Creating Awareness on Phishing Signals amongst Hospital Staff using Serious Gaming

by

Reshma Joseph

Student number: 5383595

to obtain the degree of Master of Science in

Complex System Engineering and Management,
Faculty of Technology, Policy and Management,

at the Delft University of Technology,
to be defended publicly on Thursday August 18, 2022.

Thesis Committee:	Prof. dr. ir. P. van Gelder,	TU Delft, Chairperson
	Dr. M. L. C. de Bruijne,	TU Delft, First Supervisor
	Dr. S.O. Delgado,	AUMC, External Supervisor
	D. Boschma,	TU Delft, Advisor

TU Delft Delft University of Technology

Preface

With the completion of this research, my journey at TU Delft comes to an end. At the beginning of my Masters in September 2020, I never fully thought about my time in CoSEM or my thesis. I only knew that I would pursue my thesis in a topic related to Information and Communications Technology (ICT) which was my track. The courses from the track and the corresponding academic projects that I got to work on narrowed my interests towards data privacy and regulations. Of them all, Cyber Risk Management intrigued me the most. As my interests aligned by the beginning of my second year, I decided to put them together in the healthcare domain, a domain that I have always been passionate about and enjoyed working in. I am also an ardent fan of card games, so I made the central focus of my thesis to be the design of a card game. As a person who likes to make friends and family aware of the dos and don'ts to safeguard personal data, I am happy to have been able to work on this thesis that helps hospital staff to become aware of phishing signals in their work environments through a simple card game.

In no phase of my thesis did I feel stuck and for this I am grateful to my supervisors. Thank you, Mark for guiding me through my thesis as I unraveled the complexity of the topic and for constantly reminding me of the breaks I deserved while going through with it. I believe your supervision enabled me to challenge myself in this project. Silvia, thank you so much for your constant support and approachable aura, it helped me improve my work and stay positive and determined throughout this project. Pieter, thank you for understanding my interest in this field and taking me under your wing. I also want to thank you for making the course, Cyber Risk Management interesting, which is why I am here today. Doris, thank you for always checking up on my progress and sharing your expertise to improve my game. The crash course sessions on the different games developed at the Game Lab had a significant effect on the design of my game.

Lastly, I would like to thank my family and friends for their unconditional support during the two years of my Master's. Mom, Dad, and Ranchie, I can't thank you guys enough for encouraging me, being my strength, and having my best interests. Gopal and Rishab, thank you for being my rock and believing in me. Ivo and Danny, I can't thank you guys enough for checking up on me and offering to help me when I had to sprint through deadlines. Thank you, Yves for being there and brightening time with your humor. As I look back, I am happy and proud of the personal growth that I've had in the last two years, and I am so grateful that I had you all in this journey. I wouldn't change a thing.

I hope you enjoy reading this report!

Reshma Joseph
August 2022

Executive Summary

Ransomware attacks based on social engineering have been increasing since COVID-19. Attackers have commonly used phishing as a social engineering technique to deploy a ransomware attack. Critical infrastructures such as hospitals have been the common target of these attacks due to hospitals' sudden increase in digitization and interconnectivity. Moreover, the richness of patient data housed by these organizations make them an attractive target. Although existing spam filters are efficient, the possibility of conceding to malicious website links and emails remains as sophisticated phishing methods can trick users easily. On one hand, there is no full-proof automatic system that can consistently counter ransomware attacks that propagate through innovative phishing campaigns. On the other, a survey conducted by McAfee in 2021 found that 70 percent of the ransomware attacks against organizations were attributed to the shortage of cybersecurity skills amongst employees. The basic knowledge amongst employees in recognizing suspicious signals is scarce even though there are numerous workshops, programs, and online websites to educate users of such threats. To tackle this problem, effective employee awareness methods are necessary to help hospital staff identify different phishing signals and report them. A two-layered scientific gap was identified pertaining to the lack of security awareness in in hospitals. Firstly, there is a lack of effective security education methods that focus on identifying phishing signals in hospital environments. Although serious games have been gaining popularity as effective user education methods, there are no existing cybersecurity games that focus on creating awareness by identifying phishing signals. Secondly, there is a lack of standard frameworks that can be used for designing cybersecurity awareness methods or awareness interventions. A recently developed combination of Protection Motivation Theory (PMT) and MINDSPACE frameworks have been suggested by Briggs (2017) as an effective framework for designing cybersecurity interventions. But the combination of the two frameworks have not yet been tested in the design of user education methods nor in any other field. Based on this research gap, the following research question was formulated.

How can hospital staff be made aware of phishing signals in the work environment to prevent ransomware attacks on hospitals?

The Game Design Research Approach was used to design a serious game based on the combination of the PMT and MINDSPACE frameworks to answer the main research question. The game design focuses on creating awareness on phishing signals in the players by using the elements of the PMT framework, that is, threat appraisal (making the players aware of the severity of and vulnerability to a threat) and coping appraisal (coping responses available to the player to deal with the threat). The influencers of the MINDSPACE framework are used in the design of the game to act as catalysts to improve threat appraisal and coping appraisal. A literature review was conducted to identify the challenges that prevent hospital staff from recognizing phishing cues or signals in their workplace to gather contextual knowledge for designing an effective game. Challenges arise from the hospital environment as well as from an individual's level. The former includes challenges such as high stress environment leading to workload and fatigue, lack of strict security policies around the new trend of Bring-Your-Own-Device (BYOD) in hospitals, and the lack of adequate and continuous cybersecurity awareness trainings for hospital staff. The latter arises due to the

individual's susceptibility to persuasion techniques used by attackers, high gullibility, and low cue utilization due to lack of awareness and workload fatigue.

The challenges arising from an individual's level are addressed while designing the serious game called Phish Phishy. Phish Phishy is a tabletop card game with four different sets of cards and is played in two rounds. The scenario cards used in the game are based on persuasion techniques such as authority, similarity, urgency, and commitment. The scenarios used in the cards contain real-world hospital context so that the players can easily relate to the game. The scenario cards include both legitimate scenarios as well as phishing scenarios. The phishing scenarios are used to appraise the players to threat signals or cues (threat appraisal). For example, a threat signal can be an unsecure log-in website. The Action cards include actions that a player must take for a given scenario card. The ideal action that the players must take for a phishing card is to report it as suspicious, that is, coping with the threat by reporting it (coping appraisal). Appreciation cards are given to a player if they correctly recognize and report suspicious cards. But Depreciation cards are given to a player if they trust and respond to a suspicious card. The Appreciation and Depreciation cards include messages that appraise the players of the impact of their actions on the hospital. These cards are aimed at intrinsically motivating the players to become aware of different phishing signals in their workplace. Two gameplay sessions were conducted in two large academic hospitals in The Netherlands. Based on the game design, the results from the game survey suggested an increase in the awareness levels, that is, improved understanding of phishing signals (threat appraisal) and improved response to threat by reporting them (coping appraisal). Therefore, the PMT and MINDSPACE framework combination suggested by Briggs (2017) was explored for the first time through the serious game, Phish Phishy, to make hospital staff aware of phishing signals in the work environment and report them.

This research has its limitations. For example, the gameplay was conducted only in two hospitals, that is, the sample size of the data collected is too small so it cannot validate the results to predict the behvaiour of the players in their work environment. The game only focuses on one type of social engineering threat (phishing signals) but there are many more threats that the hospital staff should be made aware of (e.g., phone scams or vishing). Moreover, the game did not consider the occupational stress that hospital staff undergo which influences their interaction with their work environment. Therefore, the results of this research should not be generalized to the hospital population without replicating and validating the gameplay with more sample groups. The recommendation for future research is to conduct more gameplay sessions with different samples to validate the effect of the game on the awareness levels of the players after the gameplay is over either through real-life observations or surveys. If the collected results can be generalized to the hospital population, then the combination of PMT and MINDSPACE frameworks can be considered as a standard framework like the NIST framework for designing awareness games on phishing signals in hospitals.

List of Figures

List of Tables

Abbreviations

AUMC - Amsterdam University Medical Center

BBC - British Broadcasting Corporation

BYOD - Bring-your-own-device

CIA - Central Intelligence Agency

CISO - Chief Information Security Officer

CS - Computer Security

DPO - Data Protection Officer

EHR - Electronic Health Record

IC - Informed Consent

ICT - Information Communication Technology

IT - Information Technology

MINDSPACE - Messenger Incentives Norms Defaults Salience Priming Affect Commitment Ego

NHS - National Health Service

NIST - National Institute of Standards and Technology

PMT - Protection Motivation Theory

RDP - Remote Desktop Protocol

SE - Social Engineering

SMS - Short Message Service

STRIDE - Spoofing Tampering Repudiation Information disclosure Denial of service Elevation of privilege

TPM - Technology Policy Management

TU - Technical University

UK - United Kingdom

URL - Uniform Resource Locator

US - United States

VPN - Virtual Private Network

TABLE OF CONTENTS

Chapter 1: Introduction

There has been a surge in cyberattacks since the advent of the COVID-19 pandemic (Beaman et al., 2021). An overnight paradigm shift to home-based workp may have led to weaker security controls making it easy for attackers to lure people into ransomware-based phishing messages (Beaman et al., 2021). Rapid launch of Remote Desktop Protocol (RDP) and Virtual Private Network (VPN) services became essential for employees' work-from-home needs across organizations (Lekshmi, 2022). Rapid work-from-home brought with it a myriad of devices to be connected to the Internet through open and unsecured RDP ports (Lekshmi, 2022).

Cyber extortion methods have existed since the 1980s and have evolved into sophisticated ransomware attacks (Beaman et al., 2021). *Ransomware attack* is referred to as "a type of malware attack designed to facilitate different nefarious activities such as preventing access to personal data unless a ransom is paid" (Khammas, 2020, p.4). A literature review conducted by Hijji & Alam (2021) on the growth trends of malicious software since COVID-19 cited ransomware as the fastest growing malicious software. The review also points out phishing – a social engineering-based cyberattack as the commonly used technique for deploying a ransomware attack. *Social Engineering* (SE) is used as an umbrella term for "a broad spectrum of computer exploitations that employ a variety of attack vectors and strategies to psychologically manipulate a user" (Heartfield & Loukas, 2016, p.6). The term *Phishing* refers to a "scalable act of deception whereby impersonation is used to obtain information from a target" (Lastdrager, 2014, p.32). Emails, websites, and text messages are the common phishing mediums employed by attackers to inject malware code into the victim's computer and network system to encrypt it and make the data inaccessible to the victim (Hijji & Alam, 2021). Thereafter, the attackers try to extort monetary payment from the victim in exchange for the key that decrypts the compromised information files (Rameem Zahra et al., 2022).

Google blocked 240 million COVID-19 related spam emails and 18 million phishing and malware emails during the rise of COVID-19 in March 2020 alone (Ahsan Pritom et al., 2020). Although existing spam filters are efficient, the possibility of conceding to malicious website links and emails still remains due to the increasing ransomware attacks through sophisticated phishing methods (Beaman et al., 2021). Currently, there exists no full-proof automatic system that can consistently counter ransomware attacks that propagate through phishing campaigns (Hijji & Alam, 2021). Critical infrastructures such as hospitals and healthcare organizations have been cited as main targets for socially engineered ransomware attacks (Hijji & Alam, 2021). The primary reason for their susceptibility to ransomware attacks is the richness of data housed by these organizations. The data includes patient names, date of birth, social security number, address, and credit card information in the hospital and insurance records (Nigrin, 2014). On the black market, an Electronic Health Record (EHR) is 100 times more valuable than a credit card information. (Akpan, 2016). Another reason is the scientific innovations due to digitalization in the health sector (Lekshmi, 2022). The 2017 WannaCry attack against the NHS in the UK affected about 80,000 hospitals, 7000 appointments including surgeries to be cancelled, and certain emergency departments to shutdown (Fierce Healthcare, 2021). But the impact of ransomware attacks is not limited to data theft and locked systems, rather it has the potential to cost the lives of patients. In 2020, The BBC reported the first ransomware attack against Düsseldorf University Hospital that cost the life of a patient when 30 hospital servers remained unreachable as a result, the doctors could not timely monitor the criticality of the patient (Tidy, 2020).

A survey conducted by McAfee in 2021 found that 70 percent of the ransomware attacks against organizations were at least partly attributed to a shortage of cybersecurity skills amongst the employees (McAfee, 2021). The basic knowledge amongst employees in recognizing suspicious signals and/or spam can save organizations from being infected. There are numerous workshops, programs, and online websites to educate users of such threats, but based on the statistics of ransomware attacks, more innovative and effective efforts are needed to prevent them (Beaman et al., 2021). In critical infrastructures such as hospitals, developing more effective user education method for employees' security awareness is essential. The weakest security link is the human element in organizations wherein the biggest security threats are either due to internal breaches or employee behaviour (Ivanov et al., 2021). Therefore, this MSc. Thesis aims to develop an effective user education method that focuses on increasing employees' security awareness and investigates its effectiveness on the hospital staff.

1.1 Scientific Relevance

Developing effective user education for building security awareness about phishing is a challenge because the research conducted in this field is very limited. This section introduces the research that has been done so far in this field to analyze the scientific relevance of the research.

1.1.1 Complexity of Awareness

The Oxford dictionary broadly defines awareness "as the knowledge or perception of a situation" (Oxford, 2022). However, this definition needs more clarity for the prevention of phishing threats. The National Institute of Standards and Technology (NIST) report on cybersecurity awareness and training as: "learning is a continuum; it starts with awareness, builds to training, and evolves into education". The goal of awareness is to ensure that individuals are aware of potential IT security concerns and know how to recognize and react to such concerns (Al-Darwish et al., 2019). But what is security awareness? The threat-mitigation distinction by Reveraert & Sauer (2022) implies that awareness of a threat should be separated from awareness of ways to mitigate the threat. While threat awareness refers to "being aware of the characteristics of the threat such as, what, why, who, when, how, etc.," mitigation awareness refers to "being aware of the mitigation measures to counter the threat" (Reveraert & Sauer, 2022). Besides the threat-mitigation distinction of security awareness, Siponen & Kajava (1998) introduced the descriptive-prescriptive distinction. According to the authors, "security experts want people to internalize and follow given guidelines (prescriptive) rather than people to be aware of them but for some reason or other fail to apply them in reality (descriptive)". Therefore, awareness can be interpreted from a descriptive perspective where the emphasis is on the degree of factual knowledge about the issue (cognitive awareness), and a prescriptive perspective where the emphasis is rather on the actor's attitude (attitudinal awareness) to follow guidelines set in an organisation (Reveraert & Sauer, 2022).

Many scholars have explored cognitive awareness with respect to knowledge to define the attributes of security awareness. Hansch & Benenson (2014) classify awareness into three different interpretations: *Awareness as a perception* considers an actor to be aware when the actor knows about the existing threat. *Awareness as protection* considers awareness to be present when one can identify the threat and subsequently knows the countermeasures to that threat. *Awareness as behaviour* goes one step further suggesting that awareness is assessed on whether one knows, in addition to the knowledge about the threat and the protection measures, how to use the protection measures. However, Siponen & Kajava (1998) argue

that knowledge alone is not sufficient since performing the counter measure depends on the actor's cognition of their environment (cognitive awareness). Similarly, Spruit (2010) classifies security awareness as understanding the importance of security, while also about being motivated to contribute to it. In addition to cognitive awareness, focusing on attitudinal awareness is equally important since the most effective awareness programs engage people by helping to change their attitude (Shahri et al., 2013). Therefore, to design security awareness programs, one must include the components of threat-mitigation, that is, have the knowledge to identify threats and mitigate them as well as descriptive-prescriptive perspectives, that is, have the motivation to follow security guidelines. Moreover, focusing on both cognitive awareness as well as attitudinal awareness can further improve effectiveness of security awareness programs.

1.1.2 Motivation, Attitude, and Awareness

A sustainable behaviour change or a long-lasting behaviour change occurs through an attitude change, and an attitude change occurs by increased motivation and self-efficacy (ability to perform an action) (Siponen & Kajava, 1998). For a behaviour change to occur, an individual must have the ability, motivation, and opportunity to facilitate the change (Fogg, 2009). While motivation tends to last for a shorter period (minutes to weeks), attitude tends to be less dynamic and lasts for a longer period (months to years) (Oinas-Kukkonen, 2013). In the context of security guidelines, people often seem to be externally motivated due to sanctions that can be imposed on an individual but this is less sustainable for a behaviour change to occur because extrinsic motivation reduces as time progresses (Siponen & Kajava, 1998). According to Tohidi and Jabbari (Tohidi & Jabbari, 2012), motivation can be cultivated extrinsically at the initial stage and transformed to intrinsic motivation if the learning process becomes deeper. But to transform the extrinsic motivation to intrinsic motivation, a high level of willpower and engagement is required otherwise it will not transform into intrinsic motivation (Tohidi & Jabbari, 2012).

Security campaigns are considered as a useful way for improving motivation and attitude towards security because the campaigns provide information on the repercussions of not following security guidelines (Peltonen, 1989). But it may also negatively impact motivation and attitude as in the case of political and advertising campaigns. For example, negative feelings, irritation, hate, and different forms of resistance can be the result of campaigns (Peltonen, 1989);(Siponen & Kajava, 1998). Another method like campaigns is to make information security an 'in' topic within the hospital as suggested by Hammer's theory (Perry et al., 1985). Hammer's theory states that depending on how a new concept is introduced in an organization, everybody becomes keen to use it (Perry et al., 1985). Both campaigns and 'in' topics can be used together within awareness programs to provide incentives for end-users and to refresh the importance of these factors in people's minds (Kajava & Siponen, 2002). However, the effectiveness of these methods is unknown, and it does not seem to correlate with the security awareness aspects as defined in section 1.1.1.

1.1.3 User Education Constructs for Creating Awareness

According to the behaviourism theory (Araiba, 2020), individuals should be provided with relevant learning experiences to change their behaviour. These learning experiences rely on single-event and event-event learning which focus on repeated learning (Araiba, 2020). On the other hand, the constructivist learning theory focuses on facilitating the discovery of knowledge by creating learning experiences and social exchange opportunities using collaborative learning activities (Alessi et al., 2001). Since ransomware uses

sophisticated phishing mediums for targeting users, repeated learning according to behaviourism theory may not allow for identifying phishing signals but rather a constructivists approach could be beneficial. Moreover, hackers use a variety of persuasion techniques such as authority, urgency, similarity, social proof, consistency or commitment, scarcity, and reciprocation to easily hook their victims and persuade them into revealing sensitive information (Zielinska et al., 2016). Among these techniques, authority, urgency, and similarity are cited to be the most commonly used strategies for deceiving users (Zielinska et al., 2016). Therefore, educating users to identify the persuasive techniques that mimic phishing attempts in their legitimate environment is crucial to prevent individuals from falling for such cues.

The National Institute of Standards and Technology (NIST) framework was developed by the US government and has a drastic impact on cybersecurity all over the world since it was accepted as the standard security framework for all organizations (Shen, 2014). This framework "adopts industry standards and best practices to provide a set of voluntary, risk-based measures that can be used by organisations to address their cybersecurity risks" (Shen, 2014). However, it does not address best practices for the end-user to stay safe online. Another common framework used in security awareness games to improve user's security attitude and cybersecurity self-efficacy is Bandura's self-efficacy design framework (Bandura, 1977). According to Bandura (1977), self-efficacy is a person's belief in the capacity to accomplish a certain goal (Chen et al., 2020). Security awareness games that have used self-efficacy in the game design have seen improvement in the players' self-efficacy in game-related security tasks (Chen et al., 2020). However, the effect of the Bandura's self-efficacy theory on security awareness, that is threat-mitigation and descriptive-prescriptive awareness is unknown. Recent research has found that the combination of Protection Motivation Theory (PMT) and the MINDSPACE frameworks to be useful in designing cybersecurity interventions focusing on behaviour change (Briggs, 2017). PMT was developed to better understand the concept of fear as a motivator for certain behaviours, and later extended to look more broadly at the psychological mechanisms underlying persuasive communication (Rogers et al., 1997). PMT proposes that individuals engage in two types of appraisal while making decisions (Rippetoe et al., 1987). First is the assessment of a threat (threat appraisal), a judgment that in turn takes in both the perceived severity of the threat and the individual's perceived vulnerability to the threat. Second is an assessment of the individual's ability to cope with the threat (coping appraisal) (Rippetoe et al., 1987). The MINDSPACE framework was developed to assist in policy making (Dolan et al., 2012) for the UK Government's Behavioural Influences Unit. It details nine behavioural influencers: messenger effects, incentives, norms, defaults, salience, priming, affect, commitment, and ego. The combination of this framework and theory can provide organizations with important insights on designing effective cybersecurity interventions focusing on long-term behaviour change strategies (Briggs, 2017). However, the research on the combination of the two frameworks is relatively new, so its practical application and effectiveness remains unavailable.

1.1.4 Serious Games as a Method for Creating Awareness

The most common method of delivering cybersecurity education and awareness trainings is either instructor-led or computer-based (Trickel et al., 2017). While these types of training provide a good theoretical start, it is not enough and practice is essential for mastering the complexity of cybersecurity concepts (Trickel et al., 2017). In recent years, serious games have been proposed as a new approach that can complement instruction-led or computer-based cybersecurity education and training. Serious games

provide learners with an enjoyable educational environment where the participants can learn theory and concepts in cybersecurity and put them into practice through the game (Hart et al., 2020). Popular serious games for cybersecurity awareness based on behaviour change theories such as self-efficacy theory and constructivists learning are Hacked Time and Riskio, respectively. However, while reviewing the two games for this research, it was found that neither of the two games focus on awareness specific to phishing nor do any of the serious games focus on the recent cybersecurity intervention theory using the combination of PMT and MINDSPACE frameworks (Chen, 2020);(Hart, 2020).

1.2 Societal Relevance

The number of phishing attacks are increasing, and organizations are incurring higher costs to deal with cybersecurity incidents. The impact of phishing attacks can range from no or limited impact to Distributed Denial of Services (DDoS), data stealing and manipulation, identity theft, or taking control of systems that can harm the physical world (de Bruijn, 2017). According to a range of polls and surveys, the public claim to be concerned about their privacy (Morar Consulting, 2016; Pike et al., 2017). However, the frequent behaviour exhibited by the public (people), places critical data at risk (Beresford et al., 2012; Felt et al., 2012). This disparity between claimed concern and empirical action is known as the Privacy Paradox (Norberg et al, 2007). The privacy paradox situation often arises due to a lack of security awareness among the people (Deuker, 2009). Likewise, hospital staff are a critical component of the hospital, and their unaware actions could introduce vulnerabilities that are subsequently exploited by hackers for cyberattacks. Unlike computers and software, employees cannot be "patched" when a new vulnerability is discovered but rather they must be made aware of new vulnerabilities (Hart et al., 2020). Therefore, it is fundamental for organizations to ensure that employees are made aware about the risks posed by even the simplest cyberattacks, and on how to make more secure decisions to avoid or mitigate security risks (Trickel et al., 2017).

1.3 Knowledge Gap

Two knowledge gaps have been identified from the lack of security awareness on identifying phishing signals in hospitals. Firstly, although education methods on security awareness exist such as security campaigns or implementation of security as an 'in' topic in organizations, they do not focus on motivating the hospital staff to become aware of security threats. Moreover, these methods do not focus on educating users about identifying persuasive phishing signals in their environment. Recognizing security education methods that focus on identifying persuasive phishing signals in hospital environments is essential. Serious games have been found to be a useful method to educate users about security awareness. Review of cybersecurity digital games by Coenraad et al (2020) found 181 digital games but their focus is more on gamifying the user-interface rather than on cybersecurity. Popular serious games that focus on behaviour change frameworks and theories such as self-efficacy and constructivist learning theory are Hacked Time and Riskio, respectively. But neither of these serious games focus on security awareness through threat-mitigation, perception of security-related risks, nor do they provide real-world scenarios to identify phishing and/or cyberthreat signals (Chen, 2020);(Hart, 2020).

Secondly, researchers believe that the combination of Protection Motivation Theory (PMT) and MINDSPACE frameworks can be useful to achieve this. However, there is no literature available that shows the practical implementation or effectiveness of using the combination frameworks for cybersecurity intervention.

Therefore, it is valuable to contribute towards developing a practical implementation of the combination frameworks in a serious game that has real-world scenarios to create an all-encompassing awareness focused on threat-mitigation and perception of security-related risks (threat appraisal and coping appraisal).

1.4 Research Question

To address the knowledge gaps that have been identified in section 1.3, the main research question has been formulated as follows:

How can hospital staff be made aware of phishing signals in the work environment to prevent ransomware attacks on hospitals?

The knowledge necessary to answer the main research question will be gained by answering the following sub-research questions.

1. What challenges prevent hospital staff from identifying phishing signals?
2. Which awareness intervention help hospital staff become aware of the various phishing signals?
    2.1. How can the identified awareness intervention be designed for a hospital environment to help hospital staff differentiate between legitimate and phishing signals?
3. What attributes of awareness are triggered by this awareness intervention?
    3.1. How effective is the identified awareness intervention in making hospital staff aware of phishing signals?

1.5 Linkage to the CoSEM Master's Program

A Complex Systems Engineering and Management (CoSEM) master thesis is focused on designing solutions for complex socio-technical problems. This research focuses on solving a complex socio-technical problem of rising ransomware attacks in a hospital environment (a complex socio-technical system) by making hospital staff aware of identifying phishing signals. To create awareness about phishing signals, an intervention is designed, developed, and tested to identify its impact on hospital staff. To be able to identify phishing signals, the awareness intervention should support cognition of hospital staff so that they can identify and cope with threats during their interaction with persuasive phishing messages. A solution for this problem is necessary due to the positive impact it could have on the digital security of hospitals. Since this research aims to intervene a complex socio-technical problem as mentioned above, the research fits with the requirements of the CoSEM Master's program.

Chapter 2: Research Methodology

This chapter presents the research approach that is used in this research to bridge the identified knowledge gap. The chapter begins with the research method selection strategy. It then progresses towards the research approach and its phases and delves into the research methods for every sub-research question. The chapter concludes with the research flow diagram to provide an overview of the research methodology.

2.1 Selecting the Research Strategy

The research strategy chosen to answer the main research question is the modelling approach using serious games for two reasons. First, a serious game modelling approach uses human role-players inside the model boundary to improve realism and knowledge transfer because serious games enable strong user involvement (Grogan & Meijer, 2017);(Kurapati et al., 2017). Second, this modelling approach allows for a safe experimentation space in realistic environments where a deliberate fiction gives players the freedom to fail, experiment, exert effort, and interpret outcomes while mitigating the cost or risk of real-world actions (Klopfer et al., 2009). It also allows to stimulate problem ownership through role adoption and allows for learning-by-doing approaches that support acquiring tacit and contextualised knowledge (Polanyi, 2009). The modelling approach using serious games is especially important for this research since answering the main research question requires the design of an awareness intervention in a hospital environment to make phishing signals seem realistic in the game environment.

2.2 Preparation for Designing the Serious Game

To design the serious game modelling approach, a design research approach is applied to create and evaluate the serious game intended to answer the main research question (Lukosch et al., 2018). Before designing the serious game, different types of knowledge are needed, and these are differentiated as follows for this research context (Lukosch et al., 2018):

2.2.1 **Contextual Knowledge**

This includes knowledge about the hospital environment that entails specific challenges that prevent the hospital staff from recognising phishing signals and the dilemma of which mitigation steps to take when faced with a threat.

2.2.2 **Scientific Knowledge**

This includes first, identifying the challenges that prevent hospital staff from recognising phishing signals. Second, using the Protection Motivation Theory (PMT) and MINDSPACE framework to educate the players (hospital staff) about phishing signals and understanding the impact of threat appraisal and coping appraisal to create awareness.

2.2.3 **Game Design Knowledge and Experiences**

This includes designing the game mechanics, elements, and interface for the players based on the research goal of creating an all-encompassing awareness about phishing signals.

2.2.4 **Knowledge about Simulations**

This includes developing game models and game rounds with high and low fidelity to test the game design.

2.3 Game Design Process

Duke & Geurts (2004) game design process is mainly aimed at analogue policy games. But the steps described for designing games are too detailed and is not relevant for this research. This is because the game design process suggested by Duke & Geurts (2004) is a sequential waterfall model and it was developed a long time ago when serious game design was not popular (Kurapati et al., 2017). Design is an iterative process that requires building, testing of mock-ups, and redesigning based on the feedback from tests, and is therefore not an isolated single step process (Klabbers, 2009). Since the research in this thesis will use the combination of behaviour change frameworks for cybersecurity intervention that have not yet been tested in any game design (or field), multiple iterations may be required to arrive at a functional game design. Therefore, the game design research process established by Kurapati et al., (2017) is selected which is based on iterative game design with emphasis on the roles of the players and maintains a balance of reality, game, and fun (*see figure 1*).

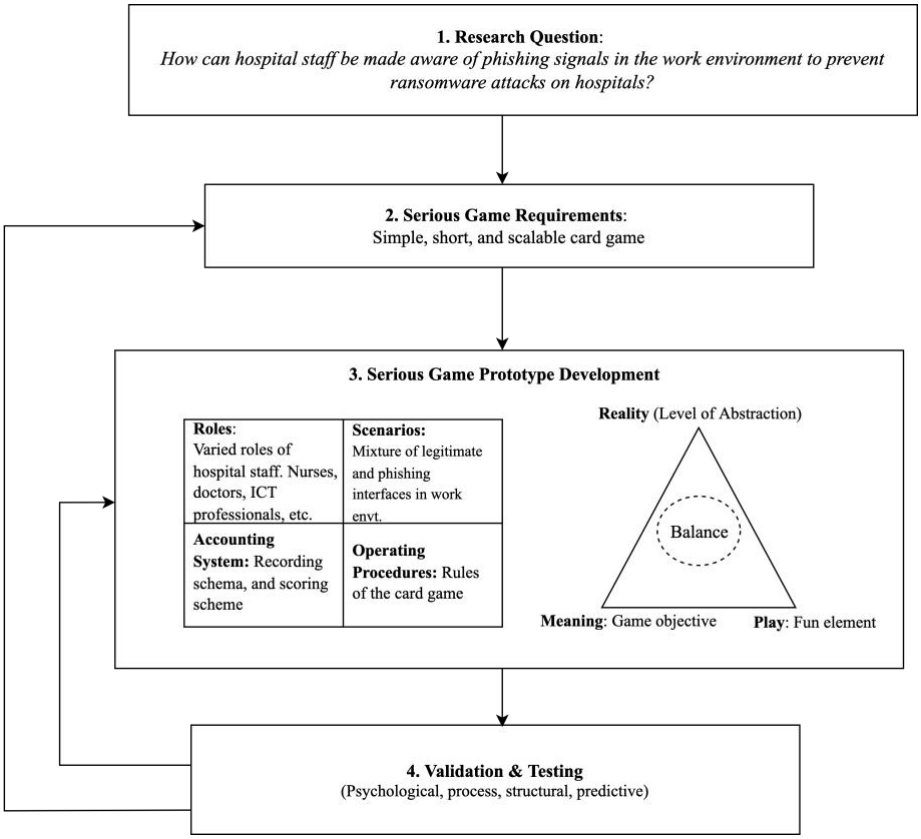


Figure 1-Game design research approach inspired by Kurapati (Kurapati et al., 2017)

Kurapati et al., (2017) iterative game design research process begins with defining the *research question*. To answer the research question, moving to the second step of defining the *serious game requirements* is important to understand the expected outcome of the game upfront. Since the target audience of this research are hospital staff who work in high stress environments with little free time, the serious game will be simple, short, and a scalable card game so that the players can fully immerse themselves and quickly learn through the game. After defining the problem and the game requirements, the third step of the process

focuses on the *game prototype development.* There are several aspects to be considered in this step. Since the hospital staff will be the players of the game, they will use their respective professional work *roles* itself to play the game. The *scenario* of the game will consist of a mixture of legitimate and phishing interfaces from their work environment. To account the gameplay, recording schema and scoring scheme will be given to the players to track their gameplay answers and performance. The gameplay will be established on a set of predefined rules of the game. The prototype development will aim to keep a balance of reality, meaning, and gameplay elements to make it equally fun. The final step of the process includes validating and testing the game from four perspectives. They are psychological, process, structural constructs, and predictiveness of expected outcome (Raser, 1969). Based on these results, the game prototype and requirements will be iteratively adjusted to incorporate feedback.

2.4 Research Approach and Methods

This research consists of three subsequent phases, aligned with the game design research process. Each phase contributes towards answering the sub-questions to together answer the main research question.

2.4.1 Research Phase I

Phase I of the research focuses on identifying the challenges that hospital staff face in their daily work environment that act as barriers against identifying phishing signals. This is an important aspect to understand so that the game can be designed in such a way that it specifically targets these barriers and teaches the players on how to overcome them. For this, a literature review using Scopus will be conducted to arrive at an answer. Therefore, the research conducted in this phase answers the first sub-question:

1. *What challenges prevent hospital staff from identifying phishing signals?*

2.4.2 Research Phase II

In this phase, the answer of the first sub-question functions as the input in this phase. The challenges acting as barriers for identifying phishing signals help in setting the requirements of the intervention. The context of the intervention will be so designed that it addresses these challenges to create an all-encompassing awareness in the hospital staff. Since the target audience are hospital staff with hectic work schedules, literature for different theories, models, and frameworks in Scopus will be explored to design a simple, short, but effective intervention. The information gathered will be used to narrow down towards an effective awareness intervention type. Therefore, the research conducted in this phase answers the second sub-research question:

2. *Which awareness intervention help hospital staff become aware of the various phishing signals?*

2.4.3 Research Phase III

In this phase, after the type of awareness intervention is identified, the aim of the awareness intervention is established (expected outcome). Followed by designing the environment, mechanics, and testing samples of the intervention. This is done based on literature research, expert advice, and creativity. Surveys are also designed for using before and after implementing the intervention so that the awareness levels and its associated attributes can be measured. This phase begins with rough sketches to connect the theories and

the creativity into a mind-map, followed by designing mock-ups using Adobe Photoshop. Therefore, the research conducted in this phase answers the sub-part of the second sub-research question:

2.1 How can the identified awareness intervention be designed for a hospital environment to help hospital staff differentiate between legitimate and phishing signals?

2.4.4 Research Phase IV

In this phase, the identified awareness intervention is applied in a simulated environment. Doing so will generate results and observations that will help in drawing conclusion about the effectiveness of this awareness intervention. The results and observations will be documented as physical copies of accounting sheets and as scanned soft copies. Therefore, the research conducted in this phase answers the third sub-research question:

3. What attributes of awareness are triggered by this awareness intervention?

2.4.5 Research Phase IV

To further validate the effectiveness of the intervention, validation is first conducted using game validation theory that is a part of the game design research approach and second, using expert interview. All the data gathered in this research will only be shared with the graduation committee as agreed in the ethics application form. Therefore, the research conducted in this phase answers the sub-part of the final sub-research question:

3.1 How effective is the identified awareness intervention in making hospital staff aware of phishing signals?

2.5 Research Flow Framework

A research flow framework is constructed to structure the research and visualize how this study will be conducted (*see figure 2*). First, a literature review is conducted on the challenges faced by hospital staff that prevent them from identifying phishing signals. This is explored in two levels: One, the challenges at the organizational level and second, at their individual levels (ability, knowledge, etc.,). This is addressed in Chapter 3. Chapter 4 identifies the awareness intervention based on literature review and helps in setting the requirements for the intervention design. It uses this information for designing the environment, mechanics, interfaces, and surveys for the intervention. Chapter 5 focuses on implementing the intervention to be able to gain insights in the contextual environment. Chapter 6 & 7 collate the findings from the implementation and validates them through expert interviews and surveys, respectively. The research concludes with Chapter 8 by answering the main research question and discussing the direction of future work.

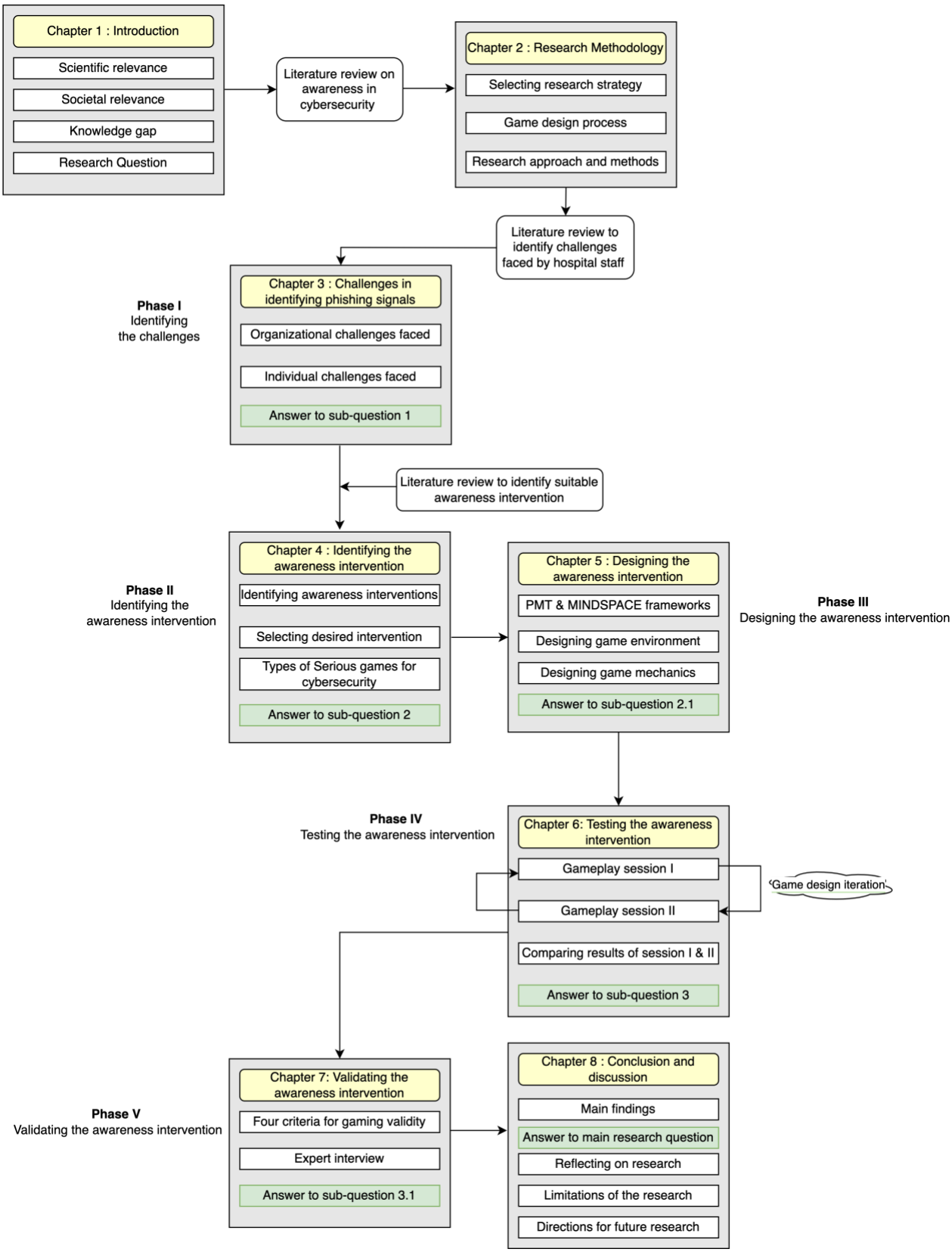


Figure 2- Research Flow Framework

Chapter 3: Challenges in Identifying Phishing Signals

This chapter dives into identifying the challenges that prevent hospital staff from recognizing phishing signals. This is done through literature review, and for a detailed literature review methodology, see Appendix A. The chapter concludes by summarizing the findings to answer the first sub-research question.

3.1 Challenges faced at an Organizational Level

The literature in this field is limited. However, from the literature available, the first category of challenges that emerged were categorized into challenges arising from an organizational level. These challenges are elaborated in this section.

3.1.1 High stress environment

According to Rizzoni et al. (2022), hospital staff fully intend to detect phishing signals and possible vulnerabilities in their work environment. However, high workload and phishing vulnerability seems to have a positive correlation (Rizzoni et al., 2022). Qualitative research also corroborates that increase in workload and stress causes fatigue so the staff are unable to prioritize cybersecurity protocols and their identification (Coventry et al., 2020). The research conducted by Nifakos et al. (2021) emphasizes patient safety as a priority of hospitals so the budget to improve IT systems security, user experience, and awareness trainings are given lesser importance than medical devices.

3.1.2 Increased interconnectivity in hospitals

Healthcare is one of the leading sectors driving Bring-your-own-device (BYOD) usage (Wani et al., 2020). The term BYOD refers to “the use of personal devices by employees for professional purposes”. These devices typically include smartphones, tablets, laptops, and Internet of Things (IoT) devices such as wearables (Wani et al., 2020). Although such a flexibility exists, healthcare authorities confirm that hospitals are unaware of the need to have strict security policies around BYOD. This can cause employees to become less vigilant while using their devices such as while visiting unsecure websites (Spok, 2018). The lack of strict policies lead to the lack of security controls around such devices since employees are not made aware of them (Wani et al., 2020).

COVID-19 decreased free mobility in hospitals for healthcare staff and patients and made them turn towards technology for consulting patients (e-consultations) and interacting within their teams (He et al., 2021). The sudden reliance on remote work with little or no experience have the made the healthcare sector more vulnerable to potential cyberattacks (He et al., 2021). Remote working policies were inadequate during such a technological change which has negatively affected employees’ security protocols with regard to remote working (He et al., 2021).

3.1.3 Lack of Awareness Trainings on Phishing

The common challenge identified in the literature points towards the lack of phishing awareness trainings, and three distinct reasons emerge for the lack of such a training. First, the employee turnover in hospitals is high and there is a constant influx of new employees who may not be well trained and are unaware of secure cyber practices (Gordon et al., 2019). Second, employees are not trained to identify fraud emails or messages so they could unknowingly violate information security policies (Georgiadou et al., 2021). Lastly, there is a

lack of skilled cybersecurity staff who can adequately train the employees to identify phishing signals in a hospital (Wani et al., 2020).

Table 1: An overview of the key challenges identified at a hospital level

Authors & Year	Theme	Challenges at Organizational Level
(Rizzoni et al., 2022) (Nifakos et al., 2021)	Stress environment	• Workload and fatigue leading to less security cautiousness. • High stress environment. • Prioritizing patient care over secure IT systems.
(Y. He et al., 2021) (Wani et al., 2020)	Increased interconnectivity in hospitals	• Increase in remote working option leaving users easier to attack. • Increase in Bring-your-own-device (BYOD) usage. • Lack of strict security policies around BYOD usage.
(Georgiadou et al., 2021) (Wani et al., 2020) (Gordon et al., 2019)	Lack of Awareness Trainings on Phishing	• Unable to identify and understand email fraud. • Lack of skilled staff in cybersecurity & budget. • Employee turnover disrupting cybersecurity training.

3.2 Challenges Faced at an Individual's Level

The second category of challenges that emerged from the literature review were categorized into challenges arising at an individual user's or employee's level. These challenges are elaborated in this section.

3.2.1 Susceptibility to 'Principles of Influence'

Phishing messages include several techniques to increase their believability. For example, they are crafted to resemble the communications of the impersonated organizations as closely as possible (Allodi, 2019). These techniques work by exploiting fundamental vulnerabilities in human cognition making it difficult for users to differentiate between spoofed messages from authentic messages (Mitnick, & Simon, 2003). Issues due to cognitive vulnerabilities can be distinguished into quick mode of cognitive processing and slower mode of cognitive processing (Kahneman, 2011). While the quick mode is an automatic reaction to detecting simple relationships in information to maintain a quick perception of our world, the slower mode is quite deliberate and is associated with subjective experiences that take time in detecting relationships in given information (Kahneman, 2011). In other words, users rely on immediate judgment and are subjected to cognitive biases (Chou et al., 2021). Since hackers exploit phishing emails using similarity to real-world

scenarios, the quick mode human cognition is unable to differentiate the message's legitimacy (Chou et al., 2021). These techniques used by hackers fall under Cialdini's 'principles of influence' and these principles suggest that the decisions taken by humans are influenced due to their momentary cognition. Hackers try to influence vulnerability in the human cognition based on Cialdini's principles of influence such as Reciprocity, Consistency, Social Proof, Authority, Liking, and Scarcity (Cialdini, 1984). The definitions and examples of these principles are elaborated in Table 2 for contextual understanding.

Table 2: An overview of the definitions and examples of Cialdini's influence principles

Principle[1]	**Definition**[1]	**Example**[1]
Reciprocity	Feeling obliged to repay favors from others. "I do something for you, you do something for me."	"While we work hard to keep our network secure, we are asking you to help us keep your account safe."
Consistency	Behaving in a way consistent with past decisions and behaviours. Acting in accordance with commitments made to certain view, company, or product.	"You agreed to the terms and conditions before using our service, so we ask you to stop all activities that violate them. Click here to unflag your account for suspension."
Social Proof	Referencing one's behaviour to that of others by using the majority behaviour to guide their own actions.	"We are introducing new security features to our services. All customers must get their verified again."
Authority	Obeying people in authoritative positions, following from the possibility of punishment for not complying with the authoritative requests.	"Best regards, Executive Vice President of <*company name*>."
Liking	Inclination to saying "yes" to the requests of people they know and like. People like those who are similar to them and who like them.	"We care for our customers and their online security. Confirm your identity …so we can continue protecting you."
Scarcity	Valuing items and opportunities when their availability is limited to not waste the opportunity.	"If your account information is not updated within 48 hours, your account will be restricted"

1. Based on Cialdini's (Cialdini, R. B., 1984) principles of influence.

3.2.2 High Gullibility

Gullibility is defined as "an individual's propensity to accept a false premise in the presence of untrustworthy cues" (Teunisse et al., 2020, p.4). In a phishing simulation research conducted by George et al., (2020), participants who attributed to a weaker sense of self and high emotionality fell for the phishing simulation emails due to lesser cognitive utilization. The *need for cognition* refers to an individual's tendency to enjoy and regularly take part in effortful cognitive activity (Cacioppo & Petty, 1982). They argue that individuals with need for high cognition will actively seek out and carefully examine information, while an individual with a lower need for cognition will tend to use other strategies such as similarity of environment or habit to make sense of their environment (Lins de Holanda Coelho et al., 2020);(George et al., 2020). This can also be corroborated through the empirical research conducted by Ndibwile et al., (2019) using smart-eye glasses. Their research used smart eyeglasses (electro-oculographic) to measure the mental effort and vigilance of their research participants while browsing websites. They found that knowledgeable participants who were gullible had insecure behaviours. For example, knowledgeable participants also opened email attachments from unfamiliar senders due to their high gullibility. However, research participants who were alert while browsing websites were able to effectively identify phishing websites. Similarly, research conducted by (Alzahrani, 2020) also found that hackers utilized users' COVID-19 anxiety as a way to gull users into clicking on phishing links. Finally, the research paper by Goethals, (2019) attempts to identify gullibility from the direction of motivation. Goethals' paper suggests that an unintentional act such as human error occurs due to an individual's lack of motivation that is in turn influenced by an individual's experience from mental capacity, knowledge level, degree of training, or commitment to a task.

3.2.3 Low Cue Utilization

Cue utilization refers to the user's perception of viewing and processing complex information (Burnkrant, 1978). Phishing messages often resort to persuasion techniques such as authority or urgency to encourage limited cognitive processing so that cue utilization in users is limited (Bayl-Smith et al., 2020). By doing so, hackers are able to inspire recipients of phishing messages to respond quickly without deliberation (Bayl-Smith et al., 2020). Users with lower cue utilization, that is, users who do not view and process the entire information have lower accuracy in differentiating phishing messages from non-phishing messages (Nasser et al., 2020). They also suggest that users who undertake concurrent tasks along with existing tasks have a decreased likelihood in detecting phishing emails because they have limited cognition to dedicate equal cue utilization for the two tasks (Nasser et al., 2020). Similarly, research conducted by (Ramkumar et al., 2020), showed that users fixate longer while reading when "the processing load is greater". For example, participants in their research spent more time on processing URLs as the length of the URL increased but only up to a point, after which the participants relied on the authority component (initial address) of URLs to trust the source of the URL.

Table 3: An overview of the key challenges identified at an individual's level.

Authors & Year	Theme	User behaviour towards phishing signal
(Chou et al., 2021) (Allodi, 2019)	Susceptibility to 'principles of influence'	• Users rely on immediate judgment and are subject to cognitive biases. • Both attractive and coercive influence can result in phishing susceptibility. • Hackers exploit phishing emails using similarity to real-world scenarios so the 'quick mode human cognition' is used and it is unable to differentiate its legitimacy. • Users fall for persuasion techniques of Authority, Scarcity, and Liking.
(George et al., 2020) (Alzahrani, 2020) (Ndibwile et al., 2019) (Goethals, 2019)	High gullibility	• Emotional users tend to be more persuadable. • Users with a weaker sense of self are more gullible. • User's anxiety is used for gulling into phishing messages. • Knowledgeable participants would also open email attachments from unfamiliar senders due to their gullibility. • User motivation influences gullibility. • Unintentional act such as human error occurs due to an individual's motivation which is in turn influenced by an individual's experience from mental capacity, knowledge level, degree of training, or commitment to a task.
(Bayl-Smith et al., 2020) (Nasser et al., 2020) (Ramkumar et al., 2020)	Low cue utilization	• Phishing messages often resort to persuasion techniques, such as authority or urgency to encourage limited cognitive processing so that cue utilization in users is limited. • Users with lower cue utilization, that is, users who do not view and process the entire information have lower accuracy in differentiating phishing messages from non-phishing messages. • Users have a cap on the number of cognitive resources they are willing to expend in recognizing a URL.

3.3 Conclusion of Chapter 3

The aim of this chapter was to identify the key challenges faced by the hospital staff that prevents them from identifying phishing signals. A literature review was conducted to identify these challenges. The challenges are summarized to answer to the first sub-research question:

1. *What challenges prevent hospital staff from identifying phishing signals?*

The challenges identified through literature review are divided into two categories: challenges at an organizational level and challenges at an individual's level.

At the organizational level, the challenges faced by the hospital staff are due to the high stress environment, increased interconnectivity in hospitals, and lack of awareness trainings on phishing. High stress environment in the hospital is the key reason for workload and fatigue leading to less security cautiousness from the staff. They also prioritize patient care as a result lesser attention is focused towards securing IT systems. The increase in the interconnectivity in hospitals is due to remote working and Bring-your-own-device (BYOD) working flexibilities. However, the security policies around working flexibility is still lacking from the hospital authorities so, the staff are also aware of their safe security behaviour that they must follow. The lack of awareness at this level is attributed to three reasons. First, there is a lack of skilled cybersecurity staff who can adequately train the staff. Second, high employee turnover in hospitals disrupts the continuous awareness training for the staff. Last, employees are not trained to identify fraud emails or messages so they could unknowingly violate information security policies.

At the individual's level, the challenges faced by the hospital staff are due to their susceptibility to 'principles of influence', high gullibility, and low cue utilization. The lack of awareness at this level is attributed to three reasons. First, individuals are susceptible to the 'principles of influence' used by hackers in phishing messages. This is because hackers exploit phishing emails using similarity to real-world scenarios so the 'quick mode human cognition' is unable to differentiate the message's legitimacy. Hackers try to influence vulnerability in the human cognition based on Cialdini's principles of influence such as Reciprocity, Consistency, Social Proof, Authority, Liking, and Scarcity. Second, individuals with a weaker sense of self and who are anxious are more easily gulled into clicking on phishing messages. An individual's motivation can also influence gullibility. Last, individuals are unable to view and process the entire information due to low cue utilization. As a result, they cannot differentiate phishing messages from non-phishing messages. Phishing messages often resort to persuasion techniques, such as authority or urgency to encourage limited cognitive processing so that cue utilization in an individual is limited.

Recognizing these challenges is essential to identify an effective intervention that can help hospital staff become aware about different phishing signals. The next chapters build on this information to identify and design a suitable awareness intervention.

Chapter4: Identifying the Awareness Intervention

This chapter dives into identifying an effective awareness intervention that can address the challenges that were identified in chapter 3. To arrive at the awareness intervention, more literature is first reviewed. After the intervention is identified, it is important to identify the shortcomings of the existing awareness interventions so that the need for a new awareness intervention can be justified. The chapter concludes by summarizing the findings to answer the second sub-research question.

4.1 Identifying Awareness Interventions

To identify the awareness interventions, this research began reviewing literature but a recent literature review conducted by Chowdhury & Gkioulos (2021) extensively covers different types of cybersecurity awareness methods used in critical infrastructures and no new type of awareness methods have emerged since. Therefore, instead of repeating the work, the literature review by the two authors is instead used.

According to the review, awareness interventions are employed in organizations via different forms of training programs. The review classifies awareness interventions into five categories based on their delivery method. The delivery methods try to create awareness by training the participants in various topics of cybersecurity. Some of the examples of the topics are general cyber threats, methods, attack techniques, and potential risks from compromising confidentiality, integrity, and availability (CIA).

According to the literature review by Chowdhury & Gkioulos (2021), the first category of delivery methods is the 'conventional awareness training method'. It includes conventional teaching and training methods that consist of onsite courses such as paper-based teachings, presentations, conferences, or exercises. They are still popular due to the familiarity and ease of content development for the participants or learners. However, this method of training is found to be tedious, time consuming, costly, and does not provide hands-on experience. The second category is the 'online and software-based awareness training method'. This includes online courses, web-accessible training material and software, and E-mail tests. They can be accessed remotely on different devices (laptop, phone, etc.,), are cost-effective, and hands-on exercises can be tailored-made. However, users may undermine the value or pay less attention during the training, it is not always scalable or adaptable, it is expensive, and does not provide instructor assistance. The third category is the 'game-based awareness training method' which includes serious games for awareness trainings. Serious games provide hands-on exercises, engages users, improves skill development in teams, is adaptable and scalable, and can be used remotely. However, older audiences may not be familiar with the game mechanics, it can be time consuming to design and develop, and it also has a high initial development cost and resource overhead. The fourth category is the 'video-based awareness trainings method' which includes educational videos. These are accessible, usable, cost effective, and time efficient. However, the video-based method is limited in content. It lacks interaction between other trainees and/or instructors and does not provide hands-on experience for the trainees. It does not guarantee active participation from the participant and requires constant updates for scalability. The final category is 'simulation and virtualization-based awareness training method' which includes simulation platforms and simulated laboratory exercises for the participants. The simulation and virtualization method provides hands-on experience on cybersecurity topics, helps replicate real-life incidents, improves team skill development. It is adaptable, can be accessed remotely, and has a high scalability. However, it is difficult to coordinate the participants and

the contents of the simulation because it requires the participants to have pre-existing knowledge. In addition, it requires a high initial development cost, resource overhead, and is time consuming to prepare and execute. Table 4 complies the different delivery methods of awareness interventions with examples for each method and their corresponding advantages and disadvantages.

Table 4: A summary of different awareness intervention methods used in organizations to raise employees' cybersecurity awareness levels based on Chowdhury & Gkioulos (2021).

Delivery Method	**Examples**	**Advantages**	**Disadvantages**
Conventional Methods	• On-site trainings • Presentations & Conferences	• Usability • Familiarity of format • Ease of communication between instructor and participants	• No guarantee of personnel active participation • Does not provide hands-on experience • High cost and resource overhead • Time-consuming
Online and Software-based	• Online courses • Web-accessible training material and software • E-mail tests	• Remote and multi-modal accessibility • Cost-effective • Hands-on exercises	• No guarantee of active participation of personnel • Not always scalable and adaptable • If personalized solution is required, it can lead to overhead cost • Does not provide instructor assistance
Video-based	Educational videos	• Accessibility • Usability • Cost-efficient • Time efficient	• No guarantee of personnel active participation • Lack of interactivity with other trainees or instructors • Limited content • Lack of hands-on experience • Requires constant updates

Game-based	Serious Games	• Team skills development • Engages users • Hands-on exercises • Demonstrated effectiveness • Adaptability • (Possible) Remote Usability • (Possible) High scalability	• Not all audiences may accept gaming • Time-consuming • High initial development cost and resource overhead
Simulation and virtualization-based	• Simulation platforms • Simulated laboratory exercises	• Hands-on experience • Replication of real-life incidents • Adaptability • (Possible) Remote Usability • (Possible) High scalability	• Hard to coordinate • Requires pre-existing knowledge • Time-consuming • High initial development cost and resource overhead

4.2 Selecting the Desired Awareness Intervention

The commonly used awareness interventions cited in the literature review are conventional methods and game-based methods (Chowdhury & Gkioulos, 2021). While conventional methods are well presented in literature due to their familiarity and ease of use, game-based methods are well presented due to their effectiveness. But game-based method is a recent development wherein specific serious games are yet to be developed that focus on individual cybersecurity topics, for example, serious games focused on phishing (Chowdhury & Gkioulos, 2021). Simulation-based methods have also been recommended for creating cybersecurity awareness along with game-based methods (Abawajy, 2014). The advantages of serious games and simulation-based methods are that they allow participants to conduct interactive, hands-on activities that develop team skills such as communication and organizational skills. As a result, these methods have demonstrated to be more engaging than conventional training methods. User engagement and motivation are two of the most significant factors for creating a successful cybersecurity awareness intervention (W. He & Zhang, 2019);(Bada et al., 2019). Experts in the cybersecurity domain also recommend the usage of simulation and game-based methods as effective cybersecurity awareness creation methods (Chowdhury et al., 2022). Using non-engaging and tedious awareness creating methods lesser chances of success in changing employees' security behaviour or attitude (W. He & Zhang, 2019).

A distinction is required between simulation and game-based methods to select an effective and usable training method. On one hand, game-based training is very engaging for the participants and has proven to stimulate self-efficacy, self-assessment, and collaboration in players or learners (Chowdhury et al., 2022);(Malone & Lepper, 2021). They can be developed for practical applications and exercises and have the ability to develop team skills (Amorim et al., 2013);(Jin et al., 2018). The development costs of game-based

trainings are also substantially smaller than simulation-based trainings (Chowdhury et al., 2022). On the other, simulation-based training is the only training method that allows participants to conduct exercises that are equivalent to possible real-life scenarios. However, simulation-based trainings are more expensive, difficult to coordinate, requires pre-existing knowledge, and is time consuming to develop and play. Since the target audience of this research are hospital staff who work under high stress environments with little time to spare for elaborate trainings, the game-based training method is selected.

4.3 Existing Serious Games for Creating Cybersecurity Awareness

Before designing a serious game as an awareness intervention, it is important to know the existing serious games in the field so that useful aspects can be reused into the serious game to be designed. Therefore, this section provides and overview of the most established security-related serious games from literature by classifying them according to their main goal and concludes by summarizing their main limitations. From this section forward, the serious game awareness intervention to be designed in this research will be called 'Phish Phishy'.

Elevation of Privilege (EoP) is a threat modelling card game proposed by Microsoft as part of the design phase of software projects (Shostack, 2014). EoP is based on the Microsoft STRIDE methodology and the aim of the game is to identify attacks using the different STRIDE threat categories (that is, Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege) (Potter, 2009). Each card represents potential real-world threat scenarios that can target a software (Hart et al., 2020), and this can be a useful input for designing 'Phish Phishy'. However, EoP is designed for software developers during the software development phase. As a result, it is considered too technical for the target audience of this research who are hospital staff.

Beckers & Pape (2016) propose a serious game to capture the underlying human behaviour that is exploited by social engineering. In this game, the players are organized into teams to learn attack and defense strategies related to human behaviours and elicit security requirements as mentioned in the game cards. Similarly, Haggman (2019) proposes a tabletop wargame based on the UK National Cyber Security Strategy for cyber security education. The board game simulates a cyberwar between UK and Russia, the two entities of the game. Each entity has a set of strategic objectives that they need to achieve using limited resources to conduct an attack or defend against it. This wargame aims to expose its players to a variety of attacks and defense strategies. However, creating awareness about attack and defense strategies is technical, beyond the scope, and consumes gameplay time for the target audience. Therefore, the 'Phish Phishy' game will focus on identifying phishing signals in phishing media that are used by attackers.

Cyber Security-Requirements Awareness Game is a tabletop card game focused on educating cyber security risks in hospital environments (Yasin et al., 2019). The players must identify vulnerabilities in the scenarios and exploit them to carry out insider or outsider attacks. The discussion among players during the game is used to evaluate and score attack scenarios. 'Phish Phishy' will also be designed around identifying phishing vulnerabilities in different scenarios based in the hospital environment. Similarly, *Decision & Disruption* (Frey et al., 2019) is a tabletop card game that

represents an industrial cyber physical system to protect. During each game round, the players must prioritize on a defense measure based on an available budget and protect the system under attack. At the end of each round, the game master rates the effectiveness of the selected defenses. The players learn the role of defenders and about security management strategies. However, this game does not focus on cyber threat awareness. To bridge this gap, the 'Phish Phishy' game will focus on identifying phishing signals in phishing media that are used by attackers.

Popular games that are based on behaviour change theories are Hacked Time and Riskio. *Hacked Time* is a computer-based game that incorporates Bandura's self-efficacy theory (*a person's belief in their capacity to accomplish a certain goal*) (Bandura, 1978) in the game design to improve player's self-efficacy towards cybersecurity tools (Chen et al., 2020). In the game, the player is a time-traveling detective who helps a college student deal with a security breach and the detective learns by making sense of the clues. The clues reveal the possible reasons for the breach and thereby improve the self-efficacy of the players by identifying the reasons that lead to security breach. For example, password written on a post-it (Chen et al., 2020). Although the self-efficacy measures improved in the players, the game did not increase their security-related risk perception, that is, capability to identify threat signals. 'Phish Phishy' will also focus on improving the self-efficacy of the players so that they can identify threat signals. *Riskio* is based on the Constructivists Learning theory and Microsoft's STRIDE threat model (Hart et al., 2020). It is a tabletop board game, and the game mechanics is structured into attack and defense phases. While one player acts as an attacker, the other player acts as a defender. The defending player is expected to choose the correct defense card for the corresponding attack card (Hart et al., 2020). Although the players learn about different attack and defense strategies, the game does not provide real-world scenarios to identify threat signals. It also focuses on technical aspects of cybersecurity, as a result, it is not suitable for the target audience. Like Riskio, 'Phish Phishy' will also focus on constructivist learning theory that is based on the belief that learning occurs when learners are actively involved in a process of meaning and knowledge construction as opposed to passively receiving information (Rolloff, 2010).

To summarize, these serious games are either tabletop board games, card games, or computer games. They allow players to learn how to think like attackers and defenders to educate the learners. Some of the games like Hacked Time and Riskio focus on behaviour change theories to target human behaviour change so that learners become aware of cybersecurity attacks and defense mechanisms. Despite the advantages of these games, they have their limitations. First, these games focus on core technical concepts of cybersecurity that are useful for software developers and IT staff but not for hospital staff. Second, very few games focus on behaviour change theories, but these games also do not create awareness about different persuasion techniques used by attackers that are useful for creating a threat perception and they do not provide real-world scenarios to identify phishing and/or cyberthreat signals. Third, these games have a generic context but to reduce the load on cognitive processing for the hospital staff, the game context should ideally be tailored to the hospital staffs' work context (emails, SMS, website logins, etc.,). Last, these games are time consuming which limits their usefulness among hospital staff because hospital staff have high workload. As a result, there is a need for a new serious game that is quick and helps hospital staff identify phishing signals.

4.4 Conclusion of Chapter 4

The aim of this chapter was to identify a suitable awareness intervention that can help hospital staff identify phishing signals. The awareness intervention was identified through an extensive literature review conducted by Chowdhury & Gkioulos (2021). The chosen awareness intervention is summarized to answer the second sub-research question:

2. *Which awareness intervention help hospital staff become aware of the various phishing signals?*

Serious games is selected as the awareness intervention that can help hospital staff become aware of different phishing signals. Although there are several awareness interventions such as conventional methods (presentations and seminars), software-based (online courses), video-based trainings (educational videos), simulation and virtualization-based (laboratory simulation exercises), serious games are chosen because it is more engaging for the users, has proven to stimulate self-efficacy, self-assessment, and collaboration among the players (Chowdhury et al., 2022). These attributes are essential to help hospital staff identify phishing signals, therefore serious games is selected as the suitable awareness intervention (Malone & Lepper, 2021).

Chapter 5: Designing the Awareness Intervention

This chapter dives into the use of Protection Motivation Theory (PMT) and MINDSPACE framework to support the foundation of the game design. Thereafter, the goal of the game is defined, the game environment, game mechanics, and gameplay are elicited to test the gameplay with the target audience. The chapter concludes by answering the sub-part of the second sub-research question.

5.1 Protection Motivation Theory and MINDSPACE Framework

From section 4.3, we find that games that use behaviour change theories (constructivist theory and self-efficacy theory) into game frameworks were successful in making the players aware about the specific cybersecurity topic that the game was intended for. Similarly, to help hospital staff overcome the challenge of identifying different phishing signals or cues used by attackers in their work contexts (email, SMS, or login websites), a combination of Protection Motivation Theory (PMT)and MINDSPACE frameworks is considered in the design of Phish Phishy. This is because recent research suggests that a combination of PMT and the MINDSPACE frameworks can be useful in designing cybersecurity interventions as they focus on behaviour change attributes (Briggs, 2017). Since the research conducted in this MSc. Thesis aims at creating an all-encompassing awareness in hospital staff about identifying phishing cues by supporting their cognition to do so, the two behaviour change frameworks are used to achieve this goal. This combination framework has not yet been used in any serious games or other fields. Phish Phishy will therefore be the first operational testing of the combined use of PMT and MINDSPACE in a hospital context.

Protection Motivation Theory (PMT) is a widely used framework to develop and evaluate persuasive communication. Rogers (1975) proposed that various environmental (e.g., fear appeals) and intrapersonal (e.g., personality) sources of information can initiate two independent appraisal processes: threat appraisal and coping appraisal. *Threat appraisal* focuses on the source of the threat and factors that increase or decrease the probability of *maladaptive responses* (e.g., avoidance, denial, wishful thinking). Individuals' perceptions of the *severity* of, and their *vulnerability* to, the threat is seen to inhibit maladaptive responses. For example, smokers may consider the seriousness of lung cancer and their chances of developing the disease. *Fear* is an additional, intervening variable, between perceptions of severity and vulnerability and the level of appraised threat. Thus, greater levels of fear will be aroused if an individual perceives him or herself to be vulnerable to a serious health threat, and this will increase their motivation to engage in a protective behaviour. While perceptions of severity and vulnerability serve to inhibit maladaptive responses, there may be several *intrinsic* (e.g., pleasure) and *extrinsic* (e.g., social approval) rewards that increase the likelihood of maladaptive responses. For example, smokers may believe that smoking helps to facilitate social interaction. *Coping appraisal* focuses on the coping responses available to the individual to deal with the threat and factors that increase or decrease the probability of an *adaptive response*, such as following the behaviour of secure online browsing. Both the belief that the recommended behaviour will be effective in reducing the threat, that is, *response efficacy* and the belief that one can perform the recommended behaviour, that is, *self-efficacy* increase the probability of an adaptive response. For example, smokers may consider the extent to which quitting smoking may reduce their chances of developing lung cancer and whether they are capable of doing so. While perceptions of response efficacy and self-efficacy serve to increase the probability of an adaptive response, there may be several response costs or barriers (e.g.,

availability of resources) that inhibit performance of the adaptive behaviour. For example, smokers may believe that quitting smoking may lead to increased craving.

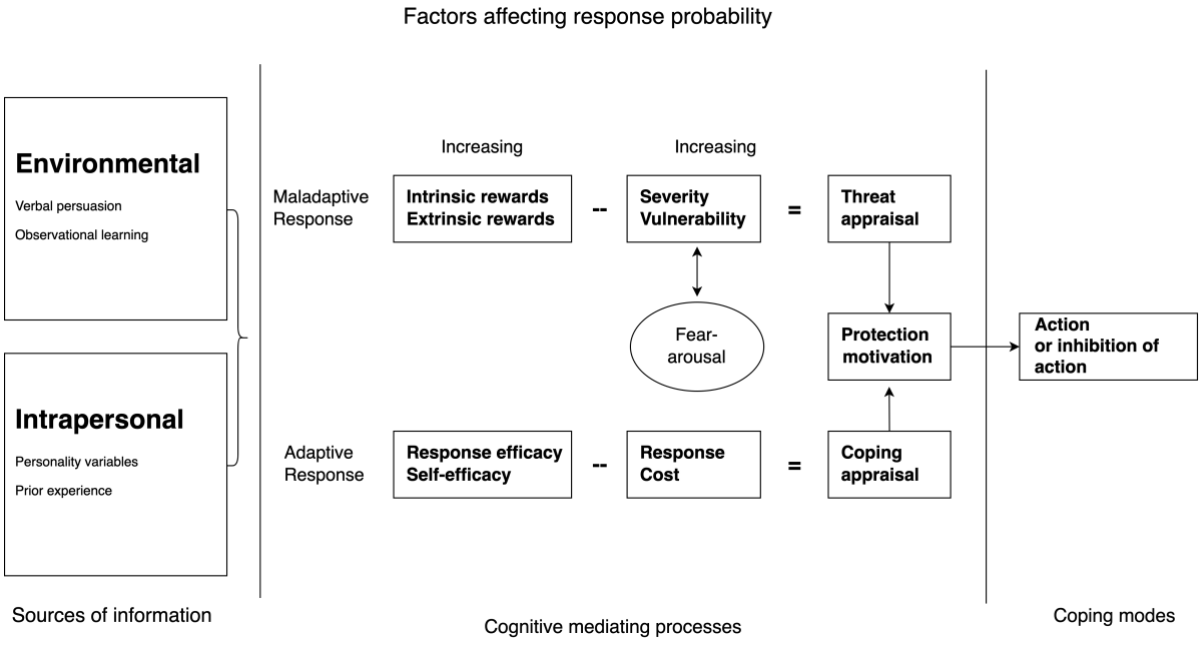


Figure 3-Depiction of the interrelating variables in the PMT framework based upon Rogers (1975)

MINDSPACE framework is defined as "a helpful mnemonic for thinking about the effects on our behaviour that result from contextual (rather than cognitive) influences" (Dolan et al., 2012, p. 265). The nine behavioural influencers are messenger effects, incentives, norms, defaults, salience, priming, affect, commitment, and ego. This framework offers new potential for managing human biases towards security awareness and seeks to alter the context of decision-making processes of System 1 cognition of the human brain (Yasin et al., 2019). It also aims to motivate human behaviour by manipulating responses to impulsive and habitual cognitive reflexes by using appropriate influencers for a given situation (Yasin et al., 2019). An example of a situational influencer is explained in section 5.1.1.

5.1.1 Combining PMT and MINDSPACE Framework in Game Design

Briggs (2017) suggests that the combination of PMT and MINDSPACE frameworks can be an effective framework for designing cybersecurity interventions, but this has not yet been tested in any field. So, in this research, the PMT framework is combined with the influencers of the MINDSPACE framework to design Phish Phishy. According to Briggs (2017), the MINDSPACE framework has the potential to influence the interrelating variables of the PMT framework and the potential of this combination will be explored through the design of Phish Phishy. This section explains the usage of different influencers in the game design.

5.1.1.1 Threat Appraisal in the Game

The most effective malware warnings are those that very clearly define the extent of risk that an individual faces if they ignored the warning (Briggs, 2017). This is translated into the game through Appreciation and Depreciation cards. The types of cards used in the game convey the extent of the risk from a malware in simple and relatable snippets of information (see section 5.3). The influencers used in these cards are *Messenger effect* (who communicates a message will affect the influence that message has on subsequent behaviour) and *ego effect* through Appreciation cards (social acceptance) and Depreciation cards (social embarrassment).

Most cybersecurity campaigns scare the user into adopting more secure behaviours but such approaches seldom work (Briggs, 2017). This is because doing so produces a maladaptive response where people tend to cope with fear appeal by denying the existence of the threat (Briggs, 2017). However, people do not experience threat unless they have some ownership or bond to the object that is threatened or under attack (Anderson & Agarwal, 2010). To overcome this problem, successful interventions that build a strong psychological relationship between worker and workplace can keep employees committed to the cybersecurity agenda at work (Leach, 2003). This is translated into the game through Appreciation and Depreciation cards. For example, the *Affect* influencer (emotional associations can powerfully shape our actions) is used in the Appreciation and Depreciation cards by linking the severity of the threat to the hospital and its patients (see section 5.3).

The organizational environment in which a threat occurs also plays an important role in shaping threat awareness in employees as well as perceived responsibility to respond to a threat (Blythe et al., 2015). Likewise, perceived vulnerability or susceptibility reflects the extent to which an individual believes that they will be affected by that threat. In other words, users who are relatively carefree about personal risk can still be called to action by critical incidents, an example being the Heartbleed security breach incident (Briggs, 2017). According to a survey conducted by the Pew Research Institute (Rainie & Duggan, 2014), 39% of Internet users changed their passwords or cancelled different accounts after the announcements on the Heartbleed hospital security breach. The user behaviour to this incident signals a change in the vulnerability perceptions (Rainie & Duggan, 2014). Therefore, in the Phish Phishy game, hospital staff will be made aware of the result of their subconscious actions on the hospital systems and their patients through real-life scenarios (Scenario cards) that they can relate to using the *Priming* influencer (acts are often influenced by subconscious cues). To be able to facilitate this, the phishing cues will be built into the game environment so that awareness creation (both of perceived severity and perceived vulnerability) can start from the game environment so that the players can relate to it in their real-world work environment (see section 5.3).

5.1.1.2 Coping Appraisal in the Game

People only engage in an action if they believe that their action will be effective, that is, if it will work to mitigate a cybersecurity threat (Woon et al., 2005). As a result, to make the hospital staff have the desire to act, it is important to incentivize the action they are expected to take, either *extrinsically* (by the organization or social group) or *intrinsically* using *ego effect* (people act in ways that makes them feel better about themselves) (Woon et al., 2005). But organizational rewards for cybersecurity behaviours are rather sparse and employees more likely see sanctions for poor security-related behaviour than rewards for good

behaviour (Briggs, 2017). Although sanctions are effective in motivating employees to read and follow a cybersecurity policy, they may only work in the short term. On the contrary, social rewards in the form of peer recognition (positive *ego effect)* and *incentives* (responses to incentives are shaped by predictable mental shortcuts such as strongly avoiding losses) can be highly influential. Similarly, social embarrassment (negative *ego effect)* can be influential in motivating behaviour. For example, even though users may become socially embarrassed by the knowledge that they have assisted in the spread of a software virus, they will more likely learn to be more cautious from that incident thereafter (Weirich & Sasse, 2001). To increase the perceived capability to act for response efficacy, both *Messenger effect* and *Group Norms* (people are strongly influenced by what others do or communicate) is considered. An example of *messenger effect* in this context is that users recognize peers or line managers to be approachable if they need advice on how to cope with a particular threat thereby making this the *Group Norm* as well (Blythe et al., 2015). These attributes translate into the game by giving the players the option to discuss with their co-players (peers) and learn from the action taken by a peer towards a phishing signal. The following sections describe the game mechanics and gameplay based on the combination of the two frameworks as shown in figure 4.

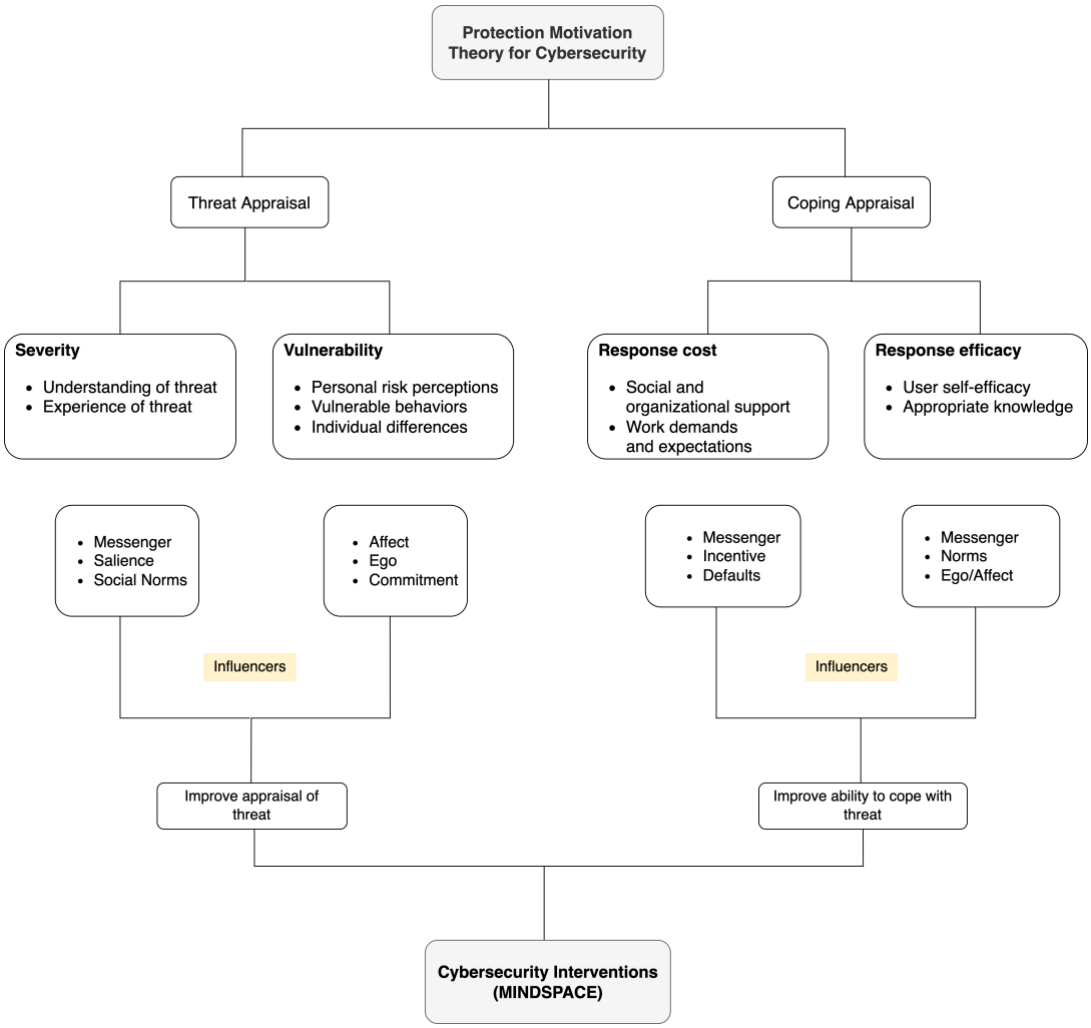


Figure 4- Combination of PMT and MINDSPACE frameworks for creating cybersecurity interventions based upon Briggs (2017).

5.2 About the Game: Phish Phishy

The goal of the game, Phish Phishy, is to create a learning environment for hospital staff to create awareness on identifying phishing signals (phishing cues) and apprise the staff of the possible actions that they can take to identify phishing signals. To meet this goal, certain requirements must be met. First, the game should facilitate discussion among the players so that the players can learn from one another to increase their self-efficacy (capability to act) from their perceived susceptibility. To facilitate this, the game will be scenario-based and action-focused for building awareness. Second, the game must be of a short duration similar to a conventional method of training such as a short presentation or seminar. Third, the game must be easy and understandable so that the hospital staff can use their cognition to identify phishing signals and to take the necessary action against these signals. To facilitate this, Phish Phishy will be a simple tabletop card game because tabletop card games have proven to be useful for creating cybersecurity awareness (Hart et al., 2020).

5.3 Designing the Game Environment

The game environment in Phish Phishy encompasses the daily online work interfaces that the hospital staff encounter such as emails, login websites, and work phone SMSs. Therefore, utilizing these avenues as the game environment for creating awareness can help hospital staff fully understand how attackers could manipulate them. The game environment is adapted into game cards. These game cards are divided into scenario cards, action cards, appreciation cards, and depreciation cards. Each card is numbered so that the actions chosen by every player can be tracked to verify the results of the game.

Scenario cards are of two types: Legitimate cards and Phishing cards. The legitimate cards contain legitimate email, SMS, or websites of the hospitals. The phishing cards contains contents that are very similar to legitimate card contents except that they include phishing cues somewhere in the card. For example, an email with Microsoft's template requesting the receiver of the email to update their Outlook password but on careful observation, one can notice that the sender's email address is neither the hospital's IT department nor Microsoft. It instead displays an individual's or another organization's email ID. To summarize, phishing cards in Phish Phishy include real-world cues used by attackers such as subtle changes in visual representations, spelling and grammatical errors, and suspicious sender's address (Bayl-Smith et al., 2020). Please note: The actual scenario cards and its context have been removed from this version of the report due to confidentiality.

The scenario cards resort to persuasion techniques such as authority, urgency, reciprocity, consistency, and commitment to encourage limited cognitive processing so that the cue utilization in users is limited (Bayl-Smith et al., 2020). This will mimic reality wherein hospital staff could be nudged to easily respond to suspicious cues.

Action cards are of six types: (i) Report as suspicious (ii) Delete or ignore the scenario (iii) Talk to you colleague and delete or ignore the scenario (iv) Talk to your colleague and report as suspicious (v) Trust and respond to scenario (vi) verify with supervisor (line manager). The action cards allow the option for verifying with peers and supervisors (line managers) to facilitate *Messenger Effect* and *Group Norm*.

Appreciation cards are used in the game to develop positive *ego effect,* and this is facilitated by appreciating the player for every phishing card identified during the gameplay. The contents of the card are aimed at intrinsically motivating the player by targeting the magnitude of impact the player prevented by correctly identifying and reporting the phishing signal. For example, an appreciation card can be read as follows: *"Thank you for taking that extra step to report this suspicious activity. Your response saved the cardiology department from shutting down due to a large cyberattack. Your action saved 15 critical heart surgeries today!"* The appreciation card is backed up with points in the form of chocolates and applause from the other players.

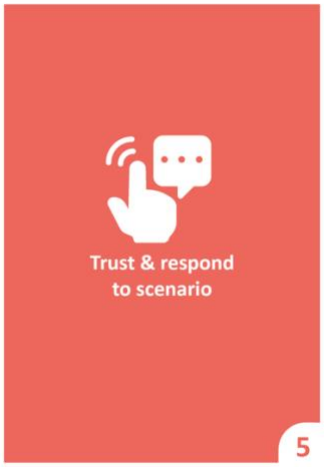


Figure 5-Example of an Action Card

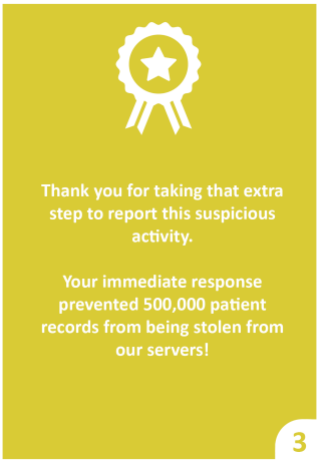


Figure 6-Example of an Appreciation Card

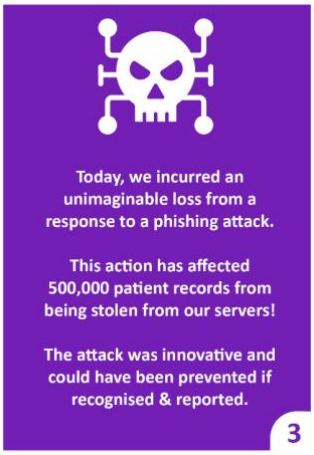


Figure 7-Example of a Depreciation Card

Depreciation cards are used in the game to develop negative ego effect, and this is facilitated by sanctioning the player through social embarrassment and a negative point for every phishing card that a player did not identify during the gameplay. For example, a depreciation card can be read as follows: Today, we incurred an unimaginable loss from a response to a phishing attack. This action has affected 15 critical heart surgeries today and the remaining surgeries for this week have been cancelled until further notice. The attack was innovative and could have been prevented if recognized & reported. The Game Master explains each phishing signal in a scenario card to justify the rationale that led to the negative outcome in the game. The depreciation cards received by a player is also an indicator of a player's gullibility and low cue utilization which makes them fall for persuasive phishing cues in the scenario cards. The depreciation cards are sufficiently dramatized to incorporate the fear arousal element of the PMT framework so that in the next gameplay round, the players carefully assess scenario cards.

In addition to the game cards, each player is asked to fill out a pseudo-anonymous paper-based pre-survey (before the start of the game) and a pseudo-anonymous paper-based post-survey (after the gameplay) to evaluate the effectiveness of the game on creating awareness in individual players about identifying phishing cues and taking the appropriate action (reporting it). A pseudo-anonymous tracking sheet is also provided to each player to capture entries of every scenario card and the corresponding action card that a player chooses. The surveys and tracking sheets are made pseudo-anonymous because each of these sheets are numbered and then distributed to the players. For example, Pre-survey 1, Post-survey 1, and Tracking

sheet 1 are given to a player and that player is thereafter referred to as Player 1 during the gameplay. This is done to check the effect of Phish Phishy on each player before and after gameplay by checking the Pre-survey 1 against the Post-survey 1 for Player 1. The questions asked in the pre-survey and post-survey aim to gather information on (i) level of understanding of phishing amongst the players (ii) prior experiences of the players with phishing scams (iii) alertness while assessing work emails, SMSs, or websites (iv) actions chosen when faced with a suspicious activity at workplace. Appendix B has all the questions that were used in the pre-survey and post-survey and includes a sample of the tracking sheet. Scenario cards are adapted based on the chosen hospital's work environment and this adaption is discussed in Chapter 6. In Chapter 6, a few examples of the game cards are also provided explaining the phishing cues.

5.4 Designing the Game Mechanics and Gameplay

The game mechanic is structured into correctly identifying phishing card from the pile of legitimate and phishing cards. A typical gameplay would request players to sit around a table with the pre-survey sheet, tracking sheet, and a set of six action cards. The game master provides a scenario card to 'Player N' while the rest of the players wait for 'Player N' to select an action card and record it in the tracking sheet. After 'Player N' selects the action card, other players receive the same scenario card, and they must record it as well. The game mechanic is designed to replicate the real-world situation wherein a hospital staff view emails, SMS, and websites at their individual workstation. All players receive twenty seconds to evaluate a scenario card (during their turn) which is the average time taken to view and believe an email, SMS, and website, and this is timed by the Game Master using a stopwatch. The gameplay is divided into two rounds. In each round, players receive different legitimate scenario cards and phishing scenario cards. The sequential order of the gameplay is described below.

A. Briefing

Before the gameplay is explained to the players, the players are expected to fill in their pre-survey. This is done to assess the awareness levels of the players before playing the game. After the players submit the pre-surveys, the Game Master briefs the game mechanics and gameplay to the players. After the players have filled in the pre-survey, the gameplay begins.

B. Gameplay Round 1

1. Game Master provides a scenario card to Player N.
2. Player N must identify if it is legitimate or suspicious (threat) card.
3. Player N selects an appropriate action card for that card based on the scenario.
4. Player N notes the action card number for the given scenario card into the tracking sheet.
5. All players then receive the same scenario card, and they too must select an appropriate action card and note it in their respective tracking sheets.
6. This process continues until every player gets their individual turn to identify a scenario card and select the appropriate action card.
7. After all the players finish their turn, the Game Master reveals the results for every card by emphasising the phishing cues.

8. The winners receive appreciation cards as well as points (chocolates) for social acceptance, recognition, and increasing self-efficacy for the next round (**Round 2**).
9. Players who were vulnerable and fell for the phishing cues receive depreciation cards and negative points for negative ego effect in the form of social embarrassment. However, this also increases self-efficacy for the next round (**Round 2**).
10. End of round one, the Game Master emphasizes on the need to report suspicious cards by choosing the 'Report as Suspicious' action card. The need to do so is emphasized so that the players understand that the IT department can investigate the activity and take necessary action to prevent suspicious activities from recurring.
11. The revelation of results is expected to stimulate conversation amongst players about the suspicious cards, phishing cues, and what they focused on.
12. All players give their opinion based on prior experience or expertise. Leading towards self-learning as well as peer-learning.

C. Gameplay Round 2

The gameplay is repeated as in Round 1 but with different set of Scenario Cards.

D. Debriefing

After the gameplay, the players are expected to fill in their post-survey. This is done to assess the awareness levels of the players after gameplay to test the effectiveness of the game. After submitting the post-surveys, the Game Master initiates an open discussion for the players based on the time left so that the players can share their learnings, feedback, experiences from the game, and prior phishing experiences, if any. The open discussion is initiated at the end of the game so that the knowledge gained from the discussion does not interfere with each player's response in their survey. This discussion concludes the gaming session.

5.5 Desired Outcome

There are two outcomes expected from the players after playing the game. First, an improvement in threat appraisal among the players especially those who were exposed to a threat scenario card or phishing card. This is to test the effectiveness of PMT and MINDSPACE frameworks in the design of a cybersecurity intervention. Threat appraisal can be measured by the increased alertness in players by comparing pre-survey with post-surveys. Second, ability to cope with the threat. This can be measured by the number of 'Report as suspicious' metric chosen by the players in their post-surveys which should ideally be a result from the increased alertness.

5.6 Conclusion of Chapter 5

The aim of this chapter was to design the game in such a way that it helps hospital staff identify phishing signals in their work environment. The game design is summarized to answer the sub-part of the second sub-research question:

2.1 How can the identified awareness intervention be designed for a hospital environment to help hospital staff differentiate between legitimate and phishing signals?

The awareness intervention, Phish Phishy is designed as a simple tabletop card game because tabletop card games have proven to be useful for creating cybersecurity awareness because they engage the players and can be easily altered to be a short game session (Hart et al., 2020). The game is based on the Protection Motivation Theory (PMT) and MINDSPACE frameworks to help the hospital staff identify phishing cues in their work environment. The game aims to bring out threat appraisal and coping appraisal in the players using MINDSPACE influencers (such as messenger effect, incentives, group norms, priming, affect, commitment, and ego effect) in the game cards.

The game environment in Phish Phishy is designed to contain scenarios from the daily online work interfaces of the hospital staff (emails, login websites, and work phone SMSs). The game environment is adapted into legitimate and phishing scenario cards and the players are expected to select an appropriate action card during their turn. Before the gameplay, the players must answer a pre-survey to assess their awareness levels before the game. The game is played in two rounds with different set of scenario cards wherein each player gets one chance to draw a card and identify whether it is legitimate or phishing and track it into a tracking sheet so that the rest of the players can draw the same card and select an action card for it. Round 1 is complete when all the players have received their chance to draw a card. The Game Master then reveals the results of the game and provides appreciation cards (positive ego effect) to the players who correctly identified the phishing cards and depreciation cards (negative ego effect) to the players who fell for the phishing card. These cards are distributed to improve the self-efficacy (capability to act) of the players for the next round. After Round 2, the players are asked to answer the post-survey so that the awareness levels of the players can be assessed after playing the game. Both the surveys are made anonymous so that the players can provide unbiased answers. Therefore, Phish Phishy is so designed that it helps players (hospital staff) to identify different phishing signals through threat appraisal and coping appraisal.

The following chapter will focus on the Phish Phishy gameplay (testing) using hospital staff (players) from two large Dutch hospitals. This is an important step in the game methodology because it will assess the effectiveness of the game. Additionally, from a research novelty standpoint, the combination of PMT and MINDSPACE framework will be explored in a practical setting for the first through Phish Phishy.

Chapter 6: Testing the Awareness Intervention

This chapter dives into the gameplay sessions with the target audience to test the effectiveness of Phish Phishy as an awareness intervention. The chapter begins with the sample selection procedure for the first gameplay session, followed by gathering inputs for designing the contents of the Scenario Cards, then the procedure of setting up the game in the sample hospital is described. The result from the gameplay is also described to draw conclusions and validate the effectiveness of the game. Based on the results from the first gameplay session, adjustments are made in the game. The chapter concludes by answering the third sub-research question.

6.1 Gameplay Session 1

To test Phish Phishy, a sequence of steps was adopted. First, a sample set of players was selected to represent a part of the hospital population. Second, the gameplay was setup for the players in the sample according to the gameplay design described in section 5.5. Third, the result from the gameplay is presented and concludes with the learnings from the gameplay session.

6.1.1 Sample Selection

A large Dutch academic hospital was selected to represent the population of hospitals in the Netherlands. This hospital is henceforth referred to as Dutch Hospital 1 in this research. To play the game, hospital staff who are the target audience is necessary. Contacting hospital staff from outside the organization for a gameplay session was difficult. But with the help of Dr. Irene Grossmann from the Safety and Security department at the TPM faculty, the attention of the Chief Information Security Officer (CISO) at Dutch Hospital 1 was made possible. The CISO's interest in the potential of such a game helped this research in getting hospital staff to play the game.

Since a hospital comprises of different kinds of employees such as doctors, nurses, admin staff, IT staff, technicians, lab assistants, academicians, having a varied set of players was considered as the best representation of a hospital sample. The CISO's team contacted various hospital staff asking if they were willing to be a part of this gameplay (research). At the end of the three weeks, ten hospital staff agreed to take part in the game. So, a heterogenous sample of hospital staff was selected. The occupation of the ten players comprised of one or more nurses, accountants, IT security experts, lab technicians, visiting health professionals, and researchers. It is important to note that the occupation of individual players is not given emphasis in this research rather the occupation of the players should collectively represent the hospital staff. This is because the aim of this research is to apprise hospital staff on identifying phishing signals rather than focusing on their individual occupations or demographics. In addition to the players, the CISO and the Data Protection Officer (DPO) at Dutch Hospital 1 requested to be a part of the game by silently observing the game so that they could gauge the need as well as the effectiveness of a game-based awareness intervention.

6.1.2 Think Like a Hacker: Gathering Inputs for Scenario Cards

To gather inputs for designing scenario cards, a web search was conducted to identify the different employee portals in Dutch Hospital 1. The search provided many inputs of how the hospital's staff portals look, of which the employee login page and employee training websites were selected. In a 30-minute Microsoft

Teams meeting with the CISO and the DPO of Dutch Hospital 1, the relevance of these webpages was validated. In addition, they also provided a sample invoice email to understand the email environment.

Based on these inputs, the scenario cards were designed in Adobe Photoshop. The entire process of data gathering and manipulation to design the final scenario cards for Dutch Hospital 1 took four full days of man-hours or effort (8 hours each day). After the scenario cards were designed, the appreciation and depreciation cards were designed in Adobe Photoshop. The cards were then printed and encased in reusable lamination sheets. To test the sequence of the gameplay starting from briefing to debriefing, trial rounds were conducted with different participants such as students, IT professionals, Engineering consultants, and Game Designers. A maximum of five participants played in each trail round. The game mechanics worked as designed and therefore proceeded towards the first gameplay session with hospital staff.

6.1.3 Setup of Gameplay I

The game participants were invited to play the game for a total time of 60-minutes on a set date based on their availability with the help of the CISO and DPO at Dutch Hospital 1. The first gameplay session was conducted in a closed meeting room wherein the players were seated in a circle around the desk facing each other. Such a seating arrangement was chosen so that the players could see each other and openly discuss their learnings. Informed Consent (IC) forms and pre-surveys were distributed to every player to receive their consent to using the game results for this research and to assess their existing awareness level about phishing, respectively (available in Appendix B). The forms were collected before the briefing of the game.

6.1.4 The Gameplay

The gameplay described in section 5.4 is followed in Dutch Hospital 1 and is therefore not repeated in this section. But hospital specific changes and/or examples are explicitly described in this section for better clarity. Since there were ten players and the gameplay had two rounds, twenty scenario cards were designed.

Gameplay Round 1

Step 7 of gameplay: After all the players finished their turn, the Game Master explained each persuasive technique used in the phishing card. For example, phishing signal based on *similarity* can be 'hospital.nl' is spelt with a double S as 'hosspital.nl'.

Step 11 of gameplay: After the Game Master revealed the results, some of the players provided their opinion on different factors that could be phishing cues. This helped facilitate peer-to-peer learning.

Debriefing

After submitting the post-surveys, the players who were not in a rush provided feedback on the game in the comments section of the post-survey. The Game Master asked the players to share their learnings or prior experiences with each other. Only the player who was able to identify all the phishing cards actively took part in the debrief session. The player summarized different phishing cues that the player had encountered in the cards to the players who stayed back for the debriefing, but it did not stimulate any conversations

amongst the players. The debrief session was less conversational than it was thought to be during the design of the game.

Observations during Gameplay 1

Most of the participants were reluctant in speaking their minds with the rest of the participants. This could be attributed to three reasons. First, the sample of participants chosen belonged to different departments in the hospital, so most participants did not know each other. Second, the presence of authorities like the CISO and DPO as observers could have made the participants feel conscious or watched. Lastly, since the game had ten participants in one session, players seem distracted a few minutes after their turn was complete. Another observation is that the players who thought they received a suspicious card during their turn seemed involved in the game.

6.1.5 Survey Results from Gameplay 1

After two rounds of gameplay at Dutch Hospital 1, the pre-surveys and post-surveys were collected from the participants to analyze the impact of the game on the players. All the results gathered from the surveys are plotted into bar charts and are available in Appendix C. However, the results that signify the impact of the gameplay on the participants are discussed further in this section. In the result figures, the blue bars and orange bars represent the results from before and after the gameplay, respectively.

I. Level of Understanding

Before factoring in the effect of the game on the players, it is important to know their level of understanding about phishing so that any change in this level can be recorded. Therefore, the pre-survey and the post-survey both ask the participants the same question and this is articulated as: *What is your level of understanding of the term phishing?*

The comparison suggests that after playing the game, the understanding of the term phishing increased among the players (see figure 8). Even though there were three participants with some and no understanding of the term phishing before the game, all the players rated themselves with good understanding or higher understanding levels after the gameplay.

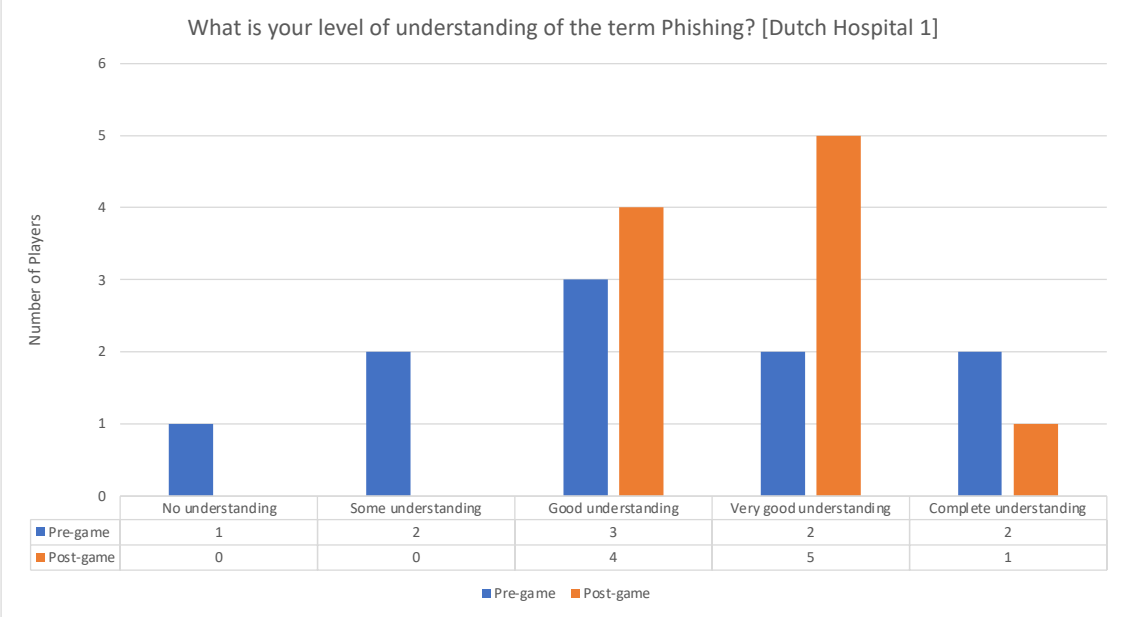


Figure 8-Depicts improvement in the level of understanding of phishing after gameplay

II. Recognizing Phishing Signals

The aim of Phish Phishy is to help its players become aware of different phishing signals and identify them in their work environment. So, it is important to capture their ability to identify phishing signals before and after the gameplay. Therefore, the pre-survey and post-survey both ask the participants the same question and this is articulated as: *How would you rate yourself in recognizing a phishing signal?* The question is based on a rating scale that is set between zero.

The comparison suggests that after playing the game, the confidence of the players to identify phishing signals has increased (see figure 9). The result also ties back to the desired outcome of the game (see section 5.5), that is, the players show an increase in the ability of threat perception after gameplay. The result in the chart suggests that the game meets the desired outcome.

It is also important to note that one among the ten players chose to not answer a few questions in the post-survey. Perhaps this player like few other players had other appointments at the latter half of the gameplay, so they had to leave right after the gameplay. Therefore, the pre-survey and post-survey of this player is considered as an anomaly and removed from the results.

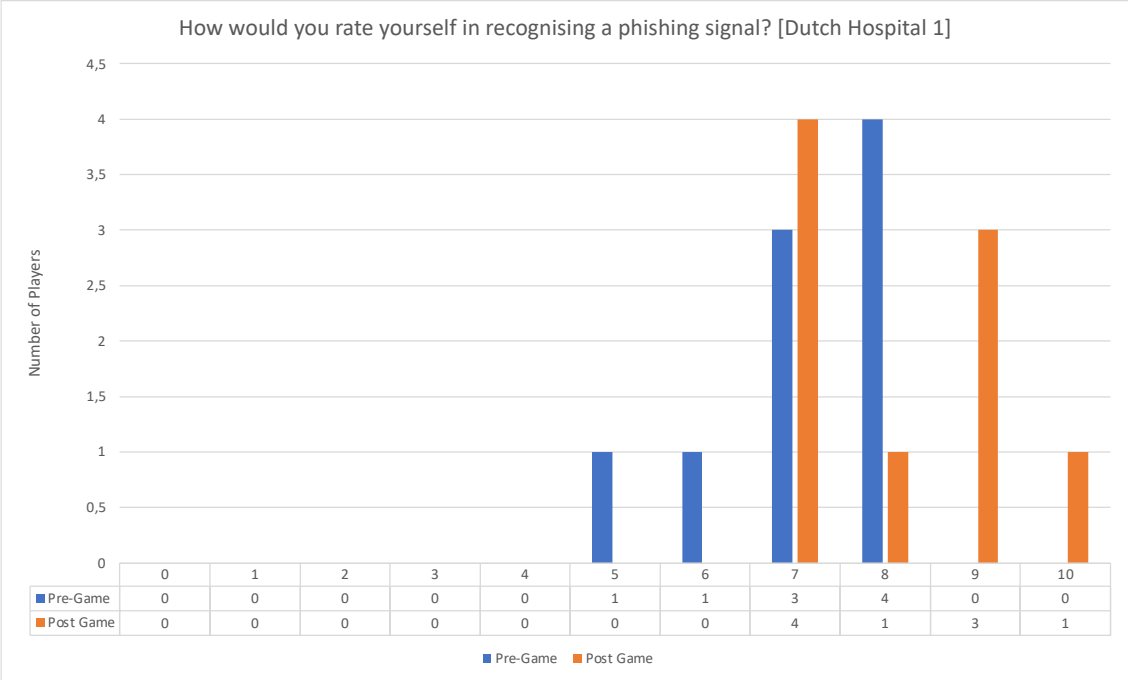


Figure 9-Depicts the impact of the game on identifying phishing signals.

III. Alertness after Gameplay

The post-survey explicitly asks the players about how alert they would be in their work environment after playing the game. The post-survey articulates this question as: *How would you rate yourself in recognizing a phishing signal?* The result to this question builds further on the findings from figure 9.

The result of this question suggests that the players would assess their work interfaces such as emails, SMSs, and login websites more cautiously (see figure 10) after the gameplay. This outcome translates to the *perceived severity* and *perceived vulnerability* of the PMT framework, that is, when faced with a threat and the extent of the risk it poses, an individual understands the severity to and vulnerability from the threat (see section 5.1.1). But those who did not receive a threat card in either of the two rounds (cross-checked using the pseudo-anonymous tracking sheets), answered that they would be as alert as they have been.

The game had no negative outcomes wherein the players' alertness levels decreased after the gameplay. Therefore, the result suggests that if every player is exposed to a threat card in more game rounds, the alertness levels of all the players can be increased.

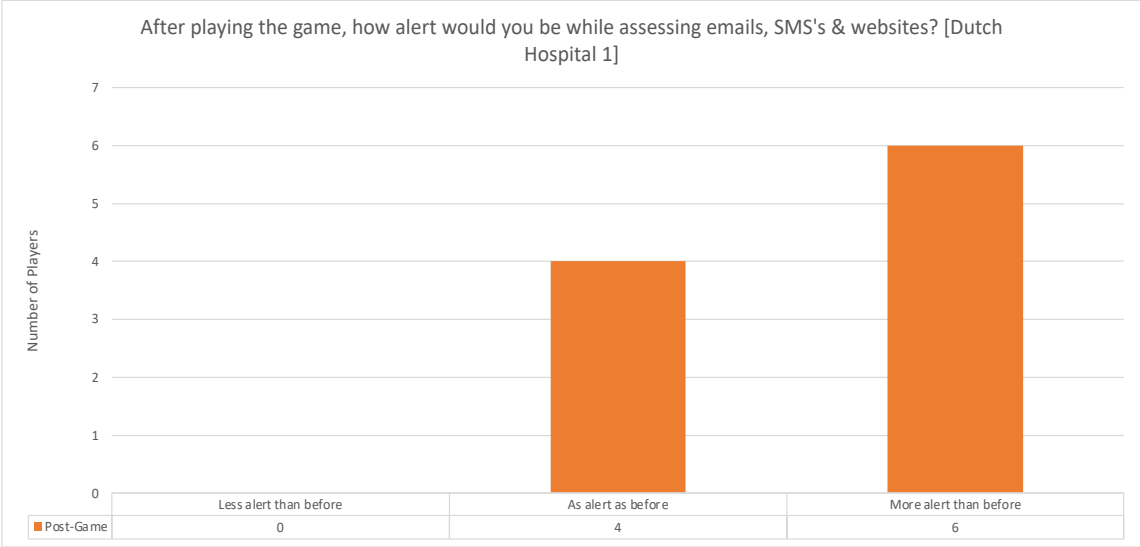


Figure 10-Depicts the alertness levels of the players after gameplay.

IV. Reporting a Suspicious Activity

The pre-survey and post-survey ask the participants as to what action they would take if they found a suspicious online activity in their workplace to observe changes in their coping appraisal. The question is articulated as: *How would you react if you found a suspicious online activity at your workplace?* The question is a multiple-choice question that duplicates the action card options.

The comparison suggests that all the participants selected the 'Report as Suspicious' option to report suspicious activities to the IT department both before and after the gameplay (see figure 11). This means that the players are aware of the ideal action that they must take with suspicious activities. But in the post-survey, some of the participants also chose other options such as (1) 'Verify with supervisor' (2) 'Talk to your colleague' (3) 'Delete the activity' (4) 'Ignore the activity'. Although choosing to verify with supervisor and colleague is a step closer to the desired coping appraisal of reporting an activity by reaching out for validation from peers, selecting the 'Delete the activity' or 'Ignore the activity' does not meet the ideal coping appraisal (to report suspicious activities). The results gathered at Dutch Hospital 1 on the participants' coping appraisals therefore contradict the ideal action. This anomaly could be due to the clashing meetings that some of the players had in the last 15-minutes of the gameplay which may have led to a hurried filling out of the post-survey. Therefore, to shed light on this metric, another gameplay session is necessary.

Figure 11-Depicts the coping appraisal of the players before and after gameplay.

6.1.6 Learnings from Gameplay 1

The gameplay at Dutch Hospital 1 was used as a starting point to test the effectiveness of the combination behaviour change models of PMT and MINDSPACE. The learnings from the gameplay are summarized below.

I. **Threat Appraisal**: The survey results suggest that using the PMT and MINDSPACE frameworks in the game design have helped to apprise the players about different threats by focusing on the threat severity (through the appreciation cards) and threat vulnerability (through the depreciation cards). The distinction in the pre-survey and post-survey results for questions I, II, III in section 6.1.3, emphasizes on threat appraisal and suggests a positive increase in the ability to identify phishing signals and being more aware about the work environment after playing the game. This improvement in threat appraisal can be seen through the results in figure 8, figure 9, and figure 10.

II. **Coping Appraisal**: The action cards provide five options of responding to threats besides the 'Trust and Respond' option. Even though the Game Master emphasizes on the importance of reporting suspicious scenarios cards, players also choose to delete or ignore the suspicious scenario. Therefore, the results shown in figure 11 do not provide a clarity on the coping appraisal of the players after the gameplay 1.

III. **Team Dynamics**: The players did not openly communicate with each other during the gameplay as described in section 6.1.4. A conducive gameplay session is essential for the effectiveness of the game. To mitigate this first, the participants should somewhat be familiar with one another. Second, formal authorities such as CISO, DPO, or higher-level managers should not be present at the gameplay session rather, a peer of the participants should be the observer so that the players do not feel judged or watched. Lastly, Phish Phishy should only have a maximum of 6 players so that the time taken for gameplay is shorter and everyone can be engaged in the game.

Therefore, these learnings are used to restructure the gameplay session at Dutch Hospital 2 (see section 6.2).

6.2 Gameplay Session 2

Based on the learnings from the gameplay in Dutch Hospital 1, alternations were made in the gameplay in Dutch Hospital 2. This section only describes the alterations that were made to Gameplay 2 because the remainder of the gameplay is based on the game design described in section 5.4. The aim of the gameplay 2 is to double-check if fewer number of players and players from similar departments (teams) improve their awareness levels to identify phishing signals after the gameplay.

6.2.1 Sample Selection

A large Dutch academic hospital that was similar in size and function to Dutch Hospital 1 was selected. This hospital is henceforth referred to as Dutch Hospital 2 in this research. With the help of one of the members of the Graduation Committee of this thesis, it was possible to reach out to the hospital staff at Dutch Hospital 2. Based on the learnings from Gameplay 1, an invitation to play the game (45-minutes) was sent out department-wise so that the players were acquainted with one another (to improve the gameplay dynamics). Seven hospital staff from the academic department of the hospital confirmed their availability to play the game. The sample comprised of research students, researchers, and lecturers. In Dutch Hospital 2, a peer member was selected as the observer of the game to make the players comfortable. The peer (observer) was introduced into Gameplay 2 to duplicate the presence of the CISO and DPO from Gameplay 1. To summarize, the gameplay consisted of seven participants of which 6 participants were players and 1 participant was an observer.

6.2.2 Gathering Inputs for Scenario Cards

The input for the scenario cards were similarly gathered as in Gameplay 1. To match the work environment of Dutch Hospital 2, twelve Scenario cards were redesigned (for six players), but the Appreciation, Depreciation, and Action cards were reused from Gameplay 1. The sequence of the gameplay was tested starting from briefing to debriefing with a group of six students before the gameplay session.

6.2.3 The Gameplay

The gameplay described in section 5.4 is followed in Dutch Hospital 2 and is therefore not repeated in this section. But hospital specific changes are explicitly described in this section for better clarity.

Debriefing

After the players filled in their post-surveys, the Game Master asked the players to share their learnings or prior experiences with each other. While some of the players shared their experiences from the game with respect to the subtle changes each phishing signal had, the others were curious about the design of the scenario cards which was explained by the Game Master. Few players also shared their real-life phishing experiences with the group. All the players keenly listened to one another and added their experiences until

the session came to an end. The debrief session was dynamic due to the conversations and competitive involvement of the players in the game.

Observations during Gameplay 2

The participants interacted with one another after identifying the action for each scenario card. The players were also competitive and engrossed in the game. The observations of Gameplay 2 are more positive than Gameplay 1 and can be attributed to the changes made based on the learnings (see section 6.1.6). Like the observation in Gameplay 1, the players who thought they received a suspicious card during their turn seemed to be involved in the game.

6.2.5 Survey Results from Gameplay 2

After two rounds of gameplay at Dutch Hospital 2, the pre-surveys and post-surveys were collected from the participants to analyze the impact of the game on the players. In addition to validating the increase in threat appraisal, the results were also collected to assess an increase in the coping appraisal, that is, reporting suspicious cards rather than deleting or ignoring them. The results from Gameplay 2 are discussed below.

I. Level of Understanding

The comparison suggests that after playing the game, the understanding of the term phishing increased among the players (see figure 12). Even before the gameplay, most players considered themselves to have a good understanding of phishing and it further increased after the gameplay.

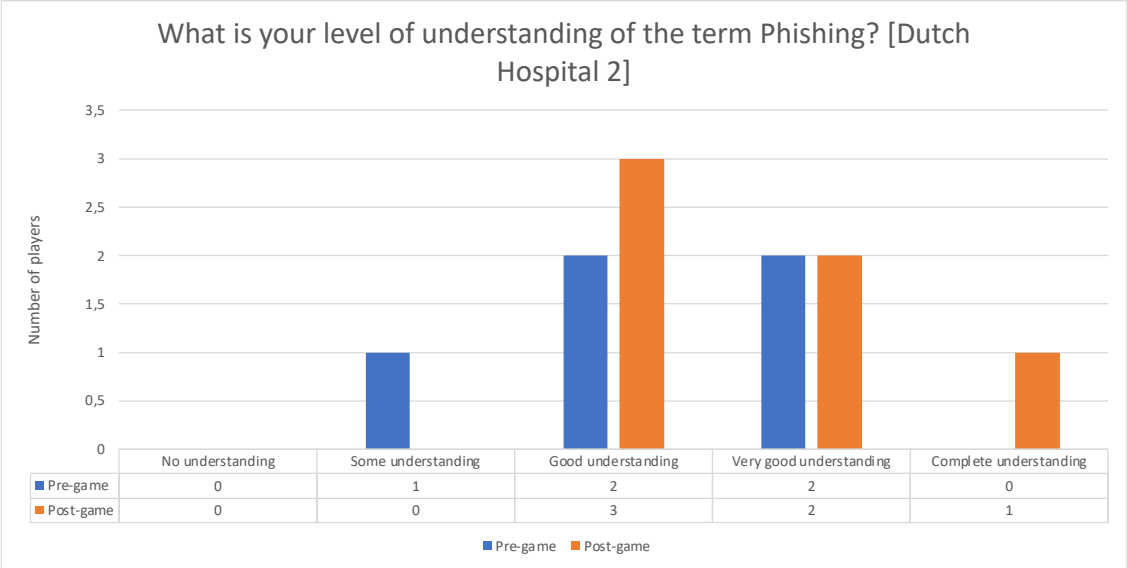


Figure 12-Depicts improvement in the level of understanding of phishing after Gameplay 2.

II. Recognizing Phishing Signals

The comparison suggests that after playing the game, the confidence of the players to identify phishing signals has increased further (see figure 13). The result also ties back to the desired

outcome of the game (see section 5.5), that is, the players show an increase in the ability to perceive threat after the gameplay.

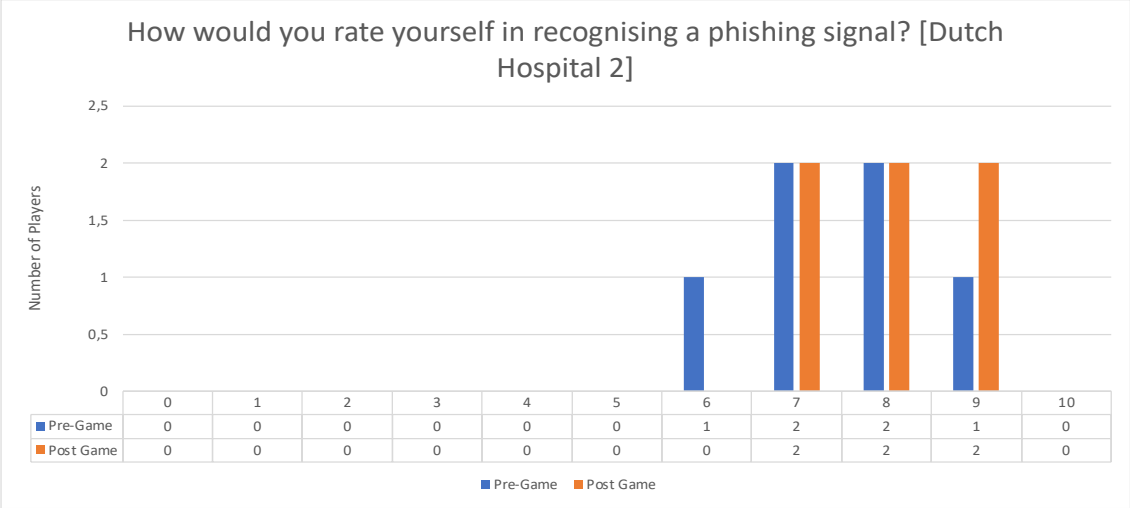


Figure 13-Depicts the impact of the game on identifying phishing signals

III. Alertness after Gameplay

The result of this question suggests that the players would assess their work interfaces such as emails, SMSs, and login websites more cautiously (see figure 14) after the gameplay. This outcome translates to the *perceived severity* and *perceived vulnerability* of the PMT framework, that is, when faced with a threat and the extent of the risk it poses, an individual understands the severity to and vulnerability from the threat (see section 5.1.1). But the player who did not receive a threat card in either of the two rounds (cross-checked using the pseudo-anonymous tracking sheets), were as alert as before.

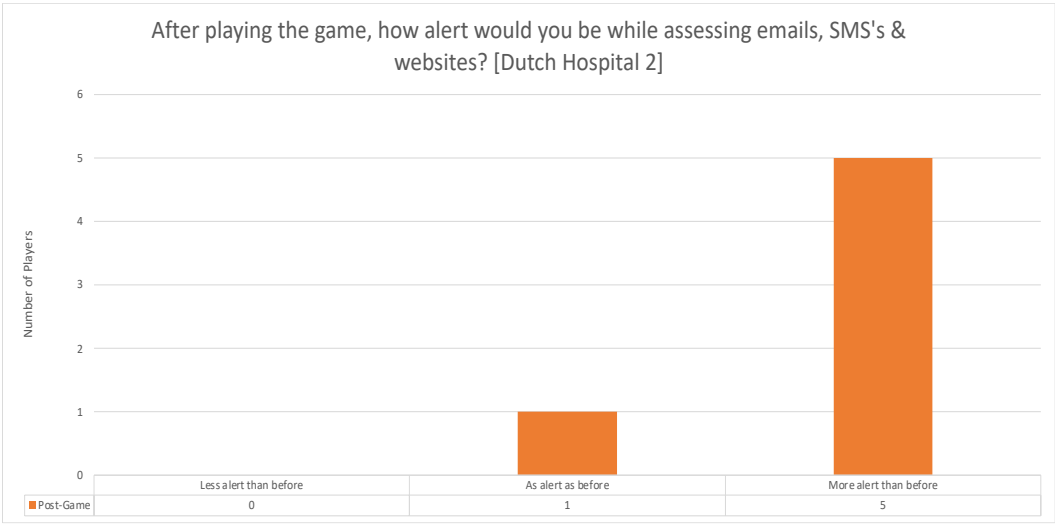


Figure 14-Depicts the alertness levels of the players after gameplay 2

IV. Reporting a Suspicious Activity

The comparison suggests that after altering the game, the participants selected the 'Report as Suspicious' option to report suspicious activities. In addition to this choice, one of the participants also chose the 'Talk to your colleague' option to double-check with a peer in case of suspicion (see figure 15). The results show that the coping appraisal increased among the players after modifying Gameplay 1.

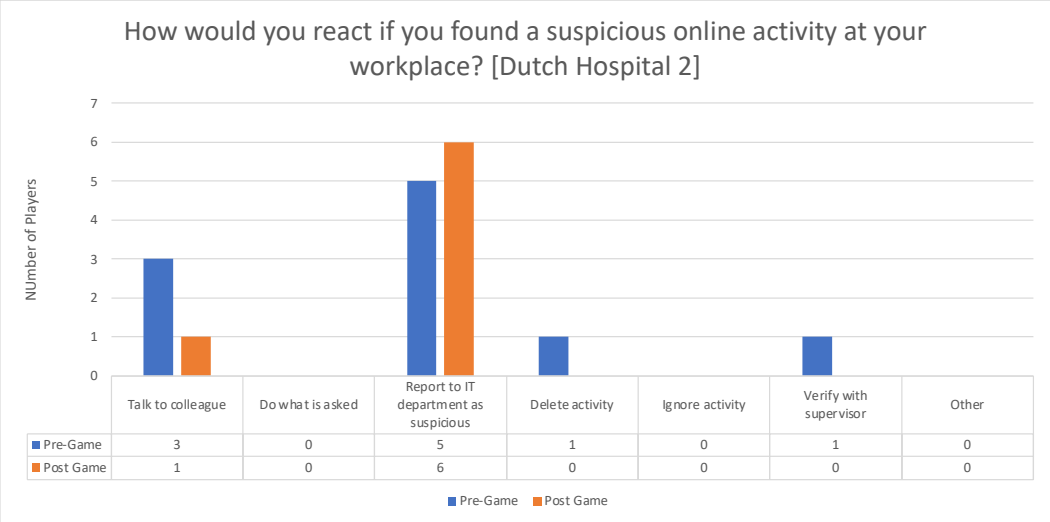


Figure 15-Depicts the coping appraisal of the players before and after gameplay.

6.3 Comparing Results of Gameplay 1 & 2

The two gameplay sessions were not identical, so a comparison is made between the two in this section to summarize the insights. The flowchart of the comparison is shown in figure 16.

I. **Game setup**: The ten participants in Gameplay 1 had to follow through every player's turn, so it reduced the level of involvement of the players after the first 30-minutes of the gameplay. While Gameplay 2 included only six participants, so the time taken for every turn much shorter, and it helped sustain the involvement of player's turn through the game. When a peer was the observer in Gameplay 2, the participants seemed to less conscious than the participants in Gameplay 1. Additionally, the changing the gameplay suggests an improvement in the coping appraisal of Gameplay 2.

II. **Team Dynamic**: The participants in Gameplay 2 were vocal about their thought processes while identifying phishing signals compared to participants in Gameplay 1. Moreover, the participants in Gameplay 2 were energetic and competitive in correctly identifying phishing signals so that they could win. This was an important gaming element that contributed towards engaging all the participants and was not evident in Gameplay 1. Therefore, in future gaming sessions, it is recommended to conduct the awareness trainings wherein the participants belong to the same unit or team for a conducive learning experience.

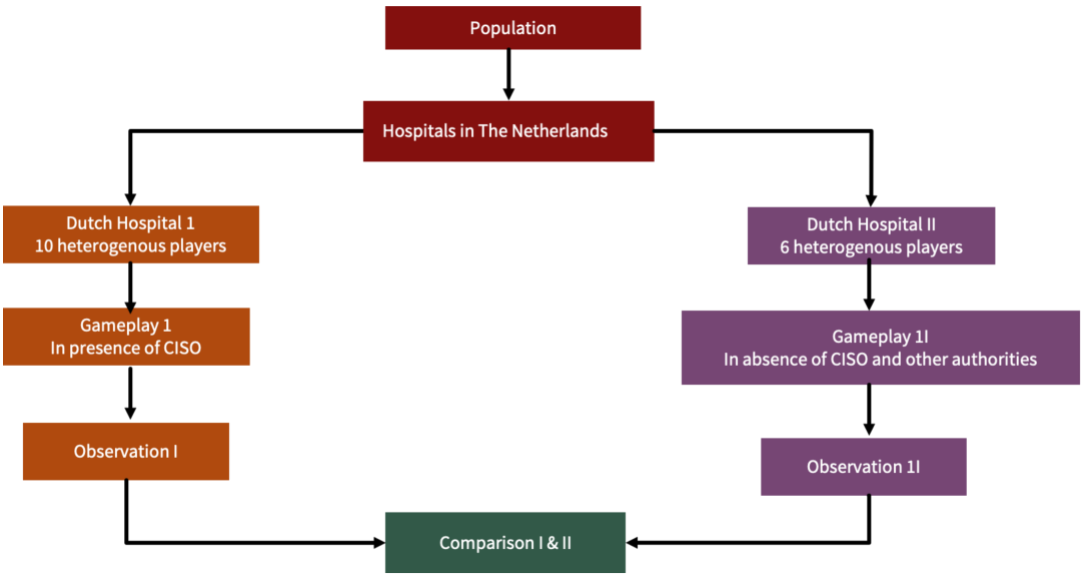


Figure 16-Flow diagram depicting the comparison of Gameplay 1 & 2

6.4 Conclusion of Chapter 6

The aim of chapter 6 was to test the game design with the hospital staff in their work environment and this is summarized to answer the third sub-research question.

3. *What attributes of awareness are triggered by the awareness intervention?*

The design of Phish Phishy was first tested in a large Dutch Hospital called Dutch Hospital 1. The participants of Dutch Hospital 1 were ten heterogenous players from different departments of the hospital such as doctors, nurses, administrative staff, IT staff, technicians, lab assistants, and academicians. The surveys of Gameplay 1 suggested an increase in the awareness levels of the participants in terms of their *threat appraisal* because both *threat severity* and *threat vulnerability*, the two attributes of threat appraisal which were addressed through the messages in the Appreciation and Depreciation cards. However, the results did not suggest an increase in the coping appraisal because the game participants not only chose 'Report as Suspicious' but also other options such as 'Delete the activity' and/or 'Ignore the activity' when asked how they would react to a threat scenario in the future. The ideal coping mechanism is to inform the hospital's IT department of suspicious activities and not to delete or ignore them. Therefore, due to the less involvement of the players during the gameplay and contradictory results of coping appraisal, the gameplay was altered.

A second large Dutch hospital was chosen to test the game again called Dutch Hospital 2. Six game participants were chosen from the same department in the hospital. Most of the participants knew other or were at least acquainted with each other. Like Gameplay 1, the participants became aware of the different threat signals used in the cards and resulted in increasing their threat appraisal. But unlike Gameplay 1, the coping appraisal increased among the participants of Gameplay II. This can be attributed to the change in

the setup of the game and the Game Master's emphasis on the importance of reporting suspicious cards so that the hospital's IT department. Both the response cost and the response efficacy, the two attributes of coping appraisal increased among the players because the participants were made to understand that the Game Master (hospital management) would recognize and reward them (appreciation cards) for correctly responding to the threat (by reporting the threat). In addition, the participants seemed to have a positive learning experience because they openly communicated and discussed their experiences during and after the gameplay.

To summarize, the Phish Phishy game has been able to explore the usage of the PMT and MINDSPACE frameworks in creating cybersecurity awareness. The following chapter validates Phish Phishy using validation criteria from gaming literature and expert interview.

Chapter 7: Validating the Awareness Intervention

Validation is the final step of the game design research approach. The concept of validity in relation to serious games as a simplified model of a complex reference system is hardly elaborated in the literature (Peters et al., 1998). A general understanding of the concept of validity in relation to serious games is the degree of correspondence between the reference system and the simulated model (Peters et al., 1998). Therefore, the aim of this chapter is to validate the research by establishing a degree of correspondence between the reference system (hospital environment with hospital staff) and the designed serious game (simulated model) called Phish Phishy. This chapter explains the two ways in which this research was validated, that is, using Game Design Research approach and interviewing a cybersecurity expert.

7.1 Four criteria for gaming validity

To be able to establish a correspondence between the reference system and the game, the four criteria for validity of gaming as suggested by Raser (1969) is used. This section explains the four-criteria game validity.

7.1.1 Psychological Reality

The first criterion for validity is psychological reality. The criterion suggests that “a game is said to be valid to the degree that it provides an environment that seems realistic to the players” (Raser, 1969, p.143). The players may behave differently than they would in real-life situations if they don't perceive the game to be genuine. The outcome would be that game behaviour of the players will not match that of the reference system’s behaviour. Phish Phishy supplements this criterion through the game cards. For both Dutch Hospital 1 and Dutch Hospital 2, the scenario cards were tailor-made to fit the email, SMS, and login website environments. This helped the players (hospital staff) to understand the different ways in which attackers could manipulate their work environment. Additionally, messages in the Appreciation and Depreciation cards were designed to help the players relate to the consequences (positive and negative) of their actions on hospital and patients.

7.1.2 Structural Validity

The second criterion is structural validity. The criterion suggests that “a game is said to be valid to the degree that its structure (the theory and assumptions on which it is built) can be shown to be similar to that of the reference system"(Raser, 1969, p.144). Phish Phishy is based on the combination of Protection Motivation Theory (PMT) and MINDSPACE frameworks. The pillars of this combination framework are threat appraisal, coping appraisal, and the MINDSPACE influencers. The game uses these elements in its cards to explore the effect on the players and this effect is observed during the game and recorded using the surveys. Messenger effect, incentives, group norms, priming, and ego effect are the common influencers used in all the cards to appraise threat severity, threat vulnerability, and coping ability (self-efficacy to report suspicious activities) among the players.

7.1.2 Process Validity

The third criterion is process validity. The criterion suggests that "a game is said to be valid to the degree that the processes observed in the game are similar to those observed in the reference system" (Raser, 1969, p.144). Serious gaming has not yet been adopted in hospitals as an awareness creation method and the

game design is not similar to the conventional awareness methods used in hospitals. So, process validity cannot be fully assessed for Phish Phishy given the target domain. For example, the Appreciation and Depreciation cards are introduced in the game is to motivate the players both intrinsically and extrinsically to actively partake in phishing awareness. But hospitals do not provide rewards or recognitions for identifying phishing signals but like other organisations may provide sanctions to the staff when led to a cyberattack. Therefore, a one-to-one correspondence cannot yet be deduced between hospital processes and game processes.

7.1.2 Predictive Validity

The fourth criterion is predictive validity. The last criterion suggests that "a game said to be valid to the degree that it can reproduce historical outcomes or predict the future" (Raser, 1969, p.144). However, the fourth criterion seems less important in serious games focused on improving knowledge or skills of the players because the desired outcome is already known (Peters et al., 1998). So, for serious games focused on improving knowledge or skills, the final validity criterion is “a game is said to be valid to the degree that the learning objectives are achieved by the participants” (Peters et al., 1998, p.4). Although the survey results from the gameplay suggest an increase in phishing awareness, the results are based on a small sample size and the longevity of the awareness levels cannot be predicted at this stage of the research.

7.2 Expert Interview

The validity criteria by Raser (Raser, 1969) suggests the theoretical aspects of validating the effectiveness of serious games. From a practical standpoint especially for understanding the effectiveness and applicability of Phish Phishy in a hospital, an expert interview was also conducted. The interviewee was Floris Duvekot, a cybersecurity consultant at Secura Consulting in The Netherlands. Floris’ consulting expertise are in managing security behaviour in client organizations (including hospitals), crisis management, and developing serious games for cybersecurity. The semi-structured interview was conducted via Microsoft Teams for 25-minutes. The research was presented first before delving into the validation. The main takeaway from the interview is highlighted in this section.

7.2.1 Game Design

“What is your opinion on the game design and its effectiveness*? Has what I aimed to do come through and has it been effective?”*

Floris was positive about the overall game design and the combination of behaviour change theories used to raise awareness among hospital staff. He suggested that customizing the cards makes it easily recognizable for the players and gives them the feeling of as though the signals are occurring in their workplace. He also found the game to be innovative due to the focus on behavioural aspects and rewards thereby having a positive impact on the success of the game. But he also critics the game design in the following ways:

1. Customization could take time for implementing it in different hospitals or even into different domains. But if the hours to develop the game cards is known upfront, then a delivery timeline for the game can be communicated to each hospital.

2. The game is perhaps designed to raise awareness for a short duration and more rounds must conducted in different intervals to assess a long-term learning effect in the participants. This is because when the participants return to their daily work life, the knowledge gained form this game could decrease.

3. In his experience, he has seen client organizations not keen on dramatizing the impact of a security negligence. In this game, he suggests reflecting on whether the depreciation cards dramatize the negative impact and that it may not be well-received by top management of organizations.

4. Currently, hospital staff email their IT department when they find suspicious activities whereas in the game the players simply pick the 'Report as suspicious' action card. He suggested that an easier solution to report suspicious activities should be a part of the work environment in hospitals before implementing the game to improve the similarity in workflows between the game and the hospital environment.

The interview with Floris Duvekot helped to think critically about the obtained results. Although the results suggest a positive impact on the awareness levels, it is possible that the positive results are due to a limited number of sample sizes. The interview also sheds light on the path towards future research. These aspects are addressed in section 8.4 and 8.5.

7.3 Conclusion of Chapter 7

The aim of chapter 7 was to validate the game design and its effectiveness on the participants (hospital staff). The findings of the validation are summarized to answer the sub-part of the third sub-research question.

3.1 How effective is the identified awareness intervention in making hospital staff aware of phishing signals?

The effectiveness of the game is validated using a two-way approach: Theoretical validation using Raser's (1969) four game validity criteria and practical validation by interviewing Floris Dukevot (an industry expert). Phish Phishy passes the first criteria of psychological reality because the scenario cards are customised for each hospital to help its participants relate to the game environment and transfer the learning from the game easily into their workplace. Floris Dukevot also found the customisation to be an important factor for making the game relatable to the work environment of the players. The game passes the second criteria of structural validity because it utilises the main pillars of the two frameworks, that is, threat appraisal, coping appraisal, and MINDSPACE influencers in the cards to increase awareness on phishing signals among the players. The third criterion is process validity, that is, the degree to which the processes observed in the game are similar to those observed in the reference system. A one-to-one correspondence cannot be deduced between the game and the hospitals because serious games have not yet gained popularity in hospitals. For example, hospitals do not have a process wherein they reward their staff for reporting phishing messages but the reward processes in the game are aimed at motivating the players to identify threats and report them. The last criterion is predictive validity, that is, the degree to which the learning objectives are achieved by the participants. Although the survey results from the gameplay suggest an increase in phishing awareness, the results are based on a small sample of players and the longevity of the awareness levels cannot be predicted at this stage of the research.

Chapter 8: Conclusion & Discussion

This research has focused on the identification of an awareness intervention in hospitals so that it can help hospital staff identify phishing signals and improve their self-efficacy to report the phishing signals. This chapter answers the main research question for the identified knowledge gap, discusses the limitations of this research, and concludes by providing the direction for future research.

8.1 Layered Knowledge Gap

Two levels of knowledge gaps were identified on security awareness in hospitals. The first level identified that there is a lack of effective security awareness education methods for hospital staff. The existing education methods are either organization-wide security campaigns or conventional training methods such as seminars, presentations, or online trainings. But these methods are not effective because hospital staff work in high stress environments and often suffer workload fatigue, so their involvement in conventional training methods cannot be guaranteed. In addition, these methods do not address the challenges that prevent hospital staff from identifying phishing signals or suspicious activities. Therefore, an effective awareness intervention is required to address this is a chicken-and-egg problem. The second knowledge gap that was found was on the lack of standard frameworks available to design an effective cybersecurity intervention that addresses these challenges. Recent research suggests that combining Protection Motivation Theory (PMT) and MINDSPACE frameworks can be effective in designing cybersecurity interventions, but there are no existing interventions that has operationally tested this research. This knowledge gap is addressed in the following section by answering the main research question.

8.2 Main Findings

The knowledge gap was addressed by answering the sub-research questions in the previous chapters. The aggregated information is used in this section to answer the main research question:

How can hospital staff be made aware of phishing signals in the work environment to prevent ransomware attacks on hospitals?

A Game Design Research Approach is used to answer the main research question. The first knowledge gap was addressed by recognizing the challenges that prevent hospital staff from identifying phishing signals. Although literature in this field is scarce, a review was conducted to identify which challenges arise from two areas: challenges due to the hospital environment and challenges due to individual's capacity. The challenges from the hospital environment are due to three reasons. Firstly, the high stress environment causes fatigue from the workload and results in less security cautiousness in hospital staff. And the hospital environment prioritizes patient care so lesser attention is focused towards securing IT systems. Secondly, there is an increase in the interconnectivity in hospitals due to remote working and Bring-your-own-device (BYOD) concept. However, security policies around flexible working is lacking, and hospital staff are not aware of the safe security behaviour that they must follow in the new working environment. Lastly, there is a lack of effective awareness trainings on phishing in hospitals. This is because there is a lack of cybersecurity experts in hospitals who can adequately train the hospital staff and due to the high employee turnover in

hospitals, wherein the latter disrupts the continuous flow of hospital specific awareness trainings. The challenges from an individual's capacity hospital staff are due to the inability to identify phishing signals. They arise due to their (hospital staffs') susceptibility to 'principles of influence', high gullibility, and low cue utilization. This is because firstly, hackers exploit phishing emails using similarity to real-world scenarios so the quick mode in the human cognition is unable to differentiate the message's legitimacy. Secondly, hackers target anxious situations such as COVID-19 pandemic to gull users into clicking on phishing signals. Lastly, hackers can succeed with persuasion techniques such as similarity, authority, and urgency. These techniques are effective on hospital staff with high workload and anxiousness because the cognition does not process the messages or cues in entirety and utilizes the quick mode of human cognition to respond to the signal with less mental effort.

These challenges are addressed using serious games as an awareness creation method since. This choice was made because serious games are argued to be more engaging, stimulate self-efficacy, self-assessment, and collaboration in the players than conventional methods of creating awareness (seminars, online courses, or presentations) (Chowdhury et al., 2022). To design an effective serious game, the combination of PMT and MINDSPACE frameworks suggested by Briggs (2017) was used. The serious game was designed as a simple tabletop card game called Phish Phishy. The combination of frameworks was used in the design of the game to make the players aware of phishing threats and how to cope with those threats. To increase the threat awareness and coping awareness among the players, MINDSPACE influencers such as messenger effect, incentives, group norms, priming, affect, and ego effect are used in the game cards. The players are intrinsically motivated using messages that emphasize the impact of the risk in the Appreciation cards and Depreciation cards. The players are extrinsically motivated using positive reward points for correctly identifying and reporting phishing signals and negative reward points for failing to do so. The game uses real-world work context in the scenario cards to make the game relatable for its players. Two gameplay sessions were conducted in two large academic hospitals in The Netherlands. Based on the game design, the results from the game survey suggest an increase in the awareness levels, that is, improved threat appraisal and improved ability to cope with threat (see section 6.2.5). Therefore, the use of the PMT and MINDSPACE framework combination suggested by Briggs (2017) was explored for the first time through the serious game, Phish Phishy, to make hospital staff aware of phishing signals in the work environment and report them.

8.3 Limitations of the Research

Like every research, this research has its limitations, and they are discussed in this section to guide towards future research.

8.3.1 Game Design

Firstly, the game design process is based on the Game Design Research Approach by Kurapati et al. (2017). A game design is an iterative process to make the game as effective as possible after every validation and testing step (Kurapati et al., 2017). But in this research, only one iteration of feedback could be implemented, that is, the game design was adjusted after Gameplay 1 in Dutch Hospital 1 due to time constraints to complete this research. More iterations of gameplay in more groups of players can support in further validating the game design and survey results. Secondly, this research is only focused on phishing signals

does not cover other types of social engineering threats (e.g., phone scams or vishing). Thirdly, the survey results of the game suggest a positive impact on the awareness levels of the game participants. However, the game design does not account for a stress factor which could duplicate workload stress that the participants would otherwise have faced in their work environment while assessing emails, SMSs, or websites. This could be why the survey results suggested an improvement in their awareness levels. Lastly, 'Think like a hacker' was the strategy used to design the game environment to place the phishing signals in the scenario cards. Based on the meetings with the employees of the two hospitals, only three work contexts were chosen to be contents of the scenario cards, that is, user interface of emails, SMSs, and Dutch Hospitals' employee log-in environments. However, using more common interfaces could strengthen the transferability of the learnings from the game. Moreover, hackers could use phishing signals in more sophisticated ways which could not be covered in this design. Only straightforward phishing signals were used based on web search.

8.3.2 Target Audience

The sample set of players were different in Gameplay 1 and Gameplay 2. The observations during Gameplay 1 suggested that the players were reserved and less involved in the game compared to players in Gameplay 2. Gameplay 1 had participants from different departments of the hospital whereas Gameplay 2 had participants from the same department which could be why the participants of Gameplay 2 seemed more involved in the game and their survey results suggested relatively improved awareness levels. However, more gameplay sessions are required with the same sample sets as Gameplay 1 (heterogenous) and Gameplay 2 (homogenous) to be able to deduce more information on the relation.

8.3.3 Serious Games

During the validation interview with Floris Duvekot, he suggested that hospital staff could be resistant to innovative corporate trainings. This is a limitation of serious games because everyone may not enjoy playing games, so a lack of interest can negatively affect the learnings from the game. Additionally, Phish Phishy requires a group of maximum 6 players and a gameplay session will require coordinating the agendas of the hospital staff, so it is not possible to get trained at any time. Hospital management must consider alternate methods to deliver security awareness trainings using similar constructs of the Phish Phishy game for those who are less interested in serious games.

8.4 Direction for Future Research

This research only considers two samples and one iteration in the game design. Future research should focus on conducting more gameplay sessions with more samples to adjust the game design and to validate the survey results. Thereafter, based on the outcome of the results, long-term effect of the game on awareness can be evaluated by verifying the behaviour of the participants after the gameplay either through real-life observations, surveys, or interviews. Knowing the long-term effect can also help gauge whether the frequency of the game sessions should be increased or decreased, and it can also be adopted to other domains (e.g., aviation industry) for creating awareness on phishing signals. Future research should also focus on including a stress factor in the game design to make the game as relatable as possible for the players so that the results of the game can be a predict the validity in a real-life scenario.

Lastly, this research focuses on using the combination of PMT and MINDSPACE frameworks suggested by Briggs (2017) in designing Phish Phishy since there is a lack of standard frameworks for designing awareness trainings. Future research should focus on using the combination of PMT and MINDSPACE frameworks in the design of new serious games and other types of awareness methods such as conventional, online software-based, videos, and/or simulation-based trainings. If the usage of this combination on large number of samples shows improvement in the awareness levels of the hospital staff, then the combination of PMT and MINDSPACE frameworks has the potential to be considered as an industry standard for designing hospital-wide cybersecurity awareness games.

References

(2020). SecurityWeek – *A Wired Business Media Publication*. https://www.securityweek.com/emerging-threats-during-times-crisis-insights-airbus-cybersecuritys-phil-jones

$21B. (2021, March 26). *Fierce Healthcare*. 2020 offered a "perfect storm" for cybercriminals with ransomware attacks costing the industry

Abawajy, J. (2014). User preference of cyber security awareness delivery methods. *Behaviour & Information Technology*, 33(3), 237–248. https://doi.org/10.1080/0144929X.2012.708787

Akpan, A. (2016). Has health care hacking become an epidemic? *The Public Broadcasting Service*. Retrieved from https://www.pbs.org/newshour/science/has-health-care-hacking-become-an-epidemic

Al-Darwish, A. I., & Choe, P. (2019, July). A framework of information security integrated with human factors. In *International Conference on Human-Computer Interaction* (pp. 217-229). Springer, Cham.

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: Methods and development*. Allyn & Bacon.

Allodi, L. (2019.). A.v.d.heijden@student.tue.nl Eindhoven University of Technology. 19.

Alzahrani, A. (2020). Coronavirus Social Engineering Attacks: Issues and Recommendations. *International Journal of Advanced Computer Science and Applications*, 11(5). https://doi.org/10.14569/IJACSA.2020.0110523

Amorim, J. A., Hendrix, M., Andler, S. F., & Gustavsson, P. M. (2013). *Gamified Training for Cyber Defence*. 12.

Anderson & Agarwal. (2010). Practicing Safe Computing: A Multimethod Empirical Examination of Home Computer User Security Behavioural Intentions. *MIS Quarterly*, *34*(3), 613. https://doi.org/10.2307/25750694

Bada, M., Sasse, A. M., & Nurse, J. R. C. (n.d.). *Cyber Security Awareness Campaigns: Why do they fail to change behaviour?* 11.

Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioural change. *Advances in Behaviour Research and Therapy, 1(4), 139-161.*

Bayl-Smith, P., Sturman, D., & Wiggins, M. (2020). Cue Utilization, Phishing Feature and Phishing Email Detection. In M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne, & M. Sala (Eds.), Financial Cryptography and Data Security (Vol. 12063, pp. 56–70). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-54455-3_5

BBC News. https://www.bbc.com/news/technology-54204356

Beckers, K., & Pape, S. (2016). A serious game for eliciting social engineering security requirements. *In 2016 IEEE 24th International Requirements Engineering Conference (RE) (Pp. 16-25). IEEE.*

Beresford, A. R., Rice, A., Skehin, N., & Sohan, R. (2011). MockDroid: Trading privacy for application functionality on smartphones. *Proceedings of the 12th workshop on mobile computing systems and applications* (pp. 49–54).

Blythe, J. M., Coventry, L., & Little, L. (n.d.). *Unpacking security policy compliance: The motivators and barriers of employees' security behaviours*. 20.

Burnkrant, E. R. (1978). "Cue Utilization in Product Perception", in NA - Advances in Consumer Research Volume 05, eds. Kent Hunt, Ann Abor, MI : *Association for Consumer Research*, Pages: 724-729.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116–131. Https://doi.org/10.1037/0022-3514.42.1.116.

Chen, T., Hammer, J., & Dabbish, L. (2019). Self-Efficacy-Based Game Design to Encourage Security Behaviour Online. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–6. https://doi.org/10.1145/3290607.3312935

Chen, T., Stewart, M., Bai, Z., Chen, E., Dabbish, L., & Hammer, J. (2020, July). Hacked Time: Design and Evaluation of a Self-Efficacy Based Cybersecurity Game. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1737-1749).

Chen, T., Stewart, M., Bai, Z., Chen, E., Dabbish, L., & Hammer, J. (2020). Hacked Time: Design and Evaluation of a Self-Efficacy Based Cybersecurity Game. *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 1737–1749. https://doi.org/10.1145/3357236.3395522

Chou, F. K.-Y., Chen, A. P.-S., & Lo, V. C.-L. (2021). Mindless Response or Mindful Interpretation: Examining the Effect of Message Influence on Phishing Susceptibility. Sustainability, 13(4), 1651. https://doi.org/10.3390/su13041651

Chowdhury, N., & Gkioulos, V. (2021). Cyber security training for critical infrastructure protection: A literature review. *Computer Science Review*, *40*, 100361. https://doi.org/10.1016/j.cosrev.2021.100361

Chowdhury, N., Katsikas, S., & Gkioulos, V. (2022). Modeling effective cybersecurity training frameworks: A delphi method-based study. *Computers & Security*, *113*, 102551. https://doi.org/10.1016/j.cose.2021.102551

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

Coventry, L., Branley-Bell, D., Sillence, E., Magalini, S., Mari, P., Magkanaraki, A., & Anastasopoulou, K. (2020). Cyber-risk in healthcare: Exploring facilitators and barriers to secure behaviour. *In International Conference on Human-Computer Interaction* (pp. 105-122). Springer, Cham.

de Bruijn, H., & Janssen, M. (2017). Building cybersecurity awareness: The need for evidence-based framing strategies. *Government Information Quarterly*, 34(1), 1-7.

Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational technology research and development*, *53*(2), 67-83.

Dictionary of Academic English at OxfordLearnersDictionaries.com. (2022). Oxford English Dictionary.https://www.oxfordlearnersdictionaries.com/definition/academic/awareness?q=awareness

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*(1), 264–277. https://doi.org/10.1016/j.joep.2011.10.009

Duke, R. D., & Geurts, J. (2004). Policy games for strategic management. Rozenberg Publishers.*Emerging Threats During Times of Crisis: Insights from Airbus Cybersecurity's Phil Jones.SecurityWeek.Com*.

Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012). Android permissions: User attention, comprehension, and behaviour. *Proceedings of the 8th symposium on useable privacy and security*.

George, M. S., Teunisse, A. K., & Case, T. I. (2020). Gotcha! Behavioural validation of the Gullibility Scale. *Personality and Individual Differences*, 162, 110034. https://doi.org/10.1016/j.paid.2020.110034

Georgiadou, A., Michalitsi-Psarrou, A., Gioulekas, F., Stamatiadis, E., Tzikas, A., Gounaris, K., Doukas, G., Ntanos, C., Landeiro Ribeiro, L., & Askounis, D. (2021). Hospitals' Cybersecurity Culture during the COVID-19 Crisis. Healthcare, 9(10), 1335. https://doi.org/10.3390/healthcare9101335

Goethals, P. L. (2019). Insider Attack Metrics for Cybersecurity: *Investigating Various Research Options*.7.

Gordon, W. J., Wright, A., Aiyagari, R., Corbo, L., Glynn, R. J., Kadakia, J., Kufahl, J., Mazzone, C., Noga, J., Parkulo, M., Sanford, B., Scheib, P., & Landman, A. B. (2019). Assessment of Employee Susceptibility to Phishing Attacks at US Health Care Institutions. *JAMA Network Open*, 2(3), e190393. https://doi.org/10.1001/jamanetworkopen.2019.0393

Gordon, W. J., Wright, A., Glynn, R. J., Kadakia, J., Mazzone, C., Leinbach, E., & Landman, A. (2019). Evaluation of a mandatory phishing training program for high-risk employees at a US healthcare system. *Journal of the American Medical Informatics Association*, *26*(6), 547–552. https://doi.org/10.1093/jamia/ocz005

Haggman, A. (2019). *Cyber wargaming: Finding, designing, and playing wargames for cyber security education*.

Hart, S., Margheri, A., Paci, F., & Sassone, V. (2020). Riskio: A Serious Game for Cyber Security Awareness and Education. *Computers & Security*, *95*, 101827. https://doi.org/10.1016/j.cose.2020.101827

He, W., & Zhang, Z. (Justin). (2019). Enterprise cybersecurity training and awareness programs: Recommendations for success. *Journal of Organizational Computing and Electronic Commerce*, *29*(4), 249–257. https://doi.org/10.1080/10919392.2019.1611528

He, Y., Aliyu, A., Evans, M., & Luo, C. (2021). Health Care Cybersecurity Challenges and Solutions Under the Climate of COVID-19: Scoping Review. *Journal of Medical Internet Research*, 23(4), e21747. https://doi.org/10.2196/21747

Hendrix, M., Al-Sherbaz, A., & Victoria, B. (2016). Game based cyber security training: are serious games suitable for cyber security training? *International Journal of Serious Games*, *3*(1).https://www.fiercehealthcare.com/tech/ransomware-attacks-cost-healthcare-industry-21b-2020-here-s-how-many-attacks-hit-providers

Jin, G., Tu, M., Kim, T. H., Heffron, J., & White, J. (2018). Evaluation of game-based learning in cybersecurity education for high school students. *Journal of Education and Learning (EduLearn), 12(1), 150-158.*

Jones, E. (2022). *What Happens if You Click on a Phishing Link? [Plus, Ways to Prevent it] Warren Averett CPAs & Advisors*. https://warrenaverett.com/insights/what-happens-if-you-click-on-a-phishing-link/

Kahneman, D. (2011). Thinking, Fast and Slow; Farrar, Straus and Giroux: New York, NY, USA, 2011.

Klabbers, J. H. (2009). The magic circle: Principles of gaming & simulation. In *The Magic Circle: Principles of Gaming & Simulation*. Brill.

Klopfer, E., Osterweil, S., & Salen, K. (2009). Moving learning games forward. *Cambridge, MA: The Education Arcade*.

Kurapati, S. (2017). *Situation Awareness for Socio Technical Systems: A simulation gaming study in intermodal transport operations*. TRAIL.

Leach, J. (2003). Improving user security behaviour. *Computers & Security*, *22*(8), 685–692. https://doi.org/10.1016/S0167-4048(03)00007-5

Lins de Holanda Coelho, G., H. P. Hanel, P., & J. Wolf, L. (2020). The Very Efficient Assessment of Need for Cognition: *Developing a Six-Item Version. Assessment*, 27(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Lukosch, H. K., Bekebrede, G., Kurapati, S., & Lukosch, S. G. (2018). A scientific foundation of simulation games for the analysis and design of complex systems. *Simulation & gaming*, *49*(3), 279-314.

Mahat, J., Alias, N., & Yusop, F. D. (2022). Systematic literature review on gamified professional training among employees. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2022.2043910

Malone, T. W., & Lepper, M. R. (2021). Making learning fun: A taxonomy of intrinsic motivations for learning. *In Aptitude, Learning, and Instruction (Pp. 223-254). Routledge.*

McAfee 2021. *Hacking the skills shortage.* [online]. Available: https://www.mcafee.com/enterprise/en-us/assets/reports/rp-hacking-skills-shortage.pdf.

Mitnick, K. D., & Simon, W. L. (2003). The art of deception: Controlling the human element of security. John Wiley & Sons.

Morar Consulting (2016). The dangers of our digital livesTech. Rep. Retrieved from https://slidex.tips/download/the-dangers-of-our-digital-lives

Nasser, G., Morrison, B. W., Bayl-Smith, P., Taib, R., Gayed, M., & Wiggins, M. W. (2020). The Effects of Cue Utilization and Cognitive Load in the Detection of Phishing Emails. In M. Bernhard, A. Bracciali, L. J. Camp, S. Matsuo, A. Maurushat, P. B. Rønne, & M. Sala (Eds.), Financial Cryptography and Data Security (Vol. 12063, pp. 47–55). *Springer International Publishing*. https://doi.org/10.1007/978-3-030-54455-3_4

Ndibwile, J. D., Luhanga, E. T., Fall, D., Miyamoto, D., Blanc, G., & Kadobayashi, Y. (2019). An Empirical Approach to Phishing Countermeasures Through Smart Glasses and Validation Agents. *IEEE Access*, 7, 130758–130771. https://doi.org/10.1109/ACCESS.2019.2940669

Nifakos, S., Chandramouli, K., Nikolaou, C. K., Papachristou, P., Koch, S., Panaousis, E., & Bonacina, S. (2021). Influence of Human Factors on Cyber Security within Healthcare Organisations: A Systematic Review. *Sensors*, 21(15), 5119. https://doi.org/10.3390/s21155119

Peltonen, M. (1989). Management in the 1990's. Aavaranta Serie. No. 14. Otava. Keuruu, Finland.

Perry, William E., (1985): Management Strategies for Computer Security. Butterworth Publisher, Boston

Peters, V., Vissers, G., & Heijne, G. (1998). The Validity of Games. *Simulation & Gaming*, *29*(1), 20–30. https://doi.org/10.1177/1046878198291003

Pike, S., Kelledy, M., & Gelnaw, A. (2017). Measuring U.S. privacy sentiment: An IDC special report (Tech. Rep.).

Polanyi, M. (2009). The tacit dimension. In Knowledge in organizations (pp. 135-146). Routledge.

Potter, B. (2009). Microsoft SDL threat modelling tool. Network Security, 2009(1), 15-18.

Rainie, L., & Duggan, M. (2014). Heartbleed's Impact. *Pew Research Center, Report.* http://www.pewinternet.org

Ramkumar, N., Kothari, V., Mills, C., Koppel, R., Blythe, J., Smith, S., & Kun, A. L. (2020). Eyes on URLs: Relating Visual Behaviour to Safety Decisions. *ACM Symposium on Eye Tracking Research and Applications*, 1–10. https://doi.org/10.1145/3379155.3391328

Raser, J. R. (1969). *Simulation and society: An exploration of scientific gaming.*

Rippetoe, P. A., & Rogers, R. W. (1987). Effects of components of protection-motivation theory on adaptive and maladaptive coping with a health threat. *Journal of personality and social psychology*, *52*(3), 596.

Rizzoni, F., Magalini, S., Casaroli, A., Mari, P., Dixon, M., & Coventry, L. (2022). Phishing simulation exercise in a large hospital: A case study. DIGITAL HEALTH, 8, 205520762210817. https://doi.org/10.1177/20552076221081716

Rogers, R. W. (1975). A Protection Motivation Theory of Fear Appeals and Attitude Change1. *The Journal of Psychology*, *91*(1), 93–114. https://doi.org/10.1080/00223980.1975.9915803

Rolloff, M. (2010). A constructivist model for teaching evidence-based practice. *Nursing Education Perspectives, 31(5), 290-293.*

S. Huntley, "Findings on covid-19 and online security threats", 2020, [online] Available: https://www.blog.google/threat-analysis-group/findings-covid-19-and-online-security-threats/

Shahri, A. B., Ismail, Z., & Rahim, N. Z. (2013). Security culture and security awareness as the basic factors for security effectiveness in health information systems. *Jurnal Teknologi*, 64(2), 7–12. https://doi.org/10.11113/jt.v64.2212

Shen, L. (2014). The NIST cybersecurity framework: Overview and potential impacts. *Scitech Lawyer*, *10*(4), 16.

Shostack, A. (2014). Elevation of privilege: Drawing developers into threat modeling. In *2014 USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14)*.

Spok. (2018). The Leader in Clinical Communication Solutions. 2018. 10 Facts About BYOD: *Healthcare Secure Text Messaging*.

Spruit, M. (2010). Bewust veilig? [Consciously secure?]. *De IT-Auditor*, 4, 15–21. https://www.deitauditor.nl/wp-content/uploads/2014/09/bewust-veilig.pdf

Teunisse, A. K., Case, T. I, Fitness, J, & Sweller, N. (2020). I should have known better: Development of a self-report measure of gullibility. *Personality and Social Psychology Bulletin*, 46(3), 408–423. Https://doi.org/10.1177/0146167219858641.

Tidy, B. J. (2020, September 18). *Police launch homicide inquiry after German hospital hack*.

Trickel, E., Disperati, F., Gustafson, E., Kalantari, F., Mabey, M., Tiwari, N., ... & Vigna, G. (2017). Shell We Play A Game?{CTF-as-a-service} for Security Education*. In 2017 USENIX Workshop on Advances in Security Education (ASE 17)*.

Wani, T. A., Mendoza, A., & Gray, K. (2020). Hospital Bring-Your-Own-Device Security Challenges and Solutions: Systematic Review of Gray Literature. *JMIR MHealth and UHealth*, 8(6), e18175. https://doi.org/10.2196/18175

Weirich, D., & Sasse, M. A. (n.d.). *Pretty Good Persuasion: A First Step towards Effective Password Security in the Real World*. 7.

Woon, I. M. Y., Tan, G. W., & Low, R. T. (2005). A protection motivation theory approach to home wireless security. *In: Proceedings of the Twenty-Sixth International Conference on Information Systems. Las Vegas, NV.*, 367–380.

Yasin, A., Liu, L., Li, T., Fatima, R., & Jianmin, W. (2019). *Improving software security awareness using a serious game.*

Appendix A: Search Methodology

A.1 Literature Review Identifying Challenges at an Organizational Level

In this section, the literature research method that is used to select the literature on phishing awareness in the healthcare sector is described in the following three phases of searching, screening, and selection.

A.1.1 Search Strategy

The research methodology began by reviewing literature on phishing awareness in the healthcare sector. This was carried out by searching through the Scopus database using keyword search in the article title, abstract, and keywords. 'Phishing awareness hospital', 'phishing awareness challenge hospital', and 'cybersecurity awareness challenge hospital' are the keywords that populated relevant articles. These keywords did not include any Boolean terms such as 'AND or 'OR' between them. Each keyword search queries resulted in 10, 1, and 7 articles, respectively. The duplicate articles were removed from the results to obtain 13 articles. For the exact search query methodology, see *figure A1*.

A.1.2. Screening for Relevant Literature

An initial screening was performed to filter out papers that did not include the theme of phishing awareness in hospitals and healthcare sector. This narrowed the sample size to seven. The time-period was also a selection criterion, that is, articles were filtered for year of publishing ranging 2019 and above, this reduced the sample size to 6. The journals that these articles belonged to were once more screened for in Scopus separately using the same keywords as used above. However, no new articles pertaining to phishing awareness in hospitals were found in these sources.

A.1.3. Selection of Relevant Literature

The selection criterion was primarily dependent on the literature available on phishing awareness to understand the different categories of challenges that hospital staff face. The 6 articles were further analyzed on the basis of any implicit or explicit mentioning of 'phishing awareness in hospitals' and 'cybersecurity awareness in hospitals''. These terms were explicitly found in all the chosen articles. The literature review, therefore, focuses on these 6 articles. A list of the selected articles and their themes are summarized in Table 2.

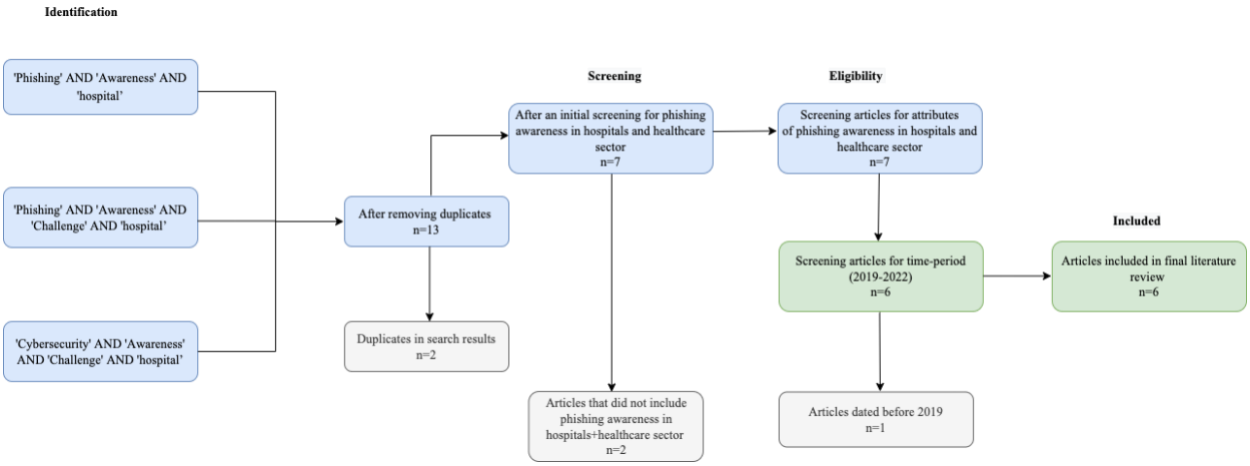


Figure A.1 Search Methodology used for finding challenges from organizational level

A.2 Literature Review identifying challenges at an individual level

In this section, the literature research method that was used for selecting the literature on user behaviour towards phishing signals is described in the following three phases of searching, screening, and selection.

A.2.1 Search Strategy

The research methodology began by reviewing literature on the human behaviour towards phishing . This was carried out by searching through the Scopus database using keyword search in the article title, abstract, and keywords. 'Phishing psychology' and 'cue phishing' are the keywords that populated relevant articles. The keywords did not include any Boolean terms such as 'AND or 'OR' between them. Each keyword search queries resulted in 52 and 59 articles, respectively. The duplicate articles were removed from the results to obtain 78 articles.

A.2.2. Screening for Relevant Literature

An initial screening was performed to filter out papers that did not include the theme of phishing awareness challenges at an individual or user level. This narrowed the sample size to 31 articles. The time-period was also a selection criterion, that is, articles were filtered for year of publishing ranging 2019 and above. This reduced the sample size to 9 articles. The journals that these articles belonged to were once more screened for in Scopus separately using the same keywords as used above. However, no new articles pertaining to human behaviour towards phishing were found in these sources.

A.3.3. Selection of Relevant Literature

The selection criterion was primarily dependent on the literature available on human behaviour towards phishing to understand the psychological challenges that users face when it comes to identifying phishing signals. The 9 articles were further analyzed on the basis of any implicit or explicit mentioning of 'human behaviour towards phishing' and 'challenges to identify cues in phishing'. This theme was implicitly found in all the chosen articles. The literature review, therefore, focuses on these 9 articles. A list of the selected articles and their themes are summarized in Table 3.

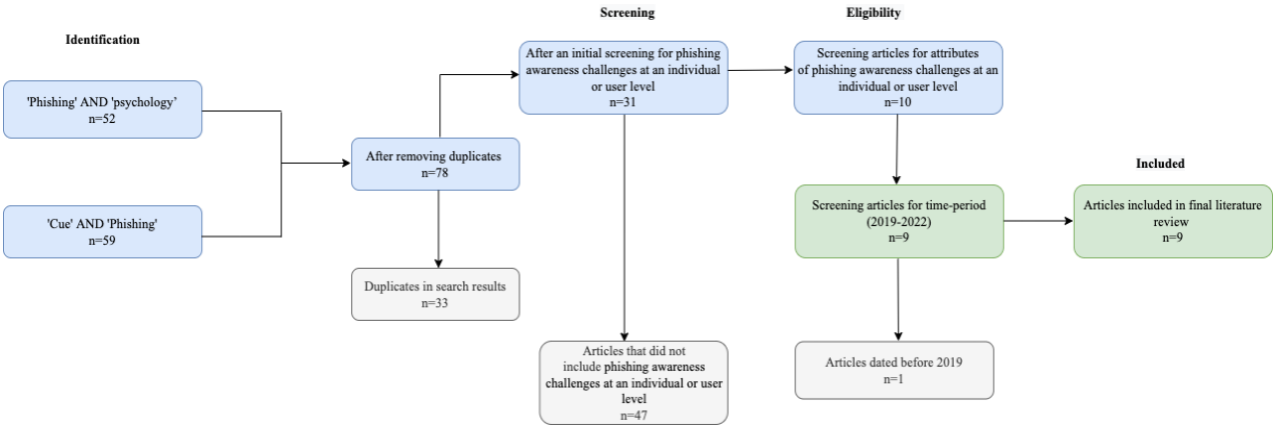
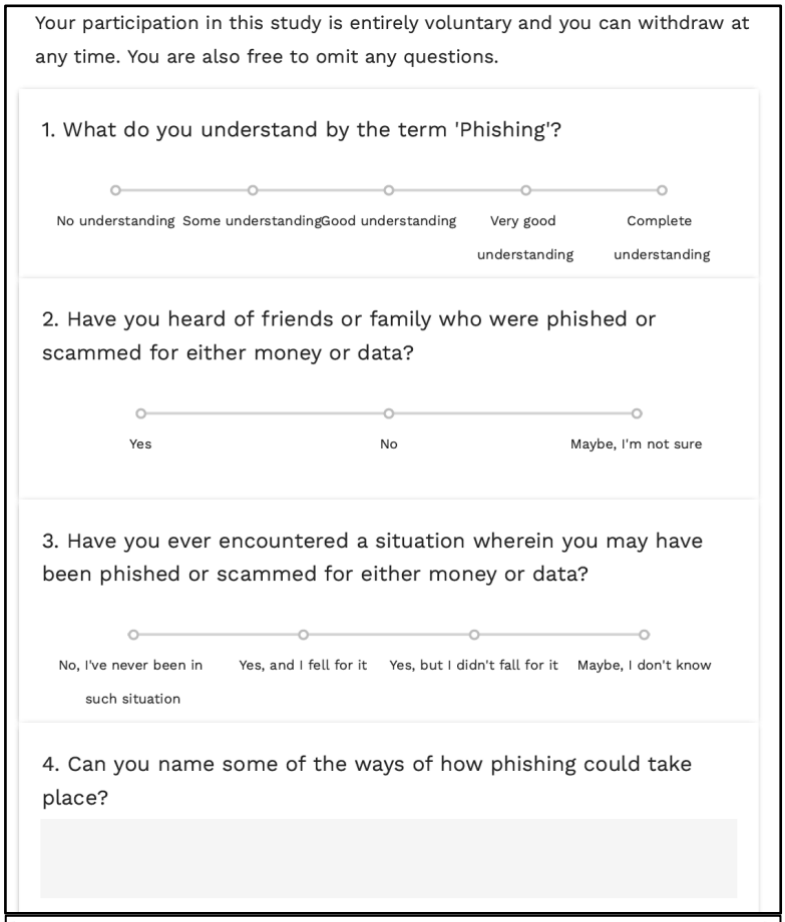


Figure A.2 Search Methodology used for finding challenges from individual's level

Appendix B: Artefacts from Gameplay 1 & 2

B1. Pre-survey Questions



Your participation in this study is entirely voluntary and you can withdraw at any time. You are also free to omit any questions.

1. What do you understand by the term 'Phishing'?

No understanding | Some understanding | Good understanding | Very good understanding | Complete understanding

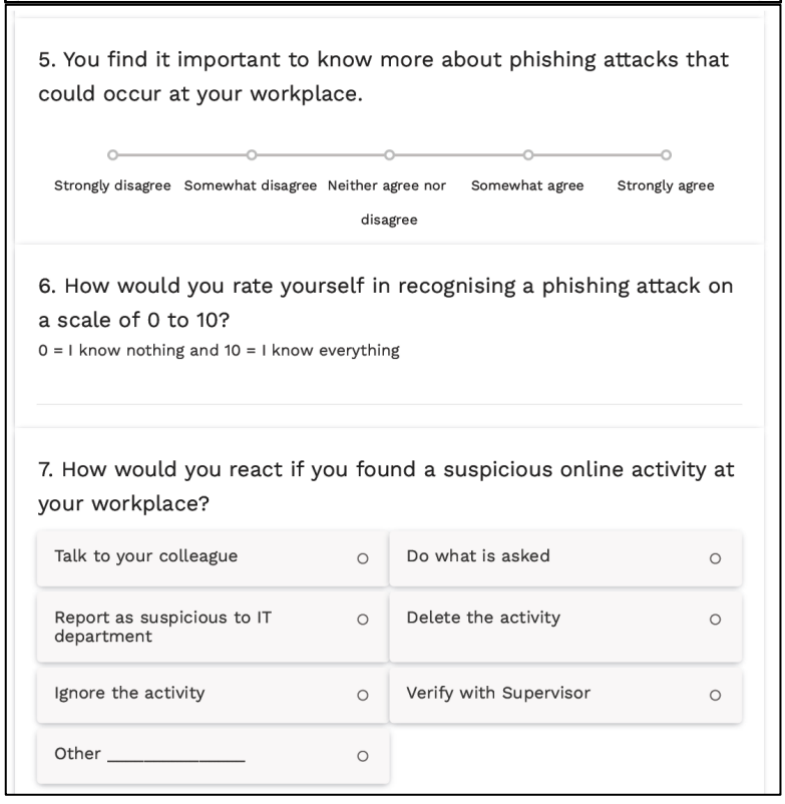
2. Have you heard of friends or family who were phished or scammed for either money or data?

Yes | No | Maybe, I'm not sure

3. Have you ever encountered a situation wherein you may have been phished or scammed for either money or data?

No, I've never been in such situation | Yes, and I fell for it | Yes, but I didn't fall for it | Maybe, I don't know

4. Can you name some of the ways of how phishing could take place?



5. You find it important to know more about phishing attacks that could occur at your workplace.

Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree

6. How would you rate yourself in recognising a phishing attack on a scale of 0 to 10?

0 = I know nothing and 10 = I know everything

7. How would you react if you found a suspicious online activity at your workplace?

- Talk to your colleague
- Do what is asked
- Report as suspicious to IT department
- Delete the activity
- Ignore the activity
- Verify with Supervisor
- Other _____

B2. Post-survey Questions

Your participation in this study is entirely voluntary and you can withdraw at any time. You are also free to omit any questions.

1. After playing the game, what is your understanding of the term 'Phishing'?

No understanding | Some understanding | Good understanding | Very good understanding | Complete understanding

2. After playing the game, how alert would you be while assessing Emails, SMSs, and websites?

More alert than before | As alert as before | Less alert than before

3. Now, how would you rate yourself in recognising a phishing attack on a scale of 0 to 10?

0 = I know nothing and 10 = I know everything

4. Now, you find it important to know more about phishing attacks that could occur at your workplace.

Strongly disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Strongly agree

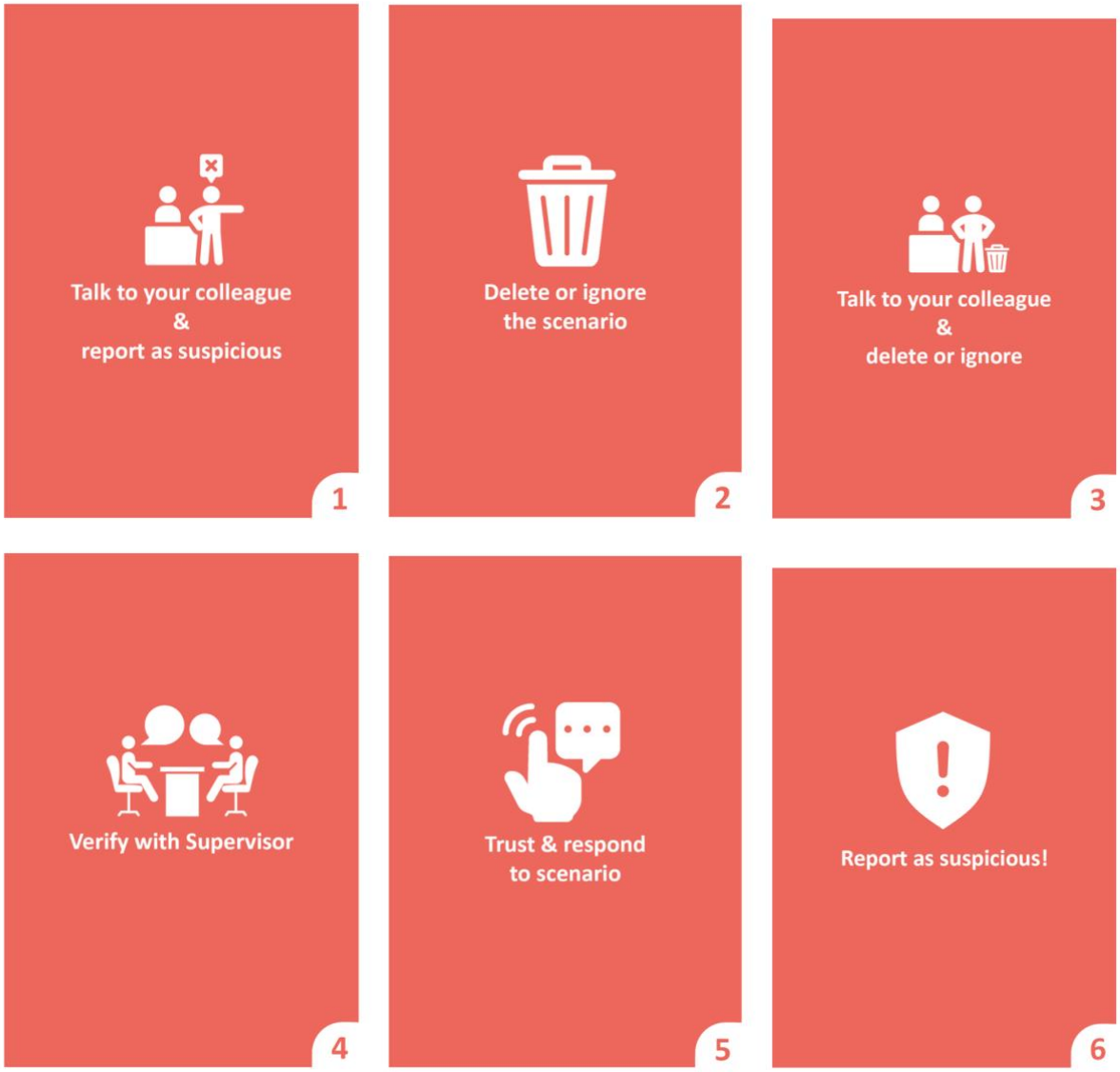
5. Now, how would you react if you found a suspicious online activity at your workplace?

- Talk to your colleague
- Do what is asked
- Report as suspicious to IT department
- Delete the activity
- Ignore the activity
- Verify with Supervisor
- Other _____

6. General comments

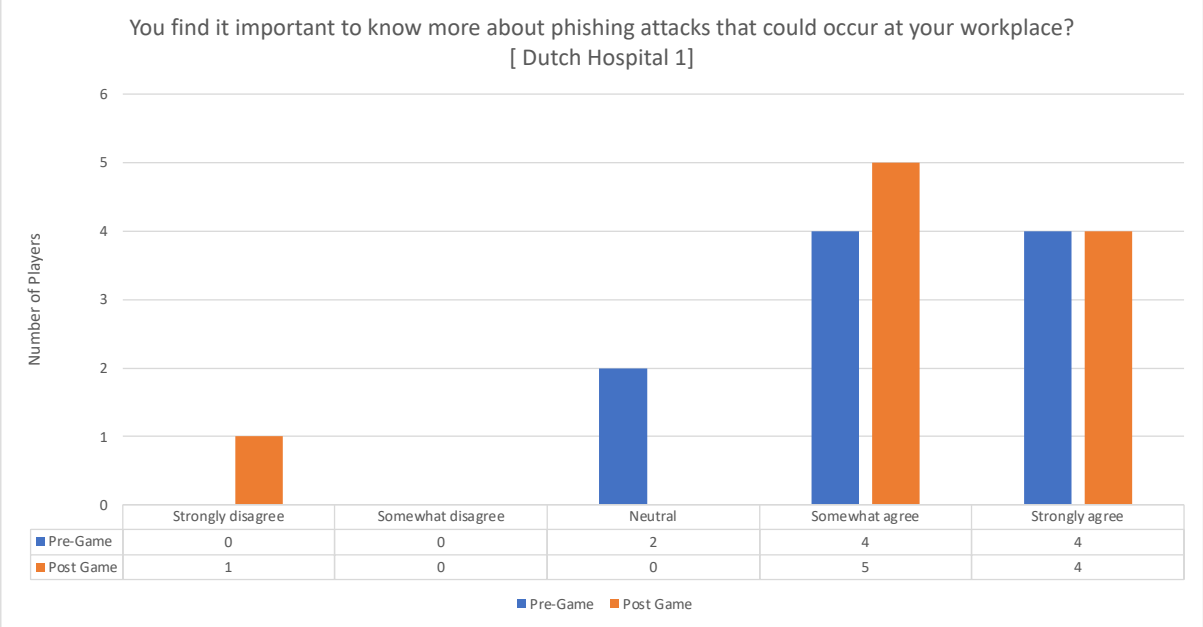
Feel free to comment on your game experience.

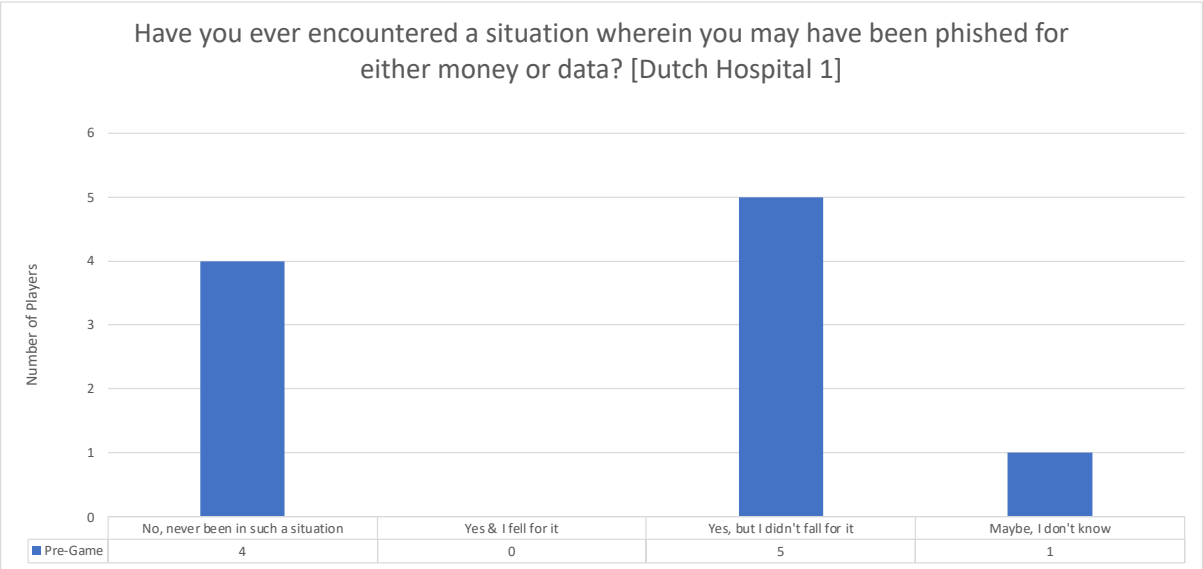
B.3 Action Cards Used in Gameplay 1 & 2



Appendix C: Survey Results of Gameplay 1 & 2

C.1 Results of Gameplay 1





C.2 Results of Gameplay 2