The disPHISHinformation Game: Creating a Serious Game to Fight Phishing Using Blended Design Approaches

Niklas Henderson¹, Helen Pallett¹, Sander van der Linden², Jake Montanarini³, and Oliver Buckley⁴

¹University of East Anglia, Norwich, Norfolk, UK

²University of Cambridge, Cambridge, UK

³Norwich University of the Arts, Norwich, Norfolk, UK

⁴Loughborough University, Loughborough, Leicestershire, UK

ABSTRACT

In 2022, 39% of all UK businesses reported identifying a cyber security attack against their own organisation, 83% of which were phishing attempts. A large body of research in cyber security focuses on technical solutions, however humans remain one of the most exploitable endpoints in an organisation. Traditional security training within organisations commonly includes point-and-click exercises and simple video media that employees are required to complete. These training exercises are often seen as unengaging and tedious, and employees are commonly pushed to complete training rather than encouraged to learn and self-educate. Simulations and games are increasingly being deployed for training purposes in organisations, however often either (a) simply raise cyber security awareness rather than deliver key security policy and content, or (b) lack accessibility with complex game pieces and rules not easily understandable by those not accustomed to playing games. We introduce the dis-PHISHinformation game: a customisable serious game to deliver phishing training specific to the threats businesses face on a day-to-day basis. Drawing on existing taxonomies, the game delivers content on email, voice, and SMS social engineering attacks, in a format that educates players in key social engineering features. In collaboration with a large service organisation, we have also developed a customised edition of disPHISHinformation game which reflects the targeted attacks faced by their staff. By creating an analog serious game to deliver key phishing training, we aim to stimulate higher employee engagement and deliver a more memorable experience.

Keywords: Serious games, Transformational games, Educational games, Gamification, Cyber security, Game design, Phishing, Inoculation theory

INTRODUCTION

Cyber security is now a key focus for organisations around the world, equally important to small and medium-sized enterprises (SMEs) as it is to large enterprises. For businesses, cybersecurity attacks risk affecting the confidentiality, availability, and integrity of data, impacting company finances and customer trust, and in the case of cyber-physical systems, threatening human life. For governments, attacks can affect sovereign structures and even influence democratic processes, a highly publicised issue in the 2016 United States presidential election. There is a large body of research focusing on technical approaches to defending against cyber security attacks, however comparatively little work attempts to combat the manipulative processes in phishing attacks. In 2022, 83% of reported cyber security attacks on UK businesses were phishing attacks (DCMS, 2022). Although technical solutions exist, these are imperfect and are not always able to filter out all messages. It is logical that the human aspects of cyber security are therefore becoming an area of great interest. By educating end-users, cyber security skills can be taught that are organisation-agnostic, and can help defend from continuously evolving threat actors. Cyber security training, often with a focus on phishing, is now commonplace in organisations. Conventional training commonly includes point-and-click exercises and simple video media that employees are required to complete, or simulations that often only test skill. These training exercises are often seen as unengaging and tedious however, and employees are commonly pushed to complete training efficiently rather than encouraged to learn, reflect, and self-educate (Proofpoint, 2023). New and innovative approaches are beginning to be explored, including the integration of serious games in this training context.

Serious Games

Serious games (also known as applied or transformational games), the focus of a burgeoning field at the intersection of gaming and education, have gained popularity for their ability to leverage the engaging and motivational nature of games for educational purposes. The boundaries between games, serious games, and gamification are often blurred, however although traditional games may *incidentally* lead to learning, serious games differ in the overall aim of the game designer(s): to educate. As the field continues to evolve, there is room to explore and connect game design methodologies to the realm of serious games. Various design methodologies have emerged, all providing different approaches to the design of serious games. When serious games are applied to another academic research area however, the game design process is rarely discussed. Games are increasingly being used in industry as a tool to increase engagement with company-specific training, particularly security policies (Proofpoint, 2023). Despite this uptake, businesses investing in 'gamification' often opt for in-house platforms over games proposed through research.

Cyber Security Serious Games

Within this related works section, we look at both digital and analog cyber security serious games, developed from both academia and industry.

Two freely accessible digital cyber security serious games are CyberCIEGE and Anti-phishing Phil. Developed by the US Naval Postgraduate School in 2005, CyberCIEGE teaches players information assurance concepts (Irvine, Thompson, and Allen, 2005). Players manage a virtual organisation with a

budget to spend on security measures, being tested with a number of scenarios. CyberCIEGE was designed with high customisability and adaptability in mind and is often cited as one of the early serious games in this area. Developed in 2007 (Sheng et al., 2007), Anti-Phishing Phil educates and encourages good habits resisting URL spoofing in phishing attacks. Players play as a fish in the 'Interweb Bay', while URLs are displayed on-screen. Players must eat the real worms (real URLs) and avoid the fake worms (phishing URLs). Anti-Phishing Phil is unique within this area as it has a higher emphasis on play.

There are several existing analog games aimed at improving cyber knowledge in organisations. Some games have been designed to be played alongside the development of a real-world system, often with developers and extended development teams in mind as target audiences. Through their designed mechanics, these games become entirely customised to the business context in which they are played. Elevation of Privilege, OWASP Cornucopia, and Protection Poker are examples of this. Developed in 2014 by Microsoft (Shostack, 2014), Elevation of Privilege (EoP) is a card game designed to help developers learn and execute software-centric threat modelling. Players play cards (based on the threat modelling STRIDE framework) based on whether they are relevant to the real-life system they are threat modelling. Designed in 2015 by the OWASP foundation, OWASP Cornucopia is heavily inspired by EoP. In contrast to EoP, cards are designed to represent security vulnerabilities compiled from the OWASP Application Security Verification Standard (rather than the STRIDE framework). Protection Poker (Williams, Gegick, and Meneely, 2009) is based on Wideband Delphi and Planning-Poker Agile Methods.

Some analog cyber security games integrate more 'play' elements and include many different game pieces. Play2Prepare (Graffer, Line, and Bernsmed, 2015) is designed to trigger discussions and knowledge exchange to strengthen incident response capabilities. Playing different roles with unique skills, players collaboratively move game pieces around a board to deal with attacks on their power grid. [d0x3d!] (Gondree and Peterson, 2013) is designed using pedagogical methods to expose young people and students to topics in cyber security. Control-Alt-Hack (Denning et al., 2013) is also loosely targeted at students with the aim of raising security awareness and change perceptions. In Control-Alt-Hack, players work for a fictitious white-hat penetration testing company. Players assume roles and use their character's skills to complete missions.

Several analog cyber security card games have more recently been developed. Riskio (Hart et al., 2020) aims to build knowledge in non-technical players within an organisation context. One player acts as the attacker with other players acting as defenders, working collaboratively to pick appropriate defence cards. Decisions & Disruptions (D&D) (Frey et al., 2019) educates players on cyber security decision making and prioritisation. Players play as security decision-makers for a utility company over 4 rounds. Each round, players decide what security defence to invest in with their limited budget. A Lego game board represents the players' facility and office, however the authors have played D&D solely with cards with little issue. D&D has become particularly popular within cyber security circles, even being explored by the Metropolitan Police's Cyber Protect team (Metropolitan Police, 2023).

Digital serious games currently receive a comparatively high focus over analog games in industry as they are easily rolled out within an organisation. This is particularly true in large-scale enterprises, where employees may work remotely and are therefore not always physically present. Serious game designers have suggested however that analog games may stimulate better learning over digital games. It is argued that analog games can encourage higher engagement, and in-person discussion and discourse can improve knowledge exchange between players (Graffer, Line, and Bernsmed, 2015; Denning et al., 2013). Determining and designing an appropriate balance between eudaimonic and hedonic enjoyment (educational and fun aspects of the playing experience) is also an important choice for serious games, and the target playing environment is an essential aspect of this (Chen et al., 2021). EoP, OWASP Cornucopia and Protection Poker have proven successful despite less integration of play mechanics. Being played as part of a business process in a supervised environment may partly subsidise much of the player motivation and engagement traits traditionally provided by play mechanics. In contrast, Play2Prepare, [d0x3d!], and Control-Alt-Hack favour increased play mechanics. This has increased their suitability to unsupervised environments and could result in higher replayability, however do not necessarily deliver as high a eudaimonic experience. Game complexity and cost are also important considerations, and if too high can introduce difficulties in game accessibility for players who are not familiar with games, or be a barrier for organisations looking to introduce gamified training.

Contribution

In this paper we propose the disPHISHinformation game, an analog serious card game designed to educate players on phishing attacks through email, SMS, and voice. Players play as employees of a fictitious advertisement business, discerning between malicious and legitimate messages delivered on cards. The development of the disPHISHinformation game is discussed with particular effort to highlight the design approach taken. The disPHISHinformation game represents a novel contribution to the field as the first phishing analog serious game as an inoculation theory intervention. At the time of writing the disPHISHinformation game has not been tested, however we hope that future work will further investigate the strengths and weaknesses of such an approach.

DISPHISHINFORMATION GAME DESIGN APPROACH

We have used a blend of game design practices from academia related to serious games and misinformation and have taken inspiration from related work in this area as discussed. The game mechanics and rules are discussed at the end of this paper.

Inoculation Theory

Originally conceptualised in 1964 by McGuire, inoculation theory represents a pre-emptive strategy to defending against counterattitudinal attacks. The theory follows the biological analogy: much like a vaccine, pre-exposing a participant to a small, altered snippet of a persuasive argument can increase their resistance to future persuasion. Inoculation interventions have traditionally been built from two key components: *threat*, and *counterarguing* (Compton and Pfau, 2005). Threat represents the participant's knowledge that an existing attitude is not immune from attack, and therefore the "shock value" of a forewarning provides motivation for the required work (McGuire, 1961). Counterarguing describes the 'weakened dose' attack arguments presented to the participant, who is required to refute the attack through examination of argument refutations (a passive defence), or to generate their own counterarguments (an active defence). Inoculation theory has been found in meta-analyses to be superior to supportive attitude-bolstering and reactive approaches (Banas and Rains, 2010).

Inoculation theory has more recently been explored as an approach to protect participants against a different type of persuasion: misinformation. Although broadly successful, most of these interventions have historically focussed on a single topic, such as climate change (van der Linden et al., 2017). This has raised questions of scalability, especially considering the constant ebb and flow of online misinformation topics (Traberg, Roozenbeek, and van der Linden, 2022). Crucially, research has found that counterarguments do not need to cover all attack arguments, and therefore an inoculation intervention can provide an "umbrella protection" within the issue domain (Parker, Rains, and Ivanov, 2016; Parker, Ivanov, and Compton, 2012). Going beyond the vaccine analogy, inoculation theory has also been found to not only protect existing attitudes, but positively affect participants with no (or a contrasting) pre-existing attitude (Ivanov et al., 2017). More recently, inoculation theory has been explored in combination with active learning in the form of serious games to create a number of misinformation games (Basol et al., 2021; Roozenbeek and van der Linden, 2020). These games have broadly moved away from inoculating on specific misinformation topics, and focus instead on misinformation techniques (Roozenbeek and van der Linden, 2019). Active and passive gamified interventions have broadly been interpreted as generative or reading tasks respectively, however this translation into game design decisions is not always consistent (Grace and Liang, 2023). Some games translate active refutation as players playing as a malicious actor generating and disseminating disinformation, although this is not always the case (Kiili, Siuko, and Ninaus, 2024).

To take advantage of the discussion and discourse available in analog games, the disPHISHinformation game is designed to be multiplayer in nature. To avoid malicious interplayer competitiveness, players will passively read and review phishing messages rather than actively attack other players with generated content. The game also follows the broader interpretation of inoculation theory by focusing on phishing techniques rather than a single attitude, thus taking advantage of inoculation theory's umbrella protection for in-domain attacks. Within the context of gamified interventions against misinformation, a potential concern is that although an intervention may reduce trust in misinformation, trust in genuine sources could also be eroded (Modirrousta-Galian and Higham, 2023). This consideration has directly informed the game's design.

Serious Game Design Approach

In the landscape of serious game development, various design methodologies are followed to design the gameplay experiences of these educational tools, however this is rarely discussed in the literature. It is important to acknowledge and outline the methodologies used in the development of a game, as doing so may begin to reveal their strengths and weaknesses. Working within the structure of an accessible and simple analog card game, as well as a technique-based active inoculation intervention, the Transformational Framework was agreed to be the main design process within this project. The Transformational Framework is a requirement-gathering approach, which prioritizes the alignment of game objectives with broader learning or behavioural goals within serious games. This pre-production-based methodology integrates explorative and reflective question-based processes for a better structured and complete serious game development approach. Importantly, the Transformational Framework involves collaboration with experts in relevant fields to ensure the accuracy and efficacy of the game's content (Culvba, 2018). This enables the development of a more carefully curated experience that balances educational content with entertainment elements to maintain player interest.

Crowdsourcing Playtesting

Within applied serious game literature, especially that of the games parallel with the disPHISHinformation game, it is uncommon for the game design process to receive much focus, particularly playtesting. Playtesting sessions give an opportunity for game designers to better understand the 'play' of a game, how the educational elements are being delivered to players, and whether the desired balance between 'play' and learning has been achieved. The research team have decided to crowdsource playtesting of the disPHISH-information game to allow a diverse group of organisations to inform future iterations of the disPHISHinformation game. A playtest document is provided alongside the game materials to incentivise game and learning feedback from players. The research team hope that this will allow development of future prototypes that meet the requirements of organisations looking to introduce gamified security training into their ecosystem.

THE DISPHISHINFORMATION GAME: RULES

When receiving a malicious phishing attack message, the context of topic, personal information, and indirectly associated information (e.g. workplace) communicated are important to help identify it as malicious (NCSC, 2017). In the disPHISHinformation game, players play as employees of the fictitious

company 'Creative Ads', working within a project management department. Players are also informed that Creative Ads uses the creativeads.co.uk domain. This added context enables the gamification of more context-specific phishing attacks. The context card can be seen in Figure 1. The disPHISHinformation game is comprised of this context card, and source and action cards.



Figure 1: The game context card and forewarning (www.disphishinformation.org).

Source Cards

The core part of the disPHISHinformation game are source cards, which represent messages players receive while working at 'Creative Ads'. The game design is centred around phishing attacks that are sent to players. A variety of different characteristics and manipulative techniques (Aleroud and Zhou, 2017; Rastenis et al., 2020; Alabdan, 2020) are featured in source cards, imitating content commonly observed by SMEs. The techniques and structures used in source cards are categorised in Figure 2. Through the lens of inoculation theory, the malicious source card content represents the 'weakened dose' message.

Attack	Engagement	Sender's	Attack Type
Platforms	Motivation	Email/Number	Behavioural
Phishing	Benefit for recipient	Owned	Shortened URL
Smishing	"Legitimate" request	Camouflaged	Bad domain name
Vishing	Important information	Hacked	Hostname obfuscation
	Possible failure	legitimate	Encoded URL
			obfuscation
			Malicious attachment

Figure 2: Catalogue of phishing attack features from which source cards are based.

There are 50 source cards that feature phishing, smishing, and vishing attacks (see Figure 3). These cards feature a message on the rear that identify it as a 'phish' (malicious), or real. If the card is a 'phish', a refutational message is included, detailing how the player could have identified it as malicious. There are 50 source cards that feature legitimate messages that are themed similarly to the malicious source cards. The disPHISHinformation

game includes the legitimate source cards to focus on improving player skill in differentiating between real and malicious messages.



Figure 3: Source cards of the disPHISHinformation game. Shown above is message content (left) and (on rear of card, right) if that message is malicious, or real.

Action Cards

This initial prototype of the disPHISHinformation game uses action cards to be a catalyst for creating 'play', with care not to compromise the educational content of the source cards. Action cards have been designed to foster humorous interactions between players, as well as supporting discussion and discourse between players on source cards. Example action cards allow players to skip their turn, pass their source card to another player, ask another player for advice, or ask all other players for advice. Two examples can be seen in Figure 4.



Figure 4: Two examples of disPHISHinformation action cards.

Gameplay

Before the game starts, Source and Action cards are shuffled together into a single deck. The deck is then placed between all players content-up, such that the content of the source/action card faces up, and the source card answers (real or "phish") face down. Players are each given a context card for reference which they keep during play (non-playable). On a player's turn, they read the card at the top of the deck. If it is an action card, they can take this

card and keep it for future use. The player then decides if the source card is genuine or malicious. Once they have come to a decision, they declare this to the group and turn the card over. If correct, the player gains the point.

Customisation

Despite the wide array of cyber security serious games within industry and academia, many large-scale organisations either continue to use traditional training approaches or develop gamified approaches in-house. As well as a natural resistance to adopting experimental approaches, medium and large-scale enterprises typically have security policies specific to their organisation. This requirement for organisation-specific training content can reduce appetite for adopting serious games that have no flexibility to be customised. The disPHISHinformation game has been designed with customisability in mind for this reason. Blank source cards have been provided alongside the original content, and users are able to edit these and use content more specific to their organisation. With support from Aviva, a unique edition of the disPHISHinformation game has additionally been created that is customised to facilitate the training content of a large business. This includes additional attack vectors including QR codes and system notifications, as well as supplychain-specific content. This customised example of the game is provided as an example of how the disPHISHinformation game can be applied to cover organisation-specific training content.

CONCLUSION

Within our broad cyber security landscape, humans remain the weak point in a mostly technological chain. Traditional training exercises including pointand-click tasks and video media are often ineffective and unmemorable, leading to an invisible security vulnerability for organisations. Although gamified approaches are increasingly being introduced in business settings, current solutions from academia and industry can be inaccessible (high cost, complicated rules) or simply raise security awareness rather than deliver key educational material. Developed using blended approaches from serious game design, inoculation theory and a crowdsourced playtesting approach we introduce the disPHISHinformation game. The disPHISHinformation game is designed to educate players in phishing techniques and designed to stimulate discussion between players in an organisation context. The dis-PHISHinformation game is also designed for potential customisability to medium- and large-scale organisations.

REFERENCES

- Alabdan, R. (2020). Phishing attacks survey: Types, vectors, and technical approaches. Future internet, 12(10), 168.
- Aleroud, A., & Zhou, L. (2017). Phishing environments, techniques, and countermeasures: A survey. Computers & Security, 68, 160–196.
- Banas, J. A., & Rains, S. A. (2010). A meta-analysis of research on inoculation theory. Communication Monographs, 77(3), 281–311.

- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & Linden, S. van der. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. Big Data and Society, 8(1).
- Chen, V., H. H., Yu, V., Koek, D., W. J., & Ho, J., S. T. (2021). Balancing Fun and Seriousness Serious: Game Design Considerations. TMS Proceedings 2021.
- Compton, J. A., & Pfau, M. (2005). Inoculation Theory of Resistance to Influence at Maturity: Recent Progress In Theory Development and Application and Suggestions for Future Research. Annals of the International Communication Association, 29(1), 97–146.
- Culyba, S. (2018). The Transformational Framework, A process tool for the development of transformational games. ETC Press, Carnegie Mellon University.
- DCMS (July 11, 2022) Cyber Security Breaches Survey 2022: https://www.gov.uk/ government/statistics/cyber-security-breaches-survey-2022/cyber-securitybreaches-survey-2022 (Accessed 25/01/2024).
- Denning, T., Lerner, A., Shostack, A., & Kohno, T. (2013). Control-Alt-Hack: The design and evaluation of a card game for computer security awareness and education. In Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security (CCS'13).
- Frey, S., Rashid, A., Anthonysamy, P., Pinto-Albuquerque, M., & Naqvi, S. A. (2019). The Good, the Bad and the Ugly: A Study of Security Decisions in a Cyber-Physical Systems Game. IEEE Transactions on Software Engineering, 45(5), 521–536.
- Gondree, M., & Peterson, Z. N. J. (2013). Valuing Security by Getting [d0x3d!] Experiences with a network security board game. 6th Workshop on Cyber Security Experimentation and Test, CSET 2013.
- Grace, L., & Liang, S. (2023). Examining Misinformation and Disinformation Games Through Inoculation Theory and Transportation Theory.
- Graffer, I., Line, M. B., & Bernsmed, K. (2015). Play2Prepare: A Board Game Supporting IT Security Preparedness Exercises for Industrial Control Organizations.
- Hart, S., Margheri, A., Paci, F., & Sassone, V. (2020). Riskio: A Serious Game for Cyber Security Awareness and Education. Computers & Security, 95, 101827.
- Irvine, C. E., Thompson, M. F., & Allen, K. (2005). CyberCIEGETM: An Information Assurance Teaching Tool for Training and Awareness. In Federal information systems security educators' association conference, North Bethesda, MD.
- Ivanov, B., Rains, S. A., Geegan, S. A., Vos, S. C., Haarstad, N. D., & Parker, K. A. (2017). Beyond simple inoculation: Examining the persuasive value of inoculation for audiences with initially neutral or opposing attitudes. Western Journal of Communication, 81(1), 105–126.
- Kiili, K., Siuko, J., & Ninaus, M. (2024). Tackling misinformation with games: a systematic literature review. Interactive Learning Environments, 1–16.
- McGuire, W. J. (1961). The effectiveness of supportive and refutational defenses in immunizing and restoring beliefs against persuasion. Sociometry, 24(2), 184–197.
- McGuire, W. J. (1964). Some Contemporary Approaches. Advances in Experimental Social Psychology, 1(C), 191–229.
- Metropolitan Police (2023). Cyber Protect: how we can help your business: https://www.met.police.uk/cyberprotect (Accessed 25/01/2024).
- Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. Journal of Experimental Psychology: General, 152(9), 2411.

- NCSC (October 10, 2017) Guidance; Small Business Guide: Cyber Security; Step 5 Avoiding phishing attacks: https://www.ncsc.gov.uk/collection/small-business-gu ide/avoiding-phishing-attacks (Accessed 25/01/2024).
- OWASP. OWASP Cornucopia. https://www.owasp.org/index.php/OWASP_Cornuc opia (Accessed 25/01/2024).
- Parker, K. A., Ivanov, B., & Compton, J. (2012). Inoculation's efficacy with young adults' risky behaviors: can inoculation confer cross-protection over related but untreated issues?. Health communication, 27(3), 223–233.
- Parker, K. A., Rains, S. A., & Ivanov, B. (2016). Examining the "blanket of protection" conferred by inoculation: The effects of inoculation messages on the cross-protection of related attitudes. Communication Monographs, 83(1), 49–68.
- Proofpoint (2023). 2023 State of the Phish Report: An in-depth exploration of user awareness, vulnerability and resilience: https://www.proofpoint.com/uk/resource s/ threat-reports/state-of-phish (Accessed 25/01/2024).
- Rastenis, J., Ramanauskaitė, S., Janulevičius, J., Čenys, A., Slotkienė, A., & Pakrijauskas, K. (2020). E-mail-based phishing attack taxonomy. Applied Sciences, 10(7), 2363.
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. Palgrave Communications, 5(1), 1–10.
- Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that "inoculates" against political misinformation. The Harvard Kennedy School Misinformation Review.
- Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-Phishing Phil: The design and evaluation of a game that teaches people not to fall for phish. Proceedings of the 3rd symposium on Usable privacy and security (pp. 88–99).
- Shostack, A. (2014). Elevation of privilege: Drawing developers into threat modeling. 2014 USENIX Summit on Gaming, Games, and Gamification in Security Education.
- Traberg, C. S., Roozenbeek, J., & van der Linden, S. (2022). Psychological inoculation against misinformation: Current evidence and future directions. The ANNALS of the American Academy of Political and Social Science, 700(1), 136–151.
- van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. Global challenges, 1(2), 1600008.
- Williams, L., Gegick, M., & Meneely, A. (2009). Protection Poker: Structuring software security risk assessment and knowledge transfer. IEEE Security & Privacy, vol. 8, no. 3, pp. 14–20.