Cognitive Deception: Cognitive Hacking Phenomenon and Cognitive Security

Mikko Luomala and Jyri Naarmala
University of Vaasa, Finland

mikko.luomala@uwasa.fi
jyri.naarmala@uwasa.fi

Abstract: Cybercriminality is an actively discussed topic in in criminology. While technical cybersecurity has evolved, increasing number of different types of technical solutions have been created to defend against cyberattacks. In security and cybersecurity, the weakest link is often considered to be a human. It seems that while malware attacks are increasing in numbers, also social engineering or cognitive hacking has increased. This paper explores the use and definition of a concept of cognitive deception in previous studies. Used dataset for this study was taken from SCOPUS database and qualitative meta-analysis (QMA) was used for analysing the content of the documents. The QMA method was selected for its suitability in content analysis of scientific-, as well as potential grey literature. Evolving technology has created new approaches for manipulating people. The analysis of research data indicates that cognitive hacking relates to definitions of cybercrime, cyberattack, warfare, and hybrid threats. This study contributes to the discussion about cognitive security.

Keywords: Cybercrime, Criminology, Cognitive deception, Cognitive hacking, Cyberattack, QMA, Qualitative meta-analysis

1. Introduction

Within cyber security, cognitive hacking has been studied (Adefabi et al. (2025). The consequences cognitive hacking has for example on human users who are being targeted to give unauthorized access to computer systems has been studied as well (Cybenko et al., 2002). This type of cyberattacks can be called social engineering where human weaknesses are being exploited (Montañez et al., 2020). Social engineering is used to gain access to confidential information (Washo, 2021). Concept of cognitive hacking can be expanded with the aid of artificial intelligence to deceive people. Cognitive hacking is a form of cyberattack, where misinformation is used to manipulate the opinions of the public in different societal aspects. This is done using psychological means to exploit human weaknesses in cognitive function. In this process, cognitive hacking refers to false information to manipulate the target using cyber-social vulnerability in the information systems. It seems, that technological approach and layer-based thinking in security or cybersecurity could greatly benefit from extending those to include controls for cognitive hacking. In this sense, previous studies have their focus on describing "what is occurring" in social interaction and what this manipulative action causes to the technology, i.e. "description of damages". For example, in a Finnish study about cybercrimes that end up in court, cases are categorized as a regular usage of computers, and not technologically that advanced (Luomala, 2025). This indicates that there is an emerging trend taking place, where the preparation of cybercrimes is shifting from purely technological approach towards more socio-technical approach, where cognitive hacking plays an important role. So, how can this evolutionary transformation be explained? This indicates that there exists a research gap, that is waiting to be found.

Layered security architecture has been seen as a methodology to secure assets (Ghasemi & Babaie, 2024). However, when layered security is penetrated with a human element (Jamuna, 2024) or an inside threat that emerges within the organisation (Anti, 2025; Anti & Vartiainen, 2024), this results a violation of the security (or cybersecurity). Even though the Department of Defence has defined layered strategy to suppress and deter threats from as far away as possible and the layers plays a key element in this defence (Lewis, 2006, p. 5). The layered security model is used to secure for example "Supervisory Control and Data Acquisition" (SCADA) systems, where defence is based on different layers (Knapp and Langill, 2015, p. 32-33). The model is also called as the defence in depth (Caballero, 2014; Morrow, 2021; Shinder & Shinder, 2005). However, it seems that the idea of layered security has a fundamental design flaw. A great amount of research has demonstrated that criminal behaviour is quite common (Howitt, 2006, p. 17), and the "threat" exists already within the system, in the very environment where humans live.

Some studies propose that there is a scam for everyone (Whitty, 2020). However, cyberthreats are created by humans (Lewis, 2006, p. 399), although there is a hypothetical possibility that in the near future artificial intelligence (AI) can generate cyberthreats or cybercrimes (Luomala, 2025). This brings up several important questions for scientific investigation. It could be asked, why are humans considered the weakest link in security and cybersecurity? How is cognitive security defined ontologically? What happens when humans are being manipulated, and do they recognise that they are being manipulated? In what mental stage human is vulnerable

to manipulation? How does manipulation take place? How to defend against cognitive hacking, and what can be done to either prevent, detect or mitigate the effects of cognitive hacking? This paper analyses why theoretically systems are never technically 100% secure (Lewis, 2006, p. 431). In order to understand this problem domain, a research problem is presented for this study: (*RQ) What is the role of cognitive security in the layered security model?*

2. Research Design

For creating a pre-theory for the research question, a qualitative meta-analysis (QMA) is used. QMA is known under three different names, and these are qualitative meta-synthesis (Berente et al., 2019; Habersang et al., 2019; Küberling-Jost, 2021), meta-synthesis (Hoon, 2013; Omanović & Langley, 2023) or qualitative meta-study (Lee et al., 2015; Tamminen & Holt, 2010). In 2024, Ignacio and Schwerer stated that the QMA can be used to examine existing knowledge for searching for new insights. Mengist et al. (2020) suggest that QMA can be used to make interpretations from literature review. According to Noah (2017), the QMA allows usage of secondary data (p. 203) and statistical analysis (p. 196-197). Hagan (2018) describes that researcher can collect field notes (p. 166) and those data can be used in the QMA, because the QMA allows usage different types of data (Noah, 2017). The QMA methodology is not about making simple summarising or cataloguing or review of already published research (Noah, 2017, p. 197). Purpose is to create a new knowledge from existing data (Noah, 2017, p. 197). The QMA methodology is about using original text to discover themes and identifying terms, and from the text interpretations and reflections are made (Noah, 2017, p. 202). “Part of the function of meta-analysis is to correct for differences between studies that are the result of bias, and to highlight differences that are unexplainable by the theory or framework being used” (Weed, 2005, p. 9). Same principle can be applied in qualitative meta-analysis (QMA) (Weed, 2005, p. 9). However, the quality of the QMA relies on researchers’ ability to reflect, make an interpretation from research data and to make conclusions and present results (Noah, 2017, p. 203). The QMA is considered a rigorous research method (Levitt, 2018; Timulak, 2009). The QMA is an inductive research method, and it was chosen for this inductive study. The QMA is used to make a qualitative interpretation from the literature review (Noah, 2017). The interpretation will then be the pre-theory, which is the answer to the research question. This study presents the role cognitive security has in the layered security model and what kind of pre-theory can describe it. In each section the use of methodology is explained in detail.

3. Literature Review with PRISMA Framework

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) literature review quality system (Page et al., 2021) was used to answer the research question. The data collection, data screening and data processing is described according to the PRISMA framework (Page et al., 2021). Used search parameter for the SCOPUS database was “TITLE-ABS-KEY ("cognitive hacking")”. The search provided 25 documents in 8.12.2025. Documents were 21 full-text articles, 3 abstracts and 1 review version of an article. Nineteen documents (n=19) were included in the research. Figure 1 illustrates the detailed process.

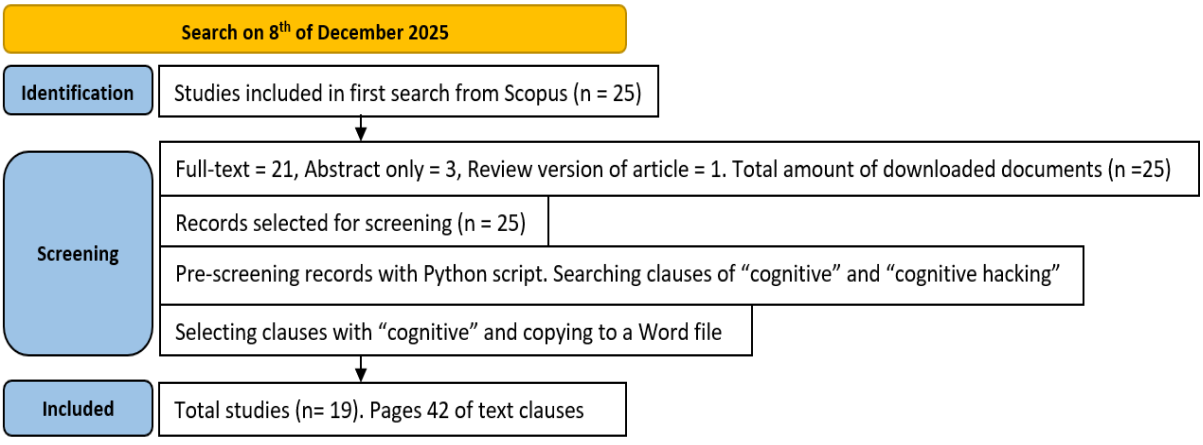

Figure 1: The PRISMA framework for collecting and processing the literature

For converting PDF files to text files, a Python code was created. The python code was run on Windows environment, and a folder was selected where the PDF files were located. The python saved the converted PDF files to the output folder. Another Python code was used for data mining and to collect clauses from the text files. The python code has a language model to detect words from clauses, and it was used to analyse all text

files in the output folder. If the lines of text file contained a word that matched search criterion, it was selected and stored into the output file. The output file had arrays, where filename was located and the collected clauses for each filename.

Sentences with words “cognitive” or “cognitive hacking” were recognized from the text files using a language model (Spacy with "en_core_web_sm"). Recognized clauses were then collected into an Excel file, and finally the output file was transferred to a Word file that made NVivo data coding possible. While analysing the data, it soon became clear, that the clauses of “cognitive” can provide a much richer view for the study, than just selecting clauses which have only “cognitive hacking” words. From there, iteration of the data was repeated. Research process continued by reading the clauses from the Word file. Depraetere et al. (2021) suggest that when critical interpretation synthesis (CIS) is used and iteration process is conducted, themes will emerge from the research data. The CIS methodology created main themes, which were: (i) Theme: attacking the mind; (ii) Theme: definition of cybercrime, cyberattack, hybrid threat, and; (iii) Theme: protecting the mind from cognitive hacking

4. Qualitative Meta-analysis Interpretation for the Proposed Pre-theory of Cognitive Hacking Theory

The layered security model can be defined in several different ways. The layered security is about creating architecture to secure assets with equipment on each layer (Ghasemi & Babaie, 2024). The layered security model includes usage of firewall, intrusion prevention systems and physical security solutions (Knapp and Langill, 2015, p. 32-33). The Department of Defence has defined layered strategy to suppress and deter threats from as far away possible and the layers plays a key element in this defence (Lewis, 2006, p. 5). The layered security model is used to secure for example SCADA (Supervisory Control And Data Acquisition) systems, where defence is based on different layers (Knapp and Langill, 2015, p. 32-33). The model is also called as the defence in depth (Caballero, 2014; Morrow, 2021; Shinder & Shinder, 2005). The purpose of the layered security model is to use technology to create security on each layer. However, when layered security is penetrated with a human element (Jamuna, 2024) or an inside threat which emerges within the organisation (Anti, 2025; Anti & Vartiainen, 2024). The cybersecurity is promised and unwanted event will occur, even it should be technically secure or otherwise it was belief that system is secure.

The three main themes have several sub-themes (Figure 2, Table 1). The contents of sub-themes are explained under the “Themes” headlines. “When presenting the results of the qualitative meta-analysis the researcher should not “boil down” the written word into numeric data, rather he should strive to keep the richness of the original text. Conducting a qualitative meta-analysis requires interpretation and reflectiveness” (Noah, 2017, p. 202). In this section interpretations from the research data are presented as researchers’ own reflection and interpretation, not just cataloguing the text from the literature. The drawn figure 1, the table 1 and three themes are the reflections and interpretations from the research data, which have been created by observing and examining the data in different ways. The proposed pre-theory is presented as synthesis of all interpretations. In this chapter the output of the research is presented, and it is based on the research methodology (Mengist et al., 2020, p. 8). It seems that cognitive hacking can be simplified to be process where human elements are manipulated by other human with the aid of technology as illustrated in Figure 2.

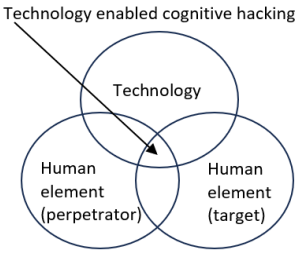

Figure 2: The interpretation of the cognitive hacking relationship to human elements and technology

During the iteration of the clauses, it soon became clear, that the discussion of the security of systems and society is merely related to development of technical solutions than inventing or discovering techniques to secure human element cognition from a form of manipulation. Even though malware and technical attacks are one phenomenon, the connection to the human element, that is manipulated to act the way the manipulator wants, receives minimal attention. Following sections, themes are described and headlines of the themes are interpreted from data.

4.1 Theme: Attacking the Mind

While the theme of attacking the mind was explored, sub-themes indicate that addressing risks from cognitive hacking is relatively scarce. Cognitive hacking targets critical infrastructure, and how to change the behaviour of the targets. In addition, it spreads misinformation and targets multiple people. There are processes that profile their victims based on available information, so the manipulated narrative can be created to cognitively manipulate them. The knowledge of internal processes makes it possible to manipulate people who are working in the organisation and to affect their daily actions.

The victims have vulnerabilities, which make them vulnerable to cognitive hacking. For example, personal traits which lead to the manipulation of beliefs. Cognitive hacking targets also victims who are already multitasking and show decision-making fatigue. One factor for vulnerability is age, but there is a wide range of manipulation for all kinds of people. When the manipulation takes place, a "malicious" action is taken. Increases in self-blame create a vulnerability to cognitive hacking by the adversary, which relates to the individual's internal fundamental assumptions. People will focus on easily accessible and understandable information, and this makes them vulnerable for manipulation if someone provides easily digestible information for them. Individuals' exploitation takes place when their subconscious thought processes are disrupted. People are not fully rational thinkers. Enrici, Ancilli and Lioy (2010) suggests, that:

> *"In particular, people who are high in normative commitment feel obligated to reciprocate gestures and favours maliciously proposed by cognitive attacks by giving up confidential and sensitive information; people who are high in continuance commitment tend to provide information to escalating requests, e.g. as part of an online game just to try to win the game; people who are high in affective commitment tend to provide information because they want to be part of a socially desirable group, or to be accepted."*

People who are trusting, will most likely fall to as victims. Enrici, Ancilli and Lioy (2010) suggests, that: "Moreover, people who are trusting were more likely to fall victim to techniques that try to cultivate trust by a process of establishing a friendly rapport or by reference to likable famous individuals, while others are more responsive to authority figures, and higher degrees of obedience to authority were an important factor in whether people responded to these types of cognitive attacks."

4.2 Theme: Definition of Cybercrime, Cyberattack, Hybrid Threat

The cognitive hacking overlaps with such definitions as cybercrime, cyberattack, information warfare, emotional warfare, hybrid warfare and hybrid threat. This due to the fact, that practitioners who deal with these "issues" classify actions of cognitive hacking in their own professional domain. After all, it is a human action that uses technology for manipulating another person to do something unwanted with technology. Table 1 illustrates how cognitive hacking has a parallel connection to other definitions.

Table 1: Parallel connections of definitions of cognitive hacking.

Cognitive hacking			
Military aspect	**Hybrid threat aspect**	**Cybercriminology aspect**	
Cognitive warfare (Bârgăoanu & Pană, 2024), (Danyk & Briggs, 2023)	Hybrid War/ hybrid warfare (Briggs et al., 2021)	False news (Java et al., 2019)	Social engineering (Java et al., 2019)
Emotive warfare (Briggs et al., 2021)	Hybrid threat (Bârgăoanu & Pană, 2024)	Insider threat (Roy & Chen, 2024)	Online manipulation (Susser et al., 2018)
Cyber warfare techniques (Briggs et al., 2021)	Political influence (Briggs et al., 2021), (Farral, 2017)	Email attacks (Guadagno & Cialdini, 2007)	Abnormal returns of the stock (Liu et al., 2021)
Weaponising the online manipulation (Susser et al., 2018)		Security attacks (Enrici et al., 2010)	Cyber attacks (Liu et al., 2021)
		Semantic attacks (Java et al., 2019)	Phishing websites / Fake online offers (Java et al., 2019)

Cognitive hacking can be used in making false news, manipulating websites for influencing wider audiences, as well as in cognitive manipulation in warfare. It can be used in email attacks, where content is used to persuade or manipulate the receiver and it can also be used to exert political influence or manipulation. Cognitive hacking

can be used in different fields, for example in the finance sector. As a summary, cognitive hacking can be seen as a weaponised method for manipulation.

Table 1 shows, that the definition of cognitive hacking is used to describe the manipulation of human elements with technology in different domains. Each domain has its own classification and scale how they technically and socially construct the meaning of the definition. From the analyses, a few theories were linked to cognitive hacking in the studies. The cognitive hacking is linked to theories of mirage theory (Rrushi, 2011), signal detection theory (Huang et al., 2020; Jiang et al., 2004; Stanislaw & Todorov, 1999), utility-theoretic (Marcum, 1947; Rrushi, 2011; Stanislaw & Todorov, 1999) and the theory of cognitive warfare (Bârgăoanu & Pană, 2024).

4.3 Theme: Protecting the Mind from Cognitive Hacking

An interesting statement in the analysed data suggested that if the technology is the weak point, why not investigate the weakest link in the supply chain, which is an extended cognition i.e. the relation between human and its cognitive artefacts. In this case, the issue lies between the relation of humans and their cognitive artefacts. The decision-making judgments are based on early perception and early cognitive processing instead of later, more systematic consideration of the information. It seems that research focuses more on building secure systems against technical threats but neglects cognitive hacking risks, even though the use of cognitive hacking techniques seems to be emerging. Machine learning technology might be capable of detecting overt threats from social cognitive hacking within content.

Countermeasures for cognitive hacking can be seen as a thought process to detect misinformation before it affects the decision and action of the user. It also includes technical solutions, that ensures information sources where information is validated and does not get manipulated. An example of this could be a situation, where website defacements are blocked by preventing unauthorised access for content manipulation. Other working countermeasures are lowering stress and cognitive overload in the organisation. The personal trait of high normative, continuance or affective commitment, as obedient to authority and wanting to trust, was found to contribute to successful cognitive attacks. In other words, this can be referred as naïve loyalty.

Combating cognitive hacking can be done by preventing unauthorised access to knowledge assets, for countering counter defacements and spoofing. In addition, detecting misinformation at an early stage can prevent it from influencing the behaviour of users. Cybenko et. al. (2002) suggests, that “In single-source cognitive hacking, the user has no independent source of information about the same topic and therefore cannot judge if the information is accurate”. Other means for combating cognitive hacking are giving cognitive training methods that help reduce impulsivity or reducing risk-taking behaviour and increasing procrastination to decrease the success of cognitive hacking. In addition, pointing out manipulative websites, which are engaging in cognitive hacking, can be used.

4.4 Cognitive Hacking: The Interpretation with the QMA

In this study, the QMA is used to create describe pre-theory of cognitive hacking. The pre-theory is proposed, because the literature review did not indicate that such a theory exists. Based on the QMA analysis of the data, a descriptive theory can be proposed. This paper presents a suggested extension to be included in layered security thinking. As such, this is a descriptive theory in a sense Gregor (2006) suggests, keeping the focus on analytically describing the observed phenomenon. In addition, Tomar and Tiwari (2025) mentioned that the Frege–Russell Descriptive Theory defines how describe meaning is constructed. The proposed descriptive pre-theory describes cognitive hacking as a form of activity done by humans. The pre-theory is called “the cognitive hacking theory”. The cognitive hacking means action where other human is manipulating another person. The cognitive hacking can include usage of technology to make manipulation. The pre-theory describes that cognitive hacking is occurring and cognitive hacking can penetrate layered security model, because of flaws existing in human nature. The pre-theory describes that cognitive hacking is used to manipulate another human element with the aid of technology. In a military environment, it is either a hybrid threat or part of hybrid warfare or information influence. In some cases, the technology might be the target itself, but the outcome is to manipulate the human element.

From the perspectives of crime science and criminology, a similar classification can be made. Purpose of the cognitive hacking is to manipulate the human element to gain, for example criminal financial gains. Technology works as a mediating device, or an aid in that environment. From the perspective of security studies and political science, technology is an aid for cognitive hacking to influence people to change their ways. It remains philosophically interesting to question what belief can be classified as truly misinformation or disinformation, because in politics, facts are not necessarily things which debate with each other. The proposal for cognitive

hacking theory includes also the basic discovery, that technology itself does not automatically protect from cognitive manipulation. It is rather a part of evolution where technology gives opportunities to manipulate human element, and the security or cybersecurity that is created inside the technology is to mitigate some cyberattacks, cybercrimes and hybrid threats.

Proposed cognitive hacking theory combines different taxonomies of security and cybersecurity, where classifications for cyber-depended manipulations or cyber enabled manipulations are used, however these are better known as cyber-depended crimes or cyber enabled crimes in criminology (Leukfeldt et al., 2020; Lusthaus et al., 2025; McGuire & Dowling, 2013). This can be derived from Alkaabi's et al. (2010) and Luomala's (2025) discussion of types of crimes. In the proposed theory cognitive hacking theory, the manipulation has a dimension named technology aided action, although the target can be influenced through manipulating human element, which have a parallel connection to Luomala's (2025) classification of type III crimes. The type III crimes mean condition where technology is a target and technology is an aid for criminal action. Technically, it means that cognitive hacking is used to manipulate a person to chance parameters of the technical artefact, which then results an alternative effect on other person to trust provided information (Type I, cyber-dependent crime). Existing technologies can be used directly to manipulate a target person (Type II, cyber-enabled crime). These types of I and II can be related to the same crime or they can be separate. When a Type I crime takes place, and it is targeted to technology, it can create an opportunity to commit a Type II crime against an individual. In addition, Luomala's (2025) extended taxonomy suggests, that cybercrime could be done by artificial intelligence. This way the presented taxonomy could be extended to cover also artificial intelligence mediated cybercrimes, and that would be Type IV crime. Reason for this is the capability of AI to be able to do both, manipulation of technology as well as cognitive manipulation against humans.

Alkaabi et al. (2010) suggests that there are Type I and Type II crimes: Type I considers technical device or software as a target of crime, while Type II considers technical device or software as aid for committing a crime. Luomala (2025) extended the taxonomy by adding Type III crimes where technical device or software is a target and aid for committing a crime (p. 50). The proposed cognitive hacking theory extends previous discoveries, where technical hacking has also elements of social manipulation, which are hidden qualitative aspects of cyberattacks, cybercrimes or hybrid threats. "The CCSTC extension with type III extension will show that cybercrime is sophisticated phenomenon and there many hidden qualitative aspects which should be measured such as have offender used social skills to commit the cybercrimes, and the most cybercrime cases are regular usage of the computer which end up to the Finnish court" (Luomala, 2025, p. 50).

Moreover, the cognitive hacking theory suggests, that the human element has internal vulnerabilities, which the layered security model has failed to address. According to the layered security model, the threat comes from the outside of the protected environment. However, the layered security environment is operated and managed by human elements. Therefore, it poses a risk of an insider threat. The layered security model mitigates the use of the system. It means that the system can be used only in certain ways and certain technical "functions" are either blocked, limited or mitigated. Cognitive hacking can penetrate these technical controls by manipulating the human element. Then the system is being used in ways, where legitimate actions are authorised from the systems perspective. However, the actions cause the "unwanted" phenomenon or impact.

5. Discussion

Cognitive hacking opens a discussion about the layered security model. When security and cybersecurity become technically so advanced, that the capabilities of the cyber offender do not necessarily find a way to bypass them, it becomes an obsolete path for cyber offender. In this case the goal shifts into finding a person, who can do everything the cyber offender needs to do. General approach has been merely on treating the symptoms, and not the cause or the discovery of where the "intention" to commit cognitive hacking merges. In this context, it is worth noticing that AI can be expanded to aid in deceiving people (Park et al., 2024). As such, the rules of the game are constantly changing.

From an engineering point of view, it is a worthy discussion to define what is an effective technical solution to "protect" the asset. However, it seems that there always exists a method to get a "malicious" act done by a person who has access and control over the system. From a legal perspective, it can be argued that the system allows only certain actions, and there is a criminal code for actions that are illegal — and this is expected to prevent "illegal" actions. However, people find ways to rationalise their actions, for example when they choose not to follow information security policy (Siponen & Vance, 2010).

There are criminological theories explaining crime (Onwuadiamu, 2025; Renzetti, 2008). Theories are such as rational choice theory, pressure theory, social learning theory, routine theory, etc. (Kivivuori et al., 2018). These theories explain from individual perspective reasons for the crime, but because cybercrime is complex phenomena involving multiple different human actors doing different "cybercrimes", and these are defined differently by different practitioners in different disciplines such as military science, security studies or criminology. A tricky question is to define a comprehensive concept for understanding "what works" against cybercriminality, because it seems to be able to an act of human behaviour. Moreover, the cognitive hacking needs more research on how artificial intelligence (AI) assisted cognitive manipulation is done. The word "artificial intelligence" was seldom mentioned in the documents, and it was mentioned as a method to detect online hoax information (Java et al., 2019) or as a part of hybrid threat (Stevens et al., 2021). In cyber domain there is a need for a theory, that explains the phenomenon of cognitive hacking what military, police or private security encounters, even if their perspective on the phenomena might be different.

The cognitive hacking relates to different types of security threats such as cybercrime, hybrid threats and cyberattacks. Cognitive hacking makes it possible to penetrate layered security model environments. The focus on security must be also on human elements, i.e. making humans aware, that they can be manipulated or become vulnerable to manipulation. The pre-theory of cognitive hacking theory, which is proposed in this study, shows that technology cannot solve issues, that exist within humanity and the human mind.

Cybercrime prevention, prevention of hybrid threats or responding to information influencing is looking for solutions to deal with this type of cognitive hacking. This qualitative meta-analysis shows, that developing effective security and cybersecurity needs evidence-based knowledge. The focus in security, cybersecurity, crime prevention, military operations and governmental security must shift towards making human element more secure by focusing more on cognitive security. It has been said that every study is as credible as the research data (Kaye & Freedman, 2011). It seems that cognitive hacking is not very widely investigated, so there is no common research tradition on how cognitive hacking should be understood (see table 1).

Different domains within the security field deal with similar human element issues, even though they classify their issue as either a cyberattack, cybercrime, hybrid threat or information warfare. It seems that this issue has been attempted to be resolved by implementing primarily technical solutions to address the weaknesses of the systems. A worthy, but understudied avenue would be to focus on understanding why humans have cognitive vulnerabilities. When does the human element become vulnerable to cognitive hacking? In addition, in contemporary society and time, the studies on what can be done to prevent cognitive hacking in an effective way are to be encouraged. Also, the role of AI should be studied — especially in AI mediated cognitive hacking?

The proposed cognitive hacking theory suggests that the layered security model has a vulnerability in cognitive security, which is not addressed. By using cognitive hacking, the technological layered security model systems can be penetrated, and unwanted actions can take place when the human elements are manipulated. The cognitive hacking phenomenon is not isolated only to cybercrime and financial crimes. It includes actions where people's public opinion can be manipulated in the political domain, or perception of the situation can be influenced through information warfare. Currently, there are no clear methods to address the issue of cognitive hacking, but in future studies, techniques to secure the human element are vital. The concept of cognitive security needs to be re-invented or re-discovered. Also, the role of AI-assisted cognitive manipulation should be studied more.

Ethics Declaration: An ethical declaration was not required for the research.

AI Declaration: Artificial intelligence was used to convert one PDF file by optical character recognition to form where strings can be copied from the PDF file. The service is available at www: https://pdf.ai/tools/ocr-pdf

References

Adefabi, R., Onimole, O., Sani, A. I., Olajide, O., Okpalanozie, V., Oseni, T., Iwuoha, C., Eniafe, F., Palmer, X., & Potter, L. (2025). Cognitive Hacking and Social Engineering in Healthcare: Exploiting Human Behaviour. *European Conference on Cyber Warfare and Security*, *24*(1), 1–8. https://doi.org/10.34190/eccws.24.1.3337

Alkaabi, A., Mohay, G., Mccullagh, A., & Chantler, N. (2010). Dealing with the Problem of Cybercrime. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, *53*, 1–18. https://doi.org/10.1007/978-3-642-19513-6_1

Anti, E. (2025). *Insider Deviant Behavior in Cybersecurity* [Article-based doctoral dissertation, University of Vaasa]. https://urn.fi/URN:ISBN:978-952-395-231-7

Anti, E., & Vartiainen, T. (2024). Explanations of Insider Deviant Behavior in Information Security: A Systematic Literature Review. *Communications of the Association for Information Systems*, *55*(1), 1–36. https://doi.org/10.17705/1CAIS.05501

Bârgăoanu, A., & Pană, M. (2024). Cyber Influence Defence: Applying the DISARM Framework to a Cognitive Hacking Case from the Romanian Digital Space. *Applied Cybersecurity and Internet Governance*, *3*(1), 92–121. https://doi.org/10.60097/ACIG/190196

Berente, N., Lyytinen, K., Yoo, Y., & Maurer, C. (2019). Institutional Logics and Pluralistic Responses to Enterprise System Implementation: A Qualitative Meta-Analysis. *MIS Quarterly*, *43*(3), 873–902. https://doi.org/10.25300/MISQ/2019/14214

Briggs, C. M., Danyk, Y. G., & Maliarchuk, T. (2021). Security Aspects of Hybrid War, COVID-19 Pandemic and Cyber-Social Vulnerabilities. *Connections*, *20*(3–4), 47–72. https://doi.org/10.11610/Connections.20.3-4.03

Caballero, A. (2014). Information Security Essentials for IT Managers. In *Managing Information Security* (Second Edition, pp. 1–45). Elsevier. https://doi.org/10.1016/B978-0-12-416688-2.00001-5

Cybenko, G., Giani, A., & Thompson, P. (2002). Cognitive hacking: a battle for the mind. *Computer*, *35*(8), 50–56. https://doi.org/10.1109/MC.2002.1023788

Danyk, Y., & Briggs, C. (2023). Modern Cognitive Operations and Hybrid Warfare. *Journal of Strategic Security*, *16*(1), 35–50. https://doi.org/10.5038/1944-0472.16.1.2032

Depraetere, J., Vandeviver, C., Keygnaert, I., & Beken, T. Vander. (2021). The critical interpretive synthesis: an assessment of reporting practices. *International Journal of Social Research Methodology*, *24*(6), 669–689. https://doi.org/10.1080/13645579.2020.1799637

Enrici, I., Ancilli, M., & Lioy, A. (2010). *A psychological approach to information technology security*. 459–466. https://doi.org/10.1109/HSI.2010.5514528

Farral, T. (2017). Nation-state attacks: practical defences against advanced adversaries. *Network Security*, *2017*(9), 5–7. https://doi.org/10.1016/S1353-4858(17)30111-3

Ghasemi, H., & Babaie, S. (2024). A new intrusion detection system based on SVM–GWO algorithms for Internet of Things. *Wireless Networks*, *30*(4), 2173–2185. https://doi.org/10.1007/s11276-023-03637-6

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611. https://doi.org/10.2307/25148742

Guadagno, R. E., & Cialdini, R. B. (2007). Persuade him by email, but see her in person: Online persuasion revisited. *Computers in Human Behavior*, *23*(2), 999–1015. https://doi.org/10.1016/j.chb.2005.08.006

Habersang, S., Küberling-Jost, J., Reihlen, M., & Seckler, C. (2019). A Process Perspective on Organizational Failure: A Qualitative Meta-Analysis. *Journal of Management Studies*, *56*(1), 19–56. https://doi.org/10.1111/joms.12341

Hagan, F. (2018). *Research Methods in Criminal Justice and Criminology* (10th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/research-methods-in-criminal-justice-and-criminology/P200000001159/9780137409020

Hoon, C. (2013). Meta-Synthesis of Qualitative Case Studies. *Organizational Research Methods*, *16*(4), 522–556. https://doi.org/10.1177/1094428113484969

Howitt, D. (2006). What is forensic and criminal psychology? In *Introduction to Forensic and Criminal Psychology* (2nd ed., p. 8). Pearson Education Limited.

Huang, W., Chen, X., Jin, R., & Ching Lau, N. K. (2020). Detecting cognitive hacking in visual inspection with physiological measurements. *Applied Ergonomics*, *84*. https://doi.org/10.1016/j.apergo.2019.103022

Ignacio, C. U., & Schwerer, J. (2024). *Qualitative meta-analysis* (pp. 325–336). https://doi.org/10.14361/9783839467343-025

Jamuna, K. M. (2024). Social Engineering and Human Factors in Penetration Testing. *International Journal For Multidisciplinary Research*, *6*(3). https://doi.org/10.36948/ijfmr.2024.v06i03.21496

Java, S., Basheer, F. L., Riaz, S., Kaur, M. J., & Mushtaq, A. (2019). *Detection of Online Manipulation to Prevent Users Victimization*. 593–599. https://doi.org/10.1109/AICAI.2019.8701330

Jiang, X., Khasawneh, M. T., Master, R., Bowling, S. R., Gramopadhye, A. K., Melloy, B. J., & Grimes, L. (2004). Measurement of human trust in a hybrid inspection system based on signal detection theory measures. *International Journal of Industrial Ergonomics*, *34*(5), 407–419. https://doi.org/10.1016/j.ergon.2004.05.003

Kaye, H. D., & Freedman, A. D. (2011). Reference Guide on Statistics. In *Reference Manual on Scientific Evidence*. National Academies Press. https://doi.org/10.17226/13163

Kivivuori, J. K. A., Aaltonen, M., Näsi, M. J., Suonpää, K. E.-M., & Danielsson, P. M. (2018). *Kriminologia: Rikollisuus ja kontrolli muuttuvassa yhteiskunnassa*. Gaudeamus.

Knapp, E. D. ., & Langill, J. Thomas. (2015). *Industrial network security : securing critical infrastructure networks for Smart Grid, SCADA, and other industrial control systems*. Syngress.

Küberling-Jost, J. A. (2021). Paths of Corporate Irresponsibility: A Dynamic Process. *Journal of Business Ethics*, *169*(3), 579–601. https://doi.org/10.1007/s10551-019-04263-z

Lee, H., Tamminen, K. A., Clark, A. M., Slater, L., Spence, J. C., & Holt, N. L. (2015). A meta-study of qualitative research examining determinants of children's independent active free play. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 5. https://doi.org/10.1186/s12966-015-0165-9

Leukfeldt, E. R., Notté, R. J. (Raoul), & Malsch, M. (Marijke). (2020). Exploring the Needs of Victims of Cyber-dependent and Cyber-enabled Crimes. *Victims & Offenders*, *15*(1), 60–77. https://doi.org/10.1080/15564886.2019.1672229

Levitt, H. M. (2018). How to conduct a qualitative meta-analysis: Tailoring methods to enhance methodological integrity. *Psychotherapy Research*, *28*(3), 367–378. https://doi.org/10.1080/10503307.2018.1447708

Lewis, T. (2006). *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471789542

Liu, J., Cheng, L., Chi, H., Liu, C., & Alò, R. A. (2021). CDetector: Extracting Textual Features of Financial Social Media to Detect Cyber Attacks. *Proceedings - International Conference on Computer Communications and Networks, ICCCN*, *2021-July*. https://doi.org/10.1109/ICCCN52240.2021.9522290

Luomala, M. (2025). A New Version of Cybercrime Taxonomy: Empirical Evidence from Finland. *International Journal of Information Security and Cybercrime*, *14*(1), 37–62. https://doi.org/10.19107/IJISC.2025.01.03

Lusthaus, J., Holt, T. J., Levi, M., Kleemans, E., & Leukfeldt, E. R. (2025). The evolution of Nigerian cybercrime: Two case studies of UK-based offender networks. *European Journal of Criminology*, *22*(4), 557–577. *https://doi.org/10.1177/14773708251329695*

Marcum, J. I. (1947). *A statistical theory of target detection by pulsed radar*. http://www.rand.org/pubs/research_memoranda/ 2006/RM754.pdf

McGuire, M., & Dowling, S. (2013). *Cyber crime: A review of the evidence - Chapter 3: Cyber-enabled crimes – sexual offending against children- Home Office Research Report 75*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246754/horr75-chap3.pdf

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, *7*, 100777. https://doi.org/10.1016/j.mex.2019.100777

Montañez, R., Golob, E., & Xu, S. (2020). Human Cognition Through the Lens of Social Engineering Cyberattacks. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01755

Morrow, A. B. (2021). Information security and cyber threats and vulnerabilities. In *Intermodal Maritime Security* (pp. 169–193). Elsevier. https://doi.org/10.1016/B978-0-12-819945-9.00010-1

Noah. (2017). A systematic approach to the qualitative meta-synthesis. *Issues In Information Systems*, *18*(2). https://doi.org/10.48009/2_iis_2017_196-205

Omanović, V., & Langley, A. (2023). Assimilation, Integration or Inclusion? A Dialectical Perspective on the Organizational Socialization of Migrants. *Journal of Management Inquiry*, *32*(1), 76–97. https://doi.org/10.1177/10564926211063777

Onwuadiamu, G. (2025). Cybercrime in criminology; A systematic review of criminological theories, methods, and concepts. *Journal of Economic Criminology*, *8*, 100136. https://doi.org/10.1016/j.jeconc.2025.100136

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n71

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. *Patterns*, *5*(5), 100988. https://doi.org/10.1016/j.patter.2024.100988

Renzetti, C. M. (2008). Criminal Behavior, Theories of. In *Encyclopedia of Violence, Peace, & Conflict* (pp. 488–498). Elsevier. https://doi.org/10.1016/B978-012373985-8.00042-8

Roy, K. C., & Chen, G. (2024). GraphCH: A Deep Framework for Assessing Cyber-Human Aspects in Insider Threat Detection. *IEEE Transactions on Dependable and Secure Computing*, *21*(5), 4495–4509. https://doi.org/10.1109/TDSC.2024.3353929

Rrushi, J. L. (2011). An exploration of defensive deception in industrial communication networks. *International Journal of Critical Infrastructure Protection*, *4*(2), 66–75. https://doi.org/10.1016/j.ijcip.2011.06.002

Shinder, T. W., & Shinder, D. L. (2005). Evolution of a Firewall: From Proxy 1.0 to ISA 2004. In *Dr. Tom Shinder's Configuring ISA Server 2004* (pp. 1–77). Elsevier. https://doi.org/10.1016/B978-193183619-7/50008-3

Siponen, M., & Vance, A. (2010). Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations. *MIS Quarterly*, *34*(3), 487. https://doi.org/10.2307/25750688

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stevens, T., Ertan, A., Floyd, K., & Pernik, P. (Eds.). (2021). *Cyber Threats and NATO 2030: Horizon Scanning and Analysis*. NATO Cooperative Cyber Defence Centre of Excellence. https://kclpure.kcl.ac.uk/portal/en/publications/cyber-threats-and-nato-2030-horizon-scanning-and-analysis/

Susser, D., Roessler, B., & Nissenbaum, H. F. (2018). Online Manipulation: Hidden Influences in a Digital World. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3306006

Tamminen, K. A., & Holt, N. L. (2010). A meta-study of qualitative research examining stressor appraisals and coping among adolescents in sport. *Journal of Sports Sciences*, *28*(14), 1563–1580. https://doi.org/10.1080/02640414.2010.512642

Timulak, L. (2009). Meta-analysis of qualitative studies: A tool for reviewing qualitative research findings in psychotherapy. *Psychotherapy Research*, *19*(4–5), 591–600. https://doi.org/10.1080/10503300802477989

Tomar, P., & Tiwari, S. K. (2025). THE SEMANTICS OF LEGAL INTERPRETATION. *International Comparative Jurisprudence*, *11*(1), 1–15. https://doi.org/10.13165/j.icj.2025.11.001

Washo, A. H. (2021). An interdisciplinary view of social engineering: A call to action for research. *Computers in Human Behavior Reports*, *4*, 100126. https://doi.org/10.1016/j.chbr.2021.100126

Weed, M. (2005). 'Meta Interpretation': A Method for the Interpretive Synthesis of Qualitative Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *6*(1). https://doi.org/https://doi.org/10.17169/fqs-6.1.508

Whitty, M. T. (2020). Is There a Scam for Everyone? Psychologically Profiling Cyberscam Victims. *European Journal on Criminal Policy and Research*, *26*(3), 399–409. https://doi.org/10.1007/s10610-020-09458-z