A HUMAN CENTRIC APPROACH TO UNINTENTIONAL INSIDER THREAT: DEVELOPMENT OF A SOCIOTECHNICAL FRAMEWORK

NEESHE KHAN, BSc (Hons)

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

March 2023

Abstract

The exploitation of so-called insiders is increasingly recognised as a common vector for cyberattacks. Unintentional insider threat– inadvertent mistakes and errors that cause cyber incidents and breaches – can enable nefarious cyberattacks to become successful resulting in a range of potential harms at an individual and organisational level. Managing unintentional insider threat is a growing challenge for organisations and businesses. Emerging work in this area has considered the phenomenon from various perspectives including the technological, the psychological and the sociotechnical. However, there is a gap in terms of (a) investigating unintentional insider threat specifically (rather than being centred on intentional or malicious insider threat) and (b) a human centric approach whereby technologies and humans are considered equally in a sociotechnical context of cyber and physical spaces in which they coexist. In order to address this deficit, this thesis investigates unintentional insider threat to uncover factors that influence it by adopting a human-centric lens to address this challenge.

A human factors theory-informed systems approach is used to evaluate and critically analyse related work. Through the application of Critical Decision Method and Theory of Planned Behaviour approaches in two linked studies, a framework is developed and validated through engagement with industry. It is suggested that unintentional insider threat is responsive to a range of factors that can be linked to the individual, the technique used in the attempted attack and, the wider work environment and culture. While attitudes towards human elements within organisational ecosystems are improving, subjective norms can be leveraged to foster the creation of innovative cybersecurity defences in the future. This thesis contributes a tool to enable organisations to reflect on the relevance of unintentional insider threat within their overall approach to cyber security, and provides contributions to human-centred theoretical and practical understanding of unintentional insider threat. Ultimately, it is argued that in order build to meaningfully tackle this threat all actors must be leveraged to take advantage of the understanding developed in this work to enhance existing systems in which the human element is critical to keeping systems safe.

Acknowledgements

My endless thanks to Sarah Sharples and Robert Houghton, my supervisors, who provided me with their unwavering support, guidance, immeasurable patience and wisdom. I recognise my incredible stroke of luck to have them agree to oversee this project, a time that was filled with out-of-the-box thinking, rethinking and, perhaps most importantly, laughter.

I am eternally grateful to the internal and external experts who offered me their time and valuable insights on this subject matter. Thanks too to the multiple industry partners who deemed me good enough to collaborate with and believed in me when I struggled in doing that for myself. I am incredibly grateful to Derek McAuley and Jason Nurse who kindly agreed to be my assessors.

Thanks to Sarah Sharples, Steve Benford and, EPSRC for funding this project. I am incredibly thankful to the Human Factors Research Group in the Engineering Department who welcomed and accepted me into their fold and thanks to the CDT's 2018 cohort for their support and general empathy.

Finally, my deepest gratitude to Aabida Khan, my mother, for encouraging me to embark on this journey, my siblings and Ryan Ahmed – all of whom have been endlessly patient and supportive.

Table of Contents

Table of Figures

1. Introduction

1. Introduction

The spread of internet-enabled services and devices into the workplace has led to significant gains in productivity and efficiency (Schuh et al., 2014). However, this technology also offers potential vulnerabilities and new attack surfaces for criminals, industrial saboteurs and extortionists to exploit. Potential vulnerabilities that result in the exposure of personal or sensitive data are also a matter of widespread concern and media interest. Aside from what might be considered traditional hacking of digital systems at a technical level, there is increasing prevalence of cyberattacks that require the unwitting participation of innocent individuals in terms of opening an attachment, clicking on a rogue link or otherwise inadvertently performing an action that compromises a system (Verizon, 2020).

This innocent facilitation of insiders to successfully cyberattack systems is considered a subset of "Insider Threat" known as unintentional or accidental insider threat. Unintentional or accidental insiders are those individuals who unknowingly or unwittingly harm the organisation through their actions due to being manipulated to click on a malicious link, install malicious software or otherwise facilitate a cyberattack. This category of unintentional insider threat is the focus of this thesis. The remainder of the category is known as intentional or malicious insider threat comprising of deliberate and malicious actions carried out by disaffected or mercenary employees within an organisation (Mundie et al. 2013).

A range of solutions have been proposed to address intentional and unintentional acts. Solutions tend to address both vulnerabilities that arise from the human element as well as technological aspects, such as those arising from software. Defences tend to focus on the technological elements rather than humans or processes (Ani et al., 2018). When systems are compromised organisations assess intentions behind insider's actions as either intentional or unintentional in order to determine the intensity of organisational response and subsequent reprimands (Predd et al., 2008). This desire to control and manage the human element, which is believed to be the generator of unacceptable risks, stems from traditional security thought whereby humans are perceived to be the weakest link in the security chain (Mittal, 2015; Ani et al., 2018).

Psychological and behavioural approaches have been utilised to further traditional security thought by devising novel solutions to assess intentions behind insider actions if systems are compromised. Such approaches entail creating individual and group psychological and behavioural profiles typically with psychometric tests used to predict stress susceptibility. Other approaches emphasise identifying rule breaking behaviour through background checks and examination of personnel records from the Human Resources department. Once developed these profiles can provide an insight into the intentions of insiders should a breach occur. Furthermore, triangulation of this personal data may also be used as early markers for potential insider threat (Cappelli et al., 2007; Greitzer et al, 2018; Kandias et al., 2010).

Where local legal regulations are in effect that bar or limit the collection of personal data on individuals, alternative behavioural approaches have emerged to tackle insider threat. These approaches disregard intentions or motivations of insiders and focus instead on controlling opportunities afforded to individuals when interacting with technologies within systems. Opportunities afforded to individuals to compromise secure systems are determined through analysing network based behaviour with access logs to determine a baseline of acceptable behaviour i.e. *normal behaviour*. Once this baseline has been identified it is used to evaluate daily actions and eradicate *abnormal behaviour* through restricting users from accessing certain parts of the system and information, unless it is justified by the organisation through a ‘case for exemption’ (Agrafiotis et al., 2015; Chattopadhyay et al., 2018; Legg et al., 2015). Whilst these techniques have brought forward a diverse set of propositions to expand solutions stemming from traditional security thought, these techniques have had limited success in addressing unintentional insider threat arising from well-meaning insiders.

To begin a meaningful discussion about unintentional insider threat, the parameters of what constitutes as insider threat must first be established. Establishing these parameters is problematic due to the multifaceted nature of insider threat which results in an abundance of definitions present in literature to define this term. However, for the purposes of this thesis insider threat is defined as follows:

'*Actions [encompassing skills, rules and knowledge-based behaviour] or inaction of individuals or groups who wittingly or unwittingly cause loss or harm to the security of an organisation, without a differentiating between cyber or physical perimeters. The individual(s) has authorised access [physical and/or cyber] to physical assets and to confidential information in order to perform a function for an organisation which results in compromised safety or a cybersecurity breach.*'

Nested in the above definition of insider threat, unintentional insider threat is defined as follows:

'*Insider threat that is not a result of intentional actions that cause loss or harm to an organisation by insiders.*'

In contrast to previous approaches that have retained a focus on technologies or on identifying weaknesses in individuals, the work in this thesis adopts a systems perspective with which to view and understand unintentional insider threat by changing the way humans are considered within systems. This will be done through including a range of established approaches i.e. The Epidemiological Triangle (Cassel, 1976), the Swiss Cheese Metaphor (Reason, 1990a), Safety II approach (Hollnagel, 2018), *S*kills, *R*ules and *K*nowledge approach known as SRK (Rasmussen, 1983) and, Generic Error-Modelling System known as GEMS (Reason, 1990b). Epidemiological Triangle (Cassel, 1976) and the Swiss Cheese Metaphor (Reason, 1990a) can provide a useful visual aid for representing and understanding the interdependent and dynamic relationship that exists between vectors within an environment. An implication of these approaches is that focusing on a single vector can be problematic and ineffective for proposing solutions to complex challenges. Safety II approach (Hollnagel, 2018) is used to categorise existing approaches and acknowledge the variability in human performance that keeps systems safe whilst learning from what works well and goes *right* as well as what goes wrong. Thus, a Safety II approach provides a further dimension to aid in understanding the environment under which unintentional insider threat occurs. Safety II approach argues that incidents or accidents are not unique events but rather an expression of the variability within human performance. Humans are the necessary element in the system that provide systems with the flexibility and resilience needed for safe operations and production of desirable outcomes. Being equipped with a Safety II approach means that learnings are acquired through understanding what goes right a vast majority of the time as well as when things go awry. The inclusion of skills, rules and knowledge based behaviour known as the SRK approach (Rasmussen, 1983) aids in examining the types of tasks that result in errors as it is important to understand the types of tasks being performed and the cognitive load on individuals during which systems are unintentionally compromised. Additionally, to gain a deeper understanding of erroneous actions which lead to unintentional insider threat Generic Error-Modelling System known as GEMS (Reason, 1990b) is utilised to classify the types of errors that occur. For instance, when a well-intentioned insider unwittingly compromises a system it could occur from a slip in attention, a lapse in their memory, a mistake in the classification of their memory or a routine violation that had been occurring in the past but never resulted in a cyber breach.

This body of work is grounded in the above approaches to understand unintentional insider threat, i.e. the types of tasks that lead to unintentional insider threat, the types of errors that result in it and, the variability in human performance that keeps systems safely operating a vast majority of the time. It is through this human centric lens that unintentional insider threat is investigated in order to propose a new approach to enhance existing understandings and solutions.

1.1 Research Questions

From this brief introduction it is clear that new attack surfaces have been created with the rapid widespread adoption and creation of connected technologies. The necessary human element that enables cyberspace operations has become a major vector that facilitates cyberattacks. Insider threat (unintentional and intentional) poses a paradox for cybersecurity whereby humans are necessary to enable operations whilst they generate or enhance vulnerabilities in systems. This paradox makes it challenging to address insider threat especially if it is unintentional in its nature. Proposed solutions address intentional *and* unintentional insider threat in tandem and appear to be focused on either protecting the technological element or leveraging it in order to control, manage or limit the human element within cyberspace. Additionally, psychological approaches have also emerged which aim to predict or ascertain intentions behind actions that result in breaches. Therefore, it is of interest to explore the extent to which humans are considered within systems, the suitability and comprehensiveness of proposed approaches to insider threat and, learn about unintentional insider threat from lived experienced of those that have had exposure to it.

Interest also arises from developing a framework that can be utilised by organisations to reflect on how unintentional insider threat can be examined, understood and defended against. Thus, this thesis seeks to explore the following research questions.

1. To what extent are current cybersecurity approaches considering operations of the human element?

Through reviewing extant literature and conducting systematic analysis of existing tools, this question aims to identify the limitations and scope of current approaches and show the opportunities of being able to apply an alternative lens with which unintentional insider threat can be understood.

2. How might a sociotechnical systems approach aid in reframing current approaches from a human centric stance?

This research question aims to explore the extent to which current approaches are suited to unintentional insider threat, the extent to which these approaches are holistic in a sociotechnical context and, opportunities for human factors domain to propose solutions.

3. What can be learned from people's experience of unintentional insider threat about factors that influence it?

This research question applies Critical Decision Method (CDM) to understand individual experiences that led to unintentional insider threat in order to validate current approaches and introduce new elements for consideration to safeguard against such a threat.

4. What user centric solutions could have a positive impact in an open environment for understanding unintentional insider threat?

This research question explores the extent to which the developed sociotechnical framework can prompt individuals to reflect on challenges posed by unintentional insider threat in organisational contexts.

1.2 Industry Engagement

Numerous industry collaborations occurred over the course of four years to help ground this research in an industry perspective. These experiences established challenges associated to unintentional insider threat and, provided distinct insights at various points that guided this research (detailed in Appendix 1). Warwick Manufacturing Group (WMG) is an academic department that facilitates collaborations between academia and industry. Since WMG provided access to the second industry partner, High Value Manufacturing Catapult, and shared insights and challenges in the context of its industry partners rather than academic perspectives, WMG was considered an industry partner in the context of this work. Industry contributions towards this research are as follows:

- WMG and High Value Manufacturing Catapult: During the first year the initial industry partner shared need-based examples of challenges pertaining to cybersecurity. This input informed the PhD proposal made to the Centre for Doctoral Training
- Connect Places Catapult (CPC): In the second year, a full-time three month placement was carried out for an immersive ‘in-the-field’ experience. This embedding in industry setting aided the author in understanding the complexity of ownership and responsibility for cybersecurity in the design of technologies
- National Cyber Security Centre (NCSC): In the third year of this project, another full-time three month placement occurred. This experience enhanced the author’s

expertise of applying models from the human factors domain to cybersecurity challenges

- NCSC Partners: Collaborations occurred with six organisations who contributed to the research findings from a study to prompt a change in behaviour through eliciting reflection

All industry partners independently appeared to be in agreement that cybersecurity and insider threat were important concerns to stakeholders and a number of insights emerged from the three different industrial placement activities. A deeper understanding was developed for the nuanced complexities that exist in real-world settings on top of which cybersecurity is designed to be implemented. For instance, despite cutting-edge cybersecurity solutions being implemented, a mismatch between individual and organisational priorities can result in the overall cybersecurity being compromised by well-meaning insiders. However, if individual priorities and reasons for performing undesirable actions are not given due consideration (i.e. the case made by Safety II, Skills-Rules-Knowledge approach and, GEMS) or wrong lens with which to examine the problem is adopted, it can result in arduous efforts that are fruitless. Additionally, while reprimands can yield short-term results individuals might revert to the same actions in the long-term or worse, create a new set of unforeseen challenges that emerge from individuals trying to achieve the same outcomes in new ways.

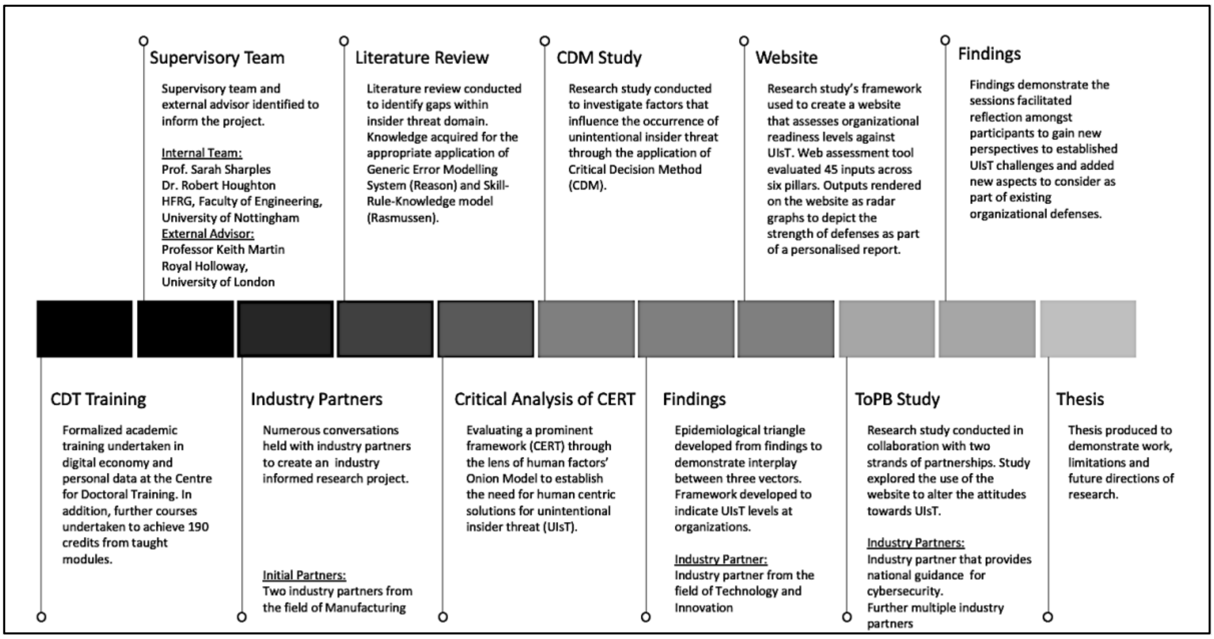
Similarly, when cybersecurity is retrofitted or superimposed on existing structures, it can reinforce the mystique associated to this domain. For instance, efforts made to build awareness and generalist knowledge about cybersecurity can result in a disconnect between top-down mandates and bottom-up efforts of how work is being performed and measured. This disconnect can contribute to a widening of the gulf between work-as-imagined and work-as-done (Hollnagel, 2017; Suchman, 1987). The relatively innocent action of not fully incorporating cybersecurity advice (which might be driven by the fear from technology companies not fully understanding cybersecurity or imposing responsibility of cybersecurity onto individuals which is intrinsically tied with their key performance indicators) can result in technology being taken to market that has not incorporated cybersecurity as part of its design.

In the context of imposing responsibility for cybersecurity elements onto individuals, which is recommended in the frameworks which will be discussed in Chapter 2, the situation becomes multifaceted when there are competing interests (such as their key performance indicators). Furthermore, while generalist cybersecurity knowledge can be developed by organisations on an individual level, a singular person responsible for the overall cybersecurity can create a sentiment of absolute authority and create a channel for reprimands. It can also serve to alienate cybersecurity knowledge and personnel from mainstream operations, add to the mystique of the domain by differing to an ‘identified, responsible expert’ and inevitably create a singular point of failure.

Embedded experiences in industry settings highlighted cybersecurity challenges in real-world settings and, provided context and informed this body of work. From the detailed experiences shared in Appendix 1, the approach adopted by industry is inclined towards the technological element for protecting against insider threat which is reflective of the relevant literature discussed in Chapter 2 i.e. technological element within sociotechnical systems is leveraged to limit or control the operation of the human element within cyberspace. As insider threat is understood to be dynamic and sudden in its nature (Nurse et al., 2014), the agility needed to respond to unintentional insider threat can be limited when cybersecurity is superimposed onto existing systems or in instances where the responsibility is wholly placed on the human element. Furthermore, placing emphasis on certain elements is not holistic and can especially fall short in its consideration of humans within systems. Thus, embedded industry experience provided a backdrop for the design of a human centric framework to tackle unintentional insider at organisations.

1.3 Statement of novelty and expected contribution

This work has been informed by a multiple disciplinary approach and by industry input through numerous partners. The diagram below reflects eleven distinct stages over the course of four years of this research project:



CDT Training

Formalized academic training undertaken in digital economy and personal data at the Centre for Doctoral Training. In addition, further courses undertaken to achieve 190 credits from taught modules.

Supervisory Team

Supervisory team and external advisor identified to inform the project.

Internal Team:
Prof. Sarah Sharples
Dr. Robert Houghton
HFRG, Faculty of Engineering,
University of Nottingham
External Advisor:
Professor Keith Martin
Royal Holloway,
University of London

Industry Partners

Numerous conversations held with industry partners to create an industry informed research project.

Initial Partners:
Two industry partners from the field of Manufacturing

Literature Review

Literature review conducted to identify gaps within insider threat domain. Knowledge acquired for the appropriate application of Generic Error Modelling System (Reason) and Skill-Rule-Knowledge model (Rasmussen).

Critical Analysis of CERT

Evaluating a prominent framework (CERT) through the lens of human factors' Onion Model to establish the need for human centric solutions for unintentional insider threat (UIsT).

CDM Study

Research study conducted to investigate factors that influence the occurrence of unintentional insider threat through the application of Critical Decision Method (CDM).

Findings

Epidemiological triangle developed from findings to demonstrate interplay between three vectors. Framework developed to indicate UIsT levels at organizations.

Industry Partner:
Industry partner from the field of Technology and Innovation

Website

Research study's framework used to create a website that assesses organizational readiness levels against UIsT. Web assessment tool evaluated 45 inputs across six pillars. Outputs rendered on the website as radar graphs to depict the strength of defenses as part of a personalised report.

ToPB Study

Research study conducted in collaboration with two strands of partnerships. Study explored the use of the website to alter the attitudes towards UIsT.

Industry Partners:
Industry partner that provides national guidance for cybersecurity.
Further multiple industry partners

Findings

Findings demonstrate the sessions facilitated reflection amongst participants to gain new perspectives to established UIsT challenges and added new aspects to consider as part of existing organizational defenses.

Thesis

Thesis produced to demonstrate work, limitations and future directions of research.

Figure 1: Stages of contribution

Whilst work has been done to address unintentional insider threat by the computer science domain and recently there has been an emergence of techniques from other disciplines to contribute solutions to this challenge, there have been limited contributions from a human centric approach. There is also a lack of exclusive examination of unintentional insider threat from its counterpart intentional insider threat. This thesis extends the application of existing approaches from risk and safety engineering and human factor domains to change how humans are considered within systems in order to enhance existing understanding of unintentional insider threat. Furthermore, a sociotechnical framework is presented that is anticipated to assess the strength of barriers in place to determine organisational readiness levels against this threat, developed through the implementation of multiple disciplinary perspectives. This work has also benefitted from numerous industry collaborations at various stages and the framework aims to provide industry with a range of novel sociotechnical factors to consider as part of their defences. It is hoped that through focusing on unintentional threats specifically (rather than more commonly studied intentional threats) and by extending the application of established approaches, it will provide a new approach with which to understand and respond to unintentional insider threat in industry.

1.4 Publications arising from this thesis

Abridged sections of this thesis have been published in the following articles. It is worth noting that the ability to engage with conferences was limited due to the Covid-19 pandemic.

Chapters 2, 5 and 8: Khan, N., J Houghton, R., & Sharples, S. (2022). Understanding factors that influence unintentional insider threat: a framework to counteract unintentional risks. Cognition, Technology & Work, 24(3), 393-421.

Chapters 1, 7 and 8: Khan, N., Sharples, S., & J Houghton, R. (*Submitted on 14/11/2022 to Cognition, Technology & Work*). A human centric approach: presenting a framework to influence understanding of unintentional insider threat.

1.5 Structure of the thesis

Having introduced the motivations to adopt a multiple disciplinary approach to investigate unintentional insider threat within this Chapter and the industry embedded nature of this work, the structure of this thesis is as follows:

Chapter 2 sets out a literature review to discuss existing approaches proposed to tackle unintentional insider threat and introduces sociotechnical theory and perspectives from the human factors domain.

Chapter 3 investigates the real-world challenges that emerge from cybersecurity recommendations offered by Computer Emergency Readiness Team (CERT) through application to SME scenarios and mapping recommendations to the onion model as a way to recontextualise and evaluate suggestions from a sociotechnical human factors perspective.

Chapter 4 discusses the methods and findings of a research study designed to investigate factors that influence unintentional insider threat. This study applies Critical Decision Method (CDM) approach to elicit knowledge from those that have been compromised to create a sociotechnical framework.

Chapter 5 details the process of creation and design of a website which is inspired by Action Design Research principles.

Chapters 6 shares the design, methodology and results of a research study that is inspired by the Theory of Planned Behaviour (ToPB) approach to examine changes in behaviour amongst participants. The website held the sociotechnical framework and produced a personalised organisational report for readiness levels against unintentional insider threat to aid in the behavioural shift amongst participants.

Chapter 7 holds a discussion of the work presented in this thesis, provides concluding thoughts in the context of the proposed research questions and, discusses contributions before presenting limitations arising from this work and recommendations for future research avenues.

Covid-19 Statement:

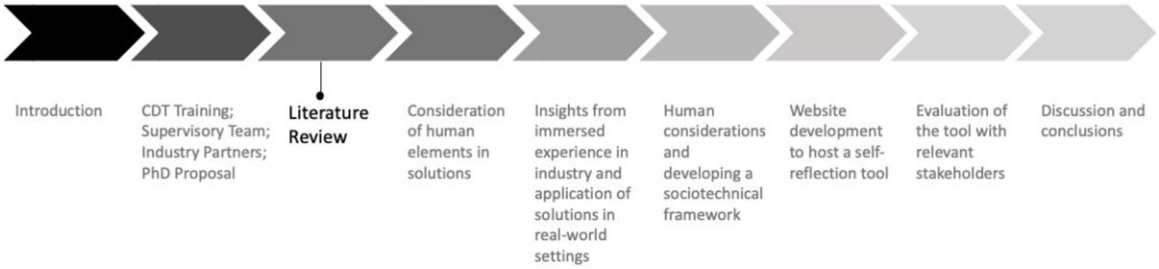
In order to acknowledge the impact of the Covid-19 pandemic, this section discusses the subsequent adaptations made to this research project. The pandemic caused severe delays to an industry collaboration with a partner due to a backlog of security clearance and consequently to the second research study inspired by the Theory of Planned Behaviour. This study was planned to be conducted in-person so as to add an additional level of comfortability for participants when sharing confidential information about their organisations. Being on premises was also believed to ease availability of senior stakeholders to be present in sessions simultaneously which was a mandatory requirement for participation. However, since lockdown regulations were still in effect at the time participants were recruited, sessions were held via online platforms at times most convenient to the participants. Session designs were adjusted whereby senior most participant in each session shared their screen with other participants present in the session including the interviewer. Due to the increased demands on diaries during remote working, sessions were also split into two sessions if requested by the participants due to existing diary commitments or clashes for availability amongst participants.

It has also not escaped the author’s notice that increased remote working as a consequence of the pandemic increased the occurrence of unintentional insider threat in personal and professional lives of individuals globally.

This Chapter introduced the challenges associated to unintentional insider threat, research questions that this thesis sets out to investigate, a statement of novelty and expected contributions, structure of this thesis and, the impact of covid-19 pandemic on this research project. The following Chapter examines extant literature to tackle unintentional insider threat and introduces human centric perspectives to offer a lens with which errors can be understood and examined.

2. Literature Review

2. Literature Review


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

The previous Chapter introduced this research and made a case was made for investigating unintentional insider threat (UIT) to enhance existing solutions.

To build an understanding of what enhanced solutions might entail, this Chapter provides a lens with which to understand cybersecurity within the context of this thesis. It presents: the types of threats that make systems vulnerable (i.e. software and human); theoretical approaches underpinning popular extant solutions and; the challenges associated in applying these techniques in the context of UIT. Three major types of cyber-attacks are presented that leverage UIT alongside their respective solutions to counteract these unintentional threats. Notable frameworks designed to identify and prevent UIT are reviewed to demonstrate the way humans or the human element is considered in systems.

Sociotechnical theory and pertinent perspectives are introduced to shift the perspective of the ways in which humans can be considered in systems. Sociotechnical perspectives include the Epidemiological Triangle and the Swiss Cheese Metaphor both of which examine interdependent relationships in environments, Safety II approach is introduced to shift away from exclusively investigating and learning from errors by acknowledging the variance in human performance, Skills-Rules-Knowledge based behaviour (SRK) approach is discussed that provides taxonomies of the types of tasks that can result in undesirable outcomes and, Generic Error-Modelling System (GEMS) that provides a taxonomy for the types of errors that can arise from tasks such as those that result in cyberbreaches.

To begin this discussion, insider threat must be defined to provide context to its subset of unintentional insider threat. All aspects that pertain to intentional or malicious insider threat are beyond the scope of this project. However, by virtue of discussing intentional insider threat within this Chapter, unintentional insider threat is provided context. Discussing both the subsets of insider threat maintains the approach adopted by literature and industry solutions to this challenge whereby intentional insider threat is considered in tandem with unintentional insider threat. The term "insider threat" in this project is defined as follows:

'*Actions [encompassing skills, rules and knowledge-based behaviour] or inaction of individuals or groups who wittingly or unwittingly cause loss or harm to the security of an organisation, without a differentiating between cyber or physical perimeters. The individual(s) has authorised access [physical and/or cyber] to physical assets and to confidential information in order to perform a function for an organisation which results in compromised safety or a cybersecurity breach.*'

Furthering the above understanding, unintentional insider threat is defined as follows:

'*Insider threat that is not a result of intentional actions that cause loss or harm to an organisation by insiders.*'

2.1 Overview of Cybersecurity

Despite the term cybersecurity penetrating almost all aspects of information technology there is a lack of agreement in literature as to its definition (Choucri et al., 2012). Cybersecurity can entail a wide range of topics that exist within the space of networked computing devices. Most commonly, cybersecurity is defined as the freedom from harm in cyberspace and involves the so-called “CIA triad”: ‘Confidentiality’ (C) where information does not suffer disclosure to anyone unintended, ‘Integrity’ (I) where the information is not modified or deleted and ‘Availability’ (A) where data is accessible in a timely manner when needed by authorised users (Weber and Studer, 2016; B von Solms and von Solms, 2018).

Real-time data, agile networks and growth in technological capabilities has created a host of new challenges, particularly those associated with controlling and safeguarding information. Existing, new and emerging technologies all consistently redesign the research landscape by expanding the cybersecurity environment making it a precarious domain and resulting in scientists, researchers, practitioners and analysts rapidly shifting their understandings and re-positioning their approach to tackle this problem (Goethals and Hunt, 2019).

2.1.1 Cyberspace Operations

There are two main categories that define defence operations within this space based on intentions: offensive or defensive. According to Goethals and Hunt (2019), ‘Offensive Cyberspace Operations’ (OCOs) are still quite understudied while Defensive Cyberspace Operations (DCOs) are better researched and approach a threat from a defensive stance. DCOs in literature can largely be categorised as: ‘passive’ or ‘active’ defences. To visualise this understanding of cyberspace operations (Goethals and Hunt, 2019) the author has created Figure 2 below.

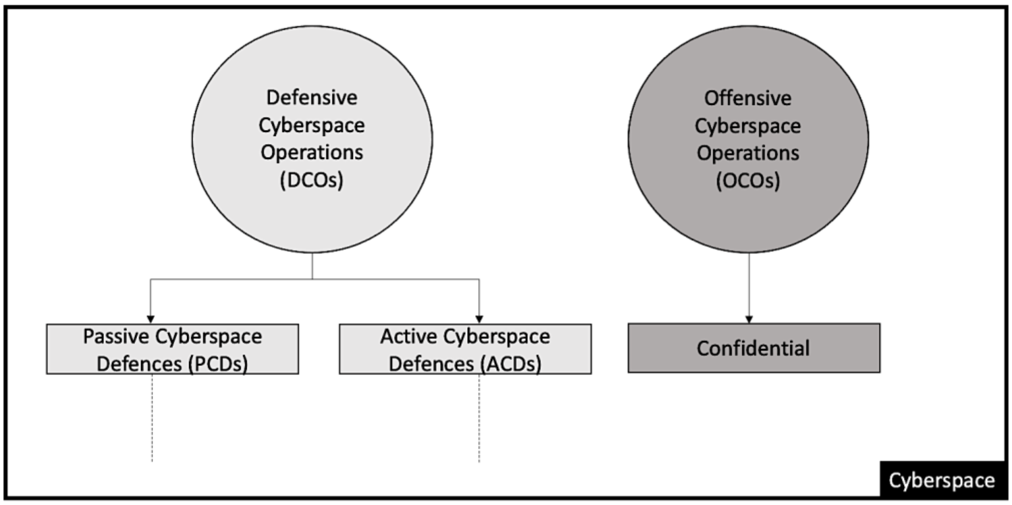

Defensive Cyberspace Operations (DCOs)
Offensive Cyberspace Operations (OCOs)
Passive Cyberspace Defences (PCDs)
Active Cyberspace Defences (ACDs)
Confidential
Cyberspace

Figure 2: Taxonomy of Cyberspace operations based on intentions

Within DCOs, ‘Passive Cyberspace Defences’ (PCDs) involve ‘best practice’ implementation for setting up systems, systems monitoring and exchanging information. This avoids vulnerabilities in the system in terms of how its set-up that can prevent attacks from penetrating the system or information being highjacked in transit. In addition to being legally permissible, PCDs do not involve covert or overt monitoring of user activities, are not concerned with individual intent, motivation, psychological disposition or behavioural patterns. Examples of PCDs include configuration management, encryption (symmetric and asymmetric), configuration monitoring, data management (storage, access and architecture). In some instances, PCDs can also include Intrusion Detection / Prevention Systems such as anomaly based, signature based and stateful protocol detection (Magklaras and Furnell, 2001). Through the use of some examples various approaches to cyber defences are represented in the ontology below (Figure 3 created by the author that is inspired from Figure 1 by Goethals and Hunt, 2019). While this list of PCDs and ACDs in Figure 3 is not exhaustive, it is presented to demonstrate the nature of cyber defences and their approach to protect or leverage the technological element for instance, for monitoring purposes.

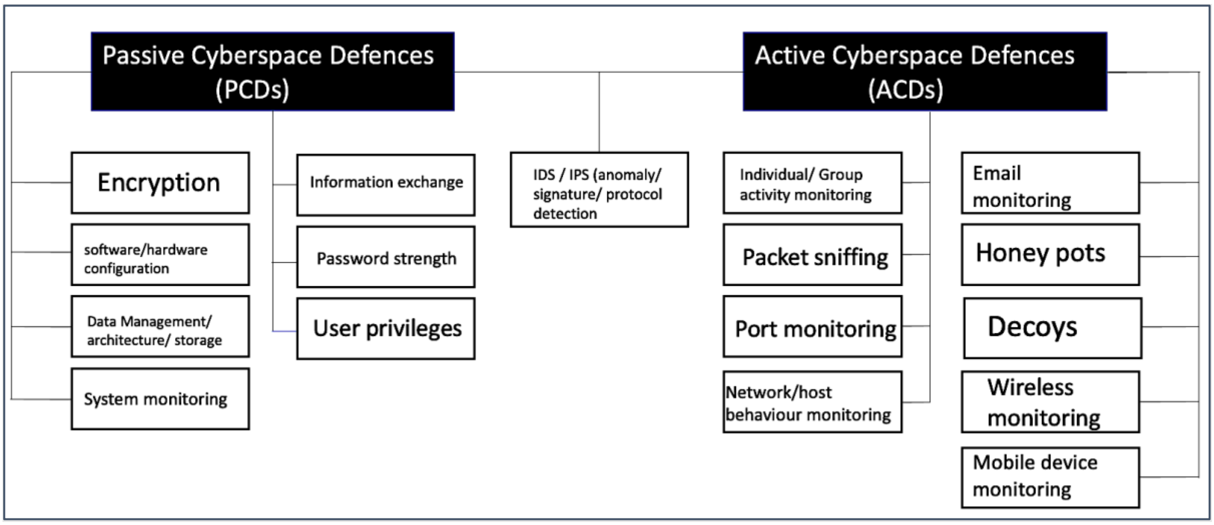

Passive Cyberspace Defences (PCDs)
Active Cyberspace Defences (ACDs)
Encryption
software/hardware configuration
Data Management/ architecture/ storage
System monitoring
Information exchange
Password strength
User privileges
IDS / IPS (anomaly/ signature/ protocol detection
Individual/ Group activity monitoring
Packet sniffing
Port monitoring
Network/host behaviour monitoring
Email monitoring
Honey pots
Decoys
Wireless monitoring
Mobile device monitoring

Figure 3: Ontology for passive and active cyber defences

2.2 Threats

Before beginning to explore the attacks that all cyberspace is perpetually at risk of, such as the popular attacks being faced today and the defences built to circumvent these attacks, core vulnerabilities must first be explored that can enable successful cyberattacks. It is with the understanding of where vulnerabilities emerge that the strategies and defences can be understood and evaluated for their effectiveness and robustness. From the discussion above about the defensive cyberspace sphere of PCDs and ACDs, two further categories of defences can be created: (i) to counteract software vulnerabilities and, (ii) to counteract human action/interaction.

2.2.1 Software vulnerabilities, threats and solutions

All software has vulnerabilities due to software developers' fallibility (Ani et al., 2018). Borrowing a scheme of categorisation from Rumsfeld (2011), these vulnerabilities can be divided into four categories: 'known', 'known unknowns', 'known knowns' or 'unknown unknowns'. 'Known' categories can include attacks that occurred from vulnerabilities known to the software developers and/or the organisation. 'White-hat hacking' also known as ‘ethical hacking’, and external penetration testing are usually conducted to help the organisation measure the obviousness and weakness of such vulnerabilities (Yaqoob et al., 2017; Sood et al., 2015). ‘Known unknowns’ category involves knowledge of vulnerabilities based on logic and so would include attacks that could not have been forecasted but expose an obvious vulnerability when an attack has occurred. ‘Known knowns’ include vulnerabilities that were known not just to the developer and/or the organisation but the wider community who has interest in this space for either protecting or attacking purposes. This can include examples of popular cyberattacks on services and systems that had previously enjoyed a reputation in the public opinion for being cyber ‘safe’. For instance the spyware attack in 2019 on a popular mobile messaging service application called WhatsApp. Nefarious parties sent malicious links to select victims who were tricked into clicking a link that would install Pegasus spyware. Once installed, this spyware collected location data, call logs, contacts and, highjack the phone’s camera and microphone (Serrano, 2021). WhatsApp was potentially the platform of choice for the hackers as potential victims would be lulled into believing that the platform is safe, primarily due to WhatsApp’s self-promotion of utilising end-to-end encryption technique. Another example of exploitation of known knowns was the ransomware attack WannaCry on the National Health Service (NHS) in the UK in 2017. WannaCry exploited a specific vulnerability in the Microsoft Windows 7 operating software if it was left unpatched i.e. updates recommended by Microsoft had not been installed that eliminate known software vulnerabilities. Through Microsoft’s public service messages to IT personnel over a period of twelve months to urgently install updates, communities with an interest in this space were aware of the fact that leaving the operating system unpatched could facilitate attacks. ‘Unknown unknowns’ are vulnerabilities that are simply unknown to everyone involved until an attack happens and there is no actionable way of building defences against it. Examples of this can include ‘0 day attacks’ or ‘zero day attacks' where no one is aware of the vulnerability that is going to be attacked until it happens and is seen to be an unexpected and surprising event to everyone and attackers were unaware of this vulnerability (i.e. a *lucky* break) and/or the scale of disruption the attack would cause.

Given this fallibility in software, constant evolution of existing software and, introduction of new and emerging software into the cyberspace environment threats in cyberspace are continuously changing. This change poses its own set of challenges for researchers in this area to create and implement effective solutions. Unsurprisingly, a majority of solutions place software at the heart of their approach as it is arguably easier to tackle software vulnerabilities than holistically address elements within complex systems. Currently popular software solutions involve monitoring of system logs to incorporate multidimensional aspects to build either passive or active defences. Examples of this include techniques such as those found in cyber-physical systems (including those from environmental sensors), stateful protocol detection and anomaly based identification (Zargar, 2016), network based and wireless based activity found in intrusion prevention systems, (end-to-end) encryption, data storage and its architecture, anti-malware and antivirus software and, regular patches and updates for existing software.

2.2.2 Human vulnerabilities and threats

It is insufficient to discuss cybersecurity that takes measures to counteract threats without discussing the human element that enables cyberspace operations. The human element is a key component in cybersecurity as humans are seen to form a second line of defence after software robustness. This means that regardless of how robust the programming language is for an application or how intelligent a counteracting software is (such as antivirus), human interaction can make executive decisions that can result in threats being realised – creating a new classification of vulnerabilities through their operation. While software-centric counter solutions to threats are complex, they are proving to be less challenging than human vulnerabilities.

> '*AIC (availability, integrity and confidentiality) security triads have been noted to be too focused on securing technology elements, and not enough to protect other elements such as people and process'*
>
> – Ani et al., 2018

Human vulnerabilities are separate and distinct from programmers fallibility, ultimately manifested in software that is discussed in the previous section. Instead here, human vulnerabilities encompass the human element's interaction within cyberspace that can result in threats being realised.

Generally, insider threat (IsT) is understood to be the human element that undertakes actions and makes executive decisions that can potentially realise threats. IsT is a well-known phenomenon dating back to the 1980s (Chinchani et al., 2005) and is believed to be the element that creates vulnerability in systems and infrastructure, assets and/or data that can emerge from the actions or inactions of 'insiders' as a consequence of their access privileges, proximity to and knowledge of systems as well as their skills and motivations.

However, a formal definition of *insiders* in literature is either absent, ambiguous or disputed (Mundie et al., 2013; Goethals and Hunt, 2019; Hunker and Probst 2011). This lack of definition hampers research efforts as approaches do not clearly indicate the specific type of insider threat they aim to detect and, limits the ability to compare approaches that exist for each type of insider threat (Bishop and Gates, 2008a). However, with the widespread global adoption of technologies that have transformed personal and professional lives defining *who* qualifies as an insider, and additionally under what cyber and physical conditions, has become problematic. In order to establish an agreed definition for the term *insider* firstly, there would need to be an agreement within the international community on its definition which is reflected in law, policies and the governance of cyberspace especially during conflict or when there are competing state interests. Secondly, numerous factors would need to be agreed upon when identifying who might qualify as an insider. These factors can range from micro to macro levels depending on the scenario being considered. For instance, individuals' cyber and/or physical access to information or assets, role of the individual, time commitment from the individual (and thus exposure to information), timings of work, legal agreement with the individual, contracted (sub-contracted) individuals, geographical location of the individual, field of work and, the jurisdiction of law and policies, are all examples of such factors (Bishop and Gates, 2008a; Nurse et al., 2014).

Categories used to define insiders and insider threat (IsT) primarily rely on distinguishing actions based on motivations and intentions of the insider. For instance, Bishop and Gates (2008a) describe insider as "*a trusted entity that is given the power to violate one or more rules in a given security policy... the insider threat occurs when a trusted entity abuses that power*." Within this definition an insider is defined through the parameters set out by organisational security policy and access controls are being implemented (i.e. access to digital and physical information and resources). Also, there are two types of insider threat presented in this definition: i) breach of security policy through authorised access and, ii) breach of access control by *obtaining* unauthorised access.

Hunker and Probst (2011) argue that the motivation for investigating insider threat subsequently influences how insiders and insider threat (IsT) are defined. For instance, they state that in the United States insider threat investigations are driven by national security incidents whereas in the European region insider threat investigations are motivated by privately employed individuals who commit (financial) crimes and break laws. They further state that the definition of the *specific type* of insider threat being discussed in literature is derived from the audience's interest. Refraining from offering a definition for insiders and insider threat, the closest definition offered by Hunker and Probst (2011) is, "*We would observe that in practice – at least to the extent that we are able to observe real incidents – the problem of real interest is the "real real insider"; an individual deeply embedded in an organization, highly trusted, and in a position to do great damage if so inclined (e.g., a high level executive, or a systems administrator). At the same time it is this kind of insider and the threats he poses that are hardest to deal with*".

Predd et al. (2008) define insiders as follows, "*Insider: someone with legitimate access to an organization's computers and networks. Notice that we don't define what "legitimate" means and thus don't provide a single bright line distinguishing insiders from outsiders. Both legitimate access and the system's perimeter are a function not only of system-specific characteristics but also of a given organization's policies and values. For instance, an insider might be a contractor, auditor, ex-employee, temporary business partner, or more. Thus, the organization itself can best determine who is an insider*". Subsequently, insider threat is defined as, "*Insider threat: an insider's action that puts an organization or its resources at risk. Different insiders can pose very different types of risk, so many types of insider threats exist. A range of factors distinguishes them, and we can categorize insider threats according to risk. We consider four dimensions to understand these risks: the organization, the individual, the system, and the environment*". These two definitions for insiders and IsT indicate a parameter of understanding drawn by individual characteristics pertaining to knowledge and motivation as well as the organisational policies, role of the systems that enable threats and, local laws and ethics.

Despite the variation in how an insider is quantified or the lack of agreement in literature to define insiders evidenced above, there is a general agreement on two types of insider threat (IsT) that exist: intentional (also known as malicious) which can be posed by an individual or a group that exist in all cyberspace operations and, unintentional (also known as accidental) (Predd et al., 2008; Hunker and Probst 2011). It is the unintentional category that is of interest to this project.

Intentional or malicious insiders are those who largely act out of a vengeful emotional state followed by a negative work related event or unmet expectations and/or can involve personal financial rewards. This fundamentally encompasses the categories of *whistle blowers* and *disgruntled employees*, both categories enjoy considerable media attention for fraud, vandalism or sabotage. An example from disgruntled employees category is when a technology firm (Uber) acquired an employee from its competitor (Google) to advance their efforts in self-driving vehicle technology. Google filed charges against Uber for theft of intellectual property (IP) as they believed the ex-employee had taken software code that he had written whilst under Google's employment. This was sensationalised in international media for several months, with one heading titled '*Silicon Valley was built on job-hopping. But when a leader of Google's self-driving-car unit joined Uber, Google filed suit. Now the Feds are on the case*' (Duhigg, 2018). Whilst this article by Duhigg *appears* to be objective, it is common practice for offenders and their alleged accomplices to be villainised in order to demonstrate the 'bad apples' who acted independently from their teammates and wider colleagues, often portrayed to be driven by insatiable ambition and greed. This approach isolates the perpetrators from their wider social contexts in which they exist and the system that enables them to act inappropriately.

Unintentional or accidental insiders might not have meant to harm organisations but their actions can put assets and operations of the organisation at risk. Actions executed by unintentional/accidental insiders can include examples of hitting ‘reply all’ that can result in triggering of a Denial Of Service (DoS) attack, or clicking an email link that can result in a ransomware or phishing attack on the organisation’s network, resources and assets. The intention behind the action becomes important as this determines the adequacy and effectiveness of the subsequent organisational response. For instance, if the action was accidental or unintentional but resulted in a temporary suspension of the employee, it can create a harmful environment that can damage productivity, trust and, morale in the workplace and disincentivise reporting of behaviours/actions that present a security risk. However, if the same action was intentional or malicious, the offender and other employees can take further liberties in the future and it can encourage risk taking behaviours that increase organisational vulnerabilities to attacks in the future.

However, it is worth noting that in real-world settings work is not conducted in insolation from other parts of life. Often effectively performing work relies on a collaboration between individuals, systems and, organisations – all of which form important aspects of insider threat. On an individual level within an organisation, it is understood that work is distinct in its nature of how it is imagined, conducted and evaluated (Hollnagel, 2017; Suchman, 1987). Humans react to their environments and adapt to new or unfamiliar conditions particularly in regard to decision making. Coupling this with complex sociotechnical systems often translates into increased demands placed on cognitive functions and a fluctuating workload experienced by individuals. In fact, there have been some preliminary links made between workload, stress and unintentional insider threat in literature (Nurse et al., 2014; Kandias et al., 2010).

Thus, in order to change the way humans are considered in systems the terms of 'insider threat' and 'insiders' in this project are defined as follows:

'*Actions [encompassing skills, rules and knowledge-based behaviour] or inaction of individuals or groups who wittingly or unwittingly cause loss or harm to the security of an organisation, without a differentiating between cyber or physical perimeters. The individual(s) has authorised access [physical and/or cyber] to physical assets and to confidential information in order to perform a function for an organisation which results in compromised safety or a cybersecurity breach.*'

Derived from the above definition, unintentional insider threat is defined as follows:

'*Insider threat that is not a result of intentional actions that cause loss or harm to an organisation by insiders.*'

This definition is developed with an aim to incorporate the multifaceted features of insider threat and its dynamic nature reflected in the discussion above. With the definition for insiders, insider threat and, unintentional insider threat established, this work moves on to explore the approaches and the subsequent solutions to tackle this threat within systems.

2.2.3 Solutions for human vulnerabilities

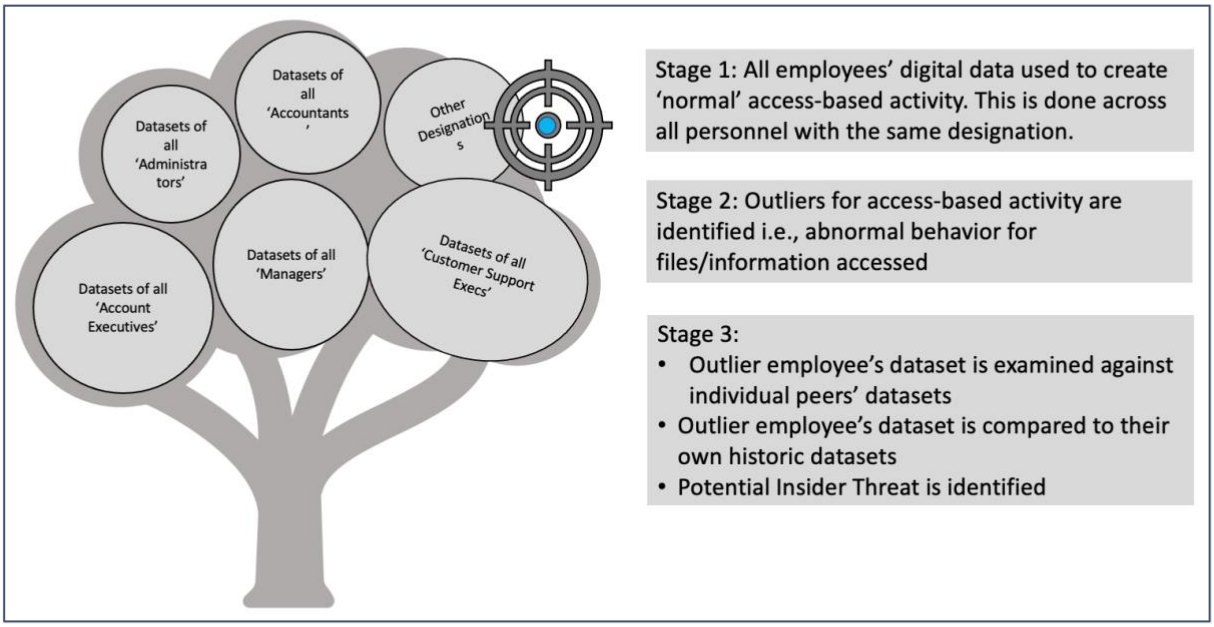
Approaches underpinning solutions:

In order to create solutions for unintentional insider threat (UIT) literature generally differentiates the 'offender' on their intentionality i.e. if they intended to do harm to the organisation or if it was accidental. A prominent framework driven from real-world breaches and incidents by US Computer Emergency Readiness Team (CERT), emphasises three primary features for a successful attack as: motive, skills and, knowledge (this is discussed in greater detail further on). This means that the evidence from an attack at a post-event forensic stage can identify the offender based on their (technical) skills, knowledge (of the company) and motives (disgruntlement/complaints/disciplinaries drawn from organisational records). The argument made by CERT is that since three elements existed in events that resulted in breaches, these traits can be reverse engineered to identify potential breaches. Thus, organisations were advised to keep a close eye on all employees that might possess the skills, knowledge of systems and processes and, are motivated to do harm. By this reasoning quite a vast net would need to be cast to identify and monitor insider threat. Depending on the nature of the organisation, people would possess an array of skills required to perform their respective tasks, knowledge about the company to operate within acceptable parameters and, ulterior or covert motives that might not be overtly exhibited for observation by others.

Where literature is not considering motivations, there is ample research considering the psychological and behavioural characteristics to identify insiders who might pose a threat which can range from detection of anomaly behaviour in employees' 'normal' day-to-day behaviour to background checks and personnel files that indicate 'rule breaking behaviour' such as violations of company policies (Bishop et al., 2008b; Greitzer and Hohimer, 2011; Kammüller and Probst, 2013; Ogiela and Ogiela 2012).

Nurse et al. (2014) further CERT's work by incorporating the dynamic and sudden nature that is understood to be a part of insider threat. They propose a framework to aid in understanding and reflecting on various aspects of insider threat. Through the use of insider threat case studies, this framework provides potential indicators for insider threat based on technical and behavioural aspects. Behavioural aspects include the use of psychological profiling through personality characteristics whereby intentional insiders are likely to be inclined to the Dark Triad traits of narcissism, Machiavellianism and, psychopathy while unintentional insiders being inclined to OCEAN Traits especially agreeableness and openness. These personality characteristics are utilised to identify the two types of insiders as well as the attackers i.e. to understand the motivations behind attacks in order to predict subsequent steps within an attack as it unfolds.

Beyond looking at motivations and the psychology of insiders, some literature moves to explore ‘opportunities’ available to employees that can facilitate insider threat (IsT), regardless of the employee’s motivation. Opportunities encompass themes such as access privilege, technical skills of perpetrators and, regulation of available software within an organisation. Legg et al. (2015) develop a ‘tree structure approach’ to examine IsT. A visual representation of this tree structure approach (created by the author of this thesis) is depicted in Figure 4 below.


Datasets of all ‘Accountants ’
Datasets of all ‘Administra tors’
Other Designation s
Datasets of all ‘Managers’
Datasets of all ‘Customer Support Execs’
Datasets of all ‘Account Executives’

Stage 1: All employees’ digital data used to create ‘normal’ access-based activity. This is done across all personnel with the same designation.

Stage 2: Outliers for access-based activity are identified i.e., abnormal behavior for files/information accessed

Stage 3:

- Outlier employee’s dataset is examined against individual peers’ datasets
- Outlier employee’s dataset is compared to their own historic datasets
- Potential Insider Threat is identified

Figure 4: Tree Structure Approach to insider threat

This approach involves creating a tree-like structure by incorporating datasets of all employees who perform the same duties at work and is usually grouped by job titles to form a ‘tree branch’. All employees’ datasets are added in a similar fashion to form multiple branches of that tree. Any outliers (who are accessing files that are not usual to their 'tree branch' or performing abnormal actions such as frequent access) are examined against their group's individual datasets, and/or their own historic datasets, to expose any possible threats. This is also known as a 'clustering approach' and commonly used as part of computational tools where the users are largely unspecified and is based on data from system logs. This individual data is then compared to their peers to develop individual behavioural patterns, where anomalies that can indicate new threats (Agrafiotis et al., 2015; Chattopadhyay et al., 2018).

Kandias et al., 2010 developed one of the models that combine techniques from computer science and psychology. This model monitors user activity in real-time to look for rule breaking behaviour or 'misbehaviour'. In addition, psychometric tests are used to identify individual susceptibility to malicious acts and stress levels that are believed to create vulnerabilities in organisational cybersecurity and enable insider threat. This model does have an important caveat to note which states that collection of such data must be legally permissible in the country of implementation but neglects to mention any ethical issues that can arise as a result of using personal information on individuals in this way.

The approaches discussed above are well suited to intentional insider threat where intentions exist prior to actions being carried out and while these approaches provide a good foundation for unintentional insider threat (UIT), there is opportunity to enhance solutions to better suit UIT. For instance, the offender would not require expertise in software development or knowledge of the internal IT department to realise an attack. On the other hand, even if the insider possess all of the above elements (motive, skills and knowledge) it would not directly correlate with them triggering a ransomware attack. As another example, an IT worker who is implementing a new software system for the organisation, could have accidently triggered a ransomware attack as they were experiencing a high workload during the time of implementation. Arming oneself with the approaches discussed above could mean crucial time lost during an investigation and relaxed efforts invested in understanding the circumstances around the incident, which can ultimately result in increasing the animosity between the employee and the organisation. When considering the use of personality characteristics some aspects presented in frameworkes can be enhanced for their application. For instance, a minimum level of expertise required from existing staff before they can conduct personality evaluations, methods for determining the motivations of insiders, guidelines for ethical collection and processing of data and, introducing additional risk assessments for potential legal and ethical challenges that can arise as a result of using personal information on individuals. With the application of clustering approach, well-intentioned insiders who might be performing additional responsibilities can repeatedly be identified as ‘malicious’ since they might access a wider set of information than their peers (who might experience lower workloads and responsibilities). Aside from the legal and ethical concerns pertaining to the creation of covert psychological profiles on individuals to *predict* harm, this approach can also foster a surveillance environment that can target innocent individuals and reinforce a range of racial, social, class, gender and, age biases that can emerge from the creators of the programme (embedded in the software) and the end-users (person of authority implementing the software).

Thus, current approaches to insider threat are centred on controlling and protecting information (Yayla, 2011, Wall 2013) through utilising the technological element (i.e. software used to make deductions and predictions) to limit the operation of the human element. Ultimately, these understandings oversimplify complex sociotechnical systems that exist in real-world settings, fail to protect and consider the human element and, fall short of protecting against unintentional insider threat.

Organisational End-user Solutions:

Emerging from these software centric security approaches solutions include human network behavioural analysis (Nguyen et al., 2003), signature based activity within Intrusion Prevention Systems and, deception techniques such as honeypots (Mokube and Adams, 2007; Spitzner, 2003; Shabtai et al., 2016), port surfing, packet sniffing and decoys within active cyber defences (ACDs). While these software centric approaches are designed to mitigate insider threat, they are designed with the aim to apprehend the attacker or the malicious insider i.e. to identify the human element or to stop it. These solutions have recently become popular with organisations but are controversial on individual privacy, legal and ethical grounds (Goethals and Hunt, 2019; Tiwary, 2011).

In other popular solutions derived from traditional security thought, all responsibility for actions is placed on the human element. This shifting of onus to the human element for intentional and unintentional actions is showcased in solutions such as Cyber-Physical Systems, Industrial Control Systems (ICS) and, Industrial Internet of Things (I-IoT) (Ani et al., 2018). Within these approaches accountability and non-repudiation are enforced as secondary security principles to improve cybersecurity, where users are believed to be able to assume full responsibility for their actions (Gollmann, 2011; Cardenas et al., 2008; Larkin, 2014; Wang et al., 2010) whilst operating within what are assumed to be complex sociotechnical systems.

The latest cutting-edge solutions to tackle insider threat include the implementation of machine learning algorithms to an individual’s network behaviour for analysis (Bowen et al., 2009, Chattopadhyay et al., 2018; Punithavathani et al., 2015) including deep learning neural networks (Tuor et al., 2017). In some approaches linguistic and personality ques are combined with signature based activity through pattern identification (Schultz 2002). Hidden

Markov Models are also being utilised that assess deviations in individual user activity patterns against the ‘blueprint’ activity models that are in place (Thompson, 2004) or individual’s own historic activities (Rashid et al., 2016; Eldardiry, 2013; Mills et al., 2017). In some instances psychological modelling (Brdiczka 2012) is being implemented including those approaches that rely on personality traits such as OCEAN (Wiggins, 1996) and The Dark Triad (Paulhus and Williams, 2002; Maasberg et al., 2015) amongst other models to predict and counteract threats emerging from human elements (Greitzer and Frincke 2010; Greitzer et al., 2012; Liu et al., 2009).

These approaches and their subsequent solutions are proving to be insufficient to counteract the maturing risk of unintentional insider threat. This is evident in the frequent coverage of cybersecurity breaches on news channels for instance, supply-chain attacks and subsequent cyberbreaches that caused disruption during the Covid-19 pandemic (Plumb, 2022). Thus, there is a growing need to implement new models to tackle this challenge that involves the exploitation of the human element (Wall 2013; Colwill 2009).

2.3 Prominent attacks

The need to implement new solutions is increasingly evident and showcased through numerous high profile attacks recently on governmental bodies, multinational corporations, educational institutes and health organisations that have fallen prey to social engineering, phishing and ransomware attacks. The following work discusses phishing, social engineering and ransomware attacks in specific to demonstrate how the solutions derived from software defence approaches discussed above are proving to be insufficient for creating effective unintentional insider threat (UIT) defences.

2.3.1 Phishing attacks

In 1996, *phishing* (a cyberpunk rendering of *fishing*) was first used to describe an attack that resulted in the loss of AOL accounts and their respective passwords (Huang et al., 2009). This means that while phishing is historically seen as an attack that steals individual identities this definition has grown substantially since then. There are many types of phishing attacks for example, malware-based, session hijacking, deceptive phishing, key-logging, web trojans, host file poisoning and, man-in-the-middle (Suganya, 2016).

Phishing is now relatively well known by the wider public with many people likely to have a rudimentary understanding of what this term means. This is primarily because individuals are more exposed to these attacks in the realm of their daily lives. In many ways, phishing is an evolving and complex problem by its nature as it is easily automated (sending numerous phishing emails in a single batch), requires little to no human resources (from the attacker's perspective), various parts of the operations can be outsourced or purchased off -the-shelf (buying a malicious code) and, all associated activities can be carried out online (Chhikara et al., 2013). Coupling this nature of phishing attacks with a relatively fast turnaround for rewards succeeds in continually attracting a new stream of attackers. Alongside this, the ingenuity used to target individuals through such attacks has been astonishing.

An example of this evolution in attack techniques is the 'African Prince' phishing scam Okosun and Ilo, 2022). In the African Prince scam an unsolicited email would indicate that a person of notoriety or influence required assistance in transferring money out of their country. If the email recipient chose to help them then the recipient would receive a reward i.e. a stated percentage of the total amount being transferred. This tactic was used to open a dialogue with the recipient who would then be tricked into surrendering some or substantial amounts of their own money. This phishing scam was positioned to manipulate human emotions such as empathy and greed. In addition, time and stress pressures were used as a crucial step to manipulate the engineered situation through stimulating urgency or the perceived risk to health or life.

This African Prince phishing scam has now been replaced with highly sophisticated impersonations of world leading banking societies that urge recipients to undertake irrational actions in order to protect their accounts in a tight timeframe. Current phishing attacks make it extremely difficult for individuals to be able to distinguish between a phishing email and a legitimate email from their bank which may lead to individuals sharing sensitive information due to a temporary lapse in judgement. Phishing attacks also rely on using various confidence tricks and game theory to make individuals fall prey to divulging private and/or sensitive information that they normally wouldn’t have done. This information can then be misused, sold or shared for gains (financial or otherwise) by the attacker . Furthermore, such attacks can cost victims financial and identity loss and create the possibility of being susceptible to an attack again whilst the impersonated party (such as the bank) might suffer reputational and financial damages. These types of attacks also create a paradox whereby banks would still need to contact their clients via online channels with important information and so this channel cannot be easily blocked entirely (Ramzan, 2010).

There are numerous anti-phishing active and passive cyber defence (ACDs and PCDs) solutions such as awareness campaigns and software algorithms either at a server level, bowser level (black and white lists), web-page and, information flow level to counteract phishing threats (Huang et al., 2009). The range of algorithm based solutions include web browser based plug-ins that prevent users from entering sensitive information to ‘untrusted’ websites, software that can detect phishing emails (auto-spam script), software to detect anomalies between the document object (DOM i.e. what is shown on screen to the user) and the HTTP transaction (request command sent to the server and the response result sent to the user), software that uses honey tokens, data mining algorithms (some of which are based on mathematical models such as Bayesian probabilistic theory or frequency or analysis of text), antimalware (Jakobsson and Stamm, 2006) and, game theory based complex designs for systems and algorithms (Woo, 2019; Kim et al., 2017). Recently, the application of machine learning algorithms and artificial intelligence have become popular to overcome this threat as they aid users in their decision making prior to engaging with harmful content.

Solutions discussed earlier (such as port surfing, packet sniffing, active decoys, linguistic and personality ques, signature based activity and, personality tests) are well-suited to indicating intentional insider threat as malicious actions would reflect ill intention as there are opportunities to catch individuals red-handed. In contrast, unintentional insider threat (UIT) is void of any pre-existing intent to harm. In the context of insider threat that can facilitate phishing attacks, these techniques remain limited in their application as cyberbreaches linked to phishing are associated to UIT. In fact, the association of such compromises to UIT could potentially be because it is problematic, if not impossible, to ascertain with certainty that an insider intentionally or unintentionally engaged with a malicious link that surfaced through an external phishing attack. However, deception techniques discussed earlier are extended in their application to afford ‘phishing simulations’. Phishing simulations are tests carried out by organisations acting as a malicious *outside* party to simulate a real attack in order to test the strength of their defences i.e. to assess the number of employees that compromise the system during a simulation and the amount of time an attack would take to penetrate organisational systems. Based on principles of accountability and non-repudiation driven from traditional security thought discussed earlier, phishing simulations are grounded in the same principles i.e. if users engaged with malicious content they were fully to blame, were negligent or should have known better. This review shows how varied technological, human and, organisational remedies may be brought to bear on the same problem.

2.3.2 Social Engineering attacks

The phenomenon of social engineering became widely known by the general public after allegations against the rigging of U.S. presidential elections in 2016. Social engineering is described as the ‘hacking of humans’ (Hadnagy, 2010) whereby sensitive knowledge can be extracted from individuals through manipulation and persuasion. This knowledge is then used to attack even the most secure of systems through four primary channels: physical, social, technical and, sociotechnical. ‘Physical’ channels include gathering information through physical or *real-world* surroundings which can include watching someone physically type passwords/PINs, collecting credentials from physical spaces such as those found on memo notes or extracting useful information from an organisation’s garbage bin. Gaining enough knowledge about victims to convince them of the legitimacy of the operation is a part of the ‘social’ aspect of social engineering. ‘Technical’ aspects rely on gathering sensitive personal information about the victims through online activities such as those available on social media platforms. The use of social media platforms thus becomes a key component of a social engineering attack (Jagatic, 2007). ‘Sociotechnical’ channels for an attack utilise multiple or all of the channels mentioned above, where social engineering usually involves small groups of people being targeted all at once. This makes the attacks very sophisticated in their nature (Krombholz et al., 2015). Social engineering is tightly knit with phishing attacks whereby social engineering is often regarded as a part of phishing including attacks such as spear-phishing. However, in this writing social engineering is discussed as a separate and distinct topic to phishing as it is more widely understood than other phishing attacks (compared to Smishing for instance), perhaps due to its popularity in media coverage.

Research being conducted for social engineering applies specific tools and proposes specific solutions to this problem that are largely unique to and distinct from the solutions presented for phishing in general. As there is willingness to communicate and share information online, with individuals sharing personal data on social platforms, humans are considered the 'weakest link' in any given system by researchers. Countermeasures for social engineering include awareness training programmes, internet browser plug-ins, use of password pathway managers where alerts are provided when users are entering sensitive information to an unsecure or untrusted website, countermeasures for known attack vectors, amongst other solutions (Ivaturi, 2011). These types of solutions are part of passive cyber defence techniques (PCDs) discussed earlier as the strategy is complaisant in its nature until a threat is identified i.e. software is used to identify malicious content when it comes across it through the user's interaction rather than actively looking for malicious content on the entire internet.

2.3.3 Ransomware attacks

Using malware to encrypt files and hold them to ransom until a fee is paid by the victim is known as ransomware. Cybercriminals use a variety of techniques that include phishing and social engineering techniques to gain access to a device, such as a computer. Once access has been triggered, for instance through accidentally clicking a URL by the victim, the malware begins to encrypt data files (Kok et al., 2019). Depending on the algorithmic code of the malware, if the device is connected to a network it can begin to act as a worm and spread to other connected devices. Whilst ransomware is not a new concept, WanaCry (also known as WanaCrypt) discussed above was one of the most notorious ransomware attacks in 2017 which affected the NHS in the United Kingdom (Mohurle and Patil, 2017).

Popular solutions discussed earlier aid in reducing the impact of ransomware attacks. This includes regularly backing up files, installing updates for software that includes patches, setting up honeypots as part of active cyber defences, machine learning algorithms that include behavioural-based monitoring and off-the-shelf intrusion prevention software discussed earlier. Despite these solutions being in place by many international organisations (Travelex, UCSF, Grubman Shire Meiselas & Sacks and, Cognizant), 2020 witnessed an exponential increase in ransomware attacks (Novinson, 2020). Arguably, this could be due to the widespread remote-working afforded to employees during the global pandemic of Covid-19. However, while these solutions can act to reduce the impact of an attack, they cannot bypass it. Honeypots can certainly aid in weeding out or misdirecting potential threats, no software can completely prevent all malicious content from coming into contact with organisational systems. In the context of unintentional insider threat, not only would it be problematic to determine if an insider intended to compromise the system but also the technological element is leveraged once again to limit the operation of the human element in order to adequately protect information and systems.

It can be observed from the discussion above that solutions fall short in protecting systems from cyberbreaches and more specifically from unintentional insider threat. Solutions approach cybersecurity challenges in a 2D fashion that are software centric and propose ‘intelligent’ algorithms that shadow individual activities in order to intervene at the exact moment before disaster strikes – saving humans from themselves. Instead of approaching threats through automation and implementation of rules, there is potential to tackle insider threat through building sociotechnical solutions that can rely on strengthening the human element by shifting the way humans are considered within systems. Afterall, humans are an integral part of the cybersecurity chain that enable cyber operations and can make executive decisions making them worthy of being given the due consideration.

2.4 Relevant frameworks

Several frameworks exist that either directly or indirectly address insider threat. For instance, NIST Cyber Security Framework (2014) which has five pillars (Identify, Protect, Detect, Respond and, Recover) to provide organisations with a baseline of cybersecurity standards to assess and manage cybersecurity risks. However, NIST is aimed at best practices and grass-root effort at organisations to create cybersecurity momentum through awareness rather than explicitly focusing on insider threat which made NIST not very well-suited for this work to be included as a focus for this work. Additionally, MERIT model by CERT was selected as there are a range of frameworks that emerge from the work carried out by CERT that build on insider's ability/skill, opportunities afforded to them in systems and, for establishing their intent. Building on this work by CERT and directly associated frameworks which seek to protect the technological element, other prominent frameworks emerge that include psychological, behavioural and/or social elements. Thus, the MERIT model by CERT was included as a relevant framework as CERT is a prominent and world-leading research institute that enjoys the reputation of providing cutting-edge solutions and has subsequently served as a foundation for numerous insider threat frameworks. While SOFIT is one example of a framework that is rooted in MERIT model, it was included in this work as it claims to be derived from a human factors-oriented ontology (HUFO) which includes an equal focus on the social and technical aspects within a system. Error Management Programme (EMP) was selected as it provides a solution directly derived from the Generic Error-Modelling System (GEMS) from the sociotechnical theory perspective and this inclusion aids the reader in understanding the how these perspectives can also be enhanced when applied to insider threat.

While it is worth noting that ‘10 Steps to Cybersecurity’ by National Cyber Security Centre is not explicitly for insider threat, it was included as it has elements that pertain to this threat. As NCSC provides cutting-edge coverage on a range of cybersecurity related topics including those related to the human element and processes, this guidance appears in numerous documents on various topics which can be varying in its coverage and left to the interpretation of the reader on the type of insider threat being discussed. ‘10 Steps to Cybersecurity’ was selected as a relevant framework to inform this discussion as NCSC is the prominent organisation in the UK that covers insider threat, it is prominent guide and, is aimed at the UK audience where this research is conducted.

Thus, the following work takes an in-depth view of three frameworks and the NCSC guide that are designed to identify and prevent insider threat (i) MERIT model proposed by Computer Emergency Readiness Team (CERT), (ii) Sociotechnical and Organizational Factors for Insider Threat (SOFIT) by Greitzer et al., (iii) Error Management Programme (EMP) by Liginlal et al. and, (iv) 10 Steps to Cybersecurity.

This is done with an aim to aid the reader in understanding the motivations behind the development of these key frameworks, to build a case for the due consideration of human elements within the cybersecurity chain and, the extent to which proposed solutions can be applied to insider threat. To aid the reader in through this in-depth discussion of frameworks a comparison table is provided below.

Framework	**MERIT** **(CERT)**	**SOFIT**	**EMP**	**NCSC**

Derived from	Real-world cases	Academic literature	Generic Error-Modelling System	Real-world cases
Model type	Descriptive	Predictive	Error focused	Guidance
Stage	Early detection	Early detection	Pre and post incidents	Pre and post incidents
Method	Game play	Assessments	Investigation	Guidance
Aim	Seeks to establish malicious intent and motive	Seeks to establish malicious intent and motive	Seeks to understand errors	Seeks to build knowledge
Audience	IT, Financial sector, Critical National Infrastructure	Business-to-business	Business-to-business	Individuals, Businesses, Critical National Infrastructure, Aerospace, Financial sector
Elements used or considered for Insider Threat				
	CERT	**SOFIT**	**EMP**	**NCSC**

Works with limited knowledge about the attack	Yes	No	No	No
Individual behavioural indicators	Yes	Yes	No	Yes
Technical/ technological aspects	Yes	Yes	Yes	Yes
Root-causes for problematic behaviour	Yes	Unknown	Yes	No
Human Resources input	Yes	Yes	No	No
PCDs (anomaly detection, secure configuration, antimalware, network behaviour)	Yes	Yes	No	Yes
Organisational Factors	Yes	Yes	Yes	Yes
Communication	Yes	Yes	No	Yes
Risk Management	Yes	Yes	No *(error management)*	No *(incident management)*
Using 3rd party admin and monitoring tools	Yes	No	Yes (monitoring)	Yes (monitoring)

Elements used or considered for Insider Threat				
	CERT	**SOFIT**	**EMP**	**NCSC**

Policies, culture, procedures	Yes (societal culture)	Yes (organisational culture)	Yes (organisational culture)	Yes (organisational culture)
Training programmes and educational materials	Yes	Yes	Yes	Yes
Access points and log use	Yes	Yes	No	Yes
Workload considered	Yes	Yes	Yes (*fatigue*)	No
Staff Satisfaction	Yes	Unknown	No	Yes
Goals, stress, deadlines, expectations, morale	Yes	Yes	No	No
Design of technologies	No	Yes	Yes	No
Consideration before implementing new technologies	No	No	Yes	No

Table to reflect comparative aspects of relevant frameworks

2.4.1 CERT's MERIT model

Carnegie Mellon University's Software Engineering Institute developed the CERT Program to study insider incidents. These incidents included those that were reported to law enforcement agencies as well as those available in the public domain. In their three major publications in 2005, 2007 and 2008, their findings developed a framework called the MERIT insider threat model. Apart from being one of the most recognised frameworks in the field of cybersecurity pertaining to insider threat, a discussion on the topic of insider threat (IsT) would be incomplete if this work is left uncharted due to its influence on how IsT is understood and approached.

Management and Education of the Risk of Insider Threat (MERIT) provides findings from the work conducted as part of a collaborative project called ‘Insider Threat Study’ between several institutes, Carnegie Mellon University Software Engineering Institute’s CERT programme and, the United States Secret Service that started in 2001. This project was funded by CyLab at the Carnegie Mellon University with an aim to tackle insider threat through proposing early indicators for this threat i.e. before this threat matures or is realised.

A cumulative one hundred and fifty cases that occurred between 1996 and 2002 involving insider threat were evaluated in the initial study published in 2005 (Keeney et al., 2005). Their methodology included cases where there was an insider (current or former employee) who purposefully enhanced their access privileges or misused their access to a network, system or company data affecting the security of the organisation’s data, processes or operations. Cases where the perpetrator attempted to view, disclose, harvest, alter, download, delete, change or add information were also included. Any incidents that were outside the critical infrastructure sector and not conducted on US soil were excluded from this study. As a result, hypothetical scenarios only included known elements drawn from real-world incidents which encapsulate the challenges associated to working with limited knowledge since companies refrain from reporting insider incidents due to the fear of reputational and financial damages that result from such breaches. This approach of operating on limited knowledge did not limit the outcomes of this project but instead for the first time provided insights into actual behaviour of perpetrators and an analysis of the incidents themselves. Analysis included all available information about the online and offline behaviour of perpetrators through various documentation (HR files, system logs etc) and covered the time from where the idea was conceived to the time of the attack, through reverse engineering the timeline from the moment the attack was triggered. This information was used to answer several hundred pre-set questions by the researchers about the insider and the behavioural and technical aspects of each case. These questions broadly encompassed themes such as the various components of the incident, detection of the incident and the perpetrator, planning and communication prior to the incident by the perpetrator, nature of harm, law enforcement and organisation's response, characteristics of the insider and the organisation, background of the perpetrator and, the perpetrator's technical skills and interests.

This project brought together experts in the fields of behavioural analysis and network systems survivability and security. MERIT developed an Interactive Learning Environment (ILE), such as role playing games, whereby hypothetical scenarios were simulated. It explored insider threat attacks linked specifically to sabotage and cases were identified through the Secret Service computer fraud department, reports from various media outlets and, criminal justice records (Lexis-Nexis database). Various simulation workshops were conducted with an aim to impart valuable lessons for participants and provide tools that helped participants understand and assess risk levels for insider threat based on organisational policies, culture, technical and, procedural factors. MERIT's scope was specifically to evaluate, understand, access and, prevent the risk of malicious or intentional insider attacks through exclusively examining sabotage and espionage incidents in specific sectors i.e. IT, financial and banking and, critical infrastructure. This programme evaluated behaviour in the cyberworld as well as offline relations, offences and reprimands that included disciplinary actions, suspensions, demotions and salary reductions to evaluate behaviour and technical aspects of the attack.

Cases reflected that the insiders were predominantly former employees with technical positions at the victim organisations. Equal attention being paid to technical and psychological aspects of the attack are reportedly the key for this model's success. This led to MERIT model being widely adopted in industry settings and it served as a foundation for numerous popular approaches discussed earlier that involve a mixture of technological and psychological profiling.

MERIT applied system dynamics modelling to assess the risks and gain insights into difficult management situations as 'intuitive solutions' were believed to be ineffective in the long-term creating a magnitude of problems as a by-product. This risk modelling was also deemed suitable as it is able to provide effective solutions and can demonstrate the solutions' benefits over a longer timeline. MERIT model captured the complexity of problematic behaviour, its underlying root causes and, included soft and hard factors so as to not render any factor(s) in the attack as negligible. This model was not predictive but rather descriptive to illustrate various trigger points that led to an attack. Simulations with participants started at the highest point of the perpetrator's career within the victim organisation where the insider enjoyed the most liberties (post the point of hiring) and ended at the point just after the attack was conducted (usually post the offender's termination or resignation).

Findings from this study (Keeney et al., 2005) reported that organisations had the opportunity to detect harm prior to an attack and, victim organisations (82%) belonged to the private sector and had similar technical controls, policies, processes and procedures in place. Findings revealed that there was no standardised profile of a malicious insider as the demographic of perpetrators varied in age (mean age of 32 years), ethnic and racial backgrounds and in their marital status. However, a vast majority of the perpetrators were male (96%), with a third of the population with a prior arrest history. An overwhelming amount of the included case studies shared a scenario where the perpetrator felt they had been treated unjustly for their hard work, had unmet or diminished expectations about their career at the victim organisation, were reprimanded for liberties they had enjoyed in the past, had experienced a change in management or reporting structure and, had reports from colleagues noticing a deterioration in perpetrator's behaviour. Attacks relied on social engineering and physical sabotage.

In light of these findings, the recommendations put forward by the research group included awareness training of employees and physical security systems to be put in place that were monitored and maintained. Awareness training included the recommendation to safeguard privacy of passwords and not disclosing personal passwords to colleagues. Password awareness included password policies to be implemented by organisations so as to limit unwarranted access by anyone other than the intended party. Recommendations also included regular audits of system logs to ensure backdoor accounts have not been created, restricting the existence of 'unknown' accounts on the system and, the organisation being knowledgeable about unauthorised privilege escalations associated to accounts within a network. The use of anomaly detection tools was suggested and importance was placed on organisations proactively dealing with insider threat through vigorous systems security, regular monitoring of those systems and resolving employee grievances in a way that doesn't provoke aggression but simultaneously addresses any concerning behaviour by employees.

In 2007, the CERT Program published another report by Cappelli et al., that built on the findings outlined in their 2005 report and described the MERIT modelling and simulation results. This report made a direct correlation between the decisions made by management regarding performance and an increased level of insider threat posed by disgruntled employees. Attacks were possible primarily due to the lack of tools available to understand and mitigate insider threat, lack of risk mitigation techniques and, an overall lack of good communication channels within an organisation.

MERIT model's proposition grew from equal attention being paid to the technical as well as psychological aspects. This incorporation of psychological elements was the first time that insider threat was not viewed with a singular lens of software solutions to overcome vulnerabilities and strengthen barriers but rather a broad approach was being adopted to understand the various components of the insider threat problem and its interdependences. Several technical and administrative controls were recommended to mitigate insider threat which included aspects such as technical monitoring of employees (access paths, resources and information accesses, online actions), tracking of employees, auditing and disabling rogue access paths, balancing termination threshold and employee intervention (Figure 4, pg. 15; Cappelli et al., 2008).

MERIT system dynamics modelling was used to simulate different company policies, their impact on the outcome and how that would affect the level of insider threat (IsT) risk for the organisation. Other factors such as culture, technical skills and procedural factors were also considered. In contrast to the previous report in 2005, this report offered succinct details about the conditions and factors that can increase the risk of IsT within organisations. Specific behavioural precursors that were exhibited by perpetrators in this study's data included high expectations from the perpetrator for technical freedom, perpetrator considered themselves as being above the rules and policies set out by the organisation and, perpetrator expected to have, or actually had, complete control of the organisation's network. To note amongst the findings is that the above behavioural precursors were claimed to usually be exhibited four weeks prior to any technical precursors being visible on system logs. Thus, findings stated that there is a high risk of IsT at an organisation if the following elements are present in a real-world setting:

1. There is a disgruntled employee following a negative work related event (potential perpetrator)
2. The potential perpetrator shows concerning social behaviour (a precursor to an imminent attack)
3. The potential perpetrator has held or is holding a technical position (skills available to conduct an attack)
4. The perpetrator is likely to or has been terminated from his designation (59% of attacks happened post perpetrator's termination)

Subsequent recommendations from these findings included building stronger defences against insider threat through regular audits of system logs pertaining to access, monitoring of any breaches to privileges, measuring employee satisfaction, evaluating concerning behaviour exhibited by employees, increasing the monitoring of employees who exhibit concerning behaviour, taking positive actions to help employees who exhibit disgruntled behaviour through HR interventions and, employee support groups. Findings stated that timely detection of possible insiders that can cause harm is critical and consistently strengthening defences against possible insider threat through technical and administrative controls is essential.

In 2008 CERT Program published the third major report as a white paper to describe the MERIT insider threat system dynamics modelling and corresponding simulation results (Cappelli et al., 2008). This modelling provided tools for understanding, assessing and analysing risk mitigation decisions that arose from insider threat within organisations.

Through the use of interactive learning environments (ILEs) based simulation workshops demonstrated how day-to-day decisions influence other components that interact with insider threat, such as technical skills, management decisions, expectations, and the subsequent paths an attack can take. ILEs also appeared to overcome Sterman's (2006) three challenges associated to learning lessons from experience in complex systems that involve humans and technology i.e. presence of good data, ability to draw conclusive lessons from complicated interdependent information and, involvement of stakeholders in the development of company policies.

Early detection was deemed key in being able to mitigate any potential insider attacks in the simulation workshops. The research team recognised that management, IT department, human resources, security as well as other parts of the business needed to be able to work together through good communication, have a firm understanding of the psychological, technical and organisational aspects that foster the emergence of insider threat and, be able to formulate responsive actions plans as countermeasures. In order to be able to achieve the above, new communication tools and training materials needed to be developed that could be utilised by various departments within an organisation. These materials were developed through a continued application of system dynamics modelling. This modelling was claimed to be effective in communicating and measuring the risk of insider threat (specifically sabotage) and its mitigation to various stakeholders. Fundamental components to simulate the application of this model were, (i) revenge or disgruntlement as motivations behind insider attacks, (ii) concerning behaviour being exhibited by perpetrators prior to attacks, (iii) perpetrators held technical positions and, (iv) a majority of attacks occurred post termination of the insider.

Key recommendations from this report in 2008 highlighted the importance of completely disabling all known access points of the insider in a timely fashion and doing regular access audits on system logs. It also recommended that all ILEs must impart the knowledge and importance of raising awareness towards proactive, continuous and thorough access management practices for IT departments within all organisations. Recommendations stated that given the workload experienced by employees, those employees who have demonstrated concerning social behaviour following a negative work related event should be carefully evaluated by management and possibly be monitored for their online interactions. It was recommended that employers must be aware of their employees satisfaction ratings and promptly evaluate concerning behaviour. Whilst the technical and admin tools helped stakeholders to work together to counteract insider threat (IsT), it was recommended that employers should take positive action to address disgruntlement such as formulating support groups and offering counselling to address the situation instead of taking punitive actions or reprimanding individuals.

Some assumptions that were made as part of this modelling that are important to note are that malicious insiders were believed to work independently in their actions to conduct an attack, perpetrators had a strong sense of entitlement, the attack was usually driven by vengeance and disgruntlement which are directly correlated with a sense of entitlement, the attack timeline started from the highest point in perpetrators' career with the organisation, poor security management practices were in place, insiders had access through granted, created and/or discovered paths that the organisation might or might not have been aware of and, poor defences were in place against unacceptable employee behaviour with an absence or lack of technical and administrative controls.

However, a study by Bell et al. (2019) discovered that individuals are reluctant to report behaviour that might be deemed 'inappropriate' in the context of predicting or preventing insider threat. This reluctance can emerge from a lack of evidence, self-ability to assess change prior to reporting behavioural indicators, seniority of the insider, confidentiality of the process and, the lack of clear reporting channels. Findings suggest multiple factors are at play in an organisation when proposing solutions to counteract insider threat. Such factors include management tensions, politics, confidentiality and rapport between the employees and the employer, policies and governance, organisational culture, training, awareness and, communicated transparency. Since these factors form a complex sociotechnical system that is the organisation, it becomes problematic to propose solutions for systems in insolation to all its other parts or to fragment the system into its parts (Hollnagel et al., 2015). Here an argument can be made that a stance adopted from a traditional security approach can propose solutions that are 'quick wins' at a first glance but through a detailed evaluation it can be argued that proposed recommendations can only enjoy limited success. This limited success for outcomes is primarily due to the lens being adopted to understand the system and the creation of undesirable outcomes within a system by technologies and humans that operate within it.

MERIT model discussed above made incremental, albeit minor, changes to their frameworks over the years. Changes include the elimination of using arrest records of employees, importance of a culture shift in organisations and, the focus on regular audits of access points to the organisational network. This framework is centred on behavioural analysis that is retrofitted on to known cases. It simplifies relationships between various departments and their ability to efficiently communicate with each other to develop an action plan that can be used as a preventative measure against insider threat. Arguably, real-life scenarios can potentially be riddled with navigating challenges such as different priorities, availability, training and, technical levels of understanding possessed by vital players such as HR, IT, CEOs and Board Members which will cost crucial time to prevent or mitigate insider threat as it unfolds.

2.4.2 SOFIT

‘*Sociotechnical and Organizational Factors for Insider Threat*’ or SOFIT is a framework developed by Greitzer et al. (2018). SOFIT combines technical aspects, individual behavioural indicators and organisational factors to identify insider threat (IsT).

Adopting MERIT’s recommendations SOFIT incorporates technical aspects which include a range of active cyber defences (ACDs) for initial mapping and then monitors the host network behaviour. Once this is done, anomalies in the network are identified and given a rating of how secure the network is in the form of a ‘yes’ or ‘no’ checklist. For instance, if a company hasn’t updated their software but have all the other nine technical measures in place then SOFIT will give this category a rating of 90% secure. These checklists are in the form of parent-child factors and overall ratings of parent factors within the technical category show stakeholders how robust the systems are against IsT and highlight areas that require further attention.

Similarly, 271 different ‘individual’ behavioural factors are utilised to establish intent and motivation to identify IsT through assigning a rating to each trait. These indicators are adopted from human factors-oriented ontology (HUFO) for cybersecurity risk and other psychology constructs such as the Dark Triad, dynamic states and personality dimensions. Similar to CERT’s MERIT model discussed earlier, SOFIT also relies on the reporting of ‘observed’ behaviours that are exhibited by individuals usually by HR personnel. Such behaviours include:

> *‘The behavioral indicators associated with the highest risk were disregard for authority, disgruntlement, anger management issues, and confrontational behavior; the occurrence of any one of these indicators would yield heightened concern about the insider threat risk of an individual’*
>
> – Greitzer et al., 2018

Alongside the technical and individual categories above, SOFIT also incorporates 49 organisational factors and provides a rating to identify IsT. Within this model organisational factors are believed to affect performance and increase errors. Factors include a range of indicators such as poor communication, inadequate training, ambiguous goals, stress, workload, blame culture, poor team management, poor system designs, environmental stressors, unrealistic deadlines, mismatch between expectations and abilities and, morale. Within this framework organisational factors are believed to be primary contributors to increased risk of insider threat as it can propagate human errors and lapses from individuals that cybercriminals can take advantage of through attacks such as social engineering and phishing. For instance, a staff survey might indicate that individuals are experiencing a high workload and thus, SOFIT will give them a rating of 10%.

Once there is a score for each of the factors within technical, individual and organisational categories SOFIT provides an output as a pie chart. The algorithm then considers a combination of factors, taking a weighted value if there are multiple factors in the same category, to provide another output as a pie chart and an overall value to indicate organisational risk levels for insider threat (IsT). With this framework certain combinations of factors might provide a higher risk level of IsT, for instance ‘disregard for authority’ and ‘poor communications’ versus ‘minor policy violation’ and ‘distractions’.

While this approach appears to be promising, SOFIT does not disclose a complete list of indicators for any of its three categories that researchers can investigate. Authors also acknowledge that indicators within each of the categories are continuously being revised with additions and exclusions which makes the reliability of such a framework problematic. Despite SOFIT admittedly being more focused on the ‘individual insider’, in contrast to technical aspects for describing the event that are imperative to MERIT (and the Insider Threat Indicator Ontology ‘ITIO’), SOFIT appears to rely heavily on individual psychological profiling. This can mean that a lot of the factors used to identify IsT might simply not be known to the organisation or the collection of various indicators might not be legally permissible in certain countries of operation. The undisclosed techniques used to gather personal data on individuals might also prove problematic. For instance, SOFIT only allows HR personnel to upload individual behavioural data which can make the data susceptible to manipulation due to real-world politics that exist in the workplace. It can also result in an ironic paradox of expectations and abilities with HR personnel not being able to make those deductions (lack of professional psychological qualifications) and the expectation from the organisation and SOFIT to do so.

2.4.3 Error Management Programme

Liginlal et al. (2009) created a sociotechnical framework known as the Error Management Programme to tackle insider threat (IsT). Through extending the application of Generic Error Modelling System (GEMS) to examine errors arising from slips, lapses, mistakes and, violations (discussed in greater detail later on in this Chapter) Error Management Programme examines root causes that lead to errors. It proposes creation of defence strategies that avoid, intercept and correct errors and recommends evaluating processes periodically for effectiveness. Liginlal et al. (2009) framework also recommends training programmes, effective design of technologies which includes displays, monitoring and alarms, timely investigation of errors, a no-blame organisational culture, careful organisational consideration being paid prior to the implementation of new systems, having effective processes in place and, monitoring work related fatigue. This approach argues that effective policies must put in place by organisations and enforced in the daily delivery of work.

While this approach adopts GEMS, it places the onus of accidents on organisations. Organisations in this approach are responsible for a range of aspects in order to avoid errors. For instance, it is the responsibility of the organisation to create and implement the use of policies that prescribe the delivery of tasks. Organisations would need to invest resources such as time and money in training and the design of software solutions being used by individuals to deliver tasks i.e. 'effective' design of technologies mentioned above. Whilst this programme recommends training people to address the lack of expertise amongst people who deliver tasks, it does not take into consideration utilising expert individuals that exist within organisations.

This writing now moves on to discuss various pieces of work by the National Cyber Security Centre (NCSC). While not all of NCSC's work is directly relevant to this project, it is still important to highlight the approach and efforts being adopted in the United Kingdom where this research project is conducted.

2.4.4 NCSC's "10 Steps to Cybersecurity"

The National Cyber Security Centre (NCSC) was established in the United Kingdom in 2016. It aims to provide a single point of contact for businesses and governmental agencies that operate in the UK for all matters pertaining to cybersecurity. NCSC is also responsible for providing a range of information and guidelines to the general public to raise awareness and utilises expertise from a range of backgrounds that includes industry and academia.

NCSC, as part of GCHQ, supports the most critical organisations, the wider public sector, SMEs and the general public to guard cyberspace operations in the UK to drive it towards a digital economy as part of Industry 4.0. NCSC's operations are not focused on a singular strand but encompass all micro and macro incidents and remits within cybersecurity. For instance, NCSC will provide individuals with a guide on how to make strong passwords for their social media accounts (micro effort at grassroot level) as well as monitor, strategize and, respond to a national cyber incidents that can involve foreign state-backed hackers as part of organised cybercrime groups that might target critical infrastructure, aviation domain or financial services. NCSC also works towards improving the cyber resilience of UK's infrastructure, managing and mitigating risks as well as providing funding for new innovative technologies for cyberspace.

In the guide by NCSC '*10 Steps to Cybersecurity*' (2019), first published in 2012, businesses were advised to incorporate ten suggestions in order for organisations to be better protected in cyberspace. The first step offered fundamental understanding of 'Network Security' and recommended setting correct perimeters for networks to operate within. This included monitoring access, removing unauthorised users and malicious content and, testing security controls within the organisational network. 'User education and awareness' was the second step that entailed the creation and distribution of security policies to all employees. It included regularly making employees aware of the various risks in their cyber interactions and communicating acceptable and secure use of company systems. 'Malware prevention' recommended having various relevant policies and anti-malware software in place for all company assets. 'Removable media controls' as the fourth step advised companies to control and limit access to removable media technologies such as USBs to organisational devices. If such were permissible, it was advised that an anti-malware software scanned the content of the device prior to importing files onto company systems. The fifth recommendation was ‘Secure configuration’ which endorsed performing regular software updates that included security patches and to properly configure organisational systems. A system inventory was recommended to track and implement the minimum baseline build for all company devices that might use different operating software. The sixth step of ‘Managing user privileges’ recommended limiting the number of privileged accounts, limiting user privileges and monitoring user activity. This included maintaining audit and activity logs. The seventh step, ‘Incident management’, outlined the need for businesses to have a response plan in the event of a successful cyberattack, organisations were advised to conduct periodic drills, provide specialist training to staff and, recommended involving local law enforcement if a cyberbreach occurred. The eighth step titled ‘Monitoring’ advised organisation to establish a strategy to monitor employee activities, create supporting policies and analyse datasets for unusual or suspicious activity that could be a precursor to a cyberattack. As the final step for cybersecurity within this guide, NCSC covered ‘Home and mobile working’ which suggested developing a company policy, training staff to understand the policy and monitoring staff compliance to this policy. At the heart of this guide organisations are to be aware and in charge of their organisational cybersecurity, prioritise it in the same way as financial or operational risks and, establish a regular risk management regime. It also highlighted the importance of organisational cybersecurity initiatives to be supported by board members, senior managers and overall throughout the organisation.

When querying ‘Insider threat’ on the NCSC website on 30 July 2020, 155 items were returned in search results. However it was apparent that instead of tackling insider threat as its own subset heading, insider threat elements were captured across various other headings.

This means that the insider threat (IsT) topic was scattered across numerous guides which were extremely varied such as ‘Cloud security guidance’ and ‘Macro Security for Microsoft Office’. One of the most relevant search result to directly tackle insider threat (IsT) was ‘User education and awareness’ contained within the guide ‘10 Steps to Cybersecurity’ (2019) discussed above. This section of the guide mentioned how IsT could arise due to dissatisfied employees or an individual’s changing personal circumstances which largely implied intentional IsT but also had undertones of unintentional IsT. Unintentional and intentional IsT was thus indirectly addressed in the guide’s scenarios and suggestions. Suggestions included:

1. Creating a user security policy
2. Conducting staff inductions which highlights that users are personally responsible for complying with the security policy and would face disciplinary action for any deviations
3. Regularly making employees aware of the security risks faced by the organisation including refresher trainings
4. Encouraging staff to attain formal qualifications to build security skills within the organisation
5. To test and evaluate user training
6. Promoting an incident reporting culture within the organisation which includes empowering staff to share poor practices and report incidents to senior managers without fear of being blamed
7. Establishing a formal disciplinary process for any offenders who do not comply with the security policy including actionable penalties that are enforceable

These suggestions were not limited to this guide but are largely prevalent, albeit in different wordings, in all the 155 results that are returned in the search querying the NCSC database for insider threat (IsT). While NCSC has taken major steps in being user centric and separated itself from the US' requirement of cybersecurity, which hinge on monitoring individual's activities, NCSC appears to contradict its stance within the search results to tackle IsT. For instance, in the suggestions listed above NCSC advices that security policies are created with consideration to different user's roles and processes and should empower individuals to share their concerns about poor practices and report incidents (including near misses). Simultaneously, the guide suggests individuals should be held personally accountable for any deviations in their actions from the security policy, action should be taken against offenders and that said action is enforceable and attainable. Such clear onus placed on individuals and swift action in the context of incidents can be seen as a reprimand when reporting an incident, foster a blame culture and viewed as 'example setting' by peers who consequently might not raise concerns or share incidents when there is a breach of the same security policy.

In real-world settings shifting the onus to end users or levying fines on organisations can potentially be a major deterrent for reporting cyber incidents, especially 'near misses' that in contrast are seen as invaluable learning experiences in nuclear and aviation industries (Bair et al., 2017). While reprimands can be a 'quick fix' to ensure individuals act reasonably and responsibly, it can provide organisations with a false sense of security. For instance, during the Covid-19 pandemic UK government rolled out a 'test and trace' or 'contact-tracking' mobile app that would identify infected people and trace others who might have been exposed unwittingly during a certain time frame. It was argued by the UK government that this could help identify 'super spreaders' of the disease. There were two options – a centralised application (app) or a decentralised app. The centralised app meant that all mobile data from individual mobile devices would be held in a national database. A decentralised app meant that Apple (iOS devices) or Google (android devices) would create an app for each city and hold the data. Both options require Bluetooth data to identify other local devices. However a decentralised app was believed to be more cybersecure because personal data was to be encrypted or ‘hashed’, which meant that doctors or nurses operating in the National Health Service (NHS) would not be able to access or view individuals’ data. In an article by McCarthy (2020) he writes:

“The other concern with the UK approach is that while it insists it will keep data private, and location data will not be stored nor attached to individuals, the truth is that it will only work as promised if that data is not kept private and location data is stored and attached to individuals… Levy [Technical Director of NCSC] repeatedly tried to square this circle, leading to some ludicrous assertions. He stated boldly in bullet points that the app "doesn't have any personal information about you, it doesn't collect your location and the design works hard to ensure that you can't work out who has become symptomatic" and that "it holds only anonymous data and communicates out to other NHS systems through privacy preserving gateways”… So long as you can rely on one piece of per-user data – like a "big random number" – everything else can be connected. And if you also have a postcode, that becomes 100 times easier. Ever heard of Facebook? It's worth billions solely because it is able to connect the dots between datasets.”

– McCarthy, 2020

While the storage and use of personal data was a particularly rampant debate in the midst of the global covid-19 pandemic, cybersecurity of the app directly included protected access to third parties, such as the NHS. Personal data being encrypted meant the decentralised app was believed to be more secure by members of the parliament and the wider public. In the context of insider threat (IsT), data held by a decentralised app with any access points to the human element in the cybersecurity chain would make the data just as vulnerable with provider’s employees (Apple and Google) as it would be for civil servants or NHS staff in case it was intentionally or unintentionally compromised.

Through the discussion of relevant frameworks that are utilised as a blueprints for existing solutions, it can be argued that real-world settings are complex environments that have a range of concurrent factors that influence decision making and how work is subsequently performed. Given the complexities of conditions that exist as part of everyday life, it becomes problematic to label people into binary categories of either *good* or *bad* or to oversimplify complex systems by taking for granted that people know the entirety of the system to make informed decisions. Equally, limiting the operation of the human element can create a gulf between how work is imagined and delivered as well as restrict innovation within the devising of processes. Thus, this work now progresses to discuss relevant approaches from the risk and safety and, human factors domain.

2.5 Alternative perspectives to undesirable outcomes

Continuing with an human centric stance adopted above to evaluate solutions proposed broadly for cybersecurity, and in some cases more specifically for insider threat, this work will now briefly introduce sociotechnical theory. Sociotechnical perspectives relevant to this work are subsequently discussed and are as follows: Epidemiological Triangle (Cassel, 1976), Swiss Cheese Metaphor (Reason, 1990a), Safety II approach (Hollnagel, 2018), Skills, Rules and Knowledge (SRK, Rasmussen, 1983) and, Generic Error-Modelling System (Reason, 1990b). These sociotechnical perspectives are discussed with the intention to provide the reader with alternative methods for considering the human element in systems through understanding the occurrence of errors, such as those that result in unintentional insider threat (UIT). Discussing taxonomies to understand errors also increases the scope of understanding to include errors that lead to near-misses (i.e. creation of undesirable outcomes such as cyber incidents) but not necessarily a cyberbreach.

The term *sociotechnical* comprises of two aspects: *socio* that pertains to humans and society and *technical* that concerns technology and machines. Sociotechnical as a term refers to the interconnectedness of the social and technical elements within a system. Sociotechnical theory rests on two primary principles (Walker et al., 2008):

1. A sociotechnical system contains dynamic relationships within and between the *socio* and *technical* elements and both elements exhibit unique behaviour to one another. System performance (success or failure) is dependent on the interactions between the social and technical elements. These interactions comprise of a mixture of relationships shared between the two elements i.e. partially linear (cause and effect) and partially non-linear, relationships are typically planned or *designed*, relationships are complex, unpredictable and, frequently unexpected. These relationships between the elements and larger systems are interdependent and sensitive to change which can aid (or hinder) the achievement of (organisational development) goals (Cooper and Foster, 1971; Appelbaum, 1997). In addition, the two elements *behave* differently i.e. socio does not behave as the technical since humans are not machines. However, increasingly the technical element has also started to display non-linear behaviour due to the complexity and interdependency of technologies.
2. The focusing on one element i.e. *either* socio or technical (as discussed earlier where the technological element is leveraged to protect against the vulnerabilities posed by the human element) can result in increased *unstable* relationships between the two elements (i.e. unpredictable, unplanned, non-linear relationships) that can harm system performance.

Therefore, sociotechnical theory emphasises mutual optimisation of both elements i.e. the socio *and* the technical. Walker et al. (2008) describe a *sociotechnical system* as the purposeful collaboration between the socio and technical elements to achieve a goal. Sociotechnical theory is adopted from general systems theory where the term *open systems* is used to describe, analyse and, design systems based on mutual optimisation of both elements and feature a level of non-linearity between the elements and the environment within which they co-exist. Subsequently, a specific set of methods and perspectives can be utilised to create open systems in organisations to make them responsive to challenges posed by complex environments (Carayon, 2006), make them dynamic and, be able to tolerate and leverage the introduction of new technologies (Walker et al., 2008). The following sections present relevant sociotechnical perspectives for understanding errors that can unintentionally result in cyberbreach or incidents.

2.5.1 Epidemiological Triangle and Swiss Cheese Metaphor

One possible way of viewing unintentional insider threat (UIT) is through an *Epidemiological Triangle* (Cassel, 1976) which is also known as the *Epidemiologic Triangle*. This triangle is commonly used as a visualisation technique to understand and demonstrate the interdependent relationship between three vectors. It is most often used in public health communications and safety science (e.g., Gordon 1949; Haddon 1968; Mpolya et al., 2009; Gulis and Fujino, 2015; Lagerstrom et al., 2016).

Amongst the triad, the first vector represents the ‘Agent’ which portrays the ‘how’ or the infectious disease like malaria, responsible for causing the disease. The second vector of ‘Host’ represents the ‘who’ or the victim who suffers the punitive damages received by the Agent. The third vector represents the ‘where’ aspect in the tripod which is presented as the ‘Environment’ within which the Agent and the Host coexist. In the example of a viral disease, the *Agent* would be malaria, *Host* would be the humans and the *Environment* might be stagnant bodies of water or a tropical climate. Within this perspective, all three vectors can be worked upon to reduce the chance of an incident occurring i.e. to contain the spread of malaria. Continuing with the example of malaria, preventative measures such as medicinal vaccines can be provided to strengthen the Host and a range of anti-disease steps can be undertaken to weaken the agent and to modify the environment. This model provides a notable insight i.e. causational factors should not be oversimplified to a singular cause but rather emerge from the interaction between various vectors which can then be strengthened to reduce negative impact. However, this approach can be limiting when trying to determine which of the three vectors has the highest contribution to a more adverse outcome (Burke, 2019) and, the Epidemiological Triangle model itself can portray an oversimplification of real-world conditions that are removed from reality (Wu and Zha, 2020).

This approach was adopted as it serves as an informative backdrop against which unintentional insider threat can be viewed. Through extending its application from public health communications and safety science to unintentional insider threat, this model can aid in understanding the dynamic and interdependent relationship that exists between the three entities that coexist in cyberspace when incidents or breaches occur i.e. the *human* element who is the operator, the type of *cyberattack* and, the *environment* within which the human and the attack coexist. This approach also demonstrates that adopting a binary approach to examine causes can potentially be limiting for understanding challenges that arise from the interaction of multiple co-dependent factors.

Similar to the Epidemiological Triangle, the Swiss Cheese Metaphor (Reason, 1990a) also provides a visualisation of the relationship between defences and the occurrence of accidents in complex systems. Swiss Cheese Metaphor approach has been popular since its first emergence and has been applied in various domains to assess and understand the generation of errors in sociotechnical systems. Each defence is represented by a slice of Swiss cheese (famous for its holes). Holes within each slice represent contributors that have inherent weaknesses and vulnerabilities that can result in, or contribute towards, the failure of that defence. Numerous amount of defences (represented in a linear way as multiple cheese slices) can be implemented by an organisation to protect itself against adverse events. With the implementation of numerous defences, even if an accident occurs within one element of a system it can be stopped from penetrating all other aspects as a subsequent defence (i.e. the following cheese slice) might block it. Inversely, there can be times when all the holes in the cheese slices align to realise an accident in defences that otherwise are believed to be robust i.e. the vulnerabilities intrinsically present in various contributors acted in a way whereby each defence was unable to limit or avoid the event from occurring. Thus, accidents that occur in complex environments can be understood as the accumulation of multiple factors and failures that worked in combination with each other. Despite this model's widespread application, specifically in safety critical domain, its limitations include an absence of how causal factors interact with each other, defences are represented as being stagnant over time (i.e. vulnerabilities represented as *holes* might change or interact with other defences' aspects), defences are viewed as being independent of each other and, it offers little instructions about its application to real-world settings (Reason et al, 2006).

This visualisation, in addition to the Epidemiological Triangle, was adopted as it offers an insight for evaluating defences i.e. despite strong defences, intrinsic vulnerabilities in complex sociotechnical systems can create systematic conditions that realise accidents or cyberattacks.

2.5.2 Safety II

The Safety II (Hollnagel, 2018) approach begins by problematising the retrospective and eliminative nature of safety science towards errors. Hollnagel (2018) states that with a ‘whack-a-mole’ attitude, established safety science techniques concern themselves with failures and their correction. However, it is argued that in *modern* systems, such as connected technologies that form complex systems discussed above, it might be more appropriate and effective to focus on emulating success.

By adopting the above stance, Safety II (Hollnagel, 2018) provides an alternative approach to understanding safety. This is done through classifying all existing safety science approaches as Safety I. This approach argues that it is through understanding Safety I that provides the contrast by which Safety II can be understood.

Safety I is the established or traditional approach to safety such as the traditional security approach adopted by the frameworks discussed earlier i.e. MERIT, SOFIT and NCSC guidance. With this view, the absence of accidents or incidents is considered ‘safe’ and, ‘safety’ is defined as a state whereby as few things as possible go awry. However, when something goes wrong, failures or malfunctions can be identified through examining three components in a system: 1. Technology 2. Procedures and, 3. Human workers. The third element, i.e. humans, are the most variable of these three components and thus viewed as a liability and creators of arising incidents or accidents. This stance is common to traditional security thought that views ‘humans as the weakest link’ in the security chain. With a Safety I approach, either a system works as desired or fails. If work is delivered in line with *work-as-imagined* (Suchman, 1987) by the developers of the system, everything will function as it is supposed to, resulting in acceptable outcomes with no adverse events. However, if malfunctions occur within the three components, such as non-compliance to procedures, insufficient procedures and system descriptions or errors in technologies, it can result in failures or unacceptable outcomes. The experienced failure or unacceptable risk prompts an accident investigation to determine the root cause with an aim to either eliminate the cause and/or implement preventative measures so as to eradicate the error in the future. Thus, a Safety I approach examines things that go wrong and reasserts ‘work as imagined’ through the avoidance of deviation to the work being performed. Accident investigations set out to identify root causes for the adverse outcome and involve risk assessments to determine the likelihood of deviations occurring in the future as ways to strengthen barriers or defences against undesirable outcomes.

However, Safety II approach argues that modern systems are not stable and increasingly interconnected. While Safety I approach seeks to control and correct human variability that result in errors (for instance the blame and punitive measures placed on humans after accident investigations), it is the same human variability in modern systems that allows adaptability necessary for systems to function in a desirable way. Furthermore, as Safety I exclusively investigates things that go wrong, it neglects the examination of things that go right (i.e. actions that have yielded desirable outcomes many times before). Therefore, a Safety I approach limits learning opportunities and the ability to replicate success or the creation of desirable outcomes that happen a vast majority of the time.

In contrast to Safety I, Safety II acknowledges that there is performance variation by humans when they deliver tasks. Through the acceptance of performance variation in the human element, a Safety II approach subsequently establishes that there is a constant variability in system performance that results from the variance in human performance. Thus, this approach argues that it is problematic to characterise components in a binary fashion as either working as desired or failing. As the performance of a system is constantly varying, it is instead classified as ‘every-day-work’. Safety II approach believes that this performance variation is the factor that allows adaptability required by a system to respond to any changes in its environment. Consequently, humans are viewed as the element necessary for a system’s flexibility and resilience. Thus, systems working correctly is not due to humans conducting ‘work as imagined’ by the creators but rather due to humans adjusting to their environment. Safety II believes that this human adaptability and flexibility becomes a cornerstone to understanding how tasks are conducted safely within complex systems. Desirable or undesirable outcomes have a common basis i.e. day-to-day performance adjustments carried out by the human element. Therefore, accidents are not perceived as unique individual events but rather an expression of everyday human performance variability.

With a Safety II lens something that goes wrong has in actuality produced desirable results numerous times in the past and will continue to produce desired results again many times in the future. Thus, Safety II approach suggests that learnings can be obtained from examining aspects that allow the system to perform as desired and not only when it fails or produces undesirable outcomes. In order to harness these learnings and produce desirable results in varying conditions, there are five principles:

1. Examining things that go well
2. Focusing on events that are occurring frequently (such as near misses) rather than the perceived severity
3. Being sensitive to the possibility of failure
4. Thoroughness is preferred to capture learning lessons than efficiency and,
5. Investing in safety also increases productivity as the focus is on learning from and replication of making systems perform to produce desirable outcomes.

As highlighted in the earlier part of this Chapter, current approaches adopt a binary stance when determining the state of a system (i.e. as *safe* that has withstood cyberattacks or *unsafe* that has resulted in a successful cyber-attack), efforts are made to eliminate errors through investigations of causal factors for cyber incidents and breaches and, technological element is utilised to predict or limit the operation of the human element as humans are believed to be the most variable component. As current approaches have demonstrated limited success in addressing unintentional insider threat and due to the perspective offered by Safety II approach discussed above, a Safety II perspective was adopted as a guiding school of thought to underpin the work conducted and presented in this thesis.

2.5.3 Skills, Rules and Knowledge

Skills, Rules and Knowledge (SRK) approach (Rasmussen, 1983) rejects simplified narratives pertaining to errors which describe error creation as part of a human condition. This narrative of errors being a part of the *human condition* has been adopted in the solutions that tackle human vulnerabilities and prominent frameworks discussed earlier that either aim to save humans from themselves or accept error creation as a human condition. Instead, SRK approach provided an insight for how decisions are made i.e. decision are made in different ways with different information, and indeed, in the cases of *novice* versus *expert* the same decision may be made in a variety of ways.

SRK introduced by Rasmussen (1983) provides a classification system for cognitive tasks that describe human behaviour and decision making within man-machine environments. According to this approach the type of task being performed can either be skill, rule or knowledge based that can potentially result in an undesirable outcome due to the physical or cognitive load being experienced by the individual whilst performing it.

The first level, called ‘Skill based behaviour’, is automated behaviour and requires very little conscious effort. This includes well-rehearsed behaviours such as riding a bicycle or a skilled musician playing an instrument. The intermediate stage is ‘Rule based behaviour’. During this intermediate stage, tasks are more cognitively demanding than skill based behaviour as tasks require actions to be guided by pre-set rules or procedures that are stored in memory. These rules can be taught or explicitly communicated. However, these sets of rules can be overwritten by ‘new rules’ that are created through individual’s learning and experience. For instance, a car is driven within pre-set ‘rules of the road’ such as no-turning at a red traffic light. However, an individual might decide to take a left-turn at a red traffic light as they have seen others do it without incurring any harm or challenge. Thus, the new rule becomes to get to the destination in the fastest time possible, overwriting the previous rule of no-turning at a red traffic light. The third and final category, which is the highest level of the three cognitive stages, is ‘Knowledge based behaviour’. This stage is the most cognitively demanding on individuals when delivering a task. Knowledge based behaviour is essential for novel situations as it occurs in environments that have no prior set of rules available for control or recovery. When faced with such a situation, an individual must have knowledge of the system, generate a range of hypothesis and, test the hypothesis through logic or trial-and-error before the situation is under control. If the state of the system has not changed after a hypothesis has been tested then another hypothesis needs to be generated and tested by the individual until the situation is resolved. For instance, Captain “Sully” Sullenberger landed US Airways Flight 1549 in the Hudson River in 2009 with 150 passengers. The novel situation of a flock of birds flying into the aircraft jet required the Captain to have knowledge of the system (various alarms and indicators), experience and, understanding of aviation rules before proceeding to generate a range of hypotheses. These hypotheses would then be logically worked out and/or implemented through actions to confirm if the state of the system had changed. These stages would be conducted consecutively until the situation was resolved i.e. safely landing of the flight with minimum to no loss of human life. Despite this model's widespread application in man-machine interaction, it has been argued that SRK can imply that there is a *preferred* or a *natural* way for creating sequences to support cognition which in some instances can diminish the *context* of situated actions (Hollnagel, 1992; Le Coze, 2015). Furthermore, there might not be such a clear delineation between behaviour types and it is challenging to predict human behaviour in complex environments (Kirwan, 1992).

SRK approach was adopted as it offers a perspective that human behaviour and decision making are complex and variable. This approach provides an understanding that human behaviour is subject to constant change due to the context of the situation, information cues being presented in the environment that inform interactions and decisions and, personal experiences of the individual. Therefore, a reductionist approach that associates unintentional errors that result in cyber incidents or breaches to an inevitable human condition can be simplistic and insufficient in understanding causal factors and in subsequently proposed solutions.

2.5.3 Generic Error-Modelling System

Introduced by Reason (1990b) Generic Error-Modelling System (GEMS) presents a taxonomy of tasks by integrating the SRK approach and cognitive psychology. This approach argues that errors can be generated from the type of task that is being performed by the human. Thus, human behaviour is dynamic in its nature and depended on the interactional context, the information being presented by the system and, experience and knowledge possessed by the human. Error Management Programme (ERP) discussed as part of relevant frameworks is founded on GEMS approach. Within this approach human errors are considered in isolation from environmental or other context related factors. According to

GEMS unsafe actions and decisions are believed to originate from unintentional or intentional actions which subsequently result in undesirable outcomes or errors. These errors are classified into four categories: slips, lapses, mistakes and violations.

‘Slips’ in memory are linked to attentional failures that occur within an individual while performing a task. For instance, when an individual is performing a task, a certain step or aspect pertaining to a task can slip the mind of the individual. This is not to say that the individual was not aware of said step but rather simply that it has slipped their memory because they were focused on another aspects within the task while it was being performed. ‘Lapses’ is the second category within GEMS responsible for producing errors. Lapses occur in memory whereby individuals know the answer but cannot locate the information in their mind, resulting in a lapse or failure of required information retrieval by the brain. This can be seen in individuals ‘drawing a blank’ when performing a task i.e. a lapse of memory. The third category of errors originate from ‘Mistakes’. Mistakes extend the SRK approach whereby errors are either rule or knowledge based. The last category of errors is generated from ‘Violations’ whereby unsafe routines are normalised or a violation occurs through a novel application of known information in exceptional circumstances. Novel application of existing information in exceptional circumstances deemed as ‘violations’ has been witnessed in a range of high-profile aviation and nuclear industry incidents. In the context of unintentional insider threat, normalisation of unsafe routines can include leaving a security protected door open to strangers for entry or leaving a fire door open for ventilation or, writing passwords on sticky-notes which are left around the desk. However, it has been argued that GEMS has not resulted in the creation of established techniques that aid in the application of this model and, it is arguably oversimplified as it does not capture the complex multitude of actions that occur in real-world settings (Johnson, 1999).

GEMS was adopted as it offers an understanding of the *types* of errors that occur in a range of situations and provides a deeper understanding for skill, rule and, knowledge based behaviours as presented in SRK approach discussed above (Levine and Woody, 2010).

2.6 Summary

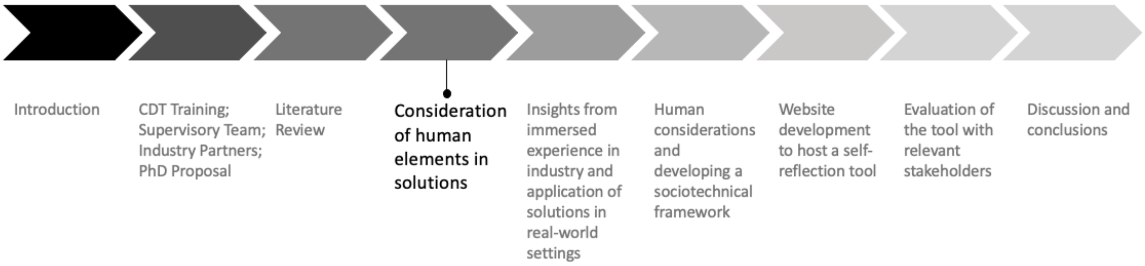
This Chapter provided an overview of cybersecurity and a paradigm through which cybersecurity can be considered. This paradigm began by categorising all actions within cyberspace as either offensive or defensive. Defensive cyberspace operations were further categorised as either active, passive or mixture of the two (active *and* passive). In order to appreciate threats within cyberspace, software and human vulnerabilities which afford cyber threats were discussed. Whilst many types of threats exist within cyberspace, prominent attacks that are the leading cause for cyberbreaches were highlighted as they are intertangled with the human element within the cybersecurity chain. After establishing an understanding of the nature of attacks, prominent frameworks were evaluated with a human factors stance as these frameworks have served as inspiration for subsequent solutions to tackle unintentional insider threat (UIT) within the field of cybersecurity. Numerous challenges that arise from these frameworks in the context of UIT were discussed as they attempt to classify, understand, monitor and, aim to avoid erroneous actions that can compromise system security.

Sociotechnical theory and perspectives were introduced to provide context with which unintentional insider threat (UIT) can be examined as ultimately, UIT also exists in complex, dynamic and, responsive environments and results in undesirable outcomes i.e. cyber incidents or breaches. The Epidemiological Triangle offered an alternative perspective to traditional security thought, psychological and behavioural approaches and, subsequent frameworks by presenting the relationship between the three vectors of: human, attack and, the cyberspace environment in which they coexist. The Swiss Cheese Metaphor offered an analogy that aids in understanding how errors can be generated in defences that exist in complex sociotechnical systems due to intrinsic vulnerabilities of the contributors. Otherwise robust defences can still lead to accidents or the generation of errors as vulnerabilities in defences can align in a way that is favourable for the attack to succeed. With the analogies of the Epidemiological Triangle and the Swiss Cheese Metaphor, investigating a singular cause (which is the case for approaches and solutions being presented to tackle unintentional insider threat) can be limiting as problems can arise from the interaction of several factors in complex sociotechnical systems that exist within cyberspace. Safety II approach was discussed that fundamentally believes that modern-day complex systems require safety science to learn from *things that work correctly* instead of focusing on eradicating errors. This approach naturally lends itself to investigating unintentional insider threat as the action that led to a cyberbreach might have been practiced many times previously without generating any adverse outcomes. Safety II also uses the contrast to Safety I in order to provide context to Safety II. This is similar to the approach in this thesis as intentional insider threat is examined and discussed in order to provide context to unintentional insider threat. Skills, Rules and Knowledge (SRK) approach is presented to provide an argument that a simplistic approach that reduces errors to a human condition is insufficient to understand UIT. Furthermore and in line with SRK's suggestion, human behaviour and decision making are not monolithic, humans can behave in different ways depending on the nature of the situation, the available informational cues that inform interactions and decisions and, their own experience which can result in UIT. Generic Error-Modelling System (GEMS) was presented which integrates SRK and cognitive psychology to classify the types of errors that are generated when tasks are performed. This is suitable in the context of unintentional insider threat as human behaviour is believed to be informed by a range of factors such as information, context, experience and, knowledge. Thus, these approaches i.e. Epidemiological Triangle, Safety II, SRK and, GEMS, further the understanding of the complexity that exists in human decision making, performance and, environments. This is in contrast to existing approaches to insider threat that reduce root-causes to binary understandings of *good* or *bad* people or decisions with an aim to eliminate them or coerce people into conforming to the desired behaviour.

Having established a human centric lens with which unintentional insider threat (UIT) can be examined through the discussion above, the following Chapter details the findings from a critical analysis which explores the extent to which a notable approach introduced by CERT holistically interweaves cybersecurity elements from a sociotechnical perspective to guard against insider threat.

3. Critical analysis of cybersecurity recommendations

3. Critical analysis of cybersecurity recommendations


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

Various approaches designed to tackle unintentional insider threat were discussed in the previous Chapter. These approaches are often presented as fused blanket solutions to defend against both subsets within insider threat (i.e. intentional and unintentional) however, are arguably limited in tackling unintentional aspects. Sociotechnical theory perspectives were also discussed to further understandings about considerations of the human element, environments, performance and decision making.

This Chapter explores the extent to which solutions consider the nuances and complexities that exist in sociotechnical environments within which work is conducted as it consequently impacts the applicability of proposed solutions. It also aims to demonstrate the extent to which these recommendations are applicable, convenient and holistic and the need for human-centric solutions to contribute towards the challenges associated to unintentional insider threat.

A critical evaluation of a notable guide introduced by CERT, titled ‘Common Sense Guide to Mitigating Insider Threats, Sixth Edition’ (Theis et al., 2019), was conducted to evaluate the extent to which it interweaves cybersecurity holistically within its recommendations for small-to-medium sized enterprises (SMEs). This guide is targeted primarily at UK organisations whereby suggestions can be incorporated quickly and conveniently to establish insider threat programmes. The aim of this critical analysis exercise was to establish the applicability and convenience of these recommendations for SMEs by considering recommendations for '*all organisations*' presented at the end of each chapter.

Whilst there are numerous options which can be used for evaluation to showcase the aims mentioned above, CERT's guide was deemed the most suitable for a number of reasons. Firstly, CERT is one the leading voices for providing guidance on insider threat related challenges including best practices, case studies to offer learning opportunities and, current trends pertaining to this threat. Due to this positioning this guide becomes suitable as recommendations are adopted by industry and provide future directions for academic research. Additionally, this guide is derived from the research and analysis of 1,500 real-world cases and is thus embedded in the context of real-world settings which aligns with the overarching context of the research presented in this thesis. In addition to the above, this edition of the guide was aimed at the UK audience, a region that this research project is based in, as it was developed to comply with European Union's General Data Protection Regulation (GDPR) law. The guide subsequently provides recommendations that limit or exclude the monitoring of individuals which are regionally appropriate. In lieu of monitoring individuals, positive incentives are introduced as part of implementing best practices to "*align the workforce with the organization*". Work in this guide recognises that insider threat is influenced by a range of sociotechnical factors such as technical abilities, behaviour inclinations and, organisational issues. To address these threats organisations are advised to closely consider their policies, procedures and technologies. This stance indicates consideration to aspects beyond the technological element which added to the suitability of this guide. The guide is aimed at businesses of all size that belong from all sectors as the *types* of insider attacks remain the same i.e. intentional or unintentional (however, *attack paths* or methods deployed by the insider might be subject to change depending on the sector) adding to the suitability of this guide for a critical analysis. Furthermore, it was of interest to determine if the understandings developed from the literature review in the previous Chapter were reflected in this guide i.e. an emphasis on technological elements to control, limit and predict human operations within cyberspace and the limited considerations paid to the operation of the human element in systems. The argument being that it might be problematic to approach elements within complex systems with simplistic views that subsequently offer oversimplified solutions. In addition, this exercise aimed to establish areas within environments that are being emphasised and held responsible for safeguarding against insider threat and identify opportunities for reframing existing thoughts from a human centric stance.

A sociotechnical systems approach called the *onion model* from the human factors domain is then applied to said recommendations in order to identify the elements responsible for mitigating insider threat. An unequal distribution of recommendations when classified by the categories presented in the onion model would be indicative of the importance placed on certain elements and the discounting of other aspects in proposed solutions.

A sociotechnical systems approach was deemed suitable as it considers social and technical factors when organisations are implementing a change, which can range from new technology to business change programmes (Cherns,1976). As organisations exist within complex sociotechnical systems that are created by them, implementing change in one aspect can affect other parts of the system and limit effectiveness (Hendrick,1997). Numerous methods that apply sociotechnical systems principles were considered prior to the selection of the onion model. For instance, Human Factors Analysis and Classification System or HFACS (Shappell and Wiegmann, 2003), Swiss Cheese Metaphor (Reason, 1990a), Systems Theoretic Accident Modelling and Process model or STAMP (Leveson,2004) and the Leavitt model (Leavitt, 1965). These methods, including the onion model, all stem from the same sociotechnical systems principles and largely aim to accomplish the consideration of all elements in a system equally and in an interconnected and interdependent manner. However, the onion model was selected as it presents a clear visualisation of a complete human-environment system in a simplified and accessible manner to its audience. The other models such as HFACS or STAMP, would need further granularity in the organisational personas whereas the Swiss Cheese Metaphor can appear linear in its representation. Leavitt's framework or the Leavitt model could appear overly complex when mapping the guide's 79 recommendations and potentially reduce the visual impact of the redistribution.

The outcomes from these two endeavours, i.e. evaluating the guide's recommendation to organisational personas and applying the recommendations to the onion model, serve to showcase two points: current approaches, such as CERT's guidelines, that aim to tackle unintentional insider threat can be enhanced in their holistic approach and, findings make a case for the human factors domain to contribute towards the challenges associated to unintentional insider threat.

This Chapter now proceeds to share the method used to critically analyse the guide. The writing progresses to discuss findings that emerge from implementing 79 recommendations to pseudo company profiles which serve as case studies. In line with document analysis method, these findings are discussed in a chapter-by-chapter format to maintain the structure of the guide. Work then proceeds to introduce a human factors perspective by reclassifying recommendations according to the eight sociotechnical categories contained within the 'onion model'. A summary is then presented to conclude this Chapter.

3.1 Method

Document analysis method (Bowen, 2009) was adopted to interpret the guide’s recommendations with an aim to analyse their applicability to SMEs. This method was selected as it would give meaning to these recommendations in light of the findings from extant literature review that indicated an inclination towards certain elements in systems. So as to avoid any preconceived ideas about the nature of recommendations a grounded approach was adopted to examine the applicability and attainability of recommendations to SMEs. Additionally, document analysis method was deemed suitable as it provides researchers with progressing through the document systematically, is less time constraining and less costly compared to other methods which were initially considered and excluded such as, empirical field-work with industry partners, interviews with individuals from industry or workshops.

Analysis began by compiling and coding all 108 recommendations to create a table (presented in Appendix 2) to identify recommendations applicable to SMEs. With this criteria i.e. recommendations applicable to SMEs, 79 recommendations were identified as being relevant to SMEs and 29 recommendations for large organisations were considered beyond the scope of this work. Large organisations were excluded as recommendations can be achievable by organisations that have resources, such as human resources and monetary funds, available to them, as is the nature of organisations that are qualified as *large* which have over 250 employees, turn-over equal to or in excess of 50 million (Euro) and a balance sheet equal to or in excess of 43 million (Euro). In addition, SMEs were suitable to reflect the intricacies of sociotechnical systems on micro or small levels which might be less pronounced in the complexities present in larger organisational structures.

In line with document analysis, recommendations were provided context from extant literature presented earlier whilst maintaining the structure of the document being analysed. A heuristic approach (Groner et al., 2014; Friess, 2015) was adopted to enrich the document analysis method by providing industry contexts and highlight usability issues in sociotechnical systems that might emerge from the implementation of these recommendations. Thus, organisational personas or scenarios were created by the author to act as case studies for each of the 21 chapters within the guide. A heuristic approach is advantageous as it provides access to information which is broadly representative of an organisation it aims to symbolise. Personas were created with the EU definition of SMEs (European Commission, 2003) and through a variety of channels by the author such as, media reports of breaches that described the victim organisation, desk-research of specific industries to establish realistic scenarios that were representative of their ways of operation, documentaries that provided insights about working in specific environments and, the experiences shared by industry partners of this research. These organisational personas were not developed to a specific theme outlined in the guide's chapter and selected prior to evaluating recommendations so as to maintain a grounded approach. In the guide's chapters where recommendations were found to be applicable or too generalist to benefit from presenting a case study, an industry persona is not presented as it was not deemed to be beneficial.

Organisational personas presented as case studies were used to evaluate three aspects: if the recommendations could be implementable, if the recommendations were easy to achieve and, if the recommendations had a high-impact to safeguard against insider threat for SMEs. Once recommendations were evaluated in an SME context, another table was created to capture the imagined ease for implementing these recommendations (presented in Appendix 3) where ratings were depicted as a range of three symbols ('✓' which indicated that recommendations would be easy to implement and achievable , ‘?’ which indicated that the recommendation was actionable but not easily achievable or, ‘✕’ which indicated that an SME would not be able to achieve or easily implement this recommendation). Brief comments that discuss the possible challenges that SMEs might encounter are listed in the column titled ‘*Potential challenges for implementation for SMEs*’.

After recommendations were evaluated through SME scenarios, they were numbered and re-categorized according to each of the aspects found in the ‘onion model’ by Wilson and Sharples (2015) maintaining a grounded approach to the analysis of data. This recategorization allowed new distributions and groupings to emerge, presented in a table format (Appendix 4) and as Figure 5 later on in this Chapter. For instance, if a recommendation related to ‘technologies’ i.e. configuring software and hardware of technologies or, ‘people’ i.e. thinking through decisions prior to action.

3.2 Application of recommendations to case studies

This section critically analyses the guide by evaluating recommendations presented under ‘Quick Wins’ and ‘High-Impact solutions’ within each of the guide’s chapters. Recommendations take into account technological, behavioural and, organisational aspects and are presented in their relevant subsections below (full list of recommendations within the scope of this exercise are presented in Appendix 2).

This guide is aimed at businesses to help them implement an insider threat programme at their organisations and pseudo organisational personas are presented as case studies to enrich the document analysis and to provide an industry context. Despite heuristic analysis being a useful tool to evaluate usability issues in systems and document analysis providing structure and context to the document being evaluated, limitations with these techniques arise from the inherent subjectivity within these methods as they rely on the evaluator's subjective knowledge and judgement to determine the severity of issues and, highlight aspects that might not necessarily be important in real-world settings (Love, 2013; Friess, 2015).

Chapter 1: Know and protect your critical assets

Recommendations presented in this chapter of the guide are as follows:

- Conduct a physical asset inventory. Identify asset owners' assets and functions and identify the type of data on the system.
- Understand what data your organization processes by speaking with data owners and users from across your organization.
- Identify and document the software configurations of all assets.
- Prioritize assets and data to determine the high-value targets.

– pg. 16-17, Theis et al., 2019

This chapter outlines the importance of organisations identifying their critical assets. Critical assets are described as: (a) anything of value or potential value to a company; (b) organisation is responsible for the security of such assets; (c) if a critical asset is destroyed or harmed in a way that could affect its confidentiality, integrity or availability it would result in a severe negative outcome for the organisation's operations. Critical assets can be physical or technological and can be comprised of a range of things including equipment, people, facilities, technologies and systems.

The guide states that with the advent of seamless technologies it is essential for companies to monitor and control data that is in rest or in transit as it can easily be removed from the organisation. It provides a list of questions to help identify and prioritise an organisation's critical assets and formulate a ranked list. Another ranked list must be created pertaining to employees who might pose a risk to these assets as insiders. This is followed by conducting a risk assessment, checking compliance to procedures in line with GDPR for organisations based in Europe and, developing compliance controls for operations when employees interact with critical assets.

Critical assets can also be identified through monitoring network traffic for digital assets and an inventory is to be developed for physical assets such as hardware. The inventory should include all the servers, their type of operating software, their ‘environment’ (integration, model, production), the applications running on each of the sever, each application’s corresponding IT support contact and the name of an employee who is the ‘owner’ for each of applications running within a system from the wider company. The guide suggests using a statistical software tool called ‘Pairwise Rankings’ to help create a ranked list and develop metrics for identified critical assets to the organisation.

The case study presented within this guide describes an incident at a small private hospital facility where a nightshift security guard accessed the server room twice, once through his security card and a second instance from a nurse’s unattended workstation, in an attempt to launch a distributed-denial-of-service (DDoS) attack. He was left unsupervised during his shifts and in his personal life acted as a leader of an online underground hacking group. Eventually his malicious actions led to the heating, ventilation and, air conditioning (HVAC) to become unstable and caused a power outage for one hour. A security researcher discovered this insider’s malicious activity. The discussion in the case study is centred on the night shift security guard as a malicious insider, ignoring the actuality of two types of insider threat that unfolded at the hospital facility (the security guard and the well-intentioned nurse), both of which should be considered in equal importance for research and analysis.

SME Case Study Evaluation

This is a heavy rescue towing and incident response company. It struggles to financially breakeven in most years as the business is climate dependent (bad weather would equate to more breakdowns and a financially lucrative year). This organisation has fewer than a 100 employees and the organisation collaborates with the highway authorities to assist with emergency breakdowns, recovery of vehicles and, provide accident response. They have a dispatch centre, a small fleet of trucks, a website and, seasonal employees for busier months. The organisation experience a high turnover. Communications are over the phone or through truck radios which are used to request back-up, seek advice from experienced drivers on difficult jobs and, give task status updates to the dispatch centre. The owner of this company is an expert in the recovery of heavy-goods-vehicles (HGVs) and inherited this business which has been family owned for two generations. The owner is responsible for administrative tasks, maintenance of equipment, accounting, HR operations (such as recruiting, training, grievances etc), strategic decisions and, ensuring smooth day-to-day operations.

Analysis, Chapter 1

With this organisational scenario to serve as a case study, recommendations put forward by the guide as ‘Quick Wins and High-Impact Solutions’ will now be considered. It might be relatively straight forward to achieve creating an inventory of the digital and physical critical assets while at rest and in transit and the type of data that exists in various systems. Such assets might include the fleet of recovery trucks, people, equipment, the website, payment details of clients etc. However, creating such an inventory would mean a lot of time and resources being spent towards developing it. In this scenario where one person is performing multiple organisational roles, it might be relatively straight forward for them to identify what data the organisation processes and list assets to create a ranked priority list. However, when identifying ‘assets owners’ i.e. people who are responsible for protecting these critical assets, the owner of this company will be responsible for a vast majority of the items on the list, making this process problematic and difficult to keep up to date. This challenge is not farfetched for SMEs who often have one person performing various job functions with numerous skills in a flat management structure (Frantz et al., 2017). A high turnover might mean that there is no one to take ownership of the critical asset if someone leaves and keeping the inventory up-to-date would require continuous diligence and time. Identifying and documenting software configurations of all assets might lie well beyond the capabilities of the CEO of an SME, especially if the CEO is adverse to information technologies (Thong and Yap, 1995) . Mapping critical asset information might also pose additional challenges if the business is outsourcing specialist functions such as using an online accounting platform or using third parties to build their website.

Chapter 2: Develop a formalized insider threat program

Recommendations presented in this chapter of the guide are as follows:

- Ensure that legal counsel determines the legal framework the team will work in.
- Establish policies and procedures for addressing insider threats that include HR, Legal Counsel, Security, Management, and IA.
- Consider establishing a contract with an outside consulting firm that is capable of providing incident response capabilities for all types of incidents, if the organization has not yet developed the expertise to conduct a legal, objective, and thorough inquiry.

– pg. 31, Theis et al., 2019

This chapter synchronises technical system logs with human action and intelligence to tackle insider threat. It primarily relies on organisational monitoring and peers alerting and reporting individuals that appear to be conducting suspicious activity or have experienced a sudden change in their personal financial circumstances. A working group is recommended to be set up consisting of employees across the organisation, with 'trusted agents' (usually line managers) to provide context or legitimacy to individual's suspicious actions that are flagged up in system logs. This chapter points out that any monitoring should be legally permissible, the organisation should use encrypted communications within the working group for confidentiality, utilise HR's 'watch list' to monitor concerning employees who can potentially pose an insider threat and, deploy good practices for terminating access for employees leaving the company.

SME Case Study Evaluation

This company provides an online software platform that helps researchers write and collaborate on articles intended for publication in academic journals. It has between 100-150 employees with teams dedicated to development of the software (50 software engineers), an IT team (3-5 people), a HR team (1-2 people), copy editors (30 people) and, a Sales & Marketing team (20-30 people). It has two CEOs who report to three actively involved owners, a Board of Trustees who offer strategic advice and approve business-critical decisions and, the company outsource an external law firm that provides them with legal counsel. This law firm charges the organisation an expensive hourly rate for consulting on any documents that are put forward to them for review. This company has been in existence for ten years and has recently managed to financially breakeven.

Analysis, Chapter 2
The first recommendation to determine the legal parameters of the working group (listed in Appendix 3) can be implemented by the company but the process might be financially expensive and time consuming. Creating a legal framework that the law firm can be consulted on would require in-house rudimentary legal skills which might not exist. If the legal firm is to create this framework, it would require for the insider threat program team to be able to understand legal terminology to interpret their operational parameters once it has been created. The second recommendation is problematic as it requires policy making skillset to exist within the company. This can create additional workload for relatively small teams (1-5 people) and can be time and effort intensive. Hiring a third-party to provide incident response capabilities could create an additional financial burden that the organisation might not be able to bear in this company scenario. For a SME company

fighting to survive, it could mean that developing incident response capabilities might be a low priority and hard to justify as SMEs often struggle to be financially stable. The Business statistics (Ward and Hutton, 2021) reported that 75% of UK businesses had 0-1 employees in 2021. While it is possible that there might be high-impact from these recommendations, each suggestion would be time and cost consuming and difficult for a SME to develop as it competes with its finite amount of existing resources.

‘Understanding and Avoiding Potential Pitfalls’ section within this chapter can be read as a cautionary warning when approaching the implementation of learnings and recommendations. These points largely undermine the recommendations presented in this chapter to set up an insider threat program within an organisation. Setting up a working group to deliver an insider threat program with importance placed on covert monitoring and reporting on selected individuals or ‘targets’ can induce a safety climate rather than a safety culture (Mearns and Flin, 1999) – lulling the organisation into a sense of robust cybersecurity state. Instead, a non-punitive proactive self-reporting culture can be developed in an organisation that can offer learnings for the entire organisation from ‘near misses’ for unintentional or accidental insider threat. Proactive reporting culture has successfully been implemented in aviation, nuclear energy, petrochemical processing, military operations and steel industries (Barach and Small, 2000). Since proactive reporting culture is closely connected to a range of sociotechnical factors (such as the attitudes of the employees), the ways in which the program is implemented and, managerial attitudes thus, it is important for an SME organisation to begin by understanding its existing practices and influencing factors prior to implementation (Douglas et al., 2014) instead of being cautioned against starting an insider threat programme of its pitfalls.

Chapter 3: Clearly document and consistently enforce policies and controls

Recommendations presented in this chapter of the guide are as follows:

- Ensure that senior management advocates, enforces, and complies with all organizational policies. Policies that do not have management buy-in will fail and not be enforced equally. Management must also comply with policies. If management does not do so, subordinates will see this as a sign that the policies do not matter or they are being held to a different standard than management. Your organization should consider exceptions to policies in this light as well.
- Ensure that management briefs all employees on all policies and procedures. Employees, contractors, and trusted business partners should sign acceptable-use policies and acceptable workplace behavior policies upon their hiring and once every year thereafter or when a significant change occurs. This is also an opportunity for your organization and employees, contractors, or trusted business partners to reaffirm any nondisclosure agreements.
- Ensure that management makes policies for all departments within your organization easily accessible to all employees. Posting policies on your organization's internal website can facilitate widespread dissemination of documents and ensure that everyone has the latest copy.
- Ensure that management makes annual refresher training for all employees mandatory. Refresher training needs to cover all facets of your organization, not just information security. Training should encompass the following topics: human resources, legal counsel, physical security, and any others of interest. Training can include, but is not limited to, changes to policies, issues that have emerged over the past year, and information security trends.

- Ensure that management enforces policies consistently to prevent the appearance of favoritism and injustice. The Human Resources department should have policies and procedures in place that specify the consequences of particular policy violations. This will facilitate clear and concise enforcement of policies.

– pg. 35, Theis et al., 2019

This chapter outlines the need for developing awareness amongst employees about the organisational procedures, policies and consequences for rule breaking behaviour (where punishment should not be disproportionate to the offence). It sets out by emphasising the importance of expectation setting whereby employees who develop IPs for the company understand that they do not own it. Consistent reinforcement of policies that are supported by clear documentation can avoid unmet expectations for rewards (e.g. recognition, promotions, bonuses etc) and lead to a sense of fairness and equality where specific individuals do not feel like they're being targeted. The guide mentions the need for every employee being held to the same standards with no exemptions based on job titles and a regular review of the policies.

SME Case Study Evaluation

The company in this scenario manufactures and supplies various types of dental implants to local clinics. This company has been in operation for over fifteen years and has successfully digitised physical records over the past year. It has an in-house IT team, two offices that host expensive specialist equipment, a couple of delivery vans for daily drop-offs and employs approximately 30 people at any given time. Majority of the processes are automated but trained technicians are required to oversee the manufacturing process and examine dental reports to produce requested implants to exact specifications and materials. Within this organisation the owner is the CEO who manages fifteen trained technicians alternating a 24-hour working shift, four IT team members, four people who are responsible for taking orders, providing customer services and managing grievances, two people are responsible for accounting and, there are four drivers. With the exception of the drivers, employees have appropriate access to sensitive information that is required to deliver their job functions such as payment details, operation of the specialised equipment and dental records that have been submitted by the local clients (i.e. dental clinics). Employee turnover in the company is low but replacing a trained technician proves to be a lengthy process that takes several months to find adequate replacements.

Analysis, Chapter 3

In order to create a safety culture in this company scenario, the guide's suggestion to hold every employee to the same standard is important and might be implemented successfully if there is buy-in from senior stakeholders (CEO and senior managers). If senior stakeholder buy-in is absent it can be a time intensive task to convert stakeholders into being actively engaged, educated in cyber awareness and good at maintaining transparent communication (Dul et al., 2012). And if this conversion of senior stakeholders needs to be achieved then the outcome would be on the contrary of the guide's positioning of this recommendation as a 'quick win'. Requesting existing, trusted business partners to sign additional legal documents could result in the loss of clients or create a precarious position for the SME if partners refuse to sign documents in addition to their existing contracts. Depending on the accountability chain and internal procedures in place at trusted business partner organisations, it might result in orders being paused or delayed due to bureaucratic procedures (all orders are on hold until 'acceptable use policy' is signed).

Similarly, trained technicians who are difficult to source, might not be willing to sign additional documents that are not the norm in their industry and can cause significant delays to an already challenging recruitment process. Making company policies accessible to all employees via internal intranet, shared folders or email is a relatively quick task. However, designing and delivering training can be a costly expense for an SME as training programmes take resources in planning as well as during delivery whilst employees attended training sessions. Mandatory sessions such as these might result in animosity between management and the trained technicians who would need to make-up for the backlog of orders while they were in training sessions. While the recommendation to implement policies consistently to avoid the appearance of favouritism and injustice is correct, it too would consistently require resources to monitor activities, oversee the enforcement of policies and reprimand violations. This as a cumulative effort can push the boundaries of what is actually manageable by this manufacturing SME whilst it delivers its day-to-day operations.

Chapter 4: Beginning with the hiring process, monitor and respond to suspicious or disruptive behavior

Recommendations presented in this chapter are as follows:

- Ensure that potential employees have undergone a thorough background investigation, which at a minimum should include a criminal background and credit check.
- Encourage employees to report suspicious behavior to appropriate personnel for further investigation.

- Investigate and document all issues of suspicious or disruptive behavior.
- Enforce policies and procedures consistently for all employees.
- Consider offering an EAP. These programs can help employees deal with many personal issues confidentially.

– pg. 40, Theis et al., 2019

This chapter recommends that organisations should proactively deal with suspicious or disruptive employees to avoid developing malicious (also known as intentional) insider threat. Should it be legally permissible in the country of operation then within the hiring process background checks should be conducted for all applicants that details how applicants approached workplace conflicts. It is recommended that applicants conviction records should also be consulted and certain job functions should have stringent checks that directly correlate to their associated risks (for example customer complaints versus accounting and finance). Managers should be trained to identify and respond to inappropriate behaviour and the organisations should consistently enforce policies. Employees should also be trained to report concerning or disruptive behaviour from their peers and a formal process should be embedded in the organisation's practices to address employee grievances. Organisations should also be alert to an employee's personal financial problems or unexplained personal financial gain. While this chapter recognises the possible decrease in employee morale due to the implementation of a reporting culture, it does not provide any suggestions about how to maintain a high morale if these practices were implemented. This chapter also pays caution for the need to be legally compliant when sharing employee information.

SME Case Study Evaluation

This scenario considers a software company that provides project management tools for teams collaborating together within organisations. This SME has been in existence for five years, employs approximately fifty people who perform various functions such as software engineers, sales, marketing, accounting and, HR. The business operates informally, offers flexible hours of work and a majority of the employees are between 24-35 years of age. Employees work on projects that are of interest to them (with each employee required to be involved in at least three projects) and all work is conducted as part of teams with assigned project leads. The employee morale is high, employees frequently engage in recreational activities together (yoga, playing foosball, shooting hoops, table tennis etc) and, share a relationship where they can rely on each other for support, advice and assistance. As this company has a flat management structure, there is considerable employee turnover due to the lack of personal career growth. The company has a short notice period of three weeks as productivity from software engineers was seen to decline during their notice period. Thus, the company faces challenges when recruiting new employees in quick succession to backfill existing positions.

Analysis, Chapter 4
A majority of recommendations in this chapter are possible to achieve for the SME scenario described above but might create interlinked challenges. For instance, while it would be good practice to conduct background checks, including criminal convictions and a financial history credit check, it would ass time delays that can have a significant impact on operations of the company. Conducting a check on each potential employee (especially software engineers who interact with the company’s IP) would also entail additional

financial costs for the organisation. In this scenario, encouraging employees to report 'suspicious behaviour' (which could be defined by any measure or to any detail) can severely damage existing interpersonal relationships, communications and, compromise day-to-day business operations. Performance and behaviour in team settings are important factors when discussing a modern day workplaces and serious thought is paid to these factors in the human factor domain, including the design, layout, performance and outcomes to instil best practices and managing risks (Becker and Steele, 1995; Salas et al., 2008). Offering confidential Employee Assistance Programme support lines, keeping a strictly confidential record of suspicious or disruptive behaviour and implementing policies and procedures consistently across all designations would also be attainable and potentially contribute to a just culture in this case study (Dekker, 2011). However, this would require time, financial resources and, consistent diligence from IT and HR resources that might already be finite.

Chapter 5: Anticipate and manage negative issues in the work environment

The following recommendations are presented at the end of this chapter:

- Enhance monitoring of employees with an impending or ongoing personnel issue, in accordance with organizational policy and laws. Enable additional auditing and monitoring controls outlined in policies and procedures. Regularly review audit logs to detect activities outside of the employee's normal scope of work. Limit access to these log files to those with a need to know.

- All levels of management must regularly communicate organizational changes to all employees. This allows for a more transparent organization, and employees can better plan for their future.

– pg. 43, Theis et al., 2019

This chapter highlights the importance of consistently enforcing policies and consequences for violations. It is advised that security related policies are clearly communicated during induction of new employees and generally over the course of the year. While organisations are instructed to be as transparent as possible to set expectations for promotions and bonuses, organisations are simultaneously warned to be alert of potential threats that might arise as a consequence of such decisions (i.e. IP theft for personal financial gain during lay-offs). Employee Assistance Programmes (EAPs) are presented as a possible solution for curbing employee's reactions that can result in insider threat. However, this suggestion simultaneously removes responsibility from the organisation as the employee's reactions (to lay-offs for instance) are inadvertently implied to be 'their problems' that would require independent help and not as those resulting from organisational decisions.

SME Case Study Evaluation

This is an SME company that publishes scientific research content online. This company employs between 30 – 50 people (primarily content editors) at any point in time and has been in existence for three years. Higher management has communicated that with the recent loss of lucrative clients the company is fighting for its survival. Despite the content still being produced to a high volume, there has been a recruitment freeze and several people have left for external opportunities. Teams have reduced in size significantly with a total of fifteen employees remaining and the company is experiencing a shortage of specialised skills such as proficient content editors.

Analysis, Chapter 5
In uncertain times such as those mentioned in the scenario above, SMEs might struggle to consistently perform day-to-day operations and experience a loss of specialised labour. Under such circumstances, enhanced monitoring might not be possible due to the IT department prioritising daily support and also being involved with terminating access of ex-employees to organisational systems. Implementation additional tasks such as auditing logs and monitoring controls during this time might prove unattainable with additional workload being experience by employees. However, the organisation being transparent during times of change and maintaining open lines of communications can leverage trust and sincerity from employees towards the organisation thus, safeguarding against insider threat. Research conducted in the human factors field indicates the importance of trust and team functioning (Spector and Jones, 2004) and the trust between peers and higher management can be a driver for employee satisfaction, loyalty and effective performance (Matzler and Renzl, 2006; Costa et al., 2001). Additional controls being recommended for audits of procedures and policies might instil an absolute organisational faith in processes and limit the leveraging of specialised employees to devise new, innovative and, efficient ways of work (Dekker, 2017).

Chapter 6: Consider threats from insiders and business partners in enterprise-wide risk assessments

Recommendations in this chapter are as follows:

- Have all employees, contractors, and trusted business partners sign nondisclosure agreements (NDAs) upon hiring and termination of employment or contracts.
- Ensure that all employees, contractors, and trusted business partners sign workplace violence prevention and/or appropriate workplace behaviors documentation upon hiring.
- Ensure each trusted business partner has performed background investigations on all of its employees who will have access to your organization's systems or information. These should be commensurate with your organization's own background investigations and required as a contractual obligation.
- If your organization is acquiring companies during a merger or acquisition, perform background investigations on all employees to be acquired, at a level commensurate with your organization's policies.
- Prevent sensitive documents from being printed if they are not required for business purposes. Insiders could take a printout of their own or someone else's sensitive document from a printer, desk, office, or from garbage. Electronic documents can be easier to track.
- Avoid direct connections with the information systems of trusted business partners if possible. Provide partners with task-related data without providing access to your organization's internal network.

- Restrict access to the system backup process to only administrators responsible for backup and restoration.

– pg. 47, Theis et al., 2019

In this chapter of the guide importance is placed on developing a risk-based security strategy. This strategy aims to protect critical assets from internal and external threats that might emerge from internal employees, external trusted business partners, consultants and, contractors with authorised access. This chapter acknowledges the natural tension that creates a paradox between core interests such as business productivity and (cyber-)security of critical assets in an organisation. It is suggested that a balance must be struck between security procedures that counteract insider threat and procedures that allow a company to accomplish its mission. This stance is captured in the following sentence:

"*Having too many security restrictions can impede the organisation's mission, and having too few may permit a security breach*"

– pg. 44, Theis et al., 2019

Signing legal agreements with external partners is advised, especially ensuring that external partners are performing the required background checks and investigations on employees who collaborate in mutual partnerships.

SME Case Study Evaluation

This SME manufactures traditional leather satchels and bags in the UK. Producing 1970's iconic style of satchels mean that neither the process nor the design are unique or under copyright by the organisation. There are approximately fifty employees which are primarily in-house designers and sewists to promote high quality British products. Originally, the company was sourcing its products from a company based in China but the partnership dissolved when they did not return a non-disclosure agreement (NDA) after numerous reminders sent over a period of fourteen months. Now, the organisation has trusted local providers who deliver within different stages of the sub-processes (raw material processing, tanning, crusting, dyeing, surface coating, etc). Once the materials are processed, sewers stitch the bags which undergo quality testing before being supplied to a couple of luxury retail stores and sold directly to customers through their website. As business has grown over the last decade with celebrities endorsing their brand, the company has been able to outsource its IT, legal and, accounting services to trusted business partners. Producing high quality leather satchels for over a decade, the SME is focused on managing the high demand and does not currently have any plans of further growth.

Analysis, Chapter 6

For this SME scenario, creating non-disclosure agreements (NDA) would entail additional financial costs. It could also be time consuming to reach an agreement which can cause significant delays or dissolution of critical partnerships such as those experienced with Chinese suppliers at the start of this business venture. As the brand enjoys popularity, the process nor the design are unique or copyrighted (a classic 1970’s satchel design). This might mean that this SME does not believe that they have something of value that would require business partners to sign a NDA. Requesting to sign an NDA at this stage for the SME might sour established relationships with trusted business partners and jeopardise business operations. While in-house employees might sign workplace violence prevention and/or appropriate workplace behaviour documentation, it might be a time consuming task to develop this for third party contractors and business partners. This SME might be aware that trusted partners, such as those that provide IT and accounting services, perform credit checks on their employees but it would not be possible for the SME to ensure that this is indeed true. As part of good practices, the company can choose not to share direct connections with the IT department of trusted business partners and provide only task related data. Partners that provide the IT functions for the organisation can only be made responsible for backing up data and restoration, managing access and provide use of cloud and local servers. However, in-house sewists might use print outs of documents to make notes about their sewing tasks which is a commonly occurring trait in this field. Restricting the of printing documents might mean that sewists make mistakes or might need to be trained in the skill of making online edits. Apart from the expenses of training, it would be financially expensive for this SME to provide sewists with technology (desktops, tablets, pens etc for digitisation) to do design specifications with, especially if sewists are not motivated to switch from physical documentation in the first place.

Chapter 7: Be especially vigilant regarding social media

The following recommendations are provided in this chapter:

- Establish a social media policy that defines acceptable uses of social media and information that should not be discussed online.
- Include social media awareness training as part of the organization's security awareness training program.
- Encourage users to report suspicious emails or phone calls to the information security team, who can track these emails to identify any patterns and issue alerts to users.

– pg. 52, Theis et al., 2019

This chapter discusses the possible correlation between the use of social media and insider threat (intentional and unintentional). Since social media platforms, personal or professional, allow information sharing opportunities with other people most information shared on these platforms are set to 'public', leaving behind a digital footprint. Consequently, a great deal of personal information can be found out about individuals through an online search. This chapter explains that any information shared maliciously or unintentionally on online platforms about an organisation can be used by attackers to design cyberattacks that can compromise critical organisational assets. This includes asking troubleshooting questions about organisational platforms where information about implemented technology can be revealed including information about operating software, the make and model numbers of devices and, Internet Protocol (IP) address. Recommendations strongly suggest that any company monitoring of social media platforms of employees must ensure they are doing so legally. Furthermore, organisations must be careful when reprimanding or penalizing employees who share working conditions, managerial complaints and other opinions that might be permissible under local law. Organisations must also avoid discrimination based on personal information gathered from online platforms such as personal perspectives on race, religion, sexual orientation etc that can result in legal lawsuits. The guide highlights the 'right to be forgotten' under GDPR for European Union citizens which can render selective search results on individuals. Finally, the guide argues that social media accounts of employees are a serious risk to the organisational cybersecurity and its use must be strictly controlled and regulated.

SME Case Study Evaluation

The company in this scenario has been in existence for a little over five years and has recently finished another successful round of funding from private investors. It employs 25 young, tech-savvy individuals and they provide filters for photographs that are shared on popular online social media platforms. The app and all its filters are free to download by users on their mobile devices and the app launches new filters every two weeks. These filters allow users to write text, insert stickers, overlay gifs, edit hues and, compile numerous photographs to create collages. The company has thousands of followers on their official social media accounts that exist on all major social platforms. In order to bypass the need to share company account passwords, employees often use their personal accounts to promptly help answer questions and to promote the release of new filters to their target audience. All employees understand the importance of confidentiality when developing filters and the use of their personal accounts when engaging with clients, should they choose to do so. Employees are mindful about never sharing any spoilers about the new filters. Usually the late deployment between agreeing on an idea and then the development of the filter means that there isn’t enough time to disclose any vital information.

Analysis, Chapter 7
In context of the guide’s recommendations, this SME could benefit from establishing a social media policy for its employees despite having a collective understanding of appropriate social media use in place. It might also benefit from employees reporting suspicious emails to the IT department that in turn could be used to raise awareness amongst employees for potential socially engineered cyberattacks. However, it could result in the IT teams being inundated with reports or data that might be ‘false alarms’ (Treisman, 1965) which can be influenced by personal perceptions and various biases of the IT personnel – false positive alarms are a major concern for current insider threat identification software (Agrafiotis et al., 2016; Martinez-Moyano et al., 2006). In addition,

identifying suitable training programmes for tech-savvy or advance skilled employees could be financially expensive, time consuming and can also risk patronising employees and damaging morale of the workplace.

Chapter 8: Structure management and tasks to minimize insider stress and mistakes

Recommendations within this chapter are as follows:

- Establish a work culture that measures success based on appropriate metrics for the work environment. For instance, knowledge workers might measure their success based on outcomes and efficiency instead of metrics that are better suited for a production line.
- Encourage employees to think through projects, actions, and statements before committing to them.
- Create an environment that encourages focusing upon one thing at a time, rather than multitasking.
- Offer employees who are under stress options to de-stress, such as massages, time off, games, or other social but non-project oriented activities.
- Routinely monitor employee workloads to make sure that they are commensurate with the employee's skills and available resources.

– pg. 55, Theis et al., 2019

This section outlines the correlation between multitasking in high-stress environments and the emergence of intentional and unintentional insider threat. The pressure faced by employees to deliver to tight deadlines can increase insider threat levels and develop a negative attitudes towards management and the organisation. The guide states that organisational overdrive for productivity can compromise (cyber)security protocols that are in place. This discussion once again acknowledges a natural tension that creates a paradox between core organisational objectives, such as productivity, and the (cyber)security of critical assets. It advises organisations to develop protective measures that are human centric, allow employees more time to achieve objectives, be responsive to human oriented management and, allocate adequate time towards planning tasks.

SME Case Study Evaluation

This company is an online publication SME which was established ten years ago and currently employs sixty people. Unlike other major publication platforms that might take up to a year to publish research content, this organisation's unique selling point is the publishing of cutting-edge advances in research to the scientific community in a matter of days. Upon receiving a manuscript the employees must perform basic checks (such as content originality), verify affiliations listed by the authors, categorise the content by discipline, identify relevant peer reviewers etc. Once the manuscript is with peer-reviewers, employees must follow up for comments, find alternative reviewers, take note of suggestions, copy-edit, maintain communications with submitting authors and, publish the content on the platform when ready. Employees are strongly encouraged to think before taking any actions as any mistakes in the process can damage organisational reputation in a highly competitive field. The SME uses 'total number of submissions published on the platform' as key process indicators (KPIs) to measure employee performance. Certain stages during the process are time sensitive and the steady stream of submissions means there is always a lot of work to be done in a busy working environment. Junior commissioners track the progress of several manuscripts that are assigned to them at various stages of review. Over the last two years, the organisation has measured employees stress levels, workloads and, resources available to staff through all staff surveys. However, these elements are normalised and this fast-paced environment is considered to come as part of the field.

Analysis, Chapter 8

Applying the recommendations from this chapter, it would be beneficial for this SME to develop appropriate metrics that measure success and efficiency in processes instead of solely using the total number of published submissions. Despite the importance placed by the organisation to think through projects before taking any actions, the inherent nature of the job which requires quick turnaround from submission to publication would mean unavoidable time pressures on employees. If time critical stages within processes are prolonged it can harm employees' KPIs and jeopardise the company's overall mission. Furthermore, since employees publish a range of scientific content and have a steady stream of submissions to process, individuals would need to multitask in order to succeed at their agreed objectives. Reducing complex environments, such as the one presented at this SME, to singular tasks being performed in a prescribed about of time as suggested by the guide can prove challenging. Similarly, this SME might not be able to financially afford offering time off and activities to de-stress employees. While this SME routinely measures employee workloads, skills and, available resources it might not have the financial ability to improve working conditions or demands. Even with disposable income available to the SME, implementation of recommendations would require major work re-design to tackle perceived workload, capacity and stress amongst its employees. In fact, the evaluation of work (Wilson and Sharples, 2015) and, understanding perceived workload and its measurement are complex phenomena in human factors domain that require detailed investigation to help inform design or redesign of workplaces. Analysis of workload involves the measurement of the cognitive demands and the capacity to respond to those demands within complex environments (Dekker, 2012) and sociotechnical systems (Sharples, 2018), where ubiquitous technology can provide data on valuable indicators (Sharples et al., 2015). Adopting the recommendations presented in this chapter would be difficult to achieve as recommendations would require human factors domain specialists, time and, financial resources in order to achieve effective outcomes. The need to involve specialists and associated financial costs might be beyond the SME's awareness and abilities.

Chapter 9: Incorporate malicious and unintentional insider threat awareness into periodic security training for all employees

The following recommendations are presented in this chapter:

- Develop and implement an enterprise-wide training program that discusses various topics related to insider threat. The training program must have the support of senior management to be effective. Management must be seen participating in the course and must not be exempt from it, which other employees could see as a lack of support and an unequal enforcement of policies.
- Train all new employees and contractors in security awareness, including insider threat, before giving them access to any computer system. Make sure to include training for employees who may not need to access computer systems daily, such as janitorial and maintenance staff. These users may require a special training program

> that covers security scenarios they may encounter, such as social engineering, active shooter, and sensitive documents left out in the open.
>
> - Train employees continuously. However, training does not always need to be classroom instruction. Posters, newsletters, alert emails, and brown-bag lunch programs are all effective training methods. Your organization should consider implementing one or more of these programs to increase security awareness.
> - Establish an anonymous or confidential mechanism for reporting security incidents. Encourage employees to report security issues and consider incentives to reporting by rewarding those who do.
>
> – pg. 60-61, Theis et al., 2019

This chapter of the guide discusses the importance of senior stakeholder buy-in to the insider threat program and its successful implementation at an organisation. The guide states that vulnerabilities in business processes are just as important as technical vulnerabilities in cybersecurity. In the absence of a stereotypical profile of an inside attacker (race, age, ethnicity, job title etc), known information and individual characteristics can be utilised to create a list of employees who might pose insider threat to the company. With the help of this created list, mitigation strategies can be implemented to counteract an attack if it occurs. Security training should include encouraging confidential peer-reporting of threatening or unacceptable behaviour (i.e. accessing company systems post termination, requesting peer employees' passwords or using company resources for personal business). Awareness training should include potential consequences of risk-taking behaviours, lack of attention to detail, multi-tasking, excessive access to personal or propriety data and, recruitment of employees by harmful external agents (i.e. through social media). Policies should be consistently enforced and periodically reviewed.

SME Case Study Evaluation: N/A

Analysis, Chapter 9

A specific scenario based on a case study is not presented in this section as recommendations in this chapter are more generalised in their nature. Developing and delivering training programmes is a challenging endeavour for organisations of all sizes. It requires content development, relevance to the audience, audience engagement, materials and, time. Thus, developing and conducting an organisation wide training program would be an immense challenge for organisations that require resources as well as buy-in from management and employees. Even if contractors are able to create training programmes, for an SME to deliver training to their staff who might possess completely different skill-sets and knowledge levels would be nearly beyond a SME’s capacity. Creating an anonymous mechanism for peer-reporting would also be extremely difficult as all devices would have something revealing about the individual reporting their concerns (IP address for instance). Additionally, IT department or another relevant person assigned to access these reports would then have additional workload, they would be required to exercise confidentiality at all times and, might be held responsible for any investigations that are being carried out on potentially harmful individuals. However, once recommended programmes have been developed and implemented, continuously training employees would be relatively easier to achieve but would still require resources such as time and money.

Chapter 10: Implement strict password and account management policies and practices

Recommendations presented in this chapter are as follows:

- Establish account management policies and procedures for all accounts created on all information systems. These policies should address how accounts are created, reviewed, and terminated. In addition, the policy should address who authorizes the account and what data they can access.
- Perform audits of account creation and password changes by system administrators. The account management process should include creation of a trouble ticket by the help desk. (Help desk staff should not be able to create accounts.) Your organization could confirm the legitimacy of requests to reset passwords or create accounts by correlating such requests with help desk logs.
- Define password requirements and train users on creating strong passwords. Some systems may tolerate long passwords. Encourage users to use passphrases that include proper punctuation and capitalization, thereby increasing passphrase strength and making it more memorable to the user.
- Security training should include instruction to block visual access to others as users type their passcodes.
- Ensure all shared accounts are absolutely necessary and are addressed in a risk management decision.

– pg. 65, Theis et al., 2019

This chapter discusses the importance of security training for all employees which includes setting strong passwords and being vigilant of protecting passwords from visual access by peers. It recommends implementing best practices to manage account access privileges, avoiding shared accounts and, performing regular account audits. Employees should report any attempts to gain access by unauthorised accounts to the IT help desk. Based on a company's termination policy access should be terminated promptly for any individuals leaving the organisation and contractors should never be granted access to the entire IT system or access to shared accounts. It is important to note that this discussion about detailed access controls, account management and other account security measures once again highlights the tension between core business objectives and the (cyber)security of critical organisational assets. It can also be challenging to keep on top of such access controls where people are conducting work on distributed workstations.

SME Case Study Evaluation

This SME delivers an online platform for organisations to host virtual meetings. It has sixty employees and has been in existence for seven years. Employees are predominantly software developers residing in two European countries who work closely together to develop new features and fix algorithm code clashes (i.e. bug fixing). There are teams based in the UK for sales, marketing, IT, accounts and, HR. Developers work to tight deadlines that factor in time dedicated to quality assurance (QA) testing which is conducted in staging environments (i.e. a platform that mirrors the live website) prior to its integration in production environment (i.e. the live website). The sales team experience a high turnover and have utilised various customised software for customer relationship management (CRM such as Salesforce) over the past seven years. As the information has not been integrated well when providers have been changed, data is distributed across platforms. Since the sales team work closely with the software developers, they also test for bugs prior to launch and troubleshoot for clients.

Analysis, Chapter 10
In the context of recommendation proposed in this chapter, it would be beneficial for the organisation to identify, manage and terminate excessive access from user accounts albeit it might be a difficult and arduous process. It might also be beneficial to ensure that software engineers who use shared accounts for QA testing are known and any shared accounts are subject to regular and frequent password change and audits. Exercising excessive account controls and limiting access privileges might disrupt business productivity and employees might be unable to perform daily tasks if the use, frequency and reasons for access are not properly understood prior to termination/restriction of access privileges. Furthermore, close consideration must be paid prior to limiting and controlling access to incorporate lessons learnt about designing productive sociotechnical systems for distributed work (Sharples and Houghton, 2016). Whilst it may be expensive, this SME would benefit from organising training workshops for employees that increase awareness for creating strong passwords, understanding the restriction of shared accounts and, blocking visual access to others as passwords are being typed.

Chapter 11: Institute stringent access controls and monitoring policies on privileged users

A single recommendation is made as part of this chapter:

- Conduct periodic account reviews to avoid privilege creep. Employees should have sufficient access rights to perform their everyday duties. When an employee changes roles, the organization should review the employee’s account and rescind permissions that the employee no longer needs.

– pg. 69, Theis et al., 2019

This chapter discusses the advantages system administrators, technical and privileged users have since they possess technical skills, access and sufficient knowledge of processes to pose insider threat. For this reason it is advised that stringent rules are implemented for privileged users. Privileged and skilled users should sign a privileged user agreement that outlines their code of conduct and operations and sets expectations for their conduct. Similar to two-factor authentication (2FA), it is recommended that software developers should have all their code approved by another developer before it is deployed to avoid malicious code being embedded in the system. It is recommended that there should be documented access termination procedures that are enforced and organisations that cannot afford two system administrators must recognise their increased risk of insider threat. The discussion in this chapter is centred on regularly conducting access audits for privilege creeping.

SME Case Study Evaluation

This SME provides an application for mobile devices to help individuals with meeting reminders and any associated documentation required for meetings. It also stores the chat logs which are available offline to track agreed follow-up actions. This organisation was created four years ago and has twenty employees with fifteen vacant positions that are being recruited. Teams operate on a skeletal framework as positions stay unfulfilled for months due to the lack of desirable candidates. The organisation also outsources cloud servers for backups and developers are challenged with tasks that are outside their expertise (HTML developer might be assigned designing user interface or writing code in JavaScript).

Analysis, Chapter 11
In light of the recommendations the SME discussed above would be able to afford hiring another system administrator. However, the IT team would not traditionally have the skills or the remit to oversee the work of software developers. Peer checking of all software code that is developed for the mobile application would increase the workload and developers might feel that they are being assigned more responsibility of approving colleagues' code and not trusted by the organisation in their skills and abilities. This might lead to low morale and resignations which can threaten operations for this SME especially as it already experiences difficulties in attracting skilled labour. With blurry lines between job roles (developers are responsible for a range of tasks such as user interface design, QA testing etc), IT team and senior management would need to invest significant time in determining the parameters of access privileges based on their constructs of tasks involved for each job title.

Chapter 12: Deploy solutions for monitoring employee actions and correlating information from multiple data sources

The following two recommendations are presented in this chapter:

- Implement rules within the SIEM system, to automate alerts.
- Create log management policy and procedures. Ensure they address log retention (consult legal counsel for specific requirements), what logs to collect, and who manages the logging systems.

– pg. 75, Theis et al., 2019

This chapter discusses the importance of fusing data from various system logs, including monitoring employee cyber actions, to tackle insider threat. It is recommended that information should be collected from a range of sources across the organisation since solely logging network activity is insufficient to safeguard critical organisational assets. Security information and event management (SIEM) system is recommended as powerful tools that continuously monitor employee actions, correlates these activities to events and eliminates background noise to highlight cases that require review or further investigation by security personnel. However, this chapter does not share how these software eliminate background noise i.e. false positives nor how the software decides which incidents need further investigation. This chapter also lists numerous other types of data that can be collected to be analysed in addition to SIEM tools.

SME Case Study Evaluation: N/A

Analysis, Chapter 12

A specific SME scenario is not created for this chapter as the recommendations pertain to the utilisation of software tools. Whilst there are numerous SIEM system tools available in the market, including those offered by IBM (QRadar) and McAfee (Enterprise Security Manager), this process would require resources such as financial investment, in-house employee skills to correctly set up the software alert thresholds and, time. Possessing and dedicating such resources towards this activity would pose major challenges for this SME. Additional legal and supplier costs associated to the retention of system logs that monitor employee interactions might also be unaffordable for most SMEs or add financial and legal pressures to the SME presented in this case study. However, once SIEM tools have been set up, it might be advantageous for identifying insider threat through correlating data from various sources into a single platform and identifying anomalies in the system.

Chapter 13: Monitor and control remote access from all end points, including mobile devices

Recommendations made in this chapter are as follows:

- Disable remote access to the organization's systems when an employee or contractor separates from the organization. Be sure to disable access to VPN service, application servers, email, network infrastructure devices, and remote management software. Be sure to close all open sessions as well. In addition, collect all company-owned equipment, including multifactor authentication tokens, such as RSA SecurID tokens or smart cards.
- Include mobile devices, with a listing of their features, as part of the enterprise risk assessment.
- Prohibit or limit the use of personally owned devices.
- Prohibit devices with cameras in sensitive areas.

– pg. 81, Theis et al., 2019

This chapter discusses providing employees with access points to work remotely through the use of ubiquitous technologies. Whilst ubiquitous technologies allow employees to 'telecommute' the guide states that this access poses its own set of organisational risks to critical assets. It is advised in the guide that access points must be known, controlled and, monitored to prevent insider threat against organisational data and systems. This includes technologies such as smart phones, remote home computers, tablets, mobile devices etc. Due to a high demand from employees to work from mobile devices, organisations have facilitated access paths. However, CERT National Insider Threat Center emphasises their stance against this facilitation as it is believed to create a high risk for malicious insider threats through remote attacks. It is acknowledged in the guide that whilst remote access can enhance productivity, organisations should be aware of associated risks and trade-offs prior to facilitating mobile or remote access. This chapter discusses the ability of personal devices to bypass system security measures in place (e.g. intrusion prevention systems, and, firewalls). For instance, mobile phones can capture sensitive information through video recording or pictures and transport this information externally (through multimedia messaging platforms like MMS or public cellular internet networks). This affordance offered by personal devices can also allow malicious insiders to go undetected and for organisational data to be transported outside authorised organisational IT networks.

SME Case Study Evaluation

This SME provides marketing services to other organisations to help develop their brand and marketing strategies. They have eighteen employees, twelve of whom are marketing experts. Each employee is a project leader for a client and they collaborate in teams of four to deliver various projects that can include brand image, online marketing campaigns, multimedia campaigns (TV, billboards, social media advertisements), marketing strategies and competitor analysis. They have one meeting a week where everyone is required to be physically present in the office, provide updates, highlights and, share any challenges they have faced in the previous week. Employees are rarely at their assigned workstations since they work remotely which has increased exponentially post the covid-19 pandemic. Employees also frequently conduct meetings with current and potential clients at their client’s offices. They often liaise with each other through mobile phone calls and text messages, emails and, other free off-the-shelf software tools. In addition, employees also choose to utilise communication platforms that are most convenient for their clients to communicate on.

Analysis, Chapter 13
For this SME disabling or terminating remote access would cause major disruptions in day-to-day operations and trigger unavailability of necessary documentation (portfolios) needed during client meetings. Promptly and diligently terminating access for ex-employees will be crucial for this SME to protect its critical assets which might include client details, rates, strategies and developed design work. However, declaring personal mobile devices for the organisation to scope its features and associated risks might not prove meaningful. In the context of implementing the recommendations within this chapter, it could mean that every employee is considered a risk because they own a cell phone with largely similar features (i.e. a camera, mic, video recording, mass storage capabilities). An organisational evaluation of personal device operating software might unduly target individuals as high risk due to the type technologies they own (for instance iOS or Android users), or perhaps identify malicious insiders as low risk based on their personal choice of devices. Completely prohibiting or limiting the use of personal devices would be extremely problematic as employees in this SME are trusted to work independently to deliver projects and their organisational culture incorporates remote working and frequent off-site client meetings. With the advent of Covid-19 pandemic and remote working, recommendation appear to have limited applicability although the challenge to protect organisational assets and networks remains a primary concern.

Chapter 14: Establish a baseline of normal behavior for both networks and employees

The following recommendations are presented at the end of this chapter:

- Use monitoring tools to monitor network and employee activity for a period of time to establish a baseline of normal behaviors and trends.
- Deny VPN access to foreign countries where a genuine business need does not exist. White list only countries where a genuine business need exists.34
- Establish which ports and protocols are needed for normal network activity, and configure devices to use only these services.
- Determine which firewall and IDS alerts are normal. Either correct what causes these alerts or document normal ranges and include them in the network baseline documentation.

– pg. 85-86, Theis et al., 2019

This chapter discusses the analysis of information rich data that can be generated and effectively utilised by organisations through ubiquitous technologies to counteract insider threat. Once SIEM tools are implemented, it discusses the importance of establishing a baseline for *normal behaviour*. Normal behaviour can include network and individual characteristics. Network characteristics include bandwidth consumption, usage patterns and protocols, while individual characteristics can include working hours, usage of resources and accessing documents that are considered to be critical assets. Several off-the-shelf software solutions can be adopted to assist with identifying normal network and individual behaviour. This chapter states that defining and enforcing access policies related to organisational virtual private networks or VPNs (for remote access) can help organisations detect insider threat. Security measures include blacklisting countries where there are no employees, implementing

VPN access controls (for e.g. limited sharing or downloading of documents), monitoring ports and protocols (as well as any port hopping) and regularly reviewing firewalls and intrusion detection systems. Individual characteristics can be evaluated against individual's own historic activity and also against their peers with similar job titles, departments and office space to identify anomalies in normal behaviour.

SME Case Study Evaluation

To evaluate recommendation an SME is now considered as a case study. This SME provides customer support for a market-leading software to other businesses who have implemented this software or are interested in purchasing it (i.e. this SME acts as a third-party broker for new sales and post-sale support to existing clients). This company has employees based in one country and clients across the world. Employees frequently attend international conferences and exhibitions where they meet potential and existing clients, manage client relationships, provide updates on ongoing cases and troubleshoot for existing clients.

Analysis, Chapter 14
For this SME, establishing baseline of normal behaviour and trends can be relatively straightforward with the help of software tools such as SIEMs. However, as noted in the guide's chapter, it would be challenging to maintain employee privacy and extremely difficult to identify ongoing malicious activity as it would be classified as normal behaviour by the software. Blacklisting international access to organisational VPNs would cause disruptions when employees are performing tasks at conferences and exhibitions. Employees would not be able to troubleshoot effectively without information previously received as part of the casefile or provide updates on reports to existing clients. This might

create a reputational risk for this SME, potentially loose new clients and undermine the objectives set out by the organisation to attend such events. Calibrating the baseline of normal behaviour to fine-tune alerts is a challenging process that can risk obvious threats going unnoticed, or the creation of far too many alerts that can result in time lost in unnecessary investigations. In fact, there is existing work within human factors literature regarding alarms that would be beneficial for designing alerts and informing recommendations (Stanton et al., 1992; Woods, 1995).

Chapter 15: Enforce separation of duties and least privilege

Recommendations for this chapter are as follows:

- Carefully audit user access permissions when an employee changes roles within the organization to avoid privilege creep. In addition, routinely audit user access permissions at least annually. Remove permissions that are no longer needed.
- Establish account management policies and procedures. Audit account maintenance operations regularly. Account activity should reconcile with help desk documentation.
- Require privileged users to have both an administrative account with the minimum necessary privileges to perform their duties and a standard account that is used for every day, non-privileged activities.

– pg. 89, Theis et al., 2019

This chapter discusses implementing ‘separation of duties’ to limit individuals ability to harm organisational processes, systems and, information without the cooperation of other employees. Least privilege should be enforced for access i.e. individuals are given access to the correct data and the exact amount of data required to do their tasks. Access should be audited for individuals who receive internal promotions or move laterally to perform other functions within the organisation. It is recommended that a two person rule (identical in its roots to two-factor authentication or 2FA) should be implemented for physical and cyber processes when changes are made to critical organisational assets.

SME Case Study Evaluation: N/A

Analysis, Chapter 15
A specific SME is not considered for this chapter as recommendations were widely applicable to all SMEs. SMEs would benefit from carefully auditing user accounts for access privileges as tasks and objectives change for individuals. It would be good practice to conduct account audits at least annually. Documentation from IT logs can be used to examine account activity in order to highlight anomalies that can indicate insider threat. However, it might be challenging for SMEs with 200 employees to issue, track and maintain multiple accounts associated to individuals and manage privileges associated to those accounts if this has not been an established practice.

Chapter 16: Define explicit security agreements for any cloud services, especially access restrictions and monitoring capabilities

The following recommendations are made in this chapter:

- Conduct a risk assessment of the data and services that your organization plans to outsource to a cloud service provider before entering into any agreement. Your organization must en-sure that the service provider poses an acceptable level of risk

> and has implemented mitigating controls to reduce any residual risks. Your organization must carefully examine all aspects of the cloud service provider to ensure the service provider meets or exceeds your organization's own security practices.
>
> - Verify the cloud service provider's hiring practices to ensure it conducts thorough background security investigations on any personnel (operations staff, technical staff, janitorial staff, etc.) before they are hired. In addition, the service provider should conduct periodic credit checks and reinvestigations to ensure that changes in an employee's life situation have not caused any additional unacceptable risks.
> - Control or eliminate remote administrative access to hosts providing cloud or virtual services.
> - Understand how the cloud service provider protects data and other organizational assets before entering into any agreement. Verify the party responsible for restricting logical and physical access to your organization's cloud assets.
>
> – pg. 93-94, Theis et al., 2019

The discussion in this chapter focuses on ensuring that critical organisational assets are secure when utilising cloud computing. Any protective measures implemented by an organisation should extend to data architecture and critical assets hosted on the cloud and not be left as the responsibility of the cloud provider. If the cloud provider attests to having defences in place to safeguard against attacks, then the organisation must carry out audits themselves or by an independent third party to act as validation. Checks must also be carried out to assess the provider's physical and logical access points and security controls. Any known risks identified as part of enterprise wide risk assessments should be shared by the provider as part of the service level agreement (SLA) and, the provider's insurance should cover potential losses for the organisation if the provider fails in its delivery (service outages etc). This chapter also discusses the importance of protecting against rouge actions by administrative accounts and insiders that can exploit cloud vulnerabilities.

SME Case Study Evaluation

This SME provides financial advice and monitors client credit score. It has been operating for a little over three years and employs 200 people. Employees perform credit history checks for clients and advise clients on ways to improve their credit score. Employees work on individual cases autonomously and systems often have two or three person controls for approvals, validation checks and authentication. As the company has grown substantially in the last two years they are considering outsourcing to a business partner to provide cloud computing that reduces the use of on-site servers.

Analysis, Chapter 16
For the SME in this scenario the recommendations presented in the guide would be beneficial to identify the data and services that will operate on the cloud and its associated risks prior to signing an agreement with a cloud provider. Ensuring that the provider has similar or better security practices in place than the SME would increase the robustness of defences for critical assets. For providers to share their hiring procedures (which include background security investigations) is unlikely and would be at the discretion of the provider. It is also worth considering that the cloud provider might have procedures in place such as background investigations which might not be enforced in practice. Identifying the party responsible for the technological and physical defences to critical organisational assets prior to entering an agreement can avoid any assumptive pitfalls.

Chapter 17: Institutionalize system change controls

A single recommendation is made in this chapter which is as follows:

- Periodically review configuration baselines against actual production systems and determine if any discrepancies were approved. If the changes were not approved, verify a business need for the change.

– pg. 97, Theis et al., 2019

This chapter highlights the importance of controlling changes to systems through system administrators. Measures should be implemented which prevent system administrators from inserting changes to the state of the system should they choose to act maliciously (for e.g. logic bombs, keystroke loggers, back doors for access and malicious code). Measures can include change controls that ensure accuracy, integrity, authorization and documentation of all changes for instance through recorded system logs.

<u>SME Case Study Evaluation: N/A</u>

Analysis, Chapter 17
As this chapter puts forward a single recommendation creating a SME case study was unnecessary. It would be in line with best practices to review systems' baseline configurations periodically against actual configurations of the system in use. If changes to initial specifications of the system exist, then must be approved by senior stakeholders. Additionally, employees should be able to put forward a case for change to senior management should this be necessary to optimise system performance. However, such

processes might be time consuming and introduce unnecessary red tape for SMEs that can hinder their operational agility.

Chapter 18: Implement secure backup and recovery processes

Two recommendations are presented in this chapter:

- Store backup media off-site. Ensure media is protected from unauthorized access and can only be retrieved by a small number of individuals. Utilize a professional off-site storage facility; do not simply send backup media home with employees. Encrypt the backup media and manage the encryption keys to ensure backup and recovery are possible.
- Ensure that configurations of network infrastructure devices (e.g., routers, switches, and firewalls) are part of your organization's backup and recovery plan as well as the configuration management plan.

– pg. 102, Theis et al., 2019

This chapter discusses system resilience in organisations through implementing and testing secure backup and recovery processes. This secure backup and recovery of data is especially needed by an organisation post a cyberbreach. Measures outlined in this guide include controlled access to storage facilities and physical media, separation of duties between personnel, two-person authentication for access to systems and, separate backup and recovery personnel. It is important to note that the guide recommends that people who have access to online copies of backed up data should be separate to those that have access to the physical copies. Mitigation strategies should include scenarios where risks to trusted business partners, such as cloud providers, are realised. Backup data should be encrypted, service licence agreements (SLAs) should state recovery periods and, name personnel with access to organisational data and critical assets. The guide recommends the use of off-site servers to backup data that is protected against corruption and destruction by insiders. Separate communication channels are recommended as a beneficial investment in the event of an attack as insiders are aware of internal communication methods.

SME Case Study Evaluation

This SME provides logistical assistance to the national electricity grid and employs a team of fifty people. In case of outage, dispatched engineers to the affected area and those at the office work together to identify the source of electrical outage. This task entails physical inspection by engineers for miles of remote areas in adverse climatic conditions to trace power supply lines and possible sources of disruption. The software used to map terrains, dispatch and track the movement of engineers, track progress of cases and, information about clients including payment details are all identified as critical assets to this SME.

Analysis, Chapter 18
It might be difficult for this organisation to make a business case for off-site servers and financially invest in storage facilities, especially if it has competing interests such as utilising the money to hire additional engineers to grow business operations. However it might be a quick win to implement the second recommendation to ensure that system configuration details for all devices are included as part of the backup and recovery plans.

Chapter 19: Close the doors to unauthorized data exfiltration

Recommendations in this chapter are as follows:

- Establish a cloud computing policy. Organizations must be aware of cloud computing services and how employees may use them to exfiltrate data. Restrict and/or monitor what employees put into the cloud.
- Monitor the use of printers, copiers, scanners, and fax machines. Where possible, review audit logs from these devices to discover and address any anomalies.
- Create a data transfer policy and procedure to allow sensitive company information to be removed from organizational systems only in a controlled way.
- Establish a removable media policy and implement technologies to enforce it.
- Restrict data transfer protocols, such as FTP, SFTP, or SCP, to employees with a justifiable business need, and carefully monitor their use.

– pg. 107, Theis et al., 2019

With the advent of industry 4.0 and technological capabilities to transfer data, this chapter focuses on the understanding how and through which channels data leaves an organisation. The chapter states that an organisation must be able to account for any points of exfiltration of data that include physical and wireless connections. Beyond controlling exit points, organisations must also regularly perform audits of media. Implemented controls should not hamper the delivery of an organisation's mission. This can be done through deploying solutions such as: i) that require two person approval; (ii) restricting removeable devices such as USBs or using a shadowing software that notes the file names and content that is transferred; (iii) barring unauthorised devices from accessing data; (iv) exclusively allowing company owned devices to transfer data; (v) encrypting files; (vi) restricting the movement of data through emails and data loss prevention systems; (vii) removing connectivity from systems (such as code writing environments); (viii) using jump boxes that segregate access for performing administrative tasks; (ix) restricting access to software solutions that are outside the organisation’s protective environment through blacklisting websites; and, (x) monitoring miscellaneous devices on the network such as fax machines, scanners, copiers and printers.

SME Case Study Evaluation: N/A

Analysis, Chapter 19

Through utilising any of the case studies shared above the recommendations presented in this chapter of the guide can be evaluated. It would be beneficial to create a cloud computing policy and to be aware of the various solutions that employees use to perform their respective tasks. However, aside from the time that will be invested in creating such a list, restricting the use of platforms might hinder employee abilities to efficiently perform work, device innovative ways of performing duties and, lead to a safety climate where employees might choose not to disclose the platforms they are utilising due to fear of reprimands. This can lead to more severe consequences for the security of SMEs where employees might utilise personal devices that are outside the organisation’s protected environment and encourage risk taking behaviours. Although cumbersome, monitoring the use of miscellaneous devices might be beneficial to investigate anomalies that can be indicative of potential insider threat. Creating and implementing a policy to remove sensitive information in a controlled way might create additional steps in performing tasks but can equally provide additional protection of critical assets. Implementing a removable media policy and software that restrict such actions might hinder internal collaborations and sharing of information for critical tasks. For instance, in order for people to collaborate they would need to share progress or information with internal or external audiences. This effort can be hampered if access controls are determined by job designations or at departmental levels, if email rules are configured that so not allow attachments or, if remote access is not supported. Restricting changes to systems through administrative accounts might be beneficial to avoid systems being compromised. However, restricting any type of data transfers throughout the organisation can be problematic for collaborative tasks. Instilling a zero-tolerance policy might result in employees using personal devices to bypass the rules, not specifically to act maliciously against the company but, to perform their assigned tasks and achieve their performance objectives. This might increase the probability of unintentional insider threat to the organisation as people might unwittingly adopt bad practices. Instead, it might be more beneficial to invest time in understanding various job functions and the tools that are seen as essential by employees to deliver their tasks efficiently in order to propose secure alternatives instead of completely restricting the effective use of technologies.

Chapter 20: Develop a comprehensive employee termination procedure

The recommendations presented in this chapter are as follows:

- Develop an enterprise-wide checklist to use when someone separates from the organization.
- Establish a process for tracking all accounts assigned to each employee.
- Reaffirm all nondisclosure and IP agreements as part of the termination process.

- Notify all employees about any employee’s departure, where permissible and appropriate.
- Archive and block access to all accounts associated with a departed employee.
- Collect all of a departing employee’s company-owned equipment before the employee leaves the organization.

– pg. 112, Theis et al., 2019

The discussion in this chapter highlights the importance of implementing best practices when employees exit organisations. This includes developing and implementing procedures such as checklists, timely action to terminate access, archiving of accounts and, exchange of organisational equipment including mobile devices and access cards. Checklists should contain the names of individuals who are responsible for various components of the checklist and for verifying completed actions. Once the checklist is complete it should be filed with the HR department prior to the employee concluding their last day of employment. Unreturned organisational equipment can be used to carry out insider attacks once an employee has left the organisation. Recommendations suggest that organisations should conduct a review of the individual’s actions through system logs over a 30-day, or 90-day period if such data is available.

SME Case Study Evaluation: N/A

Analysis, Chapter 20
A case study for this chapter was not developed as the recommendations were in keeping with best practices. Developing an organisation wide checklist for employees leaving the organisation, tracking all accounts that can provide points of access to the system for all

employees and emphasising nondisclosure and other contractual agreements to employees leaving the company might reduce the risk of malicious insider threat and instil best practices across the organisation. The organisation should communicate the departure of employees to their colleagues (where permissible), terminate access in a timely fashion, archive accounts and collect company-owned equipment prior to the employee’s departure.

SMEs tend to experience pipeline issues with attracting high calibre skilled applicants. In such cases, notice periods might be longer than the 30-day period and more likely to be around the 90-day period. Prior to the start of the notice period (assuming a 90-day window) the employee would have consciously made the decision to leave the organisation, applied to other jobs (possibly with competitors), attended interviews (possibly on multiple occasions with the same organisation and on other occasions with multiple organisations), accepted an offer before giving their 90-day notice to their organisation. This means that the risk associated to potential insider threats would be at the highest prior to the notice period due to the opportunities that exist to go unnoticed. Reviewing system logs for a 90-day period might be insufficient to detect malicious insider threat.

Chapter 21: Adopt positive incentives to align the workforce with the organization

The final set of recommendations presented in this guide are as follows:

- Organizational justice (fairness; e.g., compensation aligned internally among employees and externally with industry standards)
- Performance-based rewards and recognition (e.g., transparent criteria for promotions and discretionary rewards/recognition based on project performance)

- Transparent and respectful communication (e.g., regular employee orientation, mentoring, and expectation setting)
- Personal and professional supportiveness (e.g., employee assistance programs and professional development for furthering employee careers and sense of mastery)

– pg. 118, Theis et al., 2019

This chapter discusses the importance of implementing positive incentives for employees to adopt best practices and act in the interest of the organisation. It notes that whilst negative incentives can be favoured by management, such as enforcement of rules, forcing compliance, punishment and, reprimands, this approach is not effective against insider threat. Positive incentives can include developing job engagement, developing individual strength areas and internally promoting employees. The guide states that there is perceived organisational support if employees feel their wellbeing is important, their work is valued and, their organisation has a just culture. Positive incentives also include ‘connectedness at work’ where employees feel connected to each other. This chapter discusses the importance of striking a balance between positive incentives and traditional security measures that rely on restrictions and, rules to prevent and punish abuse derived from the Deterrence Theory. This chapter appears to be at odds with the recommendations proposed by this guide in earlier chapters that discuss enforcement of rules and policies, restricting and controlling employee actions and punishment for those who bypass the organisation’s secure IT environment. Nonetheless, this chapter acknowledges the challenges of establishing a safety culture at organisations that can entail major shifts in attitudes, perceptions and understandings of punishments and rewards as well as establishing a just culture.

SME Case Study Evaluation: N/A

Analysis, Chapter 21
The recommendations made in this chapter reflect work cultures and incentives offered by SMEs and thus a case study was not needed to demonstrate the applicability of recommendations. This includes a just culture that involves fairness and inclusiveness, performance based rewards and recognition, transparency in communications and, support from the organisation and colleagues for personal and professional needs. Instilling these values in an organisation might be invaluable but would require a substantial amount of time, effort and understanding at every level of the organisation.

Following the critical evaluation of recommendations through case studies above, this work progresses to identify the party responsibility for the design and implementation of these 79 recommendations. Responsible parties that carry the onus for preventing insider threat are identified by extending the application of the onion model.

3.3 Classification through the onion model

This section evaluates the extent to which the guide's recommendations are holistic in their nature by applying a human centric approach and subsequently demonstrating the need for the human factors domain to engage with challenges posed by insider threat. This is achieved through classifying recommendations to the onion model (Wilson and Sharples, 2015).

The onion model comprises of eight categories that form a complete human-environment system. Various facets represented in the onion model are interdependent and interconnected whilst being responsive to change. The guide's recommendations are redistributed according to the eight categories relevant to designing work according to this model. These categories are as follows: i) people; ii) artefacts; iii) technologies; iv) tasks and goals; v) personal physical and virtual workspace; vi) wider physical and virtual work environment; vii) work and organisational context and, viii) financial constraints and priorities, technical developments and capabilities, legal and regulatory framework and, social influences, expectations and norms.

Whilst it can be argued that the recommendations could be classified in numerous ways, through maintaining a grounded approach which provides context to the document being analysed, these recommendations were categorised by the category that was being enabled through the recommendation. This meant that recommendations that might obviously appear to be related to a specific category did not necessarily enable the category they appeared to empower.

For instance, recommendation in chapter 3 "Ensure that management makes policies for all departments within the organisation easily accessible to all employees…" can appear to be pertaining to the People, Technologies or Artifacts categories but this recommendation pertains to the Work and organisational context category whereby the organisation would have an accessible and/or inclusive culture (Neal and Griffin, 2004). Similarly, recommendation in chapter 4 "Encourage employees to report suspicious behavior to appropriate personnel for further investigation" can appear as though it pertains to the People category but in fact this would be dependent on and greatly influenced by the larger societal culture, norms and expectations of the region the organisation exists within. Another example is the recommendation for chapter 4 "Consider offering an EAP. These programs can help employees deal with many personal issues confidentially" can appear to be pertaining to the People category but in fact forms part of the Personal physical and virtual workspace category as it is an external identity to the individual that exists within their wider virtual workspace.

The redistribution of CERT's recommendations onto the onion model are presented in Figure 5 below and as a breakdown table in Appendix 4.

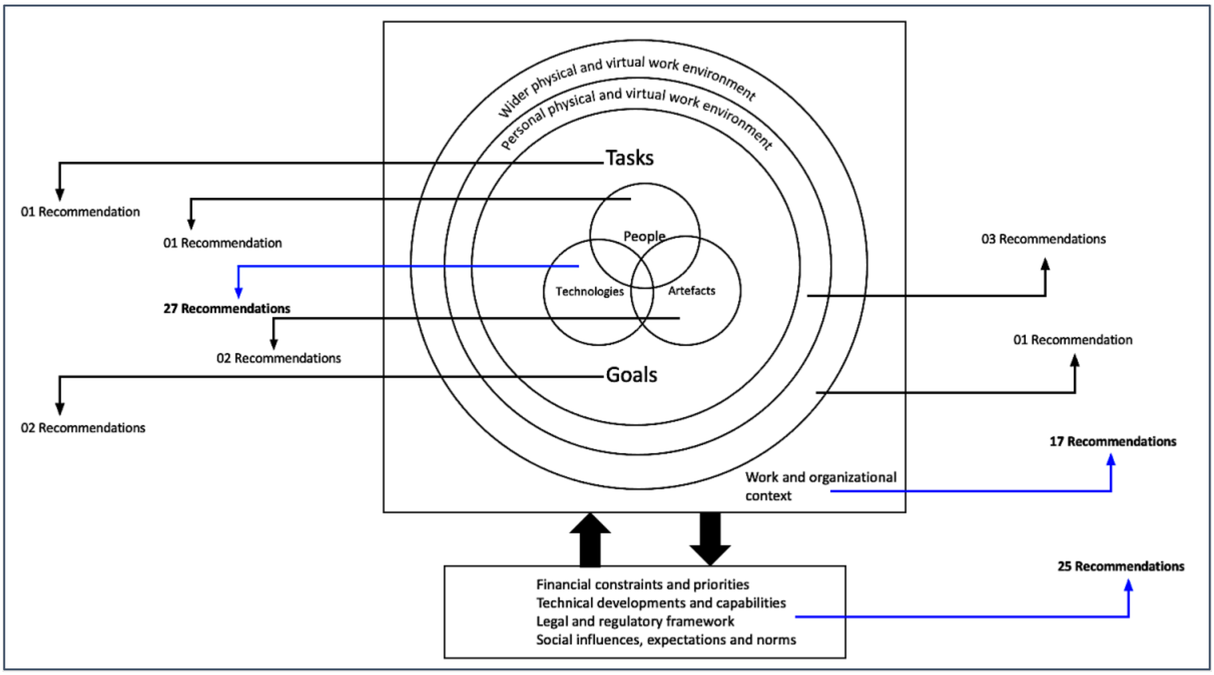

Wider physical and virtual work environment
Personal physical and virtual work environment
Tasks
People
Technologies
Artefacts
Goals
Work and organizational context
Financial constraints and priorities
Technical developments and capabilities
Legal and regulatory framework
Social influences, expectations and norms
01 Recommendation
01 Recommendation
27 Recommendations
02 Recommendations
02 Recommendations
03 Recommendations
01 Recommendation
17 Recommendations
25 Recommendations

Figure 5: Classifying recommendations onto the onion model

Category 1: People

Out of the 79 recommendations proposed by CERT's guide, 'Common Sense Guide to Mitigating Insider Threats, Sixth Edition', only one recommendation was classified as belonging to this category. Found in chapter 8, this recommendation suggests allowing people to be able to consider and choose work projects, actions and, statements prior to committing to them.

Category 2: Artefacts

Only two recommendations involved artefacts used in the workplace to mitigate insider threat. The first recommendation suggests controlling and prohibiting personally owned devices at work. The second proposes good practice procedures i.e. collection of equipment when employees depart from an organisation.

Category 3: Technologies

Out of the 79 recommendations, 27 recommendations were classified as those involving the use of technologies to mitigate insider threat. These recommendations are found in a majority of the chapters throughout the guide. Recommendations primarily propose using technologies to control, restrict and, monitor assets and human interaction with these assets in order for organisations to mitigate insider threat. This is proposed via a range of technical suggestions which include correct device configuration, controlling access to data and the organisational networks (including VPN white lists), setting up SIEM alerts for anomalies (including alerts for firewalls and IDS that are in place), system log management and retention, disabling access in a timely fashion when applicable, using monitoring tools to oversee network and employee activity, using approved ports and protocols for network activity, controlling or eliminating remote administrative access to cloud or virtual service providers, monitoring the use of printers, copiers, scanners etc and, enforcing restrictions on abilities to transfer or remove assets/data. Other suggestions within this category included controlling the distribution of sensitive documents and assets, controlling the creation and management of user accounts (including privilege creep), implementing password policies that encourage strong passwords, auditing user access permissions and, allowing a restricted number of users the ability to access to backed-up data.

Category 4: (a) Tasks (b) Goals

A total of three recommendations were classified as relating to this category. One recommendation involves tasks through the suggestion of developing appropriate metrics to quantify employee performances. Two recommendations, both found in chapter 8, are linked to goals. It is suggested that employees focus on one task at a time (as opposed to multitasking) and, organisations should monitor employee workloads against existing skill level and the resources available to them for performing tasks.

Category 5: Personal physical and virtual workspace

Only three recommendations were identified as belonging to this category. These recommendations involved factors that would influence the personal physical and virtual workspace of employees. Recommendations included multiple accounts being created for privileged users to perform relevant tasks, restricting and monitoring what employees can upload and download from the cloud and, organisations offering employee assistance programme (EAP).

Category 6: Wider physical and virtual work environment

Out of the 79 recommendations only one recommendation was classified as pertaining to this category of wider physical and virtual work environment. It recommended forbidding devices with cameras from sensitive areas within organisations.

Category 7: Work and organisational context

17 recommendations across ten chapters were identified as part of this category. Recommendations include organisations being responsible for identifying their assets (physical and virtual), classifying these assets according to their type and, prioritising these assets to determine relative security measures. Recommendations also include synchronising efforts from various departments for the deign and enforcement of policies and procedures

(including accessibility of these documents and the development and signing of NDAs), conducting background checks of employees prior to hiring and, investigating and documenting suspicious or disruptive behaviour. Organisations are also responsible for transparency in their communications, providing awareness training for the use of social media and security and, providing additional benefits to reduce stress levels amongst employees. The fourth grouping of recommendations within a work and organisational context involved the development of lists. This included listing mobile devices as part of organisation-wide risk assessments, listing all known accounts assigned to each employee for access to the system and, checklists for employees leaving the organisation.

Category 8: Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

Recommendations that involved financial, technical, legal and social factors were classified to this category. These factors were seen to influence the ability to develop, implement, execute or enforce recommendations made by CERT's guide to mitigate insider threat. 25 recommendations were identified from across 11 chapters. Recommendations appeared to place the onus for these factors on organisations. Within the recommendations it was the organisation's responsibility to seek legal counsel, develop specific contractual agreements with employees and third parties, perform background checks on all employees including those hired by business partners, develop policies, enforce policies, provide organisation-wide staff trainings and, instil a just culture. Recommendations within this category also included reporting on peers, reporting suspicious external communications, providing channels for anonymous reporting, increasing the surveillance of high risk employees, providing performance based rewards and, fostering a culture that is supportive of employee growth.

This recategorization demonstrated the emphasis by CERT's 'Common Sense Guide to Mitigating Insider Threats, Sixth Edition' on specific categories that are assumed to be the most effective in safeguarding against insider threat i.e. technologies, external factors and organisational contexts. As shown in Figure 5, recommendations pertaining to these three categories were disproportionately higher than all other categories combined. This recategorization is indicative of two things: firstly, there is trust placed in specific aspects that might outweigh the importance of other elements that are found in sociotechnical systems to safeguard against insider threat; and secondly, there is undue responsibility placed on certain aspects of sociotechnical systems in the context of insider threat, an approach that is not holistic in its nature.

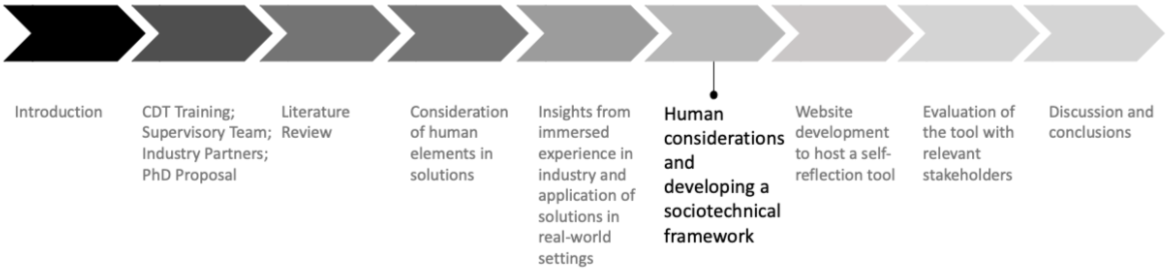
3.4 Summary

In this Chapter a guide by CERT (2019) titled 'Common Sense Guide to Mitigating Insider Threats, Sixth Edition' was critically evaluated. This was done through document analysis method to interpret recommendations. Through this exercise an understanding was developed for evaluating the real-world application of these recommendations and discuss the potential challenges that might emerge. To further this understanding, all 79 recommendations were classified according to the categories presented in the *onion model*. The transference of recommendations showcased the emphasis that is put by prominent approaches on certain elements (such as technological and external factors), perhaps unconsciously, within a sociotechnical system to prevent insider threat. The classification of recommendations to the onion model evidenced: i) solutions that propose to tackle both elements of intentional and unintentional insider threat have reduced applicability for unintentional insider threat; ii) there is a skewed division of responsibility for safeguarding against insider threat that limits approaches driven from traditional security thought from being holistic in their propositions;

and, iii) there is a need to change how humans are considered in systems when unintentional insider threat is being evaluated in order to propose new solutions to this challenge. The next Chapter discusses the design of the first research study that utilises a decision making technique to exclusively explore factors that influence unintentional insider threat.

4. Approaching unintentional insider threat with a human centric lens

4. Approaching unintentional insider threat with a human centric lens


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

The previous Chapter established a need for enhancing solutions to specifically address unintentional insider threat which are derived from a human-centric approach in order to avoid emphasising certain elements in a sociotechnical system in a bid to move away from traditional security thought.

By virtue of this knowledge above, a research study was designed to explore factors that exclusively influence unintentional insider threat from those who have experienced it through the application of Critical Decision Method (CDM) technique. Interview questions were designed to cover various aspects of the incident that would reflect the environment in which the breach occurred and subsequently avoid placing unequal onus on certain elements of a sociotechnical system. This balanced approach when designing interview questions was intended to aid in extracting hidden factors that might have been neglected or overlooked in existing solutions or to provide further evidence for existing approaches discussed in the literature review. Furthermore, since equal attention was paid to all aspects within a sociotechnical system (represented as elements in the onion model) in this study it was also believed that the subsequent findings would also contribute to creating a well-rounded framework that can aid in changing the way humans are considered in systems.

To encourage a rich discussion around naturalistic decision making (NDM) when engaging with activities that resulted in cyber breaches, Critical Decision Method for Eliciting Knowledge (CDM) (Klein et al., 1989) was chosen. CDM was particularly fitting for the design of interview questions as it retrospectively focuses on a major event with probing follow-up questions to guide discussions. These probing questions assist in eliciting expert knowledge about how decision making occurs in cognitive tasks. It is important to note that whilst individuals are not experts at falling for cyberattacks, the interest is that breaches happen in the context of expert behaviour at work or in personal lives. CDM has been widely used across various domains to help analyse decisions (Hoffman et al., 1998) and, to inform system development and design. While CDM is normally used in homogenous samples (different individuals performing the same task in the same environment) this method was applied in a novel way as all our participants performed various jobs and worked on assorted levels in organisations for different periods of time with their employers. However, they all made critical decisions in complex work or personal contexts that led to all of them experiencing the major event of a cyberbreach. All ten participants were over the age of 18 years old and eight were residents of the East Midlands and Greater London areas and two were based internationally. Further information about participants is presented in Figure 6 and discussed in Section 4.2.

4.1 Methods

Critical Decision Method for Eliciting Knowledge (CDM) begins with a general question about the incident, in this study it was the cyberbreach, to construct an initial picture of the incident from the participant. CDM then provides probing questions based on the information shared by the participant. The use of CDM to explore cyberbreaches was valuable as it allowed participant and the researcher to journey into an introspective in-depth examination of the incident. It also allowed conversation to flow naturally and provide overall consistency across discussions. Inclusion criteria consisted of participants having experienced one of the three scenarios in their personal or professional lives: i) They had accidentally sent sensitive information to the wrong recipient; ii) They had accidentally clicked a link that resulted in phishing, ransomware or gave someone access to their private information; or iii) They had clicked a link by mistake that gave someone access to their email account, social media account, bank account or personal device such as a laptop or mobile phone.

Whilst there are numerous types of unintentional insider threats (UIT) that exist, the poster designed to advertise the study included three examples that people had experienced in their personal or professional lives. While to an expert the three examples can comprise of two types of unintentional insider threat (and possible a variation of a third type), it was important for people who are less technically advanced or those that belong from a lay audience category to be able to identify and relate to the example presented. For the lay audience to understand these scenarios was also important as the study would retrospectively focus on an important event in their lives which could potentially make them re-experience feelings such as embarrassment or frustration that they felt at the time of the incident. This ability to understand the topic of the research study was also central in the participants feeling comfortable with sharing experiences which are often a taboo topic for discussions due to the stigma associated to breaches. While this thesis is focused on UIT in organisational settings individuals who had experienced UIT in their personal lives were also invited to partake as the focus of the research study was to explore factors that influence UIT, the event would have been memorable, would have required them to make critical decisions that resulted in a breach and, the breaches would have occurred in routine activities or tasks in which participants would be engaging in expert behaviour.

Structured approach and emergent themes approach are usually adopted for data analysis to compliment CDM (Wong, 2004). Structured approach assumes a pre-existing framework within which data is coded and the emergent themes approach focuses on the relationships between concepts. Presumptions about data was a particularly significant factor in the selection of methods so as not to make any assumptions prior to analysis. Consequently, in this study, a grounded theory (GT) approach was applied. Originating from sociology (Glaser and Strauss, 1967) GT has since become a widely adopted method by researchers (Muller and Kogan, 2010). Key points in the data were identified and assigned codes, known as open coding. Codes where then compared against each other in the same interview and across interview transcripts, known as the constant comparison method (Hoda et al., 2010). This was done until data reached saturation before commencing analysis.

4.2 Participants

Following ethical approval ten participants were recruited, eight from East Midlands and Greater London areas and two based internationally. All participants received a brief description of the study design (Appendix 5), participant information sheet (Appendix 6) and, a consent form (Appendix 7). The consent form was signed and returned prior to the interview and participants were given the opportunity to ask questions via email or verbally before their session began.

The inclusion criteria for participants was that they had to be over 18 years of age, have access to the internet and, had experienced one of the three scenarios in their personal or professional lives: i) They had accidentally sent sensitive information to the wrong recipient; ii) They had accidentally clicked a link that resulted in phishing, ransomware or gave someone access to their private information; or iii) They had clicked a link by mistake that gave someone access to their email account, social media account, bank account or personal device such as a laptop or mobile phone. Participants' scenario of breach, settings (personal or professional) and, their occupation at the time the breach are shown in Figure 6 below.

Participant	Scenario	Setting	Field (occupation)
P1	Accidently engaged with content that resulted in being hacked	Professional	Higher education (researcher)
P2	Accidently sent sensitive information to the wrong recipient	Professional	Healthcare (grants manager)
P3	Accidently sent sensitive information to the wrong recipient	Professional	Charity (grants manager)
P4	Accidently engaged with content that resulted in being hacked	Professional	Higher education (researcher)
P5	Accidently sent sensitive information to the wrong recipient	Professional	Charity (grants manager)
P6	Accidently sent sensitive information to the wrong recipient	Professional	Charity (international partnerships)
P7	Accidently engaged with content that resulted in being hacked	Personal	NA (student)

P8	Accidently sent sensitive information to the wrong recipient	Professional	Think Tank (internee)
P9	Accidently sent sensitive information to the wrong recipient	Professional	Food Retail (lawyer)
P10	Accidently engaged with content that resulted in being hacked	Professional	Charity (accountant)

Figure 6: Participants' scenarios of breach, settings and occupation at the time the breach

Participants had a varying degree of experience ranging from mid-level to advanced, shown in Figure 7 below (*Analogues* are when people are reminded of similar lived experiences in the past to aid in decision making and is discussed in section 5.8.2 of this Chapter). Participants were not offered any compensation for sharing their experiences as part of this study and were provided associated materials describing the motives of the study prior to recruitment. As a result of snowball sampling some of the participants were known to the author in a professional context. Consent forms were completed and participants were given the opportunity to ask any questions prior to commencing any discussions.

Participant	Cyberbreach Setting	Novice	Mid-level	Advanced	Analogue?
P1	Professional		x		Absent but context present
P2	Professional		x		Absent
P3	Professional			x	Present
P4	Professional		x		Absent

P5	Professional		x	Absent
P6	Professional		x	Absent
P7	Personal	x		Absent but context present
P8	Professional	x		Absent
P9	Professional		x	Present
P10	Professional	x		Absent

Figure 7: Participants' experience level and the presence of analogues

4.3 Data gathering process

Data were collected between March to April 2020. Discussions were held individually with participants and audio and video recorded, generating approximately 9.5 hrs of dialogue. Due to the enforcement of lockdown rules during the Covid-19 outbreak in the UK, initially agreed discussions were rescheduled to be held online through platforms that were most familiar to participants (Skype or Microsoft Teams) at a date and time convenient for them. The full set of critical decision method based questions used in interviews are listed in Appendix 8. The author of this thesis carried out discussions with participants and transcribed them verbatim from the digital recordings. The author also analysed and interpreted the data.

4.4 Analysis

Transcripts from digital recordings were produced, anonymised and uploaded to QSR-NVivo software for coding. Transcripts were highlighted with colour ink and descriptively labelled based on open coding and constant comparison method. Once labelling was exhausted to the point where no new labels could be generated, a three-stage analysis of the data was conducted, shown in Figure 8 below.

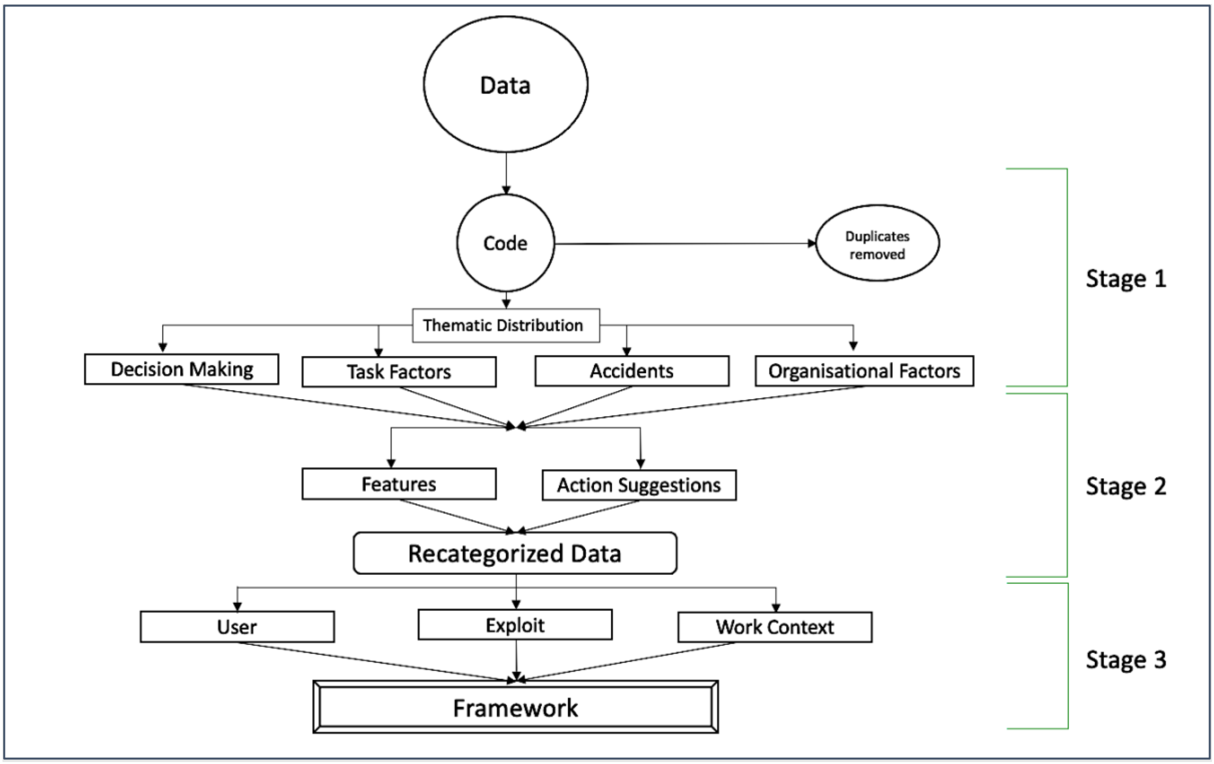


Figure 8: Three stages of data analysis process

In stage 1, labels were organised thematically as ‘Decision Making’, ‘Task Factors’, ‘Accidents’ or ‘Organisational Factors’. This thematic categorisation produced results that offered a better understanding of factors influencing unintentional insider threat (UIT). Stage 2 involved reorganizing codes as either features or actions. The results from Stage 2 were used to recategorize data according to the Epidemiological Triangle (Cassel, 1976) fields of ‘user’, ‘exploit’ or ‘work context’ as part of Stage 3. From this recategorized data a framework was developed to list ‘inputs’ and ‘outputs’ that can be used by organisations to identify, intervene and mitigate against UIT. For the purposes of this work a framework is a set of recommendations applicable in specific scenarios to reduce negative impact.

The above writing shared the methods, nature of participants, data gathering process and, the subsequent analysis of data for a study that explored factors that influence unintentional insider threat. Critical Decision Method technique was implemented to discover factors that might have previously been overlooked in other research studies that did not deploy a human factors based approach. Additionally, following the classification of recommendations onto the onion model, effort was made to design interview questions in a way that would consider all elements within a sociotechnical system in order to avoid putting the onus on specific aspects. Continuing with the same approach, the framework was developed in a way that emphasised all aspects within sociotechnical systems with distributed responsibility to enable safe cyberspace operations.

This writing now moves on to thematically discuss the findings that emerged from this study. Four themes were identified in the analysis as factors that influence unintentional insider threat, presented in the following sections as: ‘Decision Making’, ‘Task Factors’, ‘Accidents’ and ‘Organisational Factors’. Numerous codes that emerged from the data and their respective code frequencies are listed in Appendices 9 and 10. Code frequencies allowed the author to cross check the findings’ weightings. Quotes from participants are identified as: P1 for the first participant, P2 for the second participant and so on until P10. All participants are referred to as ‘she’ and ‘her’ regardless of their gender identification to maintain anonymity.

Each of the four themes i.e. Decision Making, Task Factors, Accidents and, Organisational Factors informed the sociotechnical framework. These inputs and their corresponding pillar are presented in their respective themes. Some inputs from the framework appear in two themes as data from multiple themes evidenced the same input. These inputs and their corresponding pillars are presented in order of how they appear in the framework to maintain the structure later on in this Chapter.

Limitations from this research study’s findings emerge from snowball sampling that can limit the diversity of participants and the generalisability of findings. As some participants were known to the interviewer which enabled candid and honest discussions, it might have influenced participants’ responses to some degree. As participants were asked to recall an incident in the past, whilst this incident was significant in their lived experiences, memory recall and memory bias might have unintentionally included or excluded information that could have been significant for the findings. In addition, as the study advertised for specific types of unintentional insider threat in specific contexts, which aided in recruiting lay audience participants who might not have otherwise understood the nature of the breach being investigated, it can limit the findings applicability to the broader category of unintentional insider threats that exist. Study advertisements also entailed an *action* towards a goal that resulted in an event which can be discussed as part of the Critical Decision Method technique. This could have subsequently excluded people who were victims of other types of unintentional insider threat.

4.5 Theme 1: Decision making

All participants were asked to reflect on how their lived experience and acquired knowledge affected their decision making (DM) when interacting with technologies. This was interesting as it gave an insight on how individuals might make informed decisions when identifying between malicious and non-malicious content that included utilising cues and applying knowledge-based behaviour. Cues included surface features (logo disparities or brand colours) and contextual features (typically around pre-existing expectations around who respondents would expect to hear from and the nature of their likely requests) which will be discussed below. This theme was divided into two sub-categories: lived experience and acquired knowledge. The 'lived experience' category comprises of direct personal experiences which might contribute to 'acquired knowledge'. Acquired knowledge encompasses all channels used to build knowledge which can include lived experiences as well as other channels such formal classroom teaching, word-of-mouth or, awareness campaigns.

The following findings informed the following inputs in the sociotechnical framework for Pillar 1: User Vulnerabilities to UIT and recommendations to strengthen defences:

Assess how comfortable individuals are with various technologies and platforms	Pillar 1
Assess how vulnerable users feel in their daily online interactions	Pillar 1
Assess individuals' ability to identify spear phishing scams to note vulnerabilities	Pillar 1
Assess individuals' existing experiences with malware or threats (including physical spaces)	Pillar 1
Assess individuals' knowledge base to evaluate understanding of current techniques used by hackers	Pillar 1
Assess individuals' susceptibility to spear phishing	Pillar 1
Assess the levels of how much individuals rely on their social networks (offline and online) to inform their decisions if faced with threats	Pillar 1
Assess individuals' awareness of mainstream marketing campaigns against popular attacks	Pillar 1
Assess levels of retention from basic ICT teachings to establish levels of awareness	Pillar 1
Assess and map different skill levels between individuals in a diverse workforce	Pillar 1

4.5.1 Lived experience

All our participants shared a sense of reliance on their lived experiences which formed tacit knowledge to help them differentiate between genuine and malicious content, each participant with their own set of techniques and strategies to serve as defences. To stay a step ahead of the hackers P8 expressed the importance of paying attention and reading in between the lines.

‘*But if you look at the email [ID/Domain], it's usually some crazy weird email that they've just created. And they're trying to hide behind all of the email itself. Going on, you can also kind of tell that there'll be certain slight differences between the logos and stuff and the branding that they use. They might use an old, outdated one.*’ (P8)

Techniques included clues from technical elements of the interaction such as the legitimacy of embedded web links, sender domain, font and logos. Participants also mentioned deviation in language and errors in the main body text to identify malicious content.

Apart from the techniques used to evaluate technical elements of emails, participants also shared techniques that assisted them in making snap judgements about whether something was malicious. P6, who works for a charity organisation conducting extensive work overseas with a range of different partners, shared how she applies lived experiences to quickly validate the legitimacy of online interactions.

'*When it's from an address I do know, you know, that I recognise, if it's in a way that the person wouldn't normally sort of writes or interact with me, then that's big red flags there …. If it's like something that's general but specific, like, "Oh, that thing we did last?" but it doesn't give you the information of what exactly it is, it just points to something really quite general. Firstly, if that doesn't match up with anything in my experience with that person, then that's a no-no. But then even if it does, then you put on that, sort of, other lens of 'Okay, does this look credible?*" (P6)

Techniques also included reference to an incident that didn't occur or if something was too good to be true which reflected a mixture of Type 1 (intuitive thinking which is rapid and autonomous) and Type 2 (reflective thinking which requires working memory and other resources) thinking processes (Evans, 2012) at different stages in their interactions. Having specific context to a conversation, the nature of the request from the sender and relying on a strategy similar to two-factor authentication (2FA) from establishments such as their banks also assisted participants in their DM. This is highlighted in the example below with P4, who is a researcher in the higher education sector and experienced a cyber breach by interacting with malicious content that appeared to be from her supervisor. In this part of the discussion

P4 shared how the lack of multiple channels to communicate a threat assisted her in making a snap judgement about the content being malicious.

‘*But if they were saying to me, “Oh, there's something wrong with your bank account, please log in here”, I would more likely be like, 'Oh, surely I would have received a text message' or, 'Surely I would have received a phone call'. So I'm more likely to contact the bank before I open stuff like that.*’ (P4)

While relying on a two-factor approach (2FA) to validate false alarms may appear to be a passive state, P4’s technique reflects that her lived experience, where no action was taken, has resulted in things continuing safely. Interestingly, this strategy might aid in counteracting malicious emails that use urgency or time pressure techniques to lure targets.

4.5.2 Acquired knowledge

Participants largely appeared to have more confidence in their abilities compared to older generations specifically based on how various platforms were utilised whilst participants acknowledged that this confidence could be misplaced. When speaking to P4 about how she felt her knowledge was different to other generations, P4 discussed the phenomena of fake news and how her position as a researcher allowed her to be untrusting of online interactions, which she did not feel was largely afforded to previous generations.

‘*Whereas I feel… they're[older generation] very much more different because I feel like this whole process of the internet was thrown at them without them getting taught, like, how to distinguish the truth from lies.*’ (P4)

Participants also considered informally acquired knowledge as more advantageous when safeguarding oneself against malicious content. For instance, P2 who experienced a cyber breach during an intense process of peer-reviewing grant applications in the healthcare sector, was pensive about the differences in knowledge between herself, her peers and other generations.

'*I think I feel a bit more confident about making these decisions, but that's possibly misplaced. Because my parents, being less skilled are also less likely to click anything because they're terrified of making a mistake.*' (P2)

Based on their personal and professional experiences all participants believed that they were comfortable with using technologies due to their exposure of having grown up with it. This familiarity with technologies also brought a heightened risk awareness amongst all our participants for susceptibility to being scammed themselves and difficulty in identifying sophisticated scams. However, this heightened risk awareness might have stemmed from all our participants experiencing a cyberbreach and not necessarily an attribute of comfort with technologies in general. Knowledge of cybersecurity was largely acquired through personal experiences, proactive online researching and, social networks (online and offline). Some participants also mentioned other avenues such as classroom or lab based teachings and, through posters, marketing leaflets and online bank app notifications that contributed to their knowledge but noted that these avenues were less effective.

While lived experience and acquired knowledge help guide decision making (DM) in our daily lives, the techniques deployed by individuals can subtly contribute towards indicators for assessing unintentional insider threat (UIT) risk levels. For instance, if individuals are aware of latest techniques used by attackers, are confident with the use of technologies deployed to perform daily tasks or, have internalised techniques to help identify malicious attempts it can provide a strengthening of defences (which will be discussed later in this Chapter as part of the framework). Equally, if individuals exhibit over-confidence or low levels of malicious content identification techniques it might be indicative of weaknesses in defences as end users would not possess the skillsets needed to make critical decisions in daily tasks that can result in UIT being realised.

4.6 Theme 2: Task factors

Broad themes that directly linked to task factors in the context of the incident of a cyber breach included complexity of the task, speed (of the incident, discovery and response) and, actions (to minimize impact, conclusion assumption and subsequent actions).

The following findings informed the following inputs in the sociotechnical framework for Pillar 1: User Vulnerabilities to UIT and recommendations to strengthen defences; Pillar 2: The effectiveness of processes and facilitating a continuous improvement culture; Pillar 4: Knowledge sharing and empowerment culture and; Pillar 5: Fluctuating vulnerabilities:

Assess individuals' susceptibility to rationalise abnormal behaviour or interactions	Pillar 1
Assess individuals' susceptibility to spear phishing	Pillar 1
Assess and map different skill levels between individuals in a diverse workforce	Pillar 1
Evaluate all tasks to identify missing feedback loops that indicate task completion	Pillar 1
Evaluate the effectiveness of prescribed processes amongst skilled/experienced staff	Pillar 2
Assess individuals' prioritization of processes	Pillar 2
Evaluate the effectiveness of prescribed processes amongst all designations	Pillar 2
Evaluate in-use software's limitations in prescribed processes	Pillar 2
Assess individuals' commitment to best practices set out by the company	Pillar 2
Evaluate processes for collaborative tasks that are automated	Pillar 2
Assess individuals' technical skill levels	Pillar 2
Evaluate individuals' ability to question, share and challenge abnormal interactions	Pillar 4

Assess individuals' level of attention to detail (online and physical parameter)	Pillar 5

4.6.1 Complexity of task

P2, a grant manager in healthcare industry, experienced a cyber breach while organising a peer review process. P2 described the review allocation task as being complex with many agile stages. A bulk of reviewers were approached as part of the initial assessment of grant applications. The responses triggered a cascade of subsequent actions for P2, for instance reviewers would respond at different rates and if someone had agreed to review, she would provide them with further details such as an outline of assessment process and deadlines. If someone declined, P2 would respond to them and contact another reviewer from a list. This process was recorded for each application on a spreadsheet for progress that was standardised by the employer. If the potential reviewer pool ran out, which came across as a likely scenario, P2 was responsible for finding new potential reviewers to add to the pool. The process of searching for potential reviewers had its own set of rules, such as the reviewer not having a conflict of interest or residing in a specific country. This led to many checks being performed prior to a reviewer name being added to the pool such as, not being a contributor to the application, not being affiliated to the same host institution or not being affiliated to the same lab group as the applicant. P2 described this stage as complex utilising various technologies to aid in the task such as online searches, using the application portal, emails, excel sheets and, PDFs. This entire process meant that while all application reviews started at the same time, the task was not synchronised and increased in its complexity as it progressed.

'*So you're not paying attention to drafting just one email, you're doing a lot of things at once. And it's that kind of almost like sweet spot for disaster of being repetitive and dull, but still complicated. To make sure that the right thing is going to the right person.*' (P2)

Almost all participants reported tasks as being complex with many preceding or simultaneous actions that needed to be performed in order to successfully complete the task at hand. When speaking to P6, who works for a charity with numerous overseas partners, she discussed how adopting a systematic approach allowed her some control over a complex task that is not synchronized as it progresses.

'*You've got a tonne of contracts to do. And you're scanning them and they're piling up on your desk, literally on the right-hand side and then you're trying to be systematic about the way how you're saving them. Putting the numbers there, the right codes there for the project reference numbers. Filing them in the right place and sending them on to the right people. But you're doing that at pace, probably…. You'd then go out to those people individually to ask them, you know how they wanted the contract to be set up… And so you're getting back those 45 different bits of information. And then I'd say that it's an iterative thing… And then you're trying to, one you're getting the replies in, your you're making the contracts, making them out to the right people, the right time, the right amount… But again, because there's 45 of them, they're all coming in and dribs and drabs and you're doing this process for each one.*' (P6)

All our participants described using a mixture of techniques to deliver their respective complex tasks. Techniques included using templates, manually maintaining progress spreadsheets, using pivot tables, doing email merges, performing manual checks for accuracy, searching the internet, collating information, using bespoke software platforms (CRMs) and, utilising various mass market software.

Beyond the use of technologies, participants frequently mentioned the human element that informed and influenced their task delivery. Similar to P2's case, P3 also reported reviewers feeding into the process. P3 worked for a charity dissipating grants on behalf of the government and had also experienced a cyberbreach by sending sensitive information to the wrong recipients at the peer reviewing stage. When talking about the complexity of the task P3 discussed how the human element was significant enough for her to mould the process in a way that made the human related aspects in the task run smoothly, which is reflected in the following quote.

'*So I think for everyone who agreed to review an application, we just sent all of them the PDF, as well as linking them on the [CRM] system. Because it was common enough that people would prefer the PDF.*' (P3)

Further to moulding processes to suit human elements and given that work is not conducted in a vacuum or in a linear way, all the tasks involved other people contributing in some way but almost all of the tasks involved other people actively feeding-in (internal and/or external) for the successful completion of the task. This collaboration with other people appeared to influence participants' actions for how the task was conducted.

The active feeding-in from other people for the delivery of the task also created new information that needed to be processed and managed by the operators before future steps were selected, adding another layer to the complexity of the task. P9 who is a lawyer working for an international food retailer in Asia experienced a cyberbreach where sensitive information was accidently sent to the wrong recipients. P6 collaborates with people within her country of residence (locally) and also internationally where her line manager is based in Europe. She is the data compliance officer as well as the legal representative for her employer and is often involved with contesting fines issued by the local government for alleged non-compliance of local food regulations. In order to contest the fines, P9 is responsible for filing a case in the local court if her manager and the higher management team approves. In the snippet below she describes the task that led to the cyberbreach.

‘*Then we had the calls from both of the store managers giving a bit of explanations as to how much we were fined and exactly why we were refined. Immediately after that we had a Management Team meeting… and we're all discussing what happened. And then I'm told "Okay, file the case", okay, I'll file the case. And then I had a couple of immediate phone calls, right after each other, asking everyone for evidence. I got the evidence. And then I had an online team meeting with my internal audit team. And we talked about it[raid], we discussed everything that happened and I said, "Okay, my email", and I immediately go through all of my emails, I save all of the documents, all of the evidence, and I stick all of it in one email…Pictures, screenshots, emails (IV: Okay) Everything relating to the incidents and I emailed it to my internal auditor [[John1]] and copied it to my assistant [[Janette]] and sent it off.*’ (P9)

As demonstrated in P9’s discussion above, continuously interacting with new information was another attribute for the level of complexity of the task at hand. Participants interacted with information in numerous ways such as knowledge building, dissipating information to other people/systems and, collecting information to inform their next step in the process.

Interacting with elements discussed above such as technologies, people and information contributes to the classification of complex tasks. In this study the delivering of such complex tasks facilitated new ways of working or in cases where processes were prescribed they were deemed insufficient by participants in light of new information. Findings thus concluded that the complexity of a task can contribute to unintentional insider threat (UIT) and processes should be routinely evaluated across all designations as indications of potential UIT.

4.6.2 Speed of the incident, discovery and response

While participants were engaged in delivering various complex tasks, unsurprisingly they all reported a very small window of time (seconds, minutes) over which the cyberbreach occurred. Speaking to P3, who was peer reviewing applications, she described the speed of the incident as occurring over a span of a few seconds.

'*I think it was like, I think it auto-filled after I put in the, I started typing the first name. And it auto-filled the email address. And I just sent it off.*' (P3)

P10 is an accountant who manages various portfolios and liaises with clients frequently to receive and transfer funds. She describes a fast-moving environment which relates to financial year end in the UK as well as having to regularly deal with external auditors. Existing in such an environment she has constant deadlines which reoccur annually. When performing a task, she accidentally engaged with malicious content that resulted in being hacked. Typical of our sample, P10 related the speed of the incident in the quote below.

'*I didn't realise at that time that I actually clicked the link. I realised that a week after. I probably clicked on it and then closed it straight away (IV: Yeah).*' (P10)

The speed of the incident itself was so fast that some participants did not initially recognise they had experienced a cyber breach at all. Participants' experiences reflected how users at any given time are a-click-away from a cyberbreach occurring which might be a feeling that was heightened in our participants post experiencing a cyberbreach.

Speaking to P5, who worked in a fast-paced environment that funds grant applications and accidentally sent sensitive information to the wrong recipient, she spoke of the speed of the incident and went on to say how she did not realise that a breach had occurred in the first place.

‘*And then I got an email back from somebody to say, "Hi, I don't think this was meant for me. What would you like me to do with it?". And it turns out that I'd actually sent it to another [[Joe2]] that I've previously emailed, that had come up automatically on my email list (IV: Okay).*’ (P5)

Not being able to identify if a breach had occurred straight away was common amongst our participants whilst the time taken to discover a breach varied (some cases it was within a minute, some took up to a week). Speaking to P6 who manages overseas partnerships, she described how it was her systematic approach to the task that allowed her to discover a discrepancy that identified a cyberbreach.

‘*I had a separate spreadsheet where I was documenting which stage I was at with the [[funded]] projects. Mhm, and then I would've noticed the discrepancy between what I hoped to do [laughter] versus what I actually did. And so that recording[stage] would've been an immediate check.*’ (P6)

The discovery of a breach was reliant on either on some form of a checking process or delayed feedback. Examples included the use of checklists, other people identifying an anomaly with participant’s account or, a pop-up from the malicious software itself. In many cases this feedback was delayed and serendipitous suggesting an absence of clearly identified feedback loops and the possibility that some breaches may remain undetected for an indeterminate period of time or might not be detected at all.

Once the cyberbreach was discovered participants described a quick response speed to protect their cyber defences. This time window varied between a few minutes to a couple of days but in all cases was longer than the time window for the incident to occur depicting a typical gulf of evaluation (Norman, 1986) for unintentional cyberbreaches.

As incidents were so rapid in their nature and participants were performing complex tasks, later on in this thesis a discussion is held about how this element that contributes to the fluctuating nature of unintentional insider threat (UIT) must be evaluated. Discussion also encompasses how assessing tasks to evaluate adequate feedback loops is essential and processes should be in place where individuals are familiarised with protocols in the event of a cyberbreach to respond to threats promptly and effectively. As part of the following *Reflections* section ‘stop, think, ask, action, consequence’ or ‘STAAC’ is introduced as a heuristic means of introducing reflection on actions taken.

4.6.3 Actions

Participants shared the actions they took prior to and during the breach. Speaking to P10, the accountant who had successive deadlines, she discussed how it was the start of the day and apart from having catch ups with her team she was already forward planning for the day. She spoke about how she was doing a routine action when the breach occurred:

‘*It was an email from a company who we work with on a daily basis (IV: Okay) So I'm receiving emails from them quite regularly but at that time, I didn't realise that this email is actually a phishing email because it was sent from the person who I'm dealing quite often, quite frequently (IV: Yeah) I didn't think that email could be spam.*’ (P10)

Contextual features appeared to assist participants in identifying safe content (such as an email from a known ‘safe’ sender). However, applying the same principles or the misapplication of the same contextual cues made participants vulnerable to UIT (i.e. receiving a virus from a known safe sender as their account had been compromised). This misapplication of contextual cues was often supported by assumptions being made to validate anomalies in interactions with people and systems, in turn increasing the susceptibility to UIT as users progressed in their decision making (DM) to engage with malicious content. The following quote from P4, the researcher in higher education, exhibits this as she shared her reasoning for progressing through what seemed like a usual request i.e., interaction with her supervisor.

'*And what it was, it was an email with, like, a picture of like a voice memo (IV: Okay) And I genuinely assume that my supervisor had just sent me a voice memo because I don't know, maybe that's his new way of communicating with me.*' (P4)

Results showed that during or immediately prior to experiencing a cyber breach, participants formed assumptions or had contextual cues that encouraged them to continue progressing with the task or underestimate the impact of a potential cyberbreach. In all cases participants carried on with their normal duties until a cyberbreach had actually been discovered.

When speaking to P1, who is also a researcher in higher education and was hacked as a result of interacting with malicious content, she described her actions after discovering that she had been hacked.

'*And yeah so, when the bank called me and I obviously killed the card and so on, and they blocked the transactions, so there's no harm done. But then I went to the [[Employer]] IT and they took it very seriously and… so everything was wiped out and reformatted.*' (P1)

In contrast to not knowing if a cyberbreach had occurred, once participants discovered that a cyberbreach had taken place they all performed some form of action to reduce the impact of the breach. These actions included turning off equipment, reporting the incident to managers, contacting IT specialists and, recalling the message. P10's quote highlights this below.

'*And I remember he mentioned that he had to return emails too, but it was quite a large number of emails that had been sent from my email. So he[Head of IT] had sent emails to* *people just to mention that my email has been hacked, so like an apologetic email just to reassure them that they should ignore this [participant's initial] email and if they clicked on the links or other things, then it's better to change their password (IV: Yeah)*' (P10)

Thus, actions can become critical for how to avoid or tackle a UIT should it be realised. Findings from this heading feed into the framework by assessing individuals' understanding of outcomes from a cyber breach and ties in with a culture of empowerment which will be discussed in the following section of this Chapter.

4.7 Theme 3: Accidents

The broad themes that directly linked to the incident of cyberbreach itself included training, expertise level, participants' trust in technology and errors (expecting errors from others, errors in expectations and, accepting errors in themselves).

The following findings informed the following inputs in the sociotechnical framework for Pillar 1: User Vulnerabilities to UIT and recommendations to strengthen defences; Pillar 2: The effectiveness of processes and facilitating a continuous improvement culture and; Pillar 4: Knowledge sharing and empowerment culture:

Assess individuals' trust in technologies	Pillar 1
Assess the levels of how much individuals rely on their social networks (offline and online) to inform their decisions if faced with threats	Pillar 1
Assess levels of retention from basic ICT teachings to establish levels of awareness	Pillar 1
Evaluate in-use software's limitations in prescribed processes	Pillar 2
Assess levels of stigma associated with experiences of near misses and accidents that result in cyber incidents and cyberbreaches across all levels	Pillar 4
Assess levels of communication about cyber incidents	Pillar 4
Assess individuals' understanding of outcomes that result from accidents	Pillar 4

Evaluate individuals' ability to question, share and challenge abnormal interactions	Pillar 4

4.7.1 Training

Participants were asked if training could have helped them bypass the cyberbreach. In this research study's sample, training was not seen to have much influence in deterring unintentional insider threat (UIT). The following quote by P4, the researcher in higher education, reflects this wider sentiment.

'*So you know, obviously we receive these types of trainings (IV: Yeah) But it's hard to integrate them into real life.*' (P4)

While participants believed that training was slightly useful for general and theoretical awareness, they did not feel it could be good enough to bypass the cyberbreach especially given the unintentional nature. Reinforcing earlier findings, participants identified informally acquired experience as more potent with all participants identifying sharing of knowledge with peers as valuable. Peer-to-peer sharing of knowledge and experiences through any medium (face-to-face, emails, forums, trainings) was believed to be a more effective form of learning and awareness than training for avoiding unintentional insider threat (UIT). The experience of a breach appeared to have the most influence on avoiding future UIT amongst participants. Results were indicative of a direct correlation between UIT and personalised experience, where first-hand experience can be the biggest deterrent, followed by experiences of people that are known in real-world settings and training was deemed to be the least effective. This finding is incorporated into the framework (discussed in greater later on in this thesis) to recommend auditing of personalised trainings and instilling an empowerment culture to help mitigate UIT.

4.7.2 Expertise level, trust in technology and errors

All participants identified themselves as having mid or advanced level of expertise at their jobs (listed in Figure 7 below and presented earlier in this Chapter).

Participant	Cyberbreach Setting	Novice	Mid-level	Advanced	Analogue?
P1	Professional		x		Absent but context present
P2	Professional		x		Absent
P3	Professional			x	Present
P4	Professional		x		Absent
P5	Professional			x	Absent
P6	Professional			x	Absent
P7	Personal		x		Absent but context present
P8	Professional		x		Absent
P9	Professional			x	Present
P10	Professional		x		Absent

Figure 7: Participants' experience level and the presence of analogues

Eight participants shared that at the time of experiencing the cyber incident they were not reminded of similar lived experiences in the past, which are known as *analogues* to help aid decision making (DM). P7, now a researcher in higher education, described how installing a programme on her personal computer several years ago resulted in ransomware. She spoke about the panic at seeing the ransomware notification (which appeared to be from the FBI demanding Bitcoin crypto currency) and thinking about the potential loss of emotionally valuable data such as digital family pictures. However, she was less phased by the malware itself.

‘*I didn't really think it was a genuine notification because I read about it before that this is a malware.*’ (P7)

Thus, analogues were absent in instances where context was present for the participants (for instance being able to identify a popular ransomware’s pop-up but not having similar lived experiences). Similarly, some participants shared that they had context as an analogue which meant that their lived experiences either encouraged them to proceed in the task despite reservations or identify the threat as it was unfolding (early detection). Analogues also included recalling previous cyber incidents, but none had escalated to a cyber breach. Overall, the experience of this incident was novel to a majority of the participants encapsulated in P6’s (the one who manages work with overseas partners) quote below.

‘*…So yeah, I mean there would've been plenty of things that would've reminded me [laughter] but they wouldn't have been incidents. They would've been times when I caught myself before sending the thing (IV: Yeah) rather than a breach.*’ (P6)

In fact, for two of our participants analogues were present and it aided them in identifying subsequent steps to take moving forward and/or to anticipate consequences. Speaking to P9, the lawyer, she shared how an analogue was present in her lived experience.

‘*Well, I have done this a lot [(IV: Okay) laughter]. And the thing is, by that I mean, most of it might not be important… And then the irritation of having to fix everything. It's like I do a thousand emails a day, now I have to do five more just to fix this.*’ (P9)

When discussing the cyber breach and exploring trust in technologies, participants had trust in technologies to protect users from harm such as malicious content blocked by firewalls. This is evidenced in the following quotes.

‘*But my installed antivirus software was a premium plan back then, so I was a little bit sad that it didn't work.*’ (P7)

And,

‘*There was nothing else. And those two things I think were definitely [credible]. Especially the fact that it came from someone from the [[Employer Name]] address. I mean, had it not been someone with a [[Employer Name]] address, I probably wouldn't have clicked on it. I think that was the main thing.*’ (P1)

Participants also largely described a trusting relationship with technologies for automated elements within tasks. This included not cross-checking automated actions such as recipients that are auto populated for emails or not suspecting emails from within the organisation. In fact, this very trusting nature for systems to be secure resulted in a majority of the participants being victims to a cyber breach.

However, participants reported a distrust in technologies to perform a task correctly which included concerns such as manually cleaning data when exported to make it readable, reliability issues with exports, user problems such as early time-outs and forgetting passwords, all of which led to human input to overcome software limitations. This distrust in technologies to perform a task is reflected in P5’s comments below.

‘*So pulling up the data was, I think just pulling up the data in general… was [Online platform 2] at the time, was always a nightmare… Because when you pull up data, it just didn't. All of the comments weren't[formatted], you couldn't read it[data] in a way that was easy to read. So it's up to me to format it and manipulate the spreadsheet to make it really easy to read for them[panel members].*’ (P5)

Human input also led to participants making assumptions during various points of the cyberbreach. For instance, actions that triggered the breach were seen as insignificant even as the incident unfolded and the overall significance of the action underestimated. This is exhibited in P1's comments below.

'*So that wasn't very good judgement from me at all [laughter] I was like, "Oh, whatever". And I carry on as per usual, you know… Then then after that, because in my mind after that, once I discovered the document, and I was like, "Okay, this must have been an error". And then I didn't think back.*' (P1)

Assumptions were also made about various elements of the cyberthreat by some participants. This meant that interactions that seemed out of the ordinary were normalised by the participants as they could associate a reason for why the interaction was occurring. For instance, when speaking to P4, a researcher in higher education, she discusses how after clicking the malicious link nothing happened (i.e. the lack of a feedback loop).

'*I remember emailing him to say, "Oh, like I'm not able to download what you've sent me. Can you send it to me again?" (IV: Yup). So I must have genuinely been believing this [malicious content was genuine) through the whole timeline of it [episode of the cyberbreach].*' (P4)

Lack of a feedback loop allowed participants to assume there was an error (for instance a broken link, wrongly attached file, an error in the process or, human error) but this assumption only occurred after they had been compromised.

Expertise levels can result in greater levels of analogues being present when deciding how to react to potential or unfolding unintentional insider threat (UIT). Analogues can assist in effective steps taken to contain the threat and are incorporated in the framework under evaluating automated tasks, assessing technical skill levels and, evaluating effectiveness of guidelines in the event of a cyberbreach. Individuals' trust in technologies is also a notable factor contributing to UIT and this is addressed in the framework through evaluating software limitations and, evaluating levels of trust in technologies amongst employees to strengthen defences. Errors are incorporated into the framework by evaluating individuals' ability to question, share and, challenge abnormal interactions that would indicate a culture of knowledge sharing and empowerment.

4.8 Theme 4: Organisational factors

Another major theme that emerged from the data involved factors that related to the wider context under which participants performed their tasks, relating to organisational factors. Sub-categories included individual emotional responses, employer dynamics, processes, goals, pressures, peer dynamics, physical environment and, external factors all of which appeared to interplay with the conditions that facilitated an unintentional cyberbreach.

The following findings informed the following inputs in the sociotechnical framework for Pillar 1: User Vulnerabilities to UIT and recommendations to strengthen defences; Pillar 2: The effectiveness of processes and facilitating a continuous improvement culture; Pillar 3: Workload and sufficient resource allocation; Pillar 4: Knowledge sharing and empowerment culture and; Pillar 5: Fluctuating vulnerabilities:

Assess physical working environments	Pillar 1
Assess individual's level of caution when interacting with suspicious or odd behaviour (online and physical parameters)	Pillar 1
Evaluate the effectiveness of prescribed processes amongst all designations	Pillar 2
Assess individuals' levels of personal responsibility felt when delivering tasks assigned to them	Pillar 3
Assess resources available to individuals to deliver tasks	Pillar 3

Assess individuals' motivations when delivering tasks	Pillar 3
Assess individuals' ability and willingness to take on additional tasks	Pillar 3
Assess levels of stigma associated with experiences of near misses and accidents that result in cyber incidents and cyberbreaches across all levels	Pillar 4
Assess levels of communication about cyber incidents	Pillar 4
Assess individuals' understanding of outcomes that result from accidents	Pillar 4
Evaluate effectiveness of current guidelines in the event of a cyberbreach	Pillar 4
Evaluate individuals' understanding of protocols in the event of a cyberbreach	Pillar 4
Evaluate relationships between individuals and managers across all levels	Pillar 4
Evaluate relationships between peers across all levels	Pillar 4
Assess individuals' level of attention to detail (online and physical parameter)	Pillar 5

4.8.1 Individual emotional responses

A range of emotional responses following the breach were shared by participants during the interview. For instance, when speaking to P10, the accountant, she described how the malicious virus was activated a week after clicking a link from a known client she had frequent contact with. She described receiving a phone call initially from another client to alert her about a suspicious email they had received from her account. Whilst she was on the call with them her emails and phone lines got inundated with other clients calling her to alert her or to confirm if the email is genuine. P10 goes on to share her emotional response which is exhibited in the quote below.

‘*And at the same time I was feeling really embarrassed because of the way I was acting in clicking on this email (IV: Yeah) I should be more considerate and [laughter] more observant in a situation like this.*’ (P10)

P2 who managed grant funding in healthcare shared, ‘*So yeah, there was obviously that like cold dread realising like, 'Oh, that really shouldn't have happened'.*’ (P2)

Overall, there was a feeling of disbelief that the incident happened to participants which elicited feelings of embarrassment and gullibility making them more cautious going forward. P6, who worked with overseas partners, described a feeling of disbelief which is demonstrated in the quote below.

‘*What happened when I did realise was a big "Arghhhh!" moment. Just like quite a quick, mhm, I mean I got on the email very quickly to ask them to discard it… Yeah so I mean I was aware of the pitfalls and I thought I had robust enough personal systems to deal with it. But of course not in this instance.*’ (P6)

P8, who worked at a think tank in a metropolitan city when she experienced a cyberbreach by sharing sensitive information that resulted in her account being hacked, shared her feelings at the time of the incident shown in the following quote.

‘*A lot of panic, anxiety (IV: Okay) I wasn't so much surprised. But you know when you've done something wrong, and you have this immense feeling of guilt, and you're like, "Ah, no!", and you just want it to go away. And there's nothing that you can do, you're powerless to do it, because you've already done it and there's nothing you do now, but own up to it.*’ (P8)

Participants shared having felt guilt and a sense of personal responsibility for being compromised. Furthermore, participants also shared feelings of frustration at themselves, software and, the processes that facilitated the cyberbreach. As participants developed a level of caution post the breach, the framework (presented in the following section) incorporates measuring levels of caution amongst employees to assess user vulnerabilities for unintentional insider threat (UIT).

4.8.2 Employer dynamics

Discussions at the interview also included employer’s response to the cyberbreach which was interesting as it sheds light on some of the organisational factors that might have contributed to the unintentional insider threat (UIT) incident that participants experienced.

Organisations’ actions following a cyber breach appeared to fall short of strengthening cyber defences against UIT. For instances in our findings, one example is a disclaimer ribbon on emails by the employer for employees to only interact with content that they recognise as safe, evidenced in the following quote by P10.

‘*But recently IT what they [do], any external emails which we receive there was a warning… "Do not open links or attachments unless you trust the sender and know the content is safe". We sort of, they, after all the incidents at the company they decided to do it the other way. I mean to protect the staff (IV: Yeah).*’ (P10)

While this prompt can be a useful reminder it would not have prevented the UIT experienced by P10 who recognised and trusted the sender. Beyond this trust, it would be problematic to know if the content was safe without exploring it as the malicious link did not have any identifiable anomalies to P10 i.e. it did not ‘appear’ malicious. These types of notices can create a safety climate (Neal and Griffin, 2004) as opposed to a safety culture (Reason, 1998) and be seen to place full responsibility on individuals for their actions, in turn propagating a blame culture. Overall for our participants, where applicable, IT department personnel helped to combat the threat without placing blame but there were undertones of how the incident created more work for them. IT department’s countermeasures caused participants to experience downtime which disrupted their work with additional follow up tasks such as password resetting.

Participants also discussed their relationships with their line managers as well as any senior designations that were involved once a cyberbreach had been identified. Below, P2 mentions the impact of the employer's message which appears to elicit desirable behaviour through punitive measures to circumvent unintentional insider threat (UIT).

'*So this happened over the course of several months, and then they did a big presentation basically about how terrible we've all been and how we'd failed the company in many ways [laughter]… We couldn't really add in more time [to the deadlines], never can. Well, sometimes you probably can, but that never seemed to be the option that came up… So, yeah, there were changes that came about anyway. So you kind of get the more direct ones were like the checklists and things that we partly used as training for new people and just to have documentation. The company actually had like an ops team and policy documents that everyone had to sign, you had to read them and have like an Ops[Operations] induction and testing and data protection training and all that stuff.*' (P2)

While measures such as checklists and improved processes can be a step in the right direction, in P2's case it is implemented in a way that placed responsibility on individuals if things went wrong — a classic example of blame culture stemming from traditional security thought where humans are seen as the weakest link in systems. P6 who worked with overseas partners captures organisational blame culture in her quote below.

'*But then what comes with that money is resource constraints. So you can't spend that much on staffing. And so that[incident] was definitely you know [that time], at particular crunch points like the end of year where you have quite a lot of activity happening. You spend the rest of the year planning for and then sort of implementing and then it comes down to quite a lot of administrative stuff whilst also delivering on top of your day job all like in this kind of* *period without really additional resource being there. So I guess that's where the funding side comes in. There's constraints that come with it, mhm, that force to you deliver sometimes and give you much more constrained time frame in which to do so… And without too much kind of, I don't know how to say it, at that point mm. It wasn't like a facilitative environment, it was more if something goes wrong you need to [laughter] have somebody to blame [laughter].*' (P6)

Another example of blame culture, although not directly punitive, was captured in the discussion with P5 shared below. She worked in a fast-paced grant environment and reportedly had an open and approachable relationship with her line-manager and promptly reported the cyber breach as soon as she discovered it.

'*I went to [[Joe1]] my line manager and asked him what I should do. He went to [[Jane]] who was our [[Head's Designation]]. And [[Jane]] came up to me and firstly asked me to ask [[Joe2]] who I had sent it[the sensitive email] to, to delete it. So I did that. And then once [[Joe2]] replied to me to say that he had deleted it, that was kind of just the gist of it really (IV: Okay) Which led me to getting a stern, "So, please don't do that again" [laughter]… [[Head of Ops]] would be responsible for that. He would be responsible for processes and procedures (IV: Okay) so making sure that they[policies] were being kept [followed by P5] I guess.*' (P5)

A majority of the participants shared how they were able to work autonomously with approachable line managers. This open communication correlated with participants' willingness to share the cyberbreach with their managers early in the lifecycle of the threat. Participants also expressed having good immediate relationships with their peers and managers. However, participants' experiences beyond these immediate relationships fundamentally reflected a blame culture discussed above. Additional tasks that were introduced by employers as a result of the cyberbreach to safeguard against unintentional insider threat (UIT) would be fundamentally inadequate, such as signing additional contractual documents or being told not to do that again.

All our participants shared the sentiment of limited resources to perform their tasks (such as time and people) in the organisation which was believed to be a contributing factor to the cyberbreach. For instance, P10 described how she was questioned over taking more time than anticipated for another separate task. This lead to an agreement that P5 must inform her line manager (the Director) if she needs more time on the tasks she is performing:

'*If you asked my Director, it's gonna be very easy for him [laughter] Because he's got that extra knowledge in Excel [laughter] Sometimes I struggle so, for me I have to do for example, the day I have to do reconciliation for something and it took me a bit longer so I need to get in contact with him.*' (P10)

This comment showed that there were tasks that took P10 longer than it would others. In such instances she had to inform management when she was engaging with such tasks, reflecting how time might be a limited resource within their team and require advance planning. All participants shared this sentiment about limited resources to perform their tasks such as time and people in the organisation which was a contributing factor to the cyberbreach.

P8, who worked at a think tank, reflected on the consequences of the breach in the quote below.

'*IT [department] just shut down my account and then they made me a new account. And that was it and it was like "Carry on, start again". And I don't remember there ever being like, I was ever told off, I was never given a lesson on what I did wrong, except for the fact that [P8 should use] BCC not CC [recipient fields in an email].*' (P8)

P2, who worked in healthcare grants, also said that she did not take any meaningful learnings from the experience.

'*Obviously, I suppose they don't want to come down too heavily on people because they do still want you to say that this has happened. So I had to apologise to the person and ask them to permanently delete the email. And then obviously find new reviewers and make sure that I follow the process more closely in the future.*' (P2)

Overall, in all our participants there appeared to be a lack of organisational and individual learning and accountability from the cyberbreach. This finding is incorporated in the framework by evaluating the effectiveness of prescribed processes. Findings from Employer Dynamics also contribute towards evaluating the effectiveness of guidelines in the event of the cyberbreach, assessing individual's understandings of protocols in the event of a breach, evaluating relationships between individuals and their managers, assessing stigma associated to incidents, levels of organisational communications about cyber incidents and, assessing resources available to deliver tasks.

4.8.3 Processes

Participants described vague processes in place that generally guided them in how to perform various tasks. P2, who worked in healthcare grants, reflected on the prescribed processes.

'*So they[Previous Employers] were quite good in that sense that they had a lot of things written down already, but not often the nitty gritty of how you do things and why it was important to not skip the step or whatever.*' (P2)

All our participants were relatively experienced in performing the tasks at hand and discussed how this familiarity allowed them to skip steps in the process that they did not deem important. This is reflected in the following quotes from P6 and P3 below.

‘*That was probably my second time through it[stage in the process] and the first time, again, similar thing, everything rushed, lots of things going on. It would’ve been a lot of misses, right? So that’s why you sort of develop these systems which you try to become more robust with [laughter] Or you just remember where the near misses were. And then either develop systems or just be very, very weary and then give yourself proper time to be able to focus. I think that’s the main thing really whenever you’re engaging with something quite so process orientated.*’ (P6)

And,

‘*I should say actually, in fairness, this process wasn’t really how things were supposed to work. Like, the applications were supposed to be sent securely through the grants [online] system… So, yeah, it wasn’t good practice I would say and even within the processes as they were set out… But also, yeah, then they [assigned reviewers] have to do the reviews. And some of them would do it through the system and so they would come in automatically. And some of them will email it to you if they were people who didn’t really like engaging with the system as much.*’ (P3)

Skipping steps or using unofficial channels was linked to saving time, efficiency or, convenience, indicating established routine violations (Reason et al., 1990). Participants also discussed how processes had limitations, how their existing context facilitated their error but more importantly, how they were aware of processes having limitations or potential for errors if followed as prescribed. These findings contributed towards the input of evaluating effectiveness of prescribed processes amongst skilled staff in the framework.

4.8.4 Goals and pressures

Discussing participants' goals at the time of the cyberbreach was important as it reflected their motivations for the task and how they performed it which might have contributed to the unintentional insider threat (UIT) they experienced. When asked to declare time pressures experienced on a scale of 1–5, participants reported feeling under time pressure to deliver the task with an average score of 2.8 points/participant. Speaking to P5 who worked in a fast-paced environment and P6 who worked with overseas partners they discussed their motivations and challenges at the time as follows.

'*You have a deadline of when your award meeting is, but there's so many kinds of things that come up along the way that just delay you (IV: Yeah).*' (P5)

And,

'*So it was quite pressurised only because, mm [thinking pause], like there's a lot, because you're under pressure to hit the financial deadline but a lot of the process to getting there isn't in your control necessarily (IV: Okay).*' (P6)

The quotes above reflect that pressures did not solely emerge from time constraints (time pressure listed as 2.8 on a five-point scale) but also from deadlines and feeling a lack of control. Furthermore, participants also added other goals that motivated them which included wanting to move on to another task, following the prescribed process, desire for a lower workload after successfully completing the task and, being able to achieve a larger more important goal through the completion of the task at hand.

Stemming from this discussion about pressures, participants went on to elaborate on factors that were at play at the time of the incident. P1, the researcher in higher education, elaborates influencing factors beyond time pressures below.

‘*But it’s just if you do these things in a rush, I think, and you don’t take your time to even read the URL, so, you might just do it kind of instantly… And I had a deadline coming up in a couple of weeks. So I was in, like, not stressed as such, but kind of in a rush.*’ (P1)

P10, the accountant, also elaborated on the settings of her cyber breach in the quote below.

‘*I remember that because Mondays and Tuesdays, they’re always quite busy for my workload, because I have to make sure that all the payments are prepared and everything has to be processed before the end of the week… You know, when you’re in a rush and doing things you could just accidentally, without really thinking, you’re just going through and because you’re in a rush to do other things you don’t really focus on what you’re clicking.*’ (P10)

Other than time pressures, participants experiencing factors such as planned deadlines and anticipated workload led to participants wanting to move on and rushing. In some cases this rush was supported through implementing automatic or skill based behaviour to progress through the task. Findings from this theme contributed to the input of assessing individuals motivations when delivering tasks to assist in identifying potential unintentional insider threat (UIT).

4.8.5 Peer dynamics and physical environment

When understanding the context in which cyberbreaches occurred participants discussed relationships with their peers before and after the incidents as well as their physical environments. P9 reflected on the impact of the breach on her peer dynamics in the quote below.

'*The negative repercussions of this are, is that (a) my lawyer [John2] knows that I didn't give him two cases and someone else got them and he's feeling a bit antsy. Because he called me again later and he was like, "You didn't give me those cases". And I was like, "Yeah, I didn't want to. I gave them to someone else".*' (P9)

P9's comments above exhibit how she had to justify assigning a different person to the task which was discovered by the unintended recipient as a consequence of the cyberbreach. This quote also shows undercurrents of potential friction caused in the relationship. P8, who worked at the think tank commented,

'*But within my immediate group, everyone was quite quiet, because they were doing their own work and doing their own thing. And I was just slowly panicking [laughter] Being very quiet because I didn't know what to do, like, I told my manager… As everyone else was just doing their normal thing, living their normal life, being their best them, like doing their job.*' (P8)

P8's comments reflect a sense of alienation from her peers during the cyberbreach which otherwise appear to be good interpersonal relations at a peer-to-peer level. P6, who worked with overseas partners, described her relationships with peers in the context of the cyber breach below.

'*I have a colleague sitting right by [me] and I would've explained something [laughter] "I just sent the wrong contract to the wrong people", maybe had gotten a little bit of advice of how to go about it or something like "Don't worry, happens all the time, it's okay" that kind* *of thing. And then gotten on to try and correct the error pretty promptly (IV: Yeah) after that. But yeah.*’ (P6)

Overall, all our participants generally described having a friendly relationship with their peers which included being able to openly communicate with one another for advice and provide support through the cyberbreach. They also described having a competitive relationship with their peers and everybody working autonomously to deliver their individual key performance indicators (KPIs). This led to participants feeling alienated from their peers and responsible for controlling the impact of the cyber incident as is shown in the quotes above.

Participants also described normal office environments with open plan spaces and normal noise levels. They all described physical environments where they could concentrate on tasks. In the results of this study physical environment did not seem to be a contributor to UIT but it did not also mitigate the threat from occurring. These findings contributed to evaluating relationships between peers and monitoring attention to detail in virtual tasks and physical environment within the framework. Having strong relationships at a peer level showed to have positively influenced mitigating against unintentional insider threat.

4.8.6 External factors

When speaking about their personal methods of working participants shared certain traits that might have contributed to unintentional insider threat (UIT). For instance, the quotes from P6, P8 and P9 below showcase these traits.

‘*Ideally what I like to do is as soon as I'm getting something back, sort of send it out as quick as possible. So once I know, then I'd quite like to have things go and be in process (IV: Yeah) as early as possible (IV: Yeah).*’ (P6)

And,

‘*So like, there wasn't any pressure on me to get this email out. But an eagerness for myself to do a good job, which is the irony of it all.*’ (P8)

And,

‘*And this is not just a job, it's a really good job… That is the number one pressure that's been in my head for the past four months. Not to slip up, not to do anything silly, to make sure that I'm doing the right thing, that I'm appearing to be as efficient and competent as I hope I am.*’ (P9)

Thus, interviews uncovered specific individual traits within our participants that might have been stimulated through various external factors. The traits reflected by our participants included being willing to take on and expect ad hoc work, being responsible for multiple projects, anticipating workload, taking personal responsibility for the delivery of tasks assigned to them and, being detail oriented. Participants also appeared to possess good communication skills and the ability to ask for help which allowed participants to reduce the impact of the cyberbreach. These traits also reflect a deeper connection to external factors such as job security and losing income for the organisation.

Whist there appeared to be a tension between prescribed processes and factors that might influence deviation, all our participants also showed an active commitment to best practices and compassion for others who inputted into their tasks. In this study participants appeared to compromise on prescribed processes in favour of compassion for others. For instance, a quote from the discussion with P3 below.

‘*While you do connect them through the system, you also send them a PDF, just so they can have something that they can read on the plane or something. Without faffing around with* *logging into the system, etc. So, yeah, it wasn't good practice I would say and even within the processes as they were set out.*' (P3)

Beyond compassion for others, our discussions with all participants showed self-esteem as another factor that participants were actively considerate of. This included how they appeared to themselves and to others around them. For instance, the following quotes from P6 and P2 below.

'*And this thing[programme], it was one that I had personally developed too and, so you just want to not be the [one who let it down].*' (P6)

And,

'*But it wasn't kind of directly competitive at the time, but it's like you're aware of what you're [doing] and what everyone else is doing because you're all doing the same task. So you don't want to be the one that's falling behind. Particularly if the team is understaffed, you're having to pick up slack. So there was probably some aspect of not wanting to look bad. Not only because I was new, but because I was as new as other people, we were doing the same stuff. Like not wanting to look bad within that group setting.*' (P2)

The above findings pertaining to external factors appeared to influence the conditions that facilitated unintentional insider threat (UIT). These findings contributed to inputs for assessing prioritization of processes, commitment to best practices, personal responsibility taken by individuals when delivering tasks and, levels of stigma associated to near-misses or accidents within the framework.

This research study aimed to explore the lessons that can be learnt from people’s experience of unintentional insider threat in order to examine influencing factors that enable this threat. In order to change the way humans are considered in systems, this study extended the application of the Critical Decision Method technique which is centred on the human element in complex environments. Findings discussed above evidenced four areas that interact with and influence unintentional insider threat. The next section of this Chapter shares reflections contained within these findings in the context of relevant literature discussed earlier.

4.9 Reflections

The research study investigated factors that exclusively influence unintentional insider threat (UIT) through the application of Critical Decision Method approach. Findings from this research study detailed above were characterised as either features or action suggestions. Examples of ‘features’ included data codes such as suspect logos, deviation in language, speed, compassion for others while ‘action’ suggestions encompassed actions, such as turning equipment off, performing tasks and subsequent task elements, rushing, being reprimanded amongst other codes. One way of summarising the results in an accessible form is to adopt the convention of the Epidemiological Triangle (Cassel, 1976) that was discussed earlier in Chapter 2, Section 2.5.1. Apart from the public health sector, this approach has also been used in the context of safety science (e.g., Gordon 1949; Haddon 1968; Lagerstrom et al., 2016). In this context of unintentional insider threat (UIT), the triangle is used to visualise the Host (the User) and Agent (the Exploit) both of whom may have various forms of intrinsic resistance and virulence, respectively, which are strengthened or weakened relative to each other by the environment (the Work Context). Figure 9 represents three vectors that may be militated against to reduce the chance of a cyber incident or breach. The triangle aids in positioning the probability of a breach in relation to the features of the Exploit itself, the qualities of the User and their prior experience and the Work Context in which the breach occurs.

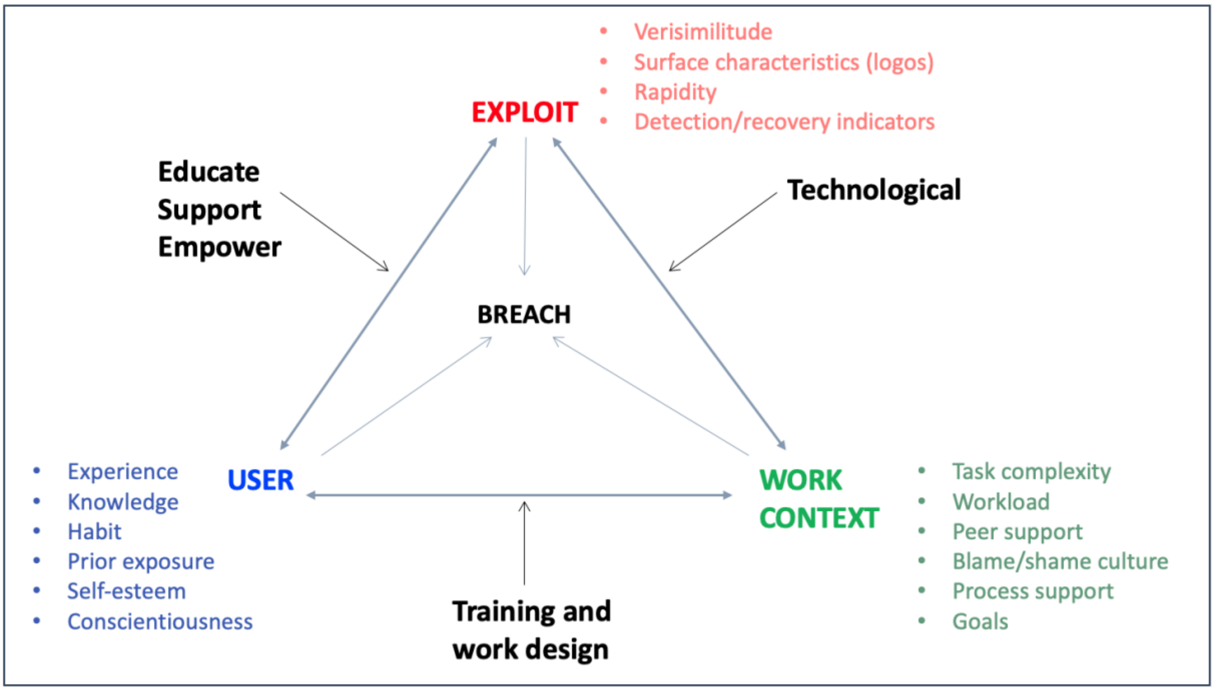


Figure 9: Epidemiological Triangle based on CDM research study

The overview exhibited through the Epidemiological Triangle in Figure 9 allowed for a deeper understanding of whom would need to take responsibility for different elements of intervention, providing a structure to delivering any interventions. Subsequently, data codes were re-classified to create a sociotechnical framework, shown in Figure 10 below, which informs the following discussion. For the purposes of this work a framework is defined as a set of recommendations applicable in specific scenarios to reduce negative impact. The framework proposes a five-pillar action plan listed as Outputs that can be achieved through 35 distinct Inputs. Based on findings discussed above, Input elements can be captured to assess the potential level of risk in a setting and therefore provide appropriate timely interventions. This framework can be implemented by organisations that are interested in starting an unintentional insider threat (UIT) programme or as an evaluation tool for organisations that currently have one. This framework is advised to be conducted bi-annually or when organisational changes occur. It is also worth noting that recommendations include tailored training programmes that go beyond traditional face-to-face teaching and are audited for their effectiveness.

Outputs	Pillar No.
User vulnerabilities to UIT and recommendations to strengthen defences	1
The effectiveness of processes and facilitating a continuous improvement culture	2
Workload and sufficient resource allocation	3
Knowledge sharing and empowerment culture	4
Fluctuating vulnerabilities	5

Inputs	Contributes to
Assess how comfortable individuals are with various technologies and platforms	Pillar 1
Assess how vulnerable users feel in their daily online interactions	Pillar 1
Assess physical working environments	Pillar 1
Assess individuals' ability to identify spear phishing scams to note vulnerabilities	Pillar 1
Assess individuals' existing experiences with malware or threats (including physical spaces)	Pillar 1
Assess individuals' knowledge base to evaluate understanding of current techniques used by hackers	Pillar 1
Assess individuals' susceptibility to rationalise abnormal behaviour or interactions	Pillar 1
Assess individuals' susceptibility to spear phishing	Pillar 1
Assess individuals' trust in technologies	Pillar 1
Assess the levels of how much individuals rely on their social networks (offline and online) to inform their decisions if faced with threats	Pillar 1
Assess individuals' awareness of mainstream marketing campaigns against popular attacks	Pillar 1

Assess levels of retention from basic ICT teachings to establish levels of awareness	Pillar 1
Assess and map different skill levels between individuals in a diverse workforce	Pillar 1
Assess individual's level of caution when interacting with suspicious or odd behaviour (online and physical parameters)	Pillar 1
Evaluate all tasks to identify missing feedback loops that indicate task completion	Pillar 1
Evaluate the effectiveness of prescribed processes amongst skilled/experienced staff	Pillar 2
Assess individuals' prioritization of processes	Pillar 2
Evaluate the effectiveness of prescribed processes amongst all designations	Pillar 2
Evaluate in-use software's limitations in prescribed processes	Pillar 2
Assess individuals' commitment to best practices set out by the company	Pillar 2
Evaluate processes for collaborative tasks that are automated	Pillar 2
Assess individuals' technical skill levels	Pillar 2
Assess individuals' levels of personal responsibility felt when delivering tasks assigned to them	Pillar 3
Assess resources available to individuals to deliver tasks	Pillar 3
Assess individuals' motivations when delivering tasks	Pillar 3
Assess individuals' ability and willingness to take on additional tasks	Pillar 3
Assess levels of stigma associated with experiences of near misses and accidents that result in cyber incidents and cyberbreaches across all levels	Pillar 4
Assess levels of communication about cyber incidents	Pillar 4
Assess individuals' understanding of outcomes that result from accidents	Pillar 4
Evaluate effectiveness of current guidelines in the event of a cyberbreach	Pillar 4
Evaluate individuals' understanding of protocols in the event of a cyberbreach	Pillar 4
Evaluate individuals' ability to question, share and challenge abnormal interactions	Pillar 4
Evaluate relationships between individuals and managers across all levels	Pillar 4
Evaluate relationships between peers across all levels	Pillar 4
Assess individuals' level of attention to detail (online and physical parameter)	Pillar 5

Figure 10: A sociotechnical framework to assess Unintentional Insider Threat

The framework is utilised to assess organisational readiness levels. Each of the five pillars is formed of respective 'Inputs'. These inputs were uncovered as part of the findings from the

Critical Decision Method based research study. This framework serves as a blueprint for identifying, intervening and mitigating UIT through the development of the website utilised by participating organisations

4.9.1 Technical defences

Passive cyber defences are undoubtedly a good measure to serve as the first line of defence to protect networks against attacks. This should include virtual and physical spaces encompassed in examples discussed in Chapter 2. While this framework lists a few technical defence elements, such as conducting penetration testing and mapping all staff's ICT skillsets, it is recommended that all passive defences within technical defences are fully inclusive in the implementation of this framework. These include best practices for software architecture, monitoring user activities and devices, configuration, encryption, managing access points (including privileges), data management, updating software and regular audits. Passive defences are in-line with NCSC's and CERT's recommendations but, in contrast to NCSC and CERT, this framework does not recommend restriction of devices or features as it can encourage users to create unauthorised or unmonitored back channels for delivering tasks. Instead, the following section proposes evaluating and noting how tasks are conducted so suitable defences can be implemented. While it would be desirable to develop in-house technical skills through formal training as suggested by NCSC, as part of the first pillar this framework recommends mapping existing skills to identify talent that can be utilised and developed across all designations. The use of active cyber defences such as those used in SOFIT are not recommended within this framework. Active defences can create complexities for implementation as they are dependent on permissibility in local laws (such as packet sniffing) as they infringe on individual privacy and can foster distrust between the organisation and individuals. Active defences also do not appear to be effective for safeguarding against UIT as active monitoring against unintentional actions is fundamentally inapplicable.

4.9.2 Sociotechnical defences

Points contributing to each outcome in the framework are not separated as belonging to individual, technological or organisational contexts but rather findings are integrated to provide effective solutions that can identify, intervene and mitigate unintentional insider threat (UIT). This work now discusses the framework, ‘inputs’ which provide objectives for evaluations and ‘outputs’ which are the five pillars to help gauge levels of vulnerability to UIT and provide recommendations to strengthen defences. The framework introduces ‘Stop, Think, Ask, Action, Consequence’ or STAAC that is used to foster Type 2 thinking that counteracts UIT. STAAC can be used prior, during or after a threat has been realised.

4.9.2.1 User vulnerabilities to UIT and recommendations to strengthen defences

The first pillar of this framework provides an assessment report to benchmark existing vulnerabilities to UIT within an organisation. This is done by evaluating 15 distinct points listed as respective ‘Inputs’ to Pillar 1 in Figure 10.

In contrast to CERT, SOFIT and other popular models that rely on psychological and behavioural profiling this framework does not assign methods or people for evaluating elements within inputs. As this framework specifically targets UIT it does not require any individuals (such as HR personnel or peers) deducing individual personality traits or reporting suspicious behaviours that are geared towards identifying unintentional insider threat (UIT). Instead, to benchmark vulnerabilities the framework evaluates individual’s comfortability with various technologies, risk awareness levels (in-line with NCSC), individual lived experience and acquired levels of knowledge to identify malicious content and individual susceptibility to rationalise anomalies in interactions. The framework also assesses physical environment for environmental stressors that are indicated in SOFIT and individual's trust in technologies to safeguard against malicious content. As part of the actions to evaluate UIT, it encompasses educating and raising awareness amongst individuals through traditional and hands-on training as suggested by NCSC and CERT.

4.9.2.2 The effectiveness of processes and facilitating a continuous improvement culture

The results from this study support a link between task processes and the risk of breaches and validates error management programme (Liginlal et al., 2009) recommendation for developing effective processes to tackle UIT. In addition to effective processes findings highlighted the importance of individuals understanding why steps within a process are important. For cases where explicit processes existed, participants who had good expertise at performing their tasks, skipped steps as the importance of following each step was not communicated. Skipping of steps within outlined procedures had resulted in near-misses (or incidents) in the past but never an actual breach.

In this section of the framework, it is recommended to evaluate processes with the help of staff who possess good expertise for performing their assigned tasks. In contrast to Liginlal et al. (2009), who focus on addressing the lack of expertise through training, this framework emphasises the importance of working with expert individuals to identify heuristics and shortcuts that can facilitate UIT. It also allows creation of processes that reflect *work-as-done* as opposed to *work-as-imagined* (Hollnagel 2017, Suchman 1987). At this stage the framework also assess the processes' effectiveness, individual prioritisation techniques that might compromise processes, conducting task analysis to device effective and improved processes as the delivery of the task changes, evaluating tasks that include automated elements and mapping software limitations that foster undesirable practices being implemented. While it is important to carefully consider implementing new systems that can facilitate new errors (Liginlal et al. 2009), it is also critically important to evaluate existing systems' effectiveness and suitability for prescribed processes. As part of this stage, the framework can also be utilised to evaluate individual commitment to best practices, trade-offs that are being made and mapping individual technical skill levels when delivering tasks as all these factors were shown to influence UIT.

4.9.2.3 Workload and sufficient resource allocation

Assessment of workload and allocating sufficient resources is the third pillar of this framework. SOFIT (Greitzer et al. 2018) includes workload as an indicator and while time pressures are the cornerstones for CERT and Nurse et al.'s frameworks for indicating UIT the results from this study highlighted additional interlinked factors. These factors included individuals feeling personally responsible for delivering tasks and allocation of sufficient resources which included *time* and *people*. Individuals' motivations for delivering tasks are also important to consider as motivations such as rushing or meeting unrealistic deadlines (in-line with SOFIT) which can support human fallibility. In addition, this framework incorporates organisational expectation for individuals to undertake new tasks that were also seen to be linked to UIT.

In contrast with SOFIT, our results did not suggest strong links between UIT and poor team management as participants reportedly enjoyed good interpersonal relationships with their direct line managers and peers. Findings did not support evidence for the mismatch between expectations and abilities listed in Liginlal et al.'s Error Management Programme as participants had mid-to-advanced level of expertise to perform assigned tasks. The study also did not find any evidence to support ambiguous goal setting or poor communication of goals (no data gathered for poor morale) all of which are factors included in SOFIT. Therefore, these elements have not been included in the framework.

4.9.2.4 Knowledge sharing and empowerment culture

NCSC, CERT, Error Management Programme and SOFIT all include blame culture as being an indicator of UIT and recommend instilling a culture of empowerment to counteract insider threat. Results from this study also found that alongside an empowerment culture, which influenced UIT, how knowledge was availed and shared between individuals had an impact on UIT risk levels.

This section of the framework evaluates the culture of an organisation through assessing stigma and levels of organisational communication associated to cyber incidents and breaches. It also evaluates individual's understanding of outcomes that are associated to cyber breaches, effectiveness of guidelines in the event of an attack and individual understanding of subsequent protocols, ability to challenge abnormal interactions and inter-organisational relationship dynamics. This framework recommends creating security protocols based on STAAC. As UIT is rapid in its nature and participants noticed anomalies that they ignored, STAAC is introduced to assist individuals in slowing down and thinking through their actions at various points in the cyber breach (prior or during) which can help identify, intervene and mitigate UIT. As suggested by NCSC, this framework also promotes the importance of practicing cyber breach protocols through drills (incorporating STAAC) and establishing an incident reporting culture through creating platforms (off-line and online) for knowledge sharing.

4.9.2.5 Fluctuating vulnerabilities

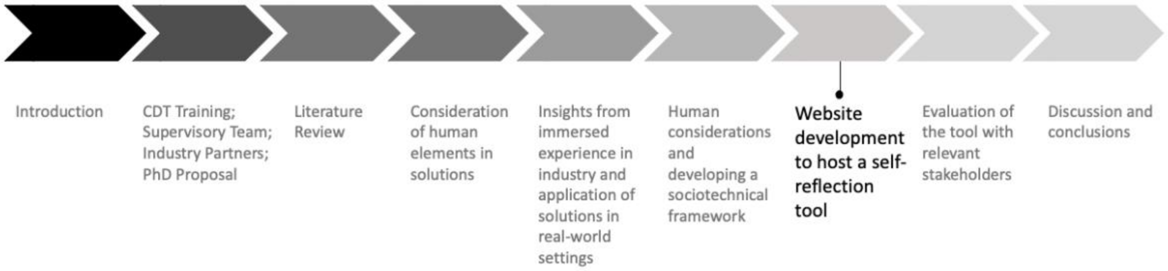
As the final pillar and further to the relevant frameworks discussed earlier in this Chapter, this framework proposes regular assessment of fluctuating vulnerabilities discussed earlier that can influence unintentional insider threat (UIT). As UIT is changing in its nature due to these fluctuating vulnerabilities, using several indicators are endorsed as part of this framework, such as evaluating attention to detail (online and physical parameter), to formulate recommendations.

4.10 Summary

This Chapter discussed the results from the research study that applied Critical Decision Method technique from the human factors domain to exclusively investigate factors that influence unintentional insider threat. Findings revealed four thematic areas known to influence unintentional insider threat namely, Decision making, Task factors, Accidents and, Organisational factors. Results aided in extending the application of the Epidemiological Triangle for understanding the dynamic relationship between vectors (human, attack and, environment) in the context of unintentional insider threat. Furthermore, findings were utilised to create a sociotechnical framework. This framework incorporated existing recommendations from extant literature that were evidenced in the data and introduced new elements which were distributed across five pillars. Additionally, passive technical defences (PCDs) were incorporated into the framework as they maintain best practices without the use of predictive modelling or personal data. The next Chapter discusses the presentation of the framework as a self-reflection tool.

5. Website Design

5. Website Design


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

The review of extant literature and the outcomes demonstrated from the onion model indicated an unequal onus being placed on the technological element to safeguard against insider threat. Insights from industry collaborations also reflected the utilisation of this element and the superimposing of cybersecurity onto existing systems that aimed to control or restrict the human element whilst alienating cybersecurity from mainstream operations. Thus, in order to change how humans are considered in systems the aim of the design was to be inclusive of diverse audiences which will be discussed in this Chapter. Inclusivity of diverse audiences meant that individuals who might or might *not* possess technical cybersecurity expertise could successfully be able to engage with the self-reflection tool through comprehending and informing the inputs presented via the tool (derived from the framework presented earlier). The utilisation of a diverse audience was believed to leverage the collective knowledge held by various stakeholders i.e. harnessing generalist intelligence to holistically inform the readiness levels against unintentional insider threat within organisations.

A website was created to host the framework in the form of a self-reflection tool as it was deemed to be an appropriate choice to enable collaboration between stakeholders, be visually appealing when evaluating a large set of defences, represent outputs such as radar graphs on the same platform and, limit confusion, deviations and errors that can stem from collaborative efforts. Finally, the aim of creating an Information Technology (IT) artefact was also to convert theory-ingrained results into a tangible output that could be evaluated by a range of relevant stakeholders. This Chapter discusses the design principles that were guided by Action Design Research to achieve said aims.

Action Design Research (ADR) inspired approach was adopted to design an information technology artefact that assisted users (technical and non-technical) in evaluating their organisational readiness levels against unintentional insider threat (UIT). For the artefact to be understood and utilised by non-technical users is of significant importance. The inclusion of non-technical users can facilitate engagement from a wider audience in the technical field of cybersecurity and specifically for this audience to be able to analyse reports pertaining to threats generated by industry evaluators. The design principles inspired by ADR and the application of this approach that occurred throughout the overarching research project will now be discussed.

5.1 Overview of Action Design Research

Design research (Reeves, 2006; Hevner et. al, 2004) adopts a technological view towards information systems artefacts and places Information Technology (IT) artefacts at the centre of its approach. Design Research includes a separate stage post the creation of the artefact to evaluate its application once it has fully been developed. However, during the time of the artefacts' creation and development, a Design Research stance can neglect organisational contexts i) within which the artefact exists, ii) which the artefact is continuously shaped by and, iii) help provide the artefact relevance.

In order to incorporate organisational contexts into the design of IT artefacts Sein et. al (2011) introduced Action Design Research (ADR). ADR aids in designing IT artefacts in their organisational contexts, during and post their development. This approach amalgamates Action Research (Avison et. al, 1999) and Design Research in order for IT artefacts to be concurrently created, introduced to and, produced within their organisational contexts and evaluated. An ADR approach involves four design stages for IT artefact creation namely, problem formulation; building, intervention and, evaluation; learning and reflection and; formalisation of learning. The first stage of ADR, guided by two principles, is concerned with identifying a problem that is experienced by organisation(s) or predicted by the researcher to occur through relevant evidence. This stage allows the framing of the problem, its scope and, potential research questions. The subsequent stage is guided by three principles and is iterative. This second stage helps inform the build of the artefact, its purpose as an intervention within the organisation and, its evaluation for effectiveness. Stage 3, reflection and learning, applies learnings to a broader class of problems. The fourth and final stage within ADR is ‘formulisation of learning’ that aims to broaden the situated learning from the earlier stages into creating solutions on a more generalised scope.

Since ADR’s introduction this approach has been applied in various domains for the design of Information System and Technology (IS/IT) artefacts. Petersson and Lundberg (2016) applied an ADR approach to the railway industry. In their article the authors utilise ADR to generate new ideas and select the most fitting solutions through an iterative process. This study engaged industry, government and, academia through the design cycles of building, testing and evaluating the artefact. Findings determined ADR to be a feasible approach and afford generalisation of findings from context-specific results.

Brooks and Alam (2015) adopted an ADR approach for the creation of an IS artefact for government (land registration ministry in Bangladesh). ADR was deemed as an appropriate choice for a complex organisational context in a developing country. The authors found that ADR could be enhanced through the introduction of an ethnographic approach being introduced for designing artefacts that are situated in complex immediate (organisational level) and overarching (societal) settings.

Within the domain of manufacturing, Hattinger and Eriksson (2015) applied ADR to the design of e-courses for experienced employees to support work-integrated learning. The design process was done iteratively and informed by an industry collaboration. Their findings suggest that the design principles pertaining to learning could be generalised more widely to the manufacturing industry.

In ‘Advocating for Action Design Research on IT Value Creation in Healthcare’, Sherer (2014) advocated for ADR to be adopted in the healthcare industry in the United States as the industry saw increased investments towards the digitisation of this sector. Another study by Maccani et. al, (2014) showed the suitability of an ADR approach to IS artefacts within the smart cities domain.

The following section discusses the application of three design processes outlined in ADR as phases 1 – 6 for the creation of a sociotechnical artefact to communicate readiness levels against unintentional insider threat (UIT) for organisations.

5.2 Method

As such, methodology for the design process to develop an IT artefact was inspired by ADR principles (Sein et. al, 2011) carried out over six phases, shown in Figure 11 below.

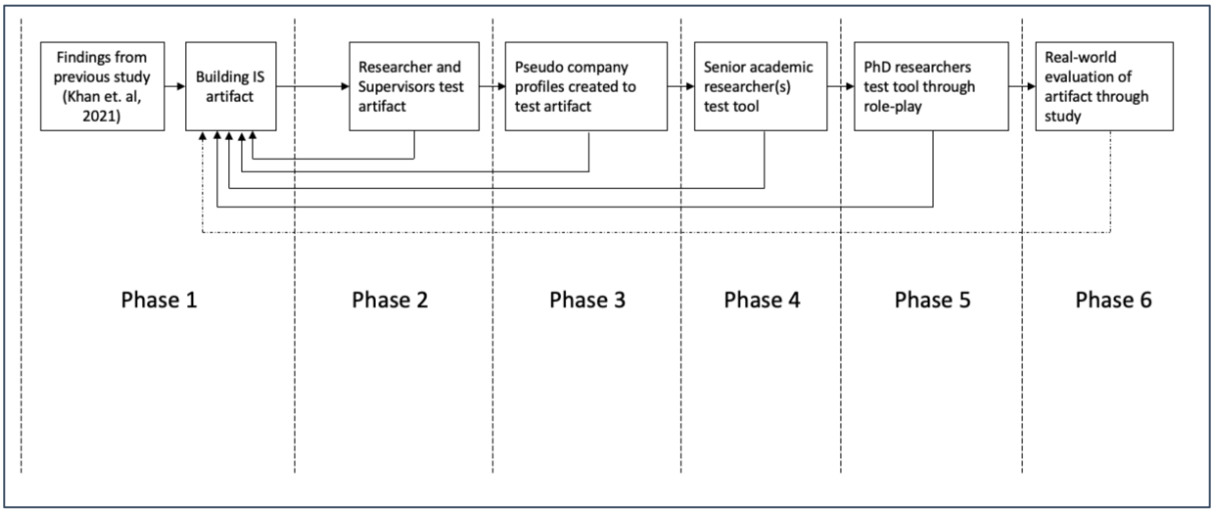

Findings from previous study (Khan et. al, 2021)
Building IS artifact
Researcher and Supervisors test artifact
Pseudo company profiles created to test artifact
Senior academic researcher(s) test tool
PhD researchers test tool through role-play
Real-world evaluation of artifact through study
Phase 1
Phase 2
Phase 3
Phase 4
Phase 5
Phase 6

Figure 11: ADR inspired phases for the design and creation of website

Stage 1 within ADR of *problem formulation* occurred from the start of the overarching project i.e. prior to the design of the website. Problem formulation was informed by academic domain experts, independent senior academics who are domain experts in cybersecurity and three industry collaborators. Stage 2 and Stage 3 of the design process to inform the *build* of the artefact, its *intervention and evaluation and reflection*, occurred iteratively over six phases shown in Figure 11. Whilst findings from qualitative dyad interviews shown in 'phase 6' to evaluate the tool are discussed in the following Chapter, feedback about the design elements were garnered post real-world application from:

1. Organic conversations between participants and/or the interviewer during the audio and video recorded sessions as participants progressed through the assessment webtool during the use of the website, and;

2. As part of a response to open-ended questions which were asked at the semi-structured interview stage following the use of the website.

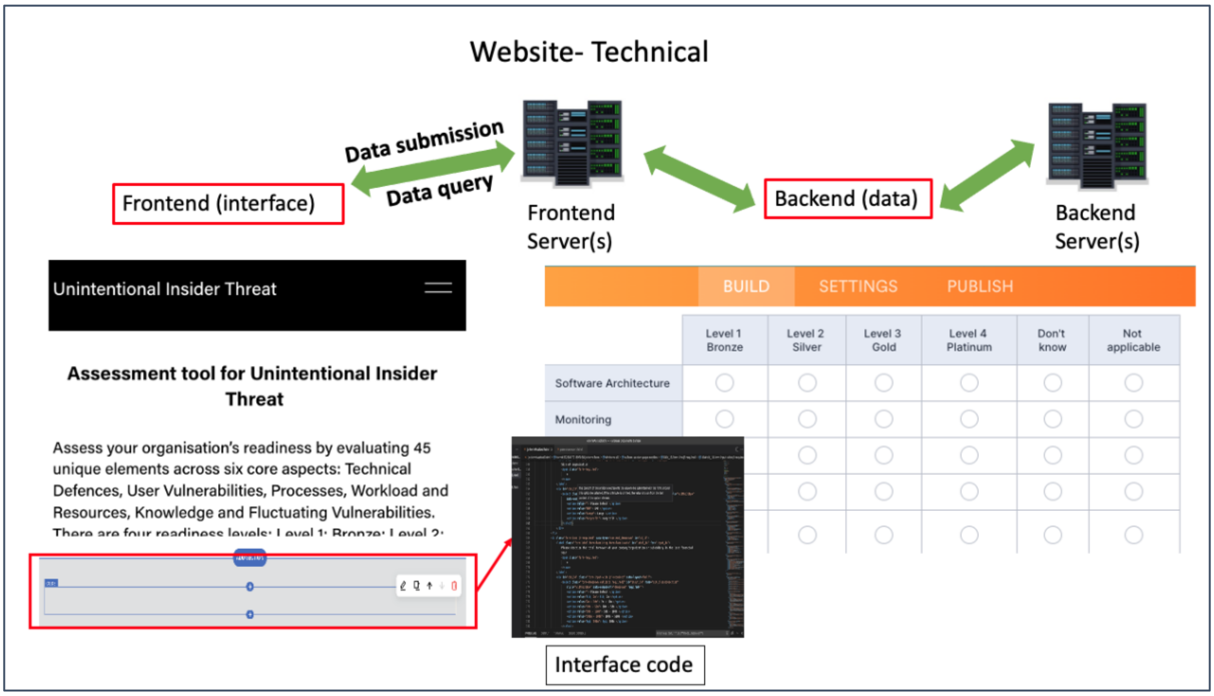
5.3 ADR inspired design process

This section discusses the design process that led to the creation of a sociotechnical IS artefact to help organisations discover, evaluate and reflect on their readiness levels against unintentional insider threat (UIT). Stage 1 within ADR, i.e. *problem formulation*, was conducted over a period of three years that involved academic ingrained and use inspired research. This research was informed by various industry partners to incorporate widely faced challenges pertaining to unintentional insider threat (UIT) i.e. *reciprocal shaping* under stage 2 of ADR. The author, supervisors and industry collaborators assumed mutually influential roles and contributed to the concurrent testing of findings as the project progressed. Efforts and findings from the above two stages informed and influenced *reflection and learning* within the project i.e. stage 3 for design. Case-specific outcomes from the Critical Decision Method (CDM) based study were generalised to a broader class of problems for understanding factors that influence UIT.

Stages 1, 2 and 3 within ADR were iteratively carried throughout the six phases depicted in Figure 11. Phases 1-5 were carried out organically as part of industry informed research that occurred during the entire length of this project. Within these phases, the author, supervisors and other academic domain experts as well as industry partners actively fed back into the design. Pseudo company profiles were developed to assess time and user interface to revise design elements. This was followed by another session with domain experts to test the use of the website which contained the assessment tool. Once feedback from this stage was incorporated into the website design another session was held with PhD researchers in group settings. Each group engaged in role-play, assuming various organisational personas to work together and test the use of the website. Feedback from these sessions were incorporated into the website design before commencing the next phase. ‘Phase 6’, which was conducted as part of a research study, also organically yielded discussions between participants and the interviewer around the design of the artefact. Findings from the research study within phase 6 can aid in the further development of the artefact as part of future research.

5.4 Website Security

As discussed above, the creation of the website was informed by six phases. Each phase informed the design and interface of the website iteratively. As the self-reflection tool required the sharing of what could be classified as *sensitive organisational data*, the tool's architecture was carefully considered and deliberated. In order to avoid compromising of this sensitive data two service providers were purchased. One service provider called SquareSpace held the interface of the website as it appeared to stakeholders (i.e. the frontend) while another service provider called Jotform was designed to host the actual data that was submitted by stakeholders (i.e. the backend). Hence, SquareSpace held user login information and their servers would submit information provided by stakeholders directly to JotForm servers (never holding the information on their servers). In order to create the outputs (i.e. radar graphs representing organisational readiness levels for defences against unintentional insider threat) SquareSpace servers were able to query JotForm servers through a valid Unique ID provided by the end user (i.e. a unique token emailed to the user upon submission). While the questions and respective options for each of the six pillars appeared correctly and completely on SquareSpace (frontend) interface, these questions and selections were represented as numerical values on JotForm servers. This partitioning of the frontend and backend is shown in Figure 12 below.



Unintentional Insider Threat

Assessment tool for Unintentional Insider Threat

Assess your organisation's readiness by evaluating 45 unique elements across six core aspects: Technical Defences, User Vulnerabilities, Processes, Workload and Resources, Knowledge and Fluctuating Vulnerabilities. There are four readiness levels: Level 1: Bronze; Level 2:

	Level 1 Bronze	Level 2 Silver	Level 3 Gold	Level 4 Platinum	Don't know	Not applicable
Software Architecture	○	○	○	○	○	○
Monitoring	○	○	○	○	○	○

Figure 12: Partitioning of the frontend and backend

Thus, this created a partition between the frontend and the backend whereby the data entered by stakeholders stored on providers' respective servers would appear to be gibberish if either of the providers were independently compromised. Since any tracking abilities were disabled on SquareSpace (such as cookies and IP addresses) stakeholders were advised not to navigate away from the assessment tool page as any data entered would be lost. To further bolster these aspects of security, only a single stakeholder was requested to register with SquareSpace and advised to do so with an account that contained a generic domain (such as Gmail, Yahoo or Hotmail).

5.5 Website Interface

This section presents the website interface that contained the self-reflection tool derived from the framework. The website contained three core pages, shown in Figure 13 below. This was purposely kept to a minimum amount of pages so as not to overwhelm the non-technical audience. Each page held the following content:

i) Home Page: This landing page provided definitions for ‘Insider Threat’, ‘Unintentional Insider threat’, information about the ‘Assessment tool’, information about the ‘Personalised Report’, and, information about ‘Registering with the website’

ii) Assessment Tool Page: This page provided an introduction to the tool, a brief questionnaire about the organisation (e.g. “Size of the organisation”, “Please describe the nature of your customers”, “Please indicate your current function” and “Please indicate the management level of your current position”) with multiple options as a drop-down list for response. This page also included six colour coded pillars which reflected the readiness levels on a four point scale against each of the factors that were found to influence unintentional insider threat, a three-point semantic confidence scale for each of the inputs to capture participant confidence level when selecting their organisational readiness levels for each of the inputs and, a ‘submit’ button for the assessment tool when completed;

iii) Personalised Report Page: This page contained a unique ID button and displayed the strength of organisational defences against unintentional insider threat (UIT) based on participants’ choices. Strength of defences were represented through output visualisations such as radar and bar graphs for each of the pillars and a narrative box to assist participants in interpreting outputs.

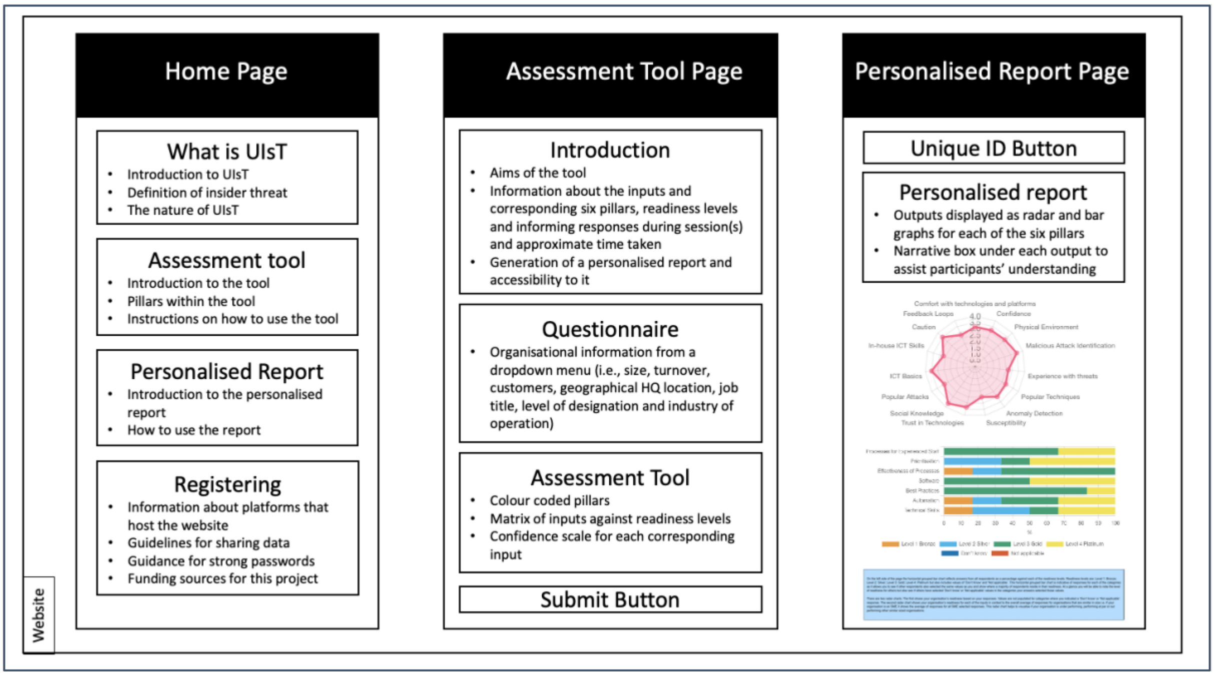
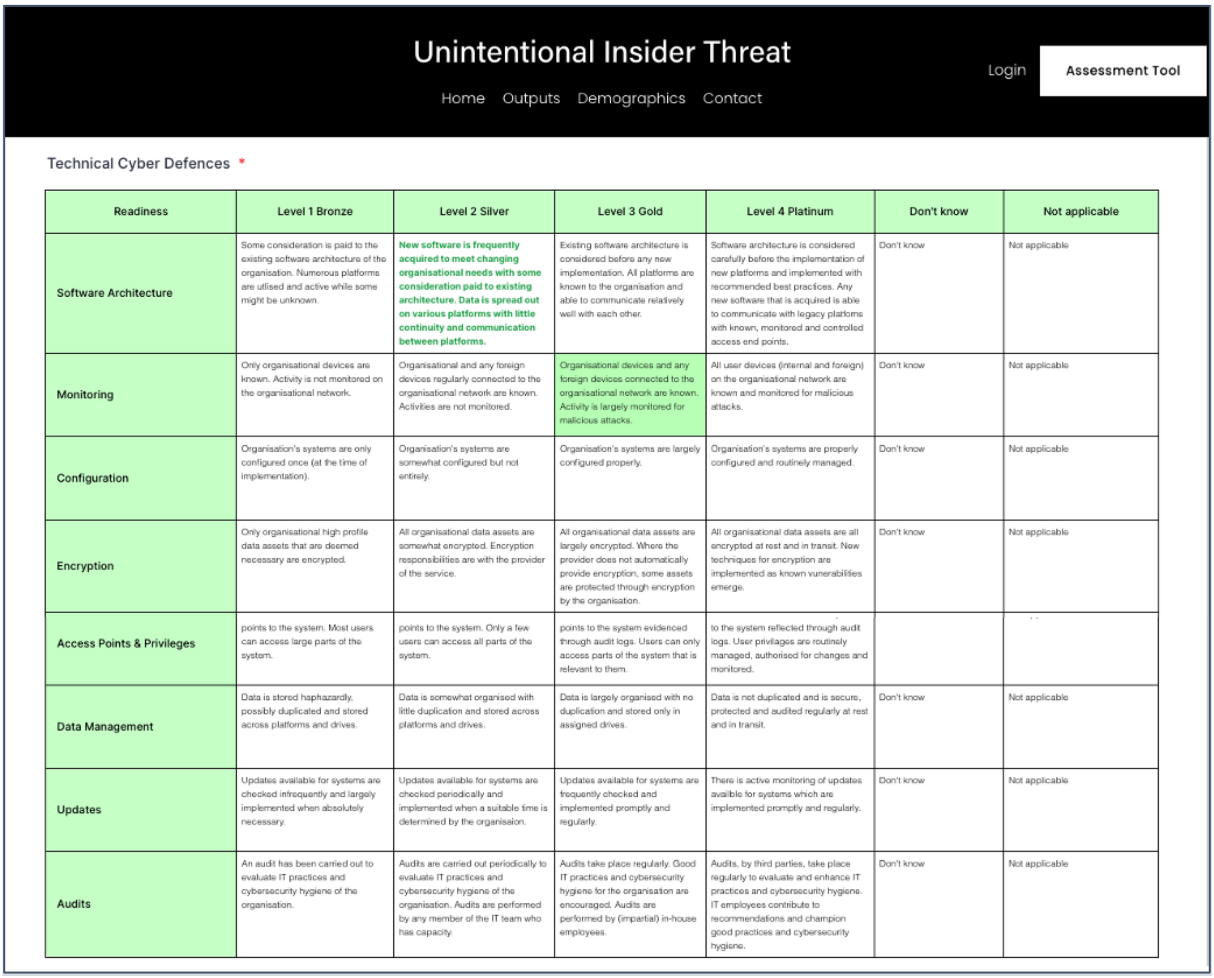

Website
Home Page
What is UIsT
Introduction to UIsT
Definition of insider threat
The nature of UIsT
Assessment tool
Introduction to the tool
Pillars within the tool
Instructions on how to use the tool
Personalised Report
Introduction to the personalised report
How to use the report
Registering
Information about platforms that host the website
Guidelines for sharing data
Guidance for strong passwords
Funding sources for this project
Assessment Tool Page
Introduction
Aims of the tool
Information about the inputs and corresponding six pillars, readiness levels and informing responses during session(s) and approximate time taken
Generation of a personalised report and accessibility to it
Questionnaire
Organisational information from a dropdown menu (i.e., size, turnover, customers, geographical HQ location, job title, level of designation and industry of operation)
Assessment Tool
Colour coded pillars
Matrix of inputs against readiness levels
Confidence scale for each corresponding input
Submit Button
Personalised Report Page
Unique ID Button
Personalised report
Outputs displayed as radar and bar graphs for each of the six pillars
Narrative box under each output to assist participants' understanding
Comfort with technologies and platforms
Feedback Loops
Confidence
Caution
Physical Environment
In-house ICT Skills
Malicious Attack Identification
ICT Basics
Experience with threats
Popular Attacks
Popular Techniques
Social Knowledge
Anomaly Detection
Trust in Technologies
Susceptibility
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
Processes for Experienced Staff
Prioritisation
Effectiveness of Processes
Software
Best Practices
Automation
Technical Skills
0
10
20
30
40
50
60
70
80
90
100
%
Level 1 Bronze
Level 2 Silver
Level 3 Gold
Level 4 Platinum
Don't know
Not applicable

Figure 13: Website design containing three core pages

Framework from the Critical Decision Method (CDM) research study was used to inform the design of the assessment tool. Each input listed in the framework was assigned four gradient levels (level one being the baseline measure which was developed in line with recommendations by NCSC, CERT and, previous research by Khan et. al 2021 and viewed as the minimum recommended level of readiness for organisations, to level four which was the best possible level that can be implemented by organisations). In addition, based on the findings from the CDM research study and immersed industry experiences of the author, non-technical language was used to express each input so it is comprehensible to and inclusive of lay audiences that might otherwise be excluded from the cybersecurity domain.

Once the stakeholders had read the information provided in the Homepage, they were directed to the Assessment Tool page. On this page stakeholders were presented with a brief questionnaire about their organisation that contained multiple choice options from drop-down lists. Upon scrolling down from the questionnaire, this page of the website presented the six pillars derived from the framework. The questionnaire and the six pillars were hosted on the same page to avoid confusion as users collaborated and navigated through various aspects in the assessment tool that captured inputs. Pillars were colour coded for the ease of identifying progression through the framework (for instance the Technical pillar was colour coded green). When users hovered their cursor over an option the box was highlighted in green colour to indicate the placing of the cursor to other collaborators. One an option was selected, the text colour turned green to indicate a selection had been made. An example of this is shown in Figure 14 below that presents the first pillar labelled *Technical Cyber Defences*, the first input's answer has been selected and the second input has the cursor hovered over it.

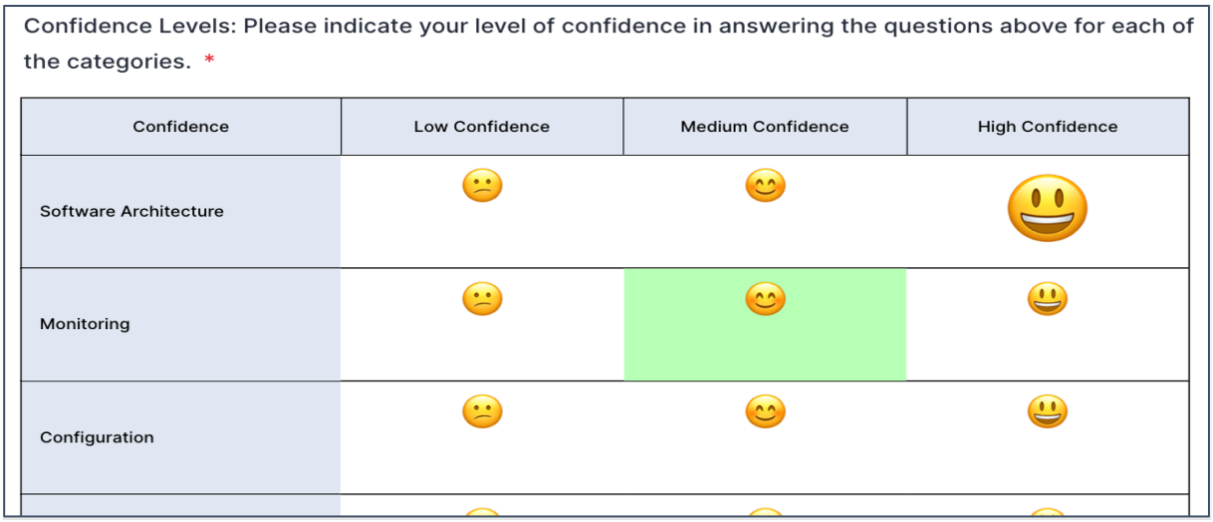


Technical Cyber Defences *

Readiness	Level 1 Bronze	Level 2 Silver	Level 3 Gold	Level 4 Platinum	Don't know	Not applicable
Software Architecture	Some consideration is paid to the existing software architecture of the organisation. Numerous platforms are utlised and active while some might be unknown.	New software is frequently acquired to meet changing organisational needs with some consideration paid to existing architecture. Data is spread out on various platforms with little continuity and communication between platforms.	Existing software architecture is considered before any new implementation. All platforms are known to the organisation and able to communicate relatively well with each other.	Software architecture is considered carefully before the implementation of new platforms and implemented with recommended best practices. Any new software that is acquired is able to communicate with legacy platforms with known, monitored and controlled access end points.	Don't know	Not applicable
Monitoring	Only organisational devices are known. Activity is not monitored on the organisational network.	Organisational and any foreign devices regularly connected to the organisational network are known. Activities are not monitored.	Organisational devices and any foreign devices connected to the organisational network are known. Activity is largely monitored for malicious attacks.	All user devices (internal and foreign) on the organisational network are known and monitored for malicious attacks.	Don't know	Not applicable
Configuration	Organisation's systems are only configured once (at the time of implementation).	Organisation's systems are somewhat configured but not entirely.	Organisation's systems are largely configured properly.	Organisation's systems are properly configured and routinely managed.	Don't know	Not applicable
Encryption	Only organisational high profile data assets that are deemed necessary are encrypted.	All organisational data assets are somewhat encrypted. Encryption responsibilities are with the provider of the service.	All organisational data assets are largely encrypted. Where the provider does not automatically provide encryption, some assets are protected through encryption by the organisation.	All organisational data assets are all encrypted at rest and in transit. New techniques for encryption are implemented as known vunerabilities emerge.	Don't know	Not applicable
Access Points & Privileges	points to the system. Most users can access large parts of the system.	points to the system. Only a few users can access all parts of the system.	points to the system evidenced through audit logs. Users can only access parts of the system that is relevant to them.	to the system reflected through audit logs. User privilages are routinely managed, authorised for changes and monitored.		
Data Management	Data is stored haphazardly, possibly duplicated and stored across platforms and drives.	Data is somewhat organised with little duplication and stored across platforms and drives.	Data is largely organised with no duplication and stored only in assigned drives.	Data is not duplicated and is secure, protected and audited regularly at rest and in transit.	Don't know	Not applicable
Updates	Updates available for systems are checked infrequently and largely implemented when absolutely necessary.	Updates available for systems are checked periodically and implemented when a suitable time is determined by the organisaion.	Updates available for systems are frequently checked and implemented promptly and regularly.	There is active monitoring of updates availble for systems which are implemented promptly and regularly.	Don't know	Not applicable
Audits	An audit has been carried out to evaluate IT practices and cybersecurity hygiene of the organisation.	Audits are carried out periodically to evaluate IT practices and cybersecurity hygiene of the organisation. Audits are performed by any member of the IT team who has capacity.	Audits take place regularly. Good IT practices and cybersecurity hygiene for the organisation are encouraged. Audits are performed by (impartial) in-house employees.	Audits, by third parties, take place regularly to evaluate and enhance IT practices and cybersecurity hygiene. IT employees contribute to recommendations and champion good practices and cybersecurity hygiene.	Don't know	Not applicable

Figure 14: Colour coded pillar interface

Each pillar was subsequently followed by a confidence indicator which was a three point semantic scale to reflect stakeholders' confidence in their selection. This part aimed to capture how confident stakeholders felt in their responses to each of the inputs. It was also aimed to make selections for each input easier as it provided stakeholders with the flexibility to select a higher input level but indicate a lower confidence level or vice versa where they could select a lower input level and indicate high confidence in their choice. Similar to the pillar interface, the corresponding confidence scale would also highlight green to indicate the position of the cursor. The emoji representing the confidence level would be magnified once a selection had been made to indicate the receival of input. This is shown in Figure 15 below.

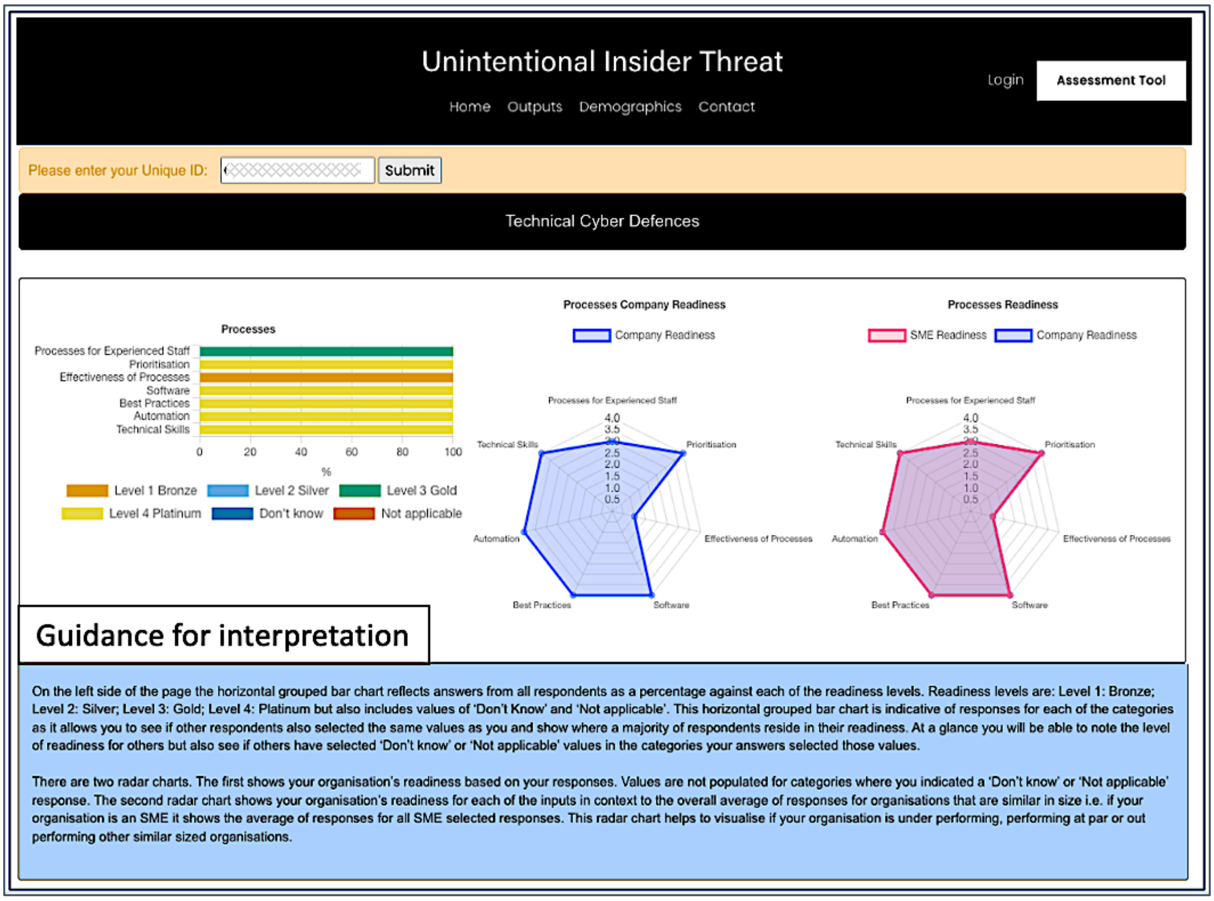


Confidence Levels: Please indicate your level of confidence in answering the questions above for each of the categories. *

Confidence	Low Confidence	Medium Confidence	High Confidence
Software Architecture	😕	😊	😃
Monitoring	😕	😊	😃
Configuration	😕	😊	😃

Figure 15: Interface for confidence indicators

All inputs were designed to require a response from the stakeholder so as to avoid unintentional skipping of inputs (non-responses). If an input had not been selected and the submit button was pressed, the interface would move the page to the position the input was missing and highlight it in red colour. Once all responses for each input had been done and the submit button was pressed, the website was redirected to the Personalised Report Page and the stakeholder was sent an email notification containing a Unique ID. Once this was entered, the page would display outputs as radar graphs for each of the pillars (Figure 16 below). Radar graphs were selected to represent outputs due to their simplicity for communicating the strength of defences so non-technical users do not get overwhelmed when interpreting the outputs.



Guidance for interpretation

On the left side of the page the horizontal grouped bar chart reflects answers from all respondents as a percentage against each of the readiness levels. Readiness levels are: Level 1: Bronze; Level 2: Silver; Level 3: Gold; Level 4: Platinum but also includes values of 'Don't Know' and 'Not applicable'. This horizontal grouped bar chart is indicative of responses for each of the categories as it allows you to see if other respondents also selected the same values as you and show where a majority of respondents reside in their readiness. At a glance you will be able to note the level of readiness for others but also see if others have selected 'Don't know' or 'Not applicable' values in the categories your answers selected those values.

There are two radar charts. The first shows your organisation's readiness based on your responses. Values are not populated for categories where you indicated a 'Don't know' or 'Not applicable' response. The second radar chart shows your organisation's readiness for each of the inputs in context to the overall average of responses for organisations that are similar in size i.e. if your organisation is an SME it shows the average of responses for all SME selected responses. This radar chart helps to visualise if your organisation is under performing, performing at par or out performing other similar sized organisations.

Figure 16: Output Interface

Our findings reflect that ADR inspired website creation was positively received by stakeholders in various sectors at technical and non-technical designations. In order to tackle unintentional insider threat it is important to design solutions that are usable and informed by a mixture of audiences.

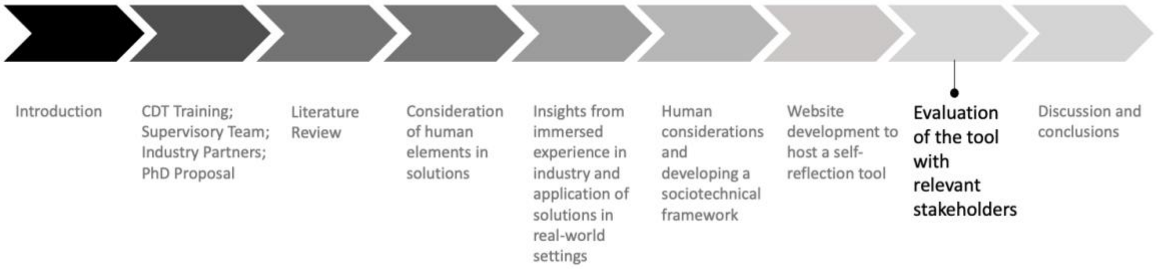
5.6 Summary

For the design of artefacts that examine, assess and, engage within specialised fields, Action Design Research stages inspired the creation of an inclusive website design within this research project which was discussed in this Chapter. Guided by Action Design Research principles domain experts and industry personnel were utilised to formulate the problem being addressed and reciprocally inform the design through six iterative phases that were carried out over a span of four years. Through *reflection and learnings* the design was enhanced as the website was progressively developed.

The website was evidenced to be easily understandable to audiences outside the domain of cybersecurity and for individuals with non-IT background. This was achieved through the use of inclusive language to communicate level descriptors for inputs and leveraging visualisation techniques such as radar graphs to communicate the strength of organisational defences as outputs. The inclusive design incorporated into the website also harnessed collective knowledge held by stakeholders as they collaborated to inform their selections. The element of incorporating diverse audiences to engage with the self-reflection tool was an important aspect because in order to change how humans are considered in systems, a diverse range of individuals (from outside the computer science domain) must first be engaged to increase the understanding of the role of human considerations pertaining to unintentional insider threat. Here, the involvement of a diverse audience is distinct from the way personnel (such as from Human Resources Department) are included as part of psychological and behavioural approaches discussed earlier. The inclusion of diverse audiences in existing literature is done with an aim to report abnormal behaviour or to contribute towards inputs that develop individual profiles. Insights from industry evidenced that it becomes problematic when cybersecurity is superimposed onto existing systems or alienated from mainstream operations. In contrast, stakeholders that engaged with the self-reflection tool were in attendance to develop their understanding of the defences in the context of *work-as-done* and to equally assess and recognise aspects that work well (learning from success as well as failures) as outlined in the Safety II approach. It is acknowledged that additional legends to support understanding of various visualisations could further improve the existing design of the website. However, improvements intrinsically entail trade-offs to be made between a clear layout and cognitively assisting users in their understandings. Having developed this tool, the next Chapter details the evaluation of this tool.

6. Evaluation of a self-reflection tool

6. Evaluation of a self-reflection tool


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

Once a framework had been developed, the previous Chapter discussed Action Design Research principles that guided the creation of a self-reflection tool which was hosted via a website.

This Chapter discusses the evaluation of this tool with relevant stakeholders through a research study that is inspired by the Theory of Planned Behaviour. The aim of this study is to explore the extent to which the developed framework can have a positive impact in an open environment for understanding unintentional insider threat. Thus, senior leaders from industry were invited to engage with the website for three-hour sessions and used the framework to identify where they believed their organisation lay in terms of readiness against unintentional insider threat. Attitudes were assessed through semantic scales and semi-structured interviews pre- and post-sessions to evaluate the impact of the self-reflection tool.

6.1 Approach

Work discussed in Chapter 4 involved the application of Critical Decision Method (Klein et al., 1989) to the accounts of individuals who had inadvertently been exposed to cyberbreaches. The findings were organised to formulate six main pillars of a framework: Technical, User Vulnerabilities, Processes, Workload and Resources, Knowledge and, Fluctuating Vulnerabilities. Additionally, a new set of linked concepts for security improvement were identified including honing in expert staff when devising processes, monitoring factors linked to time pressures, known channels of knowledge attainment and sharing and, monitoring fluctuating vulnerabilities linked to unintentional insider threat (UIT). With an Action Design Research inspired approach a website was developed to communicate the findings represented in the framework. Subsequently, in the design of this study, the main point of interested was to evaluate the positive impact the tool could offer in open environments when understanding the role of human considerations for unintentional insider threat and guide decision making by stakeholders. Inspiration was also taken from previous theory-to-practice ventures including the use of Swiss Cheese Metaphor (Reason, 1990a) for accident causation taking form as organisational decision making tool.

In order to understand how the framework informed perceptions of organisational status around unintentional insider threat, Theory of Planned Behaviour (Ajzen, 1991) was adopted as an analytic technique. This theory proposes three independent determinants to predict certain behaviours to achieve desired outcomes namely, *attitude* towards the behaviour, *subjective norms* and, *perceived behaviour control*. Theory of Planned Behaviour has been applied in several domains including environmental sciences, occupational health domain, management (Bosnjak et al., 2020) leisure choice (Ajzen and Driver, 1992) and, continuing higher education (Ingram et al., 2015).

The three determinants that indicate changes in behaviour were selected to investigate whether the tool can serve to have a positive impact on understanding and tackling unintentional insider threat for organisations of various sizes. Attitudes, subjective norms and, perceived behaviour control were measured within participants prior to and post interacting with the web based self-reflection tool, which evaluated factors that influence unintentional insider threat (UIT) within organisations. Various methods were considered to measure the three elements in the Theory of Planned Behaviour such as, focus group interviews (Rabiee, 2004) as participants from various organisations would jointly engage with the tool in sessions, structured interviews (Rogers, 2008) following interaction with the tool, circulating the tool to numerous experts who could then independently engage with the tool in an unmonitored setting and return questionnaire to the author (Rowley, 2014) or be a part of a semi-structed interview (Madill, 2011) in a one-to-one setting to share their experiences. However, these were deemed unsuitable as valuable insights might be lost as participants engaged with the tool or limit the richness of data resulting from semi-structured collaborative settings. Thus, the measurement of these three elements i.e. attitude towards the behaviour, subjective norms and, perceived behaviour control were carried out through the application of metacognition scaffolding technique (Jumaat and Tasir, 2014), seven point semantic scales, feeling thermometers and, open ended questions (Rydell and Macconnell, 2006) utilised as part of semi-structured interviews. Participants were also asked to reflect and verbally share any new learnings and assess their progress post their interaction with the website. This resulted in rich data garnered from in-depth conversations.

6.2 Method

6.2.1 Participants

Once the study had been approved by the Engineering Department Ethics Board, thirteen participants were recruited from six organisations. The call was shared via the help of multiple National Cyber Security Centre’s (NCSC) ‘Industry 100’ organisational partners and to the first author’s professional network. All participants from the same organisation were present concurrently during their respective session(s) and provided with an information sheet (Appendix 11). As shown in Appendix 12, six participants held a Chief Executive

Officer or equivalent designation, six participants were at a Director, Head or equivalent designation, while one participant held a Manager or equivalent title. Two participating organisations were small or medium-sized enterprise (SMEs), two were large and, two were non-profit organisations. Four organisations indicated the primarily nature of their customers as being ‘business-to-business’ (B2B) whilst two indicated ‘business-to-consumers’ (B2C) where products or services are taken directly to the end user. Organisational headquarters (HQ) were either located in the United Kingdom or Americas region.

6.2.2 Data Collection

Nine sessions took place between January and March 2022; the longest session was 03 hours 03 minutes, with the shortest session 1 hour 55 minutes long. Sessions cumulatively generated approximately 14 hours and 30 minutes of data. All sessions contained dyad participants with the exception of ‘company reference 5’, which contained three participants in their session(s).

Data were collected in four stages. Initially, participants were requested to fill out a questionnaire which contained free text fields and a seven point semantic scale prior to engaging in the session. Following this initial stage, data was captured during the session through audio and video recording as well as through the website which contained information about participants’ organisation, recorded their selections for readiness levels against unintentional insider threat (UIT) and confidence levels on a three point semantic scale. Once participants had interacted with the website which contained 45 inputs across six pillars and studied their organisation’s personalised report, participants were asked to fill out another questionnaire, similar to the one they had filled out prior to the session. Lastly, data was collected through semi-structured interviews conducted in the latter half of the session(s).

6.2.3 Session Design

6.2.3.1 Prior to the session

Participants filled out a short questionnaire sheet containing sixteen questions. Three questions contained a free text field for participants' responses. For the remainder of the thirteen questions, participants were asked to select their agreement levels pertaining to statements shown in the question on a seven point semantic scale that ranged from "Strongly Disagree" to "Strongly Agree". Full set of questions within the questionnaire sheet that measured participants' attitudes (Q1-Q6), subjective norms (Q7-Q11) and perceived behaviour control (Q12-Q16) as outlined in the Theory of Planned Behaviour are presented in Appendix 13.

6.2.3.2 During the session

Participants were reminded of the nature, purpose and duration of this study, content within the information sheet, confidentiality and anonymity offered to participants, handling and use of data generated from session(s) and, their right to withdraw at any time during or post their session(s). Participants were subsequently informed of the structure of the session, including ten minute breaks on the hour if needed. A flow diagram of activities within the session(s) are shown in Figure 17 below.

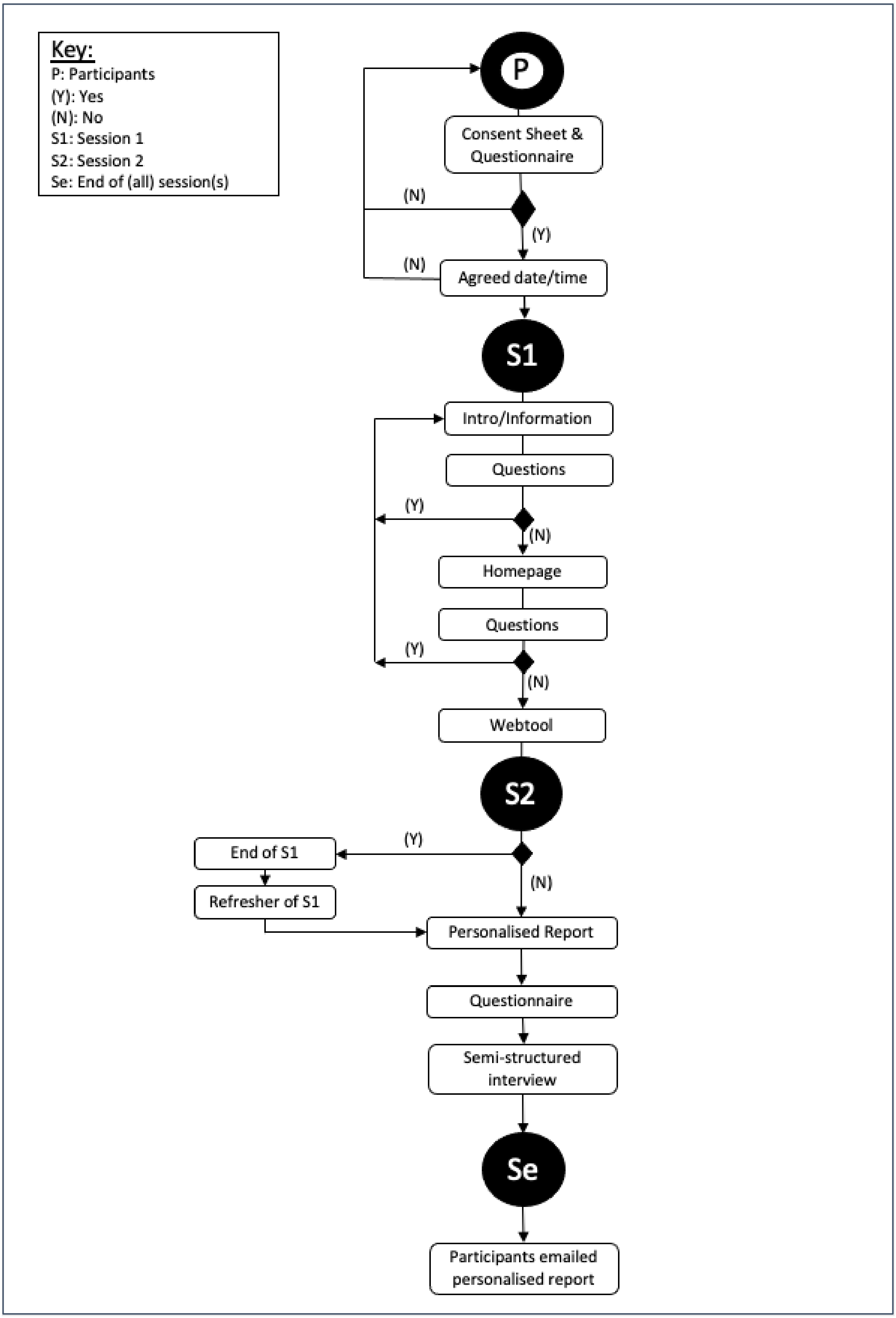


Figure 17: Flow diagram of activities within session(s)

The session began with the senior stakeholder from each organisation sharing their screen. Accompanying participants and the author conducting the session (referred to as ‘IV’) confirmed they could see and read the content being displayed. The senior stakeholder was directed to a website which contained three core pages; Home Page, Assessment Tool Page, and Personalised Report Page. Participants were given time to read over the information presented in the Home page of the website before progressing on to the Assessment Tool page. It took participants 55 minutes on average to complete the assessment tool with the shortest time being 30 minutes and the longest being 1 hour and 15 minutes. All attending participants from the same organisation were requested to agree on the readiness level for their organisation for each of the inputs, with the senior stakeholder having the final decision. Due to the nature of the business for some of the participating organisations, all participants were informed that they did not need to share any confidential information or reasoning for their selections in front of the interviewer if they did not feel comfortable. However, participants appeared to be comfortable and engaged in ‘thinking out loud’ technique as they progressed through their selections.

After participants submitted their choices they were given time to independently look at and interpret the personalised report generated for their respective organisations. Once the personalised report had been interpreted, participants were requested to fill out and return another short questionnaire sheet containing sixteen questions, similar to the one that was filled out prior to the session, shown in Appendix 14 (attitudes: Q1-Q6; subjective norms: Q7-Q11; perceived behaviour control: Q12-16). Once the questionnaire sheets were returned, participants were asked nine open-ended questions as part of a semi-structured group interview, presented in Appendix 15 (metacognition scaffolding technique: Q1-2; Q4-9; perceived behaviour control: Q3). Semi-structured interview durations were 54 minutes long on average with the longest interview lasting 1 hour 15 minutes, with the shortest being 34 minutes long.

6.2.3.3 Post session

Participants received an email of gratitude for their time which contained a PDF copy of their personalised report and authors' contact details should they have any queries in the future.

6.3 Data Processing

Quantitative data from questionnaires circulated to participants prior to and during their session(s) was compiled in a table format and colour coded by themes to respectively represent the question's category (i.e. attitudes, subjective norms and, control). Any changes to the type of words used to answer free text fields were observed. Responses to the semantic scales were averaged, the range calculated (i.e. the lowest and highest selected points within the semantic scale prior and during sessions) and, any outliers were noted. The range in responses was noted in order to provide context of overall change in behaviour and attitudes during session(s). This is shown in Appendix 16 along with author notes to assist in the interpretation of findings. This analysis and subsequent conclusion was done by the author of this thesis and independently validated by supervisors of this project.

Session recordings were transcribed verbatim. All transcripts were redacted and anonymised in parts of the transcript that were deemed appropriate. Clean transcripts were uploaded to QSR-NVivo software for coding the qualitative data. Template analysis (King, 2012) was used whereby a template containing broad themes i.e. parent nodes along with nested child nodes were utilised to code data (top-down approach). With the application of grounded theory to the transcripts, the initial template was updated with new parent and child nodes as they emerged (bottom-up approach). Template analysis provides researchers flexibility in procedures for data gathering and analysis to match their own objectives. This approach also affords time efficiency compared to other approaches (e.g. interpretive phenomenological analysis or IPA, Smith et al., 2012), is not infused with a particular methodological or theoretical position and, allows researchers flexibility within the coding structure.

This section discussed the design of a research study inspired by the Theory of Planned Behaviour. The aim of this research study was to explore the positive impact the self-reflection tool could have on relevant stakeholders – specifically to improve their understanding of human considerations in systems, guide their decision making and, acknowledge aspects that are strong in order to replicate that success which is aligned to a Safety II approach. Thus, senior leaders from industry were invited through the National Cyber Security Centre's associated partners and the author's professional network to engage with the website for three-hour sessions and used the tool to identify where they believed their organisation lay in terms of readiness levels against unintentional insider threat. Through the use of semantic scales and semi-structured interviews participant attitudes, subjective norms and, perceived control around identified cybersecurity issues were assessed. The next section discusses the analysis of the data and subsequent results that emerged from this research study.

6.4 Analysis

In line with template analysis, the initial template was created and updated through a grounded approach as the analysis progressed. Subsequently seven themes that emerged from the data, shown in Figure 18 below, were indicative of attitudes, subjective norms and, perceived behaviour control (Ajzen, 1991) as well as elements of reflective learning amongst the participants post their interaction with the website.

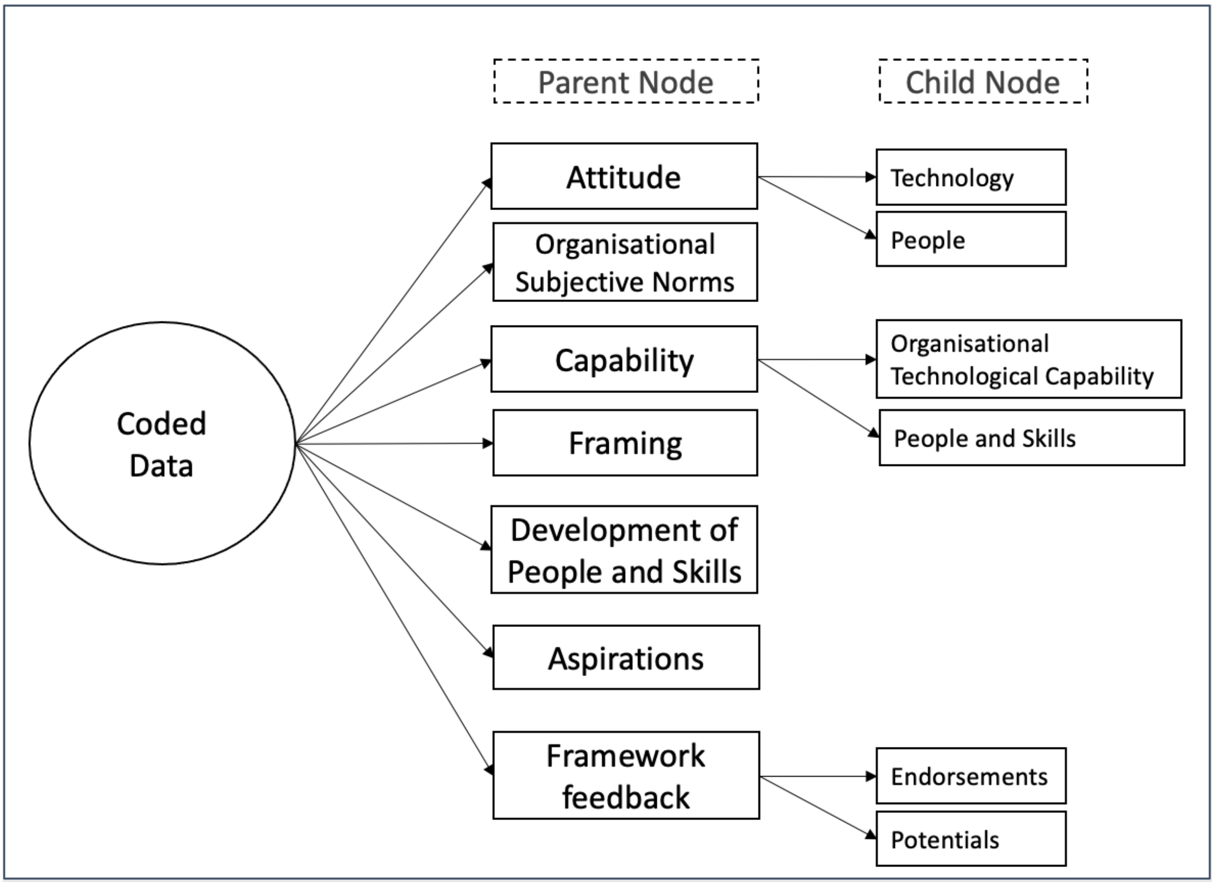


Figure 18: Seven data themes from research study

The three parent nodes derived from the Theory of Planned Behaviour were: Attitude reflective of participants' beliefs regarding technology and people; Organisational subjective norms which was indicative of participants' normative beliefs and their motivations to comply in their respective organisations and; Capability which was indicative of perceived control reflected through participants' beliefs regarding opportunities, resources and self-efficacy at an individual and organisational level. The remaining parent nodes, i.e. Framing, Development of people and skills, Aspirations and, Framework feedback, were subsequently generated through the application of metacognition scaffolding technique (Jumaat and Tasir, 2014) whereby participants reflected on their experiences, understandings and learnings during their session(s). Parent and child node frequencies are shown in Appendix 17.

Participant details are shown in Appendix 12. For the purposes of anonymisation participant genders were changed to '[[she/he/they]]' and in some parts of transcripts changed to 'she' when referring to others. The job titles held by participants were anonymised through equivalating their titles to similar positions. Throughout this Chapter participants are referred through the following system: pseudonym initials as presented in Appendix 12, company number and, designation seniority. For example, two participants from company reference 1 are referred to as "AL:1:1" and "GL:1:2". Similarly, participants from company reference 5 are referred to as "MK:5:1", "ST:5:2" and, "DS:5:3".

As is with methods, limitations of this research study emerge from targeted sampling (Watters and Biernacki, 1989) as this study was not open to the general public as the audience was difficult to reach to the social stigma associated to discussions about cybersecurity practices and weaknesses. As participants were recruited through NCSC and the author, they might possess more knowledge and awareness perhaps compared to other organisations. As organisations contacted the author, these participants were interested in cybersecurity and/or had strategically prioritised it which is reflected in the availability and participation of senior leadership staff. Whilst senior staff are a suitable sample, they might have limited visibility of day-to-day activity and consequently, some aspects of *work-as-done*. Finally, related to most results derived from qualitative research, findings can be said to be true for the state the organisations were in at the time of the study.

6.5 Attitudes – *'People are a strong defence but technology is a known friend'*

Participants' attitudes were captured prior to and during their session(s) which was discussed in detail in the previous Chapter. Attitudes reflected participants' beliefs regarding technology and people in the context of unintentional insider threat (UIT) and cybersecurity more widely.

6.5.1 Technology – *The main contender for defences despite known limitations and unequal application*

Analysis from the qualitative data showed a lack of confidence in certain technological defences and techniques such as those pertaining to encryption and phishing. Historically, a technical defence such as encryption has enjoyed immense popularity and extensive application in industry for end-use products taken to market. Similarly, phishing simulations received mixed and often contrary guidance from regulators. This is reflected in the quote below.

'*We did look at doing simulated [phishing attacks] but actually I pulled back on some of the effort we were putting into this because [[UK government]] don't advise on attack simulation in their latest guidance.*' (BP:2:1)

All participants were confident in the state of their technological defences and their organisational readiness levels (whether weak or strong) in their discussions. Participants also reflected an attitude of prioritising some assets and systems over others. This is evidenced in the following quote.

'*Yeah. There's also tiers of that [protection] as well. So, I mean, the stuff that [[team name]] [does] is the critical vulnerability stop, so we soon hear about that. There's also the less critical stuff, the updates and things.*' (DS:5:3)

Consequently some systems and assets, usually newer ones, were better protected than others resulting in an unequal application of defences across technologies. Participants' attitudes also showed a preference for using technology as opposed to humans to create and maintain defences. This is reflected in the following extract.

‘*One thing you might not be aware of [[ST]], is that we do quite in depth penetration tests and [[exercise name]] that does establish some understanding of our hygiene. So because it's quite a [[controlled]] environment, we take the approach of talking to the computers rather than the individuals.*’ (MK:5:1)

6.5.2 People – *There is strong faith in others but against the backdrop of human fallibility*

Data from quantitative questions reflected a change in participant attitudes after using the tool with an increased belief in unintentional insider threat (UIT) levels being static over short periods of time (shown in Appendix 16). Participants’ attitudes also appeared to positively shift away from believing that UIT arose as a direct consequence of users deviating from prescribed processes.

Qualitative data showed participants having trust in other people’s skills due to close-knit relationship structures that existed within their organisations, for instance in the quote below.

‘*On the other end of the scale, do they [users] feel that, you know, everyone's there, ready [to help]. That it's not just them, there's a bunch of people [to help]. I'm not going to go into process because we’ve just done processes [input] but there’s at least lots of other people around who can who can help.*’ (MM:3:1)

This structure was reflected through discussions around strong peer-to-peer support being available, informal training being offered, knowing individual skillsets and, knowledge sharing that occurred within small teams at large organisations or within smaller organisations. Overall, however, participants demonstrated low confidence in the effectiveness of formalised and external training programmes for themselves and others.

Datasets showed mixed attitudes towards communications. Participants believed people possessed good communication skills by communicating with each other and contributing openly and effectively towards enhancing and developing processes. However, participants showed a negative attitude towards people oversharing or communicating negative information (such as feedback pertaining to flaws, limitations or challenging of processes, ideas or technologies), for instance the following exchange between two participants.

'*I think we should change that [input] to 'users will, 'without fail', report processes that they consider to be inaccurate' [laughter] (KK: In this organisation definitely) [laughter].*' (GH:6:1; KK:6:2)

Participants showed a positive attitude towards empowering others, evidenced in the following dialogue between HR:3:2 and MM:3:1.

'*HR: Well. If I take the big three things that we were dealing with, I certainly would say gold [level 3]. Look at, I'm not talking about the implementation yet, right, but (MM: Learning) we didn't stigmatize anybody for SolarWinds (MM: Right) or [[another example]] or anything else. So from that perspective, we spent a hell lot of time to look at what gets better. And you know, now since we've had so many of those, I think you know. I don't think its platinum [level 4] yet because it's not across the whole organisation with everybody but it's [getting to a learning culture]. I would argue it's certainly gold [level 3].*

MM: Yeah. And I think in some cases, I'm sure there's stuff going on where people think, 'I'd rather not say anything' but, I think that's more the exception now.' (HR:3:2; MM:3:1)

This was also evidenced in discussions that emphasised others' learning, not stigmatising mistakes, being able to ask questions to understand wider strategic objectives, not having reprimands as consequences and, being able to solicit the right people for advice. There was a positive attitude of empathy amongst our participants towards others when discussing the occurrence of mistakes, unintentional errors, lack of experience or lack of knowledge.

'*I wouldn't necessarily say, depending on what's going awry, [that] everyone knows how to protect themselves and thereby, sits by using the standard safeguards that are already on the system.'* (ST:5:2)

6.6 Organisational subjective norms – *Continuous improvement*

Datasets within this theme reflected participants' subjective norms. In line with the Theory of Planned Behaviour, data codes revealed that subjective norms amongst our participants constituted of two aspects: normative beliefs within participants' organisations and, participants' motivation to comply, both of which were shaped by others around them.

All responses within the questionnaire pertaining to this theme showed a change in participants' ratings prior to and post session(s). After using the assessment tool participants indicated greater inclination towards knowledge sharing for cybersecurity practices and near-miss experiences within their organisations. While participants initially indicated that it was 'everyone's responsibility' at the organisation to be aware of cybersecurity challenges subsequently they did not indicate a wider organisational interest for insights gained during the session(s). Participants' subjective norms indicated that they intended to take action to strengthen certain defences identified in the personalised report, in line with participants' earlier indication that everyone was able to take action if a cybersecurity vulnerability was identified. Participants indicated that it was expected of them to be responsible for organisational cybersecurity by others and senior management would support their initiatives to strengthen cyber defences. Prior to a session(s) participants indicated a strong inclination to consider workload, procedures and, resources at their organisations when creating cybersecurity practices which was reflective of their subjective norms. Following their session(s) participants showed an increased inclination for this consideration to continue in the future.

Participants subjective norms indicated prestige associated to their organisations within the qualitative datasets. For instance the quote below.

'*Susceptibility [laughter] Because like [...] we rank everyone [other organisations] as well.*' (RR:4:1)

This positioning meant that participants viewed their organisations as leaders within their respective sectors of operation. Subsequently, participants' social norms indicated upholding high standards of conduct, practices and, delivery with little room for errors. Participants enjoyed a shared sense of pride associated to organisational prestige and indicated social norms for active risk awareness (such as reputational risks) that resulted from this organisational positioning.

Participants' discussions revealed that certain parts of the organisation (departments and/or processes) were understood to be performing better than others. This understanding meant that organisations were able to critically appreciate their areas of strength and limitations which is evidenced in the quote below.

'*I think it's because the risk is low [data asset value or penalties] and maybe this is the reason why it is what it is for [[department name]]. Because the data that we [that department] do hold, isn't classified as sensitive data. Whereas, you mentioned [[SN's Department]], obviously we hold sensitive data and therefore it's better protected (BP: Yeah) So it's all about proportionate risk, (BP: Yeah) what type of data do we hold? How much of it do we actually hold. So I'm still of the opinion that for the nature of the business, the amount* *of data, the type of data that we hold we are, yeah, we're doing enough [laughter] (BP: Okay).*' (SN:2:2; BP:2:1)

Participants' subjective norms showed an organisational focus to discover, understand and, improve mutually understood limitations. This subjective norm is reflected in the following quote when participants were reflecting on their organisational personalised report.

'*We have too many processes and not all of them, you know, some of them aren't fit for purpose. That may be harsh, but we need to improve our processes and that's for sure (RR: Right) And that [weakness in processes] is shining through here [in the personalised report output].*' (KL:4:2; RR:4:1)

Furthermore, participants' context was framed by their organisations being composed of highly skilled people within their industry sectors, which is evidenced in SN's quote below. This context appeared to aid participants in gaining a deep understanding of their organisational strengths and limitations.

'*Well, 'ICT basics' [output] for example. We are a [[specialised industry sector]] company so a lot of our staff or well the majority would have advanced skills rather than just basics. So definitely not surprised there [for being on level 4] 'In-house IT skills' [output] that also [is high expertise]. Yeah so we're a skilled department.*' (SN:2:2)

Participants' subjective norms indicated an active knowledge-sharing culture amongst people at their organisations. This is reflected in AL:1:1's quote below.

'*'Near misses are discussed openly and regularly across all destinations. Near-miss experiences are not stigmatized and treated as invaluable learning', I think we're very good at that. We have quite an open culture basically and if we're talking about cybersecurity* *near-misses here, if there has been [[huge oversights]] and, there have been the odd one or two, then we definitely don't hide those. They're definitely widely discussed.*' (AL:1:1)

Knowledge sharing occurred internally and externally through informal and formal channels whereby people could also challenge and question information that was being provided to them for instance,

'*I'd say most people are quite happy to question (ST: Yeah; MK: It's what we do) [...] 'Question' yeah, 'share' yeah, 'challenge' yep, absolutely. I would go gold [level 3]*' (DS:5:3)

In some instances subjective norms indicated that organisations would decide the nature and extent of knowledge being shared with others. However, this choice did not appear to be driven by malice but rather a need to safeguard people, for instance from being overwhelmed.

Subjective norms amongst participants reflected an attitude of accepting human fallibility with a shared sense of culpability within the organisation. Expecting and accepting errors stemmed from participants accepting imperfection within themselves, working with imperfect systems and, meeting organisational demands, for example the quote below.

'*Because the thing that makes me think about that [factors] is that actually people are more susceptible to make mistakes when they are overloaded, and when there's not enough capacity in the organisation. We all do it and I'm going to do it sooner or later, and, you know. As an example here, [...] And you know, that's me. I should know better. And there will be people within the organisation who, because they're very busy, click on something they shouldn't do, and it's going to happen. So it's very important that we that we manage and mitigate that threat*' (AL:1:1)

Participants' subjective norms indicated that overall processes were viewed to be effective, people were able to effectively multitask, prioritise and, manage their assigned workloads. People were also able to give feedback and feed-in towards the revision and creation of organisational processes. Subjective norms within this theme indicated that people could openly communicate and report processes that were inefficient. However, participant discussions reflected a tension between processes, workload and, capacity and the overarching organisational need for growth and maximized delivery. This tension is reflected in the following quote.

'*And with that, of course, workload is ever only increasing. Even if we just do the same [amount of current projects]. But we want to grow, right. So we're going to be constantly at a point where we have to balance employee well-being, which is our first priority, and then growth scenarios. And at times, it's just going to be a stretch*' (HR:3:2)

Amongst all our participants subjective norms indicated strong relationships between people laterally at their designations and vertically with others. This meant that people were able to ask for and receive help through formal or informal channels. GH:6:1 discussed how these strong relationships also meant that people were able to delegate responsibility to others at higher designations due to a perceived blame culture,

'*We have a real problem in the organisation with upward delegation. So, yeah, 'I'm just going to delegate that task, that responsibility, that decision making to somebody more senior than me', and it happens an awful lot of in the organisation [...] Now, that comes partly, although we kind of circumnavigated a bit, there is a bit of a blame culture in the organisation that is hypothetical [placebo] [...] And that's something we're trying quite hard to break*' (GH:6:1)

Participants also revealed a subjective norm of enjoying support from Board Members when undertaking initiatives at their organisations. Finally, participants indicated that defences were stronger than they had expected prior to their session(s) for example by saying,

‘*So what I take away from this is probably [the organisation has] more robust set of processes, having thought about it [while] going through the questions, than I would necessarily have said straight up [prior to interview]. But what I take away from it is that means that we are in a position to probably better hone in on the outliers without then having to think we have to eat the whole thing at once*’ (MM:3:1)

Lower expectations amongst our participants reflected a subjective norm of underestimating the strength of cyber defences in place and organisations feeling more vulnerable than perhaps they are in reality.

6.7 Capability – *Variance in perceived control*

Results indicated that participants believed to have strong perceived control at an individual level when dealing with others (as others were seen to be individually competent) and through the effective use of technologies. However, this perceived control was weak when discussing people as groups and over the availability of resources at their organisations.

As discussed above, sessions were designed to capture participants perceived control over cybersecurity defences in line with the Theory of Planned Behaviour. The Capability theme reflected participants’ perceived control through the availability of opportunities, resources and self-efficacy in two primary contexts: technologies and people.

Quantitative analysis indicated participants were more inclined towards a shared responsibility for cybersecurity within their organisations i.e. a distributed levels of control following their session(s). Participants indicated control over the design of procedures and processes (theoretically and in practice) and responses reflected a positive correlation between procedures/practices and cybersecurity with all ratings towards 'Strongly Agree'. Participants initially indicated influence over those around them (such as being able to start working groups and communicate ideas and findings to board members). However, following their session(s) this perceived control did not appear to permeate into practice. Finally, participants expressed increased inclination of control through technologies such as restricting user access and privileges in IT systems to limit unintentional insider threat following their session(s).

6.7.1 Organisational technological capability – *Effective use of technologies equates to increased feelings of being in control*

Participants had high perceived control to avoid unintentional insider threat (UIT) through organisational technological capability. Participant discussions reflected organisations had good technological capabilities in place which acted as passive cyber defences, for instance good configuration, ability to monitor users and devices, manage user privileges and, confidentiality of data. Participants shared they were able to consider existing technologies prior to implementing new ones and were aware of their technical capabilities (strengths and limitations) during their session(s). This contributed to participants' having high trust in technologies, self-efficacy and a strong perception of control through technologies to circumvent UIT. This is reflected in the conversation between GH:6:1 and KK:6:2,

'*So we operate the system of 'least privilege' where we can, [with]in the organisation. So it's highly unlikely that we will have a system where every user is an administrator of that system, every user is a power user etc. So, when it comes to how system access is organized, yes it's a known, auditable, laid out thing [...] But we do know who the 'super users' [users* *with increased user privileges] are (KK: Yeah) So that side of it, yes, it is known, audited and documented (KK: And we share it as well) yeah, exactly*' (GH:6:1; KK:6:2)

In contrast, participants depicted lower levels of perceived control of data management as it involved human ways of working which meant it was believed to be more challenging for participants to control, for instance

'*I think we're probably about there [level 3]. Because there's some unstructured data that's copied in [places] and we've got a tool that look kind of looks at that. But we're very confident about that [level of data duplication]*' (RR:4:1)

6.7.2 People and skills – *Higher perceived control of individuals with lower levels of control over groups and organisational resources*

Participants' discussions reflected high self-efficacy through exhibiting strong confidence in others' capabilities. This strong belief in others' abilities often circumvented the need for formalised training programmes that could be offered by the organisation, for instance the quote below.

'*The emphasis on formal training [laughter] I think to an extent it doesn't take account of the fact that when we are quite a small company, we can achieve an awful lot without some of the formal training*' (AL:1:1)

Capabilities included a range of elements. It comprised of others' abilities towards technologies for instance, proficient use, being able to identify malicious content and, spot inconsistencies in software. It also included others' ability to be proactive and take action if things were amiss. Furthermore, participants indicated strong belief in others questioning or challenging concepts, practices and, guidelines. Participants believed that people practiced procedures and actively took an interest in learning new things. Participants reflected strong faith in others being able to effectively prioritise and, be able to identify and request help and advice when needed. Participants also reflected a critical understanding of processes that were in place at their organisation which included the knowledge of processes that were effective or ineffective and those that needed further improvements. This is evidenced in the conversation between participants DS:5:3 and ST:5:2 below.

'*DS: Yeah it's that weird [[specialised industry sectors]] thing, isn't it? Actually, [[specialised]] staff in particular are quite vocal on that sort of thing [processes] (ST: Mm) that they do proactively feedback and (ST: Yeah) it's one of those things.*

ST: I mean I think it's a mix. The thing is that 'experienced users feel prescribed processes are effective' well, not necessarily, but they do report inefficient things' (DS:5:3 and ST:5:2)

However, participants showed low confidence in collective capabilities of people i.e. they were confident in a majority of people being highly competent but not everyone within their organisations which is reflected in the following exchange.

'*So [[company name]] has a lot of technically able people, but it also has a number that probably aren't IT experts in certain fields. And you're going to have that in an organisation. We're not all Google, it's not like Google where everything is IT. So yeah. So I think that's quite interesting because we're relying on people's ICT skills against insider threat vulnerabilities*' (KL:4:2)

This meant that participants were mindful about the variance in capability whereby a few outliers could compromise the overall security and integrity of their defences. This favourable evaluation on an individual scale can be representative of a person-positivity bias whereby aggregated favourable values are less appealing when represented as groups (Sears, 1983).

Discussions reflected low self-efficacy and control by participants over other's capacity and the availability of resources at their organisations. Participants also indicated lesser control over senior stakeholders such as Board Members due to time constraints and communicating in ways that resonated with them. Participant KK:6:2 discussed this challenge by sharing the following.

'*So for me, it's usually the user, 'high end users' [senior stakeholders] as I would call them [laughter] that are problematic for us in terms of implementing new security measures. As [[GH]] said, you know, that [change] would involve some systems being less flexible, or less convenient. And this is what triggers users usually. Especially if we haven't got anything to say that, 'Yeah, we're putting this [change] in place because someone has hacked [us] or someone actually did this'. If we're saying to them [users that we're implementing changes] because 'We think there is a real danger currently out there and we should protect ourselves and we should protect the organisation', they [users] don't see it as the most important [priority] thing will that we should apply [implement] it [change]. So that's the challenge we're dealing with*' (KK:6:2)

6.8 Framing – *Organisations deliver a mixture of various services*

Participants viewed their organisations as dynamic entities which existed in complex environments which is reflected in the following quote by HR:3:2.

'*Sometimes there are still tasks allocated if you haven't, it's really down to interpretation, right? If you think about, if you are in an incident response team, you will get a task allocated at short notice (MM: Always. Always) this is the business. So the question is, is the* *capacity enough to always deal with that. And since it's part of that business process, you know, you just have to have staff and avenge to deal with the impact. Versus, programmed [planned] business and/or you know day-to-day operations, which tends to be different [steadier]*' (HR:3:2)

Consequently, participants considered various aspects of the business prior to selecting the best suited readiness level for their organisation presented as part of the assessment tool. Aspects through which participants positioned their organisation entailed considering various internal or external systems. 'Systems' included software, processes, workload, skills, stakeholders and, the nature of service being delivered. Aspects such as systems legacies within the organisation (in processes and equipment), geographical location and, size were also taken into consideration by participants. This meant that organisations were believed to exist in mixed states simultaneously for readiness levels against unintentional insider threat. This is reflected in the quote below.

'*I think the other thing to add is [[organisation name]] is quite unique in how it's structured. So, I don't know [[IV]] if you know, so as we said, [[RR]] and I work in a group function. But we [also] have the individual business units, [[independent, autonomous domains]] or however you want to call them. So they [independent domains] have their own operates and processes in themselves or [they] own certain processes. So it's quite a complex environment*' (KL:4:2)

6.9 Development of people and skills – *There is a commitment to upskill people through in-house resources within functions they perform for the organisation*

Participant discussions reflected that strong, informal structures were present between people at their organisations. Organisations were believed to provide 'guided training' and development to their employees. Participant discussions exhibited team reliance for support and knowledge sharing, informal channels of communications, mentoring and teaching expert skills to others and, contributing to processes through feedback and reporting. Guided training meant that people were developed and up-skilled in more informal ways such as in-house training, inductions, on-boarding, on-the-job training, handbooks and, other forms of organisational communications. Whilst training was offered by organisations, a specific training programme was only provided if it was viewed as directly relevant to the job function employees were providing to the organisation.

‘*Clear example here of you know, that the entire technical team we've just been through [[writing safe code]] training (GL: Yeah) because it is outside their skills at the minute. There’s a lot of training that we don't offer in house, so we did lots of external stuff*’ (AL:1:1)

While communications were used to up-skill and develop people, less willingness was shown by organisations to share information fully and transparently with everyone within the organisation.

6.10 Aspirations – *Consistently improving as a goal*

Participants shared organisational aspirations to improve various technical and sociotechnical aspects. This is evidenced in the following quotes.

‘*One of the consequences of that will be that there are areas of data management that we could improve […] But it’s nevertheless something that we could do better at when resources allow*’ (BP:2:1)

And,

‘*And after seeing the questions and then analysing how our users will react to a lot of different stuff. I think this is something that we’ll definitely make users more aware of. And I think put much more effort into training, well, maybe not necessarily training, but user guidance and user awareness of what is cybersecurity, what's social engineering and all the other stuff so they [users] actually know what to expect and know how to report this to us and not to be scared of reporting this [anomalies] to us because this is a very important bit [factor]*’ (KK:6:2)

Organisational technical ambitions were framed around improving elements such as encryption, software architecture, data management, asset protection and, network security. Organisational sociotechnical ambitions included improving people’s knowledge, growth, relationships, processes, management of workloads and capacity and, communications. For instance, MM:5:1 shared aspirations to improve others’ learning and communications following the session(s),

‘*Yes, I can see that now, horizontally and vertically people would communicate, culture is going to be relatively appropriate, we've got at a medium level of protocols, but I can see that we could do more on the learning and the communications*’ (MM:5:1)

6.11 Framework feedback – *Elements that worked and those that could be better*

Throughout their sessions participants shared thoughts about the factors being considered, web-based assessment tool, personalised report and, their experiences of the session(s) overall. Discussions pertaining to these elements encompassed two themes: endorsements and potentials (which included limitations and recommendations for the assessment tool).

6.11.1 Endorsements

Participants found their session(s) to be a positive experience which is reflected in the interactions between participants below.

'*MM:3:1: Certainly [laughter] certainly made me think. I was a little concerned actually thinking, 'Oh my goodness, three hours of this thing' but it's certainly something that made me think. And we, you know, I thought I haven't, you know, you [IV] and I discussed stuff which we don't in a normal day-to-day of things, we don't take some time out to do things. So I've certainly enjoyed it, yes*

HR: Yeah me too. When you started to go through the process, I was like, I was thinking 'Seriously. All right, now, three hours of, you know, [this]', but it turned out to be quite fun, actually. I enjoyed it. I thoroughly enjoyed it. I would do it again [laughter] with different questions maybe? [laughter]' (MM:3:1; HR:3:2)

And,

'*I quite liked the challenge of being made to think about things, things that I wouldn't necessarily have thought about before. Or things being presented in a different way. So I think that would be very much around the fluctuating vulnerabilities [pillar and factors which] was the key one for me, thinking through that. And also, I don't know how to put this, erm, we do various things, we know what we do and we will present that [[organisation]] does this, this and this. Done. Boxes ticked everywhere. [When] Somebody asks you a question that doesn't quite fit into the way that you do stuff, [then] you have to really think about what you do. So I actually found that quite fascinating [...] [But] Having to think about things, somebody else that's asking me the question actually, was really helpful. And also to my mind, I think it was a little bit away from [typical assessments], so we do get audited quite* *a lot, at least three times a year we have to fill the [[irritating]] forms in for auditors, this [assessment] asked very different questions to be honest (KK: [It's a] Different point of view, yeah) Yeah*' (GH:6:1; KK:6:2)

The assessment tool and the larger session(s) prompted reflection amongst participants to reimagine their current approaches. The session(s) allowed participants dedicated time to evaluate their unintentional insider threat (UIT) defences alongside the wider cybersecurity defences and the presentation of the web assessment tool granted them the ability to examine factors in an unconventional way (i.e. a human centric devised approach).

Due to the application of metacognition scaffolding technique participants were asked to reflect on their experience whereby participants reported gaining new insights through their experience of interacting with the web assessment tool and subsequent personalised organisational report. Participants reported that their session(s) allowed them to think differently from the currently established ways of evaluating defences. Participants indicated intentions to continue with this newly acquired perception (including through acting in new directions) in the future, for instance by sharing

'*But yeah, I mean this is perhaps one for us to reflect on. It is a question of workload and whether there's a risk that sits behind that, that we need to spend a little bit more time thinking about as a business. [...] Yeah. So I think this [output] is probably something that we could call out as a as a clear take away [insight] for us. It's probably a new a new risk for us to capture in our risk bank*' (BP:2:1)

New knowledge was acquired or existing thoughts were reinforced by participants following their session(s) about factors that interact with unintentional insider threat (UIT).

Participants described the outputs within the personalised report as being accurate and in line with their expectations pertaining to the strengths and weaknesses of various organisational defences that were being evaluated. This is reflected in AL:1:1's comments below.

'*I think to be honest, that's about right for me (GL: Hmm) Because actually fundamentally we are [[industry sector]] company and actually we advise other companies on software architecture. The thing for me, so this actually tells me, I think it's a fairly realistic picture that we're very good at the architecture, very good at identifying what should be going on, very good at keeping on top of the basics. We're slightly less good at the process driven stuff and so monitoring, data management and that stuff which interestingly is the human factor*' (AL:1:1; GL:1:2)

In addition, participants shared that the personalised report effectively highlighted areas that needed continued organisational attention, emphasised areas that need improvement and, indicated areas that might contain weak defences in the near future (for instance due to organisational growth). Discussions also reflected that participants discovered new factors to include in their line of defences against UIT. This is evidenced in the discussion below between BP:2:1 and SN:2:2.

'*BP: Because SN, I can't remember going through any of the processes we've been through recently [ISO 27001] where this [workload] was a discussion point [highlight] for us. I'm not even sure whether we have this risk captured in our risk register*

SN: No, I don't think so. I don't think I have, no, I don't think I've seen this [workload being considered] before. So [this input is] one of those mental notes [SN is writing] on a piece of paper is [thinking about] workload and resources

BP: Yeah. So I think this [output] is [...] It's probably a new a new risk for us to capture in our risk bank' (BP:2:1; SN:2:2)

Thus, participants found that the outputs contained within their organisational personalised report were relevant and accurate.

6.11.2 Potential future priorities

For various elements within the assessment tool the baseline measure (i.e. level 1) for readiness was not met by organisations. Baseline measure presented as *Level 1* in the assessment tool was developed in line with recommendations by NCSC, CERT and, previous research by Khan et. al (2021). While this level was seen to be the minimum recommended level of readiness for organisations, data from this study indicated that either certain parts of organisations or the organisations overall were unable to achieve said level for inputs. This is reflected in following comment by MK:5:1.

'*Just as an observation on the questions, some of the [levels], I wonder if it might be useful to have something below bronze [level 1]. Because actually in some cases bronze [level 1] is actually a reasonable bar for some organisations. So what you might find is that as you're answering the questions you think, 'Oh, we're probably somewhere in the middle of between bad and good'. But actually, when you read the writing sometimes the worst option you have is still relatively good. So for encryption [input], I can think of some organisations that haven't got that high priority data assets encrypted yet. So offering that question, offering a level below bronze might be useful to get more accurate answers*' (MK:5:1)

Participants successfully recognised elements of established frameworks and guidance that were incorporated into the assessment tool and recommended inclusion of additional technical elements. As participants were also aware of their conscious and unconscious bias and recommended the use of automated or computerised reports and analysis for measuring certain inputs such as those indicative of technical defences.

Following domain expert input from the human factors and computer science fields, descriptors for each level were designed to be subtle in order to support 'outside the box thinking' and to limit the influence of any pre-established notions, beliefs and, perceptions. However, participants sometimes expressed difficulty in selecting an appropriate rating due to the subtle differences in meanings and expressions reflected within each of the levels. Some participants indicated a need for having levels that are singularly focused on an element within its broader category and increased choice for options that went beyond 'don't know' and 'not applicable' if none of the levels appeared to be an accurate representation of the organisation's current state of defences. Participants also made recommendations for the framework to recognise different training techniques that can achieve the same results such as those indicated within readiness levels. Such training and development techniques (for instance, formalised training programmes being replaced with in-house training, on-the-job training, mentoring etc) are likely to be utilised by specialised sectors or SMEs. Following the Covid-19 pandemic, participants also indicated a need for the framework to evaluate physical defences that are reflective of flexible, hybrid and, remote working as part of the assessment tool. This is reflected in the quote below.

'*One of the things that I wondered whether you would consider as part of this data capture is where people are working. So to [address] the point around in-house ICT skills, that can often be compensated in some non-technical environments by community knowledge. So asking the person in the next cube [desk] to you what they do, who then asks [someone], you know what I mean? (IV: Yeah) So that's a lot more difficult to orchestrate when working remotely on Teams. So I wonder whether there's a data point here in terms of the 'user* *vulnerabilities' [pillar] whether, and this could also be 'knowledge' [input] as well, whether that's a factor in determining in-house ICT skills, is that [ICT skills] a problem or isn't it? In reality when you overlay [map out] how these people tend to work, then you can understand that, actually, not only if you've got someone in a call centre, but if they're remoted [remote working], the ability for them to reach over the cube [desk] and say, 'Mm, have you seen this before?' is diminished (KL: Yeah) and we've seen attacks leveraging that kind of social construct*' (RR:4:1)

Following a research study designed to explore the effectiveness of the self-reflection tool hosted via a website, this section discussed the findings that emerged from template analysis. Inspired by Ajzen's Theory of Planned Behaviour *attitudes, subjective norms* and *perceived control* amongst participants around identified organisational cybersecurity defences were discussed. Attitudes amongst participants were reflective of being favourable to humans but a preference for technologies to defend against unintentional insider threat was noted. Organisational subjective norms evidenced a notion of continuous improvement amongst participants which also served individuals as a motivation to comply with organisational rules and provide grounds to evaluate self-efficacy. Participants reflected a variance in perceived control. High perceived control was depicted in the context of organisational technological capability and individuals. This perceived control was reduced in the context of groups and availability of resources. Findings discussed above evidenced the way organisations position themselves during assessments is critical to the accuracy of subsequent results. Findings show that organisations were interested in the development of people to up-skill them with in-house resources within the functions they perform for the organisation. Participating organisations reflected aspirations to consistently improve in all aspects of their entities. Finally, feedback about the framework presented as a self-reflection tool via the website was shared as part of the findings. In the next section an in-depth discussion is held about the reflections on findings from the research study in light of the literature review presented earlier in this thesis.

6.12 Reflections

To investigate the positive impact of the self-reflection tool developed from the framework the study discussed in the above section was designed. This was done with an aim to investigate if the framework could serve as a tool for reflection that aids stakeholders in understanding and assessing unintentional insider threat at an organisational level by changing the way humans are considered within systems. Guided by the Theory of Planned Behaviour (ToPB) findings from this research study indicated participants' attitudes, organisational subjective norms and, participants' perceived control over unintentional insider threat at their organisations. Metacognition cognition scaffolding technique rendered additional findings to ToPB i.e. organisational framing, function related development of people and skills, organisational aspirations and, feedback about the framework.

The website was shown to foster reflection amongst participants to alter attitudes towards unintentional insider threat (UIT) prior to and post session(s). For instance, participant attitudes moved away from believing that UIT was a result of individuals deviating from prescribed processes following their session(s). Whilst participants demonstrated a confident attitude towards knowing their organisational technological defences (strengths and weaknesses), participants showed a negative attitude towards specific techniques that have been widely implemented in the past (e.g. phishing simulations and encryption). Despite an inconsistent approach towards the application of technological defences, participants showed a favourable attitude towards technology more generally to create, maintain and, evaluate defences against UIT. These attitudes are reflective of technological defences enjoying popularity for their perceived ease in building and maintaining defences to circumvent insider threat.

Participants exhibited a positive attitude towards others to serve as a strong line of defence. This attitude was centred on knowing others’ skill level, having faith in others’ abilities and the existence of strong interpersonal relationships. This positive attitude was further reinforced through peer-to-peer support, informal training programmes, knowledge sharing, learning and, empowerment that occurred on an individual level. Oversharing or inappropriate communication skills, questionable effectiveness of formalised training programmes and, human fallibility at individual level limited the positive attitude demonstrated towards others. Human fallibility taken into account by participants appeared to be at a surface level which was understood as an inevitable human condition. A deeper understanding was not apparent for the type of errors that occurred i.e. slips, lapses, mistakes and, violations (GEMS, Reason, 1990b) nor the type of cognitive tasks that resulted in errors i.e. Skills-Rules-Knowledge (SRK, Rasmussen, 1983). Discussions also reflected some of the elements from MERIT (i.e. organisations knowing individual skillsets and opportunities afforded to individuals) and error management programme (pertaining to knowledge sharing, empowerment and, learning) being adopted by participants in their approach. Research conducted by Nobles (2018) found managerial favourability towards technological defences to safeguard cybersecurity vulnerabilities with a reluctance to consider human factors. While our research study found that there was still a preference for technological defences, senior management and IT professionals exhibited a positive attitude and understanding of human factors that interplay with unintentional insider threat (UIT).

In line with Theory of Planned Behaviour (ToPB), participants’ positive attitude towards technology and individuals would indicate the ability to effectively strengthen defences against UIT. Kabanda et al. (2018)'s findings suggest that attitude plays a role in the effective implementation of cybersecurity at organisations. In another study (Hadlington, 2017), positive attitudes to cybersecurity at work were shown to lower risky online behaviours in people's personal lives. The same study also reported non-planning as a significant predictor of risky cybersecurity behaviour. Within this lens of ToPB, this study's findings show that participants' positive attitude would be desirable when implementing change to strengthen organisational unintentional insider threat defences.

Participants indicated an increased inclination to share knowledge more widely (i.e. near-misses and best practices) and implement changes to strength defences within their organisations following their session(s). This inclination showcased the subjective norm of everyone being responsible for cybersecurity at organisations participating within this study. Additional findings for subjective norms included support from senior stakeholders and, responsibility and accountability from senior personnel for cybersecurity related aspects. Huang and Pearlson (2019), highlight the importance of senior leadership taking special responsibility to meet organisational cybersecurity goals, as was the case demonstrated by participants' subjective norms at their respective organisations. Subjective norms at participating organisations appeared to consider workload, procedures and resources when designing cybersecurity defences. The session(s) positively influenced participants' subjective norms through incorporating organisational factors which is evidenced in a favourable change on the seven point semantic scale rating by participants following sessions (shown in Appendix 16).

Subjective norms also included upholding high standards i.e. personal conduct, delivery, practises, risk awareness and, feelings of pride linked to organisational prestige with slim margins for error. These reported subjective norms would in turn serve as motivation for people to comply with procedures set out by organisations. Participants' subjective norms were demonstrated in their critical understanding of the strength of their unintentional insider threat (UIT) defences which included a desire to discover, understand and improve factors that affect defences. These subjective norms were in conjunction to a normative belief of sharing knowledge with others, as long as such sharing did not negatively impact audiences. This stance on knowledge sharing can be encompassed by the '*effective communications*' category entailed in CERT, SOFIT and, NCSC work discussed as part of relevant frameworks earlier.

Individual attitudes of accepting human fallibility discussed above were reinforced in participants' subjective norms. However, human fallibility as a subjective norm was linked to organisation demands and imperfect systems used to deliver tasks. This understanding is closely connected to CERT (2013) which associates unintentional insider threat (UIT) to factors such as time pressures, task difficulty and cognitive load on individuals. Participants shared a normative belief that their organisations implemented effective processes. Effective processes meant that people were able to inform processes, efficiently multitask and, manage assigned workloads which is in line with the Error Management Programme (Liginlal et al., 2009). Normative beliefs also included strong peer support being available within organisations and an underestimation of the strength of cybersecurity defences in place. Additionally, participants' subjective norms denoted an understanding of the tension between delivery of work (processes, workload, capacity) and an organisational desire for growth and maximised outputs, which in turn can weaken defences against unintentional insider threat (UIT). This tension between users delivering tasks and organisational context, shown in Figure 9 through an Epidemiological Triangle (Cassel, 1976), as a subjective norm can penetrate otherwise strong defences and bolsters the need for a human centric approach to understanding the varying strength of UIT.

Subjective norms discussed above reflect a preliminary foundation in place by participating organisations to foster robust sociotechnical unintentional insider threat (UIT) defences. This foundation can be further built upon with interventions, such as the web assessment tool that hosted the framework, to create and maintain effective sociotechnical defences.

Participants shared an inclination towards distributed levels of control to maintain strong cybersecurity defences. There was high perceived control over the design of processes and procedures amongst our participants, both of which were believed to interact with cyber defences. Thus, participants perceived to have control over defences through their ability to design and influence processes and procedures at their organisations.

High perceived control was depicted amongst participants to prevent unintentional insider threat through the use of technologies. In fact, all participating organisations reportedly had strong technological capabilities, with a mixture of passive and active cyber defences already in place. The implementation and use of active and passive defences are supported by NCSC (2012), CERT (2005, 2007, 2008) and, Nurse et al. (2014). Findings reflected that participating organisations adapted practices from Error Management Programmes' (Liginlal et al., 2009) recommendations to consider existing technologies prior to implementing new ones. Data management was the only technological element with a reduced level of perceived control as it involved others (i.e. humans). Overall, participants indicated high levels of perceived control, self-efficacy and, trust through the use of technologies to create strong defences. The heightened perceived control through technologies that is expressed by participants can also be a significant contributor to the preference for technologies by senior management as suggested in the findings by Nobles (2018).

Similar to control through technologies, our findings revealed that participants believed other's skills on an individual level contributed to participants' self-efficacy and control over defending against unintentional insider threat (UIT). Participants believed in others' individual ability to actively contribute to, practice and challenge processes, continuously learn, ask for help and solicit advice, question and challenge concepts, effectively prioritise, possess effective technological capabilities, be proactive and, take action to correct course if something was awry. This strong belief in other's abilities at an individual level which contributed to participant's own self-efficacy and control was at times noted to come at the cost of diminished belief in formalised training programmes.

However, participants exhibited lower levels of perceived control over the availability of resources and capacity within their organisations and, when people are represented as groups. Resources comprised of the availability of people (i.e. linked to turn-over) and time available to deliver outcomes. Capacity included people's ability and availability to perform functions (i.e. associated to workload and cognitive loads). Participants also demonstrated lower levels of control over their ability to defend against UIT when discussing people as groups, for instance when people are represented as departments, teams or at specific job designations. Subsequently, the perceived variance in skill emerging from collective capabilities reflected in group settings translated to little perceived control by participants.

With reference to Theory of Planned Behaviour (ToPB), participants possessed high perceptions of control that is exercised through technologies and self-efficacy through the ability of others on an individual level. However, participants indicated lower levels of perceived control over resources, capacity and groups of people. This indicates that whilst participants believe to have control through utilising technologies and other's abilities, this perception of control is limited when faced with the larger systems that exist within organisations.

Overall, within the scope of ToPB, participants demonstrated a positive attitude towards technologies and people which would assist participants in building and strengthening their organisational defences. Additionally, existing subjective norms and motivations to comply are favourable for participants in their current efforts and provide advantageous conditions for them to innovate new defences in the future. Whilst participants demonstrated high perceived control through technologies and individuals, their control was limited when interacting with larger organisational systems. Thus, ToPB suggests participants would encounter challenges when faced with devising, implementing or strengthening organisational wide defences against unintentional insider threat (UIT).

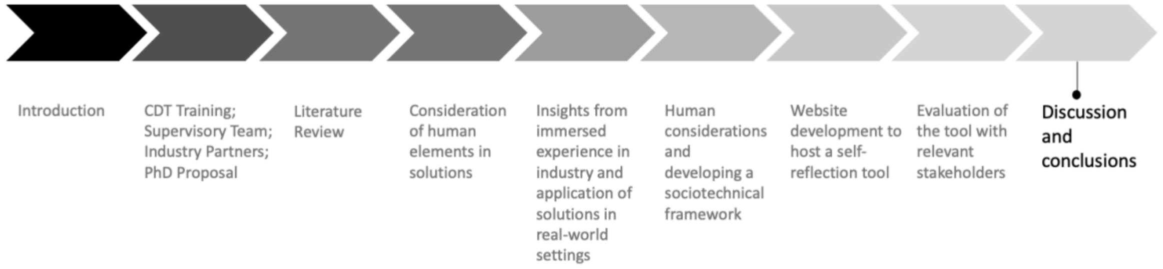
Findings from this study showcased that organisations need to position or frame themselves in a specific way when assessing their cyber defences. Considerations paid to ascertain organisational positioning when conducting assessments can include a range of aspects such as implemented software, processes, workload, stakeholders, geographical location and, size of operation. Consequently, the outcomes depicted as part of any assessment tool will only be accurate in the context of the organisational position chosen by those involved in the process. Organisations reportedly have strong interpersonal relationships that exist between people and informal structures that are believed to provide support, knowledge and, growth. People are invested in line with the job functions they perform for the organisation. Findings reflect organisational desires to improve sociotechnical cyber defences through adopting active and passive technical defences and sociotechnical defences associated to knowledge, growth, relationships, processes, workload, capacity and, communication.

Session(s) afforded participants an opportunity to reflect and introspect, fostering a change in thinking and gaining new perspectives about the way humans are considered in systems that can magnify established challenges associated to unintentional insider threat (UIT). The aspects considered were reportedly different from the more widely implemented assessments and participants described gaining insights to strengthen their defences. Outputs were deemed to be an accurate representation of current defences, effective in highlighting strength of defences and, identifying future areas of concern. Additionally, participants reported discovering new aspects to include as part of their existing unintentional insider threat (UIT) defences. The assessment tool's categories for readiness levels can be further relaxed (i.e. levels below Level 1) and more varied (more than 1-4 levels of readiness), additional technical inputs can be incorporated (e.g. ISO 27001 in Brenner, 2007) and, inputs can be further separated out into distinct levels with new options in additional to 'don't know and 'not applicable'. Varied training techniques can also be incorporated into the framework and inputs can be reflective of hybrid/remote working that has been implemented post Covid-19 pandemic.

This Chapter discussed the impact evaluation of this tool that was hosted via a website by relevant stakeholders through a research study inspired by the Theory of Planned Behaviour. The next Chapter will now present a discussion of the work presented in this thesis before sharing concluding thoughts.

7. Discussion and conclusions

7. Discussion and conclusions


Introduction
CDT Training; Supervisory Team; Industry Partners; PhD Proposal
Literature Review
Consideration of human elements in solutions
Insights from immersed experience in industry and application of solutions in real-world settings
Human considerations and developing a sociotechnical framework
Website development to host a self-reflection tool
Evaluation of the tool with relevant stakeholders
Discussion and conclusions

Introduction

The review of relevant literature reflected an approach of protecting the technological element with little consideration to the human element in cyberspace. When the human element is considered it is with the intention to protect information and control the way in which humans operate in a bid to save humans from themselves. To explore the extent to which solutions are holistic in a sociotechnical system, recommendations were applied to the onion model and revealed an unequal emphasis on technological or external elements to tackle unintentional insider threat (UIT). In order to develop an understanding of the factors that influence UIT, a research study was designed to learn from people who had experienced this threat. Findings contributed to the development of a sociotechnical framework which was taken to relevant senior stakeholders from industry for evaluation via a website in a bid to change the way humans are considered in systems. This Chapter holds a discussion on the work presented in this thesis and provides concluding thoughts.

7.1 Overview of the thesis

To serve as a reminder 'insider threat' in this thesis is defined as follows:

'*Actions [encompassing skills, rules and knowledge-based behaviour] or inaction of individuals or groups who wittingly or unwittingly cause loss or harm to the security of an* *organisation, without a differentiating between cyber or physical perimeters. The individual(s) has authorised access [physical and/or cyber] to physical assets and to confidential information in order to perform a function for an organisation which results in compromised safety or a cybersecurity breach.*'

Consequently, unintentional insider threat is defined as follows:

'*Insider threat that is not a result of intentional actions that cause loss or harm to an organisation by insiders.*'

In Chapter 1, this thesis introduced insider threat, its associated challenges and shared a definition which reflects its multifaceted features and dynamic nature. To set the scope of this thesis, research questions were shared. Partner engagement and industry driven nature of the problem was presented through insights garnered from numerous industry collaborations. This was done with an aim to share industry input that informed various elements of this project and provide an insight into real-world settings in which solutions are designed to be implemented. A detailed outline of these experiences is presented in Appendix 1. This Chapter also included a statement of novelty and expected contribution, listed the publications arising from this work and, provided the structure of this thesis.

Chapter 2 presented a literature review of relevant work. A paradigm was shared for classifying cyberspace operations based on intentions which were classified as either being offensive or defensive, where offensive operations are beyond the scope of this work. Defensive Cyberspace Operations were further classified into two categories i.e. active and passive cybersecurity defences. Through developing an understanding of the intentions behind defences to safeguard against cyberattacks, the work focused on the types of threats that exist in cyberspace which arise from two elements: technological and human. Concurrently the work discussed various types of solutions proposed to contain or counteract these threats. Prominent attacks were also discussed to demonstrate solutions that are derived from software defence approaches are insufficient for creating effective defences against unintentional insider threat (UIT). Discission progressed to examine prominent and relevant frameworks that are designed to tackle UIT. Finally, this Chapter presented sociotechnical theory and relevant perspectives such as:

- the Epidemiological Triangle (Cassel, 1976) to aid in understanding the dynamic relationship between vectors such as the host (human), agent (cyberattack) and the environment (cyberspace) in which they both exist *and*, the Swiss Cheese Metaphor to visualise the generation of errors in complex sociotechnical systems due to intrinsic vulnerabilities of the contributors. This technique provides an insight for understanding the compromising of robust defences if vulnerabilities align to realise a threat or an attack (as vulnerabilities in defences can align in a way that is favourable for the attack to succeed);
- Safety II approach that acknowledges the variance in human performance that enables the safe operation of dynamic systems and makes a case for learning lessons from when things work correctly in addition to when they don't and result in accidents (cyber incidents and breaches);
- Skills, Rules and Knowledge or SRK approach that provides a taxonomy for the types of errors to aid in understanding the complexity of human decision making and;
- Generic Error-Modelling System (GEMS) that provides a taxonomy of tasks that result in errors (cyber incidents and breaches) through amalgamating the SRK approach with cognitive psychology.

Chapter 3 presented a critical evaluation of a prominent approach to tackle insider threat in an effort to understand how current state-of-the-art solutions for unintentional insider threat can be enhanced. Recommendations by CERT (2019) in ‘Common Sense Guide to Mitigating Insider Threats, Sixth Edition’ were evaluated through the use of case studies to examine their applicability to SMEs. The work progressed to classify recommendations according to the categories presented in the onion model. Findings from this activity revealed that recommendations can be reframed to change how humans are considered within systems as results demonstrated that while some recommendations were applicable to unintentional insider threat, a majority of solutions were difficult for SMEs to achieve or apply to unintentional insider threat. In addition, results reflected an emphasis being placed on technological and external factors in systems to defend against insider threat. Thus, a case was made for human centric solutions to emerge that are exclusively tackle unintentional insider threat.

Chapter 4 discussed the design and findings from a research study that exclusively investigated factors that influence unintentional insider threat (UIT). The study applied the Critical Decision Method (CDM) approach and revealed four thematic findings pertaining to decision making, task factors, accidents and, organisational factors that were interlinked to UIT being realised. These findings were utilised to inform the creation of a framework that incorporated existing aspects within approaches and introduced new features, with a total of 35 distinct inputs that were spread across five pillars.

Chapter 5 shared the design of an information system artefact (i.e. a website) that presented the framework in the form of an assessment tool to evaluate organisational readiness levels against UIT. Using an Action Design Research inspired approach, the website was developed over six phases over the timeline of this research project. This was to aid relevant stakeholders with understanding UIT from a human centric perspective and addressing UIT at their organisations through personalised reports.

Chapter 6 discussed the design and findings of a research study inspired by the Theory of Planned Behaviour (ToPB). The aim of this study was to prompt reflection and a change in participant perspectives of unintentional insider threat and its associated defences. Results reflected attitudes towards technologies and people, organisational subjective norms and, the variance in perceived control of organisational technological capabilities and, people and their skillsets. The application of metacognition scaffolding technique rendered additional findings that included organisational self-framing, development of people and skills, organisational aspirations and feedback about the sociotechnical framework.

Having provided an overview of the thesis, this work progresses to present an overview of the research questions presented in Chapter 1.

7.1.1 Challenges of designing solutions for insider threat

This section aims to discuss the following research question that was posed at the beginning of this thesis:

1. To what extent are current cybersecurity approaches considering operations of the human element?

Through reviewing extant literature, this question explores opportunities of being able to apply an alternative human factors centric lens with which unintentional insider threat can be understood.

Designing effective solutions to address unintentional insider threat is challenging. This is due to a lack of shared understanding of the parameters with which insider threat can be defined and, it is challenging to determine and agree on who qualifies as an *insider*. Subsequently, since insider threat and insiders have their own interpretations, the emerging solutions tackle various aspects of this threat which makes it difficult to assess the nature or aspect of insider threat that is being addressed. Furthermore, approaches combine intentional and unintentional insider threat when proposing solutions that can limit their applicability to unintentional actions.

When examining the intentions behind cyber defences i.e. passive (PCDs) or active (ACDs), it can reveal if defences are set-up to guard the technological element or to leverage it against the human element in cyberspace operations. For instance, PCDs aid in optimising the technological set-up and implement best practices to establish a robust baseline for the technological element's operation. In contrast to PCDs, ACDs can be utilised or leveraged against the operation of the human element prior to any indicators of a threat being present. ACDs can involve (covertly or overtly) investigating the environment, monitoring all human activity and creating behavioural and/or psychological profiles of potentially innocent insiders represented at an individual or group level. An ontology of passive and active cyber defences is shown in Figure 3 below (created by the author that is inspired from Figure 1 by Goethals and Hunt, 2019).

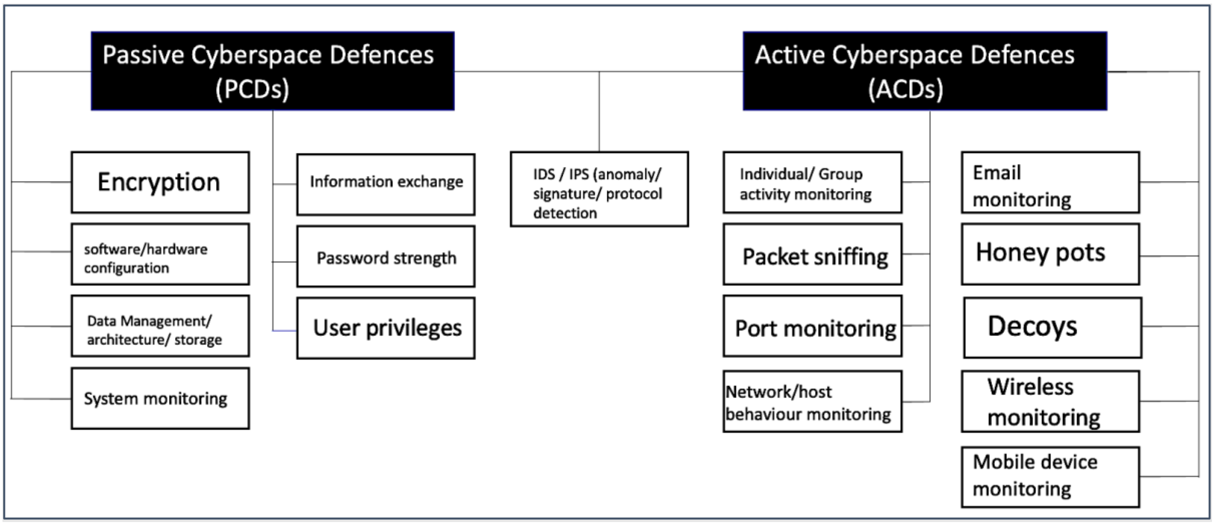

Passive Cyberspace Defences (PCDs)
Active Cyberspace Defences (ACDs)
Encryption
software/hardware configuration
Data Management/ architecture/ storage
System monitoring
Information exchange
Password strength
User privileges
IDS / IPS (anomaly/ signature/ protocol detection
Individual/ Group activity monitoring
Packet sniffing
Port monitoring
Network/host behaviour monitoring
Email monitoring
Honey pots
Decoys
Wireless monitoring
Mobile device monitoring

Figure 3: Ontology for passive and active cyber defences

Vulnerabilities that exist in software within the technological element can also be categorised into four main types: known, known unknowns, known knowns and, unknown unknowns. The *known* category comprises of vulnerabilities that were known to the developers and/or the organisation which allowed them to *anticipate* a possible route for an attack which can afford the developers a chance to pre-empt an attack if they *intend* to do so. *Known unknowns* category reflects the vulnerability in software that can include having knowledge of vulnerabilities that are obvious once an attack has occurred. However, this category is difficult to circumvent as there can be an endless amount of vulnerabilities that might be *imagined* as possible attack pathways but do not realise into cyberbreaches. The *intentions* of responsible parties (such as creators of the software or organisations) can then be used to determine some of the high priority vulnerabilities to address within this category. *Known knowns* category comprises of those set of vulnerabilities in software that were recognised by the responsible parties and the wider community that is interested in the cybersecurity domain. In the context of *intentions*, this category reflects intentional *awareness* of potential routes that can afford successful cyberattacks. Finally, *Unknown unknowns* category comprises of vulnerabilities in the technological element that were unknown to all parties unless an attack is realised through a *lucky break* i.e. the attacker might not have *intended* to attack or did not believe that an attack would be successful.

Given the above vulnerabilities that exist in the technological element, the human element adds another dimension to the cyberspace landscape. Human element is critical as it enables all cyberspace operations and can make executive decisions that can realise cyberattacks. Thus, cyber defences discussed above are set up with the intention to protect the technological and/or the human element. However, as the technological element holds data that is associated to monetary value the focus of research and subsequent solutions proposed as defences has been on protecting the technological element (Ani et al., 2018).

One type of vulnerability that arises from the human element in cyberspace is known as insider threat. Based on intentions, insider threat can be *intentional* or *unintentional*. The lack of a formal definition for the term *insider threat* and who qualifies as an *insider* poses its own set of challenges. For instance, without a clear classification and description of these two elements it is difficult to identify the specific challenge solutions are meant to be addressing. Furthermore, the definition of insider threat and insiders can be motivated by the region of originating research, the intended audience, the organisational or national security policy, the cyber and physical parameters, the type of device being used etc. This in turn impedes research efforts as it becomes time consuming to interpret from inferences the type of threat being addressed and, it limits the ability to compare approaches.

Whilst the definitions for *insider threat* and *insider* are disputed in literature, this project defined these as follows:

'*Actions [encompassing skills, rules and knowledge-based behaviour] or inaction of individuals or groups who wittingly or unwittingly cause loss or harm to the security of an organisation, without a differentiating between cyber or physical perimeters. The individual(s) has authorised access [physical and/or cyber] to physical assets and to confidential information in order to perform a function for an organisation which results in compromised safety or a cybersecurity breach.*'

Unintentional insider threat is defined as follows:

'*Insider threat that is not a result of intentional actions that cause loss or harm to an organisation by insiders.*'

Despite the ambiguity around insider threat and who qualifies as an insider there are still numerous approaches to counteract insider threat. One prominent approach by Computer

Emergency Readiness Team (CERT) identifies insider threat through individual motivations, their technical skill ability and, their knowledge about the organisation's operations. They propose close monitoring of individuals that are known or can potentially be harmful to the organisation. Whilst this approach can work well for intentional insider threat as actions would indicate an *intent* to jeopardise operations, steal valuable information or harm and deceive the organisation, this deduction has limited applicability for tackling unintentional insider threat i.e. where the individual did not *intend* to harm the organisation but inadvertently does so.

Other approaches also consider psychological and behavioural characteristics to identify insider threat. This includes the application of OCEAN Traits, Dark Triad Traits, background checks and susceptibility to rule breaking behaviour to inform individual profiles. These approaches can encourage covert monitoring of individuals, rely on unqualified personnel to make those deductions (such as individuals from the Human Resources department) and, infringe local laws that limit the gathering and use of personal data (such as GDPR in the European region). In addition, in instances where the technological element is being used to make such deductions about humans (for instance through machine learning algorithms), outcomes pertaining to individual profiles can reinforce biases that exist in algorithmic coding (*people of colour are prone to crime based on their arrest history*) and reinforce technological biases in real-world settings (*the computer cannot be wrong*). This can propagate the notion of *bad apples* and target individuals who might never harm the organisation, wasting the use of limited resources that exist in organisations.

The technological element is also leveraged to limit the opportunities afforded to insiders to realise insider threat. This can be done through managing access privileges, monitoring access and system logs, monitoring technical skills of believed perpetrators and, regulating the software being utilised by individuals in organisations, monitoring user activity to identify rule-breaking behaviour, administering psychometric tests and, evaluating stress levels. These solutions also become problematic as they collect personal data on individuals and ultimately would be unhelpful for identifying intentional insider threat. For instance, limiting individual access privileges might limit the damage from the cyberattack (depending on the *type* of attack) but it would not prevent the cyberattack from occurring especially if it was unintentional in its nature.

The approaches discussed above aim to limit and protect information that exists in systems through the technological element, i.e. software designed to make deductions or predictions, in a bid to control and limit the operation of the human element in cyberspace. This limitation of the human element can be challenging as it creates barriers for innovating efficient ways of work and oversimplifies human operations that includes decision making that occurs in complex sociotechnical systems.

The research from CERT's MERIT framework for insider threat was predominantly based on intentional or malicious attacks. However, solutions encompassed intentional *and* unintentional insider threat. As part of this framework, technical cyber defences were recommended which included a range of active (ACDs) and passive (PCDs) defences. Whilst active cyber defences (ACDs) can be effective for identifying intentional insider threat, they are limited in their applicability to unintentional insider threat due to its sudden, unexpected or *unintended* nature. This framework seeks to protect the technological element and is driven by *traditional security thought* whereby humans are believed to be the weakest link in the security chain.

Sociotechnical and Organisational Factors for Insider Threat or the SOFIT framework considers technical, behavioural and, organisational factors to identify insider threat. This framework also relies on the technological element (through churning the inputs through a Blackbox) to identify *problematic individuals* who can be addressed in an effort to tackle insider threat. This framework can also benefit from considering the human element in systems so as to avoid placing the emphasis on protecting the technological element and limiting the operation of the human element in cyberspace. Furthermore, all indicators used to evaluate this threat are not shared with the audience. The application of this framework can reinforce the biases mentioned above for individuals that utilise it. This nondisclosure can also hamper research efforts to validate the use of various indicators.

Another framework called Error Management Programme examines root causes that result in errors to tackle insider threat. The recommendations incorporate processes as they are believed to prevent, intercept and avoid errors. In addition, the design of technologies (i.e. effective design will prevent errors) alongside training programmes to develop skills are believed to avoid insider threat. Whilst this approach considers the operation of the human element, it focuses on the creation and avoidance of errors. This creates challenges for its application to unintentional insider threat as it places the onus on organisations to circumvent insider threat and ensure processes are adhered to.

Finally, National Cyber Security Centre (NCSC) provides guidance to address insider threat. Results from this centre are scattered across topics which is reflective of the challenge that originates from the lack of definition for insider threat and insiders. Additionally, it also reflects the relevance of insider threat across topics within cybersecurity as this insider threat primarily addresses the human element which is essential to enable cyberspace operations. When insider threat is being addressed, it is dependent on the audience’s interpretation of the *type* of insider threat being considered (i.e. intentional or unintentional). In this consideration of the human element, recommendations set forth emphasise the importance of considering various roles and processes when devising organisational security policy and empowering individuals to share poor practices and report incidents. However, in this consideration the full responsibility of actions undertaken in complex systems is placed on the individuals which reflects a traditional security thought approach. The reprimands associated to actions that result in cyberbreach (intentionally or unintentionally) can be viewed as a way of controlling the operation of the human element in cyberspace.

As reflected in the discussion above, while there is a strong foundation from the efforts made to consider the operation of the human element in cyberspace, approaches can be enhanced to device solutions specifically for *unintentional* insider threat through adopting an alternative perspective to understand errors in complex systems. If this understanding is human centric in its underlying nature it can result in creating solutions that are reflective of the dynamic nature of systems in order to protect the human element and enhance its operations.

7.1.2 Alternative approaches to consider the human element

There are established approaches from the human factors and risk and safety engineering domains that aid in understanding errors that result in undesirable outcomes such as cyber incidents or cyberbreaches.

One such approach is the Epidemiological Triangle (Cassel, 1976) that is implemented in public health communications and safety science. Used as a visualisation technique it aids in understanding and demonstrating the interdependent relationship between vectors. It shows the environment in which the human and the agent coexist, much like cyberspace where the attack, the human and the technological elements cohabit the same environment. The second vector in this triangle is the agent which can be understood as the cyberattack in cybersecurity. The final vector is the host which reflects the human element. This triangle is useful for understanding the dynamic relationship between entities. It makes a case that strengthening one vector is insufficient in eradicating or addressing the challenges being faced. In the context of cybersecurity and more specifically unintentional insider threat, this model would dictate that the attributes of the cyberattack, the strength of the human element and the context in which the human element and the attack are operating would all contribute to the chances of success or failure of an attack. Therefore, in order to limit or eradicate the occurrence of a successful cyberattack, the human element must be enabled to operate in the best way possible that compliments defences and the technological element, the environment must be enhanced to limit the attack (for instance through training programmes) and, the attack must be weakened (for instance people might be less likely to fall prey to phishing scams if they are equipped with the knowledge to *understand* the strategies behind such attempts and cyberspace operations).

Similarly, Safety II approach provides an alternative technique for understanding the occurrence of errors in complex sociotechnical systems. It does so by classifying prior safety science models as Safety I which are focused on identifying, limiting and eradicating errors. The absence of errors is used to indicate that a system is in a *safe* state, making the understanding binary and removed from the reality of how work is actually performed. In investigations following incidents, blame is often places on the operations of the human element as it is the most variable vector in contrast to technological and process elements. This approach to an extent appears to be similar to traditional security thought discussed above. In contrast to Safety I, a Safety II approach argues that modern day systems are dynamic and unstable due to the variance in human performance to adjust to changes in the environment. This variance in performance is what allows systems to operate in a *desirable* state. Thus, learnings can be extrapolated from examining things that result in errors *and* aspects that work well i.e. the absence of accidents or cyberbreaches can offer chances of replicating success. Furthermore, it is this variance by the human element which allows systems to be adaptive and resilient providing the cornerstone for the successful operation of systems. In a cybersecurity context, a Safety I approach is reflected in solutions and frameworks whereby the human element is controlled or limited in its operations through either pre-emptive evaluations (such as behaviour and psychological profiling) or through the leveraging of the technological elements. As these established approaches are demonstrated to have limited success with unintentional insider threat, a Safety II approach can aid in providing an alternative perspective that is human centric and can aid in enabling and strengthening the human element in cyberspace operations.

Skills, Rules and Knowledge (SRK) approach is also beneficial in understanding errors in the context of unintentional insider threat specifically as it refutes simplistic notions that associate errors to an inevitable human condition. This approach provides a classification system for tasks (i.e. skills, rules or knowledge based) that describe human behaviour and decision making that occurs in complex environments. This approach becomes important in the context of understanding unintentional insider threat as cyberspace operations involve complex landscapes and environments and rely on the human element to process complicated information and affords them making executive decisions often in short periods of time.

The SRK approach is further developed by the Generic Error-Modelling System (GEMS) approach which combines it with cognitive psychology to classify the types of tasks that can generate errors (i.e. slips, lapses, mistakes and, violations). GEMS can aid in understanding unintentional insider threat as human behaviour is recognised to be dynamic in its nature and dependent on contextual queues that are offered through interaction with the technological element and the environment. In this approach as humans learn and experience things, understood or explicitly stated rules are overwritten and transformed. With this perspective in the context of unintentional insider threat, completely placing the onus of actions on individuals can be problematic and systematically create vulnerabilities in the human element (individuals might make more errors from the fear of making errors) and the environment (creation of errors due to the fear of reprimands or a blame culture).

7.1.3 Evaluating the extent of human considerations in approaches

This section holds a discussion about the extent to which current state-of-the-art approaches consider the human element when examining insider threat. It aims to answer the following research question:

2. How might a human-environment systems approach aid in reframing current approaches from a human centric stance?

This research question explores the extent to which current approaches are suited to unintentional insider threat through the use of case studies and, the extent to which these approaches are holistic in sociotechnical contexts through the application of the onion model and identified opportunities for human factors domain to propose solutions.

In order to evaluate solutions proposed to tackle insider threat, a prominent guide by Computer Emergency Readiness Team (CERT) was critically evaluated. It was selected for evaluation as it is produced by a world leading research group that is informed by numerous collaborators and is recognised to produce state-of-the-art solutions for insider threat. In addition to this positioning which made the guide a well suited document to evaluate, the sixth edition was in response to GDPR legislation in Europe, making it even more relevant as this research project also originates from the same region.

Through applying the recommendations to developed case studies of SMEs findings reflected an oversimplification of complex environments. Subsequently, solutions were two dimensional in their approach which would prove problematic to implement or create ripple effects that can create or contribute towards systematic issues such as organisational culture. Solutions were positioned to be easily achievable to build capacity within organisations to tackle insider threat and on the surface level appear to be easy to achieve but are challenging when being implemented in complex and dynamic environments wherein technological and human elements interact to support operations in cyberspace.

To evaluate the extent of considerations being paid to the human element in cyberspace, these recommendations were classified respectively to categories within the onion model. From the human factors domain, the onion model depicts a complete human–environment system. These categories in the onion model are interdependent, interconnected and, sensitive to change. Classification of recommendations onto the onion model are presented in Figure 5 below:

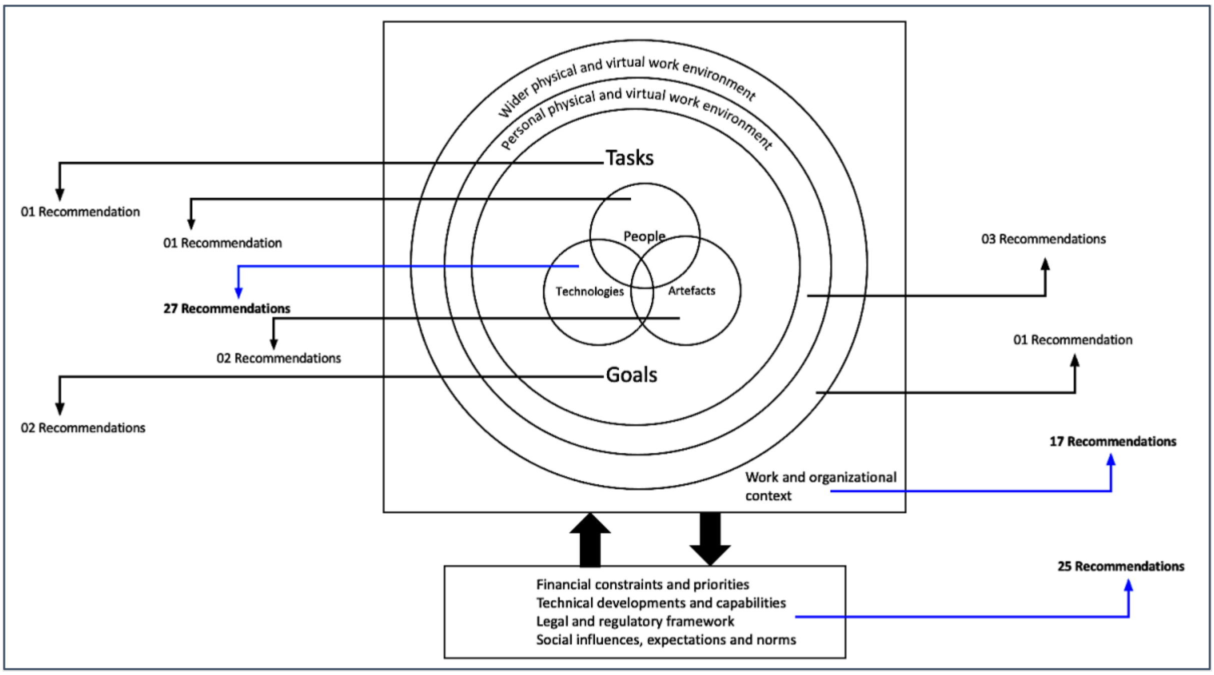

Wider physical and virtual work environment
Personal physical and virtual work environment
Tasks
People
Technologies
Artefacts
Goals
Work and organizational context
Financial constraints and priorities
Technical developments and capabilities
Legal and regulatory framework
Social influences, expectations and norms
01 Recommendation
01 Recommendation
27 Recommendations
02 Recommendations
02 Recommendations
03 Recommendations
01 Recommendation
17 Recommendations
25 Recommendations

Figure 5: Classifying recommendations onto the onion model

When being evaluated through a human centric approach, outcomes demonstrated that CERT’s recommendations within the guide placed an unequal emphasis on certain elements within a system. In addition to the recommendations being partially applicable in real-world settings experienced by SMEs, the recommendations placed substantial onus of defending against insider threat on the technological element with 27 recommendations associated to this category. External elements were also responsible for tackling insider threat with 17 recommendations pertaining to work and organisational contexts and 25 recommendations placing the onus for defences on societal aspects. Findings from this application to the onion model also reflected little consideration paid to the human element with only one recommendation pertaining to people, tasks and, the wider physical and virtual work environment that exist in complex sociotechnical systems. Furthermore, proposing solutions that fuse intentional *and* unintentional insider threat are too board and consequently lose their ability to be holistic. However, recommendations pertaining to the technological element can be enhanced in ways that do not leverage it against the human element whilst contributing to robust defences that further good cyber hygiene and best practices.

The application of current approaches such as the guide by CERT underpin solutions to tackle insider threat that are implemented in industry settings. The author’s immersed experiences in various industry partnerships demonstrated low levels of understanding for the type of human considerations that need to be taken into account when tackling unintentional insider threat. Experiences also indicated a need for these human considerations to be communicated to industry more effectively to increase their understanding of the role of human considerations when tackling unintentional insider threat.

7.2 How can the human element be considered in systems

Having established the extent of considerations paid to the human element in cyberspace and the need to reframe current approaches to inform holistic solutions in the previous section, this section examines the way in which the human element can be considered when examining unintentional insider threat.

7.2.1 What are the factors that influence unintentional insider threat?

To begin devising ways to protect and enhance understandings of the human element through adopting a human centric stance, investigation can begin by understanding factors that influence unintentional insider threat to learn from those that have experienced it. Thus, this section begins by investigating the following research question:

3. What can be learned from people's experience of unintentional insider threat about factors that influence it?

This research question applies Critical Decision Method (CDM) to understand individual experiences that led to unintentional insider threat in order to validate current approaches and introduce new elements for consideration to safeguard against such a threat.

As attacks are getting increasingly sophisticated and attack surfaces are increasing with the deployment of interconnected technologies, well-intentioned employees become prime targets for hackers to enable successful cyberattacks such as ransomware. In order to tackle unintentional insider threat (UIT) effectively organisations need to have a clear understanding of factors that influence UIT in complex environments and, human centric ways for considering the human element in cyberspace. This thesis discussed UIT as a separate and distinct phenomenon to intentional insider threat to build a better understanding of this threat through presenting an understanding of cyberspace based on intentions of those that operate within it, approaches to proposed solutions and, prominent frameworks. This understanding is utilised to evaluate the extent to which humans are considered in systems and whilst these approaches can provide a solid foundation, they can be enhanced to better enable the human element.

Bearing in mind the principles of Safety II's approach to learning and SRK's approach for contesting the proposition of simplistic notions to understand errors in complex environments that involve the human element to make executive decisions, a study was designed to investigate factors that influence unintentional insider threat. Through the application of Critical Decision Method (CDM) data was gathered and analysed through a grounded theory approach. Findings was refined to extend the application of the Epidemiological Triangle (Cassel, 1976), shown in the Figure 9 below:

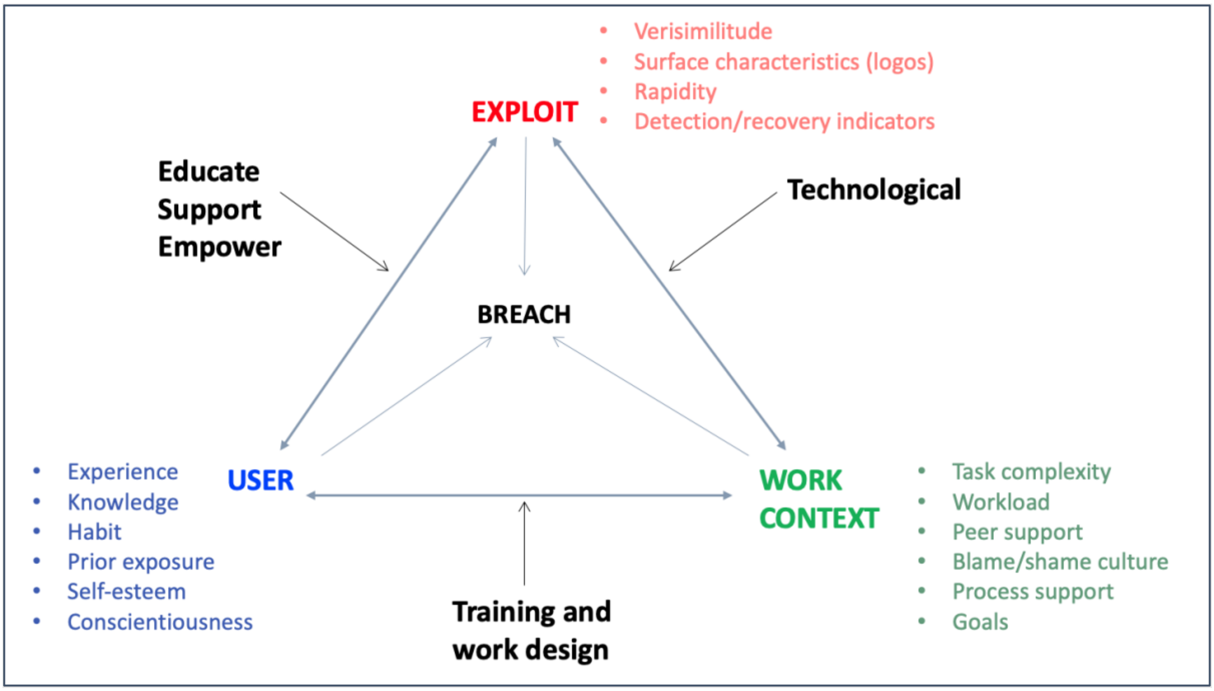


Figure 9: Epidemiological Triangle based on CDM research study

This visualisation of the three vectors in cyberspace i.e. the human, the attack and, the environment, aids in understanding how each of these elements can be strengthened or weakened in relation to each other. It represented a dynamic relationship that exists in complex sociotechnical environments that can lead to or circumvent breaches that stem from unintentional insider threat. This figure also showcases the importance of considering all three elements concomitantly to address multiple elements in cyberspace without placing the onus on a singular aspect as it would be insufficient to address challenges posed by the interaction between vectors that can increase susceptibility to unintentional insider threat.

Additionally, findings from this research study informed a sociotechnical framework (presented in Figure 10 below). For the purposes of this work a framework is defined as a set of recommendations applicable in specific scenarios to reduce negative impact in the cybersecurity of systems.

Outputs	Pillar No.
User vulnerabilities to UIT and recommendations to strengthen defences	1
The effectiveness of processes and facilitating a continuous improvement culture	2
Workload and sufficient resource allocation	3
Knowledge sharing and empowerment culture	4
Fluctuating vulnerabilities	5

Inputs	Contributes to
Assess how comfortable individuals are with various technologies and platforms	Pillar 1
Assess how vulnerable users feel in their daily online interactions	Pillar 1
Assess physical working environments	Pillar 1
Assess individuals' ability to identify spear phishing scams to note vulnerabilities	Pillar 1
Assess individuals' existing experiences with malware or threats (including physical spaces)	Pillar 1
Assess individuals' knowledge base to evaluate understanding of current techniques used by hackers	Pillar 1
Assess individuals' susceptibility to rationalise abnormal behaviour or interactions	Pillar 1
Assess individuals' susceptibility to spear phishing	Pillar 1
Assess individuals' trust in technologies	Pillar 1
Assess the levels of how much individuals rely on their social networks (offline and online) to inform their decisions if faced with threats	Pillar 1
Assess individuals' awareness of mainstream marketing campaigns against popular attacks	Pillar 1
Assess levels of retention from basic ICT teachings to establish levels of awareness	Pillar 1
Assess and map different skill levels between individuals in a diverse workforce	Pillar 1
Assess individual's level of caution when interacting with suspicious or odd behaviour (online and physical parameters)	Pillar 1
Evaluate all tasks to identify missing feedback loops that indicate task completion	Pillar 1
Evaluate the effectiveness of prescribed processes amongst skilled/experienced staff	Pillar 2
Assess individuals' prioritization of processes	Pillar 2
Evaluate the effectiveness of prescribed processes amongst all designations	Pillar 2
Evaluate in-use software's limitations in prescribed processes	Pillar 2
Assess individuals' commitment to best practices set out by the company	Pillar 2
Evaluate processes for collaborative tasks that are automated	Pillar 2
Assess individuals' technical skill levels	Pillar 2
Assess individuals' levels of personal responsibility felt when delivering tasks assigned to them	Pillar 3
Assess resources available to individuals to deliver tasks	Pillar 3
Assess individuals' motivations when delivering tasks	Pillar 3
Assess individuals' ability and willingness to take on additional tasks	Pillar 3

Assess levels of stigma associated with experiences of near misses and accidents that result in cyber incidents and cyberbreaches across all levels	Pillar 4
Assess levels of communication about cyber incidents	Pillar 4
Assess individuals' understanding of outcomes that result from accidents	Pillar 4
Evaluate effectiveness of current guidelines in the event of a cyberbreach	Pillar 4
Evaluate individuals' understanding of protocols in the event of a cyberbreach	Pillar 4
Evaluate individuals' ability to question, share and challenge abnormal interactions	Pillar 4
Evaluate relationships between individuals and managers across all levels	Pillar 4
Evaluate relationships between peers across all levels	Pillar 4
Assess individuals' level of attention to detail (online and physical parameter)	Pillar 5

Figure 10: A sociotechnical framework to assess Unintentional Insider Threat

The framework is utilised to assess organisational readiness levels. Each of the five pillars is formed of respective 'Inputs'. These inputs were uncovered as part of the findings from the Critical Decision Method based research study.

Through the evaluation of 35 distinct inputs that inform five pillars, this framework can be utilised by organisations to identify, intervene and mitigate against unintentional insider threat. Inputs were devised from capturing factors that influence unintentional insider threat and thematically categories to inform five pillars. An additional pillar was included for the technical element in cyberspace (i.e. the sixth pillar in the framework) through incorporating recommendations from current approaches that endorse good cyber hygiene and implement passive cyber defences (PCDs) discussed in the previous section. This was done to ensure that all aspects are considered equally without negating elements within systems. The inclusion of technical cyber defences was also sensible as it allowed enhancing of the strong foundation laid by existing approaches. Active cyber defences (ACDs) or any aspects in existing works that aim to limit or control the operation of the human element were excluded from being incorporated into this framework so as to avoid the leveraging of the technical element against the human element in cyberspace. Thus, the technological pillar titled *Technical Cyber Defences* included the following inputs: Software Architecture, Monitoring, Configuration, Encryption, Access Points & Privileges, Data Management, Updates and, Audits.

As a result, this framework incorporated new findings and enhanced existing elements from relevant frameworks designed to tackle insider threat (only unintentional aspects were deemed appropriate for inclusion). Existing elements that are utilised as part of the framework include the use of passive defences (adopted from NCSC and CERT), mapping in-house technical skills across all designations to build talent (incorporated from NCSC), risk awareness (adopted from NCSC), evaluating physical environmental stressors (incorporated from SOFIT), educating and raising awareness through training (from NCSC and CERT guidance), evaluating processes (adopted from Liginlal et al.), monitoring time pressures (incorporated from CERT and Nurse et al.) and, instilling an organisational culture of empowerment (adopted from NCSC, CERT, Liginlal et al. and, SOFIT). New features introduced in this framework included the following:

- Listing of rationale behind steps within processes, designing and evaluating processes with staff who possess expertise at performing assigned tasks and, periodically evaluating processes' effectiveness through 'work as conducted' are essential
- New interlinked factors to time pressures are included
- Knowledge attainment and sharing in addition to an empowerment culture is recommended
- UIT involves fluctuating vulnerabilities that must be known, monitored and, addressed

The framework was aimed to be utilised by organisations to build an understanding of the human considerations when addressing unintentional insider threat (UIT), the type of considerations that need to be appraised and, the role of human considerations when tackling UIT. This was deemed achievable through evaluating the strength of defences and self-reflection.

7.2.2 How can existing understandings about unintentional insider threat be reframed

Following the creation of the sociotechnical framework for unintentional insider threat, this section discusses the extent to which the framework can have a positive impact on audiences for understanding unintentional insider threat. It addresses the following question posed at the beginning of this thesis:

4. What user centric solutions could have a positive impact in an open environment for understanding unintentional insider threat?

This research question explores the extent to which the developed sociotechnical framework can prompt individuals to reflect on challenges posed by unintentional insider threat in organisational contexts.

A website was created to host the framework to communicate research findings to industry audiences in order to reframe understandings. This was achieved through evaluating the strength of defences for each of the inputs across six pillars. A IT artefact in the form of a website was chosen to afford collective input from dyad sessions, be understandable to lay audiences (i.e. outside the field of cybersecurity and IT), be visually appealing and, avoid confusion when relevant stakeholders were progressing through the framework. The framework was revised to be positioned as an assessment tool that evaluated the strength of defences against unintentional insider threat and provide outputs in the form of radar graphs for each of the pillars being considered. This was done to stimulate deep learning and reframe existing understandings of tackling unintentional insider threat. Additionally, a diverse set of audiences were required to participate in sessions (including technical and non-technical backgrounds) as a way to demystify this domain to organisational decision makers and, to engage audiences that have previously been excluded from the creation and evaluation of solutions. Through adopting a Safety II perspective, the outputs also designed to highlight areas that performed well in order to provide an opportunity for organisations to recognise and replicate areas of success i.e. inputs that reflected strong defences against unintentional insider threat.

Inspired by the Theory of Planned Behaviour (ToPB) a research study was designed. Thirteen participants from six organisations were recruited through National Cyber Security Centre's 'industry 100' partners and the first author's professional contacts. Participants represented senior and mid-level leadership roles within various sized organisations (SMEs, large and non-profit). Participant attitudes were measured through semantic scales and free-text fields prior and post their interaction with the tool and semi-structured interviews post engagement with the tool as this was indicative of tool's influence on their planned behaviour towards their cyber defences. Template analysis approach was applied to code approximately 14 hours and 30 minutes of data to uncover findings.

The designed sessions facilitated reflection amongst participants to gain new perspectives to established challenges posed by unintentional insider threat (UIT) and introduced participants to new aspects to consider as part of their existing defences. Participants shared gaining insights from generated outputs i.e. personalised organisational report which were thought to be accurate and relevant. Overall, findings reflected that organisations are in an advantageous position to realise plans to further strengthen their cyber defences. Elements from various prominent frameworks discussed as part of extant literature were partially incorporated (i.e. elements from existing frameworks that were evidenced in findings from the research study discussed in the previous section). Organisations appeared to possess aspirations to continuously improve their sociotechnical defences that are in place. However, organisational investment for individual's personal skill development occurred when it is aligned with the job function being performed by the person – as opposed to nurturing in-house talent. Findings from this research study to elicit reflection evidenced:

- Positive attitudes that exist towards technologies and people can serve to further strengthen defences against UIT
- Subjective norms have a strong foundation but can be further reinforced through intervention tools that allow adequate time for human centric approaches to be adopted in order to perpetuate desirable behaviour
- Strong perceived control exists through the use of technologies and on individual levels however, challenges can emerge from tackling larger systems or groups
- Organisational self-positioning (based on certain considerations) is critical in the accurate assessment of defences that are in place
- A deeper understanding must be established by organisations to understand types of errors (Reason, 1990b) and types of tasks that originate errors (Rasmussen, 1983) in order to tackle UIT in a meaningful way

Potential areas for improvement involved revising the way inputs are organised in the assessment tool, recognising alternatives to formalised training programmes in the assessment tool and, reflecting hybrid/remote work settings in inputs that have been implemented following the Covid-19 pandemic. In conclusion, the website developed from the proposed framework allowed participants to reflect and gain new perspectives for tackling unintentional insider threat at their organisations.

7.3 Contributions

The work presented in this thesis has made the following contributions:

- Contribution to the computer science domain by applying human factors domain models to the challenge of insider threat so as to identify areas that can benefit from these established approaches
- Contribution to the human factors domain by extending the application of existing approaches to the field of insider threat
- Contribution to industry through grounding this research project in industry context and the design of a tool that can serve as an intervention for unintentional insider threat
- Wider academic contribution to the community through publishing findings from research studies

7.4 Limitations

Limitations emerging from the selection of frameworks

Four frameworks were presented in Chapter 2, Section 2.4 and while these bodies of work were deemed suitable for the purposes of this work, certain limitations can arise from this selection. For instance, while the entire framework might not have been well suited it could contain an element or aspect that could be relevant to this work. As this work is conducted in the UK and while notable frameworks from the United States have been presented frameworks from other global regions could have been excluded due to regional and legal applicability.

Limitations emerging from the critical analysis of CERT's guide

The work presented in Chapter 3 conducted a critical analysis of CERT's guide through document analysis method and utilised a heuristic approach to develop organisational personas which served as case studies in order to evaluate usability issues in sociotechnical systems when implementing these recommendations in real-world settings. Case studies were used to enriched the evaluation and provide an industry context. While these methods allow a systematic approach, are less time constraining, cost effective and provide access to information which is broadly representative of the data they represent compared to other approaches, these methods i.e. document analysis and a heuristic approach are inherently subjective as they rely on the evaluator's knowledge, expertise, context and, can highlight aspects that might not necessarily be important in real-world settings (Love, 2013; Friess, 2015).

Limitations of research study based on Critical Decision Method approach

Limitations in this research study in Chapter 4 emerge from snowball sampling that might have limited the diversity of participants and in turn the generalisability of findings. Some participants were already known to the interviewer – this facilitated candid and honest discussions, but this rapport might have also influenced participants' responses to some degree. As participants were asked to recall an incident in the past, whilst this incident was significant in their lived experiences, the application of recall and memory bias might have unintentionally included or excluded information that could have been significant for the findings. In addition, as the study advertised for specific types of unintentional insider threat in specific contexts, which aided in recruiting lay audience participants who might not have otherwise understood the nature of the breach being investigated, it can limit the findings applicability to the broader category of unintentional insider threats that exist. Study advertisements also entailed an *action* towards a goal that resulted in an event which can be discussed as part of the Critical Decision Method technique. This could have subsequently excluded people who were victims of other types of unintentional insider threat.

Limitations of research study guided by the Theory of Planned Behaviour

Limitations of this work presented in Chapter 6 emerge from targeted sampling (Watters and Biernacki, 1989) whereby this study was not open to the wider general public. Targeted sampling was used as participating organisations were difficult to reach due to the social stigma attributed to discussions about cybersecurity practices and weaknesses. As participants were recruited through NCSC i100 partners and authors' personal networks, participating organisations also operated in highly specialised sectors. Due to their collaborations with NCSC, participants might have also possessed more knowledge and awareness pertaining to cybersecurity practices and implementation than perhaps other organisations. When the study was advertised interested organisations were requested to get in touch with the author thus, organisations that took part were interested in cybersecurity and/or had strategically prioritised it. This organisational interest and prioritisation was reflected in the availability and participation of senior leadership staff. However, this subsequently highlights the lack of coverage of organisations that are yet to identify cybersecurity as a strategic priority within this study. Senior staff are a suitable sample in as far as they can represent and have broad oversight of their organisations. However, it can be conceded that they might only have had limited visibility of day-to-day activity and consequently, some aspects of *work-as-done*. The approach utilised to evaluate the framework included a senior stakeholder alongside another participant who was in a junior designation to them. This interpersonal dynamic in an organisational context can limit candid conversations, although it was not outwardly visible during the sessions. Additionally, related to most results derived from qualitative research studies, findings can be said to be true for the state the organisations were in at the time of the study. Aforementioned factors can limit widespread applicability of findings resulting from this study.

7.5 Recommendations for future research

Future direction can involve the framework being adoption by more organisations as a tool for reflecting on challenges posed by unintentional insider threat. Further adoption of the framework in industry settings can provide a measure of effectiveness of the suggestions through incident rates prior and post adoption. Pertaining to the website, future work can provide an opportunity to collect feedback from end users about the design and display of outputs (UX) in the personalised organisation report. In addition to the legend box to explain generated visuals as outputs, future work involve the creation of recommendations (what steps to take) to strengthen defences if they are reportedly weak and in turn strengthen pillars and ultimately defences against UIT whilst maintaining human agency and empowerment within the system. Since participants' attitudes reflected the acceptance of human fallibility that results in errors, additional research can to be conducted to understand and subsequently classify the nature of errors that occur i.e. slips, lapses, mistakes and, violations (GEMS, Reason, 1990b) that result in unintentional insider threat. Future research can also investigate whether aforementioned errors emerge from particular cognitive tasks that involve skill, rule or knowledge based actions (SRK, Rasmussen, 1983). And finally, another possible opportunity for future research can involve examining challenges that emerge from creating and strengthening sociotechnical cyber defences at an organisational level or as part of large-scale complex systems.

7.6 Conclusion

As a sociotechnical system, cyberspace is comprised of both technological and human elements. Despite humans being an integral part of systems there has been a tendency in current approaches to seek to constrain the role humans play and to blame human operators if things go awry as they are considered the most variable element in the system. Additionally, there is limited understanding of casual factors that can influence unintended errors that result in cyber incidents or breaches due, perhaps, to cybersecurity domain being alienated from mainstream operations in industry. This limited understanding results in unintentional insider threat being feared or reduced to an inevitable human condition that must be contained, controlled and predicted. In order to enhance existing approaches and understandings, this thesis provided a systems perspective with which to view and understand unintentional insider threat by challenging the way humans are typically considered within systems. This was achieved through grounding the work in established approaches such as The Epidemiological Triangle (Cassel, 1976), the Swiss Cheese Metaphor (Reason, 1990a), Safety II approach (Hollnagel, 2018), Skills, Rules and Knowledge approach known as SRK (Rasmussen, 1983) and, Generic Error-Modelling System known as GEMS (Reason, 1990b). The Epidemiological Triangle was utilised to visualise the dynamic interdependent relationship between the human, the attack and the environment that can create vulnerabilities in defences and offer a deeper understanding of complex environments that require holistic solutions. A sociotechnical framework was presented that incorporated distinct elements that influence unintentional insider threat. The framework was presented as a web assessment tool to aid senior stakeholders in their understanding and awareness of unintentional insider threat and evaluate the strength of their defences through self-reflection. In order to meaningfully tackle the rapidly growing challenge of unintentional insider threat, all actors in systems must be leveraged to take advantage of this understanding developed from a systems perspective to enhance existing systems – where the human element is not perceived to be the weakest link but rather as critical component that helps to keep systems safe.

References

References

Agrafiotis, I., Nurse, J. R., Buckley, O., Legg, P., Creese, S., & Goldsmith, M. (2015). Identifying attack patterns for insider threat detection. Computer Fraud & Security, 2015(7), 9-17.

Agrafiotis, I., Erola, A., Happa, J., Goldsmith, M., & Creese, S. (2016, May). Validating an insider threat detection system: A real scenario perspective. In 2016 IEEE Security and Privacy Workshops (SPW) (pp. 286-295). IEEE.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Ajzen, I., & Driver, B. L. (1992). Application of the theory of planned behavior to leisure choice. Journal of leisure research, 24(3), 207-224.

Ani, U. D., Daniel, N., Oladipo, F., & Adewumi, S. E. (2018). Securing industrial control system environments: the missing piece. Journal of Cyber Security Technology, 2(3-4), 131-163.

Appelbaum, S. H. (1997). Socio-technical systems theory: an intervention strategy for organizational development. Management decision.

Avison, D. E., Lau, F., Myers, M. D., & Nielsen, P. A. (1999). Action research. Communications of the ACM, 42(1), 94-97.

Bair, J., Bellovin, S. M., Manley, A., Reid, B., & Shostack, A. (2017). That was close: Reward reporting of cybersecurity near misses. Colo. Tech. LJ, 16, 327.

Barach, P., & Small, S. D. (2000). Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems. Bmj, 320(7237), 759-763.

Bell, A. J., Rogers, M. B., & Pearce, J. M. (2019). The insider threat: Behavioral indicators and factors influencing likelihood of intervention. International Journal of Critical Infrastructure Protection, 24, 166-176.

Becker, F. D., & Steele, F. (1995). Workplace by design: Mapping the high-performance workscape. Jossey-Bass.

Bishop, M., & Gates, C. (2008a). Defining the insider threat. In Proceedings of the 4th annual workshop on Cyber security and information intelligence research: developing strategies to meet the cyber security and information intelligence challenges ahead (pp. 1-3).

Bishop, M., Engle, S., Peisert, S., Whalen, S., & Gates, C. (2008b). We have met the enemy and he is us. In Proceedings of the 2008 New Security Paradigms Workshop (pp. 1-12).

Bosnjak, M., Ajzen, I., & Schmidt, P. (2020). The theory of planned behavior: selected recent advances and applications. Europe's Journal of Psychology, 16(3), 352.

Bowen, B., Salem, M. B., Hershkop, S., Keromytis, A., & Stolfo, S. (2009). Designing host and network sensors to mitigate the insider threat. IEEE security & privacy, 7(6), 22-29.

Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative research journal.

Brdiczka, O., Liu, J., Price, B., Shen, J., Patil, A., Chow, R., ... & Ducheneaut, N. (2012, May). Proactive insider threat detection through graph learning and psychological context. In 2012 IEEE Symposium on Security and Privacy Workshops (pp. 142-149). IEEE.

Brenner, J. (2007). ISO 27001 risk management and compliance. Risk management, 54(1), 24-29

Brooks, L., & Alam, M. S. (2015). Designing an information system for updating land records in Bangladesh: Action design ethnographic research (ADER). Information Systems Frontiers, 17(1), 79-93.

Burke, R. M., Shah, M. P., Wikswo, M. E., Barclay, L., Kambhampati, A., Marsh, Z., ... & Hall, A. J. (2019). The norovirus epidemiologic triad: predictors of severe outcomes in US norovirus outbreaks, 2009–2016. The Journal of infectious diseases, 219(9), 1364-1372.

Cappelli, D. M., Desai, A. G., Moore, A. P., Shimeall, T. J., Weaver, E. A., & Willke, B. J. (2007). Management and Education of the Risk of Insider Threat (MERIT): mitigating the risk of sabotage to employers' information, systems, or networks. CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.

Cappelli, D. M., Desai, A. G., Moore, A. P., Shimeall, T. J., Weaver, E. A., & Willke, B. J. (2008). Management and education of the risk of insider threat (MERIT): System dynamics modeling of computer system sabotage. Carnegie-Mellon Univ Pittsburgh PA Software Engineering Inst.

Carayon, P. (2006). Human factors of complex sociotechnical systems. Applied ergonomics, 37(4), 525-535.

Cardenas, A. A., Amin, S., & Sastry, S. (2008, June). Secure control: Towards survivable cyber-physical systems. In 2008 The 28th International Conference on Distributed Computing Systems Workshops (pp. 495-500). IEEE.

Cassel, J. (1976). The contribution of the social environment to host resistance: the Fourth Wade Hampton Frost Lecture. American journal of epidemiology, 104(2), 107-123.

CERT Insider Threat Team Bureau, F. I. P. (2013). Unintentional insider threats: a foundational study.

Chattopadhyay, P., Wang, L., & Tan, Y. P. (2018). Scenario-based insider threat detection from cyber activities. IEEE Transactions on Computational Social Systems, 5(3), 660-675.

Cherns, A. B. (1976). The principles of organizational design. Human Relations, 29(8), 783-792.

Chhikara, J., Dahiya, R., Garg, N., & Rani, M. (2013). Phishing & anti-phishing techniques: Case study. International Journal of Advanced Research in computer science and software engineering, 3(5).

Chinchani, R., Iyer, A., Ngo, H. Q., & Upadhyaya, S. (2005, June). Towards a theory of insider threat assessment. In 2005 International Conference on Dependable Systems and Networks (DSN'05) (pp. 108-117). IEEE

Choucri, N., Daw Elbait, G., & Madnick, S. (2012). What is cybersecurity? Explorations in automated knowledge generation

Colwill, C. (2009). Human factors in information security: The insider threat–Who can you trust these days?. Information security technical report, 14(4), 186-196.

Cooper, R., & Foster, M. (1971). Sociotechnical systems. American Psychologist, 26(5), 467.

Costa, A. C., Roe, R. A., & Taillieu, T. (2001). Trust within teams: The relation with performance effectiveness. European journal of work and organizational psychology, 10(3), 225-244.

Dekker, S. (2011). Patient safety. A human factors approach, 2011.

Dekker, S. (2012). Complexity, signal detection, and the application of ergonomics: Reflections on a healthcare case study. Applied Ergonomics, 43(3), 468-472.

Dekker, S. (2017). The safety anarchist: Relying on human expertise and innovation, reducing bureaucracy and compliance. Routledge.

Douglas, E., LEVA, C., BALFE, N., & CROMIE, S. (2014). Modelling the reporting culture within a modern organisation.

Duhigg, C (2018). Did Uber Steal Google's Intellectual Property? Silicon Valley was built on job-hopping. But when a leader of Google's self-driving-car unit joined Uber, Google filed suit. Now the Feds are on the case. Retrieved from https://www.newyorker.com/magazine/2018/10/22/did-uber-steal-googles-intellectual-property (accessed 06.01.2023)

Dul, J., Bruder, R., Buckle, P., Carayon, P., Falzon, P., Marras, W. S., Wilson, J. R., & van der Doelen, B. (2012). A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics, 55(4), 377-395.

Eldardiry, H., Bart, E., Liu, J., Hanley, J., Price, B., & Brdiczka, O. (2013, May). Multi-domain information fusion for insider threat detection. In 2013 IEEE Security and Privacy Workshops (pp. 45-51). IEEE.

European Commission (2003) Document 32003H0361 Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32003H0361 (accessed 01.02.2019)

Evans, J. S. B. (2012). Spot the difference: distinguishing between two kinds of processing. Mind & Society, 11(1), 121-131.

Framework, D. N. B. D. I. (2014). DRAFT NIST Big Data Interoperability Framework: Volume 3, Use Cases and General Requirements. NIST Special Publication, 1500, 3.

Frantz, E., Dugan, A., Hinchberger, K., Maseth, B., Al Sharfa, O., & Al-Jaroodi, J. (2017, June). SMEs: The effects of strategic management. In 2017 IEEE Technology & Engineering Management Conference (TEMSCON) (pp. 388-393). IEEE.

Friess, E. (2015). Personas in heuristic evaluation: an exploratory study. IEEE Transactions on Professional Communication, 58(2), 176-191.

Glaser, B. G., & Strauss, A. L. (2017). The discovery of grounded theory: Strategies for qualitative research. Routledge.

Goethals, P. L., & Hunt, M. E. (2019). A review of scientific research in defensive cyberspace operation tools and technologies. Journal of Cyber Security Technology, 3(1), 1-46.

Gollmann, D. (2011, June). From access control to trust management, and back–a petition. In IFIP International Conference on Trust Management (pp. 1-8). Springer, Berlin, Heidelberg.

Gordon, J. E. (1949). The epidemiology of accidents. American Journal of Public Health and the Nations Health, 39(4), 504-515.

Greitzer, F. L., & Frincke, D. A. (2010). Combining traditional cyber security audit data with psychosocial data: towards predictive modeling for insider threat mitigation. In Insider threats in cyber security (pp. 85-113). Springer, Boston, MA.

Greitzer, F. L., & Hohimer, R. E. (2011). Modeling human behavior to anticipate insider attacks. Journal of Strategic Security, 4(2), 25-48.

Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2012, January). Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. In 2012 45th Hawaii International Conference on System Sciences (pp. 2392-2401). IEEE.

Greitzer, F., Purl, J., Leong, Y. M., & Becker, D. S. (2018, May). Sofit: Sociotechnical and organizational factors for insider threat. In 2018 IEEE Security and Privacy Workshops (SPW) (pp. 197-206). IEEE.

Groner, R., Groner, M., & Bischof, W. F. (2014). Methods of heuristics. Routledge.

Gulis, G., & Fujino, Y. (2015). Epidemiology, population health, and health impact assessment. Journal of epidemiology, 25(3), 179-180.

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.

Hadnagy, C. (2010). Social engineering: The art of human hacking. John Wiley & Sons.

Haddon Jr, W. (1968). The changing approach to the epidemiology, prevention, and amelioration of trauma: the transition to approaches etiologically rather than descriptively based. American journal of public health and the Nations health, 58(8), 1431-1438.

Hattinger, M., & Eriksson, K. (2015). Action design research: design of e-WIL for the manufacturing industry. In 21st Americas Conference on Information Systems, AMCIS 2015; El Conquistador Resort and Convention CenterFajardo; Puerto Rico; 13 August 2015 through 15 August 2015. (pp. 1-14).

Hendrick,H.(1997).Organizationaldesignandmacroergonomics.InG.Salvendy (Ed.),Hand book of human factors and ergonomics (pp.594-637). New York: John Wiley & Sons.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS quarterly, 75-105.

Hoda, R., Noble, J., & Marshall, S. (2010). Using grounded theory to study the human aspects of software engineering. In Human Aspects of Software Engineering (pp. 1-2).

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. Human factors, 40(2), 254-276.

Hollnagel, E. (1992, March). Coping, coupling and control: the modelling of muddling through. In Proceedings of 2nd interdisciplinary workshop on mental models (pp. 61-73).

Hollnagel, E. (2017). Why is work-as-imagined different from work-as-done?. In Resilient health care, Volume 2 (pp. 279-294). CRC Press.

Hollnagel, E. (2018). Safety–I and safety–II: the past and future of safety management. CRC press.

Hollnagel, E., Wears, R. L., & Braithwaite, J. (2015). From Safety-I to Safety-II: a white paper. The resilient health care net: published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia.

Huang, K., & Pearlson, K. (2019, January). For what technology can't fix: Building a model of organizational cybersecurity culture. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

Huang, H., Tan, J., & Liu, L. (2009, June). Countermeasure techniques for deceptive phishing attack. In 2009 International Conference on New Trends in Information and Service Science (pp. 636-641). IEEE.

Hunker, J., & Probst, C. W. (2011). Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques. J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl., 2(1), 4-27.

Ingram, K. L., Cope, J. G., Harju, B. L., & Wuensch, K. L. (2000). Applying to graduate school: A test of the theory of planned behavior. Journal of Social Behavior and Personality, 15(2), 215.

Ivaturi, K., & Janczewski, L. (2011). A taxonomy for social engineering attacks.

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

Jakobsson, M., & Stamm, S. (2006, May). Invasive browser sniffing and countermeasures. In Proceedings of the 15th international conference on World Wide Web (pp. 523-532).

Johnson, C. (1999). Why human error modeling has failed to help systems development. Interacting with computers, 11(5), 517-524.

Jumaat, N. F., & Tasir, Z. (2014). Instructional scaffolding in online learning environment: A meta-analysis. Proceedings-2014 International Conference on Teaching and Learning in Computing and Engineering, LATICE 2014, 74–77.

Kabanda, S., Tanner, M., & Kent, C. (2018). Exploring SME cybersecurity practices in developing countries. Journal of Organizational Computing and Electronic Commerce, 28(3), 269-282.

Kammüller, F., & Probst, C. W. (2013, May). Invalidating policies using structural information. In 2013 IEEE Security and Privacy Workshops (pp. 76-81). IEEE.

Kandias, M., Mylonas, A., Virvilis, N., Theoharidou, M., & Gritzalis, D. (2010, August). An insider threat prediction model. In International conference on trust, privacy and security in digital business (pp. 26-37). Springer, Berlin, Heidelberg.

Keeney, M., Kowalski, E., Cappelli, D., Moore, A., Shimeall, T., & Rogers, S. (2005). Insider threat study: Computer system sabotage in critical infrastructure sectors. National Threat Assessment Ctr Washington Dc.

Khan, N., Houghton, R.J, & Sharples, S. (2022). Understanding factors that influence unintentional insider threat: a framework to counteract unintentional risks. Cognition, Technology & Work, 24(3), 393-421.

Kim, K. N., Yim, M. S., & Schneider, E. (2017). A study of insider threat in nuclear security analysis using game theoretic modeling. Annals of Nuclear Energy, 108, 301-309.

King, N. (2012). Doing template analysis. Qualitative organizational research: Core methods and current challenges, 426(10.4135), 9781526435620.

Kirwan, B. (1992). Human error identification in human reliability assessment. Part 1: Overview of approaches. Applied ergonomics, 23(5), 299-318.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. IEEE Transactions on systems, man, and cybernetics, 19(3), 462-472.

Kok, S., Abdullah, A., Jhanjhi, N., & Supramaniam, M. (2019). Ransomware, threat and detection techniques: A review. Int. J. Comput. Sci. Netw. Secur, 19(2), 136.

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

Lagerstrom, E., Magzamen, S., Stallones, L., Gilkey, D., & Rosecrance, J. (2016). Understanding risk factor patterns in ATV fatalities: a recursive partitioning approach. Journal of safety research, 59, 23-31.

Larkin, R. D., Lopez Jr, J., Butts, J. W., & Grimaila, M. R. (2014). Evaluation of security solutions in the SCADA environment. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 45(1), 38-53.

Le Coze, J. C. (2015). Reflecting on Jens Rasmussen’s legacy. A strong program for a hard problem. Safety science, 71, 123-141.

Leavitt, H. J. (1965). Applied Organizational Change in Industry, I JG March (ed.) Handbook of Organizations. Chicago: Rand McNally, 1144, 1170.

Legg, P. A., Buckley, O., Goldsmith, M., & Creese, S. (2015). Automated insider threat detection system using user and role-based profile assessment. IEEE Systems Journal, 11(2), 503-512.

Levine, L., & Woody, C. (2010, November). System of systems analysis of catastrophic events: A preliminary investigation of unprecedented scenarios. In 2010 IEEE International Conference on Technologies for Homeland Security (HST) (pp. 467-472). IEEE.

Leveson, N. G. (2004).A new accident model for engineering safer systems. Safety Science, 42(4), 237-270. doi:10.1016/j.ssci.2011.11.009

Liginlal, D., Sim, I., & Khansa, L. (2009). How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. computers & security, 28(3-4), 215-228.

Liu, D., Wang, X., & Camp, L. J. (2009, February). Mitigating inadvertent insider threats with incentives. In International Conference on Financial Cryptography and Data Security (pp. 1-16). Springer, Berlin, Heidelberg.

Love, P. (2013). Document analysis. In Research in the college context (pp. 99-112). Routledge.

Maasberg, M., Warren, J., & Beebe, N. L. (2015, January). The dark side of the insider: detecting the insider threat through examination of dark triad personality traits. In 2015 48th Hawaii International Conference on System Sciences (pp. 3518-3526). IEEE.

Maccani, G., Donnellan, B., & Helfert, M. (2014, May). Action design research in practice: the case of smart cities. In International Conference on Design Science Research in Information Systems (pp. 132-147). Springer, Cham.

Madill, A. (2011). Interaction in the semi-structured interview: A comparative analysis of the use of and response to indirect complaints. Qualitative Research in Psychology, 8(4), 333-353.

Magklaras, G. B., & Furnell, S. M. (2001). Insider threat prediction tool: Evaluating the probability of IT misuse. Computers & security, 21(1), 62-73.

Martinez-Moyano, I. J., Conrad, S. H., Rich, E. H., & Andersen, D. F. (2006, December). Modeling the emergence of insider threat vulnerabilities. In Proceedings of the 2006 Winter Simulation Conference (pp. 562-568). IEEE.

Matzler, K., & Renzl, B. (2006). The relationship between interpersonal trust, employee satisfaction, and employee loyalty. Total quality management and business excellence, 17(10), 1261-1271.

McCarthy, K. (2020). UK finds itself almost alone with centralized virus contact-tracing app that probably won't work well, asks for your location, may be illegal. The Register, 5.

Mearns, K. J., & Flin, R. (1999). Assessing the state of organizational safety—culture or climate?. Current psychology, 18(1), 5-17.

Mills, J. U., Stuban, S. M., & Dever, J. (2017). Predict insider threats using human behaviors. IEEE Engineering Management Review, 45(1), 39-48.

Mittal, S. (2015). Understanding the human dimension of cyber security. Indian Journal of Criminology & Criminalistics (ISSN 0970–4345), 34(1), 141-152.

Mohurle, S., & Patil, M. (2017). A brief study of wannacry threat: Ransomware attack 2017. International Journal of Advanced Research in Computer Science, 8(5), 1938-1940.

Mokube, I., & Adams, M. (2007, March). Honeypots: concepts, approaches, and challenges. In Proceedings of the 45th annual southeast regional conference (pp. 321-326).

Mpolya, E. A., Furuse, Y., Nukiwa, N., Suzuki, A., Kamigaki, T., & Oshitani, H. (2009). Pandemic (H1N1) 2009 virus viewed from an epidemiological triangle model. Journal of Disaster Research, 4(5), 1.

Muller, M. J., & Kogan, S. (2010). Grounded theory method in HCI and CSCW. Cambridge: IBM Center for Social Software, 28(2), 1-46.

Mundie, D. A., Perl, S., & Huth, C. L. (2013, June). Toward an ontology for insider threat research: Varieties of insider threat definitions. In 2013 third workshop on socio-technical aspects in security and trust (pp. 26-36). IEEE.

National Cyber Security Centre, NCSC (2012). 10 steps to cyber security: Guidance on how organisations can protect themselves in cyberspace, including the 10 steps to cyber security. Retrieved from https://www.ncsc.gov.uk/collection/10-steps-to-cyber-security (accessed 11.11.2018)

National Cyber Security Centre, NCSC (2019). 10 steps to cyber security: User education and awareness. Retrieved from https://www.ncsc.gov.uk/collection/10-steps-to-cyber-security/the-10-steps/user-education-and-awareness (accessed 23.11.2019)

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work.

Nguyen, N., Reiher, P., & Kuenning, G. H. (2003, June). Detecting insider threats by monitoring system call activity. In IEEE Systems, Man and Cybernetics SocietyInformation Assurance Workshop, 2003. (pp. 45-52). IEEE.

Nobles, C. (2018). Botching human factors in cybersecurity in business organizations. HOLISTICA–Journal of Business and Public Administration, 9(3), 71-88.

Norman, D. A., & Draper, S. W. (1986). User centered system design: New perspectives on human-computer interaction.

Novinson, M. (2020). The 11 Biggest Ransomware Attacks Of 2020 (So Far).[online] CRN.

Nurse, J. R., Buckley, O., Legg, P. A., Goldsmith, M., Creese, S., Wright, G. R., & Whitty, M. (2014, May). Understanding insider threat: A framework for characterising attacks. In 2014 IEEE security and privacy workshops (pp. 214-228). IEEE.

Ogiela, M. R., & Ogiela, U. (2012). Linguistic protocols for secure information management and sharing. Computers & Mathematics with Applications, 63(2), 564-572.

Okosun, O., & Ilo, U. (2022). The evolution of the Nigerian prince scam. Journal of Financial Crime, (ahead-of-print).

Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. Journal of research in personality, 36(6), 556-563.

Plumb, T (2022). Protecting your organization from rising software supply chain attacks. Retrieved from https://venturebeat.com/security/protecting-your-organization-from-rising-software-supply-chain-attacks/ (accessed 08.01.2023)

Predd, J., Pfleeger, S. L., Hunker, J., & Bulford, C. (2008). Insiders behaving badly. IEEE Security & Privacy, 6(4), 66-70.

Punithavathani, D. S., Sujatha, K., & Jain, J. M. (2015). Surveillance of anomaly and misuse in critical networks to counter insider threats using computational intelligence. Cluster Computing, 18(1), 435-451.

Rabiee, F. (2004). Focus-group interview and data analysis. Proceedings of the nutrition society, 63(4), 655-660.

Ramzan, Z. (2010). Phishing attacks and countermeasures. Handbook of information and communication security, 433-448.

Rashid, T., Agrafiotis, I., & Nurse, J. R. (2016, October). A new take on detecting insider threats: exploring the use of hidden markov models. In Proceedings of the 8th ACM CCS International workshop on managing insider security threats (pp. 47-56).

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE transactions on systems, man, and cybernetics, (3), 257-266.

Reason, J. (1990a). The contribution of latent human failures to the breakdown of complex systems. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 327(1241), 475-484.

Reason, J. (1990b). Human error. Cambridge university press.

Reason, J. (1998). Achieving a safe culture: theory and practice. Work & stress, 12(3), 293-306.

Reason, J., Hollnagel, E., & Paries, J. (2006). Revisiting the Swiss cheese model of accidents. Journal of Clinical Engineering, 27(4), 110-115.

Reason, J., Manstead, A., Stradling, S., Baxter, J., & Campbell, K. (1990). Errors and violations on the roads: a real distinction?. Ergonomics, 33(10-11), 1315-1332.

Reeves, T. (2006). Design research from a technology perspective. In Educational design research (pp. 64-78). Routledge.

Rogers, R. (2008). Structured interviews and dissimulation. Clinical assessment of malingering and deception, 3, 301-322.

Rowley, J. (2014). Designing and using research questionnaires. Management research review, 37(3), 308-330.

Rumsfeld, D. (2011). Known and unknown: a memoir. Penguin.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. Journal of personality and social psychology, 91(6), 995.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human factors, 50(3), 540-547.

Schuh, G., Potente, T., Wesch-Potente, C., Weber, A. R., & Prote, J. P. (2014). Collaboration Mechanisms to increase Productivity in the Context of Industrie 4.0. Procedia Cirp, 19, 51-56.

Schultz, E. E. (2002). A framework for understanding and predicting insider attacks. Computers & security, 21(6), 526-531.

Sears, D. O. (1983). The person-positivity bias. Journal of personality and Social Psychology, 44(2), 233.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. MIS quarterly, 37-56.

Serrano, J (2021). WhatsApp Head Says New Pegasus Spyware Investigation Coincides With Its Findings From 2019 Attack. Retrieved from https://gizmodo.com/whatsapp-head-says-new-pegasus-spyware-investigation-co-1847356967 (accessed 05.01.2023)

Shabtai, A., Bercovitch, M., Rokach, L., Gal, Y. A., Elovici, Y., & Shmueli, E. (2016). Behavioral study of users when interacting with active honeytokens. ACM Transactions on Information and System Security (TISSEC), 18(3), 1-21.

Shappell, S. A., & Wiegmann, D. A. (2003). Reshaping the way we look at general aviation accidents using the human factors analysis and classification system.

Sharples, S. (2018, August). Workload II: A future paradigm for analysis and measurement. In Congress of the International Ergonomics Association (pp. 489-498). Springer, Cham.

Sharples, S., Brown, M., Pinchin, J., Blum, J., Nagiyev, A., Ryan, B., ... & Blakey, J. (2015, August). Ubiquitous technologies for capture of real-world performance. In Proceedings 19th triennial congress of the IEA (Vol. 9, p. 14).

Sharples, S., & Houghton, R. J. (2017). The field becomes the laboratory? The impact of the contextual digital footprint on the discipline of E/HF. Ergonomics, 60(2), 270-283.

Sherer, S. A. (2014). Advocating for action design research on IT value creation in healthcare. Journal of the Association for Information Systems, 15(12), 2.

Smith, J. A., & Shinebourne, P. (2012). Interpretative phenomenological analysis. American Psychological Association.

Sood, A. K., Zeadally, S., & Bansal, R. (2015). Exploiting trust: stealthy attacks through socioware and insider threats. IEEE Systems Journal, 11(2), 415-426.

Spector, M. D., & Jones, G. E. (2004). Trust in the workplace: Factors affecting trust formation between team members. The Journal of social psychology, 144(3), 311-321.

Spitzner, L. (2003, December). Honeypots: Catching the insider threat. In 19th Annual Computer Security Applications Conference, 2003. Proceedings. (pp. 170-179). IEEE.

Stanton, N. A., Booth, R. T., & Stammers, R. B. (1992). Alarms in human supervisory control: A human factors perspective. International Journal of Computer Integrated Manufacturing, 5(2), 81-93.

Sterman, J. D. (2006). Learning from evidence in a complex world. American journal of public health, 96(3), 505-514.

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge university press.

Suganya, V. (2016). A review on phishing attacks and various anti phishing techniques. International Journal of Computer Applications, 139(1), 20-23.

Theis, M., Trzeciak, R. F., Costa, D. L., Moore, A. P., Miller, S., Cassidy, T., & Claycomb, W. R. (2019). Common sense guide to mitigating insider threats.

Thompson, P. (2004, September). Weak models for insider threat detection. In Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense III (Vol. 5403, pp. 40-48). SPIE.

Thong, J. Y., & Yap, C. S. (1995). CEO characteristics, organizational characteristics and information technology adoption in small businesses. Omega, 23(4), 429-442.

Tiwary, D. K., & Pradesh, U. (2011). Security and ethical issues in it: an organization’s perspective. International Journal of Enterprise Computing and Business Systems, 1(2), 2230-8849.

Treisman, M. (1965). Signal detection theory and Crozier's law: Derivation of a new sensory scaling procedure. Journal of Mathematical Psychology, 2(2), 205-218.

Tuor, A., Kaplan, S., Hutchinson, B., Nichols, N., & Robinson, S. (2017). Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. arXiv preprint arXiv:1710.00811.

Verizon (2020) Data Breach Investigations Report. Retrieved from https://enterprise.verizon.com/resources/reports/2020-data-breach-investigations-report.pdf. (accessed 12.01.2021)

von Solms, B., & von Solms, R. (2018). Cybersecurity and information security–what goes where?. Information & Computer Security.

Walker, G. H., Stanton, N. A., Salmon, P. M., & Jenkins, D. P. (2008). A review of sociotechnical systems theory: a classic concept for new command and control paradigms. Theoretical issues in ergonomics science, 9(6), 479-499.

Wall, D. S. (2013). Enemies within: Redefining the insider threat in organizational security policy. Security journal, 26(2), 107-124.

Wang, E. K., Ye, Y., Xu, X., Yiu, S. M., Hui, L. C. K., & Chow, K. P. (2010, December). Security issues and challenges for cyber physical system. In 2010 IEEE/ACM Int'l

Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 733-738). IEEE.

Ward, M & Hutton (2021). Research Briefing: Business statistics. Retrieved from commonslibrary.parliament.uk (accessed 09.11.2021)

Watters, J. K., & Biernacki, P. (1989). Targeted sampling: Options for the study of hidden populations. Social problems, 36(4), 416-430.

Weber, R. H., & Studer, E. (2016). Cybersecurity in the Internet of Things: Legal aspects. Computer Law & Security Review, 32(5), 715-728.

Wiggins, J. S. (Ed.). (1996). The five-factor model of personality: Theoretical perspectives. Guilford Press.

Wilson, J. R., & Sharples, S. (Eds.). (2015). Evaluation of human work. CRC press.

Wong, B. W. (2003). Critical decision method data analysis. The handbook of task analysis for human-computer interactions. Mahwah: Lawrence Erlbaum Associates, 327-46.

Woo, T. H. (2019). Game theory based complex analysis for nuclear security using non-zero sum algorithm. Annals of Nuclear Energy, 125, 12-17.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. Ergonomics, 38(11), 2371-2393.

Wu, J., & Zha, P. (2020). Public health intervention framework for reviving economy amid the COVID-19 pandemic (2): Use of personalized mitigation measures beyond the epidemiological model limits. Available at SSRN 3592067.

Yaqoob, I., Hussain, S. A., Mamoon, S., Naseer, N., Akram, J., & ur Rehman, A. (2017). Penetration testing and vulnerability assessment. Journal of Network Communications and Emerging Technologies (JNCET) www. jncet. org, 7(8).

Yayla, A. (2011). Controlling insider threats with information security policies.

Zargar, A., Nowroozi, A., & Jalili, R. (2016, September). XABA: A zero-knowledge anomaly-based behavioral analysis method to detect insider threats. In 2016 13th International Iranian society of cryptology conference on information security and cryptology (ISCISC) (pp. 26-31). IEEE.

Appendices

Appendices

Appendix 1: Industry Insights

This Appendix describes industry collaborations that informed various elements of this research project, including a cumulative six-month period of immersed experiences in industry contexts by the author. Experiences from different types of collaborations are shared to provide the reader with insights of real-world setting under which approaches to tackle unintentional insider threat (and cybersecurity more widely) are designed to be implemented. Prior to a detailed discussion of industry experiences, the table below provides an overview of activities that helped ground this PhD in an industry perspective:

Partner No.	No. of Partners	Partnership through	Sector	Activities	Contribution to PhD
1	2	Centre for Doctoral Training	Manufacturing	Meetings with partner representatives and supervisors	Provided need-based examples of cybersecurity concerns being faced by industry
2	1	Supervisor	Not-for-profit, independent technology and innovation incubator for connected technologies	Industry placement for a period of three months	Understanding the ownership and responsibility for cybersecurity aspects when designing and

					implementing technologies
3	1	Author	Government department providing advice and support to other sectors on ways to safeguard against computer security threats	Industry placement for a period of three months	Application of various risk and safety engineering models to challenges associated to the wider cybersecurity domain
4	6	Author and Partner no. 3	SMEs, large, and, not-for-profit organisations	Various activities held during three hour session(s) shown in Figure 11	Organisations recruited as participants for multiple sessions held as part of the second research study

Table to provide an overview of industry activities

Manufacturing sector

Through the Centre for Doctoral Training two industry partners within the sector of manufacturing were established: High Value Manufacturing (HMV) Catapult and Warwick

Manufacturing Group (WMG). In the initiation stage of this project, industry partners had identified cybersecurity as the broad area of interest and relevance to them. Whilst these original industry partners were intended to be the main supporting partners for the PhD, due to capacity limitations this was not feasible. Therefore, alternative industry partners were identified to support the work from year 2 of the PhD onwards. Meetings with industry partner representatives and supervisors were held over year 1 of the PhD that were aimed at understanding the solutions being adopted and developed by the two industry partners. The aim of these discussions was to also explore specific challenge areas within the broader topic of cybersecurity that are more widespread in industry and could benefit from a human centric approach. In the context of manufacturing, a field that seeks to protect intellectual property pertaining to innovative approaches being applied to gain a competitive edge, industry partners shared the implemented of a variety of cybersecurity related solutions. Furthermore, they were devoted to enhancing off-the-shelf cybersecurity solutions in-house to protect their manufacturing related intellectual property.

Having understood the industry partners' motivations and the various cyber defences in place, the industry partner representative was requested to list the top cybersecurity related concerns they experience most frequently. Despite having implemented state-of-the-art solutions, the area of top concern emerged from well-meaning insiders propping access-controlled doors open for a variety of reasons. Reasons included insiders going outside for short breaks, increasing ventilation in corridors, to gain quick access paths to other controlled areas of the premises and to conduct informal meetings with peers from other parts of the building. Whilst industry partners had undertaken numerous actions to correct this behaviour, it continued to be a top concern for them whereby which the entire cybersecurity posture could be compromised by well-intentioned insiders who apparently failed to fully grasp the potential impact of their action.

Technology and innovation sector

Following the first research study, an industry placement for a period of three months was arranged through the Supervisor’s professional network at Connected Places Catapult. This industry partner is a not-for-profit, independent technology and innovation incubator for disruptive small-to-medium-sized enterprises (SMEs). SMEs that collaborate with the industry partner focused on providing innovative solutions to societal challenges through the application of connected technologies. The internal structure was formed of several micro-teams (10 people or less) each working on a specific theme and reporting to a Team Manager, who in turn reported to a centralised Director responsible for those portfolios. Each team had a range of specialities and skillsets (with little to no duplication) that allowed them to work closely with SMEs at various stages of their development i.e. informing the product, providing testing beds for prototypes, marketing, identifying further fundings streams, assistance with recruitment etc. Teams often provided SMEs with group sessions and one-to-one support. The peer structure within teams was supportive and inclusive as each member brought their own expertise to the table. Teams were lean and efficient as projects were mutually informed by honing in a variety of inputs. Queries were sent to the appropriate person who would be the assigned as the ‘Lead’ on specific projects. After soliciting advice for the wider team, it was the Lead’s prerogative to incorporate aspects into projects as they saw fit.

This industry placement by the author had commenced shortly after the partner’s Trustee Board identified cybersecurity as a priority area within the wider organisation and mandated a cybersecurity team. The organisational chart reflected two positions in the cybersecurity team reporting to the Director. Both positions aimed to help inform and support four teams’ efforts in all associated projects, as and when requested by the Team Manager. Two weeks after the commencement of the internship, the second cybersecurity team position became vacant and stayed vacant for the remainder of the internship. As the role being fulfilled by the internship was newly created, there was limited understanding of what the role entailed. It was ultimately seen by others as a cross-cutting position that worked across four competing micro teams. This meant that the role was consulted on an ad hoc basis for projects at short notice and the author was often less trusted as an 'outsider' who might hinder or sabotage efforts rather than further initiatives. This role was treated by team members as a position that was retrofitted and superimposed on to existing team structures.

Informing cybersecurity aspects to projects also uncovered new challenges. When working as a link in between SMEs, councils and end users, responsibility for cybersecurity aspects became highly contentious. SMEs did not believe cybersecurity to be their domain of expertise or responsibility (exhibiting a resistance to 'cybersecurity by design'); teams' skillsets did not include generalist cybersecurity knowledge; the councils did not possess the skills to identify cybersecurity robustness or vulnerabilities, and; end users did not appear to be fully knowledgeable of the consequences resulting from a breach in order to effectively prioritise this domain. Informing cybersecurity related elements through various meetings reinforced the idea amongst team members that retrofitting these aspects would propagate debates, require additional effort to implement regulations that might be inconvenient to SMEs that they support and, ultimately hinder he team's key performance indicators.

Governmental sector and subsequent partnerships

The publication of findings from the first research study led to another three-month industry placement. The author secured a partnership with National Cyber Security Centre (NCSC), a UK governmental department which provides advice and support to other sectors on ways to avoid computer security related threats. As part of this placement numerous risk and safety engineering models that have enjoyed success in other domains were applied to cybersecurity with a view to integrate, as opposed to superimpose, cybersecurity within organisations.

Simultaneously, published findings presenting a framework were recognised as innovative by this industry partner. Consequently, select portfolio of enterprises that collaborated with this industry partner were invited to participate in the second research study presented in this thesis. Numerous three-hour sessions were held with industry partners entailing activities for engaging with the self-reflection tool. Participating organisations shared the model of an exclusive role existing within their organisations that was responsible for the overall cybersecurity of all operations. This role was separate and distinct from IT functions performed by other people. However, all organisations participating in the research study shared a high baseline across all their employees for generalist cybersecurity knowledge (such as good cyber hygiene).

Appendix 2: Recommendations recategorized by organisation size

Chapter No.	Cybersecurity recommendations in CERT’s guide	
	a. All Organisations	b. Large Organisations
1	Conduct a physical asset inventory. Identify asset owners’ assets and functions and identify the type of data on the system.	-
	Understand what data your organization processes by speaking with data owners and users from across your organization.	-

	Identify and document the software configurations of all assets.	-
	Prioritize assets and data to determine the high-value targets.	-
2	Ensure that legal counsel determines the legal framework the team will work in.	Formalize an insider threat program (with a senior official of the organization appointed as the program manager) that can monitor for and respond to insider threats.
	Establish policies and procedures for addressing insider threats that include HR, Legal Counsel, Security, Management, and IA.	Implement insider threat detection rules into SIEM systems. Review logs on a continuous basis and ensure watch lists are updated.
	Consider establishing a contract with an outside consulting firm that is capable of providing incident response capabilities for all types of incidents, if the organization has not yet developed the expertise to conduct a legal, objective, and thorough inquiry.	Ensure the insider threat team meets on a regular basis and maintains a readiness state.

3	Ensure that senior management advocates, enforces, and complies with all organizational policies. Policies that do not have management buy-in will fail and not be enforced equally. Management must also comply with policies. If management does not do so, subordinates will see this as a sign that the policies do not matter or they are being held to a different standard than management. Your organization should consider exceptions to policies in this light as well.	-
	Ensure that management briefs all employees on all policies and procedures. Employees, contractors, and trusted business partners should sign acceptable-use policies and acceptable workplace behavior policies upon their hiring and once every year thereafter or when a significant change occurs. This is also an opportunity for your organization and employees, contractors, or trusted business partners to reaffirm any nondisclosure agreements.	-

	Ensure that management makes policies for all departments within your organization easily accessible to all employees. Posting policies on your organization’s internal website can facilitate widespread dissemination of documents and ensure that everyone has the latest copy.	-
	Ensure that management makes annual refresher training for all employees mandatory. Refresher training needs to cover all facets of your organization, not just information security. Training should encompass the following topics: human resources, legal counsel, physical security, and any others of interest. Training can include, but is not limited to, changes to policies, issues that have emerged over the past year, and information security trends.	-
	Ensure that management enforces policies consistently to prevent the appearance of favoritism and injustice. The Human Resources department should have policies and procedures in place that	-

	specify the consequences of particular policy violations. This will facilitate clear and concise enforcement of policies.	
4	Ensure that potential employees have undergone a thorough background investigation, which at a minimum should include a criminal background and credit check.	-
	Encourage employees to report suspicious behavior to appropriate personnel for further investigation.	-
	Investigate and document all issues of suspicious or disruptive behavior.	-
	Enforce policies and procedures consistently for all employees.	-
	Consider offering an EAP. These programs can help employees deal with many personal issues confidentially.	-
5	Enhance monitoring of employees with an impending or ongoing personnel issue, in accordance with organizational policy and laws. Enable additional auditing and monitoring controls outlined in policies	-

	and procedures. Regularly review audit logs to detect activities outside of the employee's normal scope of work. Limit access to these log files to those with a need to know.	
	All levels of management must regularly communicate organizational changes to all employees. This allows for a more transparent organization, and employees can better plan for their future.	-
6	Have all employees, contractors, and trusted business partners sign nondisclosure agreements (NDAs) upon hiring and termination of employment or contracts.	Prohibit personal items in secure areas because they may be used to conceal company property or to copy and store company data.
	Ensure that all employees, contractors, and trusted business partners sign workplace violence prevention and/or appropriate workplace behaviors documentation upon hiring.	Conduct a risk assessment of all systems to identify critical data, business processes, and mission-critical systems. (See NIST Special Publication 800-30, Risk Management Guide for Information Technology Systems for guidance [NIST 2002].) Be sure to include

		insiders and trusted business partners as part of the assessment. (See Section 3.2.1, “Threat-Source Identification,” of NIST SP 800-30.)
	Ensure each trusted business partner has performed background investigations on all of its employees who will have access to your organization’s systems or information. These should be commensurate with your organization’s own background investigations and required as a contractual obligation.	Implement data encryption solutions that encrypt data seamlessly and that restrict encryption tools to authorized users, as well as restrict decryption of organization-encrypted data to authorized users.
	If your organization is acquiring companies during a merger or acquisition, perform background investigations on all employees to be acquired, at a level commensurate with your organization’s policies.	Implement a clear separation of duties between regular administrators and those responsible for backup and restoration.
	Prevent sensitive documents from being printed if they are not required for business purposes. Insiders could take a printout of their own or someone else’s sensitive document from a printer, desk,	Forbid regular administrators’ access to system backup media or the electronic backup processes.

	office, or from garbage. Electronic documents can be easier to track.	
	Avoid direct connections with the information systems of trusted business partners if possible. Provide partners with task-related data without providing access to your organization’s internal network.	-
	Restrict access to the system backup process to only administrators responsible for backup and restoration.	-
7	Establish a social media policy that defines acceptable uses of social media and information that should not be discussed online.	Consider monitoring the use of social media across the organization, limited to looking in a manner approved by legal counsel for postings by employees, contractors, and business partners.
	Include social media awareness training as part of the organization’s security awareness training program.	-
	Encourage users to report suspicious emails or phone calls to the information	-

	security team, who can track these emails to identify any patterns and issue alerts to users.	
8	Establish a work culture that measures success based on appropriate metrics for the work environment. For instance, knowledge workers might measure their success based on outcomes and efficiency instead of metrics that are better suited for a production line.	-
	Encourage employees to think through projects, actions, and statements before committing to them.	-
	Create an environment that encourages focusing upon one thing at a time, rather than multitasking.	-
	Offer employees who are under stress options to de-stress, such as massages, time off, games, or other social but non-project oriented activities.	-

	Routinely monitor employee workloads to make sure that they are commensurate with the employee’s skills and available resources.	-
9	Develop and implement an enterprise-wide training program that discusses various topics related to insider threat. The training program must have the support of senior management to be effective. Management must be seen participating in the course and must not be exempt from it, which other employees could see as a lack of support and an unequal enforcement of policies.	The information security team can conduct periodic inspections by walking through areas of your organization, including workspaces, and identifying security concerns. Your organization should bring security issues to the employee’s attention in a calm, nonthreatening manner and in private. Employees spotted doing something good for security, like stopping a person without a badge, should be rewarded. Even a certificate or other item of minimal value goes a long way to improving

		employee morale and increasing security awareness. Where possible, these rewards should be presented before a group of the employee's peers. This type of program does not have to be administered by the security team but could be delegated to the employee's peer team members or first-level management.
	Train all new employees and contractors in security awareness, including insider threat, before giving them access to any computer system. Make sure to include training for employees who may not need to access computer systems daily, such as janitorial and maintenance staff. These users may require a special training program that covers security scenarios they may encounter, such as social engineering, active shooter, and sensitive documents left out in the open.	-

	Train employees continuously. However, training does not always need to be classroom instruction. Posters, newsletters, alert emails, and brown-bag lunch programs are all effective training methods. Your organization should consider implementing one or more of these programs to increase security awareness.	-
	Establish an anonymous or confidential mechanism for reporting security incidents. Encourage employees to report security issues and consider incentives to reporting by rewarding those who do.	-
10	Establish account management policies and procedures for all accounts created on all information systems. These policies should address how accounts are created, reviewed, and terminated. In addition, the policy should address who authorizes the account and what data they can access.	Review systems and risk to determine the feasibility of centrally managing user accounts.

	Perform audits of account creation and password changes by system administrators. The account management process should include creation of a trouble ticket by the help desk. (Help desk staff should not be able to create accounts.) Your organization could confirm the legitimacy of requests to reset passwords or create accounts by correlating such requests with help desk logs.	If using a central account management system, add contractors to groups linked to projects, organizations, or other logical groups. This allows administrators to quickly identify contractors and change access permissions. Accounts themselves might contain contractor status tipoffs, for example, putting “CONT” in the account name or description.
	Define password requirements and train users on creating strong passwords. Some systems may tolerate long passwords. Encourage users to use passphrases that include proper punctuation and capitalization, thereby increasing passphrase strength and making it more memorable to the user.	-
	Security training should include instruction to block visual access to others as users type their passcodes.	-

	Ensure all shared accounts are absolutely necessary and are addressed in a risk management decision.	-
11	Conduct periodic account reviews to avoid privilege creep. Employees should have sufficient access rights to perform their everyday duties. When an employee changes roles, the organization should review the employee's account and rescind permissions that the employee no longer needs.	Implement separation of duties for all roles that affect the production system. Require at least two people to perform any action that may alter the system.
	-	Use multifactor authentication for privileged user or system administrator accounts. Requiring multifactor authentication will reduce the risk of a user abusing privileged access after an administrator leaves your organization, and the increased accountability of multifactor authentication may inhibit some currently employed, privileged users from committing acts of malfeasance.

		Assuming that the former employee’s multifactor authentication mechanisms have been recovered, the account(s) will be unusable.
12	Implement rules within the SIEM system, to automate alerts.	Ensure that someone regularly monitors the SIEM system. Depending on the environment, this may involve multiple personnel who monitor employee activity full-time.
	Create log management policy and procedures. Ensure they address log retention (consult legal counsel for specific requirements), what logs to collect, and who manages the logging systems.	-
13	Disable remote access to the organization’s systems when an employee or contractor separates from the organization. Be sure to disable	Implement a central management system for mobile devices.

	access to VPN service, application servers, email, network infrastructure devices, and remote management software. Be sure to close all open sessions as well. In addition, collect all company-owned equipment, including multifactor authentication tokens, such as RSA SecurID tokens or smart cards.	
	Include mobile devices, with a listing of their features, as part of the enterprise risk assessment.	Monitor and control remote access to the corporate infrastructure. VPN tunnels should terminate at the furthest perimeter device and in front of an IDS and firewall. This allows for packet inspection and network access control. In addition, IP traffic-flow capture and analysis devices placed behind the VPN concentrator will allow collection of network traffic statistics to help discover anomalies. If personally owned equipment, such as a laptop or home computer, is permitted to access the corporate network, it

		should only be allowed to do so through an application gateway. This will limit the applications available to an untrusted connection.
	Prohibit or limit the use of personally owned devices.	-
	Prohibit devices with cameras in sensitive areas.	-
14	Use monitoring tools to monitor network and employee activity for a period of time to establish a baseline of normal behaviors and trends.	Establish network activity baselines for individual subunits of the organization.
	Deny VPN access to foreign countries where a genuine business need does not exist. White list only countries where a genuine business need exists.34	Determine which devices on a network need to communicate with others and implement access control lists (ACLs), host-based firewall rules, and other technologies to limit communications.

	Establish which ports and protocols are needed for normal network activity, and configure devices to use only these services.	Understand VPN user requirements. Limit access to certain hours and monitor bandwidth consumption. Establish which resources will be accessible via VPN and from what remote IP addresses. Alert on anything that is outside normal activity.
	Determine which firewall and IDS alerts are normal. Either correct what causes these alerts or document normal ranges and include them in the network baseline documentation.	-
15	Carefully audit user access permissions when an employee changes roles within the organization to avoid privilege creep. In addition, routinely audit user access permissions at least annually. Remove permissions that are no longer needed.	Review positions in the organization that handle sensitive information or perform critical functions. Ensure these employees cannot perform these critical functions without oversight and approval. The backup and restore tasks are often overlooked. One person should not be permitted to

		perform both backup and restore functions. Your organization should separate these roles and regularly test the backup and recovery processes (including the media and equipment). In addition, someone other than the backup and restore employees should transport backup tapes off-site.
	Establish account management policies and procedures. Audit account maintenance operations regularly. Account activity should reconcile with help desk documentation.	-
	Require privileged users to have both an administrative account with the minimum necessary privileges to perform their duties and a standard account that is used for every day, non-privileged activities.	-

16	Conduct a risk assessment of the data and services that your organization plans to outsource to a cloud service provider before entering into any agreement. Your organization must en-sure that the service provider poses an acceptable level of risk and has implemented mitigating controls to reduce any residual risks. Your organization must carefully examine all aspects of the cloud service provider to ensure the service provider meets or exceeds your organization's own security practices.	-
	Verify the cloud service provider's hiring practices to ensure it conducts thorough background security investigations on any personnel (operations staff, technical staff, janitorial staff, etc.) before they are hired. In addition, the service provider should conduct periodic credit checks and reinvestigations to ensure that changes in an employee's life situation have not caused any additional unacceptable risks.	-

	Control or eliminate remote administrative access to hosts providing cloud or virtual services.	-
	Understand how the cloud service provider protects data and other organizational assets before entering into any agreement. Verify the party responsible for restricting logical and physical access to your organization's cloud assets.	-
17	Periodically review configuration baselines against actual production systems and determine if any discrepancies were approved. If the changes were not approved, verify a business need for the change.	Implement a change management program within the organization. Ensure that a change control board vets all changes to systems, networks, or hardware configurations. All changes must be documented and include a business reason. Proposed changes must be

		reviewed by information security teams, system owners, data owners, users, and other stakeholders.
	-	The configuration manager must review and submit to the change control board any software developed in-house as well as any planned changes.
18	Store backup media off-site. Ensure media is protected from unauthorized access and can only be retrieved by a small number of individuals. Utilize a professional off-site storage facility; do not simply send backup media home with employees. Encrypt the backup media and manage the encryption keys to ensure backup and recovery are possible.	Implement a backup and recovery process that involves at least two people: a backup administrator and a restore administrator. Both people should able to perform either role.
	Ensure that configurations of network infrastructure devices (e.g., routers, switches, and firewalls) are part of your organization's backup and recovery plan	Regularly test both backup and recovery processes. Ensure that your organization can reconstitute all critical data as defined by the Business

	as well as the configuration management plan.	Continuity Plan and/or Disaster Recovery Plan. Ensure that this process does not rely on any single person to be successful.
19	Establish a cloud computing policy. Organizations must be aware of cloud computing services and how employees may use them to exfiltrate data. Restrict and/or monitor what employees put into the cloud.	Inventory all connections to the organization’s enclave. Ensure that SLAs and/or MOAs are in place. Verify that these connections are still in use and have a justified business need. Implement protection measures, such as firewalls, devices that capture and analyze IP traffic flow, and IDSs at these ingress and egress points so that data can be monitored and scrutinized.
	Monitor the use of printers, copiers, scanners, and fax machines. Where possible, review audit logs from these devices to discover and address any anomalies.	Isolate development networks and disable interconnections to other systems or the internet.
	Create a data transfer policy and procedure to allow sensitive company	-

	information to be removed from organizational systems only in a controlled way.	
	Establish a removable media policy and implement technologies to enforce it.	-
	Restrict data transfer protocols, such as FTP, SFTP, or SCP, to employees with a justifiable business need, and carefully monitor their use.	-
20	Develop an enterprise-wide checklist to use when someone separates from the organization.	Establish a physical-inventory system that tracks all assets issued to an employee.
	Establish a process for tracking all accounts assigned to each employee.	Conduct an inventory of all information systems and audit the accounts on those systems.
	Reaffirm all nondisclosure and IP agreements as part of the termination process.	-
	Notify all employees about any employee’s departure, where permissible and appropriate.	-
	Archive and block access to all accounts associated with a departed employee.	-

	Collect all of a departing employee’s company-owned equipment before the employee leaves the organization.	-
21	Organizational justice (fairness; e.g., compensation aligned internally among employees and externally with industry standards)	-
	Performance-based rewards and recognition (e.g., transparent criteria for promotions and discretionary rewards/recognition based on project performance)	-
	Transparent and respectful communication (e.g., regular employee orientation, mentoring, and expectation setting)	-
	Personal and professional supportiveness (e.g., employee assistance programs and professional development for furthering employee careers and sense of mastery)	-

Appendix 3: Recommendations’ ease of implementation

Recommendations in CERT’s guide assessed by ease of implementation and potential challenges			
Chapter Number	a. All Organisations	Ease of implementation	Potential challenges for implementation for SMEs
1	Conduct a physical asset inventory. Identify asset owners’ assets and functions and identify the type of data on the system.	?	
	Understand what data your organization processes by speaking with data owners and users from across your organization.	✓	In an SME one person might be the asset owner for numerous assets. This can result in vast majority of the responsibility being placed with one person. It would also add additional workload for the person responsible for these assets to keep this list updated with the passage of time.

	Identify and document the software configurations of all assets.	?	Businesses might rely on outsourcing software and so might not have the skill sets to identify software configuration of critical assets.
	Prioritize assets and data to determine the high-value targets.	✓	
2	Ensure that legal counsel determines the legal framework the team will work in.	✓	Legal council is an expensive resource and might be outsourced by SMEs. This can result in an expensive and time-consuming endeavour to develop a framework. If outsourced, it might also mean that employees who do not have a legal background cannot understand the operational parameters of the framework.

	Establish policies and procedures for addressing insider threats that include HR, Legal Counsel, Security, Management, and IA.	✓	This is attainable but might require in-house expertise and a substantial period of time. It might create additional workload for what might be small teams supporting larger core departments. Employees with policy making skills and prior implementation experience would be ideal.
	Consider establishing a contract with an outside consulting firm that is capable of providing incident response capabilities for all types of incidents, if the organization has not yet developed the expertise to conduct a legal, objective, and thorough inquiry.	✕	This might be expensive and seen as an unjustifiable expense for SMEs that are trying to financially break even.

3	Ensure that senior management advocates, enforces, and complies with all organizational policies. Policies that do not have management buy-in will fail and not be enforced equally. Management must also comply with policies. If management does not do so, subordinates will see this as a sign that the policies do not matter or they are being held to a different standard than management. Your organization should consider exceptions to policies in this light as well.	✓	

	Ensure that management briefs all employees on all policies and procedures. Employees, contractors, and trusted business partners should sign acceptable-use policies and acceptable workplace behavior policies upon their hiring and once every year thereafter or when a significant change occurs. This is also an opportunity for your organization and employees, contractors, or trusted business partners to reaffirm any nondisclosure agreements.	×	Requiring partners to sign 'acceptable-use policies' might instate a bureaucratic procedure that does not see much fruition, especially if a business collaborates with numerous trusted business partners. Similarly, new employees with technical skills who might not view this document as a 'norm' in their fields might refuse to agree to another legally binding agreement that sits separately to their employment contract.
	Ensure that management makes policies for all departments within your organization easily	✓	

	accessible to all employees. Posting policies on your organization’s internal website can facilitate widespread dissemination of documents and ensure that everyone has the latest copy.		
	Ensure that management makes annual refresher training for all employees mandatory. Refresher training needs to cover all facets of your organization, not just information security. Training should encompass the following topics: human resources, legal counsel, physical security, and any others of interest. Training can include, but is not	?	Training is valuable to companies in the long-term but bespoke refresher training might be costly. It might also take additional resources to design and deliver refresher training that is engaging for employees. Businesses might need to be able to afford lower productivity during training days.

	limited to, changes to policies, issues that have emerged over the past year, and information security trends.		
	Ensure that management enforces policies consistently to prevent the appearance of favoritism and injustice. The Human Resources department should have policies and procedures in place that specify the consequences of particular policy violations. This will facilitate clear and concise enforcement of policies.	×	This recommendation seems attainable for SMEs but the resources required for monitoring, overseeing implementation of policies and consequences for violations alongside daily operations might make it unmanageable in the long-term.
4	Ensure that potential employees have undergone a thorough background	✓	This might substantially increase the recruitment timeframe and place pressures on operations

	investigation, which at a minimum should include a criminal background and credit check.		for organisations especially when there is a lack of skilled labour to perform a particular job function.
	Encourage employees to report suspicious behavior to appropriate personnel for further investigation.	?	Reporting culture can damage morale and negatively influence peer relationships that might conduct work as part of teams. It might also foster a 'safety climate' resulting in blame placing and disproportionate reprimands rather than instilling a safety culture.
	Investigate and document all issues of suspicious or disruptive behavior.	✓	Means of investigation must be legally permissible according to the legislation of the country of operation. It might also need to be transparently shared with the concerned employee to facilitate open

			communication and offer support if required.
	Enforce policies and procedures consistently for all employees.	✓	
	Consider offering an EAP. These programs can help employees deal with many personal issues confidentially.	✓	
5	Enhance monitoring of employees with an impending or ongoing personnel issue, in accordance with organizational policy and laws. Enable additional auditing and monitoring controls outlined in policies and procedures. Regularly review audit logs to detect activities outside of the	×	This might be difficult to implement if SMEs are uncertain about their survival. If experiencing a shortage of knowledge workers and a high employee turn-over business priorities might make this recommendation unachievable.

	employee's normal scope of work. Limit access to these log files to those with a need to know.		
	All levels of management must regularly communicate organizational changes to all employees. This allows for a more transparent organization, and employees can better plan for their future.	✓	
6	Have all employees, contractors, and trusted business partners sign nondisclosure agreements (NDAs) upon hiring and termination of employment or contracts.	?	Developing various types of NDAs for employees, partners, consultants etc might be costly for SMEs. Obtaining signatures at the time of hiring or termination might prove difficult – requiring time and resources. Depending on the role of an internal

			employee this might be seen as excessive and unnecessary.
	Ensure that all employees, contractors, and trusted business partners sign workplace violence prevention and/or appropriate workplace behaviors documentation upon hiring.	✓	
	Ensure each trusted business partner has performed background investigations on all of its employees who will have access to your organization’s systems or information. These should be commensurate with your organization’s own background investigations and	×	It might prove difficult for organisations to ensure, with 100% certainty, that a background check on a partner's employees has been conducted.

	required as a contractual obligation.		
	If your organization is acquiring companies during a merger or acquisition, perform background investigations on all employees to be acquired, at a level commensurate with your organization’s policies.	?	Uncertain about the applicability of acquisitions for SMEs.
	Prevent sensitive documents from being printed if they are not required for business purposes. Insiders could take a printout of their own or someone else’s sensitive document from a printer, desk, office, or from garbage. Electronic documents can be easier to track.	×	To prevent documents from being printed might not be inclusive for employees with disabilities (known or otherwise).

	Avoid direct connections with the information systems of trusted business partners if possible. Provide partners with task-related data without providing access to your organization’s internal network.	✓	
	Restrict access to the system backup process to only administrators responsible for backup and restoration.	✓	
7	Establish a social media policy that defines acceptable uses of social media and information that should not be discussed online.	✓	
	Include social media awareness training as part of the organization’s	?	This is applicable but with reservations in the context of dealing with tech-savvy

	security awareness training program.		young employees with an advanced skill-set of using social media platforms. Training companies that can provide advanced social media and security training might be financially expensive and time consuming.
	Encourage users to report suspicious emails or phone calls to the information security team, who can track these emails to identify any patterns and issue alerts to users.	✓	
8	Establish a work culture that measures success based on appropriate metrics for the work environment. For instance, knowledge workers might measure	?	Deciding appropriate metrics might be a challenging and time consuming task that can create tensions between employees and management.

	their success based on outcomes and efficiency instead of metrics that are better suited for a production line.		
	Encourage employees to think through projects, actions, and statements before committing to them.	?	This might be problematic to implement if an SME delivers to tight deadlines for clients as it can jeopardise their operations.
	Create an environment that encourages focusing upon one thing at a time, rather than multitasking.	?	Depending on the nature of work it might be difficult to dedicate time to singular tasks without a lengthy re-design of the entire business operations and processes.
	Offer employees who are under stress options to de-stress, such as massages, time off, games, or other social but non-project oriented activities.	?	This might cost precious resources and time that SMEs might not be able to afford.

	Routinely monitor employee workloads to make sure that they are commensurate with the employee’s skills and available resources.	✓	SMEs might be able to monitor workloads but not device effective solutions.
9	Develop and implement an enterprise-wide training program that discusses various topics related to insider threat. The training program must have the support of senior management to be effective. Management must be seen participating in the course and must not be exempt from it, which other employees could see as a lack of support and an unequal enforcement of policies.	?	Developing a training program to cover a wide range of audiences might be difficult, expensive and time consuming.

	Train all new employees and contractors in security awareness, including insider threat, before giving them access to any computer system. Make sure to include training for employees who may not need to access computer systems daily, such as janitorial and maintenance staff. These users may require a special training program that covers security scenarios they may encounter, such as social engineering, active shooter, and sensitive documents left out in the open.	?	Developing bespoke training programs for employees and contractors might prove expensive.

	Train employees continuously. However, training does not always need to be classroom instruction. Posters, newsletters, alert emails, and brown-bag lunch programs are all effective training methods. Your organization should consider implementing one or more of these programs to increase security awareness.	✓	Once a program has been developed and established, continuous training would be beneficial.
	Establish an anonymous or confidential mechanism for reporting security incidents. Encourage employees to report security issues and consider incentives to reporting by rewarding those who do.	?	Anonymity might be difficult for an SME where reports can be reverse-engineered due to the small number of employees. It might also create additional workload for employees who manage these reports once they are submitted.

10	Establish account management policies and procedures for all accounts created on all information systems. These policies should address how accounts are created, reviewed, and terminated. In addition, the policy should address who authorizes the account and what data they can access.	✓	Depending on the number of existing accounts once this policy has been created implementation might prove problematic.
	Perform audits of account creation and password changes by system administrators. The account management process should include creation of a trouble ticket by the help desk. (Help desk staff should not be able to create accounts.) Your	✓	

	organization could confirm the legitimacy of requests to reset passwords or create accounts by correlating such requests with help desk logs.		
	Define password requirements and train users on creating strong passwords. Some systems may tolerate long passwords. Encourage users to use passphrases that include proper punctuation and capitalization, thereby increasing passphrase strength and making it more memorable to the user.	✓	
	Security training should include instruction to block visual access to	✓	

	others as users type their passcodes.		
	Ensure all shared accounts are absolutely necessary and are addressed in a risk management decision.	✓	
11	Conduct periodic account reviews to avoid privilege creep. Employees should have sufficient access rights to perform their everyday duties. When an employee changes roles, the organization should review the employee’s account and rescind permissions that the employee no longer needs.	?	It might be time consuming to determine sufficient access rights as it would require understanding the tasks involved for each designation. In small organisations, employees might have 'excessive privileges' associated to accounts for performing various jobs. It might also be difficult to amend permissions for new roles, as roles are either newly created or the transition between roles is fuzzy.

12	Implement rules within the SIEM system, to automate alerts.	?	Procuring SIEM tools might be expensive for SMEs and companies might lack in-house skills for a correct set up.
	Create log management policy and procedures. Ensure they address log retention (consult legal counsel for specific requirements), what logs to collect, and who manages the logging systems.	✓	While it is possible to implement this recommendation for SMEs, incurring legal costs might be unaffordable.
13	Disable remote access to the organization's systems when an employee or contractor separates from the organization. Be sure to disable access to VPN service, application servers, email, network	✓	

	infrastructure devices, and remote management software. Be sure to close all open sessions as well. In addition, collect all company-owned equipment, including multifactor authentication tokens, such as RSA SecurID tokens or smart cards.		
	Include mobile devices, with a listing of their features, as part of the enterprise risk assessment.	?	This can be problematic as employees can be identified as high risk due to owning a popular brand of mobile technology such as iPhones. Furthermore, a list created on such premises might underestimate malicious insider threats i.e. if the individual is seen to own a 'non-threatening' technology with limited connectivity capabilities

			such as a Nokia 3310 can go undetected.
	Prohibit or limit the use of personally owned devices.	?	Implementing this might largely depend on the type of organisation and the nature of the job. Many SMEs thrive on a work culture that allows employees to work flexibly through organisational or personal devices.
	Prohibit devices with cameras in sensitive areas.	×	For SMEs there might not be an assigned sensitive area but rather sensitive information found in soft critical asset. These might be distributed across the organisation so it might be difficult to explicitly eradicate the camera feature from personal devices when employees

			are interacting with these sensitive objects.
14	Use monitoring tools to monitor network and employee activity for a period of time to establish a baseline of normal behaviors and trends.	?	Whilst establishing a normal baseline can be useful such monitoring might infringe on employee privacy. It might also take a great deal of time to establish a meaningful baseline of behaviours and trends.
	Deny VPN access to foreign countries where a genuine business need does not exist. White list only countries where a genuine business need exists.34	?	SMEs with globally distributed employees or client base might find this disruptive for their operations.
	Establish which ports and protocols are needed for normal network activity, and configure	✓	

	devices to use only these services.		
	Determine which firewall and IDS alerts are normal. Either correct what causes these alerts or document normal ranges and include them in the network baseline documentation.	?	Calibrating sensitivity levels of alerts might be difficult and time consuming.
15	Carefully audit user access permissions when an employee changes roles within the organization to avoid privilege creep. In addition, routinely audit user access permissions at least annually. Remove permissions that are no longer needed.	✓	
	Establish account management policies and	✓	

	procedures. Audit account maintenance operations regularly. Account activity should reconcile with help desk documentation.		
	Require privileged users to have both an administrative account with the minimum necessary privileges to perform their duties and a standard account that is used for every day, non-privileged activities.	?	This recommendation might result in the creation of numerous access pathways to the system. These pathways might be forgotten or result in dirty and incomplete databases – making audits difficult. Employees might suffer from being locked out of the system because they have confused multiple usernames and passwords set for the same system. It might potentially create vulnerabilities in security protocols if users set the

			same passwords for multiple accounts.
16	Conduct a risk assessment of the data and services that your organization plans to outsource to a cloud service provider before entering into any agreement. Your organization must ensure that the service provider poses an acceptable level of risk and has implemented mitigating controls to reduce any residual risks. Your organization must carefully examine all aspects of the cloud service provider to	✓	

	ensure the service provider meets or exceeds your organization’s own security practices.		
	Verify the cloud service provider’s hiring practices to ensure it conducts thorough background security investigations on any personnel (operations staff, technical staff, janitorial staff, etc.) before they are hired. In addition, the service provider should conduct periodic credit checks and reinvestigations to ensure that changes in an employee’s life situation have not caused any additional unacceptable risks.	?	It might be unlikely to obtain internal hiring practices and procedures from business partners and be applicable only at the discretion of the business partner.

	Control or eliminate remote administrative access to hosts providing cloud or virtual services.	✓	
	Understand how the cloud service provider protects data and other organizational assets before entering into any agreement. Verify the party responsible for restricting logical and physical access to your organization’s cloud assets.	✓	
17	Periodically review configuration baselines against actual production systems and determine if any discrepancies were approved. If the changes were not approved, verify a business need for the change.	✓	Although this recommendation might prove beneficial, this process might be time consuming and introduce unnecessary red tape in processes for SMEs.

18	Store backup media off-site. Ensure media is protected from unauthorized access and can only be retrieved by a small number of individuals. Utilize a professional off-site storage facility; do not simply send backup media home with employees. Encrypt the backup media and manage the encryption keys to ensure backup and recovery are possible.	?	It might be expensive for SMEs to afford off-site storage facility. The decision for off-site storage might easily be influenced by competing interests such as business growth.
	Ensure that configurations of network infrastructure devices (e.g., routers, switches, and firewalls) are part of your organization’s backup	✓	

	and recovery plan as well as the configuration management plan.		
19	Establish a cloud computing policy. Organizations must be aware of cloud computing services and how employees may use them to exfiltrate data. Restrict and/or monitor what employees put into the cloud.	?	Restricting employees might create a safety climate and encourage risk taking behaviours amongst employees such as utilising personal devices to bypass an organisation's protective environment.
	Monitor the use of printers, copiers, scanners, and fax machines. Where possible, review audit logs from these devices to discover and address any anomalies.	✓	
	Create a data transfer policy and procedure to allow sensitive company	✓	

	information to be removed from organizational systems only in a controlled way.		
	Establish a removable media policy and implement technologies to enforce it.	?	This might be problematic for SMEs where work is conducted in various settings. Completely restricting the ability to share information might hamper business operations and result in work inefficiencies.
	Restrict data transfer protocols, such as FTP, SFTP, or SCP, to employees with a justifiable business need, and carefully monitor their use.	?	Restricting changes to the system with administrative privileges might be beneficial to secure the system but restricting the sharing or transferring of data more widely can hinder productivity and introduce work inefficiencies.

20	Develop an enterprise-wide checklist to use when someone separates from the organization.	✓	
	Establish a process for tracking all accounts assigned to each employee.	?	Depending on the quality of existing controls for creating, managing and tracking accounts this might prove to be a time consuming task.
	Reaffirm all nondisclosure and IP agreements as part of the termination process.	✓	
	Notify all employees about any employee’s departure, where permissible and appropriate.	✓	
	Archive and block access to all accounts associated with a departed employee.	✓	
	Collect all of a departing employee’s company-	✓	

	owned equipment before the employee leaves the organization.		
21	Organizational justice (fairness; e.g., compensation aligned internally among employees and externally with industry standards)	✓	The time required to implement this recommendation as part of instilling a safety culture would not make this recommendation a quick win.
	Performance-based rewards and recognition (e.g., transparent criteria for promotions and discretionary rewards/recognition based on project performance)	✓	The time required to implement this recommendation as part of instilling a safety culture would not make this recommendation a quick win.
	Transparent and respectful communication (e.g., regular employee orientation, mentoring, and expectation setting)	✓	The time required to implement this recommendation as part of instilling a safety culture would not make this

			recommendation a quick win.
	Personal and professional supportiveness (e.g., employee assistance programs and professional development for furthering employee careers and sense of mastery)	✓	The time required to implement this recommendation as part of instilling a safety culture would not make this recommendation a quick win.

Appendix 4: Recommendations classified by the onion model

Recommendation Number	Chapter Number	a. All Organisations	Factor responsible
1	1	Conduct a physical asset inventory. Identify asset owners' assets and functions and identify the type of data on the system.	Work and organisational context
2	1	Understand what data your organization processes by speaking with data owners and users from across your organization.	Work and organisational context
3	1	Identify and document the software configurations of all assets.	Technologies
4	1	Prioritize assets and data to determine the high-value targets.	Work and organisational context
5	2	Ensure that legal counsel determines the legal framework the team will work in.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

6	2	Establish policies and procedures for addressing insider threats that include HR, Legal Counsel, Security, Management, and IA.	Work and organisational context
7	2	Consider establishing a contract with an outside consulting firm that is capable of providing incident response capabilities for all types of incidents, if the organization has not yet developed the expertise to conduct a legal, objective, and thorough inquiry.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
8	3	Ensure that senior management advocates, enforces, and complies with all organizational policies. Policies that do not have management buy-in will fail and not be enforced equally. Management must also comply with policies. If management does not do so, subordinates will see this as a sign that the policies do not matter or they are being held to a different standard than	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

		management. Your organization should consider exceptions to policies in this light as well.	
9	3	Ensure that management briefs all employees on all policies and procedures. Employees, contractors, and trusted business partners should sign acceptable-use policies and acceptable workplace behavior policies upon their hiring and once every year thereafter or when a significant change occurs. This is also an opportunity for your organization and employees, contractors, or trusted business partners to reaffirm any nondisclosure agreements.	Work and organisational context
10	3	Ensure that management makes policies for all departments within your organization easily accessible to all employees. Posting policies on your organization's internal website	Work and organisational context

		can facilitate widespread dissemination of documents and ensure that everyone has the latest copy.	
11	3	Ensure that management makes annual refresher training for all employees mandatory. Refresher training needs to cover all facets of your organization, not just information security. Training should encompass the following topics: human resources, legal counsel, physical security, and any others of interest. Training can include, but is not limited to, changes to policies, issues that have emerged over the past year, and information security trends.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
12	3	Ensure that management enforces policies consistently to prevent the appearance of favoritism and injustice. The Human Resources department should have policies and procedures in place that specify the consequences of	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

		particular policy violations. This will facilitate clear and concise enforcement of policies.	
13	4	Ensure that potential employees have undergone a thorough background investigation, which at a minimum should include a criminal background and credit check.	Work and organisational context
14	4	Encourage employees to report suspicious behavior to appropriate personnel for further investigation.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
15	4	Investigate and document all issues of suspicious or disruptive behavior.	Work and organisational context
16	4	Enforce policies and procedures consistently for all employees.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and

			Social influences, expectations and norms
17	4	Consider offering an EAP. These programs can help employees deal with many personal issues confidentially.	Personal physical and virtual workspace
18	5	Enhance monitoring of employees with an impending or ongoing personnel issue, in accordance with organizational policy and laws. Enable additional auditing and monitoring controls outlined in policies and procedures. Regularly review audit logs to detect activities outside of the employee's normal scope of work. Limit access to these log files to those with a need to know.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
19	5	All levels of management must regularly communicate organizational changes to all employees. This allows for a more transparent organization, and	Work and organisational context

		employees can better plan for their future.	
20	6	Have all employees, contractors, and trusted business partners sign nondisclosure agreements (NDAs) upon hiring and termination of employment or contracts.	Work and organisational context
21	6	Ensure that all employees, contractors, and trusted business partners sign workplace violence prevention and/or appropriate workplace behaviors documentation upon hiring.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
22	6	Ensure each trusted business partner has performed background investigations on all of its employees who will have access to your organization's systems or information. These should be commensurate with your organization's own background	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

		investigations and required as a contractual obligation.	
23	6	If your organization is acquiring companies during a merger or acquisition, perform background investigations on all employees to be acquired, at a level commensurate with your organization’s policies.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
24	6	Prevent sensitive documents from being printed if they are not required for business purposes. Insiders could take a printout of their own or someone else’s sensitive document from a printer, desk, office, or from garbage. Electronic documents can be easier to track.	Technologies
25	6	Avoid direct connections with the information systems of trusted business partners if possible. Provide partners with task-related data without providing access to	Technologies

		your organization’s internal network.	
26	6	Restrict access to the system backup process to only administrators responsible for backup and restoration.	Technologies
27	7	Establish a social media policy that defines acceptable uses of social media and information that should not be discussed online.	Work and organisational context
28	7	Include social media awareness training as part of the organization’s security awareness training program.	Work and organisational context
29	7	Encourage users to report suspicious emails or phone calls to the information security team, who can track these emails to identify any patterns and issue alerts to users.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

30	8	Establish a work culture that measures success based on appropriate metrics for the work environment. For instance, knowledge workers might measure their success based on outcomes and efficiency instead of metrics that are better suited for a production line.	Tasks
31	8	Encourage employees to think through projects, actions, and statements before committing to them.	People
32	8	Create an environment that encourages focusing upon one thing at a time, rather than multitasking.	Goals
33	8	Offer employees who are under stress options to de-stress, such as massages, time off, games, or other social but non-project oriented activities.	Work and organisational context

34	8	Routinely monitor employee workloads to make sure that they are commensurate with the employee's skills and available resources.	Goals
35	9	Develop and implement an enterprise-wide training program that discusses various topics related to insider threat. The training program must have the support of senior management to be effective. Management must be seen participating in the course and must not be exempt from it, which other employees could see as a lack of support and an unequal enforcement of policies.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
36	9	Train all new employees and contractors in security awareness, including insider threat, before giving them access to any computer system. Make sure to include training for employees who may not need to access computer systems daily, such as	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

		janitorial and maintenance staff. These users may require a special training program that covers security scenarios they may encounter, such as social engineering, active shooter, and sensitive documents left out in the open.	
37	9	Train employees continuously. However, training does not always need to be classroom instruction. Posters, newsletters, alert emails, and brown-bag lunch programs are all effective training methods. Your organization should consider implementing one or more of these programs to increase security awareness.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
38	9	Establish an anonymous or confidential mechanism for reporting security incidents. Encourage employees to report security issues and consider incentives to reporting by rewarding those who do.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

39	10	Establish account management policies and procedures for all accounts created on all information systems. These policies should address how accounts are created, reviewed, and terminated. In addition, the policy should address who authorizes the account and what data they can access.	Technologies
40	10	Perform audits of account creation and password changes by system administrators. The account management process should include creation of a trouble ticket by the help desk. (Help desk staff should not be able to create accounts.) Your organization could confirm the legitimacy of requests to reset passwords or create accounts by correlating such requests with help desk logs.	Technologies

41	10	Define password requirements and train users on creating strong passwords. Some systems may tolerate long passwords. Encourage users to use passphrases that include proper punctuation and capitalization, thereby increasing passphrase strength and making it more memorable to the user.	Technologies
42	10	Security training should include instruction to block visual access to others as users type their passcodes.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
43	10	Ensure all shared accounts are absolutely necessary and are addressed in a risk management decision.	Technologies
44	11	Conduct periodic account reviews to avoid privilege creep. Employees should have sufficient access rights to perform their	Technologies

		everyday duties. When an employee changes roles, the organization should review the employee’s account and rescind permissions that the employee no longer needs.	
45	12	Implement rules within the SIEM system, to automate alerts.	Technologies
46	12	Create log management policy and procedures. Ensure they address log retention (consult legal counsel for specific requirements), what logs to collect, and who manages the logging systems.	Technologies
47	13	Disable remote access to the organization’s systems when an employee or contractor separates from the organization. Be sure to disable access to VPN service, application servers, email, network infrastructure devices, and remote management software. Be sure to close all open sessions as well. In addition, collect all	Technologies

		company-owned equipment, including multifactor authentication tokens, such as RSA SecurID tokens or smart cards.	
48	13	Include mobile devices, with a listing of their features, as part of the enterprise risk assessment.	Work and organisational context
49	13	Prohibit or limit the use of personally owned devices.	Artefacts
50	13	Prohibit devices with cameras in sensitive areas.	Wider physical and virtual work environment
51	14	Use monitoring tools to monitor network and employee activity for a period of time to establish a baseline of normal behaviors and trends.	Technologies
52	14	Deny VPN access to foreign countries where a genuine business need does not exist. White list only countries where a genuine business need exists.	Technologies
53	14	Establish which ports and protocols are needed for normal	Technologies

		network activity, and configure devices to use only these services.	
54	14	Determine which firewall and IDS alerts are normal. Either correct what causes these alerts or document normal ranges and include them in the network baseline documentation.	Technologies
55	15	Carefully audit user access permissions when an employee changes roles within the organization to avoid privilege creep. In addition, routinely audit user access permissions at least annually. Remove permissions that are no longer needed.	Technologies
56	15	Establish account management policies and procedures. Audit account maintenance operations regularly. Account activity should reconcile with help desk documentation.	Technologies

57	15	Require privileged users to have both an administrative account with the minimum necessary privileges to perform their duties and a standard account that is used for every day, non-privileged activities.	Personal physical and virtual workspace
58	16	Conduct a risk assessment of the data and services that your organization plans to outsource to a cloud service provider before entering into any agreement. Your organization must en-sure that the service provider poses an acceptable level of risk and has implemented mitigating controls to reduce any residual risks. Your organization must carefully examine all aspects of the cloud service provider to ensure the service provider meets or exceeds your organization's own security practices.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

59	16	Verify the cloud service provider's hiring practices to ensure it conducts thorough background security investigations on any personnel (operations staff, technical staff, janitorial staff, etc.) before they are hired. In addition, the service provider should conduct periodic credit checks and reinvestigations to ensure that changes in an employee's life situation have not caused any additional unacceptable risks.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
60	16	Control or eliminate remote administrative access to hosts providing cloud or virtual services.	Technologies
61	16	Understand how the cloud service provider protects data and other organizational assets before entering into any agreement. Verify the party responsible for restricting logical and physical	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

		access to your organization’s cloud assets.	
62	17	Periodically review configuration baselines against actual production systems and determine if any discrepancies were approved. If the changes were not approved, verify a business need for the change.	Technologies
63	18	Store backup media off-site. Ensure media is protected from unauthorized access and can only be retrieved by a small number of individuals. Utilize a professional off-site storage facility; do not simply send backup media home with employees. Encrypt the backup media and manage the encryption keys to ensure backup and recovery are possible.	Technologies

64	18	Ensure that configurations of network infrastructure devices (e.g., routers, switches, and firewalls) are part of your organization’s backup and recovery plan as well as the configuration management plan.	Technologies
65	19	Establish a cloud computing policy. Organizations must be aware of cloud computing services and how employees may use them to exfiltrate data. Restrict and/or monitor what employees put into the cloud.	Personal physical and virtual workspace
66	19	Monitor the use of printers, copiers, scanners, and fax machines. Where possible, review audit logs from these devices to discover and address any anomalies.	Technologies
67	19	Create a data transfer policy and procedure to allow sensitive company information to be removed from organizational systems only in a controlled way.	Technologies

68	19	Establish a removable media policy and implement technologies to enforce it.	Technologies
69	19	Restrict data transfer protocols, such as FTP, SFTP, or SCP, to employees with a justifiable business need, and carefully monitor their use.	Technologies
70	20	Develop an enterprise-wide checklist to use when someone separates from the organization.	Work and organisational context
71	20	Establish a process for tracking all accounts assigned to each employee.	Work and organisational context
72	20	Reaffirm all nondisclosure and IP agreements as part of the termination process.	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
73	20	Notify all employees about any employee's departure, where permissible and appropriate.	Financial constraints and priorities; Technical developments and capabilities; Legal and

			regulatory framework and Social influences, expectations and norms
74	20	Archive and block access to all accounts associated with a departed employee.	Technologies
75	20	Collect all of a departing employee’s company-owned equipment before the employee leaves the organization.	Artefacts
76	21	Organizational justice (fairness; e.g., compensation aligned internally among employees and externally with industry standards)	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms
77	21	Performance-based rewards and recognition (e.g., transparent criteria for promotions and discretionary rewards/recognition based on project performance)	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

78	21	Transparent and respectful communication (e.g., regular employee orientation, mentoring, and expectation setting)	Work and organisational context
79	21	Personal and professional supportiveness (e.g., employee assistance programs and professional development for furthering employee careers and sense of mastery)	Financial constraints and priorities; Technical developments and capabilities; Legal and regulatory framework and Social influences, expectations and norms

Appendix 5: Brief description of CDM study design

Brief Description of the Critical Decision Method (CDM) Study Design
This study is planned to start in late March 2020 and last for a period of six weeks. We aim to recruit 20 participants between the ages 18-65 who have access to connected technologies (mobiles, laptops, tablets etc) for work or personal purposes. Participants must also have had experience with insider threats that can include cyberattacks such as phishing or ransomware. Initially, participants will be recruited through posters, emails and announcements within our existing networks at the University of Nottingham. If they are deemed suitable they will be recruited for a one hour semi-structured interview which will be video and audio recorded. Participants will only be requested to reflect on a particular incident that involved insider threat and talk through their thoughts and decision making processes in a comfortable and informal setting. Participants will not be paid for partaking in this research study. To ensure participant confidentiality and anonymity through the removal of any indicators of personal attributes of the participants. This includes names, ages, ethnicity, gender and online platforms including frequently used words by participants. Data, which would consist of transcripts and recordings, will be held for seven years following the publication of findings. Digital copies of the data will be stored on a secure university server accessible through a password protected laptop. Any hardcopies of the data will be kept in the researcher's locked filling cabinet. Data will be used to inform the findings of the study and relevant exerts will be published along with the findings, without the possibility of identifying individuals. Supervisors and industry partners will have access to this data. All data will be deleted and hard copies destroyed after a period of seven years.

Appendix 6: Participant information sheet for CDM study

Participant Information Sheet

This study will be conducted by Neeshé Khan, a doctoral student at the EPSRC Horizon Center for Doctoral Training, University of Nottingham. It is sponsored by UK Research and Innovation, the University of Nottingham and industry partner investments. Data generated as part of this study will belong to the University of Nottingham and used to explore factors that might influence unintentional Insider Threat within cybersecurity. ‘Unintentional Insider Threat’ can be defined as when individuals pose a risk to themselves accidentally or when they didn’t mean to and this can result an individual’s compromised level of cybersecurity.

This study aims to explore factors that influence our vulnerability to cyberattacks such as phishing, ransomware or unauthorised access (hacking) to gain access to information we hold in our personal and/or professional lives. This study is designed to last a period of eight weeks and this research will help us understand how to design protective systems that can provide better cybersecurity during times when users are vulnerable to such factors. Participants will be part of a one hour semi-structured interview in a comfortable one-to-one setting with the researcher. Participants will be required to think back to one specific time when they were vulnerable to a cyberattack and recall what they saw, thought and information they used to make decisions during this time. As this interview will be in an informal setting, there will only be a few questions to direct participant’s thoughts and understand their perspectives when discussing this specific incident.

As part of a comfortable and relaxed setting, participants are encouraged to go in to as much detail as they can when recalling their experience. However, participants are encouraged not to not feel obliged to share details that can make them uneasy. As participants are volunteering to share their experiences no compensation is available for time or other associated costs to participants.

Participants are entitled to confidentiality and anonymity as part of this study, All findings will be anonymised to remove any identifying features of the participants. This includes any specific devices that the individual owns or specific societies they mention (for instance correspondence from their personal banking society) and includes words that are frequently used by a specific participant. Interviews will be audio and video recorded. Copies of the transcripts and recordings will be held for seven years following the publication of findings after which point they will be permanently deleted and destroyed. Digital copies of the data will be stored on a secure university server accessible through a password protected laptop. Any hardcopies of the data will be kept in the researcher's locked filling cabinet and supervisors and industry partners will have access to this data. Quotes from the interviews will be published in a journal as part of the findings.

Finally, all participation in this study is voluntary and participants can withdraw their involvement at any time. While this might affect any important findings and influence the study's conclusions, all data obtained from the participant will be destroyed.

If you have any further queries, please contact Neeshé Khan on neeshe.khan@nottingham.ac.uk

Appendix 7: Consent Form for CDM study

Participant Consent Form

Title of the Study: Factors that influence unintentional insider threat in cybersecurity

Name of Researchers:

- Mrs. Neeshé Khan, Professor Sarah Sharples & Dr Robert Houghton, University of Nottingham, Department of Engineering

Please read this form carefully and fill it in after reading the participant information sheet provided. If you are happy to participate in this study, please place your initials in the boxes if you agree with the statements and sign this form.

	Initials
I have read the information sheet and understand the nature of the study.	
I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation will be video recorded. The recording will be shown only to the study investigator or collaborators, and will be stored on an external hard-drive and a locked file cabinet.	
I understand that I can withdraw from the study at any time without giving a reason without prejudice.	

I know that I can ask the researcher for further information about the study at any time.	
I understand that all information I give will be confidential and anonymised, and that it will not be possible to identify any of the respondents in the study report.	
I understand that quotations from the study might be used in the final report and in other publications.	
I understand that quotations used will be anonymous and I will not be identifiable in any report or publication.	
I agree to take part in the above study.	
Name/Date: Signature:	
Name of person taking consent/Date: Signature:	

Appendix 8: Full set of CDM interview questions

Decision (initial question)	a) Can you describe how you know if something is genuinely from the sender or not? b) How do you think this differs from someone with less experience with technologies?
Knowledge (probe)	Where do you think you acquired this knowledge to differentiate between content that is genuine from the sender and malicious content?
Experience (probe)	Thinking back to a specific time when you were cyberattacked, could you describe the incident from the time right before you received the malicious content/virus to the time after you had/were about to click the link?
Experience (probe)	Could you explain the sequence of events as they happened including how long each stage was?
Cues (probe)	What were you seeing, reading or hearing that suggested that this content was genuine?
Analogues (probe)	Were you reminded of any previous experience?
Goals (probe)	If things went according to plan, what were you trying to achieve during the time the incident happened?
Options & Basis (probe)	a) Did you consider any other actions to take prior to clicking the malicious content? b) [if applicable] How was this option selected and others rejected?

Aiding (probe)	What training, knowledge or experience could have helped to avoid clicking this malicious content?
Time pressure (probe)	On a scale of 1-5 (1=no pressure; 5=max pressure), how much time pressure was involved in making this decision?
Externals (probe)	Do you think other/personal goals impacted how you made decisions when interacting with what might seem to be malicious content?

Appendix 9: Total code frequency for themes

Theme Name	Code frequency	%
Decision Making	340	25%
Task Factors	238	18%
Accidents	238	18%
Organisational factors	530	40%
Total Theme Frequencies	1346	100%

Appendix 10: Code frequency for thematic subcategories

Parent Category	Subcategories	Frequency
Decision Making	Lived experience	188
Decision Making	Acquired knowledge	152
Organisational factors	Individual	111
Organisational factors	External Factors	96
Organisational factors	Pressures	91
Organisational factors	Employer	88
Task Factors	Speed (incident, discovery, response)	83
Task Factors	Actions	80
Organisational factors	Processes	79
Task Factors	Complexity of task	75
Accidents	Training	73
Accidents	Expertise level	61
Accidents	Trust in Tech	56
Accidents	Errors	48
Organisational factors	Peer Dynamics	31
Organisational factors	Physical environment	20
Organisational factors	Goals	14

Appendix 11: Participant Information Sheet for ToPB study

Participant Information Sheet & Consent Form

This assessment tool is powered by the world-leading Human Factors Research Group within the Engineering Department in collaboration with the Centre for Doctoral Training (CDT) at the Computer Science Department at the University of Nottingham. This multidisciplinary research project is supported by the Engineering and Physical Sciences Research Council (EPSRC) Horizon 2020 funding. This website is part of a research study to provide organisations usable sociotechnical and holistic tools to indicate their readiness levels against unintentional insider threat within cybersecurity. It has been approved by the university's Ethics Board. Data generated from users as part of this study belongs to the University of Nottingham and used to improve and evaluate the effectiveness of the assessment tool. While we request the participation of IT, senior manager(s)/HR, this is not mandatory. Personnel from these roles will solely be present to inform the choices selected by the 'main participant'.

Participants are entitled to confidentiality and anonymity as part of this study. The assessment tool does not collect any information that could be used to identify the user or their organisation (for instance, we have turned off cookies which collect user's IP). Users are invited to share their experiences of engaging with the tool. Findings will be anonymised to remove any identifying features of the participants. This includes changing any specific details that the individual mentions as part of their feedback or words that they frequently use. Session(s) will be audio and video recorded. Copies of the transcripts and recordings will be held for seven years following the publication of findings after which point they will be permanently deleted and destroyed. Digital copies of the data will be stored on a secure university server accessible through a password protected laptop. Any hardcopies of the data will be kept in the researcher's locked filing cabinet. People on this project as well as involved industry partner(s) will have access to this data. Quotes from the feedback sessions and outputs from the website will be published in a journal as part of the findings. Users engagement in this study through the use of the assessment tool is voluntary and participants can withdraw their involvement at any time, ask questions and report any problems by emailing psxnk1@nottingham.ac.uk. While this might affect any important findings and influence the study's conclusions, all data obtained from the participant will be destroyed. This assessment tool and associated outputs are completely free to participants.

By proceeding to engage with the website after registration, users confirm and consent to the following:

- User(s) are 18 years of age or older.
- User(s) have read the information provided above and understand the nature of this study.
- User(s) have had the opportunity to consider the information provided above and asked questions prior to proceeding any further.
- User(s) understand that they are invited to participate in the usage of this tool and a feedback session. This participation will be video and audio recorded. The recording will be shown only to the study investigator or collaborators, and will be stored on an external hard-drive and a locked file cabinet.
- User(s) understand that quotations from the feedback session might be used in the final report and in other publications.

- User(s) understand that quotations used will be anonymous and they will not be identifiable in any report or publication.
- User(s) understand that they can withdraw from the study at any time without giving a reason and without prejudice.
- User(s) can ask the researcher for further information about the study at any time via email provided above.
- User(s) understand that all information provided as part of this study will be confidential and anonymised, and that it will not be possible to identify any of the respondents in the study report.
- User(s) agree to take part in this study by proceeding with the assessment tool.

Please sign or print your name below, along with consent from all the 'supporting participants' who will be accompanying you, to confirm you have read and understood the terms above and your participation in the study. Please return this form prior to the date scheduled for the session.

Name:		Date:	
Name:		Date:	
Name:		Date:	

Appendix 12: ToPB study participant demographics

*B2B: Business-to-business; B2C: Business-to-consumer

Participant Ref.	Company Ref.	Role Equivalate	Size of Organisation	Nature of customers	Regional Headquarter Location
AL	1	Chief Executive Officer	Small or medium-sized enterprise (SME)	B2B	United Kingdom
GL	1	Head	Small or medium-sized enterprise (SME)	B2B	United Kingdom
BP	2	Chief Executive Officer	Small or medium-sized enterprise (SME)	B2B	United Kingdom
SN	2	Head	Small or medium-sized enterprise (SME)	B2B	United Kingdom
MM	3	Chief Executive Officer	Large enterprise	B2B	Americas
HR	3	Chief Operating Officer	Large enterprise	B2B	Americas

RR	4	Director (Global)	Large enterprise	B2B	Americas
KL	4	Head	Large enterprise	B2B	Americas
ST	5	Director	Non-profit organisation	B2C	United Kingdom
MK	5	Chief Executive Officer	Non-profit organisation	B2C	United Kingdom
DS	5	Manager	Non-profit organisation	B2C	United Kingdom
GH	6	Director	Non-profit organisation	B2C	United Kingdom
KK	6	Head	Non-profit organisation	B2C	United Kingdom

Appendix 13: Pre-Session Questionnaire for ToPB study

Q1. What are the top three priority cybersecurity areas for your organisation?

1.	2.	3.

Q2. What three words/phrases would you use to describe your organisation’s cybersecurity state right now?

1.	2.	3.

Please select respective boxes below to indicate your preference for the following statements:

Q3. Training and technical cyber defences (such as firewalls) are the main way to prevent Unintentional Insider Threats.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q4. Cybersecurity vulnerabilities don’t tend to change drastically over short periods of time.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q5. Unintentional Insider Threat arises as a direct consequence of users not following prescribed procedures.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q6. Cybersecurity is mainly concerned with computer-based interactions.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q7. Good cybersecurity practices and near-misses are regularly shared in the organisation.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q8. It is everyone's responsibility to be aware of cybersecurity challenges faced by the organisation.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q9. People in the organisation generally take action if they identify a cybersecurity vulnerability.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q10. It is expected of me to implement best practices to make the organisation more cyber resilient.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q11. Cybersecurity is always considered in the organisation when decisions are made about changes to procedures and resource allocation.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q12. Who in the organisation is responsible for cybersecurity?

Q13. I am able to implement new procedures to streamline processes or enhance existing practices.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q14. I am able to start new group activities that are in the interest of the company (such as 'Cake Fridays' to increase morale or Equality, Diversity and Inclusion groups).

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q15. I am able to take new findings from my experiences to the Board/Senior Management Team for review to inform future organisational strategies.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q16. The main way of avoiding Unintentional Insider Threats related cyberbreaches is through restricting what users can do with organisation's IT systems.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Appendix 14: Post-Session Questionnaire for ToPB study

Q1. In the future what will be the top three priority cybersecurity areas for your organisation?

1.	2.	3.

Q2. What three words would you use to describe your organisation's cybersecurity state right now?

1.	2.	3.

Please select one box for each answer below to indicate your preference for the following statements:

Q3. Training and technical cyber defences (such as firewalls) are the main way to prevent Unintentional Insider Threats.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q4. Cybersecurity vulnerabilities don't tend to change drastically over short periods of time.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q5. Unintentional Insider Threat arises as a direct consequence of users not following prescribed procedures.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q6. Cybersecurity is mainly concerned with computer-based interactions.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q7. Knowledge sharing at the organisation will increase in the future whereby users share cybersecurity practices and (near-miss) experiences more frequently.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q8. There will be wide interest within the organisation around insights from the personalised report.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q9. I will explore ideas to strengthen defences that are low readiness levels in the personalised report.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q10. Board members and senior staff will support my initiatives to focus on specific parts of the defences to make the organisation more cyber resilient.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q11. Workload, procedures and resources are going to be considered more closely when creating cybersecurity practices within the organisation.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q12. In the future who in the organisation should be responsible for cybersecurity?

Q13. It is important to evaluate existing procedures and practices to become more cyber resilient.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q14. It is likely that the organisation will form a group that meet periodically to exchange information and experiences about cybersecurity.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q15. The Board/Senior Management Team will be interested in the findings shown in the personalised report to inform future organisational strategies.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Q16. The main way of avoiding Unintentional Insider Threats related cyberbreaches is through restricting what users can do with organisation’s IT systems.

Strongly Disagree 1 □ 2□ 3□ 4□ 5□ 6□ 7□ Strongly Agree

Appendix 15: Open-ended interview questions for ToPB study

Q1. Going pillar by pillar, can you share any insight(s) you gained from your personalised report?

Q2. Going pillar by pillar, can you share an insight(s) you gained from the type of factors that influence UIT?

Q3. What challenges do you foresee in implementing changes to strengthen defences that are highlighted in the personalised report?

Q4. Did you enjoy today’s experience?

Q5. What did you find interesting?

Q6. Can you share something you liked?

Q7. Can you share something you did not like?

Q8. Were there things you were expecting to see that were not covered?

Q9. Were there things in the website that surprised you?

Appendix 16: Semantic scale responses

Table below shows the Average and range of responses to semantic scale based questions as part of research study guided by the ToPB. All questions utilising a seven point semantic scale from ‘strongly disagree’ to ‘strongly agree’ are presented below alongside their associated themes. Responses were averaged and the range was determined to capture overall changes in attitudes and behaviour prior and during session(s). Author’s notes are also shown below to provide context to the findings.

Theme	**Q. No**	**Pre-session Avg. Response**	**Range of rating**	**Post-session Avg. Response**	**Range of rating**	**Author Notes**
Attitude	Q. 3	3.9	5	3.6	6	This question shows an overall downward shift on the semantic scale towards 'Strongly Disagree'. 6 participants changed their initial ratings. An outlier is noted [GH:6] with a four point upward shift on the scale.
			(Lower: 2 Upper: 7)		(Lower: 1 Upper: 7)	
Attitude	Q. 4	2.9	5	3.2	5	This question shows an overall upward shift on the semantic scale towards 'Strongly Agree'. Four participants changed their initial ratings. An outlier is noted [KK:6] with a five point upward shift on the scale.
			(Lower: 1 Upper: 6)		(Lower: 1 Upper: 6)	
Attitude	Q. 5	4.6	4	4.2	4	

			(Lower: 2 Upper: 6)		(Lower: 2 Upper: 6)	This question shows an overall downward shift of towards 'Strongly Disagree'. Four participants changed their initial ratings.
Attitude	Q. 6	3.7	5 (Lower: 1 Upper: 6)	3.5	6 (Lower: 1 Upper: 7)	This question shows an overall downward shift towards 'Strongly Disagree'. Four participants changed their initial ratings. An outlier is noted [KL:4] with a four point downward shift on the scale.
Subjective norms	Q.7	5.0	5 (Lower: 2 Upper: 7)	5.7	4 (Lower: 3 Upper: 7)	This question shows an overall upward shift of towards 'Strongly Agree'. One participant went down by a point.
Subjective norms	Q.8	6.5	2 (Lower: 5 Upper: 7)	5.2	3 (Lower: 3 Upper: 6)	This questions shows a downward shift towards 'Strongly Disagree'. While a majority of the participants strongly agreed with this question initially, they did not think that there'd be wider interest in their organisations on this topic or in the findings. While everyone is believed to be responsible, people are not believed to be actively interested.
Subjective norms	Q.9	5.3	5 (Lower: 2 Upper: 7)	5.8	3 (Lower: 4 Upper: 7)	This question shows an upward shift towards 'Strongly Agree'. The subjective norms in place led participants to believe that others take action and they themselves are able to do so.

Subjective norms	Q.10	6.2	3 (Lower: 4 Upper: 7)	6.1	3 (Lower: 4 Upper: 7)	Overall participants believed that they were seen to be responsible for making their organisations cyber resilient. Eight participants 'Strongly Agreed' with getting support from senior stakeholders whereby only two positively changed their ratings after the session. Five participants didn't indicate that they would receive strong support from senior stakeholders with an overall downward shift towards 'Strongly Disagree'.
Subjective norms	Q.11	5.2	6 (Lower: 1 Upper: 7)	5.8	4 (Lower: 3 Upper: 7)	Responses showed an upward shift towards 'Strongly Agree'. Participants indicated subjective norms of being able to understand and consider WL, procedures and resources. Participants' subjective norms appeared to be supportive of developing this understanding when creating cybersecurity practices.
Control	Q.13	5.3	6 (Lower: 1 Upper: 7)	6.5	1 (Lower: 6 Upper: 7)	This question shows an upward shift of towards 'Strongly Agree'. Participants indicated control over the design of procedures and processes (in belief and in practice). Participants also indicated a positive correlation between procedures/practices and

						cybersecurity with all ratings towards 'Strongly Agree'.
Control	Q.14	6.0	3 (Lower: 4 Upper: 7)	5.3	6 (Lower: 1 Upper: 7)	This question shows a downward shift towards 'Strongly Disagree'. While participants showed control over being able to start new group activities, this control belief didn't permeate into forming a group in real-world setting.
Control	Q.15	6.5	2 (Lower: 5 Upper: 7)	5.6	4 (Lower: 3 Upper: 7)	This question shows a downward shift towards 'Strongly Disagree'. While participants believed that senior stakeholders would be interested in their experiences and information they provided, participants did not feel that they could control what the senior board would be interested in and unable to influence board’s strategies.
Control	Q.16	3.4	4 (Lower: 1 Upper: 5)	4	5 (Lower: 2 Upper: 7)	Responses indicated an upward shift towards 'Strongly Agree'. One outlier is noted [SN:2] with a 6 point upward shift. Participants expressed control through restriction of users' access in IT systems in order to avoid UIT.

Appendix 17: Code frequencies for ToPB study

Parent and Child Node Frequencies from QSR-NVivo for research study guided by the ToPB

Parent Node	Child Node	Frequency	Percentage (%)
Attitude		148	14.77
	Technology	20	
	People	128	
Organisational Subjective Norms		202	20.16
Capability		263	26.25
	Organisational Technological Capability	63	
	People and Skills	200	
Framing		46	4.59
Development of People and Skills		52	5.19
Aspirations		39	3.89
Framework Feedback		252	25.15
	Endorsements	129	
	Potentials	123	
Total theme frequencies		1002	100