 2013). This hypothesis was upheld, confirming that a relationship does appear to exist between cybersecurity behaviour and impulsivity, highlighting a need to generate interventions that can help slow down the decision-making process, allowing for more time to more logically process information. For example, increasing the number of steps required to click on a link – i.e., adding implementation costs. An additional socio-psychological factor found to relate to cybersecurity behaviour was effort expectancy, measured within the acceptance of cybersecurity questionnaire. Effort expectancy is measured along a scale, defining how easy or difficult a person finds cybersecurity tasks to undertake. Participants finding such tasks easier to explicate were also found likely to report positive cybersecurity behaviour. This again supports the findings of previous literature, with effort expectancy influencing positive and secure behaviour in relation to mobile commerce (Alrawi et al., 2020), mobile payments (Ariffin et al., 2020) and mobile banking (Ivanova & Kim, 2022). None of the other factors within the UTAUT were found to significantly relate to cybersecurity behaviours, suggesting there to be no relationship between cybersecurity behaviour and the usefulness of cybersecurity technology, availability of resources to support, related social views or trust in its systems.

Taken together, findings in relation to participant socio-psychological factors suggest that *secure behaviour is more likely to be witnessed in those that take more time to consider their behaviour, are comfortable disagreeing with the behaviours of others and feel that cybersecurity behaviours are worth the effort*. Interventions should therefore focus on providing decision-making 'speed bumps', to decrease the consequences of unconscious decision-making. That said, the introduction of such speed bumps may in turn impact perceptions around effort expectancy with employees working even harder to find shadow workarounds. Such interventions should therefore be applied with caution, and research conducted it relation to their impact. What could also be useful is the introduction of a feedback tool, that makes it easier for employees to speak or act against the 'risky' shadow security behaviours witnessed, encouraging social risk-taking, as well as providing the ability for employees to discuss views on interventions that are impacting effort expectancy.

A number of propositions were also made in relation to human perceptual factors and reported cybersecurity behaviour (threat appraisal, cyber-security attitude, subjective norms, response efficacy, self-efficacy, response costs, psychological ownership, cybersecurity awareness, cybersecurity organisation policy). Of these metrics – security self-efficacy, the perception from end-users around their ability to perform cybersecurity tasks, was found to hold the largest relationship with reported behaviour: a finding supported by research involving coping appraisal within the health domain (Floyd et al., 2000; Milne et al., 2000) and their manipulation within the cybersecurity domain (van Bavel et al., 2019). Self-efficacy was also identified by Bandura (1977) as being the most influential factor determining human behaviour, with higher perceptions around task ability, resulting in an increased likelihood of exerting effort in task execution. Bandura (1977) suggests four ways to positively influence self-efficacy – increasing experience, witnessing the success of others, social encouragement, and reductions in physiological senses of stress. It is therefore important that employees are not only supported to increase their own ability in cybersecurity tasks, but that a culture is built where witnessing the success of others and social encouragement around security is also fostered. Something perhaps more challenging with people more likely to work from home. Observing the success of peers will not only increase self-efficacy but also support knowledge transfer, particularly tacit knowledge that is problematic to explicate (Elliot et al., 2011; Elliot & McGregor, 2001; Nicholls, 1984).

Information security attitude, the perception of securing information, in this case online, was also found to have a significant relationship with reported behaviour, supporting research findings found within the Safa et al. (2015) paper. This finding also reinforces Azjen’s (1991) TPB review paper where attitudes were repeatedly found to be a factor influencing human intentions and behaviours. Ajzen and Fishbein (1975) explain *attitude* as a construct relating to the expectancy-value theory, where behaviour execution rests on the expected chance of achieving the task alongside the value placed upon it. Improving end-user attitude towards cybersecurity may therefore hinge on increasing evaluation of the safety of an organisation’s systems as well as internal perception of ability.

Threat appraisal was the final factor from behaviour change theory that significantly correlated with reported cybersecurity behaviour, all be it with a smaller effect size. This reinforces the submission of behaviour change theory that choice to act or not to act is related to end-user perception of the potential likelihood and severity of the risk. A review of the related literature offered insight into why end-users may appraise threat as low, with reports of people feeling they had little of importance to hide or lose, and utilised systems already being secure without further action required at the forefront (Jones et al., 2021). Thus, increasing appraisals of threat may hinge on informing employees around system weaknesses and improving knowledge around the value that would be lost should a security breach be experienced. Together, these findings support the very essence of human behaviour change theory, with end-users required to view cybersecurity as achievable, a breach as highly possible and protecting company systems as valuable.

Significant relationships were also found between reported cybersecurity behaviour and the three antecedents of the influencing factors in the TPB. Safa et al. (2015) indicated IS experience and involvement to be a precursor of perceived behavioural control (or self-efficacy), IS awareness the antecedent for IS attitude and IS operation policy the precursor of subjective norms. IS awareness was also found to positively relate to reported cybersecurity behaviour with those more aware of how to remain up to date around security concerns also more likely to report conscious security behaviours. Finally, IS operation policy was also found to positively relate to reported cybersecurity behaviour despite subjective norms (covered later in this discussion), its potential successor, not reaching significance. This suggests that those recognising the value in company security policy will also report behaviours that have company risk in mind. The three antecedents mentioned above were in fact found within Study 1 to correlate more highly with reported behaviour than the factors they are suggested to moderate. Taken together, the findings indicate that – *increasing employee perception of their involvement in cybersecurity tasks, continually updating their knowledge around current risks and protective behaviours and helping them see value in organisation policy will contribute to improved security behaviour being achieved.*

A factor also found to have a positive relationship with cybersecurity behaviour was psychological ownership, the feeling that technology is an extension of oneself. Previous research has found increased psychological ownership to be related to higher levels of attachment and perceived responsibility of an object, although more so in home computers than mobile phones (McGill & Thompson, 2017; Peck et al., 2021). This is possibly due to computers being purchased outright within the home setting examined, and mobile phones ordinally having a monthly ‘rental’ payment plan. It is suggested that psychological ownership can be increased by investing more time in an object, controlling it how one wishes, and improving cognitive and affective evaluations of the object, with self-investment the strongest factor (Lee & Chen, 2011).

IS experience and involvement, a factor sitting beneath the model’s habitual theme, was predicted to be positively related to cybersecurity behaviour with those perceiving themselves as more involved in the domain also reporting more secure behaviour. As predicted, those more experienced and enmeshed in the cybersecurity chain, reported exhibiting positive behaviour. This can potentially be made more difficult in organisations that have a large IT department or even a separate cybersecurity team. In many organisations, employees receive a one off – e.g., annual training session on cybersecurity making it difficult for them to feel part of the solution. Including employees in as many aspects of cybersecurity as possible - such as policy creation and providing them with feedback when their behaviour has had a positive influence e.g., when they have successfully reported a phishing email, will not only increase perceptions of involvement but in turn potentially improve their level of experience.

Despite Study 1 providing confirmation of a large number of hypothesised relationships between factors included within the framework and reported cybersecurity behaviour, a number of anticipated findings were not upheld. Of the three key elements found to previously be important in the appraisal of a suggested response in behaviour change theory (self-efficacy, response efficacy, response costs) only self-efficacy was found to significantly relate. This is however not a complete surprise, as despite their inclusion within behaviour change models, a lack of clarification around their importance was evident within the literature and therefore these particular hypotheses largely exploratory. Subjective norms, another key construct found within behaviour change theory, was also not identified as significantly related to cybersecurity behaviour. Again, despite its inclusion, confirmatory literature was absent and in fact it suggested that social norms only become of importance if self-efficacy is particularly low. (Ajzen, 1991; McGill and Thompson, 2017)

In relation to hypothesised socio-emotional factors, neither intrinsic nor extrinsic maladaptive rewards were found to relate to reported behaviour. However, it must be noted that, on average, that participants rated themselves as being very unlikely to wish to gain from their organisation experiencing loss, suggesting that participants may find it difficult to express such feelings of negativity, either due to concerns for external repercussions, social desirability, or perhaps not wanting to admit such feelings to themselves. Finally, organisational commitment did not reach significance in relation to cybersecurity behaviour despite previous research suggesting a link (Ertan et al., 2020; Karim & Noor, 2017). However, in research by Reeve et al. (2020), whilst organisational commitment was found to influence cybersecurity behaviour in relation to mobile phones, this was not found to be the case in malware or phishing attacks, with perhaps the lack of relationship due to the use of a global cybersecurity behaviour scale within this study.

To summarise, the main aim of Study 1 was to better understand the experience of the human in relation to cybersecurity and the vulnerabilities that maybe putting them at risk of threat. With the key output being the first iteration of a tool that can – even without refinement (although see Study 2 and 3) – be used by organisations to better understand where susceptibility may be being experienced, to allow a better tailored application of intervention. Exploration into how a number of individual differences correlate with cybersecurity behaviour was deployed, indicating that *threat appraisal, security self-efficacy and attitude towards cybersecurity significantly relate*. This indicates that organisations should focus their efforts on helping employees remain up to date on current security risk, improve attitudes towards cybersecurity and its strategy by involving employees in improving its usability, increase end-user experience and involvement, providing support for employees to feel included and become proficient in effective cybersecurity behaviour (see Table 2). Less direct intervention may include increased allowances for personalisation within technology to intensify attachment as well as encourage a culture where knowledge is both shared and challenged. Finally, attention should be given to providing tools such as pop-ups and decision-making support systems that generate end-user ‘bumps in the road’, encouraging due diligence, driving a user into more conscious thought and providing support for those more impulsive or less conscientious.

Table 2

Main Recommendations to Help Organisations Support Mitigation of Human Risk

Metric	Recommendation
Information Security Awareness	Provide a culture where employees stay up to date on current risk and coping strategies.
Information Security Organisation Policy	Include employees in the optimisation of cybersecurity policy to increase perception of its value and increase its use.
Information Security Experience and Involvement	As well as involving employees in policy optimisation, utilise feedback around their sentiment towards cybersecurity training and provide training that supports not just education but skill proficiency.
Information Security Self-efficacy	Ensure employees can not only proficiently conduct the required cybersecurity skills, but also perceive themselves as having the personal ability to do so.
Threat Appraisal	Regularly update employees on cyber incidents that have taken place both inside and outside of the organisation.
Information Security Attitude	Help employees weigh up the benefits versus costs of cybersecurity behaviours by increasing risk perception as well as simplifying required counter actions.

Summary and Next Steps

The key objective of Study 1 was to improve understanding around key human vulnerabilities in relation to cybersecurity and the human experience, in order to address insufficiencies in current security mitigation interventions. The initial objective was to assemble a number of key constructs found in the broad inventory of behavioural theories and individual differences currently being used to investigate human decision-making and refine them into a structured assessment framework (see Figure 5). Once an initial set of metrics is defined, subsequent research can focus on a further iteration of the framework using a wider industry sample. The main objective of Study 2 will be to further advance this framework with the anticipation of corroborating the findings of Study 1. Study 2 will extend this research by investigating which of these related factors predict reported cybersecurity behaviour with the highest magnitude.

Study 2

Study 2 Aims

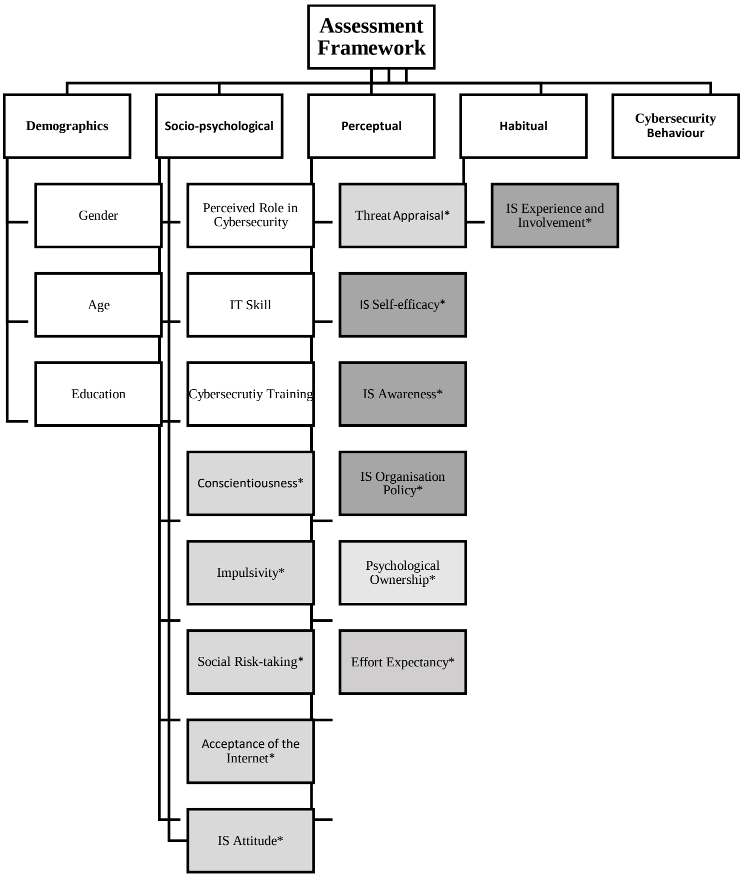
Study 2 will address the following PhD aims:

- Investigate whether a number of individual factors found to relate to reported cybersecurity behaviour in Study 1, will also associate within a sample working within one organisation (see Figure 6);
- Use exploratory factor analysis (EFA) to identify any latent variables that may sit above the large number of related factors further refining the framework;
- Extend this research by investigating which of these factors may together predict reported cybersecurity behaviour, providing support for a potential human-centric cybersecurity assessment framework;

- Generate a third and further refined iteration of the human-centric cybersecurity assessment framework that can be investigated with a larger sample of the general population (Study 3).

Figure 6

Iteration 2 of the Human-centric Cybersecurity Assessment Framework



Note *Significant correlation (within Study 1), the darker the grey the larger the effect size.

Research Hypotheses

Study 2 was designed to (a) attempt to confirm correlational findings from Study 1 and previous research (b) conduct an exploratory factor analysis for item reduction purposes and

(c) apply regression analyses to further investigate how the related constructs may fit into a predictive model. A number of hypotheses were determined from the key findings of Study 1:

Socio-psychological Factors. S2 H1 A number of socio-psychological factors found to significantly correlate with reported cybersecurity behaviour in Study 1, with it anticipated that they would also significantly relate within this study (Egelman and Peer, 2015; Gratian et al., 2018):

S2 H1a Those more conscientious were anticipated to report more secure cybersecurity behaviour.

S2 H1b Less desirable behaviour would be reported by those more impulsive.

S2 H1c Social risk-taking would significantly positively relate to reported behaviour.

S2 H1d Psychological ownership would positively correlate with reported cybersecurity behaviour.

S2 H1e IS attitude would also positively correlate.

Perceptual Factors. S2 H2 It was also hypothesised that a number of perceptual factors would significantly relate to reported cybersecurity behaviour (Burns and Roberts, 2013; Carpenter et al., 2019; Liang & Xue, 2009; Pickering et al., 2021; van Bavel, et al., 2019).

S2 H2a Threat appraisal would positively correlate.

S2 H2b Self-efficacy would positively correlate.

S2 H2c IS awareness would positively correlate.

S2 H2d IS organisation policy would positively correlate.

S2 H2e Effort expectancy would negatively correlate.

Habitual:

S2 H3 It was predicted that the more experienced and involved participants feel they are in cybersecurity, the more secure their reported behaviour (Safa et al., 2015).

Method

Participants. A sample of one hundred and fifty-six participants, 84% male and 16% female, were recruited within a multinational corporation, via their internal UK Intranet with an average age of 40.64 (*SD* 9.81) years. Participants were not rewarded for taking part in this Study and signed up in order to help further research around improving security within cyberspace.

Design. A between-participants correlational design was employed utilising Spearman's Rho non-parametric analyses to investigate whether a number of individual differences (see Figure 6 for details) significantly relate to reported cybersecurity behaviour (measured by the behavioural factor within the TPB and PMT questionnaire). Next, an investigation took place to uncover whether cybersecurity behaviour significantly differed across several user characteristics (gender, age) analysed utilising Kruskal-Wallis test of differences. An exploratory factor analysis was then conducted to determine any latent variables overarching the confirmed relationships in order to reduce the large number of factors and underlying items under analysis for simplification. Final analyses were then undertaken in the form of a stepwise regression to explore which of the latent constructs formulated from the EFA significantly explain the variance in reported cybersecurity behaviour. This particular form of regression analysis was utilised due to a lack of guidance on the level of predictive power across the variables, and therefore lack of understanding on how to enter them into the model.

Materials and Procedure. As with Study 1, participants accessed the study via *Qualtrics©*, in order to complete a battery of measures. Participants were first presented with an introduction sheet (Appendix A) and a request for informed consent to take part (Appendix B). Each participant also completed an optional demographics form collating information on their age, gender and level of education. Participants were asked to categorise their highest level of education across six categories including GCSEs, A-levels, undergraduate degree, master's degree, PhD/Doctorate, or other. Also rated by participants were work-role importance in cybersecurity, from 1 (extremely important) to 5 (not at all important), level of IT skill rated from 1 (poor) to 5 (excellent) and level of training in cybersecurity from 1 (none) to 5 (expert). All of the following questionnaires and questions within those questionnaires were randomised. Participants were asked to complete the IPIP personality traits measure (Appendix C; Goldberg et al., 2006), an instrument measuring risk-taking preferences (Appendix D; Blais & Weber, 2006). The Barratt Impulsiveness Scale (BIS-11; Appendix F; Patton et al., 1995), as well as the items measuring effort expectancy within the UTAUT acceptance of the internet questionnaire was used (Appendix G; Venkatesh et al., 2003). The combined TPB and PMT questionnaire (Appendix H; Safa et al., 2015) was also utilised to measure threat appraisal, self-efficacy, IS awareness, IS organiation policy, IS experience and involvement. Finally additional questions used within McGill & Thompson (2017) and Posey et al. (2015; Appendix I) questionnaire was used to measure psychological ownership. Cybersecurity behaviour, was measured by the behaviour construct within the PMT and TPB questionnaire, rated from 1 (strongly disagree) to 7 (strongly agree) with five statements such as *'I consider security experts recommendations in my information security manner'*. After completion of all measures participants were provided with a study debrief (Appendix J) and thanked for taking part.

Results

The key objective of Study 2, was to improve understanding around how employees experience cybersecurity, and the vulnerabilities that may lead to attack, in order to improve its measurement and intervention. This involved an investigation into relationships between several factors previously suggested in research as linked to cybersecurity behaviour, in order to work towards the generation of an assessment framework that can be iteratively reduced to form a basis for organisations to understand and measure human cybersecurity risk moving forward. Study 2 looked to further this analysis, by attempting to verify correlational findings across a wider sample of employees within an industrial organisation, clarify whether these relationships reveal any overarching unobserved variables and extend the exploratory research by investigating which of these relationships explain the greatest variance in reported cybersecurity behaviour.

Reliability of Measures. Cronbach's Alpha tests of internal consistency were applied to all measures used in the study. Good reliability was found for the Barratt Impulsivity questionnaire (α = .73) and acceptable to good reliability was calculated for all subscales of the DOSPERT risk-taking preferences questionnaire (α = .60 - .82). The IPIP personality subscales reached acceptable to good reliability (α = .61 - .82) except for conscientiousness which had poor reliability (α = .54) with no improvements should items be removed. Effort expectancy (α = .83) from the UTAUT showed good reliability. Finally for the combined TPB and PMT questionnaire all subscales displayed good reliability (α = .74 - .89) as did the set of statements used to measure psychological ownership (α = .88). The key assumptions for parametric testing were not met due to the use of ordinal data, and therefore non-parametric statistical tests were utilised. Assumptions for all statistical tests used were analysed and met. Any missing observations within the dataset were replaced with the grand mean for each question and any outliers determined by 3 IQR from the mean were windsorized to the next available value not considered extreme.

Cybersecurity Behaviour. Cybersecurity behaviour, operationalised within the study via the behaviour measure found within the combined TPB and PMT questionnaire (Safa et al., 2014), was found to have a median score across participants of six (Mdn = 6, IQR = 2) indicating, as with Study 1, that on average the sample group moderately agree that their cybersecurity behaviour is both conscious and favourable.

Demographics. A Kruskal-Wallis test of differences was first conducted between gender (*H* = 2.090, *p* = .148) and level of education (*H* = .632, *p* = .987) with cybersecurity behaviour, with no significant differences found. However, a significant difference was discovered between age group and perceived cybersecurity behaviour (*H* = 12.803, *p* = 0.030) with the age group 45-54 displaying significantly more conscious cybersecurity behaviours than both the 25-34 (*p* = .014) and 35-44 (*p* = .028) age ranges. Similarly, the 55 – 64 age range was identified as significantly more likely to report conscious cybersecurity behaviours than the 25 – 34 (*p* = .006) and 35 – 44 (*p* = .013) age groups. These findings differ from those in Study 1, where no significant differences between age groups were found.

Individual Differences. Non-parametric correlational analyses (Spearman’s Rho) were then applied to determine any significant relationships between reported cybersecurity behaviour and a number of factors found to significantly relate to reported behaviour in Study 1, as supported by previous literature. Whilst Study 1 utilised a total of 71 participants from Cardiff University (staff and PhD students) and Prolific an online participant tool. The current study employed a larger sample within a multinational corporation, with a larger proportion of males potentially impacting comparisons (see Method section of Study 2). Of particular interest is how the correlation analyses from both studies revealed the same six related factors have the largest effect sizes (see Table 3 and below).

Correlational analyses explored the relationships between reported cybersecurity behaviour and a number of socio-psychological factors (**S2 H1;** personality, impulsivity, risk-taking preferences). In relation to personality sub-types (**S2 H1a)**, associations were analysed between reported cybersecurity behaviours and levels of extraversion (Mdn = 3, IQR = 1.5), conscientiousness (Mdn = 4, IQR = 1), agreeableness (Mdn = 4, IQR = .5) neuroticism (Mdn = 2.5, IQR = 1) and openness to experience (Mdn = 4, IQR = .5). Unlike Study 1, no significant relationship was found between reported cybersecurity behaviour and level of conscientiousness (*r* = .063, n = 153, *p* = .435) nor the other personality types under investigation including extraversion (*r* = .08, n = 153, *p* = .328), agreeableness (*r* = .09, n = 153, *p* = .078), neuroticism (*r* = -.02, n = 71, *p* = .801), or openness to experience (*r* = .130, n = 153, *p* = .102).

Participant risk-taking preferences (**S2 H1c**) were also analysed to examine whether a significant relationship could be found with reported cybersecurity behaviour. As predicted from Study 1, social risk-taking propensity (Mdn = 5, IQR = 2) was found to significantly correlate with reported behaviour with a small effect size (*r* = .23, n = 155, *p* = .004). It was also found that those less likely to take ethical risks (Mdn = 1, IQR = 1) were significantly more likely to report positive behaviour, also with a small effect size (*r* = .21, n = 155, *p* = .009). However, as with Study 1, no significant relationships were found in relation to recreational risk-taking (Mdn = 3.5, IQR = 3.5; *r* = .05, n = 155, *p* = .536), financial risk-taking (Mdn = 1, IQR = 1; *r* = .14, n = 155, *p* = .089) and health/safety risk-taking (Mdn = 2, IQR = 1.5; *r* =- .05, n = 155, *p* = .554) and reported behaviour.

In respect of impulsivity (**S2 H1b**), on average participants rated themselves as likely to occasionally behave impulsively, with a large dispersion (Mdn = 2, IQR = .5). However, despite a significant relationship being found in Study 1, this was not replicated within Study 2 (*r* = .14; n = 155, *p* = .087). Next analysed was whether participant attitude (**S2 H1e**; Mdn = 5, IQR = 2) towards cyber-security, as found within the TPB, correlated with reported behaviour, with this factor again significantly relating, with a large effect size (*r* = .68, n =

155, p < .001). Also found within this study, as predicted by Study 1, was the significant link between reported behaviour and psychological ownership (Mdn = 4, IQR = 2; **S2 H1d**), the sense of attachment people feel towards their work technology (r = .30, n = 155, p < .001), with a medium effect size.

Analysed next were several perceptual factors found within a number of leading theories of behaviour change e.g., PMT. In relation to threat appraisal, results found that on average, participants strongly agree that there is a potentially high probability and severity of threat, if caution in behaviour is not taken (Mdn = 7, IQR = 2; **S2 H2a**). Upon analysis this factor was found to significantly relate to reported cybersecurity behaviour (r = .70, n = 155, p > .001) with a large effect size. Next, security self-efficacy (**S2 H2b**) was analysed with participants on average reporting that they ‘agree’ to having the skills required to protect themselves and their organisation from a cyber-attack (Mdn = 6, IQR = 1.5), this factor was also found to significantly relate to reported behaviour (r = .54, n = 155, p < .001) with a large effect size. Another factor analysed, that is specific to the TPB model, is subjective norms (Mdn = 5, IQR = 2) with this factor also significantly relating to reported behaviour but with a small effect size (r = .28, n = 155, p > .001). This differs from Study 1 where no relationship was found. Effort expectancy, from the acceptance of technology model UTAUT, was also measured, with participants on average moderately agreeing that cybersecurity tasks are easy to undertake (Mdn = 6, IQR = 1). As with Study 1, effort expectancy was found to significantly relate to reported cybersecurity behaviour with a small effect size (r = .18, n = 155, p = .029; **S2 H2e**). A number of previously defined antecedents of the factors from TPB were next analysed. ISA - the potential precursor of information security attitude (Mdn = 6.5, IQR = 1; **S2 H2d**) - was found to significantly relate to reported behaviour with a large effect (r = .68, n = 155, p < .001) as did information security experience and involvement (Mdn = 7, IQR = 1; *r* = .64, n = 155, *p* < .001; **S2 H3**) was also found to significantly relate to reported behaviours (*r* = .64) with a large effect size (see Table 3).

Finally, the habitual factor, ISOP the precursor to subjective norms was analysed (Mdn = 7, IQR = 1) to determine its relationship with reported cybersecurity behaviour. A significant correlation between these factors was expected as found within Study 1 (**S2 H3**) with this hypothesis upheld (*r* = .64, n = 155, *p* < .001) with a large effect size (see Table 3).

Table 3

Correlational Comparisons Between Studies 1 and 2

Construct	**Correlation**	**Correlation**
	Study 1	***Study 2***
Large Effect Sizes in Studies 2 (>.5)		
Threat appraisal	*r* = .36, n = 71, *p* = .002	*r* = .70, n = 155, *p* < .001
Information security awareness	*r* = .63, n = 71, *p* < .001	*r* = .68, n = 155, *p* < .001
Information security attitude	*r* = .43, n = 71, *p* < .001	*r* = .68, n = 155, *p* < .001
IS experience and involvement	*r* = .64, n = 71, *p* < .001	*r* = .64, n = 155, *p* < .001
IS organisation policy	*r* = .54, n = 71, *p* < .001	*r* = .57, n = 155, *p* < .001
Information security self-efficacy	*r* = .66, n = 71, *p* < .001	*r* = .54, n = 155, *p* < .001
Medium Effect Sizes in Study 2 (>.3 and <.49)		
Psychological ownership	*r* = .27, n = 71, *p* = .021	*r* = .30, n = 155, *p* < .001
Small Effect Sizes in Study 2 (>.1 and <.29)		
Subjective Norms	Did not correlate	*r* = .28, n = 155, *p* > .001
Social risk-taking	*r* = .33, n = 71, *p* = .004	*r* = .23, n = 155, *p* = .004
Ethical risk-taking	Did not correlate	*r* = .21, n = 155, *p* = .009
Effort expectancy	*r* = .30, n = 71, *p* = .012	*r* = .18, n = 155, *p* = .029
Did not correlate (Study 2)		
Conscientiousness	*r* = .34, n = 71, *p* = .004	Did not correlate
Impulsivity	*r* = -.30, n = 71, *p* = .011	Did not correlate

Note. Effect sizes found in Study 1 are indicated in grey, to ease visualisation. The darker the grey, the larger the effect size found.

Exploratory Factor Analysis. During correlation analyses in both Studies 1 and 2, it became evident that a large number of variables appear to correlate with both reported cybersecurity behaviour, as well as each other. It was therefore deemed necessary to conduct reduction analyses to determine whether any overarching latent variables exist that better account for the factors observed. All factors that were found to correlate with reported behaviour either within Study 1 or Study 2 were included within this analysis.

Bartlett's test of sphericity was first reviewed, to determine whether the correlation matrix is appropriate to factor analyse. Significance was present (*p* > .001) and therefore suitability was found and exploratory factor analysis (EFA) conducted. The Kaisler-Meyer-Ilkin (KMO) measure of sampling adequacy was also reviewed to determine EFA applicability with a score of .832. This score was above the required .6 and therefore meritorious for analysis (Kaiser & Rice, 1974). All individual items that were included in the matrix were found to contribute to the factor solution (items under .30) with no items under this score identified (and therefore none requiring removal). Also examined was whether multicollinearity was present with extremely high correlations causing potential problems within the correlation matrix. This was determined by examining whether any correlations fell over .80 (Watkins, 2021), with two items from the self-efficacy measure evidencing an issue - 'I have the skills to protect my business and private data' (.880) and 'I have the expertise to protect my business and private data' (.878). As the former provided variance to the factor solution, the latter was removed from analysis with the former item no longer correlating over .80 with any other factors within the model.

Upon conducting EFA, a principal axis factoring extraction method was utilised with no rotation method initially applied to generate a scree plot and determine the number of latent variables present, with two factors identified before the elbow and three factors found to account statistically for 36.34% of the variation. A varimax rotation was then applied, however a number of variables were found to cross-load and therefore a promax rotation was utilised in order to reduce the amount of cross-loadings present. This rotation resulted in the fewest number of cross-loadings evident, however despite this, two variables did cross-load irrespective of the rotation applied and therefore these were dropped from the analysis including the item ‘I understand the risk of information security incidents’ from the ISA measure and the item ‘I have suitable capability in order to manage information security risk due to my experience’ from the ISEI measure. Removal of these items only reduced the variance reported slightly from 36.34% to 35.22% (see Table 4).

Table 4

Factor Loadings for Exploratory Factor Analysis

No.	Factor	Item	Loading	Eigenvalues	Variance
1	Cybersecurity Awareness	Careful information security behaviour is necessary (ATT 1)	.777	25.400	24.267%
		My attitude towards careful information security behaviour is favourable (ATT 2)	.762		
		My experience helps me to recognise and assess information security threat (ISEI 1)	.756		
		I believe that careful information security behaviour is valuable in an organisation (ATT 3)	.734		
		Practising careful information security behaviour is useful (ATT 4)	.725		
		My experience increases my ability to have a safe behaviour in terms of information security (ISEI 2)	.722		
		I keep myself updated in terms of information security knowledge to increase my awareness (ISA 1)	.721		
		Hackers attack with different methods and I should be careful in this dynamic environment (TA 1)	.704		
		Information security policies and procedures affect my behaviour (ISOP 1)	.661		
		Behaviour in line with organisational information security policies and procedures is of value in my organisation (ISOP 2)	.653		
		I have a positive view about changing users' information security behaviour to be more considered (ATT 5)	.648		
		I know the probability of security breach increases if I do not consider information security policies (TA 2)	.646		
		I could fall victim to different kinds of attack if I do not follow information security policies (TA 3)	.645		
		Careful Information security behaviour is beneficial (ATT 6)	.631		
		I can sense the level of information security threat due to my experience in this domain (ISEI 3)	.627		

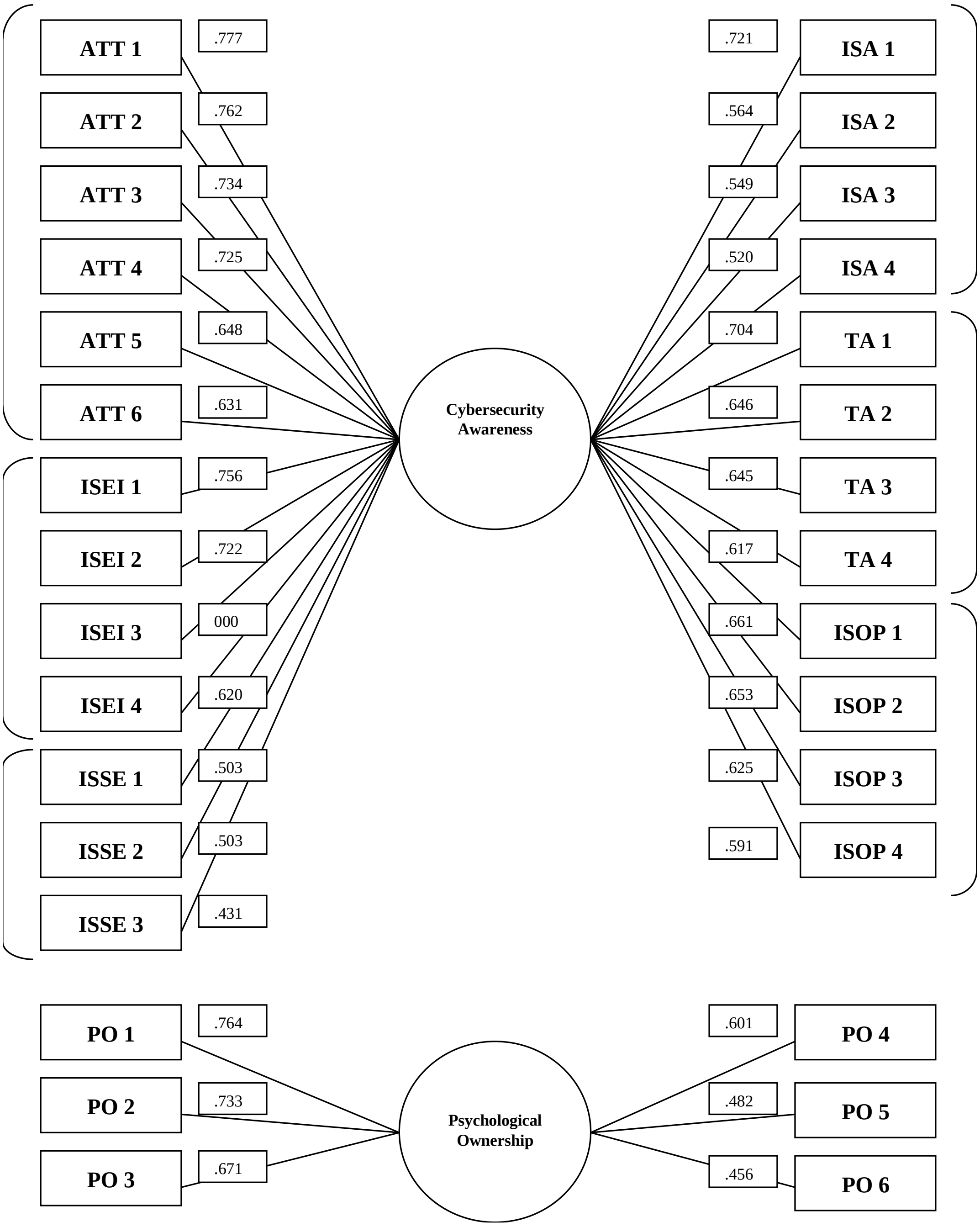
		Information security policies and procedures have attracted my attention (ISOP 3)	.625		
		I am involved with information security and I care about my behaviour in my job (ISEI 4)	.620		
		The security of my data will be weak if I do not consider information security policies (TA 4)	.617		
		Information security policies and procedures are important in my organisation (ISOP 4)	.591		
		I share information security knowledge to increase my awareness (ISA 2)	.564		
		I have sufficient knowledge about the cost of information security breaches (ISA 3)	.549		
		I am aware of potential security threat (ISA 4)	.520		
		I have the skills to protect my business and private data (ISSE 2)	.503		
		I think the protection of my data is in my control in terms of information security violations (ISSE 3)	.431		
		I have the ability to prevent information security violations (ISSE)	.405		
2	Psychological ownership	When I think about it, I see an extension of my life in my work computer (PO 1)	.764	8.114	6.953%
		I personally invested a lot in my work computer, e.g. time, effort, money (PO 2)	.733		
		I personally invested a lot in the software/applications on my work computer, e.g. time, effort, money (PO 3)	.671		
		I see my work computer as an extension of myself (PO 4)	.601		
		I feel a high degree of ownership for my work computer and its contents (PO 5)	.482		
		The information stored on my work computer is very important to me (PO 6)	.456		
3	Ethical Risk-taking	Passing off somebody else’s work as your own	.408	5.534	4.004%

Only factor loadings > .04 are presented (see e.g., Matsunaga, 2010; Watkins, 2021)

As the third factor that was identified (ethical risk-taking) only had one item (‘Passing off somebody else’s work as your own’ the ethical risk-taking measure) loading onto this latent variable, it was not included within the model resulting in only two unobserved variables considered (see Figure 7). Variable 1 has been labelled ‘Cybersecurity Awareness’, due to the underlying items including the original awareness construct, but also general attitude towards cybersecurity, how threat is appraised, experience and involvement in cybersecurity, self-efficacy in the use of its secure measures and views around cybersecurity operation policy. Together, all items generate an unobserved variable that appears to capture the true holistic experience of the human in relation to cybersecurity. The second latent factor includes six of the seven items included within the psychological ownership measure and has therefore maintained the label of ‘psychological ownership’ (see Figure 7).

Figure 7

Factor Analysis Model – Cybersecurity Awareness and Psychological Ownership



Note. Att – Information Security Attitude, ISEI – Information Security Experience and Involvement, ISSE – Information Security Self-efficacy, ISA – Information Security Awareness, TA – Threat Appraisal, ISOP – Information Security Operation Policy, PO – Psychological Ownership.

Regression Analyses. In addition to correlational analyses in order to replicate the findings from Study 1 and the EFA analysis utilised for item reduction, a stepwise regression was conducted using the two labels found within the EFA (cybersecurity awareness and psychological ownership, as well as participant age) to investigate how the key factors identified may fit into a model that significantly explains the maximum variance in reported cybersecurity behaviour. Regression analyses were not undertaken within Study 1 due to its exploratory nature and number of participants resulting in potential lack of power. In stepwise regression, all variables are entered into analysis in unison with models iterated, until the most parsimonious and significant model is found. Iteration halted at model 1 (F (1, 151) = 189.737, $p < .001$) where 55% of the variance in reported cybersecurity behaviour was explained by only one predictive variable in the model - *Cybersecurity Awareness* (adjusted $R^2 = .554$), the latent variable generated as part of the EFA. Psychological ownership and age were however extracted from the model due to neither significantly explaining additional variance to the model. (see Appendix K).

Discussion

The aim of Study 2 was to verify the factors within Study 1, found to significantly relate to reported cybersecurity behaviour but across a larger sample, specific to the industrial organisation part funding elements of this work. The current study then looked to extend upon these findings by conducting exploratory factor analysis (EFA) in order to potentially refine the large number of related variables contained within the framework. Regression analyses were then conducted utilising the refined EFA model, to better understand which of the latent variable(s) unearthed would explain the largest portion of variance in relation to reported cybersecurity behaviour.

Whilst previous literature suggests age and gender to be significantly related to cybersecurity behaviour, this finding was not confirmed within Study 1 of this thesis. However, a significant difference was discovered between age group and perceived cybersecurity behaviour within this study, with those within the 45 – 54 and 55 – 64 age groups reporting significantly more conscious cybersecurity behaviours than a number of younger age groups (namely 25-34 and 35 – 44). This differs from previous studies whereby those aged 18 – 24 appear to behave the most 'risky'. Age was also not found to have predictive power within the regression model analysed in Study 2, with similar findings experienced within the Gratian et al. (2018) paper where age differences we found but predictive power was not. As with Study 1, behaviour was not found to differ across gender types within this study.

A number of factors were predicted to significantly relate, as found within Study 1: conscientiousness, impulsivity, social risk-taking, psychological ownership, threat appraisal, self-efficacy, attitude, awareness, organisation policy, effort expectancy, experience and involvement. Results from Study 2 indicate the same eleven factors to be significantly correlated with reported behaviour, providing additional support for the conclusions of Study 1, and therefore the need to create interventions better tailored to these particular constructs moving forward. However, due to the large number of related factors identified, and inter-correlations across them, an exploratory factor analysis was undertaken to determine whether items informing these metrics load in a way that uncovers a more succinct set of unobserved variables. Results from the EFA indicate that two key latent variables exist, one that solely represents psychological ownership therefore maintaining the original construct title, and another, Cybersecurity Awareness, informed by a total of twenty-five items relating to six different observed constructs (TA, ISSE, IS attitude, ISA, ISEI, ISOP).

The number of observed constructs and determining measurement items loading onto this latent variable implied that a global construct had perhaps been identified. Encapsulating the need for an awareness of threat probability, protection ability, experiences, attitudes, policies and more, suggesting an awareness of cybersecurity generally is required to positively inform behaviour. Cybersecurity awareness is a term regularly used within the field to describe how end-users experience cybersecurity, both in relation to their understanding around threat risk *and* perceptions around their efficacy to conduct behaviours that will help prevent such risk. There have however been long term differences, even across fields, around how awareness is best defined (Chaudhary et al., 2023; Zwilling et al., 2022). It is important to note that cybersecurity awareness programmes used within organisations to provide their employees with updates and education around risk, are often also termed 'cybersecurity awareness', however this is simply describing the mode used to improve levels of awareness, and not awareness itself.

A review of past literature was conducted, with the aim of understanding how awareness was being conceptualised including the use of terms such as situational awareness, assessments of competence, perceptions and psychological aspects, policy, behaviour, task specific knowledge and interventions for improvement (Chaudhary, 2023). Even across the social sciences, the concept of awareness is still in debate, making it even more difficult to determine how cybersecurity awareness should be specifically defined. Gafoor (2012) mentions three forms of awareness in his paper – holding awareness *about* something (knowledge around a topic), awareness *of* something (subjective perceptions of a topic), and awareness around *ability* of something (being conscious of the ability to do something). Awareness has also been previously conceptualised as a lower form of knowledge, surface level, with knowledge far more accurate and specific. However, a more recent review of definitions by Travethan (2017) suggests awareness is actually related to the attention or mindfulness of a subject, in particular its dangers. Similar conceptualisations suggest awareness relates to how mindful people are around certain risks and the need to avoid them, with knowledge at its root (Khader et al, 2021; Zwilling et al., 2022). This definition appears to be particularly useful in cybersecurity awareness, due to its distinct focus on risk.

Within psychological research, awareness is more intensely known as a phenomenon used synonymously with 'consciousness' - *collective experiences within a single individual about a person, situation, item or object* (Marton, 2000). The complexity of awareness detailed by this classification, also aids the field of cybersecurity, in reference to the array of past and present experiences, perceptions, tasks and roles in play. It is believed that humans are capable of holding multiple experiences within awareness, in relation to the very same 'thing'. It is not as simple as being either 'aware' or unaware' of something. Some experiences of awareness may be related directly to the object in question; and others the way is sits within the physical world, spatially or temporally (Marton, 2000). For example, a human can experience a cyber-attack(s) as related to the physical being of a human hacker, or more generally the online environment where it exists. Cyber-attacks may feel spatially close to them, as would a physical robbery, or more distant due to the nature of cyberspace. They can feel temporally near, perhaps when a cyber-attack has been experienced recently, or temporally far due to perceiving them to be a thing of the past or indeed future (threat).

For many years, awareness was conceptualised as a state of mind whereby only a small amount of information is activated in the mind at any given time, replaced by different forms of information as soon as something falls out of use (Carr, 1979). However, awareness is now believed to influence behaviour, even when not subjectively held in mind (Merikle, 1984). Humans can be 'aware' of many things - who they are, what they do, what they are currently doing, but with most of these things not at the forefront of their mind. They are instead experienced across a continuum from implicit to explicit, moving further up the scale as they are externally triggered or become salient (Marton, 2000).

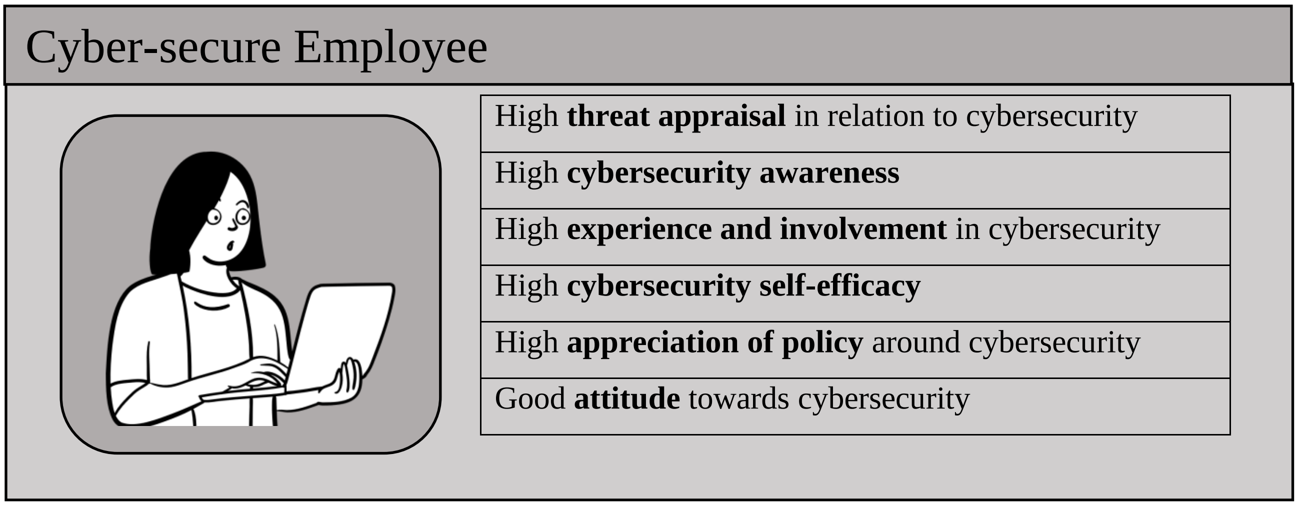
The experiences surrounding awareness will differ from person to person, and situation to situation, depending on e.g., previous exposure. However, humans do not only ‘experience’ awareness in relation to the past, but also in reference to the present context, and beliefs around the future (Marton, 2000). To consider the variables underlying cybersecurity awareness in relation to this:

a) Incidents from the past influence attitude towards cybersecurity, as well as previous experiences, and involvement;
b) The contextual present, including implicit awareness and views around the explicit knowledge held in policy;
c) Beliefs about the future including the probability and severity of threat, and whether employees have the ability to protect themselves from this threat.

As well as the EFA, a regression analysis was performed, utilising the identified latent variables, including psychological ownership and cybersecurity awareness. In order to identify the most parsimonious model available, that represents the largest variance in reported cybersecurity behaviour. cybersecurity awareness alone, was found to explain 55%, suggesting that a large portion of reported behaviour is explained by how employees experience cybersecurity across time (see Figure 8).

Figure 8

Cyber-secure Persona and Metrics for Organisations to Measure Human Vulnerability



Cyber-secure Employee

High **threat appraisal** in relation to cybersecurity

High **cybersecurity awareness**

High **experience and involvement** in cybersecurity

High **cybersecurity self-efficacy**

High **appreciation of policy** around cybersecurity

Good **attitude** towards cybersecurity

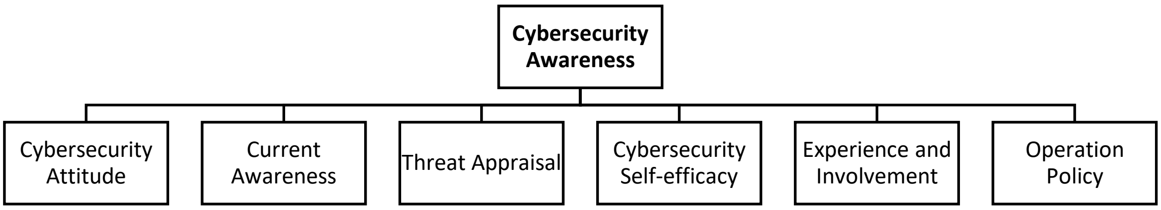
Whilst a significant model was unearthed during regression analyses, age and psychological ownership did not significantly improve its predictive value and were therefore not included. Previous literature had suggested age to be significantly related to reported cybersecurity behaviour, with those aged 18 – 25 at particular risk (Gratian et al., 2015; Parrish et al., 2009; Sheng et al., 2010; Whitty et al., 2015), as found within this current Study (however not confirmed within Study 1). When conducting regression analyses age was not found to have predictive power in relation to cybersecurity behaviour, as confirmed within Study 2 (Gratian et al., 2015). Psychological ownership, whilst found to be significantly related to reported behaviour within both Studies 1 and 2, and the instrument utilised largely validated within the EFA, it also did not add to the predictive power of the model. This could perhaps be due to the relationship being led in fact by the cybersecurity behaviours themselves, with undertaking tasks such as in updating the work computer, perhaps having an influence on the investment quality of psychological ownership. It could also be that psychological ownership as a factor, is important due to its moderating ability only, such as that seen with self-efficacy (Verkijika, 2020). Despite its lack of influence on reported behaviours found within Study 2, it is important that future research continues to understand how psychological ownerships fits with employee intentions to adopt and conduct cybersecurity tasks, and how interventions looking to increase psychological ownership can impact cybersecurity perceptions and in turn behaviour.

Summary and Next Steps

The key objective of Study 2, was to further understand how cybersecurity is experienced by the employee, and the most appropriate factors to measure and manage this experience, within organisations, to help reduce human risk. An overarching latent construct - Cybersecurity Awareness was identified, detailing how more secure behaviour is more likely to be is cybersecurity awareness is high. This includes ensuring that positive past experiences exist in relation to cybersecurity resulting in feelings of involvement and a good attitude towards its, that information security awareness is current and that employees perceive policy to be usable, and finally that perceptions around future risk are realistic, with employees that feel able to counter those risks as and when required (see Figure 9 for iteration 3 of the much reduced human-centric cybersecurity assessment framework, now title the ‘Cybersecurity Awareness Framework (CAF)’’.

Figure 9

Iteration 3 of the Cybersecurity Awareness Framework (CAF)



The CAF suggests that interventions should target the six key themes sat beneath its model, should human vulnerability in cybersecurity wish to be reduced. For example, threat appraisal could potentially be increased by providing employees with regular updates on cyber-attacks experienced within an organisation and outside of it, to ensure they have a realistic understanding of the likely probability and severity of a successful attack. The main objective of Study 3 will be to widen the participant sample further, whilst looking to verify the regression model unearthed in Study 2.

Study 3

Recap of Study 3 Aims

Following on from the findings of the previous two studies, Study 3 aims to address the following:

- Further investigate the individual differences found to predict reported cybersecurity behaviour in Study 2, in order to substantiate these findings across a much larger and wider spread sample size;

- Generate a third and final iteration of the human vulnerability to cyber-attacks framework that can provide organisations with a tool with which to measure how their employees experience cybersecurity.

Research Hypotheses

It was expected that the findings of Study 3 would largely replicate the regression analysis findings of Study 2, with the main aim of confirming the Cybersecurity Awareness Framework (CAF) across a larger and more generalised sample. Therefore, the following hypotheses were anticipated:

S3 H1 The latent factor identified in Study 2 - ‘cybersecurity awareness’ - would also significantly predict reported cybersecurity behaviour within this study.

Method

Participants. A sample of three hundred and twenty-six participants were recruited utilising the online participant tool *Prolific*. All participants were required to hold a current part-time or full-time job to take part, in order to allow them to respond to questions in specific relation to the organisation they currently work for. Of these participants, 44% were male, 55% female, 0.5% of a different identity and 0.5% declining to comment with an average age of 34.72 (*SD* 11.16). The sample was well educated (rated from GCSEs to doctorate level) with 71% of participants in receipt of an undergraduate degree or a higher qualification.

Design. As with Studies 1 and 2, a Kruskal-Wallis test was employed to test differences in cybersecurity behaviour across a number of participant demographics (gender, age, education). Regression analyses (as in Study 2) were then undertaken to explore the potential use of a model to explain reported cybersecurity behaviour, utilising the ‘cybersecurity awareness’ factor reported as predictive within Study 2. An “enter” regression analysis was used in Study 3 as findings from Study 2 provided guidance for the best, and only way, to enter factors into the model. For this study, only the combined TPB and PMT questionnaire was utilised, to measure the cybersecurity awareness construct.

Materials and Procedure. As with both previous studies, participants accessed the questionnaire via *Qualtrics©*, where they were initially presented with an introduction sheet (Appendix A) and then a request for informed consent to take part (Appendix B). Each participant also completed an optional demographics form collating information on their age, gender and level of education. Next, participants rated themselves against the combined TPB and PMT questionnaire (Appendix H; Safa et al., 2015) to measure threat appraisal, self-efficacy, IS awareness, IS organisation policy, IS experience and involvement and IS attitude, as part of the Cybersecurity Awareness Framework (CAF). After completion participants were provided with a study debrief (Appendix J) and thanked for taking part.

Results

The key purpose of Study 3 was to confirm whether the Cybersecurity Awareness Factor (CAF), found to predict reported cybersecurity behaviour in Study 2, would also predict behaviour in Study 3, but across a wider sample. Should the same results be found, further validity would be provided around the use of this human-centric cybersecurity metric as a measurement of behaviour vulnerability within organisations.

Reliability of Measures. A test of internal consistency was applied to the human-centric cybersecurity framework identified within Study 2, with Cronbach's Alpha reaching excellent within the 'cybersecurity awareness' construct ($\alpha = .91$). The key assumptions for parametric testing were not met due to the use of ordinal data, and therefore non-parametric statistical tests were utilised. Assumptions for all statistical tests used were analysed and met. During the following analysis, any missing observations within the dataset were replaced with the grand mean for each question and any outliers determined by 3 IQR from the mean were windsorized to the next available value not considered extreme.

Regression Analyses. In order to try and validate the findings of Study 2, regression analyses were conducted to investigate whether the identified cybersecurity awareness factor, would again be found to be predictive of cybersecurity behaviour. Within Study 2, a stepwise approach was used, as there was no precedent available to determine the way in which each factor should be appropriately entered, they were therefore entered simultaneously. Analysis in Study 3 utilised the enter mode, as cybersecurity awareness (Mdn = 6, IQR = 1), was the only factor under investigation. The model established in Study 2 was verified within this study, finding significance (F (1, 324) = 489.287, $p < .001$) that explained 60% of the variance in reported cybersecurity behaviour ($R^2 = .600$; see Appendix L).

Discussion

The main aim of Study 3 was to further refine and validate the findings of Studies 1 and 2, by investigating factors both related to, and predictive of reported cyber-security behaviour but across a larger working sample. The main objective being, to confirm the replicability of results unearthed in the first two Studies to ensure that those individual differences highlighted as predictive of cybersecurity behaviour are those most likely to be useful in measuring vulnerability in the real-world. The planned output of Study 3 being the generation of a validated Cybersecurity Awareness Framework (CAF) and associated novel tool, that organisations can use to measure and manage human vulnerabilities in cybersecurity, making it easier for organisations to deploy interventions that are directly tailored to these vulnerabilities. By providing organisations with an insight into how employees are experiencing cybersecurity within their company, time and budget can be better allocated in the hope to more dynamically improve behaviour.

It was also anticipated that the cybersecurity awareness latent factor, identified via EFA within Study 2, would again significantly predict reported cybersecurity behaviour, as found within regression analyses within Study 2. This was confirmed, with cybersecurity awareness significantly predicting 60% of the behaviour reported by participants. This finding, in concordance with Study 2, presents a novel overarching framework that is measured via several observed variables ordinarily contained within behaviour change theory models (and beyond). These observed factors include threat appraisal, information security experience and involvement, information security self-efficacy, information security attitude, information security awareness and information security organisation policy. Findings from Study 3 provide more clarity in relation to the term ‘cybersecurity awareness’ and how organisations can utilise this measure, and associated framework, to provide a more transparent view on how their employees are experiencing cybersecurity and therefore what can be done to potentially improve this experience.

A theoretical paper by Jeong et al., 2019 investigated the human factors related to cybersecurity behaviour, summarised a total of twenty-seven papers that had identified factors, models or frameworks of particular importance in this domain. Of these papers, two were noted as generating a cybersecurity awareness framework. The first, Metalidou et al. (2014) drew upon the findings of a number of previous studies to identify factors that may be of importance in relation to cybersecurity, considering aspects such as motivation, beliefs and inadequate use of technology. McCormac et al. (2017), rather than measuring cybersecurity awareness specifically, explored personality traits and risk propensity in relation to cybersecurity knowledge, attitude and behaviour. However, in describing cybersecurity awareness both papers list a number of important factors such as knowledge around policy, attitudes towards cybersecurity, motivations for behaviour, understanding around its importance and levels required for their particular organisation. Whilst the CAF considers similar constructs such as policy, motivation in relation to threat appraisal and attitude. It goes further, by also considering additional factors of importance such as employee security self-efficacy and experience.

Whilst a number of additional papers have also explored the use of a cybersecurity awareness framework (Khader et al., 2021; Wang et al., 2018), many focus on the generation of a process for deployment of a cybersecurity awareness tool, rather than a predictive model. Hijji and Alam (2022) generated a Cybersecurity Awareness and Training framework (CAT) that whilst considered cybersecurity awareness, it focused on raising awareness via a specific training schedule across a number of different cybersecurity topics such as cybersecurity basics and social engineering. As well as Bada et al. (2019) whose framework assesses the capabilities and maturity of a cybersecurity awareness programme. With both of these frameworks referring to cybersecurity awareness as a form of training intervention rather than an employee state of mind. The CAF is novel, in that that its aim is to measure the perceptions of employees in relation to their experience in cybersecurity and how this may influence awareness of cybersecurity. Bringing together aspects of behaviour change theory that help indicate how to help move employees towards a more enlightened level of awareness and therefore more secure behaviours.

To summarise, the main aim of Study 3 was to continue exploration into the individual differences and psychological factors that may be predictive of reported cybersecurity behaviour, in order to create a set of metrics that can be used by organisations to reduce risk and improve cybersecurity posture by understanding the experience of the human. Study 3 confirmed the regression findings from Study 2, reporting cybersecurity awareness to be a latent factor significantly influencing how employees choose to act. A construct that encapsulates - how likely employees perceive threat and their ability to protect themselves and their organisation from this threat, as well as their attitude towards cybersecurity it based on previous experience and involvement, knowledge of how to remain up to date and perceptions in relation to policy usability (see Figure 8). The finding of a principal cybersecurity awareness factor, that can help explain such a large portion of reported behaviour, will be invaluable for organisations moving forward. The cybersecurity awareness framework uncovered, and its related instrument, can be used by organisations to better understand how employees are experiencing cybersecurity, its associated vulnerabilities, and where intervention should therefore be focused moving forward.

Chapter Conclusion

The principal aim of Studies 1 to 3 within this chapter, was to better understand employee experience in relation to cybersecurity, in order to determine the best way to measure and manage human susceptibility to cyber-attacks moving forward, with the human at the centre. First, it was imperative to review human-centric cybersecurity and potentially predictive factors, across the breadth of psychological, sociological and behavioural economic literature available. Providing guidance around what key constructs may be driving behaviour and how best to intervene. By providing organisations with a framework for measuring human vulnerability to cyber-attacks, not only can interventions become more targeted but also the ability to measure the success of related interventions pre and post deployment.

Study 1 initiated the route towards a more consolidated human-centric cybersecurity framework, by exploring how several previously reported end-user demographics and individual differences, significantly relate to reported cybersecurity behaviour. To provide an initial milieu for the human experience in cybersecurity. The factors included had previously been identified as either correlating with/predictive of cybersecurity behaviour within previous literature. It was however, the first time such a breadth of constructs had been investigated collectively, in one study.

Study 2 was deigned to verify the factors identified as of interest within Study 1, but across a larger work-based sample (from the same organisation), extending further upon the original findings by refining the large number of interrelated variables into an abridged framework, more reflective of the employee experience. Regression analyses were then conducted utilising the refined EFA model, to uncover an inclusive factor – *cybersecurity awareness'* that accounted for 55% of the variance in reported behaviour. Study 3 went on to further validate the regression model, again finding the cybersecurity awareness latent variable to be predictive of behaviour, but this time accounting for 60% of the variance.

Cybersecurity awareness, the key factor found to associate with reported behaviour, indicates that, in order to improve cybersecurity behaviours within an organisation, effort must be given to ensuring employees continue to experience positive interactions with cybersecurity and feel an integral part of its process, as well as hold a positive attitude towards it. Positive experiences can be achieved through continual exposure, allowing behaviours to become ingrained within system 1. A knowledge sharing culture can then support the transfer of this knowledge across the business as well as ensuring this knowledge remains up to date. Attitudes are related to the beliefs, values and motivations of an individual, impacting the way they feel about something (Pickens, 2005). The measurement items the from six original key factors found to load onto this primary unobserved variable help generate a Cybersecurity Awareness Framework (CAF).

It is often assumed that improving cybersecurity awareness relies solely on employees becoming more informed about future risks, and what needs to be done to protect their organisation from these risks. This is logical, as awareness centres around knowledge, and this knowledge needs to be current in order for awareness to be beneficial. However, the wider literature available in relation to this topic makes clear, that the concept 'awareness' involves much more than human perceptions about the future, but also experiences from the past and current views around i.e., policy. First in reference to how cybersecurity experiences in the past can alter awareness by impacting perceptions around levels of experience and involvement and attitude towards it. Interventions such as gamification (the use of game mechanics to improve learning engagement) or phishing email reporting feedback, can help provide employees with positive past accomplishments that both improve these perceptions and encourage the growth of tacit knowledge (this point is discussed more, further on in this thesis). Present experiences such as whether an employee's implicit knowledge is current, as well as perceptions around the current usability of extrinsic knowledge held in policy and how it can impact. Finally, perceptions about future experiences and their influence on awareness, such as how threat is appraised and efficacy in relation to the ability to protect from this threat. By considering how employees holistically experience cybersecurity, organisations can better drive positive awareness above and beyond current awareness training.

These six factors, and their underlying heuristics can help provide guidance around where employee cybersecurity awareness may require support, with the overarching framework and associated measure allowing organisations to benchmark and then re-assess this status, post intervention. With such interventions perhaps focused on providing employees with regular threat updates, in order to increase their perception of future threat probability. By considering how employees holistically experience cybersecurity, organisations can better drive positive awareness above and beyond current awareness training. The experiences that determine cybersecurity awareness can exist both implicitly and explicitly, with knowledge held both outwardly in policy, as well as held directly within the mind (Marton, 2000). The experiences that are held within the mind, are also not continuously activated in working memory, but instead exist largely outside of conscious thought, potentially requiring a trigger either internally or externally such as a soft-paternalistic nudge (more to be discussed on this topic within Chapter 3 of this thesis). By understanding how humans experience all aspects of cybersecurity awareness, it becomes possible to address a large number of vulnerabilities that without attention may result in a successful malicious attack.

How employees experience cybersecurity, will in turn influence how they choose to behave. Should an employee’s level of cybersecurity awareness be less than satisfactory, perhaps due to a lack of communication around recent cyber-attacks, they will not conduct the actions that are required to protect themselves and their organisation from such risk. By measuring cybersecurity awareness, organisations can understand how their employees are currently experiencing cybersecurity, and what experiences feeding into awareness, need to be targeted by intervention in order to improve behaviour and reduce the likelihood of a successful breach. By investigating how an extensive range of individual factors influence cybersecurity behaviour, a reduction in the number of constructs of concern has been achieved, allowing organisations to better tailor and target human risk behaviours moving forward utilising the Cybersecurity Awareness Framework (CAF).

Forward to Empirical Block 2

Whilst Chapter 2 provided a novel exploration into the employee experience in cybersecurity, and how cybersecurity awareness may actively influence behaviour, it did not provide a completely holistic appreciation of the full encounter. Whilst cybersecurity awareness helps explain the employee internal experience, it does not detail how that behaviour may change when both positive and negative external prompts are being applied.

From an adverse perspective, cybercriminals are often aware of and can take full advantage of how humans experience cybersecurity, and the less conscious ways in which cybersecurity awareness exists. By utilising social engineering strategies, hackers are able to manipulate the human conscious experience negatively, and for their benefit. In direct opposition, academics and awareness organisations alike are working hard to identify ways in which to positively influence the less conscious aspects of cybersecurity awareness, with the aim of steering awareness into conscious thought as and when necessary.

The following chapter will therefore take forward the workings of Chapter 2, by continuing to investigate the human experience in cybersecurity, but how this presents itself when inevitably influenced by external factors, either positively or negatively. Chapter 3 will therefore provide a more holistic overview of the human-centric cybersecurity experience by exploring first the key social engineering techniques being employed within synthetic media and their impact on the human, as well as exploration into a number of ways in which researchers, industry vendors and organisations are attempting to positively influence the unconscious human cybersecurity experience.

Chapter Three: Human Cybersecurity Vulnerability Exploitation and Mitigation

Chapter Summary

Humans are believed to be ultimately responsible for the success or failure of most cyber-breaches, due to either intentional or unintentional human action (Gartner, 2023; Verizon, 2022). It is for this reason that more needs to be understood in relation to the employee experience in cybersecurity, and what may be influencing these alarming statistics. Whilst a large portion of human vulnerability can be explained by the internal experiences of the employee, it is important to consider external influences that may be further impacting how people ultimately behave. It is only through a holistic understanding of both internal and external experiences, that organisations can build a fair picture around the challenges employees face, in order to determine what – if anything – can be done to improve the situation. Chapter 3 investigates such external influencers, and how decision-making vulnerabilities are being influenced for both the bad and greater good. First in relation to cybercriminal persuasion techniques, and then the debiasing techniques being used mainly in academic and industrial research, to also convert behaviour, with the aim of tighter security processes and practices.

Cybercriminals influence employee (and people, citizens in general) behaviour via social engineering strategies, largely within phishing emails, that encourage automatic, intuitive thinking that increases opportunities for error e.g., employees clicking on a malicious link, and/or sharing confidential information such as passwords. Despite developments in technology that hope to identify malevolent emails, phishing has remained one of the top threat actions responsible for security breaches for many years (Techtarget, 2023; Verizon 2020, 2021, 2022). The increasingly sophisticated ways offenders are finding to bypass such interventions ultimately result in end-user responsibility for detecting email illegitimacy. Once a phishing email is opened, persuasion techniques within its content mercilessly exploit human decision-making limitations by lacing phishing emails with tactics such as authority, scarcity, reciprocity, commitment and consistency, social proof, liking and similarity (e.g., Akbar, 2014; Ferreira et al. 2015). The two studies within this chapter, that investigate these techniques were conducted to better understand the most utilised methods of persuasion, and also how employees are reacting to the key cybercriminal weapons of influence being used. The aims of Experiment 4 and 5 were to therefore understand the current (circa. 2020 when the data was collected) phishing landscape, in so far as the methods of persuasion employees find themselves experiencing and are more susceptible to. With Study 4 analysing 998 phishing emails employees have identified and reported as phishing and Study 5 utilising a simulated mailbox to investigate which methods are also most likely to be classified as suspicious – further informing the taxonomy of decision-making biases of most concern. Findings from these studies, in conjunction with the outcomes from Chapter 2, will help inform recommendations around how organisations can better target intervention in relation to employee decision-making and the need to continually analyse and track social engineering trends.

Whilst previously organisations would employ mainly technical solutions to limit the number of phishing emails landing in end-user inboxes, e.g., email filtering, this hadn’t resulted in significant reductions year on year (Verizon 2022). In more recent years a heavier focus has been given to providing employees with cybersecurity education, in an attempt to influence behaviour change, however it has not been enough to result in reductions in cyber-breaches with other strategies now required to move closer towards mitigation (Aldawood et al., 2019; Alshaikh et al., 2018; April 2018; Bada et al., 2019; Scholl et al., 2018; Skinner at al., 2018). Some cybersecurity vendors are offering alternative styles of intervention in order to try and manipulate human bias, but for the greater good. The remaining three experiments within this chapter therefore assesses the efficacy of three key interventions that can be potentially utilised within organisations to better support human decision-making, in particular relation to phishing.

One-thousand three-hundred and ten participants were requested to file a selection of genuine and phishing emails into several inbox folders with the aim of filing phishing emails into the folder named ‘suspicious’. In each of the three experimental studies, a different debiasing intervention strategy was used with hypotheses for each - including real-time soft-paternalistic nudging, motivation through education, and adapting cognitive strategies found previously to be of interest (Croskerry et al., 2013; Thaler & Sunstein, 2008). Findings from the three experiments suggest several potentially quick, relatively cheap but effective interventions that can be used by organisations to support the identification of phishing emails, particularly the use of a short mental checklist or maxim that can become habitual during an employee’s working day. Habits can then be recalled and executed by end-users when in a more unconscious mode of thought providing a stronger defence against offenders hoping their malicious communication will remain undetected. This research offers investigations into how several interventions outside of awareness training might look to support human decision-making when opening emails, in order to mitigate cyber-breaches as a result of phishing and improve the employee cybersecurity experience.

Debiasing, a term used to describe processes by which humans are supported to reduce the negative impact of violations from rational thought, is a form of intervention with possible potential. Fischhoff (1982) categorised two forms of debiasing: modification of the decision-maker, and modification of the environment. The first assumes that bias resides in the person, and therefore tools and training are needed to reduce decision-making errors within them. The latter suggests bias resides in the environment with alteration of context being the key to bias reduction.

Previous research outside of the cybersecurity domain suggests that careful modification of the decision-maker through bias training can have some success (Morewedge et al. 2015), with research both within and outside of the cybersecurity domain suggesting success in modifying the environment via soft-paternalistic nudging, explored later within this chapter (Brigg et al., 2017; Jeske et al., 2014; Petelka et al., 2019; Turland et al., 2015). Chapter 3 therefore also investigates whether a number of debiasing techniques looking to either modify the decision-maker, for example through education around new mental strategies, or by modifying the environment e.g., via soft-paternalistic nudging are useful in reducing human cybersecurity vulnerabilities or whether they inflict more cognitive burden in relation to the employee experience. Results from intervention studies (Studies 6 – 8) will help provide a more holistic overview around the human in cybersecurity, as well as inform on recommendations for future intervention.

Introduction

Cognitive psychologists have studied human decision-making, in particular decision-making theory, for over 70 years (Simon, 1947) with investigations leading to the belief that the human mind is subject to two functions; an automatic system (“system one”) and a competing reflective system (“system two”). System one utilises an unconscious process, driven by pre-determined rules of thumb, whilst system two consolidates information more slowly, and is consciously controlled and deductive. These two decision processing systems are not thought to be dichotomous, but instead move across a continuum where both are present but in differing amounts at any given moment (Croskerry et al., 2013). Due to the continuous and simultaneous choices humans need to consider, in a mind of limited capacity, the unconscious and automatic system is thought to lead around 95% of the time (Bargh et al., 2001; Simon, 1990).

Four determinants are believed to influence the automatic biases and patterns demonstrated in system one decision-making (Stanovich, 2011). First, evolution has resulted in biases that are hardwired and learned for adaptation purposes, second features emotional influences such as fear and disgust, third are drivers of regular exposure to situations resulting in social habits, and finally, patterns of choice obtained subconsciously rather than explicitly taught. More recently psychologists have focussed on two overarching sources of bias within system one, those gleaned through natural selection, and those gained developmentally through a human's environment (Croskerry et al., 2013).

Despite the usefulness of rapid and unconscious thought, particularly in life saving conditions, system one processing can often lead to bias resulting in a decision outcome that would not have been chosen should rationality have been applied. It is this system one process of decision-making that often operates when humans are undertaking cybersecurity tasks, particularly within organisations where multi-tasking and work-based interruptions are inevitable. During everyday decision-making, system one errors may have insignificant negative impact on the choices humans make, however there are some industries where unconscious biases can have particularly devastating effects for decisions made within the health domain and cybersecurity. In the field of cybersecurity, irrational decision-making and biases cause errors that can result in financial loss or impact critical national infrastructure. When humans undertake cybersecurity tasks, they are often vulnerable to problematic automatic decision-making, particularly when interacting with emails a task that has become habitual.

Additional challenges within the cybersecurity domain suggest offenders are often aware of human decision-making constraints and exploit these through social engineering methods, eliciting biases for their benefit. For example, phishing emails are often laced with several persuasion techniques actively driving recipients into intuitive decision-making leveraging the errors such rapid processing may educe. By promoting quick, heuristic thinking offenders increase the chances of end-users missing clues within their communication that with time would identify their motive as malicious.

Human Vulnerability Exploitation

In a world of unremittent information, humans have become skilled at applying conscious mental resource to just a small number of decisions (Blanco, 2017). The remainder, rather than consciously decoded, are resolved through pre-determined inferences that result in quick and less mentally taxing conclusions. Whilst the outcome of these decisions are often suboptimal, their simple approximation will suffice in many day-to-day situations (Petty & Cacioppo, 1986). Despite their benefits, heuristic decision-making can however result in error, incorrectly shaping the way people make and come to judgements (Taylor-Gooby & Zinn, 2006).

Many cybercriminals are aware of errors that can occur during heuristic thinking, hiding behind pretexts to exploit them. Offenders are becoming accomplished at driving humans into intuitive decision-making utilising a number of known weapons of influence, such as the authority principle and commitment and consistency (Akbar, 2014; Ferreira et al., 2015; Zielinska et al., 2016). Expecting employees to behave securely requires not just deviation from learned human heuristics *but also* the capability to acknowledge and rebel against cybercriminals using such techniques as weapons of influence. Educating employees not only on the vulnerabilities inherently experienced, but also the ‘influence artillery’ cybercriminals utilise to further exploit them. Chapter 3 will explore the subject of these social engineering tactics, and in particular the key ways offenders hope to influence human decision-making.

Despite developments in technology that hope to identify malevolent emails, phishing has remained one of the top threat actions responsible for security breaches for many years

(Techtarget, 2023; Verizon 2020, 2021, 2022). The increasingly sophisticated ways offenders are finding to bypass such interventions ultimately result in end-user responsibility for detecting email illegitimacy. Once a phishing email is opened, persuasion techniques within its content mercilessly exploit human decision-making limitations with tactics such as authority, scarcity, reciprocity, commitment and consistency, social proof, liking and similarity (e.g., Akbar, 2014; Ferreira et al. 2015). The study and experiments in the current chapter were conducted to better understand the most utilised methods of persuasion but also how employees are reacting to key cybercriminal weapons of influence. The aims Studies 4 and 5 were to therefore understand the current (circa. 2020 when the research took place) phishing landscape, in so far as the methods of persuasion employees find themselves more susceptible to. Study 4 focused on the analysis of 998 emails identified by employees as potentially malicious and Study 5 utilised a simulated mailbox to investigate which methods are also most likely to be classified as suspicious – further informing the taxonomy of decision-making biases of most concern discussed within the previous chapter. It is important to remember that the motive of these studies was to improve understanding around the employee experience of phishing emails and how they might respond. Findings from these studies, in conjunction with the outcomes from Chapter 2, help inform recommendations around how organisations can better target intervention in relation to employee decision-making and the need to continually analyse and track social engineering trends.

Since the arrival of mainstream internet services, revolutionary developments have resulted in ‘internet for the masses’, allowing cybercriminals to manipulate online services for personal gain. Such manipulation began early on within the Warez group who generated random credit card numbers to allow the creation of bogus America Online (AoL) accounts (Phish Protection, 2021).

The subsequent entrance and growth of eCommerce only exacerbated the situation with the first documented (but unsuccessful) eCommerce attack taking place in June 2001, targeting the E-gold online payment system. Later that year a successful phishing attack occurred using the subject of a 9/11 ID check around the attacks on the World Trade Center (Can I Phish, 2021). Cybercriminals took their next step during 2003 by registering numerous website domains with a similar appearance to companies such as eBay, Yahoo and PayPal, that included links within those emails that would drive people to spoofed sites where they were asked to enter their personal information online.

Since this time the use of emails to source user information, download malware and achieve financial gain has not abated. In 2021 phishing remained one of the top threat actions utilised by cybercriminals with it believed that around 83% of organisations experienced a phishing attack (Cybertalk, 2022; Verizon, 2022). Over time phishing has become more targeted (spear phishing – defined as a phishing email that pursues groups of individuals or companies using their background to create a more tailored message; Jari, 2022), as well as focused more on those possessing highly sensitive information (whaling – targeting c-suite executives or high-value individuals; Jari, 2022) whilst continuing to leverage on global devastation (e.g., the Covid-19 pandemic; Al-Qahtani & Cresci, 2022).

Despite continual developments in the technology required to help identify malicious emails, the threat from phishing to the integrity, privacy and accessibility to businesses around the globe has not subsided. In the UK alone, the NCSC have received as many as 13.7 million suspected phishing emails reported from external organisations April 2020 to August 2022, each providing potential attack entry to offenders using malware such as computer viruses, worms or spyware (NCSC, 2022). Actioned through a user's email by downloading an attachment, direct installation or by clicking on or even hovering over a link.

Whilst most (especially large) organisations utilise filters to minimise the exposure of end users to phishing emails, attackers are finding increasingly sophisticated ways of bypassing such technology ultimately leaving end users responsible for detecting phishing email illegitimacy. Cybercriminals lace emails with subtle manipulation strategies, convincing recipients to respond in a desired manner with strategies including persuasion techniques whereby offenders use the content of an email to exploit human decision-making biases, unconsciously influencing the way end-users respond (Akhbar, 2014; Cialdini, 1984; Lawson, 2018). An example is the authority principle that relies on the human heuristic to trust the opinions of experts, with a phishing email from a person of authority providing confidence in its contents. For this reason, human user-lead cybersecurity has become a crucial and top priority across companies around the world (e.g. Morgan et al., 2020; Pfleeger et al., 2014; Sasse et al., 2001; Zimmermann & Reanuad, 2021).

To support the risk faced by employees, many organisations now provide some form of education on cybersecurity threats that include phishing, such as online training or simulated attacks and reporting functions (e.g. report buttons within email platforms/packages). However, awareness programmes do not always result in positive behaviour change, particularly long-term (Bada et al., 2019; Scholl et al., 2018). Cybersecurity training is challenging for a number of reasons such as diversity of attack types (e.g., phishing, device securement, ransomware, password attacks) with training in one area not easily transferable to another (see e.g., Bada et al., 2019; Moschovitis, 2019). Individual differences within end-users present a further challenge for training programmes that remain a one-size-fits-all approach (Hadlington, 2017; Proctor & Chen, 2015).

In addition to these challenges, Kahneman (2011) speaks about human systematic decision-making errors that cybercriminals are aware of and further leverage for personal gain. Numerous psychologists have conceptualised the human mind as having two competing decision-making functions: the *automatic system* whereby behaviour is intuitive, driven by rules-of-thumb and heuristics, and the *reflective system* that processes information more slowly and is controlled and deductive (Caraban et al., 2019; Croskerry et al., 2013; Kahnemann, 2011; Welford, 1965). Whilst in many instances the automatic system can be useful, it may also result in bias leading to outcomes that would not have been chosen should rational thought have been applied. Around 95% of decision-making is believed to adopt the intuitive route, driven by factors such as perceiving the decision to be of little consequence, time-pressures or the need to dual task (see Bargh et al. 2001). However, many aspects of SETA rely on employees utilising a rational decision-making style when tackled with a cybersecurity decision, limiting their success. It is therefore important to investigate alternative ways of educating and influencing end users on how to identify malicious emails, particularly during intuitive decision-making. This begins by determining the underlying decision-making biases cybercriminals are looking to exploit.

Within phishing emails, offenders tend to use methods of persuasion to actively drive recipients into intuitive decision-making taking full advantage of its biases whilst in this state (Luo et al. 2013; Williams et al. 2018). Previous research by Cialdini (1984) identified six main methods of persuasion used to achieve this: *reciprocation*, *commitment and consistency*, *social proof*, *liking and similarity*, *authority*, and *scarcity*. Table 5 contains descriptions of the main biases exploited by each of these techniques, for example presenting the 'email sender' ('from') as an expert, in an attempt to activate the authority bias whereby humans attribute greater accuracy to the opinion of powerful figures allowing for a quick and intuitive decision. This makes it more challenging for employees to identify clues of potential malevolence within an email that, with time, may have been uncovered. The presence of these persuasion techniques within phishing emails has since been confirmed within literature albeit papers that are already becoming dated (Ferreira et al., 2015; Akbar, 2014; Zielinska et al., 2016).

Table 5 *The Top Methods of Persuasion Cybercriminals Exploit (Akbar, 2014; Butavicius et al., 2016; Cialdini, 1984; Ferreira et al., 2015; Lawson et al., 2020)*

Persuasion Technique	**Cognitive Bias Being Exploited**
Authority	More trusting of opinions of experts / those in power
Reciprocation	Feeling a need to return a favour or repay debt
Commitment & Consistency	Wanting to appear consistent with previous decisions
Social Proof	Following the lead of others that we are associated with
Liking & Similarity	More trusting of those we feel an affinity or connection
Scarcity	Wanting something we cannot have or that is exclusive

The authority principle within phishing emails may pose senders as working for genuine companies, pretending to be e.g., the CEO of the organisation and/or include (often numerous) accreditations and credentials. The reciprocation principle could be triggered with the offer of a ‘free gift’ or ‘discount’ alongside a suggestion that the recipient sign up to information or an event. Commitment and consistency involves the suggestion that a user (recipient) had previously conducted a behaviour relating to the email, for example donating to a good cause, in the hope that they will feel inclined to respond with interest again. An example of social proof is a statement such as ‘80% of your colleagues have already completed the survey – please follow the link below’, thus providing a social basis for fast-tracking decision-making. The principle of liking attempts to build rapport, offer praise or suggest a common interest. Lastly, scarcity within phishing emails is presented through terms such as ‘for a limited time only’ or ‘exclusive deal’ to encourage a quick response. Each of these six weapons of influence are laced into many phishing emails – sometimes in combination – to promote heuristic decision-making and detract from its ingenuity. Further details around these techniques are discussed later in this chapter.

Previous research across several countries and organisations found authority, liking/similarity, and scarcity to be among the most frequently used methods within phishing emails (Akbar 2014; Ferreira & Lenzini, 2015; Williams et al. 2018). The methods used have changed in usage over time: for example commitment/consistency and scarcity were reported as increasing in number within company targeted phishing emails between 2010 and 2015 (Zielinska, Welk & Mayhorn, 2016). In order to design interventions that will better protect people against the weapons of influence involved, there is a need to improve understanding around the *current* methods of persuasion targeting victims, how they are presented, and how to reduce their influence. Recognising patterns in persuasion techniques within phishing emails and how humans interact with them will better support the blacklisting of such emails using technical solutions as well as help to improve human-user training and knowledge.

Several previous studies have investigated the success of persuasion techniques in phishing emails, such as Rajivan and Gonzalez (2018), who found the presence of a notification, an authoritative tone (authority) and an expression of shared interest (liking and similarity) the most effective methods, with the request to change a password (potentially scarcity) and the offer of a deal (reciprocity) the least effective. Butavicius et al. (2016) also investigated the success of methods of persuasion within phishing emails, with a focus on social proof, scarcity and authority. The most effective method of persuasion was again found to be authority, followed by scarcity and the least effective – social proof. A study by Williams et al. (2018) involving emails sent to 62,000 employees over a 6-week period reported authority as the most effective method although urgency (e.g. time constraints on a suggested action such as replying), when coupled with authority, resulted in employees being particularly susceptible. A study by Parsons et al. (2019) found participants to be less susceptible to the scarcity principle and most susceptible to commitment and consistency, and reciprocity. These similarities and sometimes discrepancies in findings could potentially be due to temporal changes, with cybercriminals adapting their techniques as end-users become more aware of the methods being used over time.

The studies and experiments presented within the current chapter of this thesis reinvestigate all six methods of persuasion – authority, reciprocation, commitment and consistency, social proof, liking and similarity, and scarcity alongside an additional method uncovered during analysis within Study 5 (curiosity) plus a control condition (no method of persuasion; Akbar, 2014; Ferreira et al., 2015; Zielinska et al., 2016). The approach in Study 4 is to analyse real emails reported in a multinational corporation as suspicious by employees and then, where possible, tailor these emails for use in Study 5 (e.g., ensuring only one persuasion technique is present). Whilst reported emails are not ideal, they represent the data set for most historical phishing studies, with the novelty of this study being a collection of reported emails from a single organisation.

Findings from the current study will improve understanding around the prevalence of techniques during 2020 most likely to be encountered by employees (possibly generalisable to multiple other organisations), as well as those most likely to succeed. Noting that, according to the knowledge of the author of this thesis, the most recent study investigating this was published based on emails from 2010-2015 (Zielinska et al., 2016).

In addition to investigating the methods of persuasion most prevalent in phishing emails (reported in a large organisation) as well as those most likely to be categorised as suspicious, the current research involves analysis of threat actions endorsed within phishing emails e.g., clicking on a link, or opening an attachment, whilst also highlighting other techniques being used over the past few years. Previous research suggests links to be the threat action most likely to be suggested (~70%: Zielinska et al. 2016 and Akbar, 2014) with ancillary clues such as SPAG errors also prevalent (~81%; Zielinska, 2016).

Human Vulnerability Mitigation

Decision-making errors are believed to be systematic and non-discriminating, with possibility for effective wide-spread mitigation by interrupting or changing automatic patterns (Croskerry et al., 2013). The intention of research within the cybersecurity decision-making field should aim to discover ways to moderate these biases and drive humans into more rational thought. This may not be a simple task, with over one hundred biases reported and many of them impacting cybersecurity. Jolls and Sunstein (2006) attempt to categorise these biases into two types: judgement errors impacting everyday decisions such as the optimism bias, and departures from utility theory such as the endowment effect. Utility theory is the assumption that humans rank choice linearly based on their expected usefulness or monetary value, however deviations may occur from linearity due to additional factors such as current financial position and aversion to loss (Barberis, 2013). This suggests that several interventions may be required to mitigate the breadth of biases that are potentially impacting cybersecurity in order to ‘debias’ and reach more normative decisions. Although at times, humans can successfully create their own strategy to work against bias, for example as described by Croskerry et al. (2013) people choose to store their car keys in the same habitual place each day in order to not forget their location.

In a review of the relevant debiasing literature, Soll et al. (2014) segment such strategies into two main categories; interventions that modify the person (increasing motivation and providing alternative strategies) and interventions that modify the environment (soft-paternalistic nudging) with the latter highlighted as being more effective when people are cognitively constrained i.e., do not have the cognitive resources to modify automatic decision making themselves. As cybersecurity is often a secondary task, cognitive resources and motivations are focused elsewhere, with it possible that providing end-users with environmental tools to support their decision-making may be the most effective resolve for mitigation of security breaches. Kaufman et al. (2009) also described such debiasing interventions, categorising them as either focused on (a) creating awareness around decision-making, (b) decomposition of the decision process or (c) viewing the decisions from a different perspective. An alternative categorisation of debiasing techniques was provided by Croskerry et al. (2013) segmenting the strategies into three groups; educational strategies, workplace strategies and forcing functions. Educational strategies aim to improve knowledge around biases and aid in their detection, workplace strategies are aimed at supporting in the here and now such as offering physical checklists and forcing functions include interventions such as age-old adages or nudges that can further support strategy change. These three debiasing styles can also take a similar approached to that within Baldwin (2014) in relation to nudging, whereby a three tiers system describes the amount of awareness and/or autonomy the debiasing strategy takes away from the individual it is trying to persuade. With a Tier one debiasing strategy respecting autonomy and looks to simply improve rationality. Tier two respects autonomy less as it is looking to influence a more automatic response and tier three respects autonomy even less as it attempts to completely reframe cognition. Taken together, this suggests that perhaps three key intervention types should be reviewed:

(a) Tier 1: Nudging - A soft-paternalistic nudge that seeks to modify the environment, creating awareness in real-time;

(b) Tier 2: Motivation – An educational communication that attempts to modify the person through motivation;

(c) Tier 3: Conversion - A strategy adaptation intervention that attempts to again, modify the person, but by attempting to convert them into using alternative cognitive strategy such as decomposition of the decision-making process or viewing the decision from a different perspective.

Research to date has identified these debiasing techniques as effective in domains such as healthcare, forensic mental health and education, however research within the cybersecurity field is currently lacking (Ludolph & Schulz, 2018; Scopelliti & Morewedge, 2019; Sellier & Griffith, 2019). It is therefore important to investigate the use of the three debiasing techniques within cybersecurity, and whether such external influence can be used to positively influence cybersecurity awareness and in turn behaviour.

Humans are thought to process the majority of decisions unconsciously in order to function in a world where choice is boundless and cognitive capacity limited (Kahneman, 2011). Whilst this quickfire process of selection is practical, its underlying schemas often result in outcomes that are less than optimal. A number of methods have been suggested as useful in reducing these decision-making errors, which are of particular interest in the field of cybersecurity where unconscious oversight can result in significant financial loss. This literature review will focus on the three areas of intervention mentioned previously, soft paternalistic nudging increasing motivation and debiasing by providing alternative decision-making routes. The following research studies look to investigate each of these techniques within a cybersecurity setting in order to determine their potential effectiveness as a safety intervention.

Real-time Nudging. Soft-paternalistic nudging uses choice architecture to guide humans towards improved decisions without restricting options (Kaufman et al., 2009). Choice architecture helps present choice in a way that leads people towards the preferred option, for example positioning healthier food options at eye-level (Thaler & Sunstein, 2008). Examples of nudges that utilise choice architecture include auto-enrolment in a pension plan whereby humans are more likely to stick with the default option, and a calorie app that may trigger humans to reduce their sugar intake (Sunstein, 2014). Nudges are believed to positively utilise decision-making heuristics, by igniting those that will be most useful to arrive at the best decision within the current context (Zimmermann & Renaud, 2021). Within the cybersecurity domain, a warning nudge could be used to deter end-users from clicking on a link contained in an email by using a pop-up informing them that it may not be safe. Alike cybersecurity education, nudges are not a one-size-fits all, with the need to carefully match the nudge to the desired outcome (Zimmermann & Renaud, 2021), for example ensuring the content of the nudge connects well with the decision-making biases to be influenced. A number of studies have investigated the success of nudging as a debiasing technique within cybersecurity with both nudge content and context proving influential in improving behaviours around password generation, malicious link identification and Wi-Fi connection (Furnell et al., 2019; Petelka et al., 2019; Turland et al., 2015).

Despite the potential nudging has in guiding improved decision-making, there are several prospective barriers to its success in cybersecurity. Previous literature within behavioural theory acknowledges that continual contact with an artefact can result in habituation, whereby less attention is bestowed upon an item with increased exposure (Thompson, 2009).

Therefore, should a nudge be continually presented in the same format, the message attempting to be delivered may become ignored. There is also the potential that reductions in productivity could occur should nudging attempt to move end-users continually into more conscious thought, with it is important to consider the optimal amount of system one and system two processing required to reduce risk without significantly impacting productivity in relation to primary tasks (Croskerry et al., 2013). Despite these challenges, nudges can provide a quick, cheap and effective way of mitigating decision-making errors driven by system one biases, something of particular interest for organisations where a large employee network exists (Sunstein, 2014). Therefore, the aim of Study 6 will be to investigate whether the use of soft-paternalistic nudging can assist in the identification of phishing emails, by directly targeting human decision-making biases.

Motivation. Motivation can provide a potential form of intervention that can be used to help mitigate human susceptibility to cyber-attacks, influenced in a number of ways - for example providing incentives such as rewards, improving awareness of costs associated with inaction or instigating accountability. End-users are motivated to conduct conscious cybersecurity behaviours when they rate the benefits of cyber-safe behaviours and outcomes higher than its associated costs (Larrick, 2004). Motivated end-users will therefore perceive the time and effort it takes to remain cyber-safe worth avoiding the consequences of a security breach.

The elaboration likelihood model attempts to explain what drives humans towards either system one or system two thinking, with motivation one of its key factors. Elaboration likelihood relates to the varying degrees of thought that may be applied to the processing of a decision, with high levels resulting in more conscious thought and therefore the allocation of more cognitive resources (Petty & Cacioppo, 1986). Should humans experience a lack of motivation, elaboration likelihood will be low resulting in the avoidance of effortful thinking.

Whilst motivation may be useful in improving the outcome of decision-making, the method assumes people already possess effective conscious decision-making strategies and simply require incentives to put them into action. For those that do not possess the strategies required to make normative decisions, increasing motivation will merely encourage a different form of deficient decision-making (Larrick, 2004). Motivation can however drive people towards learning the skills or strategies required to make decision-making more effective as well as assist in new habit formation (Larrick, 2004; Soll et al., 2014).

There are thought to be two main sources of motivation to act, external factors such as rewards that are more beneficial to outcome focused activities, and internal factors such as interest in a task that are more relevant to more processed focused actions (Touré-Tillery & Fishbach, 2014). Motivation is thought to be optimal when driven by intrinsic or self-oriented motivation, most prominently resulting from internal factors such as pleasure. If pleasure in the activity is not possible, external factors such as seeing the value of the task in hand provides extrinsic motivation that is more self-determined and therefore more effective (Ryan & Deci, 2000). Cybersecurity is a set of ongoing activities that, on their own, usually derive little pleasure for humans making intrinsic motivation difficult to achieve. Interventions will need to focus on educating end-users on the value of cybersecurity tasks in order to influence a self-interested extrinsic motivation and drive more positive behaviours. Self Determination Theory (SDT) is a psychological framework that focuses on those activities that are thought to be useful but not particularly pleasurable, suggesting autonomy (ownership of one's own actions), competence (feelings of mastery) and relatedness (feelings of belonging) to be fundamental in achieving extrinsic motivation that is more self-driven (Ryan & Deci, 2020). For end-users to become motivated to conduct conscious cybersecurity behaviours this theory suggests they must believe themselves to have the ability to perform the tasks, feel a sense of choice in undertaking them and that the behaviours bring them closer to the wider community. The following set of experiments will investigate the potential use of a nudge that targets these three factors and whether this results in the identification of more phishing emails than a nudge that does not contain these factors.

Cognitive Strategy Adaptation. Whilst nudging and education around motivation can help guide users towards improved decision making, there are limitations around their ability to provide the full steps required for maintaining a more enhanced decision-making strategy. True debiasing is thought to require a full list of steps in order to be deemed successful including bias awareness (as explored in Study 6), motivation to change strategy (investigated in Study 7), and new strategy education: which underpins the main manipulation within Experiment 8 (Croskerry et al., 2013).

The adaptation of cognitive strategies is a form of debiasing that provides people with the complete process of information required to change the decision-making strategy applied (Larrick, 2004). It is however not enough for an intervention focused on adapting current cognitive strategies, to merely educate end-users on an advanced strategy, it must also help them to recognise when this strategy may need to be applied. Cognitive strategy manipulation is therefore a multi-step approach including around 5 steps (Wilson & Brekke, cited in Croskerry et al., 2013).

1. Make the decision-maker aware the bias exists;
2. Inform them on how to detect it;
3. Motivate them to want to change it;
4. Teach them how to apply the more optimal strategy;
5. Help them maintain it.

The key to altering cognitive strategies is to make individuals aware of the situational clues that may suggest a bias is taking place and then provide them with a more effective strategy to utilise moving forward. The aim of Study 8 is to investigate a number of debiasing strategies and their potential influence on end-user detection of phishing emails.

Studies 4 - 5 Aims

Studies 4 and 5 address the following thesis aims:

- To understand which methods of persuasion are currently being reported (as of 2020 – when the emails were reported) within a multi-national corporation and are potentially those most utilised (Study 4);
- To assess which threat actions offenders are encouraging end-users to undertake (Study 4);
- To determine which of these methods of persuasion are more or less likely to be categorised as suspicious (in 2020 – when Study 5 was conducted).

Findings will be discussed in conjunction with recommendations and information on next steps towards creating more targeted interventions.

Studies 6 – 8 Aims

Three studies will address the following project aims:

- Investigate whether the use of soft-paternalistic nudging that targets underlying human decision-making biases can assist end-users in the detection of phishing emails;
- Further explore the use of nudging in relation to the key factors underlying self-determination theory with the goal of verifying whether increases in motivation further improves the detection of phishing emails;
- Examine the use of two debiasing interventions (consider-the-opposite and a maxim, defined within Study 8) to determine their potential effectiveness in helping end-users identify phishing emails.

Study 4

The main aim of study 4 was to identify the methods of persuasion being utilised in phishing emails reported as suspicious by employees in a multinational corporation during February-March 2020 (noting before the global Covid-19 pandemic). In addition to analysing persuasion techniques, Study 4 aimed to identify other prevalent threat actions used within phishing emails (those that had been reported), as well as investigating any other tactics being applied.

Research Hypotheses

Hypotheses were dictated by the following research - Akbar, 2014; Butavicius et al., 2016; Rajivan & Gonzalez, 2018; Williams et al., 2018:

S4 H1 It was hypothesised that the authority principle would be the most prevalent method of persuasion in reported phishing emails;

S4 H2 The scarcity principle was predicted to be the second method of persuasion most present in reported phishing emails;

S4 H3 Social proof was expected to be the method of persuasion least utilised in reported phishing emails perhaps due to its high usage in spam emails e.g., marketing ploys.

Methods

A sample of 998 emails reported as suspicious by staff-members during February to March 2020 were analysed. Employees (anonymously) reported these emails by pressing a “report as suspicious” button within the *Microsoft Outlook* email system. The majority of phishing emails analysed in psychological research are those reported or identified by employees or from the research authors junk box (for example see Akbar, 2014; Ferreira et al., 2015). This is due to challenges acquiring a data bank of emails identified solely by a company’s security operations centre (SOC), if they have one. The current study was however novel in that it utilised a chronological and complete set of data reported within a single organisation in order to establish most prevalent techniques, whilst appreciating the limitation that the tactics unearthed are perhaps those easiest to be identified (hence being reported and noting that there will have likely been many other emails with influence techniques that were not identified and thus not reported either(. Due to the need to retain confidentiality, all emails within Study 4 were supplied in .txt format making it impossible to double check if phishing or spam, or indeed analyse elements such as image type, colour and formatting. Prior to receiving the emails, all contact details were removed as well as any other identifying features relating to the recipient and sender both in the header and the email body. Any email text files that were not in English, or that were blank were also removed. The total number of emails analysed after inclusion and exclusion criteria was 641.

Following the data cleanse, all emails were evaluated using a questionnaire by Zielinska et al. (2016), adapted from a paper by Ferreira et al. in 2015 (Appendix M). The purpose of the questionnaire was to uncover which of the six methods of persuasion have been applied within each email, with a set of questions feeding into each method. Example questions include whether the email was from an authoritative source, or whether it contained a sense of urgency. The questionnaire also analysed visual clues common in phishing emails such as spelling, punctuation and grammar - SPAG, and the email’s requested action e.g., clicking on a link or opening an attachment. During initial analysis, a further method of persuasion was found to be present, “curiosity”, whereby readers need to move outside of the email body to understand its full intention, for example to an external web page. Emails attempting to induce curiosity include the sharing of a file with an intriguing title e.g., staff salaries, or news stories that leave the reader on a “cliff hanger”. Therefore, two questions were devised to analyse the presence of curiosity within the phishing emails analysed including ‘Does the email ask the recipient to move outside of the email to 'learn more?' and 'Does the email require clicking on a link/attachment/download/replying to verify the email's full intention?'.

To accelerate data analysis, an automated search was developed to identify emails that contained links, images and attachments using common identifying strings such as "https://", ".mp4", "please see attached", and ".jpg" among others. Phrases that helped to assess the content of an email were also added to this search such as "click this link" and "please reply". Once the automatic searches had taken place, manual labelling of emails commenced with a response recorded for each question. Labelling of the emails against these questions were checked across two researchers to allow inter-rater reliability checks, with an initial 10% of ratings compared. If ratings on an item showed >10% disagreement, reasons for this were discussed, the rating criteria clarified across researchers and the item in question revisited across all emails. Another 10% of ratings were then rescored and compared until <5% disagreement was reached for each of the research questions, resulting in a >98% ratings agreement across emails. See Table 6 for examples of each technique found within emails during analysis. Note that an additional technique – curiosity – was unearthed and is therefore added to this list.

Table 6

Examples of Emails Analysed within the Current Study by Persuasion Type

Technique	**Email example (technique highlighted in bold, not in original email)**
Authority	I need a favour from you, email me back as soon as possible! Regards, **Senior Vice President**

Reciprocation	We are happy to offer you a 5-day complimentary trial access to this platform, **absolutely free**. You just need to reply to this email and we will set it up for you.
Commitment & Consistency	<NAME>, **As discussed and agreed** – we will collaborate on the document – link below to access it.
Social Proof	Conference Invitation **2000+ visitors** discovering how the latest solutions could be applied to their business for massive operational benefit
Liking & Similarity	Good day , **how are you and family**. I got your contact through Linkedin and internet search. I wish to discuss a very important partnership establishment on business and profitable investment opportunities in your country.
Scarcity	Alert Dear User, Due to your transfer to OSLA322 your access rights will be modified : **At mars 8, 2020 you will be losing your permissions for your folder**. If after mars 8, 2020 you still need access, you have to request access using http://sfs-m.securtity.corp click on 'A' to request Access.

Curiosity	Your colleague has shared a file. Click link to access it now. **https://clck.ru/Lrb7H**

Results and Discussion

The key objective of this study was to improve understanding around the potential methods of persuasion being used within phishing emails as of early (February – March) 2020, with a focus on those attempting to encourage employees to engage with them and potentially opening the way to a cyber-breach. Four main aspects were analysed: the persuasion techniques being applied, the threat actions most requested, any ancillary clues such as spelling errors, and any additional techniques identified.

Methods of Persuasion. First examined were the percentage of emails containing each of the currently recognised seven methods of persuasion resulting in a ranking of technique, ranging from 1 (most common) to 6 (least common). It is noted that some phishing emails contain more than one persuasion technique. However, and for this investigation, focus was placed on the number of emails containing an element of each technique – for example – an email with authority and scarcity would count as a data point towards each of these methods of persuasion. In terms of the original six common forms of persuasion, as hypothesised, authority was the most frequent method of persuasion utilised in the reported phishing emails analysed (69%; **S4 H1**) followed by scarcity (22%; **S4 H2**). Again, as anticipated the least frequently used method was social proof (8%; **S4 H3**). The ranking of methods was reviewed in line with previous research by Akbar (2014) and Ferreira et al. (2015), however a full ranking was not available in the Zielinska et al. (2016) paper so this was omitted.

Authority was hypothesised to be the method of persuasion most prevalent in the phishing emails analysed and appeared in almost 70% of the emails reviewed. Whilst it is appreciated that analysing reported emails may only signify those techniques easiest to identify, this finding supported by previous research. An air of authority can be evoked by use of work-based hierarchy or a claim of expertise. This is in line with findings from Akbar (2014) that analysed reported phishing emails in a crime and fraud organisation. Similar studies conducted within educational settings identified authority as the third most popular method of persuasion (e.g. Ferreira et al. 2015) and had noted an increase in authoritative educational phishing emails but a decrease in emails from banks and known companies (Zielinski et al. 2016). Despite authority presenting as the third most popular method in Ferreira et al. (2015), authority became the most successful when applied with other techniques e.g., authority and liking and similarity within one email. Emails in this study were also collated from the author's mailboxes and those found in open source, it would be expected perhaps that phishing to a multi-national corporation may target the authority principle above a technique such as social proof. Findings within this study suggest the authority principle is still utilised as a top persuasion technique within malevolent email communications and one that requires focus during intervention and future awareness campaigns.

Early in the process of analysis it became apparent that another technique was being utilised that was not covered within the six methods of persuasion put forward by Cialdini (1984). Approximately 37% of the emails analysed within the current study required the recipient to move beyond the content of the email body to fully understand its intention or learn more about the subject presented. This suggests the presence of a seventh method – *curiosity*. Human curiosity can be manipulated to encourage recipients to move outside of an email to fulfil a gap in knowledge around the email's true intent. It is a human appetite to learn more about a subject either by igniting a positive feeling of wanting to expand knowledge (joyous exploration), or a negative feeling of needing to fulfil a knowledge gap (deprivation sensitivity: Kashdan et al. 2020) that underpins this technique. Curiosity can be triggered by factors including a sense of novelty, complexity or incomplete information that motivates a human to behave a certain way (Tieben et al., 2011). Within phishing emails, curiosity can be ignited via a shared folder with an intriguing title e.g., “employee wages”.

Previous studies analysing methods of persuasion in phishing emails did not record the presence of curiosity so unfortunately comparisons cannot be made. Whilst Cialdini (1984) did not explicitly address curiosity as a technique, research findings suggest that social engineers exploit intrigue and curiosity as a form of motivation to act (Chaudhry et al., 2016; Krombholz et al., 2015). In a study by Benenson et al. (2017), curiosity was found to be the most reported reason for clicking on links, and the individual difference of curiosity was also found to increase susceptibility to phishing in a paper by Moody et al. (2011). Curiosity is not a new technique and has been used as a method of persuasion in other domains outside of cybersecurity for many years such as the marketing success of toy ‘blind bags’ for children (Grimmer and Grimmer, 2020) and the use of ‘...’ or ‘typing’ within messenger platforms to indicate that a contact is currently typing a message. However, its appreciation as a major weapon of influence in phishing emails adds to the novelty of this thesis. Curiosity is also often exploited in physical social engineering, for example, ‘baiting’ an individual to pick up and attempt to access an unknown USB stick to view its contents. Curiosity requires further investigation into its use in phishing emails and levels of success as it is clearly being utilised frequently as a method of persuasion – more so than a number of other methods reported in previous studies such as commitment and consistency and similarity and liking.

It was hypothesised that scarcity would be the second most prevalent persuasion technique within the report emails, with this finding confirmed (concurs with both the Akbar, 2014 and Ferreira et al., 2015 papers published 6-7 years prior to the recent report and based on emails analysed between 2010 and 2014; Table 7). However, the inclusion of curiosity as an important technique rendered scarcity the third most popular method exercised within the reported emails analysed. The scarcity principle is generally applied by offenders either suggesting a lack of quantity in an item or event, or a lack of time to access. Within phishing emails this could be the suggestion that an action be undertaken within a short period of time or else access to an email account or drive will be blocked. This evokes a sense of urgency within the recipient resulting in compliance with the email request prior to fully digesting its content and therefore any more obvious clues indicating its malevolence. As scarcity has been established as a technique of concern both within this and previous studies, it is important that future interventions apply particular focus to this principle.

Table 7

Methods of Persuasion Ranking Including Curiosity – Current Study and Previous Literature

Rank	Method	Current Study (%)	Akbar (2014)	Ferreira et al. (2015)
1	Authority	69 / 1st	1st	3rd
2	Curiosity	37 / 2nd	N/A	N/A
3	Scarcity	22 / 3rd	2nd	2nd
4	Reciprocity	14 / 4th	5th	4th/5th (joint)
5	Commitment	13 / 5th	4th	4th/5th (joint)
6	Similarity	10 / 6th	3rd	1st
7	Social Proof	8 / 7th	6th	6th

Note. 20% of all emails reported contained no currently defined method(s) of persuasion and thus at least some might have been genuine but reported for other reasons.

Also analysed were the number of emails that contained no method of persuasion (yet identified) to understand the percentage of potentially fraudulent phishing emails perhaps relying solely on visual clues, e.g., an image; whether there are potentially other weapons of influence being used that have not yet been recognised is a point for future research. As no previous phishing email analysis has tracked this figure it was impossible to determine any temporal changes – i.e., compared to other past studies. These emails did tend to contain clues such as spelling mistakes or potentially malicious links but were not persuasive in nature or at least not persuasive in ways currently understood.

Reciprocity, ranked fourth within the current study – moving up slightly in rank from the Akbar (2014) and Ferreiera et al. (2015) studies suggesting offenders are now more regularly utilising reciprocity (and/or staff are more likely to spot and report emails that contain it as a method of persuasion) promoting e.g., the offer of kindness or gifts in return for action. Commitment and consistency was fifth in ranking (albeit very similar in frequency to reciprocity suggesting that phishing emails containing information relating to previous contact(s) – building familiarity – remains stable in 2020 (based on the emails analysed) compared to 2010-2014. This is possibly due to the complexity of gaining commitment within one communication, often the principle requiring a two-step approach: commitment and then the individual remaining consistent with this previous commitment. Examples of commitment and consistency used in phishing emails can be as simple as ‘Dear Customer’, or suggestions of the recipient having been previously generous requiring them to feel they need to do so again.

The use of the similarity and liking principle (e.g., expressions of praise or attempts to build rapport) appeared to be utilised less/or indeed reported less in this current study over both the Akbar (2014) and Ferreira et al. (2015) papers. Social proof, for example (suggestions that similar others have already taken up an offer) remains the least prevalent method of persuasion (8%, Table 7) as was the case in both the Akbar (2014) and Ferreira et al. (2015) studies.

To summarise, authority, curiosity and scarcity were found to be amongst the most prevalent methods of persuasion used within a 2020 sample of emails reported by employees for being in some way suspicious. It must be noted that whilst the principles of reciprocity, commitment and consistency, similarity and social proof featured in the bottom four most frequent methods of persuasion within the reported emails – they are still being used by cyber-attackers and perhaps potentially more difficult for end-users to identify and report.

Taken together, these findings highlight how important it is to regularly analyse sets of reported emails to ensure organisations have a much better situational representation of the attack methods being used within phishing emails such that interventions, for example training and awareness campaigns, can be tailored accordingly and adapted quickly should the situation change.

Another important finding within this study was the number of emails containing none of the known persuasion techniques. No established methods of persuasion were found in approximately 20% of phishing emails reported, possibly highlighting the importance of finding other clues within phishing emails that can be used to alert end-users to their ingenuity. It could be that the images removed from the files analysed were persuasive in themselves. The finding of curiosity as a problematic technique indicates that there may be further persuasion techniques being used that are yet to be uncovered (using the analysis techniques employed within the current study).

It was however noted that the majority of emails with no yet known persuasion technique had spelling or grammatical errors (or similar) that made the email appear abnormal e.g., unconventional spacing or the use of copious amounts of question or exclamation marks. As previous papers did not capture statistics on phishing emails containing no persuasion techniques it is impossible to know whether this is more likely to happen in emails in an industrial organisational setting and / or whether this figure has changed across time as end-users become increasingly aware of the several persuasion techniques being used. It is also possible that phishing emails with no tactics are easier to generate, or that a number of offenders remain unaware of such methods discussed. Nevertheless, it is certainly not bad practice to report any email that for whatever reason that rouses suspicion amongst employees with an a priori Black Box Thinking approach being strongly advocated as another step towards achieving company-wide seamless cybersecurity (see Morgan & Asquith, 2021).

Additonal Findings. A number of additional findings were uncovered and recorded during email analysis. First, the use of multiple points of action, such as a link to ‘display the email within a web browser’, links to multiple end points, links to ‘unsubscribe’ should the email not be of interest, and requests to ‘open an attachment’ or ‘follow a link for more information’. Further research is required to determine how successful this is as a phishing technique (given multiple clues to potential malevolence) and/or whether it actually aids phishing detection. Taken together, these tactics highlight additional ways in which offenders are attempting to influence end-users as well as the complexity in which some phishing emails are being created. Other – more supplementary – clues that may guide intervention as well as help developers and end-users identify phishing emails were also analysed.

Threat Actions. Also analysed were actions offenders requested or implied that the recipient of the email undertake for an attack to result in a breach (see Table 8). Previous research found ‘clicking a link’ to be the top threat action contained in phishing emails, followed by the ‘request for an email reply’ and then third the ‘opening of an attachment’ (Akbar, 2014; Zielinska et al., 2016). Please note that some emails contained more than one of the below action requests so the percentages do not necessarily represent percentage total emails. Analysis of emails within the current study confirmed clicking on a link to be the most frequent action requested (71%), followed by a request to reply to the email (21%), and third to open an attachment (10%). Also analysed were the number of emails containing a request for the participant to complete a download (5%) and complete a form asking for confidential information (4%). This confirms that phishing email interventions need to remain focused on ways to stop people clicking on potentially malicious links, sending replies to emails with confidential information unless it is verified as safe to do so, as well as opening non-verified attachments.

Table 8

Top Threat Actions Requested within the Current Study and Previous Papers

Threat Action	**Current Study**	**Akbar (2014)**	**Zielinska et al. (2016)**
Click link	71%	60%	80%
Reply to email	21%	N/A	11%
Open attachment	10%	N/A	10%
Download	5%	N/A	N/A
Complete a form	4%	N/A	N/A

Note. The paper by Ferreira et al. (2015) did not measure these factors and is therefore no included in this table.

Whilst methods of persuasion are the most prominent elements within phishing emails followed by other threat actions, it is important to understand additional ancillary tactics or clues that may be equally useful in targeting human vulnerabilities or disclosing their intent. For example, 23% contained multiple question or exclamation marks, 3% spelling mistakes, 13% grammatical mistakes and 11% abnormal spacing. Akbar (2014) reported that 35% of phishing emails contained an image and this was slightly higher within the current study at 41% (Table 8). Images are used within phishing emails to quickly attract the attention of the recipient and often to evoke emotion, which in itself acts as a decision-making heuristic. SPAG errors have however seen a decrease from 81% in 2015 (Zielinski et al. 2016) to 44% within the current study involving emails from 2020 (see Table 9). SPAG is an important aspect of phishing emails both for the offender who may use it to bypass filtering, and for the defender as a tool to detect deception. It is possible that the decrease in SPAG issues is due to less tolerance within a business setting, or it could also be a temporal change where such mistakes have previously been inextricably linked with phishing emails and an increase in end-user recognition of their malevolent nature has resulted in offenders attending to this issue. Whilst this reduction does present a problem for deception detection, 44% is still considerably high and warrants focused intervention. Strong affect was present in only 6% of the emails analysed within the current study (Table 9), strong affect is defined as an attempt to induce fear, panic or excitement with an example being threats to terminate employment. It is likely that this figure is low due to strong affect appearing out of place in emails sent to employees within the workplace analysed (Table 9).

Table 9

Percentage of Additional Elements Found Within Analysed Phishing Emails

Additional Elements	**Current Study (%)**	**Akbar (2014) (%)**	**Zielinska et al. (2016) (%)**
Image Present	41	35	N/A
SPAG Errors	44	N/A	81
Affect	6	N/A	N/A

Note. The paper by Ferreira et al. (2015) did not measure these factors and is therefore no included in this table.

Further techniques were identified within this research that warrant highlighting here. The first being the use of fabricated email chains, with several emails containing what appeared to be previous conversations around an email subject. At times, the recipient did appear in the email chain (targeting the commitment and consistency method of persuasion) with others only including colleagues or peers in order to target social proof. For example, opening an email chain you appear to have previously been a part of will leave you wanting to remain consistent with that behaviour or alternatively an email chain featuring colleagues resulting in herding effects. Future research focused on email chains as an offender technique would be hugely beneficial.

The final tactic uncovered was attempts to ‘widen the web’ of deceit, with senders suggesting that the recipient forward the email to friends, family or interested colleagues. This suggests to the recipient that they should become a cybercriminal themselves without realisation - targeting a wider audience through “email forward” encouraging social proof as they move from ‘recipient’ to phishing ‘sender’.

Conclusion

The principle aim of Study 4 was to identify the top methods of persuasion being used in phishing emails, reported by employees within a multinational corporation in 2020 compared with those most recorded in the literature during 2014 (Akbar) and 2015 (Ferreira et al.) both based on reported phishing emails from 2010-2014. First hypothesised was that the authority principle would feature the most in the emails reported by employees (Akbar, 2014; Butavicius et al., 2016; Rajivan & Gonzalez, 2018; Williams et al., 2018). This finding was confirmed within Study 4 with 69% of reported phishing emails analysed containing this principle. It is possible that the emails reported contain the methods of persuasion easiest for employees to detect, however the finding of authority as the leading technique utilised does support findings from previous research and studies yet to come within this thesis (Akbar, 2014; Butavicius et al., 2016; Rajivan & Gonzalez, 2018; Williams et al., 2018). Authority appears to be a particularly robust persuasion technique with its presence and success rate in phishing emails of high concern. Organisations should focus intervention around this method, perhaps putting in place control mechanisms whereby emails from authoritative figures are unexpected e.g., communications from authority figures delivered via an alternative communication method.

The method of persuasion also hypothesised as being of key importance was the scarcity principle reported in previous research as a key technique of concern (Akbar, 2014; Butavicius et al., 2016; Ferreira et al., 2015; Williams et al., 2018). This finding was confirmed in Study 4 with the scarcity principle one of the top three methods of persuasion found in the reported emails analysed. The sense of urgency the scarcity principle evokes is of particular concern when teamed with authority, together believed to be the most potent influence tactic (Tiwari, 2020; Williams et al., 2017). It is imperative that investigations are conducted around the words or key terms used in phishing emails to induce this principle e.g., use of the word “urgent”, perhaps creating a lexicon that better supports email filtering and highlighting to users when to make decisions more consciously (research currently being conducted by the key author of this thesis).

Social proof was anticipated to be the method of persuasion less utilised in phishing emails reported by employees (Akbar, 2014; Butavicius et al., 2016). Again, this finding was confirmed within Study 4 with only 8% (the least number of emails) containing this method. It is possible that statements such as ‘80% of your colleagues have already completed the survey – please follow the link below’ are recognised as illegitimate and less likely to be expected within a working environment. Such techniques are often used in marketing campaigns and could perhaps be misidentified as spam. Whilst social proof does not often appear in phishing emails at present, it must continue to be tracked to ensure usage does not increase in the future as alternative techniques become more recognised.

Findings reveal that from the sample of reported emails analysed – authority (within 69% of emails and still the most frequently used method of persuasion based on emails reported), curiosity (within 37% of emails, newly identified as a method of persuasion in phishing emails) and scarcity (within 22% of emails – ranked second most popular in the past, third in the current study when curiosity is included) were the top methods of persuasion used. Whilst lower in prevalence–- reciprocity, commitment and consistency, similarity and liking, and social proof still appear within 8-14% of emails. Furthermore, a number of emails contained more than one method of persuasion with future investigations focused on those methods most likely to appear together as well as most likely to increase email potency when included hand in hand.

Investigations revealed that links remain the biggest threat action of concern, with organisations and awareness programmes required to place even more focus on ensuring employees know how to check if a link is malicious. Future research should investigate ways in which humans can be trained to habitually check these links with each email opened with Chapter 4 of this thesis focused on a number of potentially useful techniques. SPAG errors were identified in 44% of emails (albeit much reduced compared with 6-10 years ago), as were a number of fabricated email chains created to further persuade victims and attempt to widen the web of deceit by asking the recipient to forward the email on.

All seven methods of persuasion, threat actions and other techniques identified within the current study need to be featured as a core component of company-wide cybersecurity awareness and training and flagged to employees on a regular basis to increase vigilance. Additionally, email phishing filter software should be updated to better detect such emails that contain these features reducing the number that reach employee mailboxes. More research (especially in the field of human factors and cyberpsychology) is needed on alternative interventions (above and beyond training and awareness) to help employees better identify phishing emails. For example – nudging and email decision support systems as explored in Chapter 3.

It is also important to note that the current study has provided crucial baseline data on phishing email content as per 2020, enabling changes in techniques and prevalence over time to be tracked. Limitations were however present such that the emails analysed being those reported by employees – and therefore identified as suspicious not confirmed as phishing. Whilst similar to previous studies more must be done to create a more realistic repository of phishing emails, an action currently being undertaken by the lead author of this thesis. Due to challenges analysing data sets that are not solely reported phishing emails Study 5 will look to improve understanding around the methods of persuasion likely to be utilised by cybercriminal dues to their potency. By improving understanding around the success of each technique, alongside analysis of reported emails a novel picture can be drawn around the key influence techniques of concern and therefore those that organisations should most focus intervention.

Study 5

Interacting with emails, both at home and at work, has become habitual resulting in lower levels of suspicion and higher levels of processing emails automatically. Vishwanath et al. (2018) discuss this point as part of their Suspicion, Cognition and Automaticity Model of phishing susceptibility (SCAM) suggesting that in order to detect that a phishing email is malicious, suspicion needs to be present resulting in an email being processed more consciously compared to a more automatic and less cognitively effortful approach that generally occurs as default. When opening emails as habit, suspicion is less likely to be aroused and detection clues will possibly not be found with it therefore important to understand which of the techniques of influence used by cybercriminals are driving heuristic thinking when interacting with emails, lowering the likelihood of phishing detection. Email interaction will remain habitual, so it is important to establish interventions that highlight to an employee when conscious thought needs to be applied during such a routine behaviour.

The main aim of Study 5 was to therefore investigate, in real time (compared to the retrospective approach of Study 4) – via an online platform – which methods of persuasion within phishing emails are more likely to result in a risky response(s) from recipients. Despite previous phishing email analysis in literature, and findings from Study 4 (2020), the power of a persuasion technique may change across time as people become more aware of particular social engineering tactics, requiring the need for offenders to adapt methods and/or develop and apply others. Whilst the assumption is that cybercriminals are using techniques that are tried and tested, they will also likely be considering factors such as cost/effort to deploy versus level of success, as well as needing to adapt to changes in end-user perceptions and behaviours (see Morgan & Asquith, 2021). Taken together, the results of Study 4 combined with those of the current experiment in Study 5 can be drawn upon to design, develop and test interventions to support recipients whilst interacting with phishing emails helping them better identify potentially malevolent cyber-attack attempts.

Research Hypotheses

Study 5 (E5) H1 It was hypothesised that the authority principle will be the method of persuasion most effective in phishing emails and therefore most likely to result in attack success (Akbar, 2014; Butavicius et al., 2016; Rajivan & Gonzalez, 2018; Williams et al., 2018).

S5 H2 The scarcity principle was expected to be the second method of persuasion most likely to result in attack success (Akbar, 2014; Butavicius et al., 2016; Ferreira et al., 2015; Williams et al., 2018).

S5 H3 Social proof was anticipated to be the least likely method of persuasion to result in a cyber breach (Akbar, 2014; Butavicius et al., 2016).

Method

Participants. One hundred and ninety-two participants were recruited through the Cardiff University School of Psychology student participant pool receiving study credits in return for completion. Of these participants, 9% were male and 91% female, across an age range of 18 – 44 years with an average age of 20.67 (*SD* 2.91); the imbalanced gender and age ratio is quite common within undergraduate psychology degree programmes in the UK.

Desing and Procedure. The experiment employed a between-participants design to investigate how seven methods of persuasion (authority, commitment and consistency, social proof, similarity, reciprocity, scarcity, and curiosity) and a control factor (no method), influence the reporting phishing emails. Phishing emails were categorised as ‘successful’ if they were not identified as phishing by participants and were instead filed in a folder other than the suspicious email folder. The dependent variable was categorised as the number of phishing emails correctly identified, conceptualised as phishing emails filed within the email folder marked ‘suspicious emails’. Participants accessed the experiment through the online survey platform Qualtrics© and read a brief introduction (Appendix N) before providing informed consent. After completing a small number of demographic questions (gender, age, education) participants progressed to *Phishtray*: a browser-based email simulation tool (Joinson, Williams & Levordashka, 2018). Within Phishtray, participants were informed that they would be completing an email sorting task to help understand the level of satisfaction experienced when having actioned (see below for action options and Appendix O) all of 32 emails in their inbox (see Table 9 for examples of the emails used and Appendix P for the complete list).Appendix Pt). More trials would have been preferred in order to enable the analysis of more data point per condition; however, experiment fatigue over 30 minutes (the duration of the experiment within Study 5) was a concern. The true nature of the experiment was not revealed until completion (at debriefing; see Appendix Q) in order to avoid priming effects that would likely heighten participant suspicion above what it would be under non-experimental conditions. This was in line with ethical and risk assessment approval granted for the experiment by the Cardiff University School of Psychology Ethics Committee (SREC).

Participants were instructed to open each of the 32 randomised emails within the inbox in order in which they were presented to them and decide which folder to file each communication, across a number of folders such as ‘meetings’ and ‘reports’ as well as one named ‘deleted emails’ and one named ‘suspicious emails’. The 32 emails included four for each category of the seven methods of persuasion, with each set containing two genuine emails and two phishing emails, with a balanced mix of requested action (link/attachment) and sender type (internal contact/external contact). There were also four emails containing no method of persuasion (control emails). Once all 32 emails had been filed, participants returned to Qualtrics© and received full debriefing information including details on the main aims of the experiment as well as information on phishing emails and the persuasion emails under scrutiny so that they could improve their awareness.

Results and Discussion

Study 5 was conducted to experimentally investigate the influence techniques most likely to persuade recipients to perform an action that could result in a potential cybersecurity breach. These analyses are important as together with the findings from Study 4 (and elements of previous related studies) can be used to inform the development of interventions to better protect end-users from succumbing to the influence of phishing emails. It is important to note that a very large proportion of the participant pool identified as female (91%), perhaps due to the known issue around psychology programmes currently attracting more women than men to their courses. It is therefore difficult to generalise these findings across other samples such as men, particularly as research within the cybersecurity space does suggest that cybersecurity is experienced differently across gender types (e.g., Gratian et al., 2018; Whitty et al., 2018).

A one-way within-participants ANOVA was conducted on the effect persuasion techniques have on phishing email success. A statistically significant main effect was found on the type of technique used, accounting for a large proportion of the variance: $F(7, 1337) = 40.909$, $p <.001$, partial $\eta^2 = .18$. A significant linear trend emerged, $F(1, 191) = 35.431$, $p < .001$, with persuasion technique effect increasing above the control measure across commitment and consistency, scarcity and then authority (see Table 10 for associated means and standard deviations).

As predicted, authority, the portrayal of expertise within a phishing email, was the most successful technique used with it also found that participants are asked to click on links or open attachments in 68% of cases (**E5 H1;**Table 10). This technique also ranked high in Study 4 (within 69% of reported emails) and within a number of previous studies (e.g., Akbar, 2014; Butavicius et al. 2016; Ferreira et al. 2015; Rajivan & Gonzalez, 2018). This suggests that end-users – within the current experiment at least – were most susceptible to impressions of authority such as management titles, hints of expertise, the inclusion of qualifications, accolades and so on. This finding, together with the results from Study 4, highlight the importance of future research to investigate ways in which to specifically target authority as a phishing email method of persuasion and to find ways to minimise the lure associated with it (noting within Study 4 – emails containing this method may have been noticed and reported most often and within Experiment 5 – they were fallen for most often in terms of susceptibility).

Scarcity, the suggestion of a limitation on quantity or time in order to induce a sense of urgency, was ranked as the second most successful (not identified as suspicious) method within this experiment with participants not reporting 63% of scarcity induced phishing emails (**E5 H2;** Table 10). Scarcity ranked the second most successful method of persuasion in studies by Akbar (2014) and Ferreira et al. (2015) and sixth in the study by Butavicius et al. (2016). However, there are potentially some differences in how the scarcity principle was used within each of these studies – for example, scarcity of time or quantity, the sender of the email as internal and external to the organisation as well as emails targeting personal or business means. Whilst literature suggests scarcity has a mixed level of success as a phishing email method of persuasion, Studies 4 & 5 do however indicate that it is a technique that requires focused intervention, particularly within the business domain.

Table 10

Methods of Persuasion Reported in Study 4 and Marked as Suspicious in Study 5

Rank	**Method**	**Most Present (reported emails)** *Study 4*	**Most Successful** *Study 5*
1	Authority	69%	68% (*M* = 1.36. *SD* = .64)
2	Scarcity	22%	63% (*M* = 1.28. *SD* = .68)
3	Commitment	13%	49% (*M* = .99. *SD* = .47)
4	No Method	20%	49% (*M* = .98. *SD* = .76)
5	Similarity	10%	47% (*M* = .94. *SD* = .77)
6	Curiosity	37%	33% (*M* = .68. *SD* = .52)
6	Reciprocity	14%	33% (*M* = .67. *SD* = .54)
8	Social Proof	8%	32% (*M* = .65. *SD* = .69)

The third most successful method of persuasion within Study 5 was commitment and consistency (49%; see Table 10). This principle is achieved by an end-user wishing to be seen as consistent with a previous attitude or behaviour, or even a previously suggested personality trait. It was ranked fifth within Study 4 taking into account curiosity that was ranked second in that study. It is important to note that participants in this experiment were unaware of whether previous interaction was legitimate and therefore may have assumed this to have been the case more so than if it was their own (personal, work, etc.) email account. In the Parsons et al. (2019) study, commitment and consistency ranked top, however the emails used relied on a more personal context than the business context used in this experiment.

Phishing emails containing no known method of persuasion were the fourth most successful email type within this experiment (49%; see Table 10). Research by Butavicius et al. (2016) found emails containing no method to be the most successful tactic of all, suggesting this was due to an inoculation effect whereby humans are becoming more aware of persuasion techniques due to pre-exposure and therefore other techniques are being used. This is likely true with many methods of persuasion utilised in spam now ignored in phishing due to its notoriety.

Whilst this finding has not been fully verified within this current experiment, Study 4 did acknowledge that there was no known method of persuasion in 20% of all reported emails, and in Study 5, 45% of phishing emails containing no method were not identified. This offers some support for the fact that even without cybercriminals actively attempting to manipulate humans, natural human decision-making behaviour, as discussed in Chapter 2, can result in low level suspicion and therefore poor detection qualities.

Similarity and liking (47%) was ranked fifth in level of success with just under half of the phishing emails containing this method not filed by participants as suspicious (Table 10). Similarity and liking can be evoked through praise, and the building of rapport with its position further down the ranking possibly due to difficulty in authentically using this technique in a single communication without triggering suspicion. Whilst positioned fifth this persuasion technique still requires intervention due to nearly 50% of participants not reporting these emails as suspicious. However, with only 10% of reported emails containing this method of persuasion it is perhaps of less concern than those such as authority and scarcity. Of course, it is possible that lack of reporting could in fact be due to this persuasion technique being so successful it goes under the radar.

Curiosity, providing a lack of information whereby the recipient needs to click on a link or similar to learn more, was ranked joint sixth in the current experiment with around a third of phishing emails containing this method not reported as suspicious (33%; see Table 10). This could not be compared to previous research as curiosity has not been analysed within the methods of persuasion literature. Curiosity is ordinarily a very effective method used largely within the marketing and education domain with a person's intrigue evoking a desire to gain more information (Loewenstein, 1994). However, for curiosity to have its maximum (negative) effect it is suggested that a moderate level of information be provided, too much or too little and it is likely to lose its effect (Dubey & Griffiths, 2020). The emails used in Study 5, as well as the emails analysed in Study 4, presented the curiosity principle as either a simple link or full pages of part written articles with numerous links allowing readers to click on them to fulfil a knowledge gap. It is possible that both email styles are not truly capturing the essence of curiosity and maximising its evolutionary importance. In the year of writing this thesis (2023) phishing emails containing shared folders with intriguing titles (e.g., staff salaries) have become far more popular evoking curiosity in a potentially more impactful way. This should be investigated in future research, as well as interventions that focus on curiosity cues generally due to ease of deployment – especially given the alarming figures from both studies on the percentage of analysed emails that contained this method in Study 4 (37%), and the percentage not reported in Study 5 (33%).

Reciprocity ranked as joint sixth most successful method of persuasion within Study 5 (33%; see Table 10). Reciprocity is presented in phishing emails as the offering of kindness in order to receive action from the sender for example offering a free gift, prior to suggestion of survey completion. This is possibly due to the challenge of using this principle in email format. Reciprocity relies not just on the gift itself but also the intention of the gift giver, which is difficult to determine in text form (Falk & Fischbacher, 2006). It is also feasible that the overuse of free gifts within the marketing domain has allowed recipients to see such offers as purely spam. Whilst reciprocity was ranked in the bottom three most successful methods of persuasion within Study 5, 33% is still a very high and worrying percentage and taken together with the evidence from Study 4 that 14% of reported emails contain reciprocity cues – organisations and employees need to work together to find solutions to better detect cases and report them accordingly.

The persuasion technique found to invoke the (only just) least action within the experiment was that of social proof (32%; **E5 H3;** see Table 10). Study 4 and 5, within this set of studies, revealed social proof to be the least problematic method of persuasion (supported by Butavicius et al., 2016). Social proof in phishing emails can be the suggestion that peers have taken an action in the hope that the recipient will use this as a short-cut to deciding on their own behaviour. Social proof is a method regularly used within sales and marketing campaigns with it possible that participants were aware of the technique and therefore saw it as disingenuous. Despite this, phishing emails containing this method were still successful in just under a third of cases. Again, whilst this technique may not have the popularity and potency of authority and curiosity there is still concern should such as email reach employees with a one in three chance of it resulting in a breach.

Finally, a repeated measures Analysis of Variance (ANOVA) was conducted to investigate whether a significant main effect was present between persuasion technique utilised within each phishing email (and the control measure of no technique) with a significant difference found, Wilks' Lambda = .398, *F* (7, 185) = 40.034, *p* < .001. As hypothesised, authority (*M* = 1.36. *SD* = .64) and scarcity (*M* = 1.28. *SD* = .65) were found to be the most persuasive techniques utilised and social proof (*M* = .65. *SD* = .69) the least. Authority (*p* < .001) and scarcity (*p* < .001) were found to be significantly more successful than no persuasion technique (*M* = .98. *SD* = .76) and social proof less successful than the control condition (*p* < .001). This does suggest that clear links are present between the persuasion techniques perhaps deployed by cybercriminals (as seen in Study 4) and those the most impactful (Study 5).

Employees already experience a large number of vulnerabilities when it comes to cybersecurity, for example the natural human need to remain optimistic when it comes to risk, resulting in the downplay of threat probability. Cybercriminals add to this by purposely evoking additional heuristics that they know will result in bias and therefore attack success. Studies 4 – 5 suggest the heuristics most likely to be preyed upon are the human dedication to obey authority, and the dislike of losing autonomy. These heuristics could be evoked within a phishing email by the head of finance suggesting staff salaries will not be paid unless personal details are updated online within 24 hours. These tactics are perhaps particularly potent because social learning has trained humans to not question authority, and the urgency of the message limits this ability even further (Ferreira et al., 2015). However, despite both authority and scarcity having had a potentially high prevalence in phishing emails for some time, it is possible that different findings will be present within different organisation types. For example, cybercriminals may target/employees may be more susceptible to reciprocity phishing emails within registered charities, due to donation communications being highly expected. By better understanding the way in which employees may experience external persuasion tactics, interventions can be generated with a specific focus on reducing the vulnerable heuristics cybercriminals are looking to specifically target.

Conclusion

In summary, the experiment conducted as part of Study 5 was designed to examine to better understand how humans are interacting with the persuasion techniques, identified from reported emails in Study 4 across one hundred and ninety-two participants. As hypothesised, the authority principle was found to be the most successful method of persuasion (68% of all emails not reported as suspicious) with the technique also embedded in the highest percentage of phishing emails reported by staff at a multinational corporation and analysed in Study 4. Therefore, it appears that authority is the method of persuasion most likely to result in a security breach should phishing emails containing this method bypass other security systems and make it to the inboxes of staff working within such organisations, despite it also potentially being the case that that it is the technique identified and reported the most (Study 4).

Scarcity and commitment and consistency also featured as top weapons of influence suggesting intervention and awareness campaigns focus on these three methods in particular. Whilst curiosity ranked 6th in relation to level of success, emails were still not reported as suspicious in a third of cases alongside findings that the second most prevalent persuasion technique utilised in the reported emails in Study 4 was this particular tactic. This method also appears to be present in contemporary phishing emails through the use of shared folders with intriguing titles.

Finally, social proof – the method of persuasion found to be the least likely to be deployed by cybercriminals in Study 4 still worked as a successful method 32% of the time. It is important to also note the possibility that participants in Study 5 may have become aware of the study's true objective with therefore, real-world results likely to be higher. This behaviour was also in the absence of the typical time pressures and multi-tasking requirements within many working environments – i.e., in Study 5 the actioning of the email task was the only thing they were required to do. Overall, findings highlight the importance of deducing ways to reduce the number of phishing emails reaching employees, perhaps something that can be obtained by improving the technical filtering of phishing persuasion, or by highlighting to the recipient that emails may be evoking some of these techniques and to move into a more conscious decision-making mode. Research also needs to continue to improve SETA in relation to social engineering techniques, else all methods will continue to have a higher chance of resulting in a security breach.

Study 6 – Tier 1 Debiasing: Nudging

A number of human characteristics have previously been identified as influencing cybersecurity behaviour as referred to throughout this thesis. In Chapter 2, cybersecurity awareness and its six underlying measurement factors, were unearthed as key to encouraging secure behaviour with findings suggesting that protective behaviours will be reported by employees i.e., assessing the risk of cyber-threat to be high, maintaining awareness through knowledge-sharing, helping inform policy and feel attachment to their work devices. Should, for example, employees perceive risk as low, they will likely not be motivated to complete the security behaviours required (Rogers, 1975). The first portion of Chapter 3 focusing on vulnerability exploitation, built upon the findings of the studies in Chapter 2 by investigating additional decision-making influences, but those that cybercriminals are most likely to exploit within social engineering in phishing emails. Whilst all techniques are of concern, authority, curiosity and scarcity appear to be the tactics that are most likely to result in a cyber-attack. With a number of decision-making biases found to be of concern within Chapter 2, it is important to investigate ways in which organisations can reduce both internal vulnerabilities such as threat appraisal, as well as external vulnerabilities such as social engineering persuasion techniques.

Threat appraisal was found to inform cybersecurity awareness within Chapter 2, likely underpinned by several cognitive biases including the availability bias, a cognitive schema whereby humans evaluate the probability of an event occurring by how readily relevant instances can be brought to mind, and the unrealistic optimism (Tversky & Kahneman, 1973). Often organisations withhold information from employees around the regularity of security incidents to not cause concern, with employees left assuming such occurrences are rare and without the examples of threat that will motivate protective behaviour.

Nudge research to date has largely focused on the use of fear appeals to increase threat appraisal with findings mixed and suggestions that use of a coping message may be more effective (van Bavel et al., 2019). There is however some belief that fear appeals focus only on increasing knowledge around risk probability and that the human tendency towards unrealistic optimism mitigates the impact of this knowledge on preventative behaviour (Arkes, 1991; Jolls & Sunstein, 2006). According to the optimism bias, humans regularly overestimate personal positive outcomes and underestimate personal negative outcomes, impacting how they forecast risk (Pfleeger & Caputo, 2012; Warkentin et al., 2013). Therefore, whilst humans can be made aware of risk, they will still underestimate it in relation to themselves and their organisation against the human average (Warkentin et al., 2013). It is therefore suggested that a nudge looking to improve cybersecurity behaviours by increasing threat appraisal should also look to debias unrealistic optimism for the nudge to gain maximum effect (Rhee et al., 2005). The nudges designed within Study 6 therefore first take on the format of a fear appeal nudge, whereby a message is provided in an attempt to increase the perception of threat probability, in this case the number of cyber-attacks experienced each year. An additional nudge will then explore the addition of an optimism cue, looking to decrease participant unrealistic optimism in relation to cyber-attacks, discussed in full in Chapter 2. This nudge will include text that reminds the participant that they are at equal risk to other. The fear and optimism nudge will then also be tested in conjunction with a coping message that provides the reader with details on how to protect themselves from the risk detailed in the fear appeal, in order to increase self-efficacy. The addition of a coping message into a fear appeal allows for a far more ethical nudge than a nudge that induces fear alone.

Thaler and Sunstein in their book ‘Nudge’ (2008) provide ten key nudges that can be utilised to support humans towards making more optimal decisions with typical examples provided below:

1. **Provide a default option** (setting a printer to automatically print double-sided) – Auto suggest strong but memorable passwords for company systems to avoid the creation of those more obvious.

2. **Simplify the action** (sending a stamp addressed envelope when requesting a reply) – Ensure the inclusion of an embedded phishing reporting button to reduce the physical cost of emailing security.

3. **Utilise social norms** (stating 9/10 people actively recycle) – Advertise the number of employees that have completed cybersecurity training or are actively utilising the skills learnt to encourage buy-in.

4. **Increase ease and convenience of choice** (placing healthy food at eye level) – Ensure cybersecurity policy is easily accessible and digestible.

5. **Disclose information** (providing nutrition information on food packaging) – Educate employees on internal and external security incidents to encourage a more realistic appraisal of risk.

6. **Provide a warning** (evoking fear through graphics on cigarette packaging) – Provide emotive posters that evidence the impact of cyber-attacks, utilising strong imagery and fonts.

7. **Encourage precommitment** (setting a goal for financial savings) – Actively ask employees to set security goals such as locking their device each time they move away from their desk. Humans like to stick with predefined objectives.

8. **Set reminders** (a calendar alert reminding you of an appointment) – Send reminders or calendar invitations for training via email, Teams or Slack to nudge your employees towards training or more secure behaviours.

9. **Implementation intentions** (asking a person if they plan to vote) – Ask employees their intention to comply with policy and training, requesting feedback on how they will look to apply the skills learned, so they feel committed to implementing them during their working day.

10. **Informing on past behaviours** (how many steps taken the day before / calories burnt) – Communicate previous security behaviours and an individual cybersecurity human risk scorecard to employees, allowing them to respond to active feedback.

Study 6 will investigate whether the use of a number of soft-paternalistic nudges such as that attempting to evoke threat or fear (a warning nudge) and/or a coping message (simplify the action nudge) will improve behaviour on a phishing simulation task that is attempting to exploit decision-making vulnerabilities via persuasion. As well as to establish, whether behaviour is significantly better (i.e., less risky choices made) when participants receive a fear appeal that also informs them that they are at equal risk to those around them (optimism bias), compared to a fear appeal only condition.

Research Hypotheses

S6 H1 Interaction with a soft-paternalistic nudge, will result in participants filing significantly more phishing emails as suspicious than a control measure of no nudge (Furnell et al., 2019; Petelka et al., 2019; Turland et al., 2015).

S6 H2 A nudge containing a coping strategy nudge will result in significantly more phishing emails being filed as suspicious than a threat appraisal nudge (Bavel et al., 2019).

S6 H3 A nudge containing a threat appeal that targets both the availability bias and optimism bias will result in significantly more emails being filed as suspicious than a nudge containing a threat appeal only and a coping appraisal only (Arkes, 1991; Jolls & Sunstein, 2006).

Method

Participants. Two hundred and fifty participants were recruited via Prolific, of which 57% were male and 43% female with an average age of 35.73 years (*SD* 10.72). Participants were randomly assigned to one of seven experimental groups and asked to participate in an email sorting exercise taking ~30 minutes. Prolific pays an hourly rate (£12.05) to its users for completion of questionnaires, so participants were paid around £6.00 for taking part.

Design and Procedure. The experiment within Study 6 employed a between-participants experimental design to investigate how six categorical nudge conditions and therefore six independent variables (basic, threat appeal, optimism appeal, threat and optimism appeal, coping message, coping message and optimism appeal) and a control condition influence participant behaviour in a phishing simulation task (see Table 11 for nudge content and Appendix R for example of nudge appearance). The dependant variable was classified as the successful detection of phishing emails, conceptualised as the number of phishing emails correctly filed as suspicious by participants. Analysis also took place to identify whether nudge type significantly impacted the number of emails deleted, and were perhaps therefore identified as spam, as well as any false positives (genuine emails incorrectly filed as suspicious) that may suggest nudges were arousing suspicion generally rather than directly targeting phishing identification. False positives are regularly used within phishing research as a dependant variable, particularly in relation to email blacklisting or human individual differences (Abu-Nimeh et al., 2007; Kleitman et al., 2018; Prakash et al., 2010).

Participants accessed the experiment through Qualtrics© on PCs and tablets before being asked to read a brief introduction sheet. After completing a request for consent participants completed a demographics form recording their age, gender and level of education. Study 6 was undertaken in collaboration with ThinkCyber, an organisation that utilises soft-paternalistic nudging to guide employees towards secure behaviours. The focus of this experiment was to provide a true-to-life simulation utilising ThinkCyber’s Redflag® technology. Redflags® are computer-based pop-ups, programmed to appear in an end-users inbox providing them with context based information, such as a warning. End-users are required to close the pop-up box in order to continue working within their inbox.

Table 11

Study 6: Text for each Nudge Condition

Nudge Type	Text
Basic nudge	This inbox is at risk from fraudulent emails.
Threat appeal	This inbox is at risk from fraudulent emails. 241,324 phishing attacks took place last year.
Optimism appeal	This inbox is at risk from fraudulent emails. Your inbox is at equal risk to others.
Threat and optimism appeal	This inbox is at risk from fraudulent emails. 241,324 phishing attacks took place last year, your inbox is at equal risk to others.
Coping message	This inbox is at risk from fraudulent emails. Always check details such as sender email address and look out for clues such as spelling and grammatical errors.
Coping message and optimism appeal	This inbox is at risk from fraudulent emails. Always check details such as sender email address and look out for clues such as spelling and grammatical errors, your inbox is at equal risk others.

After completing the required demographics with Qualtrics©, participants were directed to Amazon Workspaces virtual desktop via an external link (https://clients.amazonworkspaces.com/) where they were asked to download a client for their Mac or PC. Participants were provided a registration code and asked to email the researchers through the Prolific platform in order to obtain a username and password to access the task. Upon log in, a virtual desktop was loaded that presented a readme document and an email icon providing access for the task itself. The readme document explained the purpose of the exercise ("exploring how people interact with emails when clearing an inbox") and informed participants that during the task they would play the role of manager for a fictitious company. Participants were then directed to open the mock Outlook inbox and to work through the emails to clear the inbox. As the participant clicked on each email, nudge pop-ups appeared that contained information relating to their respective experimental group (see Table 11). Nudges were presented as a simple pop-up in front of the list of emails at the start of each experiment only, where participants were required to close the pop-up box pr0ior to continuing with the task.

Contained in the inbox was a set of thirty-two emails that participants were advised to open and read each email in order before filing each in one of nine available inbox folders (urgent, follow up, finance, networking, files, IT, personal, suspicious emails, deleted emails). Emails were a mix of genuine emails and phishing emails at a ratio of 1:3 (genuine:phishing). Fewer phishing emails were sent to participants within Study 6 than Study 7 and 8 at the request of the company funding the nudging software utilised. This was to provide a more realistic inbox setting, in the hope to reduce any potential priming effects. The number of phishing emails included in Studies 7 and 8 were however increased, in order to provide more data points for analysis. It was ensured that each email would have an associated folder in which to file it outside of the "suspicious email" folder e.g., an email about a software update could be filed in the 'IT' folder or marked as suspicious, this was to detract from the experiment's true aim. Emails could be filed by clicking on and dragging and dropping into the chosen folder. The categorisation of emails as "safe" (urgent, follow-up, finance, networking, files, personal, deleted) or "dangerous" (suspicious) were a marker of cybersecurity decision making.

A selection of emails were chosen from Study 5 (with any sensitive data removed) that contained a mix of persuasion techniques, internal and external sources and links and attachments. Of the thirty-two emails utilised, eight contained clues of phishing such as malicious links and SPAG errors, and sixteen represented genuine emails. By presenting a larger number of genuine emails, it was expected that participants would be less likely to become aware that one of the main aims of the experiment was to focus on phishing. All emails were presented in a randomised order. Should links or attachments within the emails be clicked participants were directed to a web page that informed them that their desired action had been successfully completed. Unfortunately, it was not possible to record this data, however allowing this action made for a more immersive experiment.

The dependant variable within this study was the number of phishing emails currently filed as suspicious conceptualised as the number of phishing emails participants filed in the 'suspicious' mailbox folder. Analysis also took take place on the number of phishing emails filed as 'deleted' and perhaps identified as spam, and the number of legitimate emails filed as suspicious (false positives), perhaps suggesting that suspicious had been raised across emails and not just phishing. Participants were advised that once all emails were filed the task was deemed complete and participants asked to return to the Qualtrics© online tab to be debriefed.

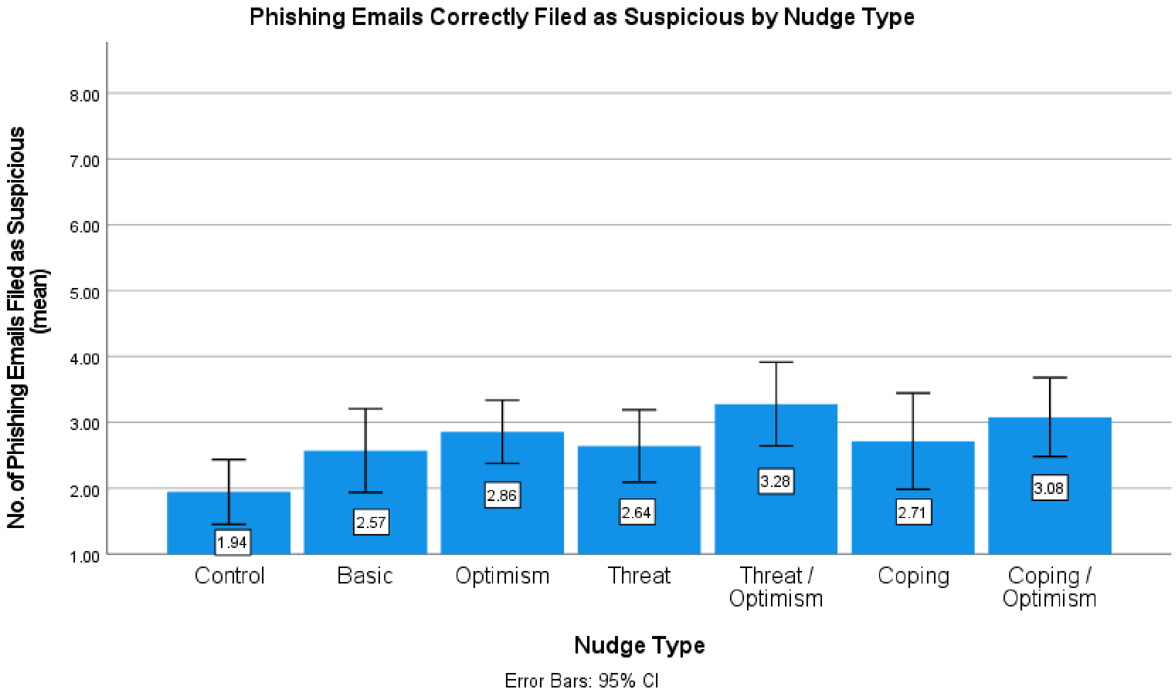
Results

Study 6 was designed to investigate whether behaviour on a phishing simulation task, would be significantly safer for those in receipt of a nudge (basic, threat appeal, optimism appeal, threat and optimism appeal, coping message, coping message and optimism appeal) Of particular interest was whether a threat appraisal nudge targeting the availability bias *as well as* the optimism bias (two examples of human decision-making vulnerability potentially influencing threat appraisal), would be significantly more effective than no nudge, as well as the other nudges under analysis. Any observations that were missing in the dataset were replaced with the grand mean value for each question and any outliers determined by 3 IQR from the mean were windsorized to the next available value not considered extreme.

As data was highly negatively skewed, it did not meet parametric assumptions and therefore a non-parametric test of difference was applied. A Kruskal-Wallis test was undertaken to investigate the influence nudge type (basic, threat appeal, optimism appeal, threat and optimism appeal, coping message, coping message and optimism appeal) and a control has on the number of phishing emails participants correctly filed as suspicious during a true-to-life inbox simulation (**S6 H1;** see Figure 10). A significant difference was found, χ^2, N = 250 = 14.737, p = .022, with a moderate effect size (ϵ^2 = 0.06). Post-hoc pairwise comparisons were then analysed to identify which nudge types were significantly more effective, adjusted by the Bonferroni correction for multiple tests (Figure 9). Comparisons across these tests identified a significant difference between the threat appraisal and optimism statement and no nudge only (**S6 H2; S6 H3**) with this nudge therefore the most effective of the experimental conditions (χ^2 = 10.314, p = .027; **E6 H1c**). Whilst the other nudge conditions (basic, threat appeal, optimism appeal, coping message, coping message and optimism appeal) and a control condition did not reach significance, all nudge conditions did result in more phishing emails being filed as suspicious than the control condition.

Figure 10

Phishing Emails Correctly Filed as Suspicious by Nudge Type



Also analysed were the number of false positives experienced, measured by the number of legitimate emails that were incorrectly filed as suspicious, by nudge type. No significant differences found (χ^2 = 10.938, *p* = .090). Next analysed was whether nudge type had a significant effect on the number of phishing emails deleted during the experiment with again, no differences found (χ^2 = 4.856, *p* = .562) Although when analysing whether a significant difference could be found in the number of phishing emails correctly filed as both phishing and deleted by nudge type, a significant difference was identified, χ^2, *N* = 250 = 12.961, *p* = .044, with a small to moderate effect size (ϵ^2 = 0.05). Post hoc analysis utilising Bonferroni correction also revealed only the threat and optimism nudge to significantly differ from the control condition when both deleted and reported phishing emails were analysed (*p* = .039).

Discussion

The main aim of this experimental study was to validate several previous findings that suggest soft-paternalistic nudging to be useful in aiding phishing email detection (Furnell et al., 2019; Petelka et al., 2019; Turland et al., 2015). However, this particular experiment was also interested in a different angle, in that it was also interested in understanding whether such nudges were actively useful in targeting human cybersecurity decision-making vulnerabilities (namely the availability and optimism biases within threat appraisal) or whether their use was possibly adding cognitive burden to the human cybersecurity experience.

It was first hypothesised that the use of a soft-paternalistic nudge, generally (and across all nudge types included within this study) would result in significantly more phishing emails identified as suspicious than no nudge. This finding was upheld within the current experiment whereby participant behaviour was significantly different (for the better) when a nudge was deployed. It was also predicted that the use of a nudge containing a coping message would be significantly more effective for participants than one containing a threat appraisal (for example., see Bavel et al., 2019), however this finding was not substantiated. Neither was the hypothesis that a nudge containing a threat appeal, targeting both the availability bias and optimism bias, would result in significantly more emails being filed as suspicious than a nudge containing a threat appeal only and a coping appraisal only (Arkes, 1991; Jolls & Sunstein, 2006).

However, what was established from the analysis of experimental data, was that the fear appeal nudge that was generated to increase both the availability bias and reduce the optimism bias to alert employees to the potential risk of phishing emails, resulted in significantly more phishing emails being filed as suspicious than no nudge. A finding not substantiated across any of the other nudge types. Significant differences were however not found *across* nudge types, nor those emails that were genuine but incorrectly filed as phishing (false positives). These findings do suggest that perhaps a threat and optimism nudge is useful in raising suspicion around phishing emails only. Many email providers and/or organisations now supply end-users with an added phishing ‘reporting’ button, contained within electronic mailbox, often in the form of a red fish shape to attract attention. The purpose being to simplify the process of announcing to the email provider or security team managing the inbox that a cyber-attack is possibly being attempted. False positives can cause an issue in for those at the end of the reporting button having to sieve through the emails to understand which are of concern and which are merely spam. Any phishing identification interventions applied within organisations must therefore ensure that whilst it increases the number of phishing emails being reported, it does not also inflate the number of false positives such teams receive (Jenson et al., 2022; Nakayam et al, 2009).

Also investigated was whether a nudge containing a threat appeal alongside an attempt to reduce unrealistic optimism resulted in significantly more phishing emails being deleted than no nudge, as well as whether significantly more phishing emails were being both deleted and reported across nudge types. Whilst deletion of phishing emails alone did not reach significance, it was found that the threat and optimism nudge did result in significantly more phishing emails being filed as suspicious or deleted over no nudge. Three types of emails are known to exist – those that are sent for malicious purposes and require reporting (phishing), those that are usually attempting market a product or service and are often not desired and require deleting (spam) and those that a recipient may genuinely expect/wish to reply to and are therefore more desired (ham; Bassiouni et al., 2018). Whilst this finding does provide further support for the use of a threat and optimism nudge, it may also suggest that (a) whilst the nudge increased suspicion, the participants were unable to decide whether the phishing emails were in fact phishing or spam (b) this nudge may have aroused suspicion to the fact that the email was phishing but participants did not want to mark it as such through fear of being incorrect (c) participants were not personally concerned about where the email was filed so long as it was ‘removed’ from their inbox. It is likely that the first option was at play, due to the fact that a large proportion of spam emails also contain persuasion techniques (Gamez, 2018, Tallard, 2000) and may therefore be difficult to decipher from the body of something more malicious.

Ethical concerns do however need to be considered when nudging end-users with a threat appeal that does not consist of details around a coping mechanism for preventing the risk that has aroused such fear, with it possible that the benefit of evoking fear outweighs the potential negatives (Nagai et al., 2022; Sharot, 2011; White et al., 2011). As well as investigating the use of a motivation debiasing strategy, Study 7 will also consider the inclusion of a threat appraisal and coping appraisal nudge alongside as a form intervention.

Whilst findings within Study 6 offer some support for the use of soft paternalistic nudging, particularly when used to increase threat appraisal (including the availability and optimism biases), further work is required to understand a) nudge content that is most influential, b) nudge context that is most influential e.g., shape, colour, font, and c) at what point end-users become desensitised to these nudges and what can be done to maintain alertness to their message.

Study 7 – Tier 2 Debiasing: Education Around Motivation

Study 7 is aimed at further exploring interventions that could prove useful in increasing end-user phishing email detection. Whilst Study 6 investigated the potential benefit of increasing threat appraisal and reducing optimism via soft-paternalistic nudging, Study 7 will explore whether phishing detection can be further enhanced through increased motivation. With cybersecurity tasks unlikely to induce pleasure in end-users, interventions looking to increase motivation to report phishing emails must instead find ways to increase perceptions of value in the task itself in order for extrinsic motivation to be as self-determined as possible. Self-Determination Theory is a psychological framework that suggests humans become more motivated to undertake uninteresting tasks should they not feel obliged to do so, feel competent in doing so and when the action enhances interpersonal attachment (Ryan & Deci, 2020). Often cybersecurity tasks are enforced on employees that do not have the capabilities to undertake them and are not fostered by those around them. Menard et al. (2017) investigated the benefits of including an SDT appeal to one focused on PMT and found the addition of motivational factors to be of significance. Yang et al. (2020) continued this work by presenting participants with either a self-determined appeal, a fear appeal or both, finding that both together were optimal in helping explain 58.4% of behaviour.

With this in mind, Study 7 continues the work of Study 6 by investigating the influence of a fear appeal and optimism nudge alone (to increase threat appraisal), as well as one that incorporates a coping message (to increase information security self-efficacy) to encapsulate more of factors underlying the Cybersecurity Awareness Framework (CAF). Study 7 also aims to understand whether an SDT appeal focused on increasing motivation is also effective as suggested, and whether both together can have the impactful results found in Yan et al. (2020). The main aim of Study 7 is therefore to investigate the use of a debiasing technique that looks to increase motivation to report phishing emails and whether it can encourage end-users to detect phishing emails, as well as extending the findings from Study 6.

Research Hypotheses

S7 H1 It was hypothesised that a nudge containing a threat appeal also attempting to reduce unrealistic optimism would result in significantly more phishing emails being filed as suspicious than no nudge (targeting threat appraisal and confirming the findings of Study 6).

S7 H2 The inclusion of a coping appraisal into a threat appeal would significantly increase nudge success, as well as provide a more ethical nudge providing participants with a way in which to protect them from the fear such an appeal may evoke (targeting threat appraisal and information security self-efficacy; see. van Bavel et al., 2019 on coping appraisals)

S7 H3 A nudge containing a motivation statement targeting SDT would result in significantly more phishing emails being filed as suspicious than a threat appraisal and optimism nudge alone (Ryan & Deci, 2020).

S7 H4 A nudge targeting both the two aspects if the Cybersecurity Awareness Framework (CAF; threat appraisal and information security self-efficacy) and a motivational statement would result in significantly more phishing emails being filed as suspicious than both nudge alone (Yang et al., 2020).

Method

Participants. Five-hundred and twenty-five participants were recruited via the Prolific online marketing tool, of which 44% were male, 55% female, 0.5% of a different identity and 0.5% declined to comment with an average age of 37.42 years (*SD* 10.59). A larger sample was afforded above Study 6 due to an increased budget within the multi-national organisation supporting this thesis, as well as Study 6 requiring far more administration and analysis time per participant (working with the collaborating organisation, ThinkCyber), due to its true-to-life scenario style. A larger sample results in data that is more representative of the population, limiting the influence of outliers or extreme observations compared with experiments with smaller samples. Participants were randomly assigned to one of five experimental groups (threat appraisal nudge, threat appraisal and information self-efficacy nudge, motivational statement, threat appraisal and information security self-efficacy nudge plus a motivational statement) before being asked to take part in a task investigating how people unclutter an inbox. Prolific (at the time of this research) pays an hourly rate (£12.05) to its users for completion of questionnaires, so participants were paid accordingly (circa £6 for a 30-minute experiment).

Experimental Design and Procedure. This experiment employed a between-subjects experimental design to investigate how four levels of debiasing (threat appraisal nudge, threat appraisal and information self-efficacy nudge, motivational statement, threat appraisal and information security self-efficacy nudge plus a motivational statement) and a control condition influence participant behaviour in a phishing simulation task (see Table 12 for nudge content).

Participants accessed the experiment via Qualtrics© as with previous experiments, where they were presented with an instruction sheet informing them that they would be carrying out a simulated email task to help understand how people interact with emails when clearing an inbox. Playing the role of a fictious manager they would be presented with a number of emails that they would need to read and then file into one of nine available folders (urgent, follow up, finance, networking, files, IT, personal, suspicious emails, deleted emails).

Table 12

Study 7: Text for each Nudge Condition

Nudge Type	Text
Control	Thank you for agreeing to take part.
Threat Appraisal Nudge (threat and Optimism nudge from Study 6)	241,324 phishing attacks took place last year, your inbox is at equal risk to others.
Threat Appraisal and Information Security Self-efficacy Nudge	241,324 phishing attacks took place last year, your inbox is at equal risk to others. Always check details such as sender email address and look out for clues such as spelling and grammatical errors.
Motivational Statement	We recommend to our colleagues always checking the sender email address and look out for clues such as spelling and grammatical errors. Feel free to do so.
Threat Appraisal and Information Security Self-efficacy Nudge and a Motivational Statement	241,324 phishing attacks took place last year, your inbox is at equal risk to others. We recommend to our colleagues always checking details such as sender email address and look out for clues such as spelling and grammatical errors. Feel free to do so.

Prior to commencing the task, participants were randomly assigned to one of the five experimental conditions and were presented with the text of either one of the four nudges or, for the control condition, a statement that thanked them for agreeing to take part. Participants were then presented with thirty-two emails, randomised and across a genuine:phishing ratio of 2:2, with content informed by the reported phishing emails unearthed in Study 4. These emails included a mix of persuasion techniques, internal and external senders and links or attachments. Of the thirty-two emails, sixteen contained clues of phishing and sixteen were presented as genuine. Once each email had been assigned to a folder location the experiment was deemed complete with participants debriefed and thanked for taking part.

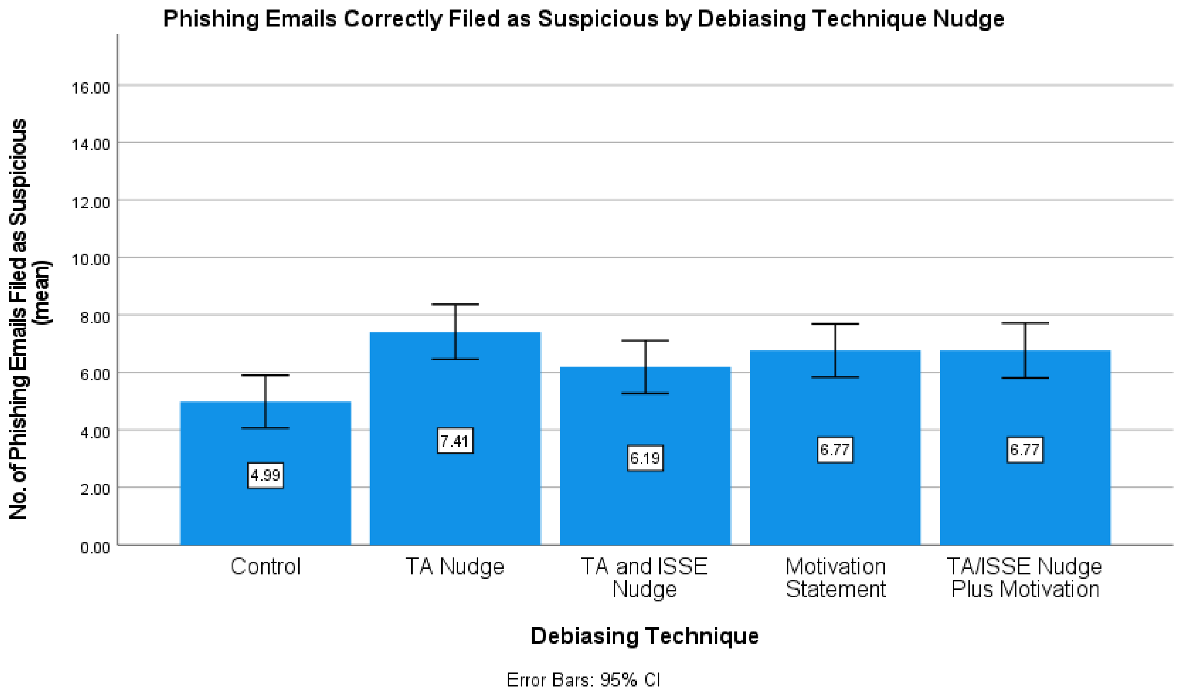
Results

The main aim of Study 7 was to investigate whether a motivational communication would significantly influence behaviour, and whether the inclusion of motivational aspects into a nudge targeting a number of key aspects from the Cybersecurity Awareness Framework (CAF) could further boost the merit of the soft-paternalistic nudge found to be significantly effective over no nudge in Study 6. Any missing observations, such as non-completion of email classification, were replaced with the grand mean value for each question and any extreme outliers determined by 3 IQR from the mean were windsorized to the next available value not considered extreme. Data was highly negatively skewed and therefore did not meet parametric assumptions. A Kruskal-Wallis test of difference was applied to investigate whether four nudge types and a control condition significantly differed in the number of times participants correctly filed a phishing email as suspicious.

A significant difference was found between experimental debiasing conditions and a control condition, χ^2, $N = 525 = 24.511$, $p < .001$ (**E7 H1**), with a small to moderate effect size ($\epsilon^2 = 0.5$), in relation to how many phishing emails were correctly filed as suspicious, adjusted by the Bonferroni correction for multiple tests. The threat appraisal and information security self-efficacy nudge containing both a fear appeal and optimism content significantly differed from the control condition (χ^2 = -98.933, *p* < .001; **E7 H1c**), and this nudge also resulted in significantly higher numbers of phishing emails correctly filed as suspicious than the threat appraisal and information security self-efficacy nudge (χ^2 = -75.402, *p* = .040; **E7 H1a**) with slightly less emails correctly filed in this condition. The combined threat appraisal and information security self-efficacy nudge plus the motivational statement condition significantly differed from the control condition (χ^2 = -55.336, *p* = .048; **E7 H1d**) with those in receipt filing higher numbers of phishing emails into the suspicious emails folder in relation to the control condition. However, and despite clear differences in numbers of emails filed as suspicious across debiasing conditions, no other significant differences were found between experimental conditions (see Figure 11) – i.e. there were only differences between the experimental conditions and the control condition.

Figure 11

Phishing Emails Correctly Filed as Suspicious by Condition Type



Note. TA – Threat Appraisal, ISSE – Information Security Self-efficacy.

Analysis also took place to identify whether the number of genuine emails incorrectly identified as phishing (false positives) differed across nudge, perhaps suggesting that nudges were simply arousing global suspicion rather than to phishing emails only. No significant difference between nudge types was identified (χ^2 = .756, p = .944). Also analysed was whether nudge type had a significant effect on the number of phishing emails deleted rather than reported as suspicious during the experiment with no differences found (χ^2 = 6.944, p = .139). In addition, it was analysed whether there would be significant differences in the number of phishing emails correctly filed as both phishing and deleted by nudge type, with no significant differences found (χ^2 = 4.025, p = .403).

Discussion

As with Study 6, it was hypothesised that a nudge containing a threat appeal that also attempted to reduce unrealistic optimism would result in significantly more phishing emails being filed as suspicious than the control condition (**S7 H1**). This hypothesis was upheld, suggesting that a soft-paternalistic nudge attempting to influence threat appraisal significantly influenced behaviour, compared to control (no nudge). It was however anticipated within this experiment, that the inclusion of a coping strategy (targeting information security self-efficacy) into a threat appraisal nudge would significantly increase nudge success from a threat appraisal nudge only (**S7 H2;** Dupuis et al., 2021) when in fact the direct opposite was found. The threat appraisal only nudge was significantly more effective than a similar nudge that also included a coping statement.

This finding raises a number of ethical concerns in relation to evoking fear without providing a suggested method to reduce this fear. Cavanagh et al., (1981) speaks of three main ethical considerations when contemplating action (a) that you would be happy to receive the same treatment, (b) that humans are not used as a means to an end, (c) the action should be suitable for all people. By adding a coping statement to a fear appeal, the nudge becomes more palatable in relation to all of these ethical considerations, as all people would benefit from the information it entails without evoking fear alone. However, it is unlikely that anyone would have to read a statement looking to evoke fear, with it certainly not suitable for a number of humans suffering from conditions such as anxiety. It is important that a nudge makes clear the benefits of cybersecurity protective action, and that any deception utilised is justified (Dupuis & Renaud, 2021). Such a fear appeal alone would not support the first point, however ensuring the message was proportionate to true risk and not inflated may play to the second. It is possible that the nudge including a coping statement was too long for the participants to want to read its content, simply clicking away from the nudge without taking in its key points. This however needs to be investigated fully in future research. It is also imperative that future research finds better ways of eliciting the same results as a fear appeal nudge, but with a better ethical grounding (this will be investigated within Study 8).

It was also hypothesised that the use of a debiasing technique, developed to increase cybersecurity motivation, would be significantly more effective than the threat appraisal nudge discussed above (**S7 H3;** Menard et al., 2017; Ryan & Deci, 2020). There were no findings to substantiate this hypothesis. However, a nudge that added these motivational factors into the threat appraisal and information security self-efficacy nudge mentioned above (**S7 H4**) was significantly more effective than the control condition, however, this was the case for all debiasing conditions presented. As a lot of text was presented in this debiasing technique particularly, this does suggest that size of intervention content was perhaps not a confounding factor.

Study 8 – Tier 3 Debiasing: Cognitive Strategy Adaptation

A number of debiasing strategies have previously been considered within the decision-making literature with a focus on providing techniques that can be easily remembered and implemented despite cognitive limitations (Larrick, 2004). Soll et al., (2014) suggests that the choice of strategy to be included in debiasing interventions are reflective of experimental research, cognitive limitations experienced, user competence and the frequency and complexity of the decision. Croskerry et al. (2013) posit three groups of debiasing strategies, those more educational, such as the awareness training currently being utilised within organisations, workplace interventions, such as the nudges investigated within Studies 6 and 7 and forcing functions (such as checklists now utilised in aviation and the medical domain where behaviours cannot take place until safety or security checks have been acknowledged).

Two intervention examples of forcing functions strategies are ‘consider-the-opposite’ and a ‘maxim or checklist’. Consider-the-opposite is a strategy whereby people are required to consider what evidence is available for an alternative outcome: for example, viewing all emails as phishing and unearthing clues that confirm that the communication is genuine (Hirt & Markman, 1995). Considering alternative outcomes is supported by the literature as a useful technique for interrupting automatic bias, showing with success that it may encourage more conscious thought. However, its use is yet to be investigated in the cybersecurity domain (Arkes, 1991; Hirt et al., 2004; Mussweiler et al., 2000).

The implementation of mental checklists provides another potential debiasing intervention, using information that can be easily brought to mind through acronyms or maxims. The brain can naturally become accustomed to skill repetition until it becomes ingrained in system one behaviour (Croskerry et al., 2013). Heuristics can be created through similar forms of behaviour repetition, with it possible that replacing these strategies with new forms can help reduce susceptibility to social engineering whilst not experiencing reductions in productivity (Croskerry et al., 2013). An example of a successful maxim used within the carpentry field is ‘measure twice, cut once’, that looks to reduce measurement errors occurring during System one processing. Challenges do however exist around the use of mental checklists within the cybersecurity domain due to the fast-paced changes endured whilst adaptations in technology and offender strategies mature. It is however important to initiate research around the potential success of mental checklists specifically in relation to social engineering whereby the decision-making vulnerabilities discussed in Chapter 2, face the additional challenge of offender manipulation as investigated earlier within this Chapter.

The main aim of Study 8 is to therefore explore the potential use of two cognitive adaptation interventions - consider-the-opposite and a mental checklist - and whether they can assist end-users in the identification of phishing emails as additional decision-making support strategies to the nudging interventions explored within Studies 6 and 7.

Research Hypotheses

S8 H1 Whilst Studies 6 and 7 indicate the usefulness of a soft-paternalistic nudge focused on increasing threat appraisal and reducing unrealistic optimism however clear ethical implications apply, it is hypothesised within the experiment in Study 8, that the use of a debiasing technique believed to influence intuitive decision-making will be significantly more successful than a control condition and an ethical nudge containing a nudge targeting both threat appraisal and information security self-efficacy (Arkes, 1991; Corskerry et al., 2013; Hirt et al., 2004; Mussweiler et al., 2000; Soll et al., 2014).

S8 H1a It was predicted that a debiasing technique utilising a consider-the-opposite strategy would result in significantly more phishing emails being filed as suspicious than the control condition and the threat/optimism/coping nudge.

S8 H1b It was also anticipated that a debiasing technique utilising a maxim would result in significantly more phishing emails being filed as suspicious than the control condition and the threat/optimism/coping nudge.

Method

Participants. Five-hundred and thirty-five participants were recruited using Prolific, of which 47% were male, 52% female, 1% of a different identity with an average age of 37.31 (*SD* 12.21) years. Participants were randomly assigned to one of four experimental groups (control, threat/optimism/coping, consider-the-opposite, maxim) before undertaking an email sorting task. Prolific (at the time of this experiment) pays an hourly rate (£12.05) to its users for completion of questionnaires, so participants were paid accordingly (circa £6 for a 30-minute experiment).

Experimental Design and Procedure. A between-participants experimental design was employed to investigate how three levels of intervention (threat/optimism/coping, consider-the-opposite, maxim) and a control condition, influence participant behaviour in a phishing simulation task (see Table 13 for intervention content). Participants accessed the experiment via Qualtrics© on PCs and tablets before being asked to read a brief introduction sheet, complete a request for consent as well as a demographics form including gender, age and education. Participants followed the same procedure as Experiment, however in Experiment 8, participants were randomly assigned to one of the four experimental conditions detailed and presented with the text of either one of the three interventions or the control condition. Participants, who were prescribed one of the two debiasing interventions, were first provided with instructions around the availability bias and the challenges with humans processing decisions heuristically before information around the alternative strategy. The thirty-two emails used in Study 8 were identical to those used in Study 7 with the same genuine: phishing ratio (2:2), as were the folder locations participants were asked to file the emails. The dependant variable was classified – as in Experiment 7 – as the number of phishing emails correctly filed as suspicious. As in Experiments 6 and 7, also analysed were the number of phishing emails deleted and perhaps assumed to be spam, and ham emails incorrectly filed in the suspicious email folder (false positives).

Table 13

Study 8: Text for Each Intervention Condition

Intervention Type	Text
Control	Thank you for agreeing to take part.

PMT Nudge (threat incl. optimism and coping appeal)	241,324 phishing attacks took place last year, your inbox is at equal risk to others. Always check details such as sender email address and look out for clues such as spelling and grammatical errors.
Consider-the-opposite Intervention	This strategy is called consider-the-opposite, whereby with each email viewed you must consider it phishing and identify three pieces of evidence within its text that convince you the email is genuine. For example, consider the sender email address, the authenticity of links and attachments, and whether the email presents well grammatically.
Maxim Intervention	This strategy is considering and repeating a short maxim called “Who, What and How”. The maxim suggests that before interacting with an email you must first look at who sent it (is their email address legitimate?), what they are asking you to do (are there links or attachments that are not genuine?) and how the email is presented (are there any spelling or grammar errors?). Acknowledging this maxim with each email

	viewed should assist in identifying those that are fraudulent. WHO – WHAT – HOW (repeat five times before proceeding)

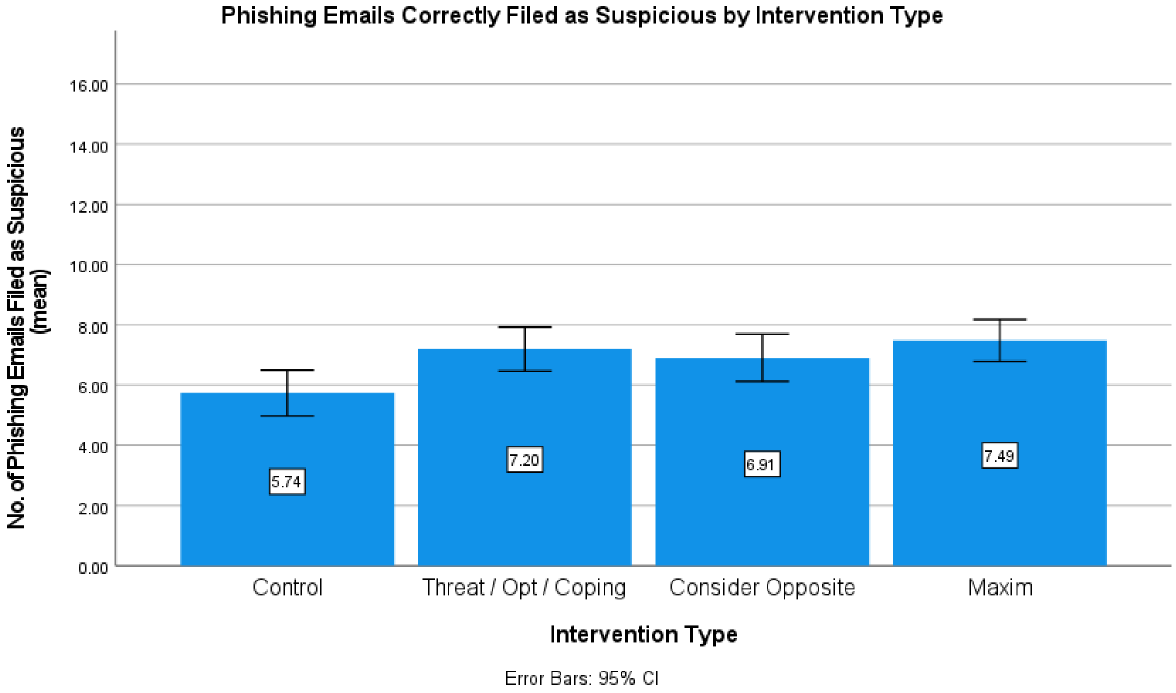
Results

The main aim of Experiment 8 was to investigate whether the use of two potential debiasing interventions – consider-the-opposite, and a mental checklist – could be effective in supporting end-users to better identify phishing emails moving forward, or whether they result in more cognitive burden for little or no return. Any missing observations were replaced with the grand mean value for each question and any outliers determined by 3 IQR from the mean where they were windsorized to the next available value not considered extreme. Missing observations included those emails that were not filed into any of the folders provided.

As data was slightly negatively skewed with a high kurtosis, as well as a significant Levene's test of homogeneity (p = .40) a non-parametric test of ranks was applied. A Kruskal-Wallis test of differences was undertaken to investigate the influence intervention type has on the number of phishing emails participants correctly filed as suspicious across debiasing conditions, adjusted by the Bonferroni correction for multiple tests. A significant difference was found across all four conditions, χ^2, N = 535 =13.759, p < .003 (**E8 H1**; see Figure 12), with a small effect size (ϵ^2 = 0.3). Post-hoc analysis found two conditions (maxim and threat/optimism/coping nudge) out of the three manipulated interventions to be significantly more effective at supporting end-users to correctly file suspicious emails in relation to the control condition. The debiasing strategy intervention containing a maxim produced significant results in relation to a control (χ^2 = -65.261, p > .001; **S8 H1b**), as did the threat/optimism/coping nudge (χ^2 = -56.608, *p* = .021; **S8 H1**). However, the consider-the-opposite intervention did not significantly differ from the control (**E8 H1a**). See Figure 11 for a visual representation of these results (please note that significant differences across interventions were not found).

Figure 12

Phishing Emails Correctly Filed as Suspicious by Intervention Type



The number of genuine emails incorrectly filed as suspicious (false positives) were also analysed by nudge type, with no significant difference found (χ^2 = 4.079, *p* = .253). Also investigated was whether intervention type had a significant effect on the number of phishing emails deleted during the experiment with no difference found (χ^2 = 7.060, p = .070). Though, a significant difference was found with emails filed as both phishing and deleted, χ^2, *N* = 250 = 12.872, *p* = .005, with a small effect size (ϵ^2 = 0.02). Post hoc analysis revealed only the maxim intervention to significantly differ from the control condition when both deleted and reported phishing emails are analysed (p = .003).

Discussion

The key aim of Study 8 was to conduct an experiment that investigated the use of cognitive adaptation debiasing strategies, that hoped to improve participant detection of phishing emails by encouraging the use of new cognitive strategies. It was hypothesised that participant use of a consider-the-opposite strategy, whereby people were required to consider what evidence is available to identify an email as genuine (rather than phishing; Arkes, 1991; Hirt et al., 2004; Mussweiler et al., 2000) or the implementation of maxim that aids skill repetition until it becomes ingrained in system one behaviour (Corskerry et al., 2013) would significantly aid phishing identification. However, of the two cognitive adaptation techniques analysed, it was only the maxim that resulted in significantly more phishing emails be . This is perhaps due to the maxim strategy being easier and less cognitive demanding to deploy than the consider-the-opposite strategy.

The consider-the-opposite strategy informs a top-down approach whereby the email is considered phishing, and the participant needs to then study the detail to see if it fits with the bigger picture. However, the maxim encourages a bottom-up approach whereby details are gathered in relation to the suggested clues and a bigger picture built in relation to whether or not the email is phishing. The challenge with the top-down approach required for the consider-the-opposite strategy, is that the ‘bigger picture’ where the process of the strategy begins will not have perhaps been natural for the participants. The bigger picture representation is ordinarily influenced by a person’s stored knowledge or expectations (Nisbet and Weiss, 2010). As we know from Chapter 2, awareness of a situation can be influenced by many things such as low expectations of risk. Whilst this experimental condition aimed to adapt the representation and expectations around the email to that of initially phishing, this will likely not have been an automatic and natural process for them making such change in strategy challenging. It must also be noted, that whilst the maxim offered a quick and intuitive three stepped approach to phishing identification, the consider-the-opposite strategy was left far more open in so far as how many positive features a participant should search for. Providing participants with a more succinct list of genuine email clues may have motivated them to more regularly attempt this strategy.

The use of a maxim also resulted in significantly more phishing emails being filed as suspicious or deleted over the control condition, yet no significant differences were found between the maxim and the control condition in relation to false positives (genuine emails incorrectly filed as suspicious). This provides additional support for the use of a maxim, in that whilst the debiasing technique may also encourage employees to consider phishing as perhaps spam and therefore delete it, it will not result in ham emails being incorrectly reported as phishing, that would increase the workload of security operation centres (SOCS) and negatively impact the human experience.

As with Studies 6 – 7, the experiment conducted within Study 8 aimed to learn more about the human experience in cybersecurity when being actively persuaded by supportive interventions, perhaps deployed within their organisation in the hope to reduce cybersecurity vulnerability. Whilst these particular debiasing strategies are not yet receiving much focus within the cybersecurity space, it was important to consider whether they could offer benefits to the employee cybersecurity experience, particularly in countering the manipulation tactics employed by cybercriminals to further encourage vulnerability. Despite the benefits such strategies have evidenced in domains such as healthcare, forensic mental health and education (Ludolph & Schulz, 2018; Sellier et al., 2019), far more research is required, before being actively deployed within cybersecurity.

Chapter Discussion

The principal purpose of Studies 4 – 8 were to supplement the work of Studies 1 – 3, by exploring how the employee experience in relation to cybersecurity, is also continually influenced by external influences – particularly in the form of phishing emails, the methods of persuasion used within them, and potential interventions to mitigate their negative effects. In addition to the conscious and unconscious dialogue taking place internally, employees are regularly being persuaded to change or adapt their behaviour, either consciously or unconsciously, in order to increase the success of either offender or defender strategy. The objective of this set of five studies was to therefore understand how this was being achieved, and how successful such external messages are at the persuasion they are trying to induce.

Human Vulnerability Exploitation

With a key focus on Cialdini's (1984) six methods of persuasion (authority, commitment and consistency, liking and similarity, scarcity, social proof and reciprocation) the main aim of Studies 4 and 5 were to gain an up-to-date (as of 2020/2021) understanding of the social engineering techniques being used to influence email recipients that are identified and reported within a multinational corporation (Study 4) and the extent to which such emails are reported, or neglected to be reported, as suspicious under experimental conditions (Study 5 – student participant sample). Similar studies are already becoming dated especially given the fast moving and ever-changing landscape of cybersecurity (Akbar 2014; Ferreira & Lenzini, 2015) allowing research to explore interventions that may help technical solutions and employees better detect social engineering and perhaps the fact that they are being phished.

Findings from both studies provide important insights into the methods currently being used, as well as an indication (all be it experimentally) of the likelihood of resulting in a security breach should they be deployed. Investigating these two aspects alongside each other has offered an appreciation of the recent techniques of concern and will help to underpin further research into the main human decision-making biases being targeted. Based upon the findings, interventions can be constructed centring on the biases that take place during unconscious and instinctive decision-making mode in which phishing emails are finding success. This research also acts as a baseline in which to continue to track these techniques and identify others that will allow organisations to remain one step ahead of offenders moving forward.

One of the most powerful findings from both studies was the prominence and impact of the authority principle and its use within phishing emails. Elements of authority were found to be present in 69% of the reported phishing emails analysed in Study 4, potentially suggesting this to be the most popular weapon of influence for cybercriminals to use – although noting it seems to be the technique identified the most – based on the suspicious reporting nature of this study. In Study 5, 68% of phishing emails were not reported as suspicious – therefore only 32% of these emails were perceived by participants as potentially phishing. Authority, as a technique, can be used in phishing emails via the display of prestigious titles and accolades suggesting to recipients that such expertise can be used as a short-cut to decision-making.

Obedience to authority is a long-standing human heuristic driven by social reinforcement that is used as a rule of thumb to optimal decision-making particularly when there is a lack of knowledge or time (Ghafir et al., 2018). Most humans are trained from a very young age to obey authority in order to maintain communal living, with evidence that humans can be conditioned into performing acts that they

would likely never have considered without authority being present (Milgram, 1974). As children, humans are soft wired to obey the rules of their parents, teachers, law enforcement and so on, learning to trust the view and opinions of experts as instinct. This trust in experts is present even when the topic in question is not the ‘expert’s’ area of expertise. This is due to the general belief that those in authoritative positions make better decisions that will result in fewer mistakes (Hinnosaar and Hinnosaar, 2012). The authority bias described takes place prior to conscious thought and can be evoked by the subtlest of clues such as a smart suit or lab coat, an impressive business title or advertised awards and accolades. Authority can be threaded through emails by imitating known companies, high-level business titles or company credentials. The evolutionary and instinctive composition of the authority bias creates a huge challenge for intervention and explains in part why current training focused on conscious thought can have little impact.

Another persuasion technique found to be potentially potent is the scarcity principle, the third most pervasive tactic within reported emails in Study 4 (22%), and the second most likely to result in a clicked link or an opened attachment in Study 5 (63%). The scarcity principle involves the suggestion that something is restricted in either quantity or time, such as a ‘limited time offer’ that elicits a sense of urgency moving people into quick intuitive decision-making. Humans are fearful of losing freedom of choice so the suggestion of such restrictions will result in a sense of urgency in the need to regain this freedom and therefore autonomy (Aggarwalm et al., 2011). The scarcity principle relies on the commodity bias, whereby humans value objects or experiences based on their availability, with value increasing as supply, or time to access the supply decreases (Brock, 1968). A lack of availability hints at high demand leaving humans to believe that the object or event is perceived as of value to others (Aggarwal, 2011). As with authority, the scarcity bias provides a decision-making shortcut that is experienced outside of conscious cognition demanding intervention that supports humans during the unconscious decision-making stage driving them, where required, into more conscious thought.

Across both studies in Chapter 2, the curiosity principle was identified as a technique of concern, ranked as the second most likely to be deployed within 37% of reported emails in Study 4. Despite this finding, curiosity was one of the easiest tactics to detect during Study 5, with breach success in around 33% of emails (noting that 33% is still very high, however some of the other techniques had higher success rates with authority at almost 70%). The disparity between likelihood to be deployed, and probability to succeed is potentially due to the ease in which curiosity emails can be generated, allowing for a blanket approach that could result in similar numbers of breaches as those smaller and more tailored campaigns due purely to numbers. As mentioned previously, curiosity is the human want to learn more about a subject due to its novelty, complexity, or incomplete information. It evolved within humans to aid survival whereby uncertainty in the world would evoke anxiety, and curiosity would help alleviate this by the human investigating the stimuli (Shin & Kim, 2019). It is therefore experienced by end-users as recognition of a gap in knowledge that reduces self-control and increases impulsivity in order to bridge this knowledge gap (Loewenstein, 1994).

During email analysis in Study 4, curiosity was presented in a number of ways, such as a simple link or perhaps newspaper bulletins promoting links to ‘find out more’. Emails utilised in Study 5 covered a range of these curiosity techniques. In order to evoke curiosity, offenders need to find the sweet spot between providing enough information to pique interest, but not too much as to complicate and reduce motivation. It is possible this intermediate position has been found with a potential increase in shared folders across the cloud landing in email boxes. Curiosity is a long-understood exploitation technique used by social engineers to motivate their victims to act but is yet to be added to Cialdini’s (1984) original six methods of persuasion by other researchers in the field, resulting in a general lack of understanding in relation to both its prevalence and success rate. This research therefore presents curiosity as the seventh method of persuasion/weapon of influence and a technique that must be tracked and investigated, in terms of emails reported, as well as investigations into interventions that can support people in not falling victim to it. Current available research, in relation to curiosity, is focused on *increasing* its level, due to its positive association with areas such as learning, creativity and relationship development (Kashdan & Silva, 2009; Schutte & Malouff, 2020). Research around *reducing* curiosity is therefore sparse, with the added challenge of reductions in curiosity needing to remain domain-specific in relation to cybersecurity only. Further research is therefore required within this very particular context.

Commitment and consistency, another principle pinpointed as a potential tactic of concern (the third most successful method used in Study 5; 49% of all emails), was however only present in 13% of the phishing emails reported in Study 4, ranking sixth. Inconsistency is perceived in society as a negative personality trait, and therefore a social norm is in place for humans to remain consistent with their prior commitments i.e., behaviours, beliefs or even characterisations placed upon them (Cialdini & Trost, 1998). For example, should an end-user be labelled as a ‘good customer’ in a phishing email (despite no previous contact) many recipients will feel compelled to try and become that ‘good customer’ in order to reduce cognitive dissonance. Cognitive dissonance is a state in which humans feel discomfort when two cognitive processes do not match, with an attempt to reduce this unease by bringing the disparate thoughts, beliefs or attitudes back in line. Lower levels of success rate and level of utilisation of this technique does suggest a potential lack of understanding in the appreciation of this bias as a weapon of influence, perhaps due to perceptions of it being difficult to achieve in a single communication. It must be noted that as Study 5 was a simulated task it is possible that participants assumed genuine pre-commitment to email senders inflating its success.

Outside of the four methods of persuasion previously discussed, three other techniques were identified as lower in ranking both in prevalence and level of success. Reciprocity, the fifth most utilised persuasion technique analysed in Study 4 (14%) and the seventh most successful in Study 5 (33%) is the behavioural response to an act of kindness whereby humans will feel indebted to return the favour. Its relatively low ranking is possibly due to difficulties in convincing (especially compared to times past) recipients of a genuine enough gift and motive during a one-off electronic communication. Liking and similarity was the seventh most prevalent persuasion technique analysed in Study 4 (10%) and the fifth most successful in Study 5 (47%), with the similarity and liking principle suggested as most likely to convince participants to respond positively to the requests of someone who shares similarities to them or is likable through attractiveness or paying people compliments. An email attempting to build rapport or offer praise will have more success from an in-group member than those sent by an apparent out-group which is challenging for an offender to determine. Again, it is possible that rapport or praise appears less genuine during a non-face to face single communication.

Social proof was found to be the least prevalent persuasion technique in Study 4 (8%), as well as the least successful within Study 5 (32%); a finding supported by previous literature. Social proof relies on heuristics such as herding effects where humans will follow decisions that fall in line with those of their group or peers, offering an unconscious short-cut to decision-making. Social proof is utilised in a lot of advertising campaigns such as customer testimonials and reviews and may possibly be assumed to be a marketing appeal and therefore spam. It is again important to note that those methods suggested as least important still found success in around a third to a half of emails within Study 5, so some of the discussion points above are based on relative differences when actually these percentages are very worrying in terms of how susceptible the Study 5 sample seemed to be. These methods may be even more successful when combined with other forms of decision-making influence, with ‘mixed methods of persuasion’ in phishing emails not uncommon.

Another interesting finding across both studies was the moderate reporting (20%) and moderate success under experimental conditions (49%) of emails containing no single method of persuasion or possibly no currently known method. It is possible that due to the experimental situation, over reporting may have occurred with participants less sure whether bland emails were innocuous. It is also imperative to note, that in the emails analysed more than one technique may have been present with it impossible to verify which of these methods may have led to detection. Future research must focus on the experience of the employee at the time of email interaction to improve understanding around what made them choose to report, something currently being undertaken by the author of this thesis and academic collaborators. These future investigations will not only look at the potentially unconscious decision-making processes (eye tracking) being undertaken by end-users but also qualitative analysis around their conscious experience, helping set the scene around the human experience during phishing interaction.

Whilst interventions aimed at protecting employees from the threat of cyber-attacks is important, the human cognitive constraints previously mentioned (decision-making biases that have been developed across evolution and a human’s lifespan as well as those manipulated by cybercriminals) suggests that wherever possible technical intervention should be devised to reduce the number of phishing emails reaching an employee’s inbox. Email filtering software and associated algorithms should be updated to ‘flag’ cues within the text that suggest authority, scarcity, curiosity and so on may possibly being evoked – that these emails are either filtered, depending on its contents, or at least highlighted that elements of persuasion are present, for example, the word ‘urgent’ underlined to encourage an employee to double check and verify that the email is genuine before actioning it (e.g., opening an attachment, clicking on a link, replying). Interventions targeting such behaviours will be further examined in Chapter Four of this thesis.

Despite technical interventions, employees continue to receive malicious emails. This ultimately requires improved ways to support the identification of phishing emails moving forward, with a number of key interventions deemed as useful in order to better guide human cybersecurity decision-making (Arkes, 1991; Croskerry et al., 2013; Larrick, 2004; Soll et al., 2014). Interventions may include improving motivation to encourage employees to adhere to suggested cybersecurity behaviours (Larrick, 2004; Soll et al., 2014), soft-paternalistic nudging, modifying the decision-making environment guiding more secure decisions (Furnell et al., 2019; Petelka et al., 2019; Turland et al., 2015). As an example, a computer pop-up that highlights the presence of urgency in an email and the need to choose ‘yes’ or ‘no’ to confirm the enclosed link is safe.

Another suggested technique is cognitive debiasing strategies (or modifying the user) with interventions that help end-users understand the biases being manipulated by cybercriminals e.g., the authority bias, recognise it at its intuitive stage and then apply an alternative strategy (Croskerry et al., 2013). A number of different debiasing strategies can be used such as ‘consider-the-opposite’ where employees are asked to assume all emails are phishing and detect clues to suggest an email is genuine, or a maxim such as that used in the carpentry domain of ‘measure twice cut once’ to avoid error. An example within cybersecurity could be ‘External, Unknown, Link’, attempting to create a simplistic tick box exercise for emails from an external source, an unknown contact, and a suspicious link.

The reporting of 49% of emails containing no persuasion techniques within Study 5 is of high interest. Findings suggest that in experimental circumstances, perhaps participants view such non-offensive emails as a 50/50 option of suspicious or not. It is possibly due to the presence of persuasive tactics not yet understood or ancillary techniques such as high amounts of SPAG. Therefore, whilst interventions should focus on those methods most prevalent and successful, intervention also needs to address more generally the ease in which humans can sometime succumb in lieu of these efforts.

Within Study 4, several peripheral clues were highlighted including the presence of a link in 71% of phishing emails, an image in 41%, and SPAG in 44%. In Study 5 all emails were designed without images, all inclusive of SPAG errors and 50% with links and 50% with attachments to ensure a fair experimental control base. These findings suggest the need to also educate end-users on how to check the authenticity of a link, the potential need for spelling and grammar checks to be available on emails received to highlight any errors as well as potentially blocking (or indeed highlighting for inspection) images from external sources. Images are known to better attract attention than text alone with the ability to drive people into action through emotional appeals, as well as conceal malicious links (Matz et al., 2019). Several new techniques were also identified that should be further investigated such as ‘widening the web’ where recipients are asked to forward emails and therefore evoke social proof, multiple points of threat entry within one email (e.g., lots of links) and the use of fabricated email chains evoking either social proof or commitment and consistency principles.

To summarise, authority, curiosity and scarcity were found to be the most concerning methods of persuasion used in phishing emails in Studies 4 and 5, suggesting an urgent need for intervention to focus on detecting emails where the sender positions themselves as a person of power, where incomplete information is available within the email and where a sense of urgency is attempting to be evoked. There should also remain a focus on both technical solutions and employee interventions that help identify ancillary clues such as malicious links, persuasive imagery containing hidden links and spelling and grammar errors that may detect phishing when no method of persuasion is being utilised.

Human Vulnerability Mitigation

Whilst many organisations now employ technical solutions to limit the number of phishing emails landing in end-user inboxes, e.g., email filtering, the number of phishing emails received by employees has not appeared to significantly reduce (Verizon 2022). Whilst education is an important aspect of behaviour change it has not been enough to result in reductions in cyber-breaches with other strategies required to move closer towards mitigation (Aldawood et al., 2019; Alshaikh et al., 2018; April 2018; Bada et al., 2019; Scholl et al., 2018; Skinner at al., 2018). The current set of studies assesses the efficacy of three interventions that are being utilised within organisations in the hope to better support human decision-making or are being discussed or starting to be discussed within the literature in particular relation to phishing.

One-thousand three-hundred and ten participants were recruited to file a selection of genuine and phishing emails into several inbox folders with the aim of filing phishing emails into the folder named ‘suspicious’. In each of the three experimental studies, a different debiasing intervention became the focus (soft-paternalistic nudging, motivation, and strategy modification) in order to understand the influence external debiasing strategies can have on the employee cybersecurity experience and resulting behaviour (Croskerry et al., 2013; Thaler & Sunstein, 2008). Findings from the three studies suggest several potentially quick, relatively cheap but effective interventions that can be used by organisations to support the identification of phishing emails, such as a nudge targeting the threat appraisal factor found within the Cybersecurity Awareness Framework (CAF). Although most interventions found to be effective within Studies 6 – 8 where only more effective than a control condition where no intervention is applied. It is perhaps easy for any intervention reminding employees to remain alert to phishing emails, to result in better identified than no such communication. Future research must investigate their effect across time, how soon employees become desensitised to their content (Petelka et al., 2019), with the need for both content and context to remain dynamic so that employees continue to digest their message. Habits can then be recalled and executed by end-users when in a more unconscious mode of thought providing a stronger defence against offenders hoping their malicious communication will remain undetected. This research offers investigations into how several interventions outside of awareness training might look to support human decision-making when opening emails, in order to mitigate cyber-breaches as a result of phishing should more academic support be applied.

Debiasing, a term used to describe processes by which humans are supported to reduce usual violations from rational thought, is a form of intervention with possible potential. Fischhoff (1982) categorised two forms of debiasing: modification of the decision-maker, and modification of the environment. The first assumes that bias resides in the person and therefore tools and training are needed to reduce decision-making errors within them. The latter suggests bias resides in the environment with alteration of context being the key to bias reduction. Previous research outside of the cybersecurity domain suggests that careful modification of the decision-maker through bias training can have some success (Morewedge et al. 2015), with research both within and outside of the cybersecurity domain suggesting success in modifying the environment via soft-paternalistic nudging can potentially be successful (Brigg et al., 2017; Jeske et al., 2014; Petelka et al., 2019; Turland et al., 2015). A number of interventions suggested as influential in behaviour change could also be useful in increasing cybersecurity threat appraisal in organisations, and therefore improve cybersecurity behaviour. The novelty of Experiments 6 – 8 is therefore improving understanding around how external influence can help support human decision-making vulnerabilities in relation to cybersecurity – potentially through real-time nudging, adaptation of cognitive strategies such as use of a maxim or mental checklist. However, the question remains unclear around whether in fact the burden caused by such interventions e.g., impact to productivity when need to make decision more regularly in conscious mode also having substantial costs related to it.

Chapter 3 has helped identify the persuasion methods most likely to be reported and perhaps received (if reporting is linked to the numbers of emails received) by employees (Study 4) as well as those most likely to result in a security breach (Study 5). It is important, from these findings, to allocate ways to support end-users in identifying these techniques as well as the strategies required to help avoid their influence. Current cybersecurity training and awareness programmes (in general – including those focused on phishing attempts and methods of persuasion) are not working sufficiently, largely due to a lack of appreciation of the number of decisions made intuitively, particularly when under cognitive strain (Bada et al., 2019; Caraban et al., 2019; Scholl et al., 2018). End-users are not only driven to heuristic decision-making naturally, but offenders also actively use the techniques discussed within this chapter to further prevent conscious thought from being applied reducing deception detection. Interventions therefore need to focus on supporting humans whilst using heuristic and more default automatic modes of thought, identifying key biases of concern and identifying ways in which to educate end-users on strategies to help trigger suspicion (Croskerry, 2013). Chapter 2 within this thesis also highlighted a number of decision-making biases experienced by employees in relation to cybersecurity, with outcomes from both chapters supporting the importance of analysing human decision-making at its root cause. Chapter 4 therefore focuses on interventions that can be potentially used to support human decision-making biases both naturally occurring and those manipulated by cybercriminals.

Study 6 investigated a number of soft-paternalistic nudges focused on targeting the availability and optimism biases in order to increase human appraisal of threat and improve the identification of phishing emails. It was found that participants were more likely to file phishing emails as suspicious if they received a nudge over no nudge at all, and particularly if that nudge targeted the availability and optimism biases together (targeting threat appraisal within the Cybersecurity Awareness Framework (CAF). Study 7 investigated whether the inclusion of the antecedents of self-determination theory (SDT; competence, autonomy and relatedness) into a motivation nudge would further encourage receivers of a soft-paternalistic nudge above and beyond what a nudge targeting threat/optimism could offer. The motivation text alone offered no significant difference to the control, however when linked with the threat appraisal and information security self-efficacy nudge, it became significantly more useful in phishing detection against the control. Finally, Study 8 investigated the use of two debiasing techniques that attempt to adapt the cognitive strategies used (consider-the-opposite and a maxim) in helping end-users identify malicious emails. Of the two debiasing strategies investigated, the maxim was the only technique found to result in significantly more phishing emails identified than no intervention. However, the use of a threat/optimism/coping nudge also resulted in more phishing emails being filed as suspicious.

These Experiments continued investigations into the use of several techniques to support human decision-making whilst in a more intuitive mode of thought. Recent interventions tend to utilise educational training that require end-users to switch to conscious mode and actively apply their learnings whilst busy at work, multi-tasking or when being socially engineered. Findings from the five studies within this chapter suggest a number of potentially quick and cheap debiasing techniques that can be used by organisations to support the identification of the phishing emails investigated within Chapter 3: Human Vulnerability Exploitation.

The primary aim of Chapter 3, was to improve understanding around the employee experience in relation to continual external persuasion, and the extra layer of vulnerability this brings to every cybersecurity encounter. Investigating only the internal vulnerabilities of the human, does not paint a complete picture of the numerous aspects influencing behaviour, and the challenges this creates when relying on employees to shield organisations from cyber-threat. With a more transparent view of the human cybersecurity experience, interventions can be generated that are better tailored to employee vulnerabilities, holistically, in the hope to make genuine strides towards cyber-attack mitigation.

Chapter Five: General Discussion and Future Directions

Despite the benefits to society technology and the internet has yielded, it has also provided the opportunity for anonymous and often remote cybercriminals to gain access to end-user finances, technology and personal data. In 2021, over 79,635 security incidents were reported to have taken place across the globe, with the human left responsible for the success or failure of a large number of these attacks (Verizon, 2022). Whilst several technical interventions have been deployed to better support the human, such as email filtering, cybercriminals continue to find ways in which to bypass these technical efforts, leaving the human ultimately responsible for the outcome of an attack.

Over recent years, many organisations have become aware of the important role employees play in the fight against cyber-attacks, now often applying time and budget to human-related control mechanisms, in the hope to better protect companies from the devastation cyber-attacks often cause. However, despite these largely educational interventions, cyber-attacks involving the human element do not appear to have resulted in significant mitigation, with organisations unclear on what they can do next to improve the current situation (Verizon, 2021; Verizon 2022). In order to advise organisations around how to better target intervention, it is important to first improve understanding around the employee experience in cybersecurity, their vulnerabilities and challenges. This information can then help build a picture around what employees are actually capable of achieving in relation to cybersecurity, what is fair to ask of them, and whether they actually want to be actively involved. By better understanding the human factors of most import, tools can be devised to support organisations in the measuring of cybersecurity awareness, and interventions tailored to these vulnerabilities in order to produce genuine results.

The key aim of this PhD was to therefore bring together a broad range of psychological, sociological and behavioural economics research (including studies, reviews, position pieces, models, and theories) that have the potential to aid understanding around the human experience in cybersecurity, not only considering their internal experiences, but also the added cognitive pressure placed on them through both malicious and sympathetic attempts to influence this experience externally. After a comprehensive literature review, the first empirical block (Chapter 2) within this thesis, had a main objective of drilling into the underlying internal vulnerabilities of the human, to provide an improved understanding around the multitude of individual differences and perceptions that interact with and ultimately define the human cybersecurity experience and the challenges it may bring. A key output being a cybersecurity awareness framework and associated measurement tool that can provide organisations with a more transparent view of how their employees are encountering cybersecurity - the Cybersecurity Awareness Framework (CAF), and the associated tool - the Cybersecurity Awareness Measure (CAT) that together can help inform organisations on where risk to cybersecurity awareness exists.

Empirical block 2 changed tack by focusing instead on the external influences of the human experience, deployed by both cybercriminals and researchers and awareness leads offering a number of interventions hoping to support vulnerabilities. Cybercriminals have been utilising social engineering strategies, within emails, to successfully target humans for many decades (Phish Protection, 2021). Without an improved understanding around the persuasion techniques used, and how they are influencing end-users, it will be impossible to take strides towards migration. Similarly, organisations, academics and vendors are working hard to generate cybersecurity interventions that have the ability to counter such manipulation, also utilising techniques with the ability to persuade end-users but with a more sympathetic motive. It was therefore also important to gain an understand the benefits such interventions can bring, or whether they are adding an additional layer of vulnerability for little gain. The principal aim being, to utilise findings from both empirical blocks to understand the continual cognitive push and pull humans are experiencing in cybersecurity, what is working and what is not, and how this can impact cybersecurity awareness.

Empirical Block 1: Human Cybersecurity Vulnerabilities

First explored, were the human factors informing the employee experience in cybersecurity, and how they generate employee vulnerability. With such a wide number of behavioural change theories and individual differences under recent empirical investigation, it has become problematic for organisations to understand where to even begin in relation to measuring human risk in cybersecurity, and in turn, how to successfully intervene. Most organisations do not have specialists (e.g., psychologists, human factors experts) embedded within their company to work with them on trying to improve human-centred cybersecurity. The Open Systems Interconnection Model (OSI) framework has been generated to help better understand and categorise types of cyber-attacks when it comes to the architecture of data communications for networked computers, helping define technical vulnerabilities across seven layers - physical, data link, network, transport, session, presentation and application (Mughal, 2020). This has been, in some way, extended by the Pedagogical Cybersecurity Framework (PCF; Swire, 2018) to also include an eight layer – the human, in order to begin consideration into the needs of the user (perception, cognition, memory) and how they interact with the hardware and software during attack (user interface; Bauer & Patrick, 2004). The PCF considers all players in cybersecurity, from end-users to operators across modes such as training, policies and information sharing (Aloseel et al., 2020). However, despite the PCF stating the importance of ‘human vulnerability management’ in the eighth layer, it does not detail how that framework might look. Studies 1 -3 hoped to help bridge that gap by providing a framework and tool that can help organisations measure and manage human vulnerabilities, but with the employee experience in mind.

As research continues to evolve in relation to the key factors influencing human cybersecurity behaviour, organisations require some guidance today on what they can do to better protect their employees and business from the often-devastating impact of a cyber-attack. It was therefore important to bring together the vast number of socio-psychological, habitual, perceptual and socio-emotional factors previously found to relate to cybersecurity behaviour and investigate those most influential within one set of studies. Once factors presenting the most influence over cybersecurity behaviour have been identified, and the human-decision making biases at the root of the issue understood, work can begin on developing interventions that are far more targeted to the specific needs of the human.

To understand where intervention should be focused, especially over the longer-term, Studies 1 - 3 were conducted to help create an assessment framework and set of associated metrics that could be used by organisations to measure and manage cybersecurity vulnerability, in specific relation to awareness. A literature review was conducted to identify the large number of psychological models and individual differences that required exploration. First it was anticipated that significant differences would be unearthed in relation to both age and gender, with those younger (18-24), and female more likely to report less secure behaviour (Branley-Bell et al., 2022; Gratian et al, 2018; Parrish et al., 2017; Whitty et al., 2015). Gender was not found to significantly differ in relation to reported cybersecurity behaviour in any of the three studies, however age did in two of the studies, with those younger found to be at more risk. However, age was not found to significantly predict reported behaviour within regression analyses and therefore does not form part of the framework. Despite this, research must continue to determine more about the relationship between age and cybersecurity, and whether it moderates a number of predictive factors.

The assessment framework was then explored , spanning across a large number of factors including several socio-psychological constructs (level of IT skill, level of cybersecurity training, perception of importance of role in cybersecurity, personality, risk-taking preferences, decision-making styles, impulsivity, acceptance of the internet, information security attitude), perceptual attributes (threat appraisal, information security self-efficacy, subjective norms, response efficacy, response costs, information security awareness, information security organisation policy), a habitual factor (information security experience and involvement) and socio-emotional factors (intrinsic and extrinsic maladaptive rewards, organisational commitment, psychological ownership). Analysis of this iterative framework took place across three studies involving five-hundred and fifty-three participants, utilising correlational analyses, an exploratory factor analysis and regression analyses to identify and develop a Cybersecurity Awareness Framework (CAF). The over-arching latent factor that informs this framework, cybersecurity awareness, is conceptualised as - *the collective experiences employees hold when approaching cybersecurity, and how it impacts their protective behaviours* (Gafoor, 2012; Marton, 2000; Travethan, 2017). Six observed factors feed into this framework including - threat appraisal, information security self-efficacy, information security awareness, information security attitude, information security operation policy, information security experience and involvement.

Threat appraisal refers to how an employee perceives the probability and potential severity of a cyber-attack, with higher probability and severity resulting in more conscious behaviour (McGill and Thompson, 2017). Threat appraisal features as an important factor in most behaviour change theories, with regular attempts to manipulate it within interventions through fear appeals and was therefore also included within empirical block two when considering external manipulation of the human experience. It is believed to be the availability bias that informs this factor, assisting humans with a quick calculation on the probability of risk, by adding up the number of relevant instances of cyber-attacks held within the mind (Taylor-Gooby & Zinn, 2006; Tversky & Kahneman, 1973). The more instances cognitively available, the higher the perceived probability of an attack taking place, resulting in higher motivation to act to reduce this probability. Should an organisation identify threat appraisal as low amongst their employees, they can attempt to actively influence it by providing regular and salient updates around recent cyber-incidents. The external manipulation of threat appraisal (with optimism) was investigated most notably within Study 6 of this thesis, finding its positive manipulation helping improve behaviour beyond a control condition. There are a number of apparent concerns when considering external threat appraisal persuasion (a) Supplying employees with additional details of security incidents will place even more cognitive strain upon their experience, for perhaps little benefit, (b) There are ethical implications for increasing concerns around risk perhaps inducing anxiety, (c) Humans, may choose to not fully digest the information they receive anyway, tending to particularly avoid information in relation to negative events (Sunstein, 2020). It may therefore be more practical and ethical to subtly prime such a heuristic, possibly through the use of an alert vibration on an employee smartwatch each time an attack occurs. Smart nudges delivered through biotechnology can be useful for cybersecurity awareness more generally, buy providing reminders, updates and more, in real-time promoting very quick behaviour adaptation (Mele, 2021).

Information Security Self-efficacy is another factor underlying cybersecurity awareness, referring to the skills and capabilities a person believes are required to bring about a certain course of action, and whether they perceive themselves as capable of those skills (Maddux & Gosselin, 2012). High self-efficacy can be achieved in a number of ways, such as self-mastery in a skill, witnessing others achieve it, commendation of achievement of the skill by peers, and affective physical feedback (Maddux & Gosselin, 2012; Ryan & Deci, 2020). Humans ordinarily judge ability in two ways, by improvements in self-ability (self-referenced) and in relation to the ability of others (other referenced), with the latter believed to be the most useful (Nicholls 1984). Self-efficacy, amongst other factors within the Cybersecurity Awareness Framework (CAF) can potentially be manipulated through the creation of serious games (the use of a game to encourage education rather than simply fun), and gamification of these games (the application of points, awards, leaderboards and more to encourage engagement). Gamification has more recently been identified as particularly helpful in improving self-efficacy but can also benefit information security experience and involvement also featured within the framework (Scholefield & Shepherd, 2019; Steen & Deeleman, 2021).

First, in relation to the use of serious games to improve the employee experience, outside of the more simplistic and easier to deploy interventions such as the debiasing techniques investigated within this thesis. Games are structured forms of play, that are usually undertaken by humans for the purpose of fun. They can take many forms, but most popular applications include physical sports such as football, boardgames such as Chess, and online gaming such as the currently popular *FORTNITE*. The online gaming industry is a huge market, with around 40% of the total world population online gamers, and 88% of young adults immersed in the online gaming world (Uswitch, 2023). Games can however be utilised for more than just pleasure, they have found great use as an educational tool, particularly within healthcare (Amab, 2013; Gamberini et al., 2008, Ma & Zheng, 2011), with this particular variety termed serious games.

In more recent years, the cybersecurity industry have taken an interest in serious games, in attempts to train both end-users and security specialists on anything from cryptography to phishing. A number of examples include Anti phishing Phil – a mobile application used to educate on the identification of malicious links, CyberCEIGE training within a 3D virtual world, and Control – Alt- Hack a puzzle card and board game, all with positive results (Hendrix et al., 2016). The aim of such games are to educate on one or more cybersecurity skills, for example software updates, avoiding unknown and untrusted networks and identifying social engineering. They are believed to result in feelings of cybersecurity skill mastery, and at a must faster pace (Hart et al., 2020). Serious games utilise a number of mechanics to increase player ability in cybersecurity skill training, including educational instructions, observation, strategy planning, response simulation, time pressures, feedback, action reflection, level advancement and so on (Kulshrestha et al., 2021). The purpose of these mechanics are to help users consolidate the many elements of cybersecurity behaviours in one space, allowing users to learn naturally through experimentation (Salazar et al., 2013).

Previous research details the lack of success in current security education, training and awareness programmes, perhaps due to a lack in theoretical basis in their choice of delivery (Alshaikh, 2019) as well as an absence of appreciation around how humans make decisions (Larrick, 2004). Studies 6 – 8 of this thesis therefore investigated a number of low cost interventions that are being used to directly target human vulnerabilities in decision-making during intuitive thought. Investigations into these interventions revealed great promise, indicating that these easily deployable control measures can be useful at highlighting to employees that more conscious thought is required. However, there are concerns there are concerns around the maintenance of their effect across time, with employees perhaps becoming desensitized to a number of them unless continually reminded or their context changed e.g., colour or position (Petelka et al., 2019; Shah et al., 2021). Serious games can perhaps support the interventions explored, by allowing employees to practice e.g., identifying phishing emails under the intervention conditions, until the desired behaviours become automatic. Augmented reality, combining real and computer generated worlds, have been found to further increase the benefits of serious games as an intervention by fully immersing participants in the experience (Salazar et al., 2013). There will however be a cost to utilising such technical interventions that perhaps make them difficult for smaller organisations to deploy.

Gamification is a complementary yet different concept, whereby the mechanics of a game are again applied to non-gaming contexts, but with a focus on engagement and increasing motivation to interact with the game (Scholefield & Shepard, 2019). A common example of gamification is seen within the mobile application *Duolingo©*, that uses mechanics such as badges, levels, leaderboards, virtual currency, awards, progress bars and challenges to provide language education. Gamification in cybersecurity awareness intervention has been found to increase self-efficacy and perception around ability and experience, by supporting user perception of both self-referenced ability (progress bars) and other-referenced ability (leaderboards; Scholefield & Shepard, 2019). Gamification mechanics have also been found to increase attitudes towards cybersecurity and intentions and behaviour directly (van Steen & Deeleman, 2021). In fact, gamification mechanics applied to a serious game is a particularly powerful blend when used to increase cybersecurity awareness (Weitl-Harms et al., 2023). The use of serious games and gamification in cybersecurity awareness is a topic that continues to receive research focus, with positive results found (Barendse, 2023; Batzos et al., 2023; Troja, 2023).

Also informing cybersecurity awareness, is *information security awareness*: employee perceptions around their ability to remain up to date around current risks to online data and what needs to be done to better protect them from those risks. High information security awareness can be obtained through regular communication, a knowledge sharing culture and cross-company collaboration (Safa et al., 2015; Zwilling et al., 2022). It needs to be simple and obvious to employees how they can keep their knowledge current, this can be accomplished by providing employees with access to an online community for easy knowledge collaboration. Such an online community would not only help increase perceptions around awareness, but also help induce generative responses that could inform intervention in relation to risky shadow workarounds (Faraj et al., 2011).

The integration of an online community either within an e-learning platform, policy software, serious games or as an application in its own right, can assist organisations in taking full advantage of employee knowledge, whilst further supporting vulnerability reduction by i.e., increasing feelings of involvement. An online community is ordinarily developed as a way to help construct, compare and share knowledge (De Laat, 2023). Such crowdsourced activities have resulted in knowledge creation across a number of opensource platforms, such as Wikipedia. As in all platforms of this kind, there is an issue with policing content, particularly in relation to negative information and therefore this requiring further attention (Altman et al., 2019; Kretschmer et al., 2022; Nickerson et al., 2017). An online community has been found to be particularly useful within a gamified platform, as the competitive atmosphere created results not only in higher numbers of contributors, but also higher motivation from those contributors to collaborate (Loh and Kretschmer, 2023). Online communities are so successful due to the power of social dynamics. Social networks have become more and integral to everyday life, with the ability to share knowledge never easier. Carley (2020) discusses the importance of applying the same processes to benefit cybersecurity, with the emerging science of ‘social cybersecurity’ requiring far more attention. Online communities can be used not only to explicitly share knowledge, but to also increase threat appraisal, improve employee perceptions of involvement, and help better shape policy.

Computer supported cooperative work (CSCW) or collaborative virtual environments (CVE) are examples of online communities with the very specific aim of sharing knowledge. They focus on supporting communication and collaboration but within an online space (Grundin & Poltrock, 2012). Due to its sensitive nature, cybersecurity has often remained a very solitary activity, with end-users, security operation centres (SOCs) and organisations working in isolation, and without larger collaboration towards defending attack. Work is now being conducted to help crowdsource in relation to data and information acquired by SOCs (see for example, Kabil et al., 2018), but not yet from an end-user, social cybersecurity perspective. Cybersecurity collaboration within organisations is absolutely imperative, should perceptions towards cybersecurity wish to be changed. By deploying them within a gamified environment, not only will it be for employees easier to knowledge share, but the competitive atmosphere will actually increase collaboration mentality.

Information security experience and involvement acknowledges the importance of perceptions around interactions with cybersecurity in the past, and how they influence how employees choose to interact with cybersecurity (Safa et al., 2015). If employees do not feel they have previously been involved in cybersecurity and that this involvement was positive, they will not see value in future interactions, evoking the bystander effect whereby inaction will result in the belief that others will use their knowledge to protect security, e.g., the IT department (Garcia et al., 2002). By involving employees in the creation and adaptation of cybersecurity policy, perhaps again through an online community, the IKEA effect will occur resulting in employees that place higher value on the policy and culture they have spent time helping shape (Franke et al., 2010; Norton et al., 2012). However, the question must be asked, what is actually reasonable for organisations to expect from their employees, with something technical and outside of their employment remit. Would they be expected to exert equivalent efforts into aspects of physical security. Are expectations proportional to the amount of protection it personally provides employees, for example they wouldn’t be expected to spend hours monitoring the security of a building that they do not own (Strawser & Joy Jr, 2015). Therefore, whilst it is important that employees remain involved in cybersecurity, it is important that that involvement is proportionate and considers their overall experience.

Information security attitude refers to the way in which an employee has evaluated cybersecurity, based on their feelings, beliefs and emotions towards it (Maio & Haddock, 2007). Attitudes exist to help guide behaviour and to simplify reasoning around how to act (Maio & Haddock, 2007). It is therefore important that employees have a positive attitude towards cybersecurity and why it is needed. The attitudes that employees will experience in relation to cybersecurity may not always be logical, and in fact may be completely unrelated such as affective warmth towards something solely because the sun is shining (Bohner & Dickel, 2011). Multiple attitudes around cybersecurity can also exist, and be conflicting, with a mental calculation required to sum up these feelings and ultimately decide on the correct attitude valance (Maio & Haddock, 2007). Attitudes can also be implicit or explicit, with no current agreement on how and when attitudes are held in memory, or when they are constructed on the spot (Bohner & Dickel, 2011; Gawronski, 2007). What is known is that attitudes can be difficult to change due to humans constantly searching out confirmatory information, feeling uncomfortable when considering any belief that may differ from the attitude they currently hold (Bohner & Dickel, 2011). Persuasion has however been found to be useful in encouraging attitude change, either negatively as found within phishing emails or more positively within debiasing (Bada et al., 2019). It is however important to note that even when successful persuasion has occurred, old attitudes remain stored at the back of the mind, almost as cognitive residue (Bohner, Dickel, 2011). It is perhaps again, the social aspect that will support the largest change in cybersecurity attitude, with people feeling more connected to others when they hold the same attitudes towards something (Albarracin & Shavitt, 2018). A supportive online community that fosters positive discourse in relation to cybersecurity could have the largest influence on cybersecurity attitude within organisations.

Finally. *Information Security Operation Policy* relates to the perceptions around the policies and procedures that governments, compliance agencies and organisations create to inform employees on the behaviours required to protect information from cyber-attacks. The application of such policy often results in a ‘them versus us’ attitude, with employees adapting the processes suggested, to fit their own agendas (Ashenden and Sasse, 2013; Hedstrom et al., 2011; Lin and Wittmer, 2017). By including employees in the creation of policy and strategy and listening to their thoughts on what is working and what is not, feelings of empowerment will develop and higher value in the policies perceived. As with many of the factors that form part of the Cybersecurity Awareness Framework (CAF), online communities can be useful in collating employee feedback on the usability of policy, helping understand where security workarounds are likely occurring. Sentiment analysis, the use of natural language processing to identify affective states in relation to a topic, can be used to highlight quickly from the collaborative text, where negative sentiment exists in relation to policy and extract information on how best to intervene.

Together, the six key factors that formulate the Cybersecurity Awareness Framework (CAF) – developed as a key outcome of this PhD thesis – formulate a way in which to measure, manage and report on human vulnerabilities in cybersecurity moving forward. Organisations must work towards developing a cybersecurity awareness culture that considers their cybersecurity experience and what is therefore fair and reasonable to expect from them. A cybersecurity awareness culture can be achieved by ensuring employees remain up to date around risks, and what can be done to protect themselves from these risks, perhaps through the use of debiasing techniques such as nudging. That they have regular opportunities to experience the behaviours they are required to undertake, possibly via serious games, and discuss any challenges openly through online communities, in turn positively influencing employee attitude towards cybersecurity. With a focus shift towards the Cybersecurity

Awareness Framework (CAF), organisations can begin a more advantageous journey towards securing their information held within cyberspace.

Empirical Block 2: Vulnerability Exploitation and Intervention

The second set of studies (4 and 5 – Chapter 3) within this thesis were developed and designed to examine the social engineering strategies that cybercriminals are using to further exploit some of the human decision-making constraints previously identified and highlighted within studies 1-3 and across the wider literature. Helping improve understanding in relation to how external malicious persuasion can influence the human cybersecurity experience. The main aim was to gain a novel and up-to-date understanding of the techniques being used and reported at the time the experiments were conducted (2020-21), as well as the techniques to which humans are potentially more vulnerable, with an aim of better understanding where intervention should be focused. In order to determine the techniques reported and perhaps more regularly used by cybercriminals, Study 4 involved analysing 641reported phishing emails from a multinational company (initially 1000 prior to data cleansing as explained within the methods section of Study 4). Results indicated authority, scarcity, and curiosity to be the key (in terms of likely prevalence – Study 4; and susceptibility – Study 5) persuasion tactics of concern and where intervention should focus.

Authority and scarcity, in particular, have remained two of the top techniques identified in phishing email research stemming ~ten years and were also found to be the techniques most likely to result in a cyber-attack (see Study 5 and Akbar, 2014; Ferreira et al., 2015). Curiosity as a method of persuasion has not previously been included in similar research yet was found to be the second most reported technique (within Study 4); however, was one of the least likely to result in a breach. This was possibly due to the simplistic way in which curiosity was being induced in the analysed phishing emails, with just links or less tailored content increasing detection of deception, however it is a quick and dirty phishing template for offenders to action. More recently research is starting to apply more focus to the importance of curiosity as a weapon of influence, finding it to results in faster 'clicks' in phising emails (Kuraku, 2022; Sarno et al., 2023).

Authority, as a technique, can be used in phishing emails through the display of prestigious titles and accolades suggesting to recipients that such expertise can be used as a short-cut to decision-making. Obedience to authority is soft wired into humans from a young age, resulting in a general belief that experts make better decisions, that result in fewer mistakes (Hinnosaar & Hinnosaar, 2012). In emails it can be simple for cybercriminals to present themselves as a person of experience for example via an authoritative title, with humans more than happy to follow their guidance when in intuitive decision-making mode. Scarcity is the second most concerning principle used in phishing emails, with the suggestion of something limited in either quantity or time evoking a sense of urgency, moving people into making quick decisions due to fear they may lose freedom of choice (Aggarwalm et al., 2011). This principle relies on the commodity bias, whereby humans value objects or experiences based on their availability, with value increasing as supply, or time to access the supply, decreases (Brock, 1968).

Curiosity, an additional method of persuasion flagged during initial email analysis, requires recipients to move outside of an email to learn more e.g., by clicking on a link or perhaps opening an attachment. Humans have a need to fill gaps in their knowledge or learn more about a subject that they see as novel, with this need reducing self-control and increasing impulsivity (Loewenstein, 1994; Shin & Kim, 2019). This is often now educed through a phishing email containing a shared folder with an intriguing title e.g., staff salaries.

Other popular elements of phishing emails include links (71% of emails analysed), spelling and grammar errors (44% of emails) the increased use of images as well as the use of some affect (positive or negative emotive language). Fabricated email chains and requests to forward emails were found within the data set in Study 4 used to increase trust and further widen the web of deceit. Multiple threat actions were also often contained in the emails analysed, perhaps in the hope that at the very least recipients would "click to unsubscribe". Control mechanisms must therefore be found to protect against these particular email elements, with employees aware that finding these in an email may be a clue to the fact that a communication may not be genuine. However, whilst it is recommended that organisations use email filtering software and algorithms to 'flag' some of the phishing clues mentioned above, research has not yet understood how often these elements appear in genuine email, including those that are sent for marketing purposes. This makes it difficult to solely rely on technology and AI, that have limited ability when it comes to understanding context. The answer to better supporting the human when being externally influenced, may lie in a man-machine solution, leaning on the strengths and weaknesses of both applications to generate mitigation.

Human augmentation, the application of science and technology to optimise or enhance human cognitive, sensory and/or physical capabilities is believed to be the 'future' in cybersecurity intervention (Naik et al., 2022). For example, in the case of a decision support system, helping computer emergency response teams (CERTs) respond quickly to an incident by providing biotechnology such as smart glasses that offer real-time information or statistics as they continue to work on a solution in the physical world. This human-machine co-ordination of skills, allows for the benefits of both forms of collaborators to be realised. For example, when tasks are highly repetitive or involve large amounts of data AI is best suited, however tasks that require creativity, strategy, empathy and innovation humans must lead.

This assembly of ‘minds’ is best utilised in cases of high risk and high urgency, perfectly aligned to the needs of cybersecurity, where the machine can work to decrease timescales and the human consider the context in which to innovate (Human Layer Security, 2023). An example intervention could include - AI repetitively checking emails for persuasion techniques, highlighting them within an email and producing a ‘persuasion score’ based on an algorithm that the human can then digest. In opening the email, the human can be nudged, as investigated in Studies 4 – 8, by the highlighted persuasion techniques and the score (perhaps colour coded from amber to red), driving employees from intuitive to conscious thought. This would not only result in the more time-sensitive application of nudges, but nudges that differ within each email, due to the ever-changing email body, reducing nudge desensitisation (Malkin et al., 2017). The lexicon informing this procedure can then be used to continually inform email filtering and AI algorithms moving forward.

Three experimental interventions were also carried out as part of the second empirical block, to understand the employees experience in relation to intervention, and perhaps the benefits of utilising control measures currently available to organisations as technology continues to improve and expand. The first two set of studies within this thesis highlighted first the internal cognitive constraints experienced by employees, and then how these vulnerabilities are further exploited by cybercriminals, adding an additional vulnerability layer to the employee experience. The purpose of the final set of studies was to add the additional element of intervention persuasion, that is often/can be used by organisations to counter malicious communications. The three experiments examine the potential benefits such interventions afford, in order to understand whether this additional (and third) layer of vulnerability is causing additional cognitive confusion but for little gain. Also, whether interventions need to take a different form if true mitigation is to be achieved. One-thousand three-hundred and ten participants took part across the three intervention experiments, focused on the use of debiasing in the form of nudging, motivation and by attempting to adapt cognitive strategies.

Soft-paternalistic nudging is believed to be a tier one debiasing intervention, focused on guiding humans towards optimal choice without restricting options (Thaler & Sunstein, 2008). An example includes a text from the dentist, reminding the recipient that they have an up-and-coming appointment (Sunstein, 2014). Within Study 6, nudges were used to highlight to participants that an email may be phishing, with a fear appeal nudge generated to increase the availability bias and reduce the optimism bias resulting in significantly more malicious emails being filed as suspicious than no nudge.

The experiment found within Study 7, was designed to explore whether the use of a self-determined motivational statement would also be a significantly useful intervention, against a control conditional (motivation believed to be a key driver across a number of factors featuring in the Cybersecurity Awareness Framework (CAF) such as threat appraisal (McGill & Thompson, 2017; Rogers, 1975). The motivational communication was not found to significantly improve phishing email detection, although it did offer some benefits when added to a fear appeal nudge (with threat, optimism and coping elements) but no more than a fear appeal nudge alone. These findings suggest that a soft-paternalistic nudge with the aim to increase threat appraisal may offer the most benefit to employees, with the least amount of additional cognitive burden.

The final experiment within Study 8, investigated two techniques that attempt to adapt human cognitive strategies to those more security focused (Croskerry et al., 2013; Larrick, 2004); consider-the-opposite and a maxim. *Consider the opposite* teaches employees to assume each email is phishing with a need to detect clues to the contrary. The *maxim* was a memorable acronym that participants were required to remember that reminded them to check elements of each email for suspicious clues. A nudge looking to directly target two aspects of the Cybersecurity Awareness Framework (CAF; threat appraisal and self-efficacy) the maxim debiasing strategy were found to be significantly more effective than no intervention. However despite the maxim technique resulting in more phishing emails being correctly filed as suspicious than the other experimental conditions, there were no significant differences found between the intervention types applied.

It is evident from Studies 6 - 8 that bringing employee attention to potential phishing email presence can be a useful intervention strategy, perhaps alleviating the external pressures of vulnerability manipulation by cybercriminals. Traditional education looks to influence the human experience in cybersecurity via system two thinking, and debiasing via system one with perhaps this route effective when something such as interacting emails is often undertaken habitually. Whilst debiasing can be useful in attracting attention back to system two thinking, this process will have an impact on productivity, with organisations needing to consider the costs implied. It is therefore important to find ways to ensure that a nudge is delivered when required, and not continuously, perhaps through a man-machine solution as previously discussed (Human Layer Security (2023; Naik et al., 2022).

An additional supporting intervention could include, the use of biotechnology, the integration of the natural and engineering sciences, to alert employees to vulnerability, for example when their stress levels, emotions or level of cognitive load may suggest that error might be imminent. Such biotechnology is already being used by employees in their personal lives to count steps, monitor sleep and remind them about events, it is not implausible that the same technology could be of use in cybersecurity such as automatic door entry and device securement when a certain distance exists between a person and the technology/device. Research to date has been positive in this area, suggesting such devices are useful in impacting behaviour change (Hartment et al., 2018; Ringeval, 2020). This technology could therefore also help reduce the cognitive burden of many cybersecurity tasks for humans, such as making humans aware of when e.g., affect is being aroused in them and error may be imminent. Studies 4 – 8 evidenced the benefit of nudges in the cybersecurity domain, to perhaps increase threat appraisal, with biotechnology perhaps helping deliver such debiasing techniques when most required.

What is clear from Studies 4 – 8 is the need for organisations to become cognizant not only to the individual differences and perceptions impacting their employees, but the continual persuasion delivered by external influencers and how this may be holistically influencing their behaviour. Employees do not only experience vulnerability in relation to their own attitudes and perceptions, but this experience is then continuously influenced by external factors that can change this experience and impact awareness at any given time moment. A universal understanding of the employee experience in cybersecurity is required if genuine mitigation to cyber-attacks is to be achieved.

The Holistic Human Cybersecurity Experience

The fundamental aim of this research was to better understand the human in cybersecurity, through a series of studies, experimental research and critical analysis considering the breadth of psychological, social and behavioural economics literature available. With the objective of providing organisations – as well as other communities (e.g. academics) with an improved understanding around how the human may experience cybersecurity and what, if anything, can be done to better protect them from risk.

Findings from both empirical blocks paint an interesting picture on how employees are interacting with cybersecurity, and the need to provide them with more than just compliance-led education around several curriculum-based attack vectors e.g., malware, how to recognise it, and how to remain secure. Awareness programmes must also consider the attitudes and perceptions causing most vulnerability within the Cybersecurity Awareness Framework (CAF), still likely influencing behaviour even when thorough education has been delivered. Also important is understanding and considering the external cognitive persuasion tactics that are delivered to employees both maliciously and sympathetically. Whilst humans may approach cybersecurity positively, vulnerabilities will still occur when malicious cybercriminals are attempting to evoke cognitive bias without warning. In addition, organisations and cybersecurity awareness vendors are working hard to restrict the influence malicious action can have on human decision-making, yet not enough is known about how much benefit they deliver employees, particularly over the longer term (Venema, 2018). Interventions analysed within this thesis have evidenced some support for debiasing techniques, particularly in relation to the identification of phishing emails.

To conclude, Cyber-attack mitigation is not currently being achieved despite the deployment of multiple technical interventions and more focus applied to employee compliance training. This is perhaps due to a lack of appreciation around how employees are holistically experiencing cybersecurity and the vulnerabilities that this can bring. Moving forward, organisations must consider not only the internal vulnerabilities faced by the human (measured via the tool generated as part of this PhD), but also external influences that look to encourage decision-making error, such as manipulation of the authority and scarcity biases. It is important that organisations stay abreast of those techniques most impactful in their particular industry (as conducted in Studies 4 – 5). Whilst a number of debiasing techniques can be useful in countering some of this bias, such as the use of a soft-paternalistic nudge targeting threat appraisal or supporting cognitive strategy adaptation such as memorising a maxim to help detect phishing email (as explored in Studies 6 – 8), more is required to ensure such interventions do not reduce productivity and are applied when error is more likely. This can perhaps be achieved by utilising human augmentation or biotechnology that can help deliver such interventions when actively required. The main objective of this PhD was both novel and timely – to furnish organisations with a more holistic understanding around the employee experience in cybersecurity, considering both internal and external influencers. With some guidance around how to reduce this risk moving forward.

Future research is required to not only address some of the key limitations discussed in relation to this research, but also to investigate ways in which to support the human decision-making errors experienced by defenders (Chapter 2) and those that are externally manipulated for both positive and negative motives (Chapter 3). Human decision-making heuristics have developed across evolution and a human's lifetime, so to make decisions without them would result in encumbrance. Day-to-day life error often comes from quick decisions with little consequence, however in a domain like cybersecurity, error can result in serious impact to an organisation, its staff and customers. It is therefore important to understand ways to better support human decision-making and reduce bias when e.g., opening emails or sharing passwords at work.

In specific relation to the Cybersecurity Awareness Framework (CAF), it is important to determine whether the framework can be validated across different samples and populations, as well as its ability to measure actual behaviour, rather than just intentions. The literature suggests that a number of additional individual differences may be related or predictive of cybersecurity behaviours, with future research perhaps looking to extend the framework's predictive power above and beyond its current 60% through the inclusion of additional expounding variables. Whilst Studies 1 – 3 identified an overarching cybersecurity awareness factor and six informing human vulnerabilities, this does not provide a complete picture around the decision-making challenges experienced by employees. Examples could include situational factors such as stress, burnout and cybersecurity fatigue (Corradini, 2020; Nobles, 2022; Reegård et al., 2019). Research should continue to evolve the framework and refine the metrics included.

Whilst Studies 4 -5 attempted to provide universal insight into how cybercriminals may be manipulating victims into interacting with a phishing email, around 20% of the phishing emails reported contained no method of persuasion. Future research should further analyse phishing emails that appear to contain no method to better understand whether there are further techniques not yet unearthed within them, or whether cybercriminals are really utilising such simplistic formats. Curiosity was also uncovered as an additional technique of concern, with future research required to consider this tactic alongside the previous six methods considered. This research looked to build on the small number of previous literature that has investigated this topic, providing new data points to add to previous findings (Akhbar, 2014; Ferreira et al., 2015; Harris & Yates, 2015, Lawson et al., 2020; Parsons et al., 2019; Rajivan & Gonzalez, 2018; Zielinska et al., 2016). Future research should look to continue this work by analysing ongoing social engineering trends as well as how susceptible humans are to each of them as time progresses, allowing intervention to prioritise risk in relation to individuals and groups of individuals. Susceptibility could be biologically measured via eye-tracking, EEG, heart rate monitoring and more in order to inform biotechnical interventions.

The final set of experiments involved exploration into a number of easily deployable interventions that could potentially offer employees support when working in a more intuitive decision-making mode. Whilst some promise was found for the use of debiasing techniques in cybersecurity, more research is required to determine their use across other vulnerabilities, as well as whether employees will become desensitised to their effect over time. However, research around the desensitisation to nudging is still in its infancy (Petelka et al., 2019).

Several alternative debiasing strategies are available that can also be tested for use within the cybersecurity domain such as cognitive tutoring, affective debiasing and simulation training (Croskerry et al., 2013). It is also important to investigate more technical interventions as they continue to evolve within the space such as the use of technology to provide online serious games that are gamified and involve an online collaborative community. In addition, research should continue to investigate how human augmentation can support the human in cybersecurity. Many people are now using (wearing) smart devices to improve their experience within the world e.g., smart watches. A wearable device can naturally progress into also protecting the human from cyber-attacks whilst reducing the cognitive pressure cybersecurity places on the employee. Future research should investigate their use within the cybersecurity domain, in additional to any ethical implications around its use.

Limitations

As always anticipated with any research, particularly of this size, a number of limitations to the studies and experiments presented within this thesis will now be presented, in order to better inform future work. The first limitation that must be mentioned is the challenges faced due to all studies and experiments taking place during the Covid 19 pandemic (from March 2020). Due to challenges such as the need to isolate from others by staying at home and avoid standing in close proximity to reduce the risk of spreading the disease further, the ability to conduct (including small and large scale) in-person experiments (for almost 2.5-3 years) was largely prevented. This resulted in the need to rely more heavily on online testing and the use of self-reported measures, as well as the inability to ensure any experiments were undertaken without confounding variables e.g., multi-tasking. In addition, across this period it became a challenge to employ large numbers of participants from i.e., students who were no longer required to gain any or perhaps smaller numbers of participant pool credits, as well as the organisation collaborating on this PhD requiring employees to increase focus on maintaining service. Therefore, future objective, in-person laboratory-based studies would help to assess the extent to which the findings of these thesis can be replicated and whether any confounding variables influenced some of the findings.

Limitations Studies 1-3

Self-reported measures were utilised across all three studies, including the metric of cybersecurity behaviour, with potentially different results arising should more objective measures have been used. Self-reported measures were however useful – (1) the objective of understanding the more subjective employee experience as part of the exploratory nature across the first three studies, (2) the need to measure a large sample of participants and (3) due to the inability to collect objective data face to face during the early stages of the Covid-19 pandemic. Cybersecurity governs an extremely diverse set of behaviours that are continually under development and change, presenting a challenge when wanting to gain objective measurements for each required behaviour. The focus of this thesis was to therefore gain a more global understanding around human perceptions of cybersecurity, irrespective of which particular topics are of current importance or en vogue. A self-reported global measure of behaviour was therefore required; however this did limit findings in that outcomes and recommendation could not be applied to specific cybersecurity tasks.

An instrument measuring a complete, up to date and available index of the specific cybersecurity behaviours requirement is however not currently available. Investigations could perhaps commence with utilising the benefits of the validated Security Behaviour Intention Scale that focuses on updating, device securement, passwords and general awareness (SeBIS; Egelman & Peer 2015) to include the more complete list of skills detailed by Carlton (2016) in the cybersecurity skills index. This cybersecurity behaviour model can then be investigated in relation to the Cybersecurity Awareness Framework (CAF) and whether the employee experience is influencing cybersecurity behaviour both globally and specifically. Perhaps also looking to generate a more global measure of behaviour alongside. However, despite the use of self-reported measures within Studies 1 - 2, a person's intentions and behaviour have previously been found to highly correlate, with intention signifying the effort a human is willing to apply to a behaviour should ability be present (Ajzen, 1991; Conner & Armitage, 1998).

Limitations Studies 4-5

A number of limitations were also present within both Studies 4 and 5. First, all emails included in Study 4 were *reported* by employees within a multinational corporation, and therefore will have excluded phishing emails that did not raise suspicion. This may explain some disparities between levels of use (according to emails reported at least) and success (according to numbers of emails not reported) found in these studies. Due to a current lack of data repositories that contain both filtered and reported emails, analyses such as these do tend to rely solely on reported emails (see also Zielinska et al., 2016).

Limitations were also experienced due to the confidential and potentially hazardous nature of the emails being analysed in the second set of studies. Requirements (ethical) to send emails to the researchers as txt. files resulted in the removal of colour, font and images that may have added to or helped explain some of the findings. Sender information was also removed, and links deactivated removing the ability to check if the reported emails were in fact malicious. Therefore, a portion of the emails may have not been determined as legitimate or spam. This is however the nature of most phishing email analyses due to a current lack of repository containing reported emails as well as those uncovered by email filtering (e.g., Zeilinska et al., 2016). Analyses did however replicate many of the findings of previous studies where this information was made available (Akbar, 2014;; Ferreira et al., 2015; Zielinska et al., 2016). It is also important to note that within these experiments, over reporting may have occurred with participants less sure whether ham emails, those without malicious intent, were innocuous. It must however be remembered that this was the first known examination of potential phishing emails from a multi-national organisation rather than universities, personal inboxes or cross-company (Akbar, 2014;; Ferreira et al., 2015; Zielinska et al., 2016). Current work is being undertaken by the authors of this thesis and the collaborating organisation to address these limitations.

Study 5 included the use of a simulated study whereby participants had to imagine themselves to be the owner of an imitated inbox, and thus assumptions may have been made by them in regard to previous communications with the sender, legitimacy of companies and so on. The fact that it was not a real inbox (e.g., personal, work) could also increase the likelihood of participants worrying less about the consequences of making incorrect decisions and actioning a potentially malevolent phishing email. Also, participants in Study 5 were university undergraduate students all studying the same course and not employees – making generalisation to an employee sample needing to be met with some caution (although confirmation of at least a part-time job was required). Study 5 was also conducted online and as such it is impossible to control for confounds such as being able to ensure participants were only engaging in the study and not experiencing distractions and interruptions. A future in-person laboratory-based experiment would help to assess the extent to which these and other potential factors may have confounded some of the findings..

Limitations Studies 6-8

Some limitations were also present within Studies 6 – 8 that must be considered when interpreting findings, implications and considering future directions. Whilst the statement "your inbox is at equal risk to others" was used to evoke more realistic optimism, the statement could also have targeted psychological ownership by stating the inbox as theirs. Previous research has previously identified that a change from the word "the" to "your" to be enough to significantly increase psychological ownership and resultant behaviour (Peck et al., 2021). Psychological ownership has indeed previously been found to increase the effectiveness of the use of fear appeals in increasing threat appraisal (Briggs et al., 2017). Future research should investigate which of the two underlying decision-making biases are at the root of improved behaviour.

Effect sizes for Study 6 and 7 were also small to moderate, and Study 8 small. This is potentially due to a maximum of sixteen phishing emails available to file within the studies, with the analysis of larger numbers of emails potentially resulting in larger effect size. Within this set of experiments, genuine emails (no signs of malevolence, methods of persuasion etc.) were also present to ensure deception was maintained limiting the number of phishing emails made available in order to keep participation within a reasonable time frame. Future research could however include phishing emails only to allow more instances for emails to be reported possibly increasing effect size. Whilst it is assumed that the significantly larger number of phishing emails correctly filed as suspicious was due to interventions targeting threat appraisal, it is important that more research is conducted to confirm whether the availability and optimism biases are indeed targeting this factor and that these interventions do actually result in increases of participant appraisal of threat.

It is essential to note that the nature of the interventions themselves could possibly have made participants within those groups aware that the purpose of the exercise was to identify phishing and therefore more significant results would be seen. However, should this be so, the difference in number of emails filed as suspicious across experimental groups still poses interesting findings to further be explored. It would also be of interest to explore eye fixation, blinks and perhaps pupillometry as well as time and click rates to provide a more holistic picture of participant experience (research due to be undertaken by the lead author as part of her current role this year). Despite the limitations highlighted, this was the first known set of experiments investigating different ways in which to support intuitive decision-making in the cybersecurity domain offering precedence for the continuation of research into interventions supporting intuitive and habitual decision-making. Finally, due to the aim of creating a set of recommendations for organisations that span across all aspects of human risk, only threat appraisal (availability and optimism bias) was analysed in relation to the three examined interventions. It is possible that the findings will be different when other human biases are examined. Future research should look to examine these interventions in relation to other human decision-making biases that are believed to impact human risk in relation to cybersecurity.

The current research focused on threat appraisal and the decision-making biases underlying it, however this is just one of the factors found to influence cybersecurity behaviour with many more to be explored. Future research should investigate the use of such strategies involving the other variables and their success in supporting end-user detection of phishing emails. Findings in Experiments 6 to 8 should also be replicated using a larger number of emails, more participants and a more true-to-life environment such as within an organisation's phishing simulation programme. Whilst some research has taken place on the desensitisation of nudges and debiasing techniques, this work needs to continue to find ways in which to ensure employees within organisations take note of the messaging when continuously exposed.

Conclusion

Whilst developments in technology have offered great benefits to communication, productivity and information sourcing attacks towards online data and system integrity are continuing to evolve in both number and level of intelligence. Yet, despite the introduction of a number of technical interventions to help support the issue, significant reductions in the number of cyber-attacks experienced across the world is not decreasing. With 82% of security breaches believed to involve a human element, many organisations are now setting aside time and budget to provide their employees with some form of cybersecurity compliance training. However, in spite of these further efforts, mitigation in relation to human risk is yet to be significantly witnessed (Techtarget, 2023; Verizon, 2022).

Whilst psychological research in relation to human risk in cybersecurity is continuing to mature, the challenges faced within the space have previously been experienced across a number of alternative domains without much comparison to date. This has resulted in a breadth of research that can be drawn upon to better support organisations in mitigating cybersecurity human risk and further investigated with cybersecurity challenges in mind. The aim of the PhD linked to this thesis was to therefore bring together a wide scope of potentially relevant literature across psychology, sociology and behavioural economics extending its findings within cybersecurity through a number of explorative experiments. The result – a human-centric cybersecurity playbook and Human Susceptibility to Cyber-attacks tool that will provide organisations with accessible steps they can take to measure, manage and mitigate human risk from today.

Human-centric cybersecurity risk refers to the intended or unintended human behaviours that increase the probability of a cyber-attack resulting in success. This can include risks posed by employees, groups within an organisation or the cybercriminals themselves. To manage human risk, an organisation needs to record key metrics that can help identify and continuously measure employee vulnerability as well as cybercriminal trends to anticipate potential susceptibility moving forward. Building intelligence around the systematic and idiosyncratic risks posed by humans, or groups of humans, affords the first step towards building a stronger human line of defence. Once intelligence has been collated, both systematic and dynamic interventions can be applied in a cyclic manner resulting in a robust human risk management framework reinforming the support required from technology and the organisations themselves.

In order for employees to maintain organisation security it is not enough to simply educate on what needs to be done to protect company systems and data. Employees also need to know how to perform the skills being asked of them, and be able to do so proficiently (Carlton, 2016). Employees becoming proficient in cybersecurity skills will not only enable them to conduct the skills when busy at work but avoid unnecessary decreases in productivity that would be experienced should they be required to conduct each behaviour consciously.

In addition to employees having the skills required to keep an organisation secure, they must also feel motivated to want to put the skills learned into action. For example, an employee may have the ability to create a strong password but may not feel motivated to do so if they perceive the risk of an attack to be low. Motivation is the cognitive energy that drives behaviour believed to run from amotivation through to intrinsic motivation where tasks are undertaken for the pleasure of doing so (Heckhausen & Heckhausen, 2008; Touré-Tillery & Fishbach, 2014). In cybersecurity, pleasure or interest in its tasks will unlikely be experienced, with extrinsic motivation and seeing the true value of the task the level of motivation organisations should strive for. Motivation can be increased in a number of ways, such as communicating to employees the number of incidents that have recently occurred in order to increase their appraisal of threat. Another key way may be to increase their sense of connection to work technology and data, for example allowing more customisation, to increase aversion to any loss as a result of a cyber-attack.

Whilst increasing skill through employment experience and involvement and maintaining motivation by increasing threat appraisal and psychological ownership will be beneficial, there is little use in the information informing these aspects out of date. For awareness to remain contemporary, knowledge needs to naturally flow around an organisation so that even outside of awareness training employees and company cybersecurity assets remain up to date. Effective awareness can be managed through current and consistent awareness programmes, a knowledge sharing culture and motivation for collaboration. To be aware is to hold knowledge and understanding around a situation or fact which can only exist is information is actively shared. Knowledge can be communicated in a number of ways; implicit knowledge can be shared through collaborative meetings and online portals that allow employees to fill in the knowledge gaps of others and explicit knowledge e.g., that held in training documents and policy, continually optimised via employee feedback. Employees should be encouraged to share knowledge that can be declaratively communicated and given the opportunity to practice and observe knowledge that is more procedural and complex. For an effective cybersecurity culture to be achieved, organisations must therefore provide employees with the tools required to become proficiently skilled in cybersecurity, feel motivated to put those skills into action and have the opportunity to continue to learn and educate others on the knowledge required to ensure these skills are up to date.

There are a number of interventions outside of conventional training that can be utilised to increase employee skill and motivation and encourage a risk-aware culture. The first is the use of soft-paternalistic nudging a way of indirectly influencing behaviour without the need to set commands or prohibitions (Thaler & Sunstein, 2008). Nudging can offer support to employees working in intuitive mode, guiding them towards safe action, for example providing a warning that a link in an email may have malicious intent. Humans can however become desensitised to nudging, with the need for both content and context to remain dynamic so that employees continue to digest its important message (Petelka, Zou & Schaub, 2019).

Another form of intervention that is possibly useful in supporting intuitive employee decision-making is debiasing, optimising their current utilised cognitive strategies. To debias, organisations need to educate employees on the bias of concern, inform them around how to detect it occurring, motivate them to want to change it, teach them an improved strategy and support them in its maintenance (Croskerry et al., 2013). The debiasing strategy that appears to have the most potential in the cybersecurity domain is perhaps that of a mental checklist/maxim such as Who – What – How, that highlights three key things an employee needs to become habitually used to scanning when opening each email if they are to remain secure. However, more research is required to validate this.

It has recently been indicated that by 2025, many working within cybersecurity will suffer burnout and fatigue due to the lack of mitigation currently being witness, resulting in a lack of cybersecurity skills within the space (Techtarget, 2023). The playbook is therefore a tool for all involved, but in particular those fighting hard day on day to keep employees and organisations secure. The time is *now* to utilise the full armoury of information psychological research has to offer to support organisations in reducing cyber-attacks related to the human before these predictions become a reality. The main objective of this playbook – to provide organisations with a snapshot of the breath of research currently available in relation to the human in cybersecurity (including experiments undertaken within this thesis) allowing organisations to take purposeful steps towards cyber-attack mitigation from today.

Final Words

The key objective of this PhD thesis and associated research programme was to bring together the hard work of many forefathers and empirical contemporaries both within and outside of the cybersecurity domain, multiple specialist cybersecurity teams, key influencers in the cybersecurity industry and a lot of hours of investigation and experimentation as used to inform an improved understanding around the human cybersecurity experience. Cybersecurity is a sensitive and extensive subject making it difficult to extrapolate the multitude of information that is required to truly make a difference. By integrating the work of key people and teams across psychology, cybersecurity and beyond, a more concise understanding has been gained on the experiences of the human in cybersecurity.

Whilst the mind of the human is extraordinary, is it not infallible. Expectations set by company awareness teams need to be realistic, with a collective understanding on what makes employees an organisation's greatest cybersecurity asset - trialling training, investigating new challenges, feeding back insights and sharing knowledge across teams build a risk-aware culture into the very fabric of the organisation. Human cognition is being used by cybercriminals to weaken cybersecurity defences, it is now time to fight back by armouring organisations with the knowledge they require to upskill, motivate and collaborate with their employees in order to allow them to become their strongest cybersecurity line of defence.

It is time to develop interventions that holistically consider the human experience in cybersecurity, with the support of current and future technology. if cyber-attack mitigation is to be achieved.

References

Aarts, H., Custers, R., & Holland, R. W. (2007). The nonconscious cessation of goal pursuit: when goals and negative affect are coactivated. *Journal of personality and social psychology*, *92*(2), 165.

Abroshan, H., Devos, J., Poels, G., & Laermans, E. (2018). Phishing attacks root causes.

Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007, October). A comparison of machine learning techniques for phishing detection. In Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit (pp. 60-69).

Aggarwal, P., Jun, S. Y., & Huh, J. H. (2011). Scarcity messages. *Journal of Advertising*, *40*(3), 19-30.

Agha, S., Tollefson, D., Paul, S., Green, D., & Babigumira, J. B. (2019). Use of the Fogg behavior model to assess the impact of a social marketing campaign on condom use in Pakistan. Journal of health communication, 24(3), 284-292.

Aivazpour, Z., & Rao, V. S. (2018). Impulsivity and risky cybersecurity behaviors: A replication. *Twenty-fourth Americas Conference on Information Systems, New Orleans, 2018.*

Ajzen, I. (1991). The Theory of Planned Behavior, Organızational Behavior And Human Decision Processes. 50, 179-211. *Aras, M., Özdemir, Y., ve Bayraktaroglu, S.(2015). Insan Kaynaklari Bilgi Sistemlerine Yönelik Algilarin Teknoloji Kabul Modeli Ile Incelenmesi/The Investigation of Perceptions for Human Resource Information Systems via Technology Acceptance Model. Ege Akademik Bakis, 15*(3), 343.

Ajzen, I., & Fishbein, M. (1975). A Bayesian analysis of attribution processes. *Psychological Bulletin*, *82*(2), 261.

Ajzen, I., & Fishbein, M. (2000). Attitudes and the attitude-behavior relation: Reasoned and automatic processes. *European review of social psychology*, *11*(1), 1-33.

Akbar, N. (2014). *Analysing persuasion principles in phishing emails* (Master’s thesis, University of Twente).

Akkad, A., Wills, G., & Rezazadeh, A. (2023). An information security model for an IoT-enabled Smart Grid in the Saudi energy sector. *Computers and Electrical Engineering*, *105*, 108491.

Alawida, M., Shawar, B. A., Abiodun, O. I., Mehmood, A., & Omolara, A. E. (2023). Unveiling the Dark Side of ChatGPT: Exploring Cyberattacks and Enhancing User Awareness.. Renaud, K., Warkentin, M., & Westerman, G. (2023). From ChatGPT to HackGPT: Meeting the Cybersecurity Threat of Generative AI. MIT Sloan Management Review.

Albladi, S. M., & Weir, G. R. (2018). User characteristics that influence judgment of social engineering attacks in social networks. *Human-centric Computing and Information Sciences*, *8*(1), 1-24.

Aldawood, H., & Skinner, G. (2019). Reviewing cyber security social engineering training and awareness programs—Pitfalls and ongoing issues. *Future Internet*, *11*(3), 73.

Alhalafi, N., & Veeraraghavan, P. (2023). Exploring the Challenges and Issues in Adopting Cybersecurity in Saudi Smart Cities: Conceptualization of the Cybersecurity-Based UTAUT Model. *Smart Cities*, *6*(3), 1523-1544.

Aljazzaf, Z. M., Perry, M., & Capretz, M. A. (2010, September). Online trust: Definition and principles. In *2010 Fifth International Multi-conference on Computing in the Global Information Technology* (pp. 163-168). IEEE.

Aloseel, A., He, H., Shaw, C., & Khan, M. A. (2020). Analytical review of cybersecurity for embedded systems. *IEEE Access*, *9*, 961-982.

Al-Qahtani, A. F., & Cresci, S. (2022). The COVID-19 scamdemic: A survey of phishing attacks and their countermeasures during COVID-19. *IET Information Security*, *16*(5), 324-345.

Alrawi, M. A. S., Samy, G. N., Yusoff, R. C. M., Shanmugam, B., Lakshmiganthan, R., Maarop, N., & Kamaruddin, N. (2020). Examining factors that effect on the acceptance of mobile commerce in malaysia based on revised UTAUT. *Indonesian Journal of Electrical Engineering and Computer Science*, *20*(3), 1173-1184.

Alshaikh, M., Maynard, S. B., Ahmad, A., & Chang, S. (2018). An exploratory study of current information security training and awareness practices in organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Alshaikh, M., Naseer, H., Ahmad, A., & Maynard, S. B. (2019). Toward sustainable behaviour change: an approach for cyber security education training and awareness.

Althonayan, A., & Andronache, A. (2018, September). Shifting from information security towards a cybersecurity paradigm. In *Proceedings of the 2018 10th International Conference on Information Management and Engineering* (pp. 68-79).

Altman, E. J., Nagle, F., & Tushman, M. (2019). Managed ecosystems and translucent institutional logics: Engaging communities. Working Paper

Amah, E., & Ahiauzu, A. (2013). Employee involvement and organizational effectiveness. *Journal of Management Development*, *32*(7), 661-674.

Amirpur, M. (2017). *The Role of Cognitive Biases for Users' Decision-Making in IS Usage Contexts*. Unpublished Doctoral Thesis). Technischen Universität Darmstadt, Gabrovo, Bulgaria.

Anwar, M., He, W., Ash, I., Yuan, X., Li, L., & Xu, L. (2017). Gender difference and employees' cybersecurity behaviors. *Computers in Human Behavior*, *69*, 437-443.

Ariffin, N. H. M., Ahmad, F., & Haneef, U. M. (2020). Acceptance of mobile payments by retailers using UTAUT model. *Indonesian Journal of Electrical Engineering and Computer Science*, *19*(1), 149-155.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological bulletin*, *110*(3), 486.

Arnab, S. (Ed.). (2012). *Serious games for healthcare: Applications and implications: applications and implications*. IGI Global.

Ashenden, D., & Sasse, A. (2013). CISOs and organisational culture: Their own worst enemy?. *Computers & Security*, *39*, 396-405.

Bada, M., Sasse, A. M., & Nurse, J. R. (2019). Cyber security awareness campaigns: Why do they fail to change behaviour?. Retrieved from: https://www.researchgate.net/profile/Maria-Bada/publication/330276525_Cyber_Security_Awareness_Campaigns_Why_do_they_fail_to_change_behaviour/links/6201157b6b16d97aed2667aa/Cyber-Security-Awareness-Campaigns-Why-do-they-fail-to-change-behaviour.pdf

Bada, M., Von Solms, B., & Agrafiotis, I. (2019). Reviewing national cybersecurity awareness in Africa: an empirical study.

Baldwin, R. (2014). From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review*, *77*(6), 831-857.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American psychologist*, *37*(2), 122.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of social and clinical psychology*, *4*(3), 359-373.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics*, *107*(3), 797-817.

Barash, V. (2011). The dynamics of social contagion (Unpublished Doctoral Thesis). Cornell University, Ithaca, New York.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, *27*(1), 173-196.

Barendse, S. W. (2023). Exploring Gamification and Cybersecurity: How Could Gamification Increase the Cybersecurity Awareness.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of personality and social psychology, 81*(6), 1014.

Barnfield, M. (2020). Think twice before jumping on the bandwagon: Clarifying concepts in research on the bandwagon effect. *Political studies review*, *18*(4), 553-574.

Bassiouni, M., Ali, M., & El-Dahshan, E. A. (2018). Ham and spam e-mails classification using machine learning techniques. Journal of Applied Security Research, 13(3), 315-331.

Bauer, B., & Patrick, A. S. (2004). A human factors extension to the seven-layer OSI reference model. *Retrieved January*, *6*, 2004.

Baxter, W. L., Aurisicchio, M., & Childs, P. R. (2015). A psychological ownership approach to designing object attachment. *Journal of Engineering Design*, *26*(4-6), 140-156.

Benenson, Z., Gassmann, F., & Landwirth, R. (2017). Unpacking spear phishing susceptibility. In *Financial Cryptography and Data Security: FC 2017 International Workshops, WAHC, BITCOIN, VOTING, WTSC, and TA, Sliema, Malta, April 7, 2017, Revised Selected Papers 21* (pp. 610-627). Springer International Publishing.

Bion, W. R. (2021). *Learning from experience*. Routledge.

Black, P. E., Scarfone, K., & Souppaya, M. (2008). Cyber security metrics and measures. *Wiley Handbook of Science and Technology for Homeland Security*, 1-15.

Blais, A. R., & Weber, E. U. (2006). A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision making*, *1*(1), 33-47.

Blanco, F. (2017). Cognitive bias. Encyclopedia of Animal Cognition and Behavior, 1-6.

Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual review of psychology*, *62*, 391-417.

Bonhoeffer, D. (2012). *Ethics*. Simon and Schuster.

Bottemanne, H., Morlaàs, O., Fossati, P., & Schmidt, L. (2020). Does the coronavirus epidemic take advantage of human optimism bias?. *Frontiers in Psychology*, *11*, 2001.

Bradley, P., Chambers, W., Davenport, C., & Saner, L. (2017). A National Research Agenda on Insider Threat.

Branley-Bell, D., Coventry, L., Dixon, M., Joinson, A., & Briggs, P. (2022). Exploring Age and Gender Differences in ICT Cybersecurity Behaviour. *Human Behavior and Emerging Technologies, 2022*.

Brasel, S. A., & Gips, J. (2014). Tablets, touchscreens, and touchpads: How varying touch interfaces trigger psychological ownership and endowment. Journal of Consumer Psychology, 24(2), 226-233.

Briggs, P., Jeske, D., & Coventry, L. (2017). Behavior change interventions for cybersecurity. In *Behavior change research and theory* (pp. 115-136). Academic Press.

Brock, T. C. (1968). Implications of commodity theory for value change. In *Psychological foundations of attitudes* (pp. 243-275). Academic Press.

Burn, S. M. (2017). Appeal to bystander interventions: A normative approach to health and risk messaging. In *Oxford research encyclopedia of communication*.

Burns, S., & Roberts, L. (2013). Applying the theory of planned behaviour to predicting online safety behaviour. *Crime Prevention and Community Safety, 15*, 48-64.

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

Can I Phish (2022). The History of Phishing. Retrieved from: https://caniphish.com/phishing-resources/history-of-phishing.

Caraban, A., Karapanos, E., Gonçalves, D., & Campos, P. (2019, May). 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1-15).

Carlton, M. (2016). *Development of a cybersecurity skills index: A scenarios-based, hands-on measure of non-IT professionals' cybersecurity skills* (Doctoral dissertation, Nova Southeastern University).

Carpenter, D., Young, D. K., Barrett, P., & McLeod, A. J. (2019). Refining technology threat avoidance theory. *Communications of the Association for Information Systems*, *44*.

Carr, D. (1979). The logic of knowing how and ability. *Mind*, *88*(351), 394-409.

Carroll, J. S. (1978). The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of experimental social psychology*, *14*(1), 88-96.

Cavanagh, G. F., Moberg, D. J., & Velasquez, M. (1981). The ethics of organizational politics. Academy of management review, 6(3), 363-374.

Ceric, A., & Holland, P. (2019). The role of cognitive biases in anticipating and responding to cyberattacks. *Information Technology & People*, *32*(1), 171-188.

Chaudhary, S., Gkioulos, V., & Katsikas, S. (2023). A quest for research and knowledge gaps in cybersecurity awareness for small and medium-sized enterprises. *Computer Science Review*, *50*, 100592.

Chaudhry, J. A., Chaudhry, S. A., & Rittenhouse, R. G. (2016). Phishing attacks and defenses. *International journal of security and its applications*, *10*(1), 247-256.

Chen, H., Turel, O., & Yuan, Y. (2022). E-waste information security protection motivation: the role of optimism bias. *Information Technology & People, 35*(2), 600-620.

Chenoweth, T., Minch, R., & Tabor, S. (2007). Expanding views of technology acceptance: seeking factors explaining security control adoption. AMCIS 2007 Proceedings, 321.

Cialdini, R. B. (1984). The psychology of persuasion. *New York: Quill William Morrow*.

Cialdini, R. B. (2001). Harnessing the science of persuasion. *Harvard business review, 79*(9), 72-81.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology (pp. 151–192).* McGraw-Hill.

Cismaru, M., Nagpal, A., & Krishnamurthy, P. (2009). The role of cost and response-efficacy in persuasiveness of health recommendations. *Journal of Health Psychology*, *14*(1), 135-141.

Conetta, C. (2019). Individual differences in cyber security. McNair Research Journal SJSU, 15(1), 4.

Conner, M., & Armitage, C. J. (1998). Extending the theory of planned behavior: A review and avenues for further research. *Journal of applied social psychology*, *28*(15), 1429-1464.

Corradini, I. (2020). *Building a cybersecurity culture in organizations* (Vol. 284). Berlin/Heidelberg, Germany: Springer International Publishing.

Coventry, L., Briggs, P., Jeske, D., & Van Moorsel, A. (2014). SCENE: A structured means for creating and evaluating behavioral nudges in a cyber security environment. In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part I 3* (pp. 229-239). Springer International Publishing.

Cox, A., Zagelmeyer, S., & Marchington, M. (2006). Embedding employee involvement and participation at work. *Human Resource Management Journal*, *16*(3), 250-267.

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: origins of bias and theory of debiasing. *BMJ quality & safety*, *22*(Suppl 2), ii58-ii64.

Cutello, C. A., Walsh, C., Hanoch, Y., & Hellier, E. (2021). Reducing optimism bias in the driver's seat: Comparing two interventions. *Transportation research part F: traffic psychology and behaviour*, *78*, 207-217.

Cybertalk (2022). History of Phishing. Retrieved from: https://www.cybertalk.org/2022/03/30/top-15-phishing-attack-statistics-and-they-might-scare-you/

Dada, E. G., Bassi, J. S., Chiroma, H., Adetunmbi, A. O., & Ajibuwa, O. E. (2019). Machine learning for email spam filtering: review, approaches and open research problems. Heliyon, 5(6).

Darban, M., & Polites, G. L. (2020). Why is it hard to fight herding? The roles of user and technology attributes. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *51*(4), 93-122.

De Bona, M., & Paci, F. (2020, August). A real world study on employees' susceptibility to phishing attacks. In *Proceedings of the 15th International Conference on Availability, Reliability and Security* (pp. 1-10).

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. *Handbook of Theories of Social Psychology*, 20.

Deci, E. L., Ryan, R. M., Deci, E. L., & Ryan, R. M. (1985). Conceptualizations of intrinsic motivation and self-determination. *Intrinsic motivation and self-determination in human behavior*, 11-40.

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, *127*(3), 455.

Dupuis, M., & Renaud, K. (2021). Scoping the ethical principles of cybersecurity fear appeals. *Ethics and Information Technology*, *23*(3), 265-284.

Dupuis, M., & Renaud, K. (2021). Scoping the ethical principles of cybersecurity fear appeals. Ethics and Information Technology, 23(3), 265-284.

Dwivedi, Y. K., Rana, N. P., Chen, H., & Williams, M. D. (2011). A Meta-analysis of the Unified Theory of Acceptance and Use of Technology (UTAUT). In *Governance and Sustainability in Information Systems. Managing the Transfer and Diffusion of IT: IFIP WG 8.6 International Working Conference, Hamburg, Germany, September 22-24, 2011. Proceedings* (pp. 155-170). Springer Berlin Heidelberg.

Egelman, S., & Peer, E. (2015, April). Scaling the security wall: Developing a security behavior intentions scale (sebis). In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (pp. 2873-2882).

El-Din, R. S. (2012, September). To Deceive or Not to Deceive! Ethical Questions in Phishing Research. In The 26th BCS *Conference on Human Computer Interaction 26*. 1-4.

Elliot, A. J., & McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of personality and social psychology*, *80*(3), 501.

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3× 2 achievement goal model. *Journal of educational psychology*, *103*(3), 632.

Ertan, A., Crossland, G., Heath, C., Denny, D., & Jensen, R. (2020). Cyber security behaviour in organisations. arXiv preprint arXiv:2004.11768.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293-315.

Fei, Z., Kassim, N. M., & Mohamad, W. N. (2022). Factors Influencing the Adoption of IoT Based Mobile Health Services in China: A Conceptual Framework. Global Business and Management Research, 14(3s), 1094-1104.

Fernet, C., Senécal, C., Guay, F., Marsh, H., & Dowson, M. (2008). The work tasks motivation scale for teachers (WTMST). *Journal of Career assessment*, *16*(2), 256-279.

Ferreira, A., & Lenzini, G. (2015, July). An analysis of social engineering principles in effective phishing. In *2015 Workshop on Socio-Technical Aspects in Security and Trust* (pp. 9-16). IEEE.

Finn, P., & Jakobsson, M. (2007). Designing ethical phishing experiments. *IEEE Technology and Society Magazine*, *26*(1), 46-58.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment Under Uncertainty: Heuristics and Biases (pp. 422–444). *Science Volume 185.* Cambridge, United Kingdom: Cambridge University Press.

Floyd, D. L., Prentice‐Dunn, S., & Rogers, R. W. (2000). A meta‐analysis of research on protection motivation theory. *Journal of applied social psychology*, *30*(2), 407-429.

Fogg, B. J. (2009, April). A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology* (pp. 1-7).

Fontenelle, G., & Howell, W. C. (1984). *A Replication and Extension of the Inducement of the Availability Heuristic*. RICE UNIV HOUSTON TX DEPT OF PSYCHOLOGY.

Franke, N., & Schreier, M. (2010). Why customers value self-designed products: The importance of process effort and enjoyment. *Journal of product innovation management*, *27*(7), 1020-1031.

Friedman, B., Khan Jr, P. H., & Howe, D. C. (2000). Trust online. *Communications of the ACM*, *43*(12), 34-40.

Furnell, S. M., Alotaibi, F., & Esmael, R. (2019). Aligning security practice with policy: Guiding and nudging towards better behavior. Retrieved from: https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/12764/Paper%202452%20-%20Final.pdf?sequence=1

Gafoor, K. A. (2012). Considerations in the Measurement of Awareness. *Online Submission*.

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., Bellerose, J., Benabou, C., Chemolli, E., Güntert, S. T., Halvari., H., Indiyastuti, D. L., Johnson, P., Molstad, M. H., Naudin, M., Ndao, A., Olafsen, A. H., Roussel., P, Wang, Z., & Westbye., C. The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. European Journal of Work and Organizational Psychology, 24(2), 178-196.

Gamberini, L., Barresi, G., Maier, A., & Scarpetta, F. (2008). A game a day keeps the doctor away: A short review of computer games in mental healthcare. *Journal of CyberTherapy and Rehabilitation*, *1*(2), 127-145.

Gamez, J. (2018). Persuasive design and the web: How Cialdini Principles are used in online successful companies.

Ganti, N., & Baek, S. (2021). Why People Stand By: A Comprehensive Study About the Bystander Effect. *Journal of Student Research*, *10*(1).

Garcia, S. M., Weaver, K., Moskowitz, G. B., & Darley, J. M. (2002). Crowded minds: the implicit bystander effect. *Journal of personality and social psychology*, *83*(4), 843.

Gartner (2023). Gartner Predicts 2023: Cybersecurity Industry Focuses on the Human Deal. Retrieved from: https://www.bitsight.com/resources/gartner-predicts-2023-cybersecurity-industry-focuses-human-deal

Gawronski, B. (2007). Attitudes can be measured! But what is an attitude?. *Social Cognition*, *25*(5), 573-581

Gefen, D., Benbasat, I., & Pavlou, P. (2008). A research agenda for trust in online environments. *Journal of Management Information Systems*, *24*(4), 275-286.

Gekoski, A. (2017). Bystander intervention and the bystander effect. *Psychology Review*, *23*(2017), 20-22.

Ghafir, I., Saleem, J., Hammoudeh, M., Faour, H., Prenosil, V., Jaf, S., Jabber., S & BAKER., T. (2018). Security threats to critical infrastructure: the human factor. *The Journal of Supercomputing*, *74*, 4986-5002.

Gheyas, I. A., & Abdallah, A. E. (2016). Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. *Big data analytics*, *1*(1), 1-29.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, *40*(1), 84-96.

Gratian, M., Bandi, S., Cukier, M., Dykstra, J., & Ginther, A. (2018). Correlating human traits and cyber security behavior intentions. *computers & security*, *73*, 345-358.

Greitzer, F. L., Imran, M., Purl, J., Axelrad, E. T., Leong, Y. M., Becker, D. E., ... & Sticha, P. J. (2016). Developing an Ontology for Individual and Organizational Sociotechnical Indicators of Insider Threat Risk. In STIDS (pp. 19-27).

Grimmer, L., & Grimmer, M. (2020, March). Blind bags: How toy makers are making a fortune with child gambling. In *Phi Kappa Phi Forum* (Vol. 100, No. 1, pp. 10-14). Honor Society of Phi Kappa Phi.

Grudin, J., & Poltrock, S. (2012). Taxonomy and theory in computer supported cooperative work.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion*, *24*, 175-213.

Gupta, M. (2015). A study on employees perception towards employee engagement. *Globsyn Management Journal*, *9*(1/2), 45-51.

Hadlington, L. (2017). Human factors in cybersecurity; examining the link between Internet addiction, impulsivity, attitudes towards cybersecurity, and risky cybersecurity behaviours. *Heliyon*, *3*(7), e00346.

Hadlington, L. J. (2018). Employees attitudes towards cyber security and risky online behaviours: an empirical assessment in the United Kingdom.

Halevi, T., Lewis, J., & Memon, N. (2013, May). A pilot study of cyber security and privacy related behavior and personality traits. In *Proceedings of the 22nd international conference on world wide web* (pp. 737-744).

Harris, A., & Yates, D. (2015). Phishing attacks over time: a longitudinal study (Emergent Research Forum Paper). *Twenty-first Americas Conference on Information Systems, Puerto Rico, 2015*.

Hart, S., Margheri, A., Paci, F., & Sassone, V. (2020). Riskio: A serious game for cyber security awareness and education. *Computers & Security*, *95*, 101827.

Hartman, S. J., Nelson, S. H., & Weiner, L. S. (2018). Patterns of Fitbit use and activity levels throughout a physical activity intervention: exploratory analysis from a randomized controlled trial. *JMIR mHealth and uHealth*, *6*(2), e29.

Hassandoust, F., & Techatassanasoontorn, A. A. (2020). Understanding users' information security awareness and intentions: A full nomology of protection motivation theory. In *Cyber influence and cognitive threats* (pp. 129-143). Academic Press.

Heckhausen, J. E., & Heckhausen, H. E. (2008). *Motivation and action*. Cambridge University Press.

Hedström, K., Kolkowska, E., Karlsson, F., & Allen, J. P. (2011). Value conflicts for information security management. *The Journal of Strategic Information Systems*, *20*(4), 373-384.

Hendrix, M., Al-Sherbaz, A., & Victoria, B. (2016). Game based cyber security training: are serious games suitable for cyber security training?. International Journal of Serious Games, 3(1), 53-61.

Heslin, P. A., & Klehe, U. C. (2006). Self-efficacy. *Encyclopedia Of Industrial/Organizational Psychology, SG Rogelberg, ed*, *2*, 705-708.

Hijji, M., & Alam, G. (2022). Cybersecurity Awareness and Training (CAT) Framework for Remote Working Employees. Sensors, 22(22), 8663.

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin*, *19*(5), 501-512.

Hinnosaar, M., & Hinnosaar, T. (2012). Authority bias. *Retrieved from Academia: https://www. academia. edu/2108445/Authority Bias*.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of personality and social psychology*, *69*(6), 1069.

Hirt, E. R., Kardes, F. R., & Markman, K. D. (2004). Activating a mental simulation mind-set through generation of alternatives: Implications for debiasing in related and unrelated domains. *Journal of Experimental Social Psychology*, *40*(3), 374-383.

Hodas, N. O., & Lerman, K. (2014). The simple rules of social contagion. *Scientific reports*, *4*(1), 4343.

Hortensius, R., & De Gelder, B. (2018). From empathy to apathy: The bystander effect revisited. *Current Directions in Psychological Science*, *27*(4), 249-256.

Hsieh, H. L., Kuo, Y. M., Wang, S. R., Chuang, B. K., & Tsai, C. H. (2017). A study of personal health record user's behavioral model based on the PMT and UTAUT integrative perspective. International journal of environmental research and public health, 14(1), 8.

Hudson, J. M., & Bruckman, A. S. (2004). The bystander effect: A lens for understanding patterns of participation. *The Journal of the Learning Sciences*, *13*(2), 165-195.

Human Layer Security (2023). *The Rise of Human Augmentation*. Retrieved from: https://www.humanlayersecurity.com/blog/the-rise-of-security-augmentation/

IBM (2021). Cost of a Data Breach Report 2022. Retrieved from: https://www.ibm.com/reports/data-breach?lnk=hm

In *Risks and Security of Internet and Systems: 12th International Conference, CRiSIS 2017, Dinard, France, September 19-21, 2017, Revised Selected Papers 12* (pp. 187-202). Springer International Publishing.

Ivanova, A., & Kim, J. Y. (2022). Acceptance and use of mobile banking in Central Asia: Evidence from modified UTAUT model. *The Journal of Asian Finance, Economics and Business*, *9*(2), 217-227.

Jampen, D., Gür, G., Sutter, T., & Tellenbach, B. (2020). Don't click: towards an effective anti-phishing training. A comparative literature review. *Human-centric Computing and Information Sciences*, *10*(1), 1-41.

Jari, M. (2022). An Overview of Phishing Victimization: Human Factors, Training and the Role of Emotions. *arXiv preprint arXiv:2209.11197*.

Jensen, M. L., Wright, R. T., Durcikova, A., & Karumbaiah, S. (2022). Improving phishing reporting using security gamification. Journal of Management Information Systems, 39(3), 793-823.

Jeong, J., Mihelcic, J., Oliver, G., & Rudolph, C. (2019, December). Towards an improved understanding of human factors in cybersecurity. In *2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)* (pp. 338-345). IEEE.

Jeske, D., Coventry, L., Briggs, P., & van Moorsel, A. (2014). Nudging whom how: Nudging whom how: IT proficiency, impulse control and secure behaviour. *In: Personalizing Behavior Change Technologies.* CHI Workshop, 27 April.

Joinson, A., Williams, E. J., & Levordashka, A. (2018, July). PHISHTRAY: A modifiable, open-source email sorting task for research and training applications. In *International Conference on Behavioural and Social Sciences in Security*.

Jolls, C., & Sunstein, C. R. (2006). Debiasing through law. *The Journal of Legal Studies*, *35*(1), 199-242.

Jones, K. S., Lodinger, N. R., Widlus, B. P., Namin, A. S., & Hewett, R. (2021). Do warning message design recommendations address why non-experts do not protect themselves from cybersecurity threats? a review. *International Journal of Human–Computer Interaction*, *37*(18), 1709-1719.

Kabil, A., Duval, T., Cuppens, N., Le Comte, G., Halgand, Y., & Ponchel, C. (2018). 3D cybercop: A collaborative platform for cybersecurity data analysis and training. In *Cooperative Design, Visualization, and Engineering: 15th International Conference, CDVE 2018, Hangzhou, China, October 21–24, 2018, Proceedings 15* (pp. 176-183). Springer International Publishing.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kahneman, D. (2011). Thinking, fast and slow. *Macmillan*.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and psychological measurement*, *34*(1), 111-117.

Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019). A comprehensive survey for intelligent spam email detection. IEEE Access, 7, 168261-168295.

Karim, N. H. A., & Noor, M. N. H. N. M. (2013). Investigating The Correlate And Predictors Of Affective And Continuance Organisational Commitment Among Malaysian Academic Librarian.

Kashdan, T. B., & Silvia, P. J. (2009). Curiosity and interest: The benefits of thriving on novelty and challenge. *Oxford handbook of positive psychology*, *2*, 367-374.

Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The Five-Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt and covert social curiosity. *Personality and Individual Differences*, *157*, 109836.

Kaufmann, L., Michel, A., & Carter, C. R. (2009). Debiasing strategies in supply management decision-making. *Journal of Business Logistics*, *30*(1), 85-106.

Keller, C., Siegrist, M., & Gutscher, H. (2006). The role of the affect and availability heuristics in risk communication. *Risk analysis*, *26*(3), 631-639.

Keller, P. A. (2006). Regulatory focus and efficacy of health messages. *Journal of Consumer Research*, *33*(1), 109-114.

Kessler, S. K., & Martin, M. (2017). How do potential users perceive the adoption of new technologies within the field of Artificial Intelligence and Internet-of-Things?-a revision of the UTAUT 2 model using voice assistants.

Khader, M., Karam, M., & Fares, H. (2021). Cybersecurity awareness framework for academia. Information, 12(10), 417.

Kirk, C. P., & Swain, S. D. (2018). Consumer psychological ownership of digital technology. Psychological ownership and consumer behavior, 69-90.

Kirlappos, I. (2016). *Learning from" shadow security": understanding non-compliant behaviours to improve information security management* (Doctoral dissertation). University College London.

Kirlappos, I., Parkin, S., & Sasse, M. A. (2014). Learning from “Shadow Security”: Why understanding non-compliance provides the basis for effective security.

Kirlappos, I., Parkin, S., & Sasse, M. A. (2015). " Shadow security" as a tool for the learning organization. *Acm Sigcas Computers and Society*, *45*(1), 29-37.

Kleitman, S., Law, M. K., & Kay, J. (2018). It’s the deceiver and the receiver: Individual differences in phishing susceptibility and false positives with item profiling. PloS one, 13(10), e0205089.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Kretschmer, T., Leiponen, A., Schilling, M., & Vasudeva, G. (2022). Platform ecosystems as Metaorganizations: Implications for platform strategies. Strategic Management Journal, 43(3), 405–424

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security and applications*, *22*, 113-122.

Krupić, D., Žuro, B., & Corr, P. J. (2021). Anxiety and threat magnification in subjective and physiological responses of fear of heights induced by virtual reality. Personality and Individual Differences, 169, 109720.

Kshetri, N., & Chhetri, M. (2022). Gender asymmetry in cybersecurity: socioeconomic causes and consequences. *Computer*, *55*(2), 72-77.

Kulshrestha, S., Agrawal, S., Gaurav, D., Chaturvedi, M., Sharma, S., & Bose, R. (2021). Development and validation of serious games for teaching cybersecurity. In *Serious Games: Joint International Conference, JCSG 2021, Virtual Event, January 12–13, 2022, Proceedings 7* (pp. 247-262). Springer International Publishing.

Kumaraguru, P., Cranshaw, J., Acquisti, A., Cranor, L., Hong, J., Blair, M. A., & Pham, T. (2009, July). School of phish: a real-world evaluation of anti-phishing training. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (pp. 1-12).

Kuraku, S. (2022). *Curiosity Clicks: The Need for Security Awareness* (Doctoral dissertation, University of the Cumberlands).

Kwak, Y., Lee, S., Damiano, A., & Vishwanath, A. (2020). Why do users not report spear phishing emails?. *Telematics and Informatics*, *48*, 101343.

Laplante, P. A., & Laplante, N. (2016). The internet of things in healthcare: Potential applications and challenges. *It Professional*, *18*(3), 2-4.

Larrick, R. P. (2004). Debiasing. *Blackwell handbook of judgment and decision making*, 316-338.

Lawson, P. A., Crowson, A. D., & Mayhorn, C. B. (2019). Baiting the hook: Exploring the interaction of personality and persuasion tactics in email phishing attacks. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume V: Human Simulation and Virtual Environments, Work With Computing Systems (WWCS), Process Control 20* (pp. 401-406). Springer International Publishing.

Lawson, P., Pearson, C. J., Crowson, A., & Mayhorn, C. B. (2020). Email phishing and signal detection: How persuasion principles and personality influence response patterns and accuracy. *Applied ergonomics*, *86*, 103084.

Lee, S., Atkinson, L., & Sung, Y. H. (2022). Online bandwagon effects: Quantitative versus qualitative cues in online comments sections. *New Media & Society*, *24*(3), 580-599.

Lee, Y., & Chen, A. N. (2011). Usability design and psychological ownership of a virtual world. *Journal of Management Information Systems*, *28*(3), 269-308.

Lewis, I. M., Watson, B., & White, K. M. (2010). Response efficacy: The key to minimizing rejection and maximizing acceptance of emotion-based anti-speeding messages. *Accident Analysis & Prevention*, *42*(2), 459-467.

Li, Y., & Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy Reports*, *7*, 8176-8186.

Liang, H., & Xue, Y. (2009). Avoidance of information technology threats: A theoretical perspective. *MIS quarterly*, 71-90.

Liang, N. P., Biros, D., & Luse, A. (2016). Taxonomy of Malicious Insiders: A Proof of Concept Study.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, *116*(1), 75.

Loewenstein, G., & Lerner, J. S. (2003). The role of affect in decision making.

Loske, A., Widjaja, T., & Buxmann, P. (2013). Cloud computing providers' unrealistic optimism regarding IT security risks: A threat to users?. *Thirty Fourth International Conference on Information Systems, Milan 2013*.

Lowry, P. B., Moody, G. D., Parameswaran, S., & Brown, N. (2023). Examining the Differential Effectiveness of Fear Appeals in Information Security Management Using Two-Stage Meta-Analysis. Journal of Management Information Systems (JMIS)(10-Apr-2023).

Ludolph, R., & Schulz, P. J. (2018). Debiasing health-related judgments and decision making: a systematic review. *Medical Decision Making*, *38*(1), 3-13.

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.

Ma, M., & Zheng, H. (2011). Virtual reality and serious games in healthcare. *Advanced computational intelligence paradigms in healthcare 6. Virtual reality in psychotherapy, rehabilitation, and assessment*, 169-192.

Maclnnis, D. J., Moorman, C., & Jaworski, B. J. (1991). Enhancing and measuring consumers' motivation, opportunity, and ability to process brand information from ads. *Journal of marketing, 55*(4), 32-53.

Maddux, J. E., & Gosselin, J. T. (2012). *Self-efficacy*. The Guilford Press.

Maio, G. R., & Haddock, G. (2007). Attitude change. *Social psychology: Handbook of basic principles*, *2*, 565-586.

Malkin, N., Mathur, A., Harbach, M., & Egelman, S. (2017, April). Personalized security messaging: Nudges for compliance with browser warnings. In *2nd european workshop on usable security. internet society*.

Mamra, A., Sibghatullah, A. S., Ananta, G. P., Alazzam, M. B., Ahmed, Y. H., & Doheir, M. (2017). A proposed framework to investigate the user acceptance of personal health records in Malaysia using UTAUT2 and PMT. International Journal of Advanced Computer Science and Applications, 8(3).

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, *14*, 81-95.

Markey, P. M. (2000). Bystander intervention in computer-mediated communication. *Computers in Human Behavior*, *16*(2), 183-188.

Markey, R., & Townsend, K. (2013). Contemporary trends in employee involvement and participation. *Journal of Industrial Relations*, *55*(4), 475-487.

Marton, F. (2000). The structure of awareness. *Phenomenography*, *10216*, 102-116.

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. International journal of psychological research, 3(1), 97-110.

Matz, S. C., Segalin, C., Stillwell, D., Müller, S. R., & Bos, M. W. (2019). Predicting the personal appeal of marketing images using computational methods. *Journal of Consumer Psychology*, *29*(3), 370-390.

McCormac, A., Zwaans, T., Parsons, K., Calic, D., Butavicius, M., & Pattinson, M. (2017). Individual differences and information security awareness. Computers in Human Behavior, 69, 151-156.

McGill, T., & Thompson, N. (2017). Old risks, new challenges: exploring differences in security between home computer and mobile device use. *Behaviour & Information Technology*, *36*(11), 1111-1124.

Mele, C., Spena, T. R., Kaartemo, V., & Marzullo, M. L. (2021). Smart nudging: How cognitive technologies enable choice architectures for value co-creation. *Journal of Business Research*, *129*, 949-960.

Menard, P., Bott, G. J., & Crossler, R. E. (2017). User motivations in protecting information security: Protection motivation theory versus self-determination theory. *Journal of Management Information Systems*, *34*(4), 1203-1230.

Metalidou, E., Marinagi, C., Trivellas, P., Eberhagen, N., Skourlas, C., & Giannakopoulos, G. (2014). The human factor of information security: Unintentional damage perspective. Procedia-Social and Behavioral Sciences, 147, 424-428.

Meyer, J. P., & Allen, N. J. (1991). A three-component conceptualization of organizational commitment. *Human resource management review*, *1*(1), 61-89.

Michie, S. F., West, R., Campbell, R., Brown, J., & Gainforth, H. (2014). *ABC of behaviour change theories*. Silverback publishing.

Milgram, S. (1974). Obedience to Authority: An Experimental View. *Harper & Row*.

Milne, S., Sheeran, P., & Orbell, S. (2000). Prediction and intervention in health-related behavior: A meta-analytic review of protection motivation theory. *Journal of applied social psychology*, *30*(1), 106-143.

Moody, G. D., Galletta, D. F., & Dunn, B. K. (2017). Which phish get caught? An exploratory study of individuals′ susceptibility to phishing. *European Journal of Information Systems*, *26*, 564-584.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129-140.

Morgan, P. L., & Asquith, P. M. (2021a). Airbus cyber securitycybersecurity behaviours tool: Experimental findings and recommendations. *Airbus Cyber Lab Pillar II Internal Report. Airbus and Endeavr Wales*, 1-76.

Morgan, P. L., & Asquith, P. M. (2021b). Black Box Thinking for company-wide cyber securitycybersecurity: A dynamic, integrated and human-user driven perspective. *Airbus and Eneavr Wales Internal Deliverable*. 1-69.

Morgan, P. L., Asquith, P. M., Bishop, L. M., Raywood-Burke, G., Wedgbury, A., & Jones, K. (2020). A new hope: human-centric cybersecurity research embedded within organizations. In *HCI for Cybersecurity, Privacy and Trust: Second International*

Conference, HCI-CPT 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22 (pp. 206-216). Springer International Publishing.

Morrison, B., Coventry, L., & Briggs, P. (2021). How do Older Adults feel about engaging with Cyber-Security?. *Human Behavior and Emerging Technologies*, *3*(5), 1033-1049.

Moschovitis, C. (2019). Why do cyber security programmes fail?. *Cyber Security: A Peer-Reviewed Journal*, *2*(4), 303-309.

Mowrer, O. H. (1939). A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychological review, 46*(6), 553.

Mughal, A. A. (2020). Cyber Attacks on OSI Layers: Understanding the Threat Landscape. *Journal of Humanities and Applied Science Research*, *3*(1), 1-18.

Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin*, *26*(9), 1142-1150.

Nagai, H., Nakazawa, E., & Akabayashi, A. (2022). The creation of the Belmont Report and its effect on ethical principles: a historical study. *Monash Bioethics Review*, 1-14.

Naik, B., Mehta, A., Yagnik, H., & Shah, M. (2022). The impacts of artificial intelligence techniques in augmentation of cybersecurity: a comprehensive review. *Complex & Intelligent Systems*, *8*(2), 1763-1780.

Nakayama, S., Echizen, I., & Yoshiura, H. (2009, September). Preventing false positives in content-based phishing detection. In 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 48-51). IEEE.

Naqshbandi, M. M., Tabche, I., & Choudhary, N. (2019). Managing open innovation: The roles of empowering leadership and employee involvement climate. *Management Decision*, *57*(3), 703-723. NCSC. (2020). *Annual Review 2020*. (Retrieved from: https://www.ncsc.gov.uk/files/Annual-Review-2020.pdf.

NCSC (2021). *NCSC Annual Review 2021*. Retrieved from: https://www.ncsc.gov.uk/collection/ncsc-annual-review-2021

NCSC (2022). *NCSC Annual Review 2022*. Retrieved from: https://www.ncsc.gov.uk/collection/annual-review-2022?trk=public_post-text

Neves, J. (2004). Organisational Climate and Culture. *In: Manual of Psychosociology of the Organisations, Ferreira, J.M.C., J. Neves and A. Caetano (Eds.)*. McGraw-Hill Portugal, Lisbon, pp: 431–468.

Nicholls, J. G. (1984). Achievement motivation: conceptions of ability, subjective experience, task choice, and performance. *Psychological review*, *91*(3), 328.

Nichols, T. W., & Erakovich, R. (2013). Authentic leadership and implicit theory: a normative form of leadership?. *Leadership & Organization Development Journal*.

Nickerson, J. A., Wuebker, R., & Zenger, T. (2017). Problems, theories, and governing the crowd. Strategic Organization, 15(2), 275–288

Nickols, F. (2000). The knowledge in knowledge management. *The Knowledge Management Yearbook, 2000–2001*, 12, 21.

Nisbet, E., & Weiss, R. (2010). Top-down versus bottom-up. Science, 328(5983), 1241-1243.

NIST Cybersecurity (2023). *Cybersecurity Glossary*. Retrieved from: https://csrc.nist.gov/glossary/term/cybersecurity

Nobles, C. (2022). Stress, Burnout, and Security Fatigue in Cybersecurity: A Human Factors Problem. *HOLISTICA–Journal of Business and Public Administration, 13*(1), 49-72.

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of consumer psychology, 22*(3), 453-460.

Obiekwe, O., Zeb-Obipi, I., & Ejo-Orusa, H. (2019). Employee involvement in organizations: benefits, challenges and implications. *Management and Human Resource Research Journal*, *8*(8), 1-11.

Oh, J. C., & Yoon, S. J. (2014). Predicting the use of online information services based on a modified UTAUT model. *Behaviour & Information Technology*, *33*(7), 716-729.

Oliveira, D. (2017). Security for Vulnerable {Populations—On} the Interplay of Weapons of Influence and Life Domains in Predicting Older Adults' Susceptibility to {Spear-Phishing} Emails. Retrieved from: https://www.usenix.org/conference/enigma2017/conference-program/presentation/oliveira

Osborne, S., & Hammoud, M. S. (2017). Effective employee engagement in the workplace. *International Journal of Applied Management and Technology, 16*(1), 4.

Parsons, K., Butavicius, M., Delfabbro, P., & Lillie, M. (2019). Predicting susceptibility to social influence in phishing emails. *International Journal of Human-Computer Studies*, *128*, 17-26.

Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., & Jerram, C. (2013). Phishing for the truth: A scenario-based experiment of users' behavioural response to emails. In *Security and Privacy Protection in Information Processing Systems: 28th IFIP TC 11 International Conference, SEC 2013, Auckland, New Zealand, July 8-10, 2013. Proceedings 28* (pp. 366-378). Springer Berlin Heidelberg.

Patterson, J. (2017). Cyber-security policy decisions in small businesses (Doctoral dissertation, Walden University).

Peck, J., Kirk, C. P., Luangrath, A. W., & Shu, S. B. (2021). Caring for the commons: Using psychological ownership to enhance stewardship behavior for public goods. *Journal of Marketing, 85*(2), 33-49.

Petelka, J., Zou, Y., & Schaub, F. (2019, May). Put your warning where your link is: Improving and evaluating email phishing warnings. *In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-15).*

Peters, E. M., Burraston, B., & Mertz, C. K. (2004). An emotion-based model of risk perception and stigma susceptibility: Cognitive appraisals of emotion, affective reactivity, worldviews, and risk perceptions in the generation of technological stigma. *Risk Analysis: An International Journal*, *24*(5), 1349-1367.

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

Pfleeger, S. L., & Caputo, D. D. (2012). Leveraging behavioral science to mitigate cyber security risk. *Computers & security*, *31*(4), 597-611.

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.

Phish Protection (2021). History of Phishing. Retrieved from: https://www.phishprotection.com/resources/history-of-phishing

Pickens, J. (2005). Attitudes and perceptions. *Organizational behavior in health care*, *4*(7), 43-76.

Pickering, B., Boletsis, C., Halvorsrud, R., Phillips, S., & Surridge, M. (2021, July). It's not my problem: how healthcare models relate to SME cybersecurity awareness. In *HCI for Cybersecurity, Privacy and Trust: Third International Conference, HCI-CPT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings* (pp. 337-352). Cham: Springer International Publishing.

Pirocca, S., Allodi, L., & Zannone, N. (2020). A Toolkit for Security Awareness Training Against Targeted Phishing. In *Information Systems Security: 16th International Conference, ICISS 2020, Jammu, India, December 16–20, 2020, Proceedings 16* (pp. 137-159). Springer International Publishing.

Posey, C., Roberts, T. L., & Lowry, P. B. (2015). The impact of organizational commitment on insiders' motivation to protect organizational information assets. *Journal of Management Information Systems*, *32*(4), 179-214.

Prakash, P., Kumar, M., Kompella, R. R., & Gupta, M. (2010, March). Phishnet: predictive blacklisting to detect phishing attacks. In 2010 Proceedings IEEE INFOCOM (pp. 1-5). IEEE.

Prentice-Dunn, S., & Rogers, R. W. (1986). Protection motivation theory and preventive health: Beyond the health belief model. *Health education research*, *1*(3), 153-161.

Proctor, R. W., & Chen, J. (2015). The role of human factors/ergonomics in the science of security: decision making and action selection in cyberspace. *Human factors*, *57*(5), 721-727.

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. *Trends in cognitive sciences*, *13*(10), 420-428.

Rachman, S. (1976). The passing of the two-stage theory of fear and avoidance: Fresh possibilities. *Behaviour Research and Therapy*, *14*(2), 125-131.

Raddatz, N. I., Coyne, J. G., & Trinkle, B. S. (2020). Internal motivators for the protection of organizational data. Journal of Information Systems, 34(3), 199-211.

Rahi, S., Khan, M. M., & Alghizzawi, M. (2021). Factors influencing the adoption of telemedicine health services during COVID-19 pandemic crisis: an integrative research model. Enterprise Information Systems, 15(6), 769-793.

Raineri, E. M., & Resig, J. (2020). Evaluating Self-Efficacy Pertaining to Cybersecurity for Small Businesses. Journal of Applied Business & Economics, 22(12).

Rajivan, P., & Gonzalez, C. (2018). Creative persuasion: a study on adversarial behaviors and strategies in phishing attacks. *Frontiers in psychology*, *9*, 135.

Reegård, K., Blackett, C., & Katta, V. (2019). The concept of cybersecurity culture. In *29th European Safety and Reliability Conference.* 4036-4043.

Renaud, K., Otondo, R., & Warkentin, M. (2019). "This is the way 'I'create my passwords"... does the endowment effect deter people from changing the way they create their passwords?. *Computers & Security*, *82*, 241-260.

Renaud, K., Searle, R., & Dupuis, M. (2021, October). Shame in cyber security: effective behavior modification tool or counterproductive foil?. In *New Security Paradigms Workshop* (pp. 70-87).

Rhee, H. S., Ryu, Y., & Kim, C. T. (2005). I am fine but you are not: Optimistic bias and illusion of control on information security. *ICIS 2005 proceedings*, 32.

Ringeval, M., Wagner, G., Denford, J., Paré, G., & Kitsiou, S. (2020). Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. *Journal of medical Internet research*, *22*(10), e23954.

Rizzoni, F., Magalini, S., Casaroli, A., Mari, P., Dixon, M., & Coventry, L. (2022). Phishing simulation exercise in a large hospital: A case study. *Digital Health*, *8,* 20552076221081716.

Roberts, B. W., Lejuez, C., Krueger, R. F., Richards, J. M., & Hill, P. L. (2014). What is conscientiousness and how can it be assessed?. *Developmental psychology*, *50*(5), 1315.

Rogers, R. W. (1975). A protection motivation theory of fear appeals and attitude change1. *The journal of psychology*, *91*(1), 93-114.

Rohan, R., Pal, D., Hautamäki, J., Funilkul, S., Chutimaskul, W., & Thapliyal, H. (2023). A systematic literature review of cybersecurity scales assessing information security awareness. *Heliyon*.

Rosén, J., Kastrati, G., Reppling, A., Bergkvist, K., & Åhs, F. (2019). The effect of immersive virtual reality on proximal and conditioned threat. Scientific reports, 9(1), 17407.

Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health education monographs*, *2*(4), 328-335.

Rosenstock. (1990). The health belief model: Explaining health behavior through experiences. *Health Behavior and Health Education: Theory, Research and Practice*. 39-63.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Safa, N. S., Sookhak, M., Von Solms, R., Furnell, S., Ghani, N. A., & Herawan, T. (2015). Information security conscious care behaviour formation in organizations. *Computers & Security*, *53*, 65-78.

Salazar, M., Gaviria, J., Laorden, C., & Bringas, P. G. (2013, March). Enhancing cybersecurity learning through an augmented reality-based serious game. In *2013 IEEE global engineering education conference (EDUCON)* (pp. 602-607). IEEE.

SANS (2021). 2021 Security Awareness Report. Retrieved from: https://www.sans.org/security-awareness-training/resources/reports/sareport-2021/

Sarno, D. M., Harris, M. W., & Black, J. (2023). Which phish is captured in the net? Understanding phishing susceptibility and individual differences. *Applied Cognitive Psychology*.

Sasse, M. A., Brostoff, S., & Weirich, D. (2001). Transforming the 'weakest link'—a human/computer interaction approach to usable and effective security. *BT technology journal*, *19*(3), 122-131.

Schein, E. H. (2010). *Organizational culture and leadership* (Vol. 2). John Wiley & Sons.

Schenk, D. H. (2011). Exploiting the salience bias in designing taxes. *Yale J. on Reg.*, *28*, 253.

Scherer, C. W., & Cho, H. (2003). A social network contagion theory of risk perception. *Risk Analysis: An International Journal*, *23*(2), 261-267.

Scholefield, S., & Shepherd, L. A. (2019). Gamification techniques for raising cyber security awareness. In *HCI for Cybersecurity, Privacy and Trust: First International Conference, HCI-CPT 2019, Held as Part of the 21st HCI International Conference, HCII 2019,* *Orlando, FL, USA, July 26–31, 2019, Proceedings 21* (pp. 191-203). Springer International Publishing.

Scholl, M. C., Fuhrmann, F., & Scholl, L. R. (2018). Scientific knowledge of the human side of information security as a basis for sustainable trainings in organizational practices.

Schuetz, S. W., Benjamin Lowry, P., Pienta, D. A., & Bennett Thatcher, J. (2020). The effectiveness of abstract versus concrete fear appeals in information security. *Journal of Management Information Systems*, *37*(3), 723-757.

Schutte, N. S., & Malouff, J. M. (2020). Connections between curiosity, flow and creativity. *Personality and individual differences*, *152*, 109555.

Schutte, N. S., & Malouff, J. M. (2020). Connections between curiosity, flow and creativity. *Personality and Individual Differences, 152*, 109555.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and psychological measurement*, *55*(5), 818-831.

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological science*, *30*(9), 1371-1379.

Serpa, S. (2016). An overview of the concept of organisational culture. *International business management*, *10*(1), 51-61.

Shaari, R., Rahman, S. A. A., & Rajab, A. (2014). Self-efficacy as a determined factor for knowledge sharing awareness. *International Journal of Trade, Economics and Finance*, *5*(1), 39.

Shah, J., Shah, R., & Liang, P. (2021). Too Much Nudging: Can it Cause a Decrease in the Desired Response?. *Journal of Student Research*, *10*(4).

Shaikh, F. A., & Siponen, M. (2023). Information security risk assessments following cybersecurity breaches: The mediating role of top management attention to cybersecurity. *Computers & Security*, *124*, 102974.

Shalev, E., Keil, M., Lee, J. S., & Ganzach, Y. (2014). Optimism bias in managing it project risks: a construal level theory perspective. *ECIS 2014 Proceedings.*

Sharot, T. (2011). The optimism bias. *Current biology*, *21*(23), 941-945.

Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L. F., & Downs, J. (2010, April). Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 373-382).

Shillair, R., & Dutton, W. H. (2016). Supporting a cybersecurity mindset: getting internet users into the cat and mouse game. *Available at SSRN 2756736*.

Shin, D. D., & Kim, S. I. (2019). Homo curious: Curious or interested? *Educational Psychology Review, 31(4),* 853-874.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 99-118.

Skinner, T., Taylor, J., Dale, J., & McAlaney, J. (2018, April). The development of intervention e-learning materials and implementation techniques for cyber-security behaviour change. ACM SIG CHI.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, *177*(3), 1333-1352.

Snyman, D., & Kruger, H. (2021). Group Behavior in Cybersecurity. *Encyclopedia of Cryptography, Security and Privacy*.

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. *The Wiley Blackwell handbook of judgment and decision making*, *2*, 924-951.

Standage, M., & Treasure, D. C. (2002). Relationship among achievement goal orientations and multidimensional situational motivation in physical education. *British Journal of Educational Psychology*, *72*(1), 87-103.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Sulaiman, N. S., Fauzi, M. A., Wider, W., Rajadurai, J., Hussain, S., & Harun, S. A. (2022). Cyber–information security compliance and violation behaviour in organisations: A systematic review. Social Sciences, 11(9), 386.

Sun, Y., Wang, N., Guo, X., & Peng, Z. (2013). Understanding the acceptance of mobile health services: a comparison and integration of alternative models. Journal of electronic commerce research, 14(2), 183.

Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy*, *37*, 583-588.

Sunstein, C. R. (2020). *Too much information: understanding what you don't want to know*. MIT Press.

Swire, P. (2018). A pedagogic cybersecurity framework. *Communications of the ACM*, *61*(10), 23-26.

Taherdoost, H. (2022). Cybersecurity vs. Information Security. *Procedia Computer Science*, *215*, 483-487.

Taillard, M. O. (2000). Persuasive communication: the case of marketing. Working Papers in Linguistics, 12, 145-174.

Tannenbaum, M. B., Hepler, J., Zimmerman, R. S., Saul, L., Jacobs, S., Wilson, K., & Albarracín, D. (2015). Appealing to fear: A meta-analysis of fear appeal effectiveness and theories. Psychological bulletin, 141(6), 1178.

Taylor-Gooby, P., & Zinn, J. O. (2006). Current directions in risk research: New developments in psychology and sociology. *Risk Analysis: An International Journal*, *26*(2), 397-411.

Techtarget (2023). *34 Cybersecurity Statistics to Lose Sleep Over in 2023*. Retrieved from: https://www.techtarget.com/whatis/34-Cybersecurity-Statistics-to-Lose-Sleep-Over-in-2023.

Thaler, R. H., & Sunstein, C. R. (2008). Nudge: Improving decisions about health, wealth, and happiness. In Thaler, R. H., & Sunstein, C. R. (eds). *Nudge: Improving Decisions about Health, Wealth, and Happiness.* Penguin.

Thomas, J. (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. *Thomas, JE (2018). Individual cyber security: Empowering employees to resist spear phishing to prevent identity theft and ransomware attacks. International Journal of Business Management*, *12*(3), 1-23.

Thompson, R. F. (2009). Habituation: a history. *Neurobiology of learning and memory*, *92*(2), 127.

Tieben, R., Bekker, T., & Schouten, B. (2011, July). Curiosity and interaction: making people curious through interactive systems. In *Proceedings of HCI 2011 The 25th BCS Conference on Human Computer Interaction 25.* 361-370.

Tiwari, P. (2020). *Exploring Phishing Susceptibility Attributable To Authority, Urgency, Risk Perception And Human Factors* (Doctoral dissertation). Purdue University Graduate School.

Touré-Tillery, M., & Fishbach, A. (2014). How to measure motivation: A guide for the experimental social psychologist. *Social and Personality Psychology Compass*, *8*(7), 328-341.

Touré-Tillery, M., & Fishbach, A. (2018). Three sources of motivation. *Consumer Psychology Review*, *1*(1), 123-134.

Trevethan, R. (2017). Deconstructing and assessing knowledge and awareness in public health research. *Frontiers in public health*, *5*, 194.

Troja, E., DeBello, J. E., & Truong, L. M. (2023, March). Teaching Effective and Gamified Cybersecurity using the Metaverse: Challenges and Opportunities. In *2023 IEEE World Engineering Education Conference (EDUNINE)* (pp. 1-6). IEEE.

Turland, J., Coventry, L., Jeske, D., Briggs, P., & Van Moorsel, A. (2015, July). Nudging towards security: Developing an application for wireless network selection for android phones. In *Proceedings of the 2015 British HCI conference* (pp. 193-201).

USwitch (2023). *Online Gaming Statistics Report*. Retrieved from: https://www.uswitch.com/broadband/studies/online-gaming-statistics/

Vallerand, R. J. (1997). Toward a hierarchical model of intrinsic and extrinsic motivation. *Advances in experimental social psychology*, *29*, 271-360.

Van Bavel, R., Rodríguez-Priego, N., Vila, J., & Briggs, P. (2019). Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies*, *123*, 29-39.

Van Bommel, M., Van Prooijen, J. W., Elffers, H., & Van Lange, P. A. (2012). Be aware to care: Public self-awareness leads to a reversal of the bystander effect. *Journal of Experimental Social Psychology, 48*(4), 926-930.

Van den Broeck, A., Vansteenkiste, M., De Witte, H., Soenens, B., & Lens, W. (2010). Capturing autonomy, competence, and relatedness at work: Construction and initial validation of the Work‐related Basic Need Satisfaction scale. *Journal of occupational and organizational psychology*, *83*(4), 981-1002.

Van Dyne, L., & Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 25*(4), 439-459.

Van Steen, T., & Deeleman, J. R. (2021). Successful gamification of cybersecurity training. *Cyberpsychology, Behavior, and Social Networking*, *24*(9), 593-598.

Venema, T. A., Kroese, F. M., & De Ridder, D. T. (2018). I'm still standing: A longitudinal study on the effect of a default nudge. *Psychology & Health*, *33*(5), 669-681.

Venkatesh, V. (2022). Adoption and use of AI tools: a research agenda grounded in UTAUT. *Annals of Operations Research*, 1-12.

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences, 39*(2), 273-315.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly,* 425-478.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.

Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.

Verizon (2019). *2019 Data Breach Investigations Report*. Retrieved from: https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report-emea.pdf.

Verizon (2020). *2020 Data Breach Investigations Report*. Retrieved from: https://enterprise.verizon.com/en-gb/resources/reports/2020-data-breach-investigations-report.pdf.

Verizon (2021). *2021 Data Breach Investigations Report*. Retrieved from: https://www.verizon.com/business/resources/reports/dbir/2021/masters-guide/

Verizon (2022). 2022 Data Breach Investigations Report. Retrieved from: https://www.verizon.com/business/resources/reports/dbir/

Verkijika, S. F. (2020, November). Employees' Cybersecurity Behaviour in the Mobile Context: The Role of Self-Efficacy and Psychological Ownership. In 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 1-5). IEEE.

Vishwanath, A., Harrison, B., & Ng, Y. J. (2018). Suspicion, cognition, and automaticity model of phishing susceptibility. *Communication Research, 45*(8), 1146-1166.

Von Solms, R., & Van Niekerk, J. (2013). From information security to cyber security. *computers & security, 38*, 97-102.

Waddell, T. F., & Sundar, S. S. (2020). Bandwagon effects in social television: How audience metrics related to size and opinion affect the enjoyment of digital media. *Computers in Human Behavior, 107*, 106270.

Wang, J. L., Jackson, L. A., Wang, H. Z., & Gaskin, J. (2015). Predicting social networking site (SNS) use: Personality, attitudes, motivation and internet self-efficacy. *Personality and Individual Differences*, *80*, 119-124.

Wang, Y., Qi, B., Zou, H. X., & Li, J. X. (2018, October). Framework of raising cyber security awareness. In 2018 IEEE 18th International Conference on Communication Technology (ICCT) (pp. 865-869). IEEE.

Wanner, J., Herm, L. V., Heinrich, K., & Janiesch, C. (2022). The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets, 32*(4), 2079-2102.

Warkentin, M., Xu, Z., & Mutchler, L. A. (2013). *I'm safer than you: the role of optimism bias in personal IT risk assessments*. In Proceedings of (pp. 1-32).

Watkins, M. W. (2021). A step-by-step guide to exploratory factor analysis with SPSS. Routledge.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology, 39*(5), 806.

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health psychology, 14*(2), 132.

Weitl-Harms, S., Spanier, A., Hastings, J., & Rokusek, M. (2023). A systematic mapping study on gamification applications for undergraduate cybersecurity education. *Journal of Cybersecurity Education, Research and Practice, 2023*(1), 9.

Welford, A. T. (1965). Stress and achievement. *Australian Journal of Psychology, 17*(1), 1-11.

White, M. J., Cunningham, L. C., & Titchener, K. (2011). Young drivers' optimism bias for accident risk and driving skill: *Accountability and insight experience manipulations. Accident Analysis & Prevention, 43*(4), 1309-1315.

Whitty, M., Doodson, J., Creese, S., & Hodges, D. (2015). Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking, 18*(1), 3-7.

Whitty, M., Doodson, J., Creese, S., & Hodges, D. (2015). Individual differences in cyber security behaviors: an examination of who is sharing passwords. *Cyberpsychology, Behavior, and Social Networking*, *18*(1), 3-7.

Wikipedia (2023). *Computer Security*. Retrieved from: https://en.wikipedia.org/wiki/Computer_security

Willcoxson, L., & Millett, B. (2000). The management of organisational culture. *Australian Journal of Management and Organisational Behaviour*, *3*(2), 91-99.

Williams, E. J., Hinds, J., & Joinson, A. N. (2018). Exploring Susceptibility to Phishing in the Workplace. *International Journal of Human-Computer Studies, 120*, 1-13.

Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health education & behavior, 27*(5), 591-615.

Yang, N., Singh, T., & Johnston, A. (2020). A Replication Study of User Motivation in Protecting Information Security using Protection Motivation Theory and Self Determination Theory. *AIS Transactions on Replication Research, 6*(1), 10.

Yeoh, W., Huang, H., Lee, W. S., Al Jafari, F., & Mansson, R. (2022). Simulated phishing attack and embedded training campaign. *Journal of Computer Information Systems, 62*(4), 802-821.

Zainal, H. Y. (2022). *Examining the factors affecting users' cybersecurity behaviour in mobile payment contactless technologies: A hybrid SEM-ANN approach* (Doctoral dissertation, The British University in Dubai (BUiD)).

Zhao, Q., Chen, C. D., & Wang, J. L. (2016). The effects of psychological ownership and TAM on social media loyalty: An integrated model. Telematics and Informatics, 33(4), 959-972.

Zhou, T., Lu, Y., & Wang, B. (2010). Integrating TTF and UTAUT to explain mobile banking user adoption. *Computers in human behavior*, *26*(4), 760-767.

Zielinska, O. A., Welk, A. K., Mayhorn, C. B., & Murphy-Hill, E. (2016, September). A temporal analysis of persuasion principles in phishing emails. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 60, No. 1, pp. 765-769). Sage CA: Los Angeles, CA: SAGE Publications.

Zimmermann, V., & Renaud, K. (2021). The nudge puzzle: matching nudge interventions to cybersecurity decisions. *ACM Transactions on Computer-Human Interaction (TOCHI), 28(1),* 1-45.

Zimmermann, V., & Renaud, K. (2021). The nudge puzzle: matching nudge interventions to cybersecurity decisions. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *28*(1), 1-45.

Zwilling, M., Klien, G., Lesjak, D., Wiechetek, Ł., Cetin, F., & Basim, H. N. (2022). Cyber security awareness, knowledge and behavior: A comparative study. *Journal of Computer Information Systems*, *62*(1), 82-97.

Appendices

Appendix A: Study 1-3 Introduction Sheet

School of Psychology, Cardiff University

Participant Information Sheet

How Individual Differences and Environmental Context Predict Cyber-Security Perceptions and Behaviours

Introduction

You are being invited to take part in a research experiment. Before you decide to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please contact and ask the lead researcher if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part and thank you for reading this.

This research is being conducted by Laura Bishop (Psychology PhD Student, Cardiff University and Airbus), Dr Phil Morgan (Reader / Associate Professor in Cognitive Science and Human Factors Excellence (HuFEx) Research Group Director at Cardiff University, and, Senior Researcher and Technical Lead in Cyber Psychology and Human Factors at Airbus)

What is the purpose of this study?

The main objective of this survey is to better understand how specific human characteristics can predict cyber-security behaviours as well as how they may differ across different contexts. Findings from this survey will help improve insight into how cyber-security interventions can be tailored to both the individual and the environment helping mitigate the number and severity of attacks experienced.

How will the research be done and what will I have to do?

The survey should take no longer than 60 minutes (Study 1) or 20 minutes (experiments 2 and 3) and will commence after you have made an informed decision to participate or not, through informed consent. Once you have completed the survey you will be debriefed and thanked for taking part.

Why have I been asked and do I have to take part?

Due to the nature of the study, participants are required to be >18 years old and will need to have normal or normal-corrected vision. An interpreter and/or translator is not available for these experiments and therefore a good level of the English language is required.

Your participation is completely voluntary, and it is up to you to decide whether or not you would like to take part. If you do decide to participate you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part, you are still free to withdraw at any time without giving any reason. However, after your participation in the study you may only request the withdrawal of your data up to the point of data analysis (15 working days after the experiment). After this point it may be difficult to trace your data, and the removal of your data may possibly impact the ongoing data analysis and the write-up of the project.

Is this information confidential and held securely, and what will be done with the results?

The personal information collected in this research project (e.g., any form/questionnaire/survey) will be processed by the University in accordance with the terms and conditions of the 1998 Data Protection Act, GDPR regulations (2018). We will hold your data securely and not make it available to any third party unless permitted or required to do so by law. Your personal information will be used/processed as described within this information sheet. All the data you provide will be stored in password-protected computer files under an anonymous identifier and used on a confidential basis.

Non-personal data collected from your questionnaire responses will be held if it retains research value although this will not exceed 7 years. The anonymised data may be made available for further appropriately approved research at the University. No data will be published in a way that could lead to the identification of the individual participants.

The findings from this research may be used in publications in academic journals and also presented at academic conferences. Your personal details will never be included in any of these publications, and your data will only be used anonymously.

Do you have any further questions?

If you have questions about the research - either now or at some future date - please contact either:

Laura Bishop (Lead Researcher): bishoplm2@cardiff.ac.uk

Dr Phil Morgan (Lead Researcher): morganphil@cardiff.ac.uk

Privacy Notice:

The information provided on the consent form will be held in compliance with GDPR regulations. Cardiff University is the data controller and Matt Cooper is the data protection officer (inforequest@cardiff.ac.uk). This information is being collected by Dr Phil Morgan. This information will be held securely and separately from the research information you provide. Only the researcher will have access to this form and it will be destroyed after 7 years. The lawful basis for processing this information is public interest.

The data controller is Cardiff University and the Data Protection Officer is Matt Cooper CooperM1@cardiff.ac.uk. The lawful basis for the processing of the data you provide is consent.

Appendix B: Study 1-8 Consent Form

School of Psychology, Cardiff University

Consent Form - Anonymous data

<Title of Experiment>

I understand that my participation in this project will involve the completion of a survey, whereby I will complete an online and computer based battery of individual test and measures. This requires approximately 60 minutes of my time.

I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason.

I understand that I am free to ask any questions at any time. I am free to withdraw or discuss my concerns with the lead researcher, Laura Bishop or the researcher, Dr Phil Morgan.

I understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study.

I understand that the research information provided by me will be held totally anonymously, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely or published.

I, ____________________________________(NAME) consent to participate in the study conducted by Laura Bishop School of Psychology, Cardiff University with the supervision of Dr Phil Morgan.

Signed: ___________________________

Date: ___________________________

Privacy Notice: [The following should also be the information sheet and debrief.]

The information provided on the consent form will be held in compliance with GDPR regulations. Cardiff University is the data controller and Matt Cooper is the data protection officer (inforequest@cardiff.ac.uk). This information is being collected by Dr Phil Morgan. This information will be held securely and separately from the research information you provide. Only the researcher will have access to this form and it will be destroyed after 7 years. The lawful basis for processing this information is public interest.

Appendix C: Study 1-3 Personality Traits Measure (IPIP; Goldberg et al., 2006)

Please indicate to what extent each of the following statements applies to you.

(1) Very Inaccurate, (2) Moderately Inaccurate, (3) Neither Inaccurate nor Accurate, (4) Moderately Accurate, (5) Very Accurate.

During the survey the below statements will be randomised and titles omitted

Extraversion

1. I feel comfortable around people
2. I make friends easily
3. I am skilled in handling social situations
4. I am the life of the party
5. I know how to captivate people
6. I have little to say*
7. I keep in the background*
8. I would describe my experiences as somewhat dull*
9. I don’t like to draw attention to myself*
10. I don’t talk a lot*

Agreeableness

1. I have a good word for everyone
2. I believe that others have good intentions
3. I respect others

4. I accept people as they are

5. I make people feel at ease

6. I have a sharp tongue*

7. I cut others to pieces*

8. I suspect hidden motives in others*

9. I get back at others*

10. I insult people*

Conscientiousness

1. I am always prepared

2. I pay attention to details

3. I get chores done right away

4. I carry out my plans

5. I make plans and stick to them

6. I waste my time*

7. I find it difficult to get down to work*

8. I do just enough work to get by*

9. I don’t see things through*

10. I shirk my duties*

Neuroticism

1. I often feel blue

2. I dislike myself

3. I am often down in the dumps

4. I have frequent mood swings

5. I panic easily

6. I rarely get irritated*

7. I seldom feel blue*

8. I feel comfortable with myself*

9. I am not easily bothered by things*

10. I am very pleased with myself*

Openness to experience

1. I believe in the importance of art

2. I have a vivid imagination

3. I tend to vote for liberal political candidates

4. I carry the conversation to a higher level

5. I enjoy hearing new ideas

6. I am not interested in abstract ideas*

7. I do not like art*

8. I avoid philosophical discussions*

9. I do not enjoy going to art museums*

10. I tend to vote for conservative political candidates*

Risk-Avoidance

1. I would never go hang-gliding or bungee jumping*

2. I would never make a high-risk investment*

3. I avoid dangerous situations*

4. I seek danger

5. I am willing to try anything once

6. I do dangerous things

7. I enjoy being reckless

8. I seek adventure

9. I take risks

10. I do crazy things

Appendix D: Study 1-3 Risk-taking Preferences (DOSPERT; Blais & Weber, 2006)

Part 1: For each of the following statements, please indicate the likelihood that you would engage in the described activity or behaviour if you were to find yourself in that situation.

(1) Extremely Unlikely, (2) Moderately Unlikely, (3) Somewhat Unlikely, (4) Not Sure, (5) Somewhat Likely, (6) Moderately Likely, (7) Extremely Likely.

1. Admitting that your tastes are different from those of a friend. (Social)
2. Going camping in the wilderness. (Recreational)
3. Betting a day’s income at the horse races. (Financial)
4. Investing 10% of your annual income in a moderate growth mutual fund. (Financial)
5. Drinking heavily at a social function. (Health/Safety)
6. Taking some questionable deductions on your income tax return. (Ethical)
7. Disagreeing with an authority figure on a major issue. (Social)
8. Betting a day’s income at a high-stake poker game. (Financial)
9. Having an affair with a married person. (Ethical)
10. Passing off somebody else’s work as your own. (Ethical)
11. Going down a ski run that is beyond your ability. (Recreational)
12. Investing 5% of your annual income in a very speculative stock. (Financial)
13. Going whitewater rafting at high water in the spring. (Recreational)
14. Betting a day’s income on the outcome of a sporting event. (Financial)

15. Smoking cigarettes. (Health/Safety)

16. Revealing a friend’s secret to someone else. (Ethical)

17. Driving a car without wearing a seat belt. (Health/Safety)

18. Investing 10% of your annual income in a new business venture. (Financial)

19. Taking a skydiving class. (Recreational)

20. Riding a motorcycle without a helmet. (Health/Safety)

21. Choosing a career that you truly enjoy over a more prestigious one. (Social)

22. Speaking your mind about an unpopular issue in a meeting at work. (Social)

23. Sunbathing without sunscreen. (Health/Safety)

24. Bungee jumping off a tall bridge. (Recreational)

25. Piloting a small plane. (Recreational)

26. Walking home alone at night in an unsafe area of town. (Health/Safety)

27. Moving to a city far away from your extended family. (Social)

28. Starting a new career in your mid-thirties. (Social)

29. Leaving your young children alone at home while running an errand. (Ethical)

30. Not returning a wallet you found that contains £200. (Ethical)

Appendix E: Study 1-3 Decision-making style (GDMS; Scott and Bruce, 1995)

Please indicate to what extent you agree or disagree with each of the following statements, according to the five-point scale below ranging from Strongly Disagree to Strongly Agree.

(1) Strongly Disagree, (2) Disagree, (3) Neutral, (4) Agree, (5) Strongly Agree.

1. When I make decisions, I tend to rely on my intuition. (Intuitive)

2. I rarely make important decisions without consulting other people. (Dependent)

3. When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it. (Intuitive)

4. I double check my information sources to be sure I have the right facts before making decisions. (Rational)

5. I use the advice of other people in making my important decisions. (Dependent)

6. I put off making decisions because thinking about them makes me uneasy. (Avoidant)

7. I make decisions in a logical and systematic way. (Rational)

8. When making decisions I do what feels natural at the moment. (Spontaneous)

9. I generally make snap decisions. (Spontaneous)

10. I like to have someone steer me in the right direction when I am faced with important decisions. (Dependent)

11. My decision-making requires careful thought. (Rational)

12. When making a decision, I trust my inner feelings and reactions. (Intuitive)

13. When making a decision, I consider various options in terms of a specified goal. (Rational)

14. I avoid making important decisions until the pressure is on. (Avoidant)

15. I often make impulsive decisions. (Spontaneous)

16. When making decisions, I rely upon my instincts. (Intuitive)

17. I generally make decisions that feel right to me. (Intuitive)

18. I often need the assistance of other people when making important decisions. (Dependent)

19. I postpone decision-making whenever possible. (Avoidant)

20. I often make decisions on the spur of the moment. (Spontaneous)

21. I often put off making important decisions. (Avoidant)

22. If I have the support of others, it is easier for me to make important decisions. (Dependent)

23. I generally make important decisions at the last minute. (Avoidant)

24. I make quick decisions. (Spontaneous)

25. I explore all of my options before making a decision. (Rational)

Appendix F: Study 1-3 Impulsivity (Barratt Impulsiveness Scale, BIS-11; Patton, Stanford and Barratt, 1995)

Please indicate how regularly you experience the following statements according to the five-point scale below ranging from Rarely/Never to Always.

Rarely/Never, Occasionally, Often, Almost Always, Always

1. I “squirm” at plays or lectures
2. I am restless at the theatre or lectures
3. I don’t pay attention
4. I concentrate easily
5. I am a steady thinker
6. I act on impulse
7. I act on the spur of the moment
8. I buy things on impulse
9. I make up my mind quickly
10. I do things without thinking
11. I spend or charge more than I earn
12. I am happy go lucky
13. I am a careful thinker
14. I am self-controlled
15. I plan trips well ahead of time

16. I plan for job security

17. I say things without thinking

18. I like to think about complex problems

19. I like puzzles

20. I save regularly

21. I am more interested in the present than the future

22. I get easily bored when solving thought problems

23. I change residences

24. I change jobs

25. I am future orientated

26. I can only think about one problem at a time

27. I often have extraneous thoughts when thinking

28. I have racing thoughts

29. I change hobbies

30. I plan for job security

Appendix G: Study 1-3 Acceptance of Technology Questionnaire (UTAUT; Venkatesh, Thong and Xu, 2012)

1 – Strongly disagree, 2 – Moderately disagree, 3 – Slightly disagree 4 – Neither disagree nor agree, 5 – Slightly agree, 6 – Moderately agree, 7 – Strongly agree

During the survey the below statements will be randomised and titles omitted

Performance Expectancy

1. I find cybersecurity tasks useful in my daily life.
2. Cybersecurity tasks help me accomplish things more quickly
3. Cybersecurity tasks increase my productivity
4. If I conduct cybersecurity tasks, I will increase my chances of getting a raise

Effort Expectancy

1. The cybersecurity tasks I need to undertake are clear and understandable
2. I find the cybersecurity tasks easy to undertake
3. It is easy for me to become skilful at cybersecurity tasks
4. Learning new cybersecurity tasks is easy for me

Social Influence

1. People who are important to me think that I should conduct cybersecurity tasks
2. People who influence my behaviour think that I should conduct cybersecurity tasks
3. People whose opinions that I value prefer that I conduct cybersecurity tasks

Trust

1. Trusting cybersecurity is not difficult

2. I trust cybersecurity to be secure

3. My tendency to trust cybersecurity is high

Facilitating Conditions

1. I have the resources necessary to conduct cybersecurity tasks

2. I have the knowledge necessary to conduct cybersecurity tasks

3. Cybersecurity tasks are not compatible with the technologies I use

4. I can get help from others when I have difficulties conducting cybersecurity tasks

Hedonic Motivation

1. Conducting cybersecurity tasks is fun

2. Conducting cybersecurity tasks is enjoyable

3. Conducting cybersecurity tasks is very entertaining

Price Value

1. Cybersecurity software is reasonably priced

2. Cybersecurity software is good value for the money

3. At the current price, cybersecurity software provides good value

Habit

1. Conducting cybersecurity tasks has become a habit for me

2. I am addicted to conducting cybersecurity tasks

3. I must conduct cybersecurity tasks

Behavioural Intention

1. I intend to conduct cybersecurity tasks in the future

2. I will always try to conduct cybersecurity tasks in my daily life

3. I plan to continue to conduct cybersecurity tasks frequently

Please choose your usage frequency for each of the following:

1. SMS

2. MMS

3. Ringtone and logo download

4. Java games

5. Browse websites

Perceived Risk

- I feel totally safe conducting cybersecurity tasks

- I do not feel secure conducting cybersecurity tasks

- I feel secure conducting cybersecurity tasks

- I am concerned about the safety cybersecurity tasks

Anxiety

- I feel apprehensive about conducting cybersecurity tasks

- It scares me to think that I could lose a lot of information conducting cybersecurity tasks
- Cybersecurity tasks are somewhat intimidating to me
- I find conducting cybersecurity tasks scary

Appendix H: Study 1-3 Combined PMT and TPB Questionnaire (Safa et al., 2015)

During the survey the below statements will be randomised and titles omitted

Thinking about your organisation please answer the following questions using the below scale:

1 – Strongly disagree, 2 – Moderately disagree, 3 – Slightly disagree 4 – Neither disagree nor agree, 5 – Slightly agree, 6 – Moderately agree, 7 – Strongly agree

Information Security Awareness (ISA)

- I am aware of potential security threat
- I have sufficient knowledge about the cost of information security breaches
- I understand the risk of information security incidents
- I keep myself updated in terms of information security knowledge to increase my awareness
- I share information security knowledge to increase my awareness

Information Security Organisation Policy (ISOP)

- Information security policies and procedures are important in my organisation
- Information security policies and procedures affect my behaviour
- Information security policies and procedures have attracted my attention
- Behaviour in line with organisational information security policies and procedures is of value in my organisation

Information Security Experience and Involvement (ISEI)

- My experience increases my ability to have a safe behaviour in terms of information security
- I am involved with information security and I care about my behaviour in my job
- I can sense the level of information security threat due to my experience in this domain
- My experience helps me to perform more considered information security behaviour
- I have suitable capability in order to manage information security risk due to my experience

Attitude (ATT)

- Careful information security behaviour is necessary
- Careful Information security behaviour is beneficial
- Practising careful information security behaviour is useful
- I have a positive view about changing users’ information security behaviour to be more considered
- My attitude towards careful information security behaviour is favourable
- I believe that careful information security behaviour is valuable in an organisation

Subjective Norms

- Information security policies in my organisation are important for my colleagues
- My colleagues’ information security behaviour influences my behaviour
- Information security culture in my organisation influences my behaviour • My boss’s information security behaviour influences my behaviour

Perceived Behavioural Control (PBC)

- I believe that careful information security behaviour is not a difficult practice

- I believe that my experiences help me to behave carefully around information security

- Following information security policies and procedures is easy for me • Careful information security behaviour is an achievable practice

Threat Appraisal

- I know the probability of security breach increases if I do not consider information security policies

- I could fall victim to different kinds of attack if I do not follow information security policies

- The security of my data will be weak if I do not consider information security policies

- Hackers attack with different methods and I should be careful in this dynamic environment

- To reduce the risk I do not open unexpected and out of context email

Information Security Self-Efficacy

- I have the skills to protect my business and private data

- I have the expertise to protect my business and private data

- I think the protection of my data is in my control in terms of information security violations

- I have the ability to prevent information security violations

Information Security Conscious Care Behaviour (ISCCB)

- I consider security experts recommendations in my information security manner
- Before taking any action that affects information security, I think about its consequences
- I talk with security experts before I do something that relates to information security
- I consider my previous experience in information security to avoid repeating prior mistakes
- I always try to change my habits to security conscious behaviour

Appendix I: Study 1-3 Additional Factors Found in Thompson and McGill (2015) and Posey, Roberts and Lowry (2015)

During the survey the below statements will be randomised and titles omitted

Thinking about your organisation please answer the following questions using the below scale:

1 – Strongly disagree, 2 – Moderately disagree, 3 – Slightly disagree 4 – Neither disagree nor agree, 5 – Slightly agree, 6 – Moderately agree, 7 – Strongly agree

1. Intrinsic Maladaptive Rewards

a. I would receive personal gratification for purposefully not protecting my organisation from its information security threats

b. I would feel a sense of internal satisfaction for allowing information security threats to harm my organisation

2. Extrinsic Maladaptive Rewards

a. I could be rewarded financially for choosing not to protect my organisation's information and information systems from security threats

b. I believe others would be willing to reward me financially for intentionally failing to protect my organisation's information and information systems

3. Organisational commitment

a. I would be very happy to spend the rest of my career with this organization.

b. I do not feel like "part of the family" at my organization. (R)

c. I do not feel "emotionally attached" to this organization. (R)

d. This organization has a great deal of personal meaning for me

e. I do not feel a strong sense of belonging to my organization. (R)

15. Psychological ownership

a. I feel a high degree of ownership for my work computer and its contents

b. The information stored on my work computer is very important to me

c. I personally invested a lot in my work computer (e.g. time, effort, money)

d. I personally invested a lot in the software/applications on my work computer (e.g. time, effort, money)

e. When I think about it, I see an extension of my life in my work computer

f. I have personalised my work computer to better suit the way I use it

g. I see my work computer as an extension of myself

Appendix J: Study 1-3 Research Debrief

School of Psychology, Cardiff University

Research Debrief

How Individual Differences and Environmental Context Predict Cyber-Security Perceptions and Behaviours

Firstly, thank you very much for taking part in the study!

This study involved participants completing a number of questionnaires on individual differences (e.g. personality, impulsivity), and their cyber-security perceptions and behaviours. This research is being undertaken to find ways to better protect computer users from a cyber-attack and will hopefully lead to the development of more tailored interventions. Outcomes will not be individualised but analysed together to better understand how different groups of individuals (e.g. gender, age group, personality) perceive and behave towards cyber-security in order to better protect these groups moving forward.

This study is part of a larger-scale research project that is addressing several research questions. The current study aims to:

1. Establish which individual differences predict cyber-security behaviours in order to better tailor intervention within the workplace and increase intervention success.
2. Improve understanding around how cyber-security perceptions and behaviours differ across work-based environments in order to better tailor cyber-security intervention.
3. Bring together a number of psychological models and theories that are currently being used within the cyber-security domain in order to (with the above) crease a human factors cyber-security assessment framework that can be used to better tailor intervention.

This study is particularly focused on building a cyber-security assessment framework that is based around the specific human factors that influence cyber-security perceptions and behaviours. This framework can then be utilised to assess an individual’s needs and highlight which specific intervention is required for them to better protect themselves and their workplace.

If you require further information about the study, please do not hesitate to contact the researchers below;

Lead Researcher: Laura Bishop (PhD Student)

or

Supervisor: Dr Phil Morgan*

School of Psychology

Office: 9.18, Tower Building

70 Park Place

Cardiff University

Cardiff

CF10 3AT

Email: bishoplm2@cardiff.ac.uk

Tel: 02922510784 (Extn. 10784)

Secretary of the Ethics Committee

School of Psychology

Cardiff University

Tower Building

Park Place

Cardiff

CF10 3AT

Tel: 029 2087 0360

Email: psychethics@cardiff.ac.uk

The data controller is Cardiff University and the Data Protection Officer is Matt Cooper CooperM1@cardiff.ac.uk. The lawful basis for the processing of the data you provide is consent.

*Reader/ Associate Professor in Cognitive Science and Human Factors

Research Group Director: Human Factors Excellence (HuFEx) at Cardiff University

Senior Researcher and Technical Lead in Cyber Psychology & Human Factors at Airbus

Appendix K: Regression Model Explaining 55% of the Variance in Reported Cybersecurity Behaviours

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.746[a]	.557	.554	3.112

a. Predictors: (Constant), CSA

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.249	2.136		-.116	.908
	CSA	.195	.014	.746	13.775	.000

a. Dependent Variable: Reported Cybersecurity Behaviour (combined PMT and TPB questionnaire)

Appendix L: Regression Model 2 Explaining 60% of the Variance in Reported Cybersecurity Behaviours

Model Summary

			Adjusted R	Std. Error of the
Model	R	R Square	Square	Estimate
1	.776[a]	.602	.600	3.363

a. Predictors: (Constant), CSA

Coefficients[a]

				Standardized		
		Unstandardized Coefficients		Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.065	1.146		.929	c.354
	CSA	.190	.009	.776	22.120	.000

a. Dependent Variable: Reported Cybersecurtiy Behaviour

SSSSS

Appendix M: Study 4 Email Analysis Questionnaire

- Email ID
- Date Listed
- Attachment Present – Yes/No
- Link Present – Yes/No
- Reply Requested (Reciprocation) – Yes/No
- Is the email from the following: (Authority, Liking/Similarity)
    - Government
    - Educational Institution
    - Banking Agency
    - Other
    - None
- Is there a logo? (Authority, Liking Similarity) – Yes/No
- Is the email asking the user to perform an action? (Authority, Consistency/Commitment, Reciprocity)
    - Click here/ Click link/ “Click”
    - Update form
    - Confirm form
    - Open the attachment
    - Confirm personal information
    - Upgrade account information
    - Other action asked to be performed
    - None
- Does the contain information regarding known contacts? (Social Proof, Liking/Similarity)?

  - Friends
  - Colleagues
  - Family
  - Other information regarding known contacts
  - None
- Does the email refer to actions performed by other users? (Social Proof, Liking/Similarity)
  - Customer complaints
  - Others expecting your input
  - Other actions performed by other users
  - None
- Does the email contain the following identifying information? (Liking/Similarity, Authority)
  - Email address
  - Physical address
  - Telephone number
  - Other identifying information
  - None
- Are there details included in the email? (Liking/Similarity)
  - Invoice number
  - Requested service details
  - Payment details
  - Other details of service
  - None

- Are there elements in the first person stating “I am this or that”? (Liking/ Similarity) – Yes/No
- Are there elements in the first person describing the behavior around others? (Social Proof, Liking/Similarity) – Yes/No
- Is the email referring to other elements outside the email to look more reliable (Adobe Reader, etc.)? (Liking/Similarity, Consistency/Commitment) – Yes/No
- Is the email asking commitment from the user? (Commitment/Consistency, Reciprocation)
    - “Can I trust you?”
    - “Can you do this for me?”
    - Other commitments
    - None
- Does the email have visual cues? (Liking/Similarity)
    - Colors
    - Unusual font
    - Abnormal use of capital letters
    - Big images
    - Exclamation and/or interrogation marks
    - Spelling mistakes
    - Grammar mistakes
    - Other visual cues
    - None
- Does the email convey a sense of urgency? (Scarcity)
    - Time restrictions
    - “Urgent”

  - “Must be done”
  - Other items conveying sense of urgency
  - None
- Does the email list a consequence if user does not comply? (Authority) – Yes/No
- Does the email ask the recipient to move outside of the email to 'learn more'? (Curiosity) – Yes/No
- Does it require clicking on a link/download/replying to verify the email's intention? (Curiosity) – Yes/No
- Does the email ask to forward to a friend/colleague (Widening the Wed - Social Proof) – Yes/No
- Does the email attempt to evoke strong affect? If so, which affect type?
  - Fear – Yes/No
  - Excitement – Yes/No
  - Panic – Yes/No

Appendix N: Study 5-8 Introduction Sheet

School of Psychology, Cardiff University

Participant Information Sheet

The Impact of Disorganised Inboxes on Society

Introduction

You are being invited to take part in a research experiment. Before you decide to take part, it is important for you to understand why the research is being conducted and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please contact and ask the lead researcher if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part and thank you for reading this.

This research is being conducted by Laura Bishop (Psychology PhD Student, Cardiff University), Dr Phoebe Asquith (Senior Research Associate, Cardiff University).

What is the purpose of this study?

The main objective of this study is to investigate how computer users organise and clear a selection of emails from their inbox and how they rate satisfaction, post organisation. Findings from this survey will help provide insight into the impact of disorganised inboxes on society.

How will the research be done and what will I have to do?

The study should take no longer than 30 minutes and will commence after you have made an informed decision to participate or not, through informed consent. Once you have answered some initial questions you will be moved across to the email sorting task. Upon completion of the task you must return back to the survey and answer the remaining questions.

Why have I been asked and do I have to take part?

Due to the nature of the study, participants are required to be >18 years old and will need to have normal or normal-corrected vision. An interpreter and/or translator is not available for these experiments and therefore a good level of the English language is required.

Your participation is completely voluntary, and it is up to you to decide whether or not you would like to take part. If you decide to take part, you are still free to withdraw at any time without giving any reason. However, after your participation in the study you may only request the withdrawal of your data up to the point of data analysis (15 working days after the experiment). After this point it may be difficult to trace your data, and the removal of your data may possibly impact the ongoing data analysis and the write-up of the project.

Is this information confidential and held securely, and what will be done with the results?

The personal information collected in this research project (e.g., any form/questionnaire/survey) will be processed by the University in accordance with the terms and conditions of the 1998 Data Protection Act, GDPR regulations (2018). We will hold your data securely and not make it available to any third party unless permitted or required to do so by law. Your personal information will be used/processed as described within this information sheet. All the data you provide will be stored in password-protected computer files under an anonymous identifier and used on a confidential basis.

Non-personal data collected from your questionnaire responses will be held if it retains research value although this will not exceed 7 years. The anonymised data may be made available for further appropriately approved research at the University. No data will be published in a way that could lead to the identification of the individual participants.

The findings from this research may be used in publications in academic journals and also presented at academic conferences. Your personal details will never be included in any of these publications, and your data will only be used anonymously.

Do you have any further questions?

If you have questions about the research - either now or at some future date - please contact either:

Laura Bishop: bishoplm2@cardiff.ac.uk

Dr Phoebe Asquith: asquithpm@cardiff.ac.uk

Privacy Notice:

The information provided on the consent form will be held in compliance with GDPR regulations. Cardiff University is the data controller and Matt Cooper is the data protection officer (inforequest@cardiff.ac.uk). This information is being collected by Dr Phil Morgan. This information will be held securely and separately from the research information you provide. Only the researcher will have access to this form and it will be destroyed after 7 years. The lawful basis for processing this information is public interest.

The data controller is Cardiff University and the Data Protection Officer is Matt Cooper CooperM1@cardiff.ac.uk. The lawful basis for the processing of the data you provide is consent.

Appendix O: Study 5 Study Welcome within Google Spaces

Welcome to the study.

Please read the following information.

This is a simulated email task to understand how people interact with emails when clearing their inbox, a task rarely undertaken today.

Your participation is very much appreciated!

Introduction

During this exercise we will be gathering information on how you choose to respond to each email and how you to choose to file it.

The main objective of this study is to investigate how computer users interact with and clear a selection of emails from their inbox and how they rate satisfaction, post organisation.

Findings from this survey will help provide insight into the impact of disorganised inboxes on society.

Instructions

You are playing the role of a procurement manager, at fictional company BlueStar Technologies.

You will be presented with a simulated inbox filled with emails. You need to click on each email to open it and decide whether or not to take action. If you decide to take action please wait until a pop up appears stating “success, your action has been completed”.

- Once decided you must then move the email to one of the folders found to the left of the inbox (folders include - inbox, urgent, follow up, finance, networking, files, IT, personal, suspicious emails, deleted emails). This can be achieved by clicking on the email and dragging it to the appropriate folder.

  - You have completed the task when you have opened all emails, chosen whether to take action, and filed them in their most suitable folder. Please ensure your inbox is empty before exiting.

- When you complete the task, close the virtual desktop down as you would a normal computer.
- Once exiting this study, you should then return to your Qualtrics online tab, to complete a few short questions.

Taking Part

If you are happy to participate in this task in accordance with the below conditions, please click on the email button on the desktop to begin the exercise.

Your responses will be used in research reports, conference presentations and journal publications. However, you will not be personally identifiable and your responses will not be linked to you in any way.

When you have completed the email sorting task, please return to the Qualtrics tab and complete the rest of the survey.

If you have any questions about this research, please contact one of the leading researchers prior to completing this survey:

Laura Bishop, School of Psychology, Cardiff University (bishoplm2@cardiff.ac.uk)

Dr Phil Morgan, School of Psychology, Cardiff University (morganphil@cardiff.ac.uk)

Appendix P: Complete Email List

Authority

External - Genuine

From: <Jamie Foxon> jp.foxon@churchlanemedical.co.uk

To: sam.poole@bluestar.co.uk

Hi

I have been forwarded your records by your GP surgery and after reviewing them in full I agree that you need to be seen at our clinic.

Please click the below link to book an appointment with us:

<http://www.churchlanemedical.co.uk/calendar>

Dr Foxon
BDS FDS RCS MB ChB FRCS(Ed) FRCS(Plast)

Plastic Surgeon
Church Lane Hospital

External - Phishing

Sender: <Jessica Smith> Jessiesmith@buslnessoftheyear.de

To: sam.poole@bluestar.co.uk

Invitation to the 2021 Business of the Year Conference

Date: Thursday , 30 September 2021 | Time : 10:00 AM PST , 01:00 PM EST

This rewarding event is recomended by a number of leading industry experts including the CEO of Twitter and Chief Financial Officer at PayPal.

Key Learning Objectives :

Fundementals of business

How do you take advantage of contract negotiations?

Building your empire piece by piece

Click here to register too join the session, or decline the invitation.

http://r.windsre.de-checking.bitnet/de/ID.php?u=TbsreRTrsdhiexs?u=7325y8192s

Jessica Smith

Conference Lead

www.businessoftheyear.com

387 Tyson Street Suite 21 London, UK

Call: 0208 426 783

customersupport@businessoftheyear.com

© Copyright Business of the Year

Internal - Genuine

From: <David Matthew> david.matthews@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi,

I writing to invite you to the Annual Bluestar Employee Engagement Summit 2021, set to take place in Amsterdam on November 23rd - 24th, 2021.

Please find attached the event agenda including information on this year's keynote speaker:

<Agenda2021.pdf>

This unique event comes highly recommended by company CEO Gerald Fowler, and provides an opportunity for employees to get really involved. I look forward to seeing you in Amsterdam!

Professor David Matthews Ph.D.

Organisational Psychologist and expert in Employee Engagement

Bluestar

Internal - Phishing

From: <Gerald Fowler> gfowler@safersecurity.ma

To: sam.poole@bluestar.co.uk

Hi,

Could you please review attached report from last company meetings and send me across any feadback? ?

<Compny Report.pdf>

Thanks in advance

Gerald Fowler

CEO

Bluestar

Commitment and Consistency

External - Genuine

From: <Customer Services> customerservices@brislingtonanalytics.co.uk

To: sam.poole@bluestar.co.uk

Hi,

Thank you for your previous interest in our services. We are pleased to attach our latest Whitepaper as requested:

<Company Whitepaper.pdf>

Kind regards

Customer Services
Brislington Analytics

External - Phishing

Sender: <David Washington> info@swanagepublications.bitnet

To: sam.poole@bluestar.co.uk

Dear Customer

As a reader of our magazine we would like to know which products and services you are looking to purchase over the next 12 months .

Please click here to enter our brief survey –it will take less than a minute to complete .

<http://h.nawina.de-checking.bitnet/de/ID.php?u=LhsdoOKJfsjdsdvg?u=8493j3040i>

Thanks you for your timing

David Washington

Swanage Publications

Internal Genuine

From: <simon Taylor> simon.taylor@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi,

As discussed, below is a link to the minutes collated at our latest meeting. Can you check that I captured everything discussed?

Minutes: <http://bluestar.co.uk/draftdocuments/>

Simon Taylor

Bluestar

Internal Phishing

From: <John Richards> johnrichards4@safesecurety.com

To: sam.poole@bluestar.co.uk

Hello,

I complied the company data you requested into one file and have attached it here , hopefully it is all needed. If I left anything please let me know

<Compay Data.xlsx>

Thank you

John

Reciprocity

External - Genuine

From: <Nicola Edwards> ns.edwards@NBMCE.com

To: sam.poole@bluestar.co.uk

Hi,

We would like to offer you free registration to the National Business Management Conference & Exhibition, held over two days on the 29th & 30th October 2021 in the Citywest Convention Centre.

Please find attached details for how to sign up!

<Sign Up Form.docx>

Nicola Edwards
NBMCE

External - Phishing

From: <Linda Blackwood> lb32423@fastgroup.cn

To: sam.poole@bluestar.co.uk

Hi?

We would like to offer you a free gift from our product range worth around £20 for you to try. This should be with you shortly.I wonder whether you would complete this short survey to let us know how we are doing as a business?

Survey: < http://63.17.167.23/pc/verification.htm?>

Linda Blackwood

FAST Group

Internal - Genuine

From: <David Walker> david.walker@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi,

We have entered you into a free competition to win a Samsung Galaxy S10!

Would you mind clicking on the below link to take part in a very brief survey on the quality of our companies offerings?

<https://www.bluestar.co.uk/survey-samsung>

David Walker

Marketing
Bluestar

Internal - Phishing

From: <Jane Reed> janereeed@bluestar.cn

To: sam.poole@bluestar.co.uk

Hi

We have just posted you a £10 gift voucher bonus for the shop of your choice !!!

We would also like to make you aware of a short online staffs training course we have coming up this week. This is NOT mandatory, but will overall customer experience.

Please view the attached agenda. The sign up form can be found on page 3.

<Agendda.pdf>

Jane Reed

Training
Bluestar

Similarity and Liking

External – Genuine

From: <Business Funding> Melanie.Thorne@businessfunding.com

To: sam.poole@bluestar.co.uk

Hello, dear friend

I wish you and your family health, happiness, peace and prosperity at this challenging time. At businessfunding.com we will do our best to assistant you and your company should you find it compliments your current focus. We feel our current business strategies align and that a meeting could be mutually beneficial.

Please click the link below to arrange a quick call!

<http://businessfunding.com/appointments/>

Best Regards

Melanie
Business Funding Co., Limited

External - Phishing

From: <Mike Britton> mike.britton@eievate.com

To: sam.poole@bluestar.co.uk

Good day, how are you and the family? all well I hope? i got your contact details through Linkedin and internet search and I wish to use this opportunity to briefly introduce myself to you before moving forward to this mail. I am Michael Britton from Elevate, part of the Fusion Group, and have been very impressed with your work to date. I wish to discuss a very important partnership with you that I think would benefit both of our businesses.

Do you have some time over the next few days to meet? I have attached my availability

<Mike Briton Diary.xlsx>

Thank you,
Mike

Internal - Genuine

From: <Brianna Andrews> brianna.andrews@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi there,

Hope this find you well? How are things?

I have attached details of our new services and pricing matrix for your perusal.

<Pricing Matrix.xlsx>

Thank you in advance!

Brianna Andrews

Account Department

Bluestar

Internal Phishing

From: <John Roberts> john.roberts@blue-star.co.uk

To: sam.poole@bluestar.co.uk

Hi,

How are you ??? Looking forward to a prosperous and fulfilling second half of 2021? I work in the company procurement division and you were recently recommended as someone with high expertise in my current work.

Do you have 5 minutes to discuss a project we are currently working on? Book something in my dairy below and I'll shoot an email right back to you. <

<http://diaryuendoe.com@10.19.32.4/o/clicku=5ooefkfoss099ss>

Speak soon!

John Roberts

Procurement
Bluestar

Scarcity

External - Genuine

From: <Marketing> marketing@themoneymagazine.com

To: sam.poole@bluestar.co.uk

Hello,

We are reaching out because we want to help you keep your finger on the pulse of the happenings and activities in your respective sector. Sign up <u>today only</u> for immediate access to our online reports, recommended readings and useful resources within your sector.

If you agree to start the journey with us, please sign up here

<https://themoneymagazine/signup>

The Money Magazine

External – Phishing

From: <Mary Milson> @instantcommunicat1ons.com

To: sam.poole@bluestar.co.uk

Urgent : Request for Quotation

Due to an urgent requirement our end, I would like to request an urgent quotation. Please find full details attatched.

Instruction to Bid: Please refer to page 4 highlighted in BLUE

<Instructon to Bid.docx>

Your cooperation in this regard is highly appreciated and thanks in advance.

Mary Milson

Instant Communications

Internal - Genuine

From: <IT Department> ithelpdesk@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Alert: You are about to lose access rights to the G Drive.

If you still need access to this folder then you need to request an extension within the next 24 hours. Please complete the attached form and send back to us as soon as possible.

<Folder Access.docx>

Best Regards

IT Department

Internal - Phishing

From: <IT Department> ithelpdesk@helpdeskbluestar.co.uk

To: sam.poole@bluestar.co.uk

ID: @bluestar.com Immediate Action Required: Verify email address 'Your@bluestar.co.uk' password expires today at 17:00 PM. Use the link below to continue with the same password

<https://akabomed.us11.list-manage.com/track/click?u=8946djooid>

Best,

IT Department

Social Proof

External – Genuine

From: <Sally Cains> exhibitions@unitedbusinesssociety.com

To: sam.poole@bluestar.co.uk

Dear Sir/Madam,

United Business Society Exhibition is taking place 21st – 26th June 2021 at the Richmond Arena, Coventry. The exhibition will be a comprehensive display of all the latest business theories and solutions.

Who's attending the Exhibition:

130+ exhibitors within your sector, demonstrating their solutions as well as their practical application

2000+ colleagues and industry competitors, discovering how the latest solutions could be applied to their business for massive operational benefit

Each session will have an open forum at the end so you can get the answers to your burning questions and discover how the latest business solutions can be applied to your business.

To book click here <https:/unitedbusinesssociety.com/bookings>

We look forward to seeing you there.

Sally Cains
Exhibition Associate
United Business Society

External - Phishing

From: <James Clarke> exhibitions@omus98088.dz

To: sam.poole@bluestar.co.uk

Hi!!!!!!

We are delighted to invite you to celebrate the 10th anniversary of Operations Management UK Symposium (OMUS) and be part of the largest most influential event in the UK. For 10 years, OMUS has been gathering place for many of your colleagues and competitors, sharing the latest industry innovations but also explore partnership and investment opportunities in region.

This year OMUS will bring together more than 27,000 professionals from more than 45 countries. Last year 96% of attendees scored the event as unmissable! ! ! !

To attend complete and send back page 7 of the attached agenda

<Syposium Agenda.pdf>

James Clarke
Exhibitions
OMUS

Internal - Genuine

From: <Gloria Brown> gloria.brown@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi all

Please find attached our latest company newsletter.

In 2020 the Bluestar newsletter was rated 5 out of 5 by 82% of your colleagues.

“A must for anyone looking to improve their career prospects and satisfaction within the business!”

<Company Newsletter.pdf>

Kind Regards

Ms Gloria Brown

Marketing
Bluestar

Internal - Phishing

From: <Richard Martin> richardmartin@bluestaruk.co.uk

To: sam.poole@bluestar.co.uk

Hi all,

Do you have time in your calendar to attend the following course: “ HR for managers ” ??

Across our business 80% of managers have so far attended this course reporting huge improvement in both their management skills and employee satisfaction.

To book yourself on to the course click here <https:/ h.bluestar.de-checking.net/de/ID.php?u=LhsdoOKJfsjdsdvg/>

Best,

Richard Martin

HR Development Rep.

Bluestar

Intrigue

External – Genuine

From: <Newsroom> newsroom@kingswoodpost.co.uk

To: sam.poole@bluestar.co.uk

Hi

Have you seen this article? We thought you might be interested!

<Article.pdf>

Best regards

Yungblud The indefinable Yorkshire artist wants to "change culture", while making anthems for Generation Z. Spending time not working can spark the best business ideas, says top P&G exec Coming up with original ideas can be extremely costly in terms of both time and resources. Why P&G's skin care president says getting out of the office to purposely spend time not working can spark the freshest, most innovative ideas. Derek Acorah, TV psychic medium and 'Most Haunted' star, dies at 69 Derek Acorah, a popular TV psychic medium and former host of British reality show "Most Haunted," has died at age 69.

External – Phishing

From: <Info> ineeof435@bizgroup.a

To: sam.poole@bluestar.co.uk

_____ ___ _____ http://bizgroup.com.de.cgi-bin.webscr.cmd-login-submit.dispatch.sicherkontrolle.su/cgi-bin/

Internal - Genuine

From: <Bernard Sampson> bernard.sampson@bluestar.co.uk

To: sam.poole@bluestar.co.uk

bernard.sampson@bluestar.com has shared the following LINK:

<http://www.bluestar.co.uk/internalmarketing>

Internal - Phishing

From: <Sabrina Davidson> sdavidson@ukbluestar.bf

To: sam.poole@bluestar.co.uk

Hi - Please watch the attached video and let know your thoughts.

<videoo.mov>

Sabrina Davidson

Bluestar

None

External – Genuine

From: <Charles Draper> charlesdraper@interatlantic.com

To: sam.poole@bluestar.co.uk

Hi

I would like to place my first order with your business.

Please find attached my completed order form including product numbers, could you possibly let me know an estimated delivery time?

<Order Form.pdf>

Thank you

Charles Draper

InterAtlantic Inc

External - Phishing

From: <Tom Porter> tomporter3@sanovi.de

To: sam.poole@bluestar.co.uk

Hi,

Your industry has recently gone live on Editorial Manager (EM) , our online submission and peer review tracking system and you have been registered. The first time you log in, you need to select classifications and enter words, so editors know your areas of expertise before and invite you as a reviewer.

Full details on how to log into system for the first time you will need view this document:

<sanovig35453j33545.pdf>

Kind Regards,

Tom Porter

Editorial Manager UK

Internal - Genuine

From: <Jennifer Underwood> jennifer.underwood@bluestar.co.uk

To: sam.poole@bluestar.co.uk

Hi

I have attached an invoice for the recent training sessions your team attended.

Please click on the link below to check the invoice and e-sign

http://www.docusign.com/bluestar

Jennifer Underwood
Accounts Clerk

Bluestar

Internal - Phishing

From: <James Bright> james.bright@bluesstar.co.uk

To: sam.poole@bluestar.co.uk

Hi,

This is a test mail for validating your email. Please click link to validate

< http://h.saffesecurity.de-checking.net/de/ID.php?u=LhsdoOKJfsjdsdvg>

Thanks ! !

James Bright
IT Trainee

Appendix Q: Study 5-8 Research Debrief

School of Psychology, Cardiff University

Research Debrief

Human Susceptibility to Persuasion Tactics in Phishing Emails

Firstly, thank you very much for taking part in the study!

As you may realise, this study was initially titled “The Impact of Disorganised Inboxes on Society". However, you may have noticed that the survey you completed post exercise asked questions on cyber-security perceptions and behaviours, which did not link directly with the task.

This research was actually interested in how elements within phishing emails persuade recipients to consider the email genuine, as well as investigate which of these elements are the most influential. The survey itself was looking at whether these methods of persuasion are more successful across a number of individual differences including how high you appraise the threat of a cyber-attack and how you rate your ability to protect yourself from this threat.

Due to the number of ways in which offenders attempt to persuade email recipients to click on links, open attachments, or provide confidential information it is important to build knowledge around which of these elements are the most effective in order to focus intervention in these areas and better support computer users in the identification and reporting of phishing emails moving forward.

You may have experienced some information prior to taking part in the experiment that had the purpose of attempting to increase your ability to detect those emails that were phishing. Should the information provided have increased participant phishing detection, it is possible that similar messages can be used in the workplace to guide employees towards more secure behaviour.

If you require further information about the study, please do not hesitate to contact the researchers below;

Lead Researcher: Laura Bishop (PhD Student)

or

Supervisor: Dr Phoebe Asquith*

School of Psychology

Office: 9.18, Tower Building

70 Park Place

Cardiff University

Cardiff

CF10 3AT

Email: bishoplm2@cardiff.ac.uk

Tel: 02922510784 (Extn. 10784)

Secretary of the Ethics Committee

School of Psychology

Cardiff University

Tower Building

Park Place

Cardiff

CF10 3AT

Tel: 029 2087 0360

Email: psychethics@cardiff.ac.uk

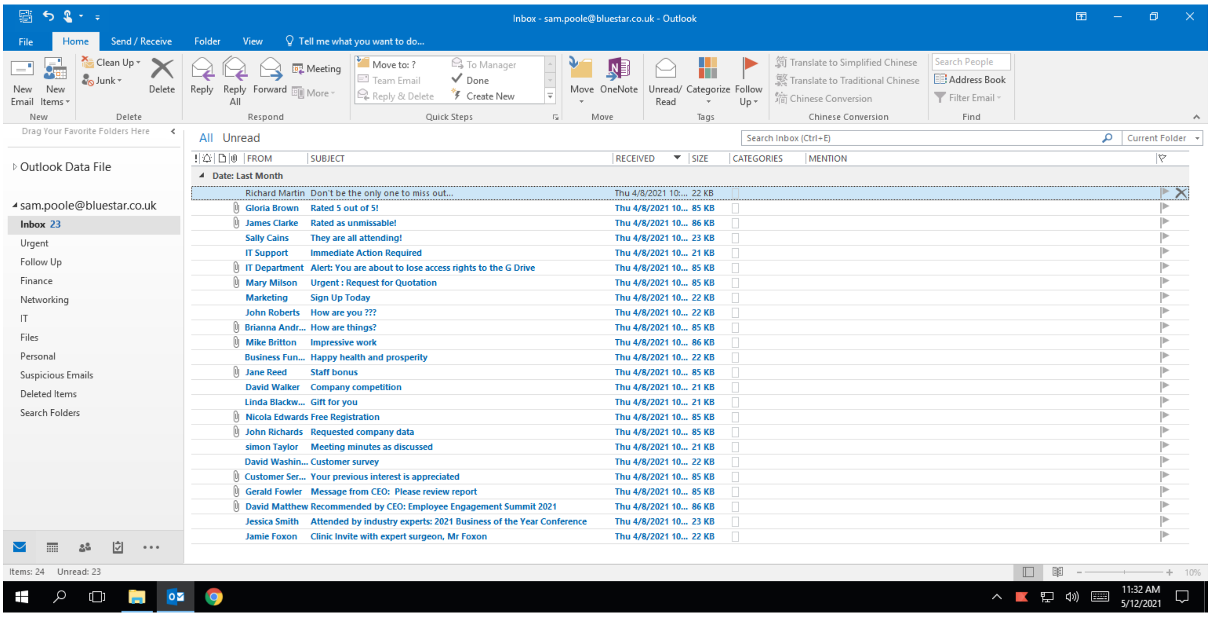
The data controller is Cardiff University, and the Data Protection Officer is Matt Cooper CooperM1@cardiff.ac.uk. The lawful basis for the processing of the data you provide is consent.

*Reader/ Associate Professor in Cognitive Science and Human Factors

Research Group Director: Human Factors Excellence (HuFEx) at Cardiff University

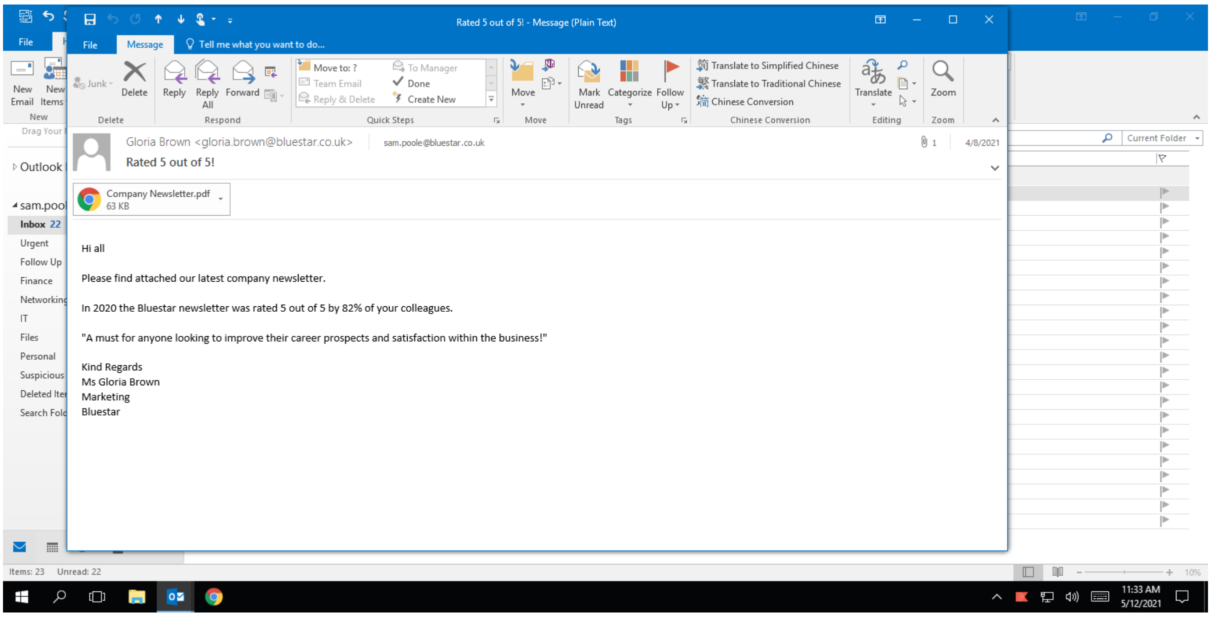
Senior Researcher and Technical Lead in Cyber Psychology & Human Factors at Airbus

Appendix R: Study 6 Screenshots of Procedure



Inbox - sam.poole@bluestar.co.uk - Outlook
File Home Send / Receive Folder View Tell me what you want to do...
New Email New Items Clean Up Junk Delete Reply Reply All Forward Meeting More
Move to: ? To Manager Team Email Done Reply & Delete Create New
Move OneNote Unread/ Read Categorize Follow Up
Translate to Simplified Chinese Translate to Traditional Chinese Chinese Conversion
Search People Address Book Filter Email
New Delete Respond Quick Steps Move Tags Chinese Conversion Find
Drag Your Favorite Folders Here
Outlook Data File
sam.poole@bluestar.co.uk
Inbox 23
Urgent
Follow Up
Finance
Networking
IT
Files
Personal
Suspicious Emails
Deleted Items
Search Folders
All Unread
Search Inbox (Ctrl+E) Current Folder
Date: Last Month
Items: 24 Unread: 23
11:32 AM 5/12/2021

FROM	SUBJECT	RECEIVED	SIZE
Richard Martin	Don't be the only one to miss out...	Thu 4/8/2021 10:...	22 KB
Gloria Brown	Rated 5 out of 5!	Thu 4/8/2021 10...	85 KB
James Clarke	Rated as unmissable!	Thu 4/8/2021 10...	86 KB
Sally Cains	They are all attending!	Thu 4/8/2021 10...	23 KB
IT Support	Immediate Action Required	Thu 4/8/2021 10...	21 KB
IT Department	Alert: You are about to lose access rights to the G Drive	Thu 4/8/2021 10...	85 KB
Mary Milson	Urgent : Request for Quotation	Thu 4/8/2021 10...	85 KB
Marketing	Sign Up Today	Thu 4/8/2021 10...	22 KB
John Roberts	How are you ???	Thu 4/8/2021 10...	22 KB
Brianna Andr...	How are things?	Thu 4/8/2021 10...	85 KB
Mike Britton	Impressive work	Thu 4/8/2021 10...	86 KB
Business Fun...	Happy health and prosperity	Thu 4/8/2021 10...	22 KB
Jane Reed	Staff bonus	Thu 4/8/2021 10...	85 KB
David Walker	Company competition	Thu 4/8/2021 10...	21 KB
Linda Blackw...	Gift for you	Thu 4/8/2021 10...	21 KB
Nicola Edwards	Free Registration	Thu 4/8/2021 10...	85 KB
John Richards	Requested company data	Thu 4/8/2021 10...	85 KB
simon Taylor	Meeting minutes as discussed	Thu 4/8/2021 10...	21 KB
David Washin...	Customer survey	Thu 4/8/2021 10...	22 KB
Customer Ser...	Your previous interest is appreciated	Thu 4/8/2021 10...	85 KB
Gerald Fowler	Message from CEO: Please review report	Thu 4/8/2021 10...	85 KB
David Matthew	Recommended by CEO: Employee Engagement Summit 2021	Thu 4/8/2021 10...	86 KB
Jessica Smith	Attended by industry experts: 2021 Business of the Year Conference	Thu 4/8/2021 10...	23 KB
Jamie Foxon	Clinic Invite with expert surgeon, Mr Foxon	Thu 4/8/2021 10...	22 KB



Rated 5 out of 5! - Message (Plain Text)
File Message Tell me what you want to do...
Junk Delete Reply Reply All Forward
Move to: ? To Manager Team Email Done Reply & Delete Create New
Move Mark Unread Categorize Follow Up
Translate to Simplified Chinese Translate to Traditional Chinese Chinese Conversion
Translate Zoom
Delete Respond Quick Steps Move Tags Chinese Conversion Editing Zoom
Items: 23 Unread: 22
11:33 AM 5/12/2021

Gloria Brown <gloria.brown@bluestar.co.uk> sam.poole@bluestar.co.uk 4/8/2021

Rated 5 out of 5!

Company Newsletter.pdf
63 KB

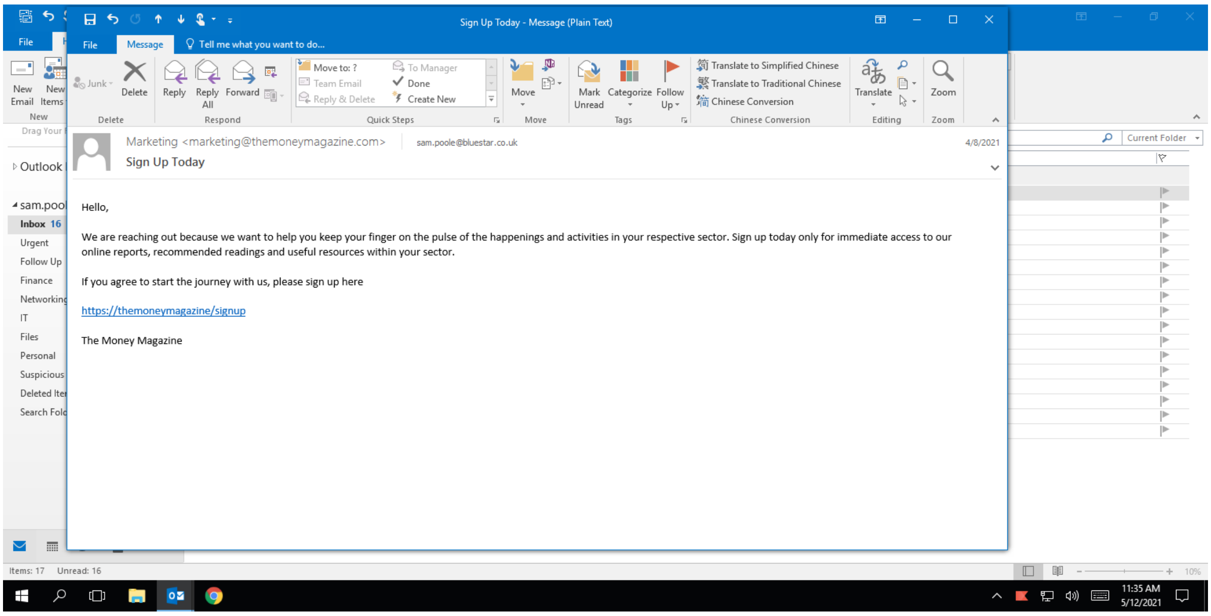
Hi all

Please find attached our latest company newsletter.

In 2020 the Bluestar newsletter was rated 5 out of 5 by 82% of your colleagues.

"A must for anyone looking to improve their career prospects and satisfaction within the business!"

Kind Regards
Ms Gloria Brown
Marketing
Bluestar



Marketing <marketing@themoneymagazine.com> sam.poole@bluestar.co.uk 4/8/2021

Sign Up Today

Hello,

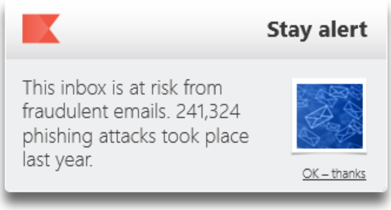
We are reaching out because we want to help you keep your finger on the pulse of the happenings and activities in your respective sector. Sign up today only for immediate access to our online reports, recommended readings and useful resources within your sector.

If you agree to start the journey with us, please sign up here

https://themoneymagazine/signup

The Money Magazine

Example Nudge (ThinkCyber Redflags®):



Stay alert

This inbox is at risk from fraudulent emails. 241,324 phishing attacks took place last year.

OK – thanks