Defining Comprehensive Cognitive Security in the Digital Era: Literature Review and Concept Analysis

H Grahn*[1], T Taipalus[2]

[*] Corresponding author
Tampere
[1]University of Jyväskylä
P.O. Box 35
FI-40014 University of Jyväskylä,
Jyväskylä, Finland

E-mail: hilkka.grahn@jyu.fi

[2]Tampere University
FI-33014 Tampere University
Tampere, Finland

E-mail: toni.taipalus@tuni.fi

Abstract: *Cognitive security, an emerging field, spans disciplines and contexts but often lacks clear definitions. In the digital age, where disinformation spreads rapidly via the Internet, cognitive security is particularly crucial. Drawing on theoretical background, a definition is proposed: comprehensive cognitive security can be defined as a state and a process in which undesired malign influence or manipulation is incapable of altering human cognition and which can be achieved through a combination of knowledge and situation awareness with purposeful actions. This definition provides a framework that can be applied across various disciplines and can support societies in educating citizens on cognitive security.*

Keywords: *Psychological Information Influence, Malign Influence, Cognitive Warfare, Information Operations, Disinformation, Situation Awareness*

Introduction

In today's complex information environment, the Internet, large language models, and social media platforms like Facebook, Instagram, and TikTok enable the rapid, efficient, and widespread creation and dissemination of false information (that is, misinformation and disinformation; for more information, see for example Altay *et al*. 2023 and Kruger *et al*. 2024). For instance, Russia used TikTok to spread disinformation regarding the Ukrainian war (Evans 2022) and Facebook (McCombie, Uhlmann & Morrison 2020) and X (formerly Twitter) (Linvill *et al.* 2019) to undermine the 2016 U.S. presidential election. Additionally, Facebook and X have been used to spread disinformation, for instance, about the COVID-19 virus (Yang, KC *et al.* 2021). Even brief exposure to such fake news can unconsciously influence individuals' behaviour (Bastick 2021). Not

only state actors, but also individuals, utilize social media platforms to disseminate disinformation. Reasons for this behaviour may include, for example, anger towards the government (Wintterlin *et al.* 2023). These factors can contribute to the formation of echo chambers—although some research suggests their impact may be overstated (see for example Dubois & Blank 2018)—as well as filter bubbles, mistrust, and societal polarization, while also exerting a significant influence on human cognition (see, for example, Kozyreva, Lewandowsky & Hertwig 2020). In this era, social media shapes public opinion and can lead to issues like political polarization (Gorodnichenko, Pham & Talavera 2021). Furthermore, the proliferation of misinformation and disinformation on these digital platforms erodes public trust (Lovari 2020; Nahum, Drekonja & Alpern 2021) and diminishes trust in traditional media (Ognyanova *et al.* 2020). The spread of fake news online has even led to scepticism about science, truthful news, and societal norms (Olan *et al.* 2022). Moreover, the World Economic Forum's Global Risk Report (2024) identified misinformation and disinformation as the most severe global risks over the next two years, exacerbating societal and political divides. While there is no universally accepted term for all these activities, some authors have suggested the term disinformation, misinformation, and malign influence (DMMI) to capture the wide spectrum of deceptive, manipulative, exploitative, and coercive activities involved (Kruger *et al*. 2024). In addition to these influences, modern threats such as online scams, phishing attempts, identity theft, and social engineering further jeopardize individuals' cognitive processes. For example, a 2020 Google report revealed that an astonishing 126 million COVID-19-related phishing scams were blocked in just one week (Hanoch & Wood 2021). This number highlights the sheer volume of scams that preyed on people's fears, with tactics such as fake donation requests and false COVID-19 treatments. These scams posed a serious threat to human cognition, particularly in the realm of decision making.

Strategies and actions designed to alter the opinions, attitudes, emotions, behaviour, and decision making of individuals, groups, and societies have been referred to as psychological information influence (Starbird, Arif & Wilson 2019; Tashev, Purcell & McLaughlin 2019). The relatively new concept of cognitive security has been associated with this psychological information influence and is also an emerging concern in national security discourses around the world (for example, Buzzell 2024). However, cognitive security (like cognitive safety) is a term and phenomenon that lacks a universally established definition and is used across different disciplines and contexts with varying meanings.

According to Terp and Breuer (2022), the term 'cognitive security' is employed in two distinct communities: the machine learning in information security community and the disinformation community. In the former, it encompasses Artificial Intelligence technologies, attacks and defences of information systems, and vulnerabilities in machine learning algorithms. The latter community concentrates on manipulating human beliefs through misinformation and disinformation. Beyond these two communities, cognitive security (or safety) also appears in other contexts, including ransomware detection (Silva & Hernandez-Alvarez 2017), data protection (Ogiela & Ogiela 2020), and warfare (Waltzman 2017). Furthermore, it is explored in medical fields, such as the study of dementia (Pulsford, Duxbury & Carter 2016) and clinical drug development (Roiser *et al.* 2016). In some cases, the term is used without a definition. Therefore, depending on the context, cognitive security can mean a myriad of things.

In summary, this study aims to provide a definition for cognitive security which 1) spans different disciplines (that is, is not limited to one discipline, such as military or healthcare), 2) is based on

previous theories of human cognition, appraisal theory, and situation awareness, 3) is rooted in a systematic analysis, and 4) is provided in detail regarding, for instance, the elements enhancing and hindering cognitive security. This definition is aimed to ensure more consistent understanding among researchers on what cognitive security is. Hence, the following research goals have been posited:

1) Explore how previous literature has defined or used the term *cognitive security,*
2) Clarify the meaning of the concept of *cognitive security* from a multi-disciplinary perspective,
3) Clarify what enhances cognitive security,
4) Identify antecedents, attributes, and consequences of cognitive security, and
5) Develop a working definition for comprehensive cognitive security that is meaningful across different disciplines, and which can be universally understood and applied across research, policy, and practice.

These research goals are pursued through a systematic literature review followed by a concept analysis based on the findings.

The remaining paper is organized as follows: first, it outlines the theoretical foundations relevant to cognitive security, including human cognition, appraisal theory, and security. Next, it details the method used for the systematic literature review and presents the findings, along with a comprehensive concept analysis. Based on these insights, the paper defines *comprehensive cognitive security* by integrating the theoretical background with the literature review findings and concept analysis. It also introduces a framework for comprehensive cognitive security. Finally, the paper discusses the study's practical implications, highlights its limitations, and provides concluding remarks.

Theoretical Background

The purpose of this section is to briefly clarify concepts related to cognitive security from a multidisciplinary perspective, establishing a scientific foundation for understanding the topic. It provides a basic understanding of key theoretical concepts, highlighting their relevance to cognitive security. The following subsections will introduce human cognition, appraisal theory, and security.

Human cognition

Cognition, a highly specialized set of mental operations (Loh & Kanai 2016), plays an integral role in virtually all facets of human activity and behaviour. Cognitive processes mediate the effects of stimuli on a response (Nisbett & Wilson 1977). However, as is customary in science, the definition of 'cognition' lacks consensus in the scientific community (Allen 2017). Perhaps one of the most widely recognized definitions dates back to 1967, stating that "The term cognition refers to all the processes by which sensory input is transformed, reduced, elaborated, stored, recovered, and used" (Neisser 2014, p. 4). Additionally, the functions of cognition can be divided in two: basic functions and higher-level functions. The basic cognitive functions include attention, perception, and memory, whereas higher-level functions encompass, for instance, speech and language processing, decision making, and executive control (Glisky 2007). Executive control, further, consists of a range of different processes involved in activities such as planning (Pessoa 2008), problem solving (Newell, Shaw & Simon 1958), organization, coordination, implementation, and

evaluation (Glisky 2007). Overall, cognition refers to a myriad of processes that enable humans to comprehend and to make sense of the world (Frith 2008).

Appraisal theory

In general, emotions are closely connected to cognitive processes and this cooperation is essential for humans to function in their environment (Ochsner & Phelps 2007). Emotions also significantly impact reasoning and contribute to the fixation of belief (Dolan 2002). In relation to appraisal theory, Lazarus (1991) has described the relationship between cognition and emotions as functional: when something happens to an individual, emotions result from the evaluation of how significant that event is for the individual. Therefore, emotions are always a response to a cognitive process. In other words, thoughts have the capacity to trigger emotions, and emotions cannot occur without some form of thought. This evaluation, whether conscious or unconscious, of events is termed 'appraisal' (Pessoa 2008), and it pertains to the appraisal theory. According to Moors *et al*. (2013), appraisal theory posits that emotions function as adaptive responses, reflecting evaluations of environmental features crucial for the well-being of the organism. Expanding on this, Scherer (1999) highlights that at the core of appraisal theory is the notion that emotions are produced and distinguished through an individual's subjective assessment of the personal significance of a situation, object, or event. In essence, this means that the same situation, object, or event may lead to different emotions in two people since one person can see it as highly significant while the other may not. In the context of cognitive security, the appraisal process becomes significant when individuals assess potential threats or manipulative influences.

Security

Security is described as both multidimensional in nature and diverse in practice, making it a difficult term to define (Brooks 2010). Zedner (2003) offers one definition, suggesting that security serves a dual role—both as a state of being and as a means to an end. The 'state of being' encompasses both objective and subjective conditions. The objective condition can be further divided into three forms: the absence of threats (that is, a state of absolute security), the neutralization of threats (that is, being protected from threats), and the avoidance or non-exposure to danger. The subjective condition refers to the positive feeling of being safe and free from anxiety or apprehension. Essentially, it is the sense that we have of our safety. According to Zedner (2009), security as a subjective state primarily exists in the mind and is partly influenced by material and social conditions. There is often an association between objective and subjective conditions of security. Furthermore, security can also be seen as a means to achieve a specific goal, which can be either the objective freedom from risk or the subjective feeling of safety. However, the pursuit of security can also be an end in itself. According to Zedner (2003), there are challenges in identifying whether the end has been achieved or not. This is partly because absolute security is said to be an illusion, and partly because, even if security were attainable today, it may not be tomorrow. The pursuit of security is seen as an ongoing endeavour that demands continuous vigilance.

Security beliefs are formed by individuals' perceptions of threats in their environment. Through cognitive processes, individuals evaluate their level of security, resulting in subjective interpretations (Bar-Tal & Jacobson 1998). These perspectives align with the principles of the appraisal theory. Naturally, the opposing concept of security is insecurity. When individuals believe they are unable to cope with or face difficulties coping with a perceived threat, it leads to the formation of insecurity beliefs (Smith & Lazarus 1993). The perspective that security encompasses both a state of being

and a process to attain is adopted here, with recognition of the role appraisal plays in individuals' perceived security.

Materials and Method

Search strategy and study selection

Four databases—ProQuest, Science Open, Scopus, and Web of Science—were used to conduct the systematic literature review. Two separate searches were performed using the terms "cogni* security" and "cogni* safety". These terms were searched within the titles, keywords, and abstracts of the articles. The results were filtered to include only peer-reviewed articles or conference proceedings written in English. As a result, commercial reports, grey literature, books, and similar sources were excluded. The searches were conducted from 22 August 2023, to 6 September 2023.

Using the search term "cogni* security", 197 results were initially found, while "cogni* safety" yielded 250 results. These were subsequently narrowed down to a final selection of 25 papers, with four of them falling into the category of known works. The inclusion criteria consisted of five points that the abstract, or subsequently, the full text needed to meet: 1) the article is about human cognition (not machine cognition or similar), 2) the article is about security in the context of human behaviour, 3) the article is written in English, 4) the article is published in a peer-reviewed scientific forum, and 5) the article is available online. The exclusion criterion specified that articles concerning the effects of drug testing on specific aspects of cognition were not considered.

One researcher conducted the initial database searches and selected the articles based on their titles. After the title-based selection, two researchers independently reviewed all the abstracts or full texts, depending on the selection phase, and autonomously determined whether each article met the established criteria. If their assessments differed, they discussed the discrepancies to reach a consensus on whether the article should be included. A more detailed description of the search process is provided in **Figure 1**. The selected articles were published between 2002 and 2023, with a focus on the period between 2020 and 2023. The selected research was published across 24 outlets.

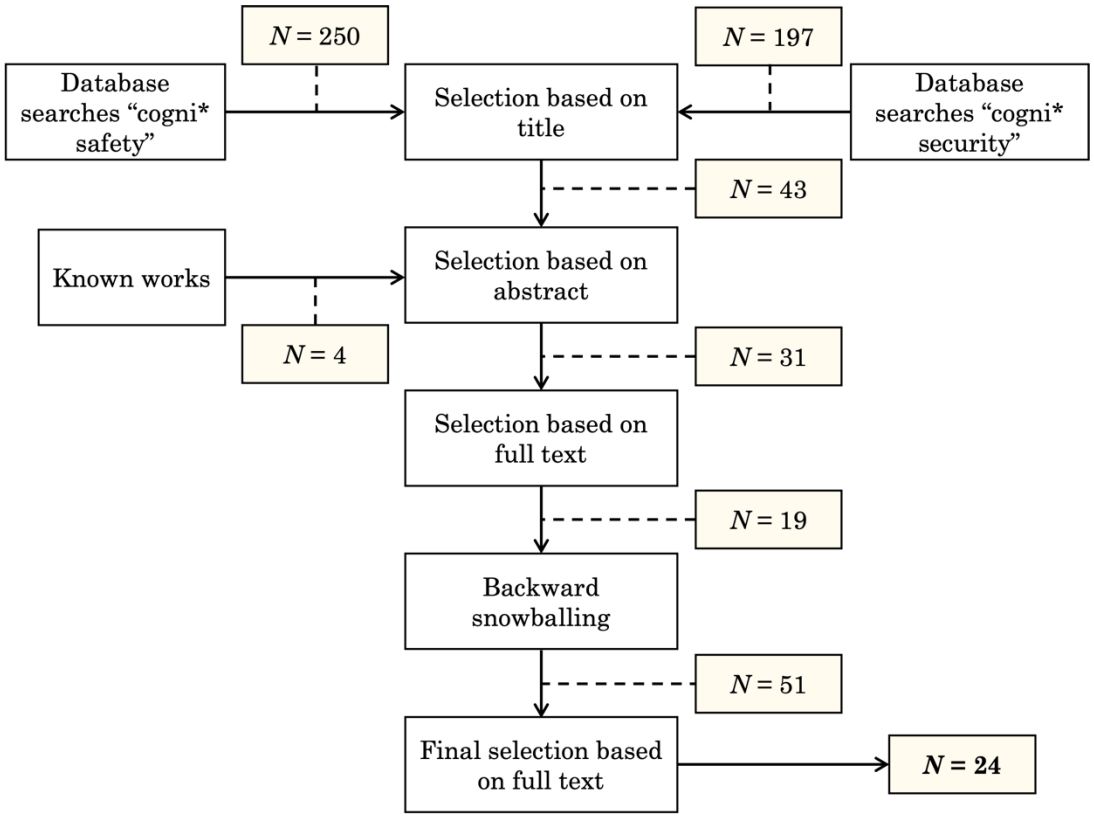


Figure 1: The literature selection process

Analyses

The selected articles were examined to determine how cognitive security or cognitive safety was defined, with definitions reported in **Table 1**. For articles lacking a clear definition, the discussion of phenomena interpreted as cognitive security were analysed. Additionally, the articles were reviewed to identify references to human skills or knowledge that could enhance individuals' cognitive security or related concepts. Articles addressing these aspects were categorized using inductive content analysis (Elo & Kyngäs 2008). This process involved compiling all mentions of enhancing factors into a single table, grouping similar mentions under broader headings to consolidate the data. The original 21 mentions were reduced to 11 categories, with some categories containing only one reference. These results are presented in **Table 2**.

To analyse the concept of cognitive security, Rodger's evolutionary method, as outlined by Tofthagen and Fagerstrøm (2010), was applied. The method is particularly suited for concepts spanning multiple disciplines, such as cognitive security. It seeks to clarify, describe, and explain the concept by analysing its usage across various fields. Following this method, surrogate terms, related terms, antecedents, attributes, examples, and consequences of the concept were identified. These elements, along with explanations of their meanings, are presented in **Table 3**. The concept analysis was informed by the reviewed literature and the identified enhancing factors.

Findings of the Systematic Literature Review

Definition of cognitive security

Following a thorough examination of the selected articles containing any of the search terms in the

title, abstract, or key words, the articles were divided into two groups: those providing definitions of cognitive security or cognitive safety, and those that, while not offering formal definitions, used or discussed terms interpreted as related to cognitive security or its dimensions. In total, nine of the selected articles included a definition of cognitive security or cognitive safety, which are presented in **Table 1**. The definitions are presented here as they appear in the original articles to highlight the various meanings of the term.

Definition	Source
"Cognitive security considers four components: processes, knowledge, technology, and cognitive abilities to establish mental maps, complex data fusion, the handling of massive data and the maintenance of knowledge".	Andrade & Yoo 2019
"Cognitive security is the ability to generate cognition for efficient decision making in real-time by modelling human thought processes to detect cyber security attacks and develop defence strategies."	Andrade *et al*. 2022
"Cognition Security refers to the potential impacts of fake news on human cognition, ranging from misperception, untrusted knowledge acquisition, targeted opinion/attitude formation, to biased decision making"—"and investigate[s] the efficient ways to debunk fake news and maintain human cognition security".	Guo *et al*. 2021
"Cognitive security is the ability to detect, characterize, and counter misinformation, disinformation, and other information-based harms and forms of malign influence among people. Resilience, as part of cognitive security, includes the structural context that protects humans from exposure to disinformation in the first place, as well as the ability to identify it, limit its spread, and mitigate its effects once exposed." And, "We also include the ability to effectively, securely, and reliably utilize and preserve critical information in our definition of cognitive security".	Janzen, Orr & Terp 2022a; Janzen, Orr & Terp 2022b
"Cognition security approaches the impact disinformation has on human cognition with the goal of changing attitudes and influencing the decision-making process through misperception and distrustful information and knowledge, and seen as a process, it is meant to ensure humans' safety from being affected by false content.–Cognition Security has been called upon as a means to detect and understand the hostile acts and plan a response strategy against those that act to bring confusion and weaken trust between state institutions and society".	Popescu 2021
Cognitive security "focuses on the exploitation of cognitive biases in large public groups".	Rajtmajer & Susser 2020
Cognitive security is "the application of information security principles, practices, and tools to misinformation, disinformation, and influence operations. Cognitive security can be seen as a holistic view of disinformation from a security practitioner's perspective".	Terp & Breuer 2022

"Cognitive security, by extension, can be understood as efforts aimed at strengthening the defence of individuals–military or otherwise–and social groups against threats and/or vulnerabilities to cognitive warfare tactics and strategies of adversaries".	Doherty 2023

Table 1: Found definitions of cognitive security or cognition security

Fourteen of the selected articles that met the inclusion criteria did not provide definitions for the terms 'cognitive security' or 'cognitive safety'. Despite this, all the articles used these terms.

Kozyreva, Lewandowsky & Hertwig (2020) discuss manipulative techniques on the Internet designed to influence, for example, individuals' decision-making and voting behaviour. They suggest that individuals can gain control over their digital environments by enhancing their information literacy and cognitive resistance to manipulation. This was interpreted as a form of cognitive security. Roozenbeek and van der Linden (2019) discuss 'cognitive immunity' and how it can be enhanced by exposing, warning, and familiarizing people with the strategies used in the production of fake news before they encounter real misinformation. Cybenko, Giani and Thompson (2002) use the term 'overt or covert cognitive hacking'. 'Covert cognitive hacking' refers to the subtle manipulation of perception and the use of misleading information, while 'overt cognitive hacking' involves defacing or spoofing legitimate communication to influence "(Internet) users". In the military context, cognitive security is discussed in relation to cognitive warfare—defined, for example, as "controlling others' mental states and behaviours by manipulating environmental stimuli" (Hung & Hung 2022, p. 1)—and influence operations, which have been described as efforts "to exert power by influencing the behaviour of a target audience" (Brangetto & Veenendaal 2016, p. 114). In this context, cognitive security is primarily defined as protection against such forms of warfare (Brangetto & Veenendaal 2016; Danet 2023; Hung & Hung 2022). Huang *et al*. (2023) state that online interaction has become the primary channel for influencing individuals' cognition and that AI technologies pose challenges to cognitive security. However, they do not provide an explanation of what cognitive security entails.

Vulkan *et al*. (2015) use the term 'cognitive security' in the context of job well-being, suggesting it is synonymous with perceived risk and effects on mental wellbeing. Wang *et al*. (2019) also use psychological security and define it as an individual's emotional experience derived from external stimuli. It is determined by both the intensity of the stimulus and the psychological qualities of the individual. As they conceive it, 'psychological security' also refers to an individual's sense of certainty, control, and ability to anticipate risks.

The construction industry uses the term 'cognitive safety' which refers to construction workers' safety behaviour on the job. Failures in cognitive processes can lead to unsafe behaviour (Deng, Peng & Pan 2022; Deng *et al.* 2023; Liang *et al*. 2022). Specifically, Deng, Peng & Pan (2022) identify four types of cognitive failures resulting in unsafe behaviour: 1) being unaware of the hazard, 2) not knowing the hazard, 3) choosing unsafe behaviour, and 4) being unable to avoid the hazard or not knowing how to avoid it. In the context of autonomous vehicles, the term 'cognitive safety' is also used. Cao *et al*. (2021) define it as an aspect of overall perceived safety, specifically relating to factors such as the perceived controllability of the autonomous vehicle. Prasetio and Nurliyana (2023) also consider cognitive safety to be a component of perceived safety, emphasizing that it involves users' understanding of the risks associated with autonomous vehicles and their

confidence in the technology's ability to ensure their safety. In addition, in one article, the term is used in the text "to ensure that one's cognitive security remains intact" (Yeh 2016, p. 542) but not defined.

Enhancing factors of cognitive security

Some of the articles also discussed factors—such as human skills and knowledge—that could enhance individuals' cognitive security. These skills and areas of knowledge are summarized in **Table 2**.

Enhancing factor	**Source**
Situation awareness	Andrade *et al*. 2022; Andrade & Yoo 2019; Cybenko, Giani & Thompson 2002; Huang *et al*. 2023
Cyber awareness	Brangetto & Veenendaal 2016
Safety awareness: sense of safety, hazard identification, accurate risk judgement, risk identification knowledge, and correct understanding of dangerous information	Deng *et al*. 2022, 2023; Liang *et al*. 2022
Digital literacy ("ability to distinguish between false and mainstream news")	Hung & Hung 2022; Janzen, Orr & Terp 2022a, 2022b; Kozyreva *et al*. 2020
Critical thinking	Hung & Hung 2022
Understanding the way humans' access, share, and repost information	Popescu 2021; Guo *et al*. 2021
Bias mitigation	Doherty 2023
Anticipation, refuting and responding to disinformation, training and education	Danet 2023
Prebunking (that is, preventing false narratives from taking root in memory)	Roozenbeek & van der Linden 2019
Inoculation technique (that is, "exposing people to a weakened version of a misleading argument, and by preemptively refuting this argument, attitudinal resistance can be conferred against future deception attempts")	Roozenbeek & van der Linden 2019
Cognitive interventions and tools aiming to foster people's cognitive and motivational competencies in their digital environments and boosting competencies of reasoning and resilience to manipulation	Kozyreva, Lewandowsky & Hertwig 2020

Table 2: Factors enhancing or protecting cognitive security

Concept analysis—Identifying antecedents, attributes, and consequences of cognitive security

To develop a more comprehensive understanding of cognitive security, given that the articles

indicate the phenomenon extends beyond a single definition, a concept analysis (Tofthagen & Fagerstrøm 2010) was conducted based on the reviewed articles.

Surrogate terms Do other words say the same thing as the chosen concept? (Tofthagen & Fagerstrøm 2010)	Not exactly, other terms are too narrow
Related terms Do other words have something in common with the concept? (Tofthagen & Fagerstrøm 2010)	Resilience, psychological security, cognitive immunity, perceived safety, sense of security
Antecedents Which events or phenomena have been associated with the concept in the past? (Tofthagen & Fagerstrøm 2010) Antecedents must occur prior to the occurrence of the analysed concept (Henneman, Lee & Cohen 1995).	Misinformation, disinformation, exploitation of cognitive biases occurring in human thinking, manipulation, experience of risk or threat, malign influence, cognitive engineering, misperception, influence operation, cognitive warfare, information warfare, poor safety attitude, cognitive hacking, polarization
Attributes What are the concept's characteristics? (Tofthagen & Fagerstrøm 2010) Characteristics without which the concept cannot occur (Henneman, Lee & Cohen 1995).	Being aware of the situation (that is, situation awareness: perception, comprehension, projection), critical thinking, digital literacy, information literacy, reasoning, bias mitigation, training and education, identifying misinformation, disinformation and fake news, identifying malicious content and acts, resilience, feeling of overall security
Consequences What happens after or as a result of the concept? (Tofthagen & Fagerstrøm 2010) End-points that occur as a result of the antecedents and attributes of the concept (Windle 2011).	Human cognition (for example, decision making and opinion forming) cannot be altered by malign influence or manipulation

Table 3: Antecedents, attributes, and consequences of cognitive security and questions used during the analysis

This analysis fulfills Research goal 4, 'identify antecedents, attributes, and consequences of cognitive security'.

Discussion

Cognitive security from a multi-disciplinary perspective

The systematic literature review addressed that the term 'cognitive security' is indeed employed across various disciplines and contexts, each with slightly different interpretations. 'Cognitive security' is used, for instance, in the context of the military, cyber security, and communications. In the military context, cognitive security is discussed either indirectly through the concept of cognitive warfare, which stands in contrast to cognitive security, or in relation to the goals and

objectives associated with cognitive warfare (Brangetto & Veenendaal 2016; Hung & Hung 2022). In the communications literature (for example Guo *et al*. 2021; Janzen, Orr & Terp 2022a, 2022b; Popescu 2021; Terp & Breuer 2022), cognitive security is naturally associated with the context of fake news, misinformation, and disinformation. In the cyber security context, cognitive security is understood as the ability of security analysts (for example, Andrade *et al*. 2022) or relates to information generated with AI (for example, Huang *et al*. 2023) or manipulated data (Cybenko, Giani & Thompson 2002). It is noteworthy that 14 of the selected articles used the term 'cognitive security' without providing a definition. These findings underscore the need to clearly define the term and fulfill Research goal 1, to 'explore how previous literature has defined or used the term *cognitive security'*.

Enhancing factors of cognitive security

In addition to the varying definitions, the reviewed articles were examined to explore ways to enhance cognitive security. Among these factors, situational awareness received considerable attention in the literature (Andrade & Yoo 2019; Andrade *et al.* 2022; Huang *et al.* 2023), alongside related concepts such as cyber awareness (Brangetto & Veenendaal 2016) and safety awareness (Liang *et al.* 2022; Deng et al. 2023). The exploration of factors enhancing cognitive security also encompassed concepts like identifying and mitigating disinformation (Popescu 2021; Janzen *et al.* 2022a, 2022b; Terp & Breuer 2022), digital literacy (Kozyreva, Lewandowsky & Hertwig 2020; Hung & Hung 2022; Janzen, Orr & Terp 2022) as well as critical thinking (Hung & Hung 2022). Digital literacy (along with other forms of literacy, such as information literacy) and critical thinking can be considered components of metacognitive skills, which are broadly defined as the ability to reflect and regulate one's own thinking processes (Lai 2011). Additionally, the literature highlighted other pertinent factors such as training and education (Danet 2023), prebunking (Roozenbeek & van der Linden 2019), bias mitigation (Doherty 2023), and the inoculation technique (Roozenbeek & van der Linden 2019). These findings fulfill Research goal 3, to 'clarify what enhances cognitive security'.

Situation Awareness and Cognitive Security

Situation awareness, mentioned as an enhancing factor, is a theory developed by Endsley (1995) and refers to a person's understanding of the state of the environment for succeeding in a task in a dynamic situation. More precisely, Endsley (1988, p. 792) has defined situation awareness as follows: "Situation awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Situation awareness consists of three levels. Level 1 involves the perception of elements in the environment; Level 2 entails the comprehension of the current situation; and Level 3 encompasses the projection of future status of the environment's elements.

Situation awareness can indeed be applied in the context of cognitive security. For instance, consider a situation where an individual encounters possible disinformation. In this context, Level 1 refers to the perception of information, which involves basic cognitive functions and abilities (Glisky 2007) such as perception, attention, and memory. These processes contribute to the individual's awareness that perhaps a source is attempting to influence or manipulate him or her, signifying an understanding of the existence of influence and manipulation in general. Level 2 refers to comprehending the situation the individual is facing. In this context, this level requires the individual to possess domain-specific knowledge (or the ability to acquire it), awareness of cognitive biases existing in human thinking, and an acknowledgment of the techniques by which human perception

and behaviour can be manipulated (for example misinformation, disinformation, deep fakes, usage of bots, robots, and AI agents)—as indicated in **Table 2** under 'enhancing factors'. Level 3, in this context, involves the individual understanding the potential aim of the malign influence or manipulation. Subsequently, the individual plans for coping with the situation, utilizing higher-level cognitive functions such as speech and language processing, decision making, organization, coordination, evaluation, and implementation (Glisky 2007). These cognitive processes materialize as actions, such as analysing the situation for potential cognitive bias exploitation or manipulation, searching for reliable sources, and ultimately implementing necessary actions while developing resilience to move forward. On the contrary, a sense of insecurity arises when individuals perceive their inability to cope with a threatening situation (Smith & Lazarus 1993).

However, according to Endsley (1995), the ability to acquire situation awareness differs among individuals even when exposed to identical information input. This variability is assumed to arise from differences in information processing mechanisms, which are influenced by individuals' capabilities, experiences, and training. This perspective aligns with the view that there is a diversity in the ability to detect disinformation among individuals (Leding & Antonio 2019). Also, personality factors and news consumption have been noted to be related to the ability to discern between true and false news (Calvillo *et al.* 2021). Related to these, the systematic literature review also highlighted that digital literacy, critical thinking, training and education, and bias mitigation are factors that could enhance cognitive security.

Overall, the findings of the systematic literature review suggest that cognitive security requires a more comprehensive definition that considers various interpretations of the phenomenon and is grounded in theoretical knowledge. Therefore, next, a working definition and a framework for comprehensive cognitive security is proposed.

Working Definition for Comprehensive Cognitive Security in the Digital Era

Based on the theoretical background, systematic literature review, concept analysis, and Endsley's (1995) situation awareness, a working definition for comprehensive cognitive security is proposed. Cognition, which involves fundamental functions such as attention, perception, and memory, plays a crucial role in shaping human activity and behaviour. According to the appraisal theory, individuals assess the situations they encounter, and this assessment triggers emotions in response to the event (Scherer 1999). Hence, emotions and cognitive processes are closely intertwined (Ochsner & Phelps, 2007). Additionally, security, as defined by Zedner (2003), involves both a state and a process. The state refers to being without threats, being protected from threats, and avoiding or not exposing oneself to danger. The process involves achieving specific goals, such as objective freedom from risk (although usually not fully achievable) or the subjective feeling of safety. This feeling of safety can be viewed as an outcome of the appraisal process.

Based on the concept analysis, the attributes (characteristics without which the concept cannot occur [Henneman, Lee & Cohen 1995]) of cognitive security include situation awareness, critical thinking, information literacy, reasoning, bias mitigation, training and education, identifying misinformation, disinformation and fake news, identifying malicious content and acts, resilience (that is, "the ability to maintain or regain mental health, despite experiencing adversity" [Herrman *et al*. 2011, p. 258]), and an overall sense of security. These attributes involve diverse knowledge, cognitive skills, and human capabilities. The theory of situation awareness was previously expanded

to include the implementation of actions to develop resilience and progress. Furthermore, as noted by Yang, J *et al*. (2021), resilience facilitates the adoption of secure behaviour.

With this line of thought, comprehensive cognitive security can be defined as *a state and a process in which undesired malign influence or manipulation is incapable of altering human cognition and which can be achieved through a combination of knowledge and situation awareness with purposeful actions.*

In addition to the working definition, a framework of the state and process of cognitive security was also developed (see **Figure 2**).

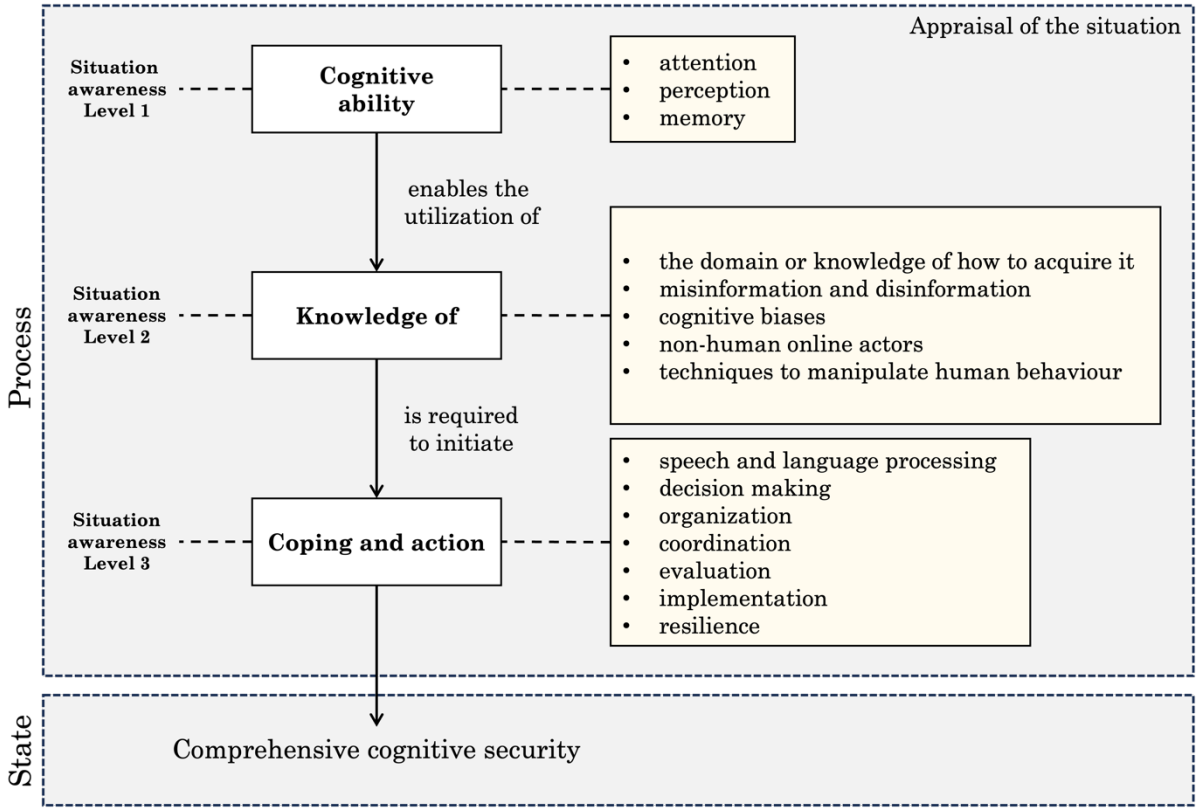


Figure 2: The framework of comprehensive cognitive security

The framework describes a process in which individuals encounter malign influence or manipulation, progress towards understanding and comprehending it, and culminate in a state of comprehensive cognitive security through purposeful actions in the given situation. The presented framework is built upon the situation awareness (Endsley 1995), functions, abilities, and processes of human cognition (Glisky 2007), the appraisal theory (Scherer 1999), and Zedner's (2003) perspective on security. The 'knowledge of' component of cognitive security is drawn from the articles examined during the systematic literature review and reported in **Table 2**, 'enhancing factors. In addition to the enhancing factors, previous research has found that domain knowledge and education have positive impact on detecting disinformation (Zrnec, Poženel & Lavbič 2022), but the effectiveness remains an area of ongoing research (Courchesne *et al*. 2021). Finally, the 'coping and action' component is grounded in Endsley's (1995) theory of situation awareness Level 3, enriched by the incorporation of coping actions and resilience to move forward.

These findings fulfill Research goals 2, to 'clarify the meaning of the concept of cognitive security from a multi-disciplinary perspective' and 5, to 'develop a working definition for comprehensive cognitive security that is meaningful across different disciplines and can be universally understood and applied across research, policy, and practice'.

Practical Implications

The proposed working definition of cognitive security offers a more comprehensive and versatile approach than previous definitions. It builds on existing literature and scientific theories, aiming to define cognitive security without merely describing its functions or its counterpoints. Additionally, this definition is applicable across various disciplines, including national security, psychology, and media studies. Given that many related terms in this field, such as 'misinformation' and 'disinformation', lack universally agreed upon definitions, the concept analysis approach used in this paper, commonly applied in nursing science, could help clarify also these terms and ensure greater consistency in research. A well-defined concept is essential for measurement. This working definition facilitates the operationalization of comprehensive cognitive security, enabling its quantification at both the individual and national levels. Standardized definitions ensure consistency in assessment, making it possible to track cognitive security over time and across different contexts. This, in turn, allows researchers and policymakers to identify trends, vulnerabilities, and key intervention points—such as enhancing information literacy or metacognitive skills—that could improve overall cognitive security. Tracking changes over time can also help determine which factors contribute to its strengthening or decline, ensuring that educational interventions remain effective, and evidence based.

Additionally, understanding cognitive security, including its antecedents and attributes, can support efforts to mitigate the spread of disinformation. For example, this knowledge can inform the development of more sophisticated algorithms capable of identifying not only misleading content but also its psychological impact on users. Such insights could guide the design of educational tools that help users recognize and avoid disinformation while mitigating cognitive biases.

Ultimately, this approach contributes to strengthen cognitive security not only at the individual level but also across entire populations, making it possible that populations are better equipped to navigate an increasingly complex information environment.

Limitations and Further Research

The literature search was exclusively focused on English language, peer-reviewed publications. Incorporating sources, for example, in Finnish or other Nordic languages could have provided additional cultural context. Additionally, articles from organizations such as NATO on cognitive warfare were excluded due to the lack of peer review. However, the inclusion of non-peer-reviewed sources, such as those from NATO, would have compromised the systematic nature of the search. Additionally, alternative publications, beyond peer-reviewed literature, could be explored using different exclusion and inclusion criteria. Additionally, in this review, only the search terms "cogni* security" and "cogni* safety" were utilized. Omitting the quotation marks or adding more search terms would have yielded significantly more results, potentially altering the overall outcome. However, the focus of this article is specifically on understanding how the terms "cognitive security" or "cognitive safety" have been employed in previous English language literature.

For the analysis, both inductive content analysis and concept analysis were employed. Each of these methods has its own potential threats to validity, and different researchers may arrive at varying outcomes based on their approaches and preconceptions. For these reasons, the procedures for conducting the analyses were described in Section 3.2. Despite reporting the systematic literature selection and analyses, it is likely that another group of researchers would synthesize the literature at least slightly differently. The reason behind reporting our research in such detail is not supposed to be taken as steps to reproduce our results (that is, our definition for cognitive security), but to provide transparency to how we reached our definition. For future research, the operationalization and measurement of cognitive security would help societies combat the disinformation in our turbulent information environment.

Conclusions

In this paper, a systematic literature review on cognitive security, a term and phenomenon lacking a commonly understood meaning, was conducted. Its importance is especially evident in the digital era, where the Internet, along with platforms like Facebook, Instagram, X, and TikTok, enable the rapid, effective, and widespread fabrication and dissemination of information. Drawing on theoretical background, systematic literature review, and concept analysis, the authors proposed a working definition for comprehensive cognitive security: *comprehensive cognitive security can be defined as a state and a process in which undesired malign influence or manipulation is incapable of altering human cognition and which can be achieved through a combination of knowledge and situation awareness with purposeful actions.* Moreover, the necessary knowledge and abilities individuals need to pursue comprehensive cognitive security were described. By adopting a comprehensive viewpoint, this revised definition provides a framework that yields insights across various disciplines. It can, for instance, help social media platforms develop measures against disinformation and can support societal efforts to educate citizens on factors influencing individual cognitive security. Educational interventions, such as training in information literacy and metacognitive skills, can further strengthen individuals' abilities to critically assess information and to recognize manipulation attempts. Ultimately, this contributes to strengthening cognitive security at the population level. Additionally, it lays the groundwork for strategic planning aimed at mitigating malign influence at the national level.

Acknowledgements

Grahn's work was funded by the Finland's Ministry of Education and Culture (OKM/93/523/2022).

Declarations of interest

None

References

Allen, C 2017, 'On (not) defining cognition', *Synthese*, vol. 194, no. 11, pp. 4233-49.

Altay, S, Berriche, M, Heuer, H, Farkas J & Rathje, S 2023, 'A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field', *Harvard Kennedy School Misinformation Review*, vol. 4, no. 4.

Andrade, RO, Fuertes, W, Cazares, M, Ortiz-Garcés, I & Navas, G 2022, 'An exploratory study of cognitive sciences applied to cybersecurity', *Electronics*, vol. 11, no. 11, p. 1692.

—& Yoo, SG 2019, 'Cognitive security: A comprehensive study of cognitive science in cybersecurity', *Journal of Information Security and Applications*, vol. 48, p. 102352.

Bar-Tal, D & Jacobson, D 1998, A psychological perspective on security', *Applied Psychology*, vol. 47, no. 1, pp. 59-71.

Bastick, Z 2021, 'Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation', *Computers in Human Behavior*, vol. 116, p. 106633.

Brangetto, P & Veenendaal, MA 2016, 'Influence cyber operations: The use of cyberattacks in support of influence operations', *Proceedings of the International Conference on Cyber Conflict, CYCON*, pp. 113-26.

Brooks, DJ 2010, 'What is security: Definition through knowledge categorization', *Security Journal*, vol. 23, pp. 225-39.

Buzzell, A W 2024, *The ethics of cognitive security*, York University, Toronto, Canada.

Calvillo, DP, Garcia, RJB, Bertrand, K & Mayers, TA 2021, 'Personality factors and self-reported political news consumption predict susceptibility to political fake news', *Personality and Individual Differences*, vol. 174, p. 110666.

Cao, J, Lin, L, Zhang, J, Zhang, L, Wang, Y & Wang, J 2021, 'The development and validation of the perceived safety of intelligent connected vehicles scale', *Accident Analysis and Prevention*, vol. 154, p. 106092.

Courchesne, L, Ilhardt, J & Shapiro, JN 2021, 'Review of social science research on the impact of countermeasures against influence operations' *Harvard Kennedy School Misinformation Review,* vol. 2, no. 5.

Cybenko, G, Giani, A & Thompson, P 2002, 'Cognitive hacking: A battle for the mind', *Computer*, vol. 35, no. 8, pp. 50-6.

Danet, D 2023, 'Cognitive security: Facing cognitive operations in hybrid warfare', *Proceedings of the European Conference on Information Warfare and Security, ECCWS*, pp. 161-8.

Deng, S, Cai, Y, Xie, L & Pan, Y 2023, 'Group management model for construction workers' unsafe behavior based on cognitive process model', *Engineering, Construction and Architectural Management*, vol. 30, no. 7, pp. 2928-46.

—, Peng, R & Pan, Y 2022, 'A cognitive failure model of construction workers' unsafe behavior', *Advances in Civil Engineering*, vol. 2022, no. 1, p. 2576600.

Doherty, G 2023, 'Cognitive security: An architecture informed approach from cognitive science', *Proceedings of the International Conference on Human-Computer Interaction*, pp. 395-415.

Dolan, RJ 2002, 'Emotion, cognition, and behavior', *Science*, vol. 298, no. 5596, pp. 1191-4.

Elo, S & Kyngäs, H 2008, 'The qualitative content analysis process', *Journal of Advanced Nursing*, vol. 62, no. 1, pp. 107-15.

Dubois, E & Blank, G 2018,'The echo chamber is overstated: The moderating effect of political interest and diverse media', *Information, communication & society*, vol. 21, no. 5, pp. 729-45.

Endsley, MR 1988, 'Situation awareness global assessment technique (SAGAT)', *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference,* pp. 789-95.

—1995, 'Toward a theory of situation awareness in dynamic systems', *Human Factors*, vol. 37, no. 1, pp. 32-64.

Evans, J 2022, 'War in the age of TikTok', *Russian Analytical Digest*, vol. 280, pp. 17-19.

Frith, CD 2008, 'Social cognition', *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 363, no. 1499, pp. 2033-9.

Glisky, E 2007, 'Changes in cognitive function in human aging', *Brain aging: Models, methods, and mechanisms*, ed. DR Riddle, CRC Press, Boca Raton, FL, US, pp. 3-20.

Gorodnichenko, Y, Pham, T & Talavera O 2021, 'Social media, sentiment and public opinions: Evidence from #Brexit and #USElection', *European Economic Review*, vol. 136, p. 103772.

Guo, B, Ding, Y, Sun, Y, Ma, S, Li, K & Yu, Z 2021, 'The mass, fake news, and cognition security', *Frontiers of Computer Science*, vol. 15, pp. 1-13.

Hanoch, Y & Wood, S 2021, 'The scams among us: Who falls prey and why', *Current Directions in Psychological Science*, vol. 30, no. 3, pp. 260-6.

Henneman, EA, Lee, JL & Cohen, JI 1995, 'Collaboration: A concept analysis', *Journal of Advanced Nursing*, vol. 21, no. 1, pp. 103-9.

Herrman, H, Stewart, DE, Diaz-Granados, N, Berger, EL, Jackson, B & Yuen, T 2011, 'What is resilience?', *The Canadian Journal of Psychiatry*, vol. 56, no. 5, pp. 258-65.

Huang, R, Zheng, X, Shang, Y & Xue, X 2023, 'On challenges of AI to cognitive security and safety', *Security and Safety*, vol. 2, p. 2023012.

Hung, TC & Hung, TW 2022, 'How China's cognitive warfare works: A frontline perspective of Taiwan's anti-disinformation wars', *Journal of Global Security Studies*, vol. 7, no. 4, p. ogac016.

Janzen, S, Orr, C & Terp, SJ 2022a, 'Building cognitive security and resilience to enhance crisis preparedness: A social ecological approach to minimizing information-based harms', *Proceedings of the SIG SI & IEP 2022: The 18th Annual Social Informatics Research Symposium and the 4th Annual Information Ethics and Policy Symposium: Resilient Sociotechnical Systems for Social Good*.

—2022b 'Cognitive security and resilience: A social ecological model of disinformation and other harms with applications to COVID-19 vaccine information behaviors', *Proceedings of the CEUR Workshop*, pp. 48-88.

Kozyreva, A, Lewandowsky, S & Hertwig, R, 2020 'Citizens versus the Internet: Confronting digital challenges with cognitive tools', *Psychological Science in the Public Interest*, vol. 21, no. 3, pp. 103-56.

Kruger, A, Saletta, M, Ahmad, A & Howe, P 2024, 'Structured expert elicitation on disinformation, misinformation, and malign influence: Barriers, strategies, and opportunities', *Harvard Kennedy School Misinformation Review*, vol. 5, no. 7.

Lazarus, RS 1991, 'Cognition and motivation in emotion', *American Psychologist*, vol. 46, no. 4, p. 352.

Lai, ER 2011, 'Metacognition: A literature review', *Pearson Research Reports,* vol. 24, pp. 1-40.

Leding, JK & Antonio, L 2019, 'Need for cognition and discrepancy detection in the misinformation effect', *Journal of Cognitive Psychology*, vol. 31, pp. 409-15.

Liang, Q, Zhou, Z, Ye, G & Shen, L 2022, 'Unveiling the mechanism of construction workers' unsafe behaviors from an occupational stress perspective: A qualitative and quantitative examination of a stress-cognition-safety model', *Safety Science*, vol. 145, p. 105486.

Linvill, DL, Boatwright, BC, Grant, WJ & Warren, PL 2019, '"THE RUSSIANS ARE HACKING MY BRAIN!" investigating Russia's Internet research agency Twitter tactics during the 2016 United States presidential campaign', *Computers in Human Behavior*, vol. 99, pp. 292-300.

Loh, KK & Kanai, R 2016, 'How has the Internet reshaped human cognition?', *Neuroscientist*, vol. 22, no, 5, pp. 506-20.

Lovari, A 2020, 'Spreading (Dis)trust: Covid-19 misinformation and government intervention in Italy', *Media and Communication*, vol. 8, no. 2, pp. 458-61.

McCombie, S, Uhlmann, AJ & Morrison, S 2020, 'The US 2016 presidential election & Russia's troll farms', *Intelligence and National Security*, vol. 35, no. 1, pp. 95-114.

Moors, A, Ellsworth, PC, Scherer, KR & Frijda, NH 2013, 'Appraisal theories of emotion: State of the art and future development', *Emotion Review*, vol. 5, no. 2, pp. 119-24.

Nahum, A, Drekonja, DM & Alpern, JD 2021, 'The erosion of public trust and SARS-CoV-2 vaccines—More action is needed', *Open Forum Infectious Diseases*, vol. 8, no. 2.

Neisser, U 2014, *Cognitive psychology: Classic edition*, Psychology Press, New York, NY, US.

Newell, A, Shaw, JC & Simon, HA 1958, 'Elements of a theory of human problem solving', *Psychological Review*, vol. 65, no. 3, p. 151.

Nisbett, RE & Wilson, TD 1977, 'Telling more than we can know: Verbal reports on mental processes', *Psychological Review*, vol. 84, no. 3, pp. 231-59.

Ochsner, KN & Phelps, E 2007, 'Emerging perspectives on emotion-cognition interactions', *Trends in Cognitive Sciences*, vol. 11, no. 8, pp. 317-8.

Ogiela, L & Ogiela, MR 2020, 'Cognitive security paradigm for cloud computing applications', *Concurrency and Computation: Practice and Experience,* vol. 32, no. 8, p. e5316.

Ognyanova, K, Lazer, D, Robertson, RE & Wilson, C 2020, 'Misinformation in action: Fake news exposure is linked to lower trust in media, higher trust in government when your side is in power', *Harvard Kennedy School Misinformation Review*, vol. 1, no. 4.

Olan, F, Jayawickrama, U, Arakpogun, EO, Suklan, J & Liu, S 2022, 'Fake news on social media: The impact on society', *Information Systems Frontiers,* vol. 26, no. 2, pp. 443-58.

Pessoa, L 2008, 'On the relationship between emotion and cognition', *Nature Reviews Neuroscience*, vol. 9, no. 2, pp. 148-58.

Popescu, MM 2021, 'Disinformation deconstructed—Cognition security and digital control', *Bulletin of the Transilvania University of Braşov, Series VII: Social Sciences and Law*, vol. 14, pp. 123-34.

Prasetio, EA & Nurliyana, C 2023, 'Evaluating perceived safety of autonomous vehicle: The influence of privacy and cybersecurity to cognitive and emotional safety', *IATSS Research*, vol. 47, no. 2, pp. 160-70.

Pulsford, D, Duxbury, J & Carter, B 2016, 'Personal qualities necessary to care for people with dementia', *Nursing standard*, vol. 30, no. 37.

Rajtmajer, S & Susser, D 2020, 'Automated influence and the challenge of cognitive security', *Proceeding of the 7th Symposium on Hot Topics in the Science of Security ACM International*, pp. 1-9.

Roiser, JP, Nathan, PJ, Mander, AP, Adusei, G, Zavitz, KH & Blackwell, AD 2016, 'Assessment of cognitive safety in clinical drug development', *Drug Discovery Today*, vol. 21, no. 3, pp. 445-53.

Roozenbeek, J & van der Linden, S 2019, 'Fake news game confers psychological resistance against online misinformation', *Palgrave Communications*, vol. 5, no. 1, pp. 1-10.

Scherer, KR 1999, 'Appraisal theory', *Handbook of cognition and* emotion, T Dalgleish & M J Power (eds), Wiley, Chichester, UK, pp. 637-63.

Silva, JAH & Hernandez-Alvarez, M 2017, 'Large scale ransomware detection by cognitive security', *Proceedings of the 2017 IEEE 2nd Ecuador Technical Chapters Meeting, ETCM 2017*, pp. 1-4.

Smith, CA & Lazarus, RS 1993, 'Appraisal components, core relational themes, and the emotions', *Cognition and Emotion*, vol. 7, no. 3-4, pp. 233-69.

Starbird, K, Arif, A & Wilson, T 2019, 'Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations', *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, pp. 1–26.

Tashev, B, Purcell, M & McLaughlin, B 2019, 'Russia's information warfare: Exploring the cognitive dimension', *MCU Journal*, vol. 10, no. 2, pp. 129-47.

Terp, SJ & Breuer, P 2022, 'DISARM: A framework for analysis of disinformation campaigns', *Proceedings of the 2022 IEEE International Conference on Cognitive and Computational Aspects of Situation Management, CogSIMA 2022*, pp. 1-8.

Tofthagen, R & Fagerstrøm, LM 2010, 'Rodgers' evolutionary concept analysis—A valid method for developing knowledge in nursing science', *Scandinavian Journal of Caring Sciences*, vol. 24, pp. 21-31.

Waltzman, R, 2017, *The Weaponization of Information: The need for cognitive security*, viewed 22 August 2023, <https://www.rand.org/pubs/testimonies/CT473.html>.

Wang, J, Long, R, Chen, H & Li, Q 2019, 'Measuring the psychological security of urban residents: Construction and validation of a new scale', *Frontiers in Psychology*, vol. 10, p. 2423.

Windle, G 2011, 'What is resilience? A review and concept analysis', *Reviews in Clinical Gerontology*, vol. 21, no. 2, pp. 152-69.

Wintterlin, F, Schatto-Eckrodt, T, Frischlich, L, Boberg, S, Reer, F & Quandt, T 2023, '"It's us against them up there": Spreading online disinformation as populist collective action', *Computers in Human Behavior*, vol. 146, p. 107784.

World Economic Forum 2024, *Global Risks Report 2024*, viewed 10 January 2024, <https://www.weforum.org/publications/global-risks-report-2024/>.

Yang, J, Ye, G, Xiang, Q, Kim, M, Liu, Q & Yue, H 2021, 'Insights into the mechanism of construction workers' unsafe behaviors from an individual perspective', *Safety Science*, vol. 133, p. 105004.

Yang, KC, Pierri, F, Hui, PM, Axelrod, D, Torres-Lugo, C, Bryden, J & Menczer, F 2021, 'The COVID-19 infodemic: Twitter versus Facebook', *Big Data and Society*, vol. 8, no. 1, p. 20539517211013861.

Yeh, HY 2016, 'Using an awakening narrative to leave behind a former national-identity: An investigation of the conversion of national-identity in Taiwan', *Nations and Nationalism*, vol. 22, no. 3, pp. 542-60.

Zedner, L 2003, 'The concept of security: An agenda for comparative analysis', *Legal Studies*, vol. 23, no. 1, pp. 153-76.

—2009, *Security*, Routledge.

Zrnec, A, Poženel, M & Lavbič, D 2022, 'Users' ability to perceive misinformation: An information quality assessment approach', *Information Processing and Management*, vol. 59, no. 1, p. 102739.