Computers & Security 134 (2023) 103435

Contents lists available at ScienceDirect

Computers & Security

journal homepage: www.elsevier.com/locate/cose

Learning from safety science: A way forward for studying cybersecurity incidents in organizations

Nico Ebert [a,*], Thierry Schaltegger [a], Benjamin Ambuehl [a], Lorin Schöni [b], Verena Zimmermann [b], Melanie Knieps [c]

[a] ZHAW School of Management and Law, Institute of Business Information Technology, Theaterstrasse 17, 8400 Winterthur, Switzerland
[b] ETH Zurich, Professorship for Security, Privacy & Society D-GESS, Stampfenbachstrasse 69, 8006 Zurich, Switzerland
[c] University of Zurich, Digital Society Initiative, Rämistrasse 69, 8001 Zurich, Switzerland

ARTICLE INFO

Keywords:
Cybersecurity
Incident
Safety science
Human error

ABSTRACT

In the aftermath of cybersecurity incidents within organizations, explanations of their causes often revolve around isolated technical or human events such as an Advanced Persistent Threat or a "bad click by an employee." These explanations serve to identify the responsible parties and inform efforts to improve security measures. However, safety science researchers have long been aware that explaining incidents in socio-technical systems and determining the role of humans and technology in incidents is not an objective procedure but rather an act of social constructivism: what you look for is what you find, and what you find is what you fix. For example, the search for a technical "root cause" of an incident might likely result in a technical fix, while from a sociological perspective, cultural issues might be blamed for the same incident and subsequently lead to the improvement of the security culture. Starting from the insights of safety science, this paper aims to extract lessons on what general explanations for cybersecurity incidents can be identified and what methods can be used to study causes of cybersecurity incidents in organizations. We provide a framework that allows researchers and practitioners to proactively select models and methods for the investigation of cybersecurity incidents.

1. Introduction

Prevailing explanations put forth by organizations for successful cyberattacks typically revolve around a specific threat, human error or technical failure. Organizations are believed to be "one bad click away" from a cybersecurity incident (Ahmad et al., 2012, Banga, 2020, Tøndel et al., 2014). While being intuitive on the surface, attributing the cause of an incident to seemingly obvious explanations (e.g., human error) has been criticized as artificial and trivial (Leveson, 2016, Canfield and Fischhoff, 2018, Lipner and Pescatore, 2023) and can result in the blame being placed on individual employees (Renaud et al., 2021). However, this observation raises questions about more suitable, alternative explanations for a successful cyberattack, the accountability of people involved, and the effectiveness of measures to prevent similar incidents in the future. These questions have not yet been adequately addressed in the cybersecurity discourse. Although new ideas, such as the role of the "security culture" (von Solms, 2000), have recently been emerging, there is still no clear idea of an approach to systematically studying these cyber incidents. Given that explaining and analyzing incidents in socio-technical systems is not an objective procedure (Woods et al., 1994), but rather a social constructivist activity in which the perspective of the investigator shapes what is ultimately found and fixed (Catino, 2008, Heraghty et al., 2018, Lundberg et al., 2009), awareness about the implicit assumptions underlying an explanatory model is crucial to ensure the validity of the resulting conclusions. However, the theoretical foundation and empirical toolbox of examining these events is meager at best.

The purpose of this narrative review in the field of incident causation is to demonstrate similarities between the emerging field of cybersecurity and the established discipline of safety science, especially regarding general explanations and corresponding approaches to understanding cyber incidents. We also seek to provide a framework for analyzing cyber incidents, allowing cybersecurity researchers and practitioners to choose an explanatory model that more actively aligns with their goals, to understand its limitations, and to anticipate and prevent unintended consequences. By establishing a sound, evidence-

* Corresponding author.
E-mail address: nico.ebert@zhaw.ch (N. Ebert).

https://doi.org/10.1016/j.cose.2023.103435
Received 2 June 2023; Received in revised form 15 July 2023; Accepted 14 August 2023
Available online 19 August 2023

based foundation that equips researchers and practitioners with a greater awareness of their options, we aim to facilitate more informed and systematic research agendas in the emerging field of cybersecurity.

This article is structured as follows: in the second section, we show parallels between safety science and cybersecurity, look at the history of safety science, and show that safety science concepts can also be found in cybersecurity. In the third section, we show various models and methods from safety science that can contribute to a better understanding of cybersecurity incidents. In section four, we ask how the models can be applied and when which model is suitable. Finally, we summarize the findings in the fifth chapter and consider the need for further research.

2. From the First to the Fourth Industrial Revolution: how 20th century safety debates resemble ongoing cybersecurity challenges

At first glance, security, cybersecurity and safety appear to be distinct concepts. While security is typically associated with deliberate events, safety[1] focuses on accidental incidents (Herrmann and Pridöhl, 2020). In cybersecurity practice, this distinction is of little use because successful cyberattacks frequently occur due to a combination of both deliberate, non-random events like Advanced Persistent Threats (APTs), and unintentional events such as failure to adhere to security policies. Indeed, it is important to bear in mind that security and safety rely on the idea that risks can be mitigated, regardless of whether they arise from intentional or unintentional causes (Amundrud et al., 2017, Aven, 2014). Researchers have argued that cybersecurity can benefit greatly from insights derived from safety science (Bair et al., 2017, Brostoff and Sasse, 2001, Young and Leveson, 2013, Zimmermann, 2023), and cybersecurity researchers have frequently adopted concepts from safety science (e.g., the Swiss Cheese Model (Lawrence Pfleeger et al., 2014)).

The origins of safety science can be traced back to the beginning of the 20th century, when the First Industrial Revolution transformed how work was done (Dekker, 2019). The evolving work environment, which required employees to operate new heavy equipment, led to novel types of incidents. To understand and address the causes of these incidents, safety science emerged as a discipline that has since undergone various paradigm shifts (Dekker, 2019). Today, safety science is an interdisciplinary field that draws on social science, psychology, population health, physical sciences, and engineering to study incidents and their prevention. It encompasses multiple research traditions and has a broad impact, including in fields such as software engineering and cyber-physical systems. The digital transformation, or Fourth Industrial Revolution (WEF 2016), is now once again redefining how work is performed in contemporary societies, and is once again leading to novel types of incidents. Efforts to explain these novel types of incidents using previous methods have been further complicated due to the unprecedented speed at which the complexity of socio-technical systems is advancing in the digital era (Leveson, 2016). The difference is that this time around, we can call upon the blueprints of a century of safety research to help us to proactively address the challenges associated with cybersecurity.

Many ongoing debates in cybersecurity resemble those that have already been discussed at length in safety science (see Table 1) (Dekker, 2019). One example is the concept of "human error", which dates back to the 1930s. Human errors are a common theme in the contemporary discourse on cybersecurity (Schneier, 2015, Vroom and von Solms, 2004, Yan et al., 2018), in which humans are often referred to as "the weakest link". For example, some Chief Information Security Officers

[1] There is no commonly agreed definition among safety scientists about what safety is (Aven, 2014, Hollnagel, 2014). For example, while in its most simple definition safety is defined as the absence of incidents, in a more complex definition safety is seen as a social construct ("safe operations").

Table 1
Chronological development of concepts in safety science (Dekker, 2019) and today's counterparts in cybersecurity.

Original concept in safety science	Today's concepts in cybersecurity
"Taylor and Proceduralization" (the 1910s and later) Human operators were framed as a problem to be controlled by enforcing compliance with rules and penalizing violations.	Humans as the "weakest link" in cybersecurity (Schneier, 2015, Vroom and von Solms, 2004, Yan et al., 2018) and compliance as a model of appropriate behavior (Cichonski et al., 2012, Moody et al., 2018, Siponen and Vance, 2010)
"Accident-prone" (the 1920s and later) Emerging sciences like psychology and eugenics attributed incidents to individual predispositions (while neglecting context) to identify "bad apples".	Insecure personalities (Canfield and Fischhoff, 2018, Parrish et al., 2009, Wright and Marett, 2010) and individuals' susceptibility to threats such as social engineering (Uebelacker and Quiel, 2014)
"Heinrich and Behavior-Based Safety" (the 1930s and later) Incidents were explained as linear causal chains ("dominos") with a single root cause. At first, this approach focused on improving technical aspects and later on eliminating unsafe behavior (human error, behaviorism).	Analyzing root causes (Dolezel and McLeod, 2019, Wangen et al., 2017), attack vectors (Landwehr et al., 1994), human errors (Kraemer et al., 2009, Liginlal et al., 2009, Wood and Banks, 1993), lessons learned (He et al., 2014) or behaviors (Abroshan et al., 2021) based on causal chains
"Human Factors and Cognitive Systems" (the 1940s and later) The focus shifted from human error to the operator's context (e.g., technology, tools, and working environments). Tools and machines used by the operator were redesigned based on mental and social phenomena.	Approaches such as the People, Processes, and Technology Framework (Andress, 2003), Usable Security (Cranor and Garfinkel, 2005, Sasse et al., 2002, Angela Sasse and Flechais, 2005), Security by Design (Duncan, 2020), and Situation Awareness (Ahmad et al., 2012, Ofte and Katsikas, 2023)
"System Safety" (the 1950s, 1960s, and later) Safety was holistically considered as an aspect of socio-technical systems with its interrelationships. The focus shifted from assuring the absence of events to providing the capacity to handling unforeseen events.	Cybersafety (Khan et al., 2023) and Systems Thinking for Safety and Security (Young and Leveson, 2013) approaches
"Man-Made Disasters" (the 1970s and later) Incidents were increasingly understood as social and organizational phenomena rather than engineering problems. Latent problems were thought to accumulate in an incubation phase before actual incidents (e.g., the Three Mile Island).	Security as an organizational problem (Dhillon and Backhouse, 2000); information security controls (Baker and Wallace, 2007) and phishing as a problem of organizational norms (Petrič and Roer, 2022)
Normal Accidents and High-Reliability Organizations (the 1980s and later) Normal Accidents Theory (NAT) suggested that incidents are normal and predictable in complex socio-technical systems. However, other socio-technical systems did not generate incidents and were thought to have specific characteristics ("high-reliability theory").	Zero Trust Architectures (Bush and Mashatan, 2023) (which views breaches as normal incidents), high-reliability theory in information systems, and redundancy in security (Littlewood and Strigini, 2004)
Swiss Cheese and Safety Management Systems (the 1990s and later) The Swiss cheese model conceptualized influences in incidents on different levels, ranging from latent organizational forces to active, unsafe behavior. This was conceptuality aligned with administrative safety management ("safety bureaucracy and culture of compliance").	Technical defenses (Al-Shaer and Hamed, 2004), the NIST Computer Security Incident Handling Guide (Cichonski et al., 2012), information security governance (von Solms, 2006, Da Veiga and Eloff, 2007), and information security management systems (e.g., ISO 27001 (ISO 2023))
Safety Culture (the 2000s and later) Encouraging organizations and leaders to build a "good" or "just" safety culture rather than focusing on things that cannot be fixed (with limited consensus on what safety culture is).	Information security culture (von Solms, 2000), climate (Goo et al., 2014) and Just Culture in cybersecurity (Pollini et al., 2022)

(*continued on next page*)

Table 1 (*continued*)

Original concept in safety science	Today's concepts in cybersecurity
Resilience Engineering (the 2010s and later) Enhancing capacities of people and organizations that allowed them to adapt under varying circumstances instead of reducing negative outcomes. Resilience engineering was inspired by organic systems and their adaptive capacity.	The view of people as an adaptive cyber capability is incorporated into Cyber Resilience (Björck et al., 2015), Cyber Resilient Behavior (Kleij and Leukfeldt, 2020), Systems (Ross et al., 2021), and Management (Christine and Thinyane, 2022)

(CISOs) blame employees and their non-compliance with security policies - rather than inadequate security technology - for corporate security incidents (Hielscher et al., 2023). In other words, the current cybersecurity situation exhibits pronounced similarities to the approaches that were prevalent in safety science until the mid-twentieth century, before a paradigm shift towards a more systemic understanding of safety began to replace this view. Therefore, the extensive research conducted in the field of safety science provides valuable lessons that can be also used to inform research on the occurrence of incidents (etiology) in the cybersecurity domain.

3. Lessons from safety science: Exploring models and methods to study cybersecurity incidents

The type of perspective individuals take in investigating causes of incidents (e.g., hierarchical level of analysis (Leveson, 2011)) defines where they look for explanations and determines the conclusions they will ultimately draw. People's explanatory models are seldom made explicit (Lawrence Pfleeger et al., 2014) and are often only implicitly invoked (Cichonski et al., 2012). While there is never only one "true" or "correct" theoretical explanation for an incident (Hollnagel, 2009:12), it is very important to be aware of one's underlying explanatory model, as this significantly shapes the reasons that are identified, the groups of people that are held accountable, and the measures that are subsequently taken (Catino, 2008, Lundberg et al., 2009). The predominant examination of incidents through a technical lens (e.g., technical system failures), will inevitably prompt the search for technical solutions such as "strengthening network perimeter security" for preventing future incidents (see, for example, the NIST Computer Security Incident Handling Guide (Cichonski et al., 2012)). By highlighting possible alternatives, we expand the toolbox available to researchers and professionals seeking to analyze and explain cyber incidents.

As demonstrated in Table 2, four types of general incident models and methods[2] have emerged from safety science, each with their own implicit assumptions about the origins of incidents (Khanzode et al., 2012, Leveson, 2016). Single-factor models attribute incidents to one distinct event (e.g., an attack, a technical or human error), while linear models assume failures at multiple (but interconnected) levels. As socio-technical systems became more complex, more holistic causation models were developed to replace previous post hoc descriptions of incidents by cause-effect modelling (Qureshi, 2008). Cultural and Management Models focused on how organizational culture and management contribute to incidents or prevent them, while systemic models viewed incidents as sometimes inevitable and attributed their occurrence and severity to inadequate control systems or insufficient organizational resilience. Importantly, these later causation models largely disregarded individual human errors as the primary cause of incidents.

[2] For an overview of models, see (Fu et al., 2020); for an overview of methods, see (Sklet, 2004); for a mapping between models and methods, see (Katsakiori et al., 2009).

[3] We do not consider statistical and mathematical incident models and methods (e.g., as used to model road traffic accidents (Abdulhafedh, 2017)).

The following section will highlight (a) how the implicit assumptions underlying each explanatory model shape the analysis of incidents, and (b) what method(s) correspond with each approach.

3.1. Single-factor and linear models

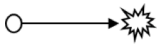
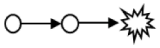
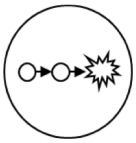
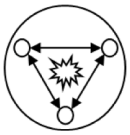
The strength of single-factor and linear models lies in their simplicity. They are particularly useful for scrutinizing incidents caused by failures of technical components or human errors in simple socio-technical systems. However, it is important to recognize that they are not infallible and may not always include all contributing factors to an incident. For example, their capacity to explain incident causation in more complex systems is limited (Qureshi, 2008). Historically, these models often label what cannot be attributed to technical failures as "human error." (Hollnagel, 2001).

Single-factor models involve using a straightforward technical or human factor as the root cause of an incident (e.g., sharp edges of a machine). When these models were first developed in the early 20th century, the human involved in the incident was often deemed solely responsible (Safety thinking and safety methods 2023). Today, also the idea of blaming a single factor for major cybersecurity incidents such as APTs or human error is prevalent in cybersecurity (Banga, 2020, Lipner and Pescatore, 2023). This mindset may be the reason behind the common usage of the "weakest link" phrase (Arce, 2003, Schneier, 2015, Vroom and von Solms, 2004, Yan et al., 2018), which is typically associated with human error (Kraemer et al., 2009, Liginlal et al., 2009, Moody et al., 2018, Siponen and Vance, 2010, Wood and Banks, 1993). Conversely, a single technical component or heroic action may be credited with preventing a security incident (Lawrence Pfleeger et al., 2014).

Linear models operate on the assumption that incidents occur because of a chain of events and despite lines of defenses. At the beginning of this chain of events lies a root cause or multiple causes triggering a sequence of events ultimately resulting in the incident (Underwood and Waterson, 2013). These models originate from the field of industrial safety, and aim to protect workers from injuries or illnesses. For instance, Heinrich's domino model from 1931 suggests that a single cause sets off a linear progression of events leading to an incident, and stopping any one event from happening would prevent it (Heinrich, 1931, Lehto and Salvendy, 1991). In more complex linear models, also known as "epidemiological models," causes may not always be manifest but instead may be latent, and "defenses" or "barriers" may prevent incidents from happening, like a functioning immune system. Some researchers have argued that organizational incidents do not arise solely due to a single human error but instead result from the interconnectedness of several latent factors originating at many levels within an organization (Qureshi, 2008). In the "Swiss Cheese Model," an incident may also have latent organizational factors that are difficult to observe, such as time pressure, understaffing, or inadequate equipment, that only become visible when combined with active failures (James Reason 1990). Each layer of defense, such as technical safety measures, is situated between hazards and incidents. In this model, the layers are represented by slices of Swiss cheese, and if flaws (holes) in each layer align, an incident may occur.

The notion of a singular root cause of incidents that can be fixed is also popular in cybersecurity. For example, a recent study on security breaches found that organizations were generally able to determine the cause of, and fix, the "weakness" (Department for Digital, Culture, Media and Sport 2022). The Cyber Kill Chain is based on a linear model that shows the chain of offensive actions of cyber attackers (Yadav and Rao, 2015). The idea of using technical or human defenses, such as a firewall, to halt the negative chain of events is also present (Al-Shaer and Hamed, 2004, Mosteiro-Sanchez et al., 2020). For example, NIST's Computer Security Incident Handling Guide conceptualizes incidents as "a violation or imminent threat of violation of computer security policies, acceptable use policies, or standard security practices" (Cichonski

Table 2
General incident models and methods[3] (based on (Qureshi, 2008, Safety thinking and safety methods 2023)).

Model	 Single-factor Models	 Linear Models (e.g., Domino, Swiss Cheese)	 Cultural & Management Models	 Systemic Models Control Models (e.g., STAMP)	Resilience Models (e.g., FRAM)
Perspectives on incidents	Incidents result from a single event (e.g., attack, technical or human error).	Incidents result from a series of linked events (e.g., combined technical and human errors)	Incidents result from the cultural problems and management failures of an organization.	Incidents result from dysfunctional interactions among system components that were not adequately handled by control loops.	Incidents are an inevitable part of operations and result from the inability of a system to adapt to changing hazards.
Illustration of model	A successful cyberattack is the result of insecure behavior (e.g., "bad click"). It can be prevented by stopping the behavior.	A successful cyberattack is the result of a chain of events and has a root cause. It can be prevented by eliminating the root cause and by setting up barriers to prevent events.	A successful cyberattack is the result of a broken security culture or ineffective management. It can be prevented with a better security culture (e.g., "just culture") or management (e.g., governance, risk management)	A successful cyberattack is the result of a control problem (e.g., mismanagement). It can be prevented by enforcing security constraints.	A successful cyberattack is the result of an organization's limited capacity to adapt. A resilient system has capacities to function despite a cyberattack.
Examples of model usage in cyber-security	Single threats • APT (Lipner and Pescatore, 2023) Human factors • Malicious Insider (Johnston et al., 2016) • Human error (Kraemer et al., 2009, Liginlal et al., 2009, Moody et al., 2018, Siponen and Vance, 2010, Wood and Banks, 1993) and weakest link (Arce, 2003, Schneier, 2015, Vroom and von Solms, 2004, Yan et al., 2018) • Heroism (Lawrence Pfleeger et al., 2014) • Human as a security sensor (Heartfield and Loukas, 2018, Heartfield et al., 2016, Jensen et al., 2022, Stembert et al., 2015) • Human as a Solution (Zimmermann, 2023) Technical factors • Attack vectors (Landwehr et al., 1994) (e.g., missing 2FA)	Linear Models • Cyber Kill Chain (Yadav and Rao, 2015) • Attack surface (Banga, 2020) • Attack fault trees (Fovino et al., 2009) • Root cause analysis (Department for Digital, Culture, Media and Sport 2022) • Lessons learned (He et al., 2014) Human defenses • Humans as the weakest link (Arce, 2003, Schneier, 2015, Vroom and von Solms, 2004, Yan et al., 2018) • Security Policies (Cichonski et al., 2012, Moody et al., 2018, Siponen and Vance, 2010) Technical defenses • Defence-in-Depth (Mosteiro-Sanchez et al., 2020) • Firewall (Al-Shaer and Hamed, 2004)	Cultural Models • Organizational culture and security compliance (Hu et al., 2012) • Security culture (von Solms, 2000) and climate (Goo et al., 2014) Management Models • Information security management (e.g., ISO 27001 (ISO 2023)) • Security governance (von Solms, 2006, Da Veiga and Eloff, 2007)	• Cybersafety (Khan et al., 2023) • System Dynamics in Social Engineering (Greitzer et al., 2014)	• Cyber Resilient Systems (Ross et al., 2021) • Cyber resilience management frameworks (Christine and Thinyane, 2022)
Methods to analyze incidents	• Individual level: Decomposition (e.g., Root cause analysis, chain of events, 5 Whys) • Population-level: Epidemiological Methods (e.g., Experiments, Cohort studies, Cross-sectional study, Cross Case Analysis)		Sociological, organizational, and anthropological methods (e.g., Cultural Immersion, Survey Studies, Case Studies).	• Safety: STAMP, STPA, CAST, AcciMap • General: Simulation, Cybernetics, System Dynamics	• Safety: FRAM, Resilience Analysis Grid
Illustration of method	• Individual level: After a phishing attack in an organization, a root cause analysis is performed to understand what went wrong. • Population-level: Longitudinal, observational study of employees' behavior related to phishing emails (prospective cohort study)		• Interviews with different employee groups (end users, IT staff, management) are conducted to understand the security culture after a large phishing incident	• Control and feedback loops between information security officers and employees are studied (e.g., effective phishing reporting by employees)	• Studying employees' ability to perform work despite incidents • Studying the capabilities of the IT staff to handle security threats
Examples of method usage in cybersecurity	• Individual level ○ Root cause analysis (Dolezel and McLeod, 2019, Sundaramurthy et al., 2014, Wangen et al., 2017) ○ Attack Fault Trees (Fovino et al., 2009) • Population-level ○ Phishing susceptibility (Lain et al., 2022) (experiment) ○ Security breaches in organizations (Department for Digital, Culture, Media and Sport 2022) (pseudo-longitudinal) ○ IT investment and data breaches (Li et al., 2023) (retrospective cohort study)		• Cultural immersion in SOC teams (Sundaramurthy et al., 2014, Sundaramurthy et al., 2016) • Security climate cross-sectional survey (Goo et al., 2014) • After-breach case studies (Department for Digital, Culture, Media and Sport 2022)	• STPA-Sec (Schmittner et al., 2016, Young and Leveson, 2013) • Cybersafety (Khan et al., 2023)	• FRAM in Security Engineering (Hlaing and Ochimizu, 2018) • System Security Engineering (Ross et al., 2021) • Resilient cyber incident mgmt (Aoyama et al., 2015)

et al., 2012). Safety science approaches following this rationale have led to the development of a handful of methods to determine the root cause, and these might also inform research on cybersecurity.

3.1.1. Methods

Methods to study incidents based on single-factor and linear models can be categorized as those used for studying single incidents and those used for studying multiple incidents in a given population (e.g., annual incident rate among forklift operators, clicks on phishing mails in an employee population). To analyze single incidents, researchers disentangle the functional and behavioral aspects of the socio-technical system (Leveson, 2016, p.61). This procedure involves breaking down the technical system into its individual components in order to identify failures and reconstructing the discrete events over time that led to an incident. The view that these components are operating independently from each other is reflected in this methodological approach (Leveson, 2016, p.61). Common techniques used for this purpose include 5 Whys, chain-of-event, root cause analysis, or similar methods (Leveson, 2011). This means that possible interactions and interdependencies between components are largely disregarded, setting root cause models apart from systemic approaches. Given that these have also been used to examine incidents in cybersecurity, especially concerning technical system aspects (Dolezel and McLeod, 2019, Fovino et al., 2009, Wangen et al., 2017), suggests that they are regarded as valuable for uncovering the underlying causes of cyber incidents.

When a researcher decides to look at an accumulation of multiple incidents in a given population to understand their shared cause(s), safety science typically applies methods of epidemiology (Waller, 1977). These methods can be used to study visible causes of incidents (e.g., lack of safety equipment, click behavior of users) as well as latent causes (e.g., self-reported stress level). Epidemiological methods (Levin, 2005) encompass various methods to study the causes of incidents that are equally relevant for cybersecurity. One common method is the experimental study examining the effect of an intervention (or independent variable) on a given outcome (dependent variable), such as the effect of a safety measures on the prevalence or severity of incidents. The common use of simulated phishing emails to evaluate the effectiveness of security training for employees falls into this category of methods (Lain et al., 2022). Another way to study incidents is through observation, as in cross-sectional, case-control, and cohort studies. A cross-sectional study looks at the outcome (e.g., incident) at any given point in time and is often used to study prevalence (e.g., the share of employees with incidents). For example, the British government conducts yearly surveys related to cybersecurity breaches in organizations using cross-sectional studies (Department for Digital, Culture, Media and Sport 2022) ("pseudo-longitudinal study with varying samples"). In a case-control study, entities with and without incidents but otherwise similar attributes are compared (e.g., employees with and without incidents). In a prospective cohort study, researchers start to monitor a cohort of entities until incidents occur (e.g., studying the influence of work-related stress on individual security incidents). In retrospective cohort studies, historical data containing risk factors or protective factors (e.g., IT investment) is used to find relationships to actual outcomes (IT investments and data breaches (Li et al., 2023)). Although these population-level studies have begun to emerge in cybersecurity, they are still relatively rare and inconsistencies in terminology (e.g., constructs, study designs) makes them difficult to compare (Drogkaris and Bourka, 2019).

3.1.2. Strengths & limitations

The strength of single-factor and linear models lies in their simplicity. They are particularly useful for scrutinizing incidents caused by failures of technical components or human errors in simple socio-technical systems. However, it is important to recognize that they are not infallible and may not always include all contributing factors to an incident. Their capacity to explain incident causation in more complex systems is limited (Qureshi, 2008). For example, APTs have been criticized as a too simple explanation for the success of cyberattacks (Lipner and Pescatore, 2023).

3.2. Cultural & management models

Cultural and management models take into various factors (e.g., culture, power, politics) (Qureshi, 2008) while still assuming a linear chain of events that ultimately lead to an incident. They aim to comprehend not only the immediate causes of individual incidents, but also the broader meaning of safety or unsafety within a socio-technical system (Pidgeon and O'Leary, 2000). Likewise, in cybersecurity, a more holistic view of security is developing, moving beyond the study of individual incidents to organizational aspects and their institutionalization (Renaud, 2011, von Solms, 2000, von Solms, 2006). For example, large organizations have established institutionalized "security bureaucracies" including management structures, governance frameworks and security controls that aim to increase security through security roles and responsibilities (e.g., Chief Security Officer), corporate policies (e.g., information classification), dedicated security budgets, and standards such as ISO 27001 (Mirtsch et al., 2021).

Safety and organizational culture are believed to play a significant role in incidents in cultural models. For instance, following the Columbia space shuttle incident, the Columbia Accident Investigation Board determined that the leading edge of one of the space shuttle's wings was punctured. However, beyond this singular explanation, the board also examined NASA's organizational, historical, and cultural factors that contributed to the incident (NASA | Columbia Accident Investigation Board 2023). The board identified a "broken safety culture" that led to "structural secrecy", causing decision-makers to overlook the threat of previous foam debris strikes. Researchers have studied different cultural aspects, ranging from organizational culture and safety culture/climate to power and conflicts (Cooper, 2000, Guldenmund, 2000). While organizational culture refers to corporate values that affect and influence members' attitudes and behavior, safety culture is a sub-facet related to an organization's safety performance (Cooper, 2000). Nowadays, safety science distinguishes more granularly between the concept of safety culture as a broader concept of shared attitudes, behaviors, and values and the narrower concept of safety climate, which expresses the individual perceptions of safety in the workplace (Petitta et al., 2017, Steven, 2003). Researchers assume that cultural safety aspects and individual safety behavior influence each other (Neal and Griffin, 2006) and that cultural approaches to improve safety could be effective in reducing incidents and improving safety performance indicators (Hlaing and Ochimizu, 2018).

Cybersecurity researchers have also begun to study cultural aspects in socio-technical systems and incidents and their relationship with the individual's security practices. Focusing on constructs such as organizational culture, security culture, security climate and individuals' policy compliance (Da Veiga and Eloff, 2010, Dalal et al., 2022, Goo et al., 2014, Hu et al., 2012, Van Niekerk and Von Solms, 2010), they conclude that factors such as top management support, policy and procedures, and awareness are critical in engendering cybersecurity culture (Uchendu et al., 2021). Companies have also picked up the security culture trend, providing clear-cut definitions and impressive illustrations of how their solution can supposedly eradicate their security culture problem (NCSC 2023). Meanwhile, evidence-based research on this topic is still in its infancy (Uchendu et al., 2021) and has only recently begun to describe approaches to change cultural aspects (Alshaikh, 2020). Cybersecurity research therefore clearly stands to benefit from the approaches that have emerged from the safety sciences.

Unlike cultural models, management models focus on a set of formal roles, responsibilities, structures, and policies. Safety management systems are examples of management models in safety science. They are also referred to as "safety bureaucracy", as they delegate safety efforts to an administrative function in the organization (e.g., a safety officer)

(Dekker, 2019). Besides their purpose of controlling incidents, losses, and defenses, another purpose is to ensure compliance with standards, laws, and regulations (Li and Guldenmund, 2018). Following the logic of safety management systems, an incident is the result of non-compliance of the organization or its members (for example, an organization has not properly implemented the safety management system or an employee has not followed policies).

Cybersecurity has established similar models, such as information security management systems (e.g., ISO 27001 (ISO 2023)). As a consequence, administrative functions such as CISOs, security operation centers (SOCs), or computer emergency response teams have been created to ensure a secure state of the organization (Allen et al., 2015, Da Veiga and Eloff, 2007). These organizational systems and structures are not only responsible for improving security but also for ensuring security compliance of the organization and its individuals (Moody et al., 2018). Similar to safety, compliance is seen as a way to prevent incidents, and non-compliance is seen as a cause of incidents (Cichonski et al., 2012, Moody et al., 2018, Siponen and Vance, 2010).

3.2.1. Methods

Sociological, organizational, and anthropological methods are used in safety science such as cultural immersion, case studies or surveys to study cultural or management aspects. For example, in one five-year longitudinal study, consecutive surveys were used to understand the safety climate, which was identified as a major and lagged factor influencing the individual's safety motivation in an organization (Neal and Griffin, 2006).

Similar methods are used in cybersecurity. For example, in one research project immersive methods were applied to understand incident handling (Sundaramurthy et al., 2014, Sundaramurthy et al., 2016). Computer science students were trained in anthropological methods and embedded as analysts in SOCs over a period of 3.5 years. The embedded students performed the same analyst job and observed the world from the analyst's viewpoint. This revealed a number of things, such as the conflicting goals of the SOC members, as well as potential health issues due to repetitive tasks. In another study using a cross-sectional survey, researchers found that information security climate has a significant positive impact on employees' conformity with the security policy (Goo et al., 2014). Other researchers have developed specific survey instruments such as the Information Security Climate Index (Kessler et al., 2020). Case studies conducted in British companies of various sizes affected by security breaches examined security in the individual companies before and after the breach (Department for Digital, Culture, Media and Sport 2022). In this study the majority of employees interviewed stated their organizations put more of an emphasis on technology than them to stay secure.

3.2.2. Strengths & limitations

The advantage of cultural and management models is their greater appreciation of organizational aspects compared to single-factor and linear models, while they still assume an individual incident has one or multiple causes resulting from a sequential chain of events. They take into account various organizational factors (Qureshi, 2008) and aim to comprehend not only the immediate causes of individual incidents, but also the broader meaning of safety or unsafety within a socio-technical system (Pidgeon and O'Leary, 2000). However, this is associated with greater complexity of investigations compared to the simpler incident models (e.g., interviews with many parties directly and indirectly involved in the incident).

3.3. Systemic models

Systemic models take a different approach to understanding incidents by recognizing that they naturally arise from a working environment, rather than dwelling on speculation about what might have been. Therefore, the emphasis is not on preventing incidents, but on anticipating their predictable occurrence and mitigating their consequences. Systemic models are rooted in systems theory, which posits that incidents are more than the sum of singular root-cause associations, but that they emerge from interactions among a system's components (Qureshi, 2008). Similar to cultural and management models, they aim to comprehend safety and unsafety within a socio-technical system but extend beyond its mechanisms (Leveson, 2011). For example, instead of understanding the chain of events that led to a single traffic incident, a systemic approach would seek to grasp how incidents arise in the socio-technical system from an individual's driving behavior (Ranney, 1994) or a specific traffic system (e.g., road tunnel (Kazaras et al., 2012)). Debates about "cyber resilience" and "cyber capabilities" that have been emerging in recent years reflect a systemic mindset (Christine and Thinyane, 2022).

Control models assume that incidents result from ineffective control and feedback loops (e.g., missing feedback from an operational process to the management). Two examples of models in the tradition of control theory stemming from system engineering are the "Socio-technical Framework" (Rasmussen, 1997) and the "Systems Theoretic Analysis Model and Processes Model" (STAMP) (Leveson, 2016). In both models, an incident is described as a hierarchical control problem involving different organizational levels (e.g., mismanagement on the operational and strategic levels) which can be prevented by enforcing security constraints. While the Socio-technical Framework predefines different hierarchical levels (e.g., government, regulator, company, management), STAMP establishes a flexible hierarchical control structure consisting of multiple (and possibly overlapping) control loops within a socio-technical system (Leveson, 2016).

Models originally developed in safety science, such as STAMP, have made their way into cybersecurity as a tool for identifying vulnerabilities (Khan et al., 2023, Salim, 2014). Initial research has begun to model complex socio-technical systems and reached different conclusions than those reached with traditional models (e.g., linear models). For example, in the case of the 2019 Capital One data breach, academics used publicly available data to find reasons for the incident that go beyond singular explanations (e.g., a misconfigured firewall, a single employee) to managerial and organizational flaws in the whole organization (Khan et al., 2023). Control models therefore allow for more holistic insights into incident occurrences while not singling out the one responsible party (e.g., a specific employee).

Resilience models focus on capabilities of people and organizations to fulfill their goals (e.g., work performance, business goals) despite varying circumstances (Dekker, 2019, p. 391f and p56). Resilience models assume that negative and positive outcomes are caused by the same mechanisms (Hollnagel, 2017). Incidents are only indirectly prevented, to a certain extent accepted as inevitable, and sometimes even result in the adaption of a system (Dekker et al., 2011). For example, in the case of road traffic, the behavior of cyclists and vehicles can be modeled to enhance their interaction capabilities in autonomous driving settings (Parnell et al., 2023). In this example, prevention of incidents is, therefore, the product of an effective communication capability.

Following this adaptive understanding of a system, relying solely on passive methods (e.g., multiple layers of defenses) for safety is inadequate, as these methods can be compromised in situations where productivity is prioritized over safety concerns (Woods, 2003). Resilience, on the other hand, necessitates an ongoing emphasis on preserving a system's abilities to recognize, adapt to and absorb variations, changes, disturbances, disruptions, and surprises (Woods and Hollnagel, 2017). Resilience-based models are helpful as they enable us to examine and affect safety or security, even in the absence of incidents or a minimal number of incidents, since unlike linear models they do not focus solely on incidents (Hollnagel, 2017).

In cybersecurity, resilience approaches are traditionally applied in relation to technical components (e.g., redundancy of infrastructure to improve availability). However, in recent years the notion of cyber resilience has also begun to embrace organizational aspects. For

example, NIST's approach to developing cyber-resilient systems (Ross et al., 2021) focuses on developing primarily technical and organizational capabilities (e.g., architectural diversity, segmentation) of the security organizations. In a wider sense, resilience-based models in cybersecurity also focus on non-technical aspects of the socio-technical system such as general risk management or business continuity (Christine and Thinyane, 2022, Dupont, 2019). Despite initial model proposals on the individual and organizational levels (Kleij and Leukfeldt, 2020, Ross et al., 2021), the discourse on resilient models in cybersecurity is currently still in its infancy and there is little agreement on what resilience means, let alone how it can be assessed.

3.3.1. Methods

A range of methods has been developed to support incident investigation based on systemic models. The advantages of these methods are that they have a good theoretical basis, are highly systematic, and are widely applied (i.e., practical experience is available). In some cases, there is a clear distinction between a model and a corresponding method (e.g., "Socio-technical Framework" and the related "AcciMap" method) (Rasmussen, 1997). However, the model and associated method can also be identical (e.g., "Systems Theoretic Analysis Model and Processes Model" (STAMP) (Leveson, 2016)). Despite being resource-intensive and requiring considerable domain and theoretical knowledge, systemic methods are a dominant concept in safety science (Underwood and Waterson, 2013) and have also found their way into cybersecurity (Khan and Madnick, 2022).

STAMP is an example of a method that understands safety or security as a problem of uncontrolled relationships between components of socio-technical systems ("control methods") (Leveson, 2016). It allows visual modelling of control loops on different levels of an organization (e.g., individual, team, company, government) and identification of problematic areas. An example of an ineffective control loop in security would be the missing feedback from an employee to the IT department about suspicious emails, which would prevent the department from taking further action (e.g., withdraw the email from the mailboxes of all employees). Another example of an ineffective loop would be employees who receive training after clicking on a simulated phishing email but are scared and therefore do not learn, contrary to what the IT department intended. STAMP is related to approaches from systems thinking as well as control engineering, applying the concepts (e.g., actuators, feedback) not only to purely technical, but socio-technical systems. STAMP also provides specific methods for hazard analysis (Systems-Theoretic Process Analysis, STPA) and incident analysis (Causal Analysis based on Systems Theory, CAST) (Leveson, 2016). Based on STAMP, specific cybersecurity-related methods have been developed, such as STPA-Sec (Schmittner et al., 2016, Young and Leveson, 2013) and Cyber Safety (Khan and Madnick, 2022).

An example of a resilience method is the "Functional Resonance Analysis Method" (FRAM), which is used to model the functions that are needed for everyday performance to succeed in a socio-technical system. The model can explain specific events by demonstrating how functions can interconnect and how the fluctuations in regular performance can result in outcomes that exceed expectations, either positively or negatively. Instead of searching for faults and breakdowns, the FRAM clarifies outcomes by examining how functions link together and how everyday performance variability can create a ripple effect (Hollnagel, 2016). For example, FRAM could be used to model the normal incident response process of a computer security incident response team (e.g. weekdays vs. weekends, relaxed working vs. working under pressure) to understand the variability in incident handling (e.g., false negatives vs. false positives). FRAM has been progressively developed scientifically and increasingly adopted by professionals with successful results (Patriarca et al., 2020). In cybersecurity, FRAM has been used in the engineering of information security requirements (Hlaing and Ochimizu, 2018).

The Resilience Analysis Grid (RAG) is another resilience model that looks at a system from a higher level than FRAM to assess the capability of a system for "resilient performance". This includes the assessment of the system's capability to respond to changes, disturbances, or opportunities, to monitor what could affect the system's performance, to learn from experience, and to anticipate future developments (Hollnagel, 2011). These capabilities have also been used in cybersecurity to develop a framework for organizations that promises to better prepare for emerging cyber threats (Kleij and Leukfeldt, 2020). RAG could be also used to model business-critical parts of an organization (e.g., securities settlement of a bank's branches) and assess its resilience in case of cyberattacks.

3.3.2. Strengths & limitations

The advantage of systemic models is the holistic approach to the socio-technical system and its interrelationships. Incidents can arise from the interaction of system components and normal activities. However, the modeling and analysis of incidents with systemic models, as well as with cultural and management models, is very complex compared to simple linear models like 5 Whys.

3.4. Applying models and methods

In the following, we briefly address two important aspects of the application of models and methods, which have been discussed in safety science (and are equally important for cybersecurity): Retrospective vs. prospective analysis and the selection of an appropriate model.

3.4.1. Retrospective vs. prospective analysis?

It is a common assumption that the best way to reduce incidents is study them after they occur (Leveson, 2011). In cases where socio-technical systems are very static (e.g., nuclear power plants), reactive learning from incidents has been very successful (Leveson, 2011, Rasmussen, 1997). The retrospective analysis is commonly applied in organizations in cybersecurity as well (Cichonski et al., 2012).

However, relying solely on retrospective analysis has several limitations (Leveson, 2011, Woods et al., 1994). For example, an investigator can be biased in the ability to assess the likelihood of an outcome (e.g., hindsight bias) (Dekker, 2004). Also, socio-technical systems might not be static but constantly subject to change (Leveson, 2011, Rasmussen, 1997). For example, cybersecurity is a highly dynamic environment in which threats continuously emerge: security vulnerabilities like Log4j (Cyber Safety Review Board 2022), seasonal social engineering attacks like holiday scams (Mitnick Security, 2023), or risks emerging from the latest technologies like ChatGPT (ChatGPT and large language models: what's the risk? 2023). To address these limitations, safety science researchers advocate for combining retrospective and prospective approaches (Leveson, 2011). This means studying socio-technical systems a) in the design (e.g., the implementation of security mechanisms), but also b) before incidents occur (e.g., identifying control loops in an organization) and c) while incidents occur (e.g., in cohort studies). Therefore, we argue for a more comprehensive approach to cybersecurity incident investigation that considers both prospective and retrospective analysis.

3.4.2. Which model to choose?

Unfortunately, safety science does not provide clear guidance on when to choose what model, or on methods to be used to explain incidents. Attempts to provide frameworks for the selection of models based on criteria such as "complexity/tractability of a socio-technical system", "coupling of the components of a socio-technical system", or "impact of an incident" (Hollnagel and Speziali, 2008, Rasmussen, 1997) have been criticized as not objectively measurable and not applicable in a real world situation (Hopkins, 1999). For example, there is little agreement about the complexity of given socio-technical system that would (in the logic of the frameworks) justify or require the use of more complex, systemic incident causations models.

Still, an investigation of a cybersecurity incident can benefit from the awareness of different models and their respective benefits and limitations. As we have already argued, researchers and practitioners have often limited awareness of their implicit incident causation models and methods (Kjellén and Albrechtsen, 2017:25,82). They may even have different accounts of why incidents happen. For example, while some investigators would try to find a root cause, others would conduct interviews to understand the security culture. These implicit models make it difficult to reflect on the benefits and shortcomings of the chosen model. Choosing a model proactively and explicitly would therefore increase external validity of the findings.

Table 3 provides a taxonomy that can help to reflect on the incident causation model in the context of cybersecurity. While, for example, systemic models are currently considered the state of the art in safety science (Grant et al., 2018), this does not necessarily mean that these are useful in every given case. In some cases, a simpler explanation following a linear chain of effects logic might be sufficient and require less effort.

A first criterion for selecting a model is the purpose of an investigation. Depending on the type of the incident and its impact, legal requirements might also dictate a certain type of investigation. If the purpose is to simply fix what comes to first sight, linear models might be well suited. However, if the purpose is to holistically learn about a socio-technical system ahead of an incident or after an incident (why does the same incident keep happening?) and to develop a set of follow-up measures (e.g., new security policies, trainings, or a reporting system), systemic models might be better suited. For example, interviews with multiple stakeholders facilitate not only understanding why things happened but also collecting ideas for improvement in a forward-looking way. However, if the purpose is to find who was responsible, complex models might be less suited than simpler models because they do not point to a single root cause (e.g., mismanagement). However, the history of safety science has also demonstrated that these simple explanations can easily lead to blaming individuals instead of looking at the bigger picture (Catino, 2008).

A second criterion for model selection are the competencies and resources available for the investigation. The application of more complex models might require a large variety of skills and resources. For example, the application of a sociological model may go more smoothly and reveal more insights when supported by a person who is familiar not only with the approach but also a sociological way of thinking. During the investigation, people may present different accounts of what happened (Dekker et al., 2011). Particularly in consideration of the complexity of reality, the goal of the analysis of incidents cannot be to uncover the "true" story of what happened but rather to consider different perspectives on what happened. These perspectives might give overlapping but potentially even contradictory accounts of how outcomes emerged (Dekker et al., 2011). Also, systemic approaches are often more resource-intensive than simple ones, as multiple perspectives from different actors must be considered (e.g., interviews with various stakeholders).

Eventually, the scope influences the model selection. A limited scope on technical or sociological aspects is well aligned with a simple model. The question is then when to stop looking for a root cause ("stopping rule") (Lundberg et al., 2009). If the scope is broader and involves larger parts of the socio-technical system, complex models might be well suited. In this case, the question is what the boundaries of the socio-technical system are (e.g., a team, an organization, society) (Årstad and Aven, 2017).

Fig. 1 illustrates a potential approach for the selection of an adequate model in the retrospective analysis of a cybersecurity incident. Adequate in this context means choosing the simplest and most economical model that is sufficiently explanatory. The simplest possible model will be used first to explain an incident, and only if this does not contribute sufficiently to understanding will more complex models be selected. One driver for selecting a more complex model is the impact of the incident. Incidents with low impact are explained with simple models. For example, an isolated system vulnerability is explained by a missing patch (single-factor model). In the case of a local malware infection on a PC, the linear chain of events leading to the infection is investigated until the root cause is found. (e.g., clicking on an email attachment a few days earlier). When a large-scale security breach occurs, for example, a systemic model is used and various stakeholders are interviewed. Another driver for using more complex models is whether an incident is unique or recurring. A single phishing incident involving one employee may be sufficiently explained by a simple model (e.g., incorrectly set email filter), but if phishing incidents are recurring in the organization, more complex models will be used to explain them (e.g., inadequate security culture).

Table 3
Taxonomy of incident causation models in cybersecurity.

Criterion	Simple and Linear Models	Cultural, Management and Systemic Models
Purpose of the investigation	• Quickly fix what comes to first sight • Finding who is responsible (at the cost of potentially blaming this person)	• Holistically learn about an incident (e.g., why does it occur repeatedly?) • Develop a broad set of follow-up measures
Competencies and resources available	• A disciplinary skillset is available (e.g., primarily technical skills). • Limited resources are available	• An interdisciplinary skillset is available including social science skills • Many resources are available
Scope of the investigation (e.g., boundaries of the socio-technical system).	• Technical or non-technical scope • Limited scope with clearly defined stopping rule in case of a root cause analysis.	• Socio-technical scope • Broader scope with clearly defined boundaries of the socio-technical system (e.g., team, organization, society)

4. Conclusion

Cybersecurity often uses concepts from safety science partially or implicitly without being aware of the associated school of thought. While in safety science there has been an explicit debate about incident causation models, researchers in cybersecurity researchers have (implicitly) used many of these models with far less critical analysis on their suitability. However, researchers in safety science have pointed out that the general assumptions about incidents shape what is later derived as measures to prevent future incidents (Lundberg et al., 2009). We argue that with regard to cybersecurity of socio-technical systems, there is no "true" story of what happened in a successful cyberattack, but that there are different "realities". The purpose of this paper is to demonstrate similarities related to understanding incidents between the established discipline of safety science and the comparably young domain of cybersecurity. We used the method of a narrative review in the field of incident causation to summarize safety science literature. We further collected examples and case studies from cybersecurity and transferred the models and methods to these case studies. By doing so, we seek to provide an initial framework for cybersecurity researchers and practitioners to be more aware of the models (both implicit and explicit) and methods they use to explain the factors and/or environment from which an incident emerged and the implications of their choice.

Successful cyberattacks on organizations can be explained with very simple single-factor models. For example, an APT, a misconfigured firewall or an unreported incident of a junior SOC analyst can be held responsible for an incident. In the slightly more complex logic of a linear model, a ransomware attack can be traced back to a malware infection on an employee's computer and the lack of an endpoint security solution, or latent factors such as low security awareness. Cultural and

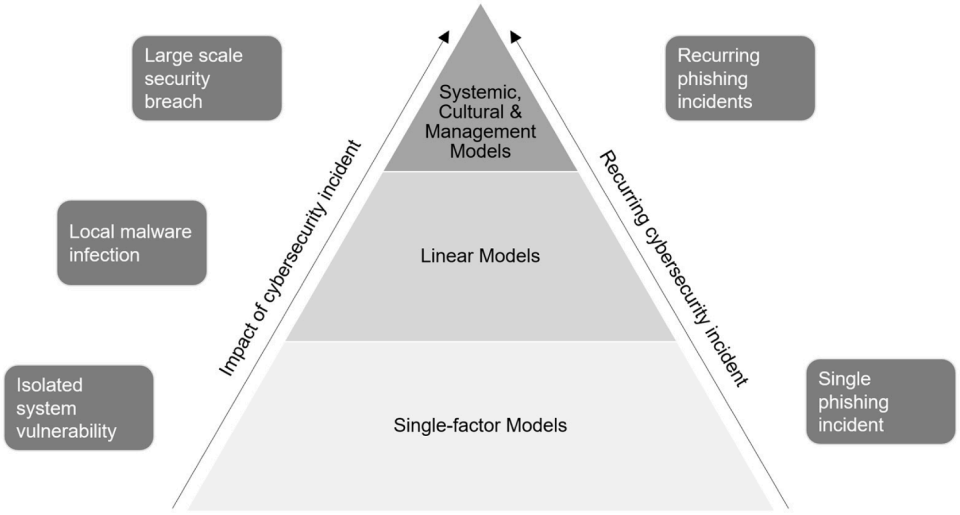


Fig. 1. Incident model continuum.

management models would probably diagnose limited cyber risk awareness among management and end users, or a "blame culture" rather than a "learning culture" when security incidents occur. Eventually, systemic models might try to improve the resilience of an organization and its individuals by providing capabilities for the case that a cyberattack is successful.

Researchers have emphasized that reliable knowledge about actual incidents is scarce (Hove et al., 2014, Maschmeyer et al., 2021) and that organizations need to do more to investigate and learn from cyber incidents (Patterson et al., 2023). While we have provided an initial overview of different models and methods to support learning from incidents, future research should address their applicability and usefulness in practice. For instance, there is not only the question of when to apply which model but also what resources are needed to apply them. This may concern the skills of the people involved or the required knowledge about incidents. Finally, also evidence as to what extent the models and methods contribute to the actual improvement of cybersecurity in organizations is needed.

Author contributions

Nico Ebert had the idea for the manuscript and its preliminary structure (manuscript conceptualization). He also coordinated the work of the author team and, more generally, was responsible for project administration. All authors conducted literature searches, wrote sections of the initial draft of the manuscript, and contributed to the review and editing of the manuscript as well as the responses to reviewers.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nico Ebert reports financial support and article publishing charges were provided by Swiss National Science Foundation. Nico Ebert reports financial support was provided by Digitalization Initiative of the Zurich Higher Education Institutions (DIZH).

Data availability

No data was used for the research described in the article.

References

Årstad, Ingrid, Aven, Terje, 2017. Managing major accident risk: concerns about complacency and complexity in practice. Saf. Sci. 91 (January 2017), 114–121. https://doi.org/10.1016/j.ssci.2016.08.004.

Abdulhafedh, Azad, 2017. Road crash prediction models: different statistical modeling approaches. J. Transport. Technol. 07 (2017), 190. https://doi.org/10.4236/jtts.2017.72014, 02.

Abroshan, Hossein, Devos, Jan, Poels, Geert, Laermans, Eric, 2021. Phishing happens beyond technology: the effects of human behaviors and demographics on each step of a phishing process. IEEE Access 9 (2021), 44928–44949. https://doi.org/10.1109/ACCESS.2021.3066383.

Ahmad, Atif, Hadgkiss, Justin, Ruighaver, A.B., 2012. Incident response teams – Challenges in supporting the organisational security function. Comput. Security 31 (2012), 643–652. https://doi.org/10.1016/j.cose.2012.04.001, 5.

Al-Shaer, Ehab S., Hamed, Hazem H., 2004. Modeling and management of firewall policies. IEEE Trans. Netw. Serv. Manage. 1 (2004), 2–10, 1.

Allen Julia, Crabb Gregory, Curtis Pamela, Fitzpatrick Brendan, Mehravari Nader, Tobar David. 2015. *Structuring the chief information security officer organization*. doi:10.13140/RG.2.1.1242.6967.

Alshaikh, Moneer, 2020. Developing cybersecurity culture to influence employee behavior: a practice perspective. Comput. Security 98 (November 2020), 102003. https://doi.org/10.1016/j.cose.2020.102003.

Amundrud, Øystein, Aven, Terje, Flage, Roger, 2017. How the definition of security risk can be made compatible with safety definitions. Proc. Inst. Mech. Eng. O J. Risk Reliab. 231 (June 2017), 286–294. https://doi.org/10.1177/1748006x17699145, 3.

Andress, Amanda, 2003. Surviving Security: How to Integrate People, Process, and Technology. CRC press.

Angela Sasse, M., Flechais, Ivan, 2005. Usable Security: Why Do We Need It? How Do We Get It? O'Reilly.

Aoyama, Tomomi, Naruoka, Hidemasa, Koshijima, Ichiro, Machii, Wataru, Seki, Kohei, 2015. Studying resilient cyber incident management from large-scale cyber security training. In: 2015 10th Asian Control Conference (ASCC), pp. 1–4. https://doi.org/10.1109/ASCC.2015.7244713.

Arce, I., 2003. The weakest link revisited. IEEE Security Privacy 1 (March 2003), 72–76. https://doi.org/10.1109/MSECP.2003.1193216, 2.

Aven, Terje, 2014. What is safety science? Saf. Sci. 67 (August 2014), 15–20. https://doi.org/10.1016/j.ssci.2013.07.026.

Bair, Jonathan, Bellovin, Steven M., Manley, Andrew, Reid, Blake, Shostack, Adam, 2017. That was close: reward reporting of cybersecurity near misses. Colo. Tech. LJ 16 (2017), 327.

Baker, Wade H., Wallace, Linda, 2007. Is information security under control?: investigating quality in information security management. IEEE Security Privacy 5 (January 2007), 36–44. https://doi.org/10.1109/MSP.2007.11, 1.

Banga, Gaurav, 2020. Why is cybersecurity not a human-scale problem anymore? Commun. ACM 63 (March 2020), 30–34. https://doi.org/10.1145/3347144, 4.

Björck, Fredrik, Henkel, Martin, Stirna, Janis, Zdravkovic, Jelena, 2015. Cyber resilience – fundamentals for a definition. *New Contributions in Information Systems and Technologies* (Advances in Intelligent Systems and Computing). Springer International Publishing, Cham, pp. 311–316. https://doi.org/10.1007/978-3-319-16486-1_31.

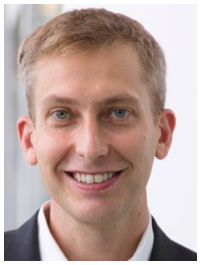
Brostoff, Sacha, Sasse, M.Angela, 2001. Safe and sound: a safety-critical approach to security. In: Proceedings of the 2001 workshop on New security paradigms, pp. 41–50.

Bush, Matthew, Mashatan, Atefeh, 2023. From zero to 100. Commun. ACM 66 (January 2023), 48–55. https://doi.org/10.1145/3573127, 2.

Canfield, Casey Inez, Fischhoff, Baruch, 2018. Setting priorities in behavioral interventions: an application to reducing phishing risk. Risk Anal. 38 (2018), 826–838. https://doi.org/10.1111/risa.12917, 4.
Catino, Maurizio, 2008. A review of literature: individual blame vs. organizational function logics in accident analysis. J. Conting. Crisis Manage. 16 (2008), 53–62. https://doi.org/10.1111/j.1468-5973.2008.00533.x, 1.
ChatGPT and large language models: what's the risk? Retrieved May 11, 2023 from https://www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk.
Christine, Debora Irene, Thinyane, Mamello, 2022. Socio-technical cyber resilience: a systematic review of cyber resilience management frameworks. In: Marx Gómez, Jorge, Lorini, Maria Rosa (Eds.), Digital Transformation for Sustainability: ICT-supported Environmental Socio-economic Development. Springer International Publishing, Cham, pp. 573–597. https://doi.org/10.1007/978-3-031-15420-1_28.
Cichonski, Paul, Millar, Thomas, Grance, Tim, Scarfone, Karen, 2012. Computer Security Incident Handling Guide. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2.
Cooper, M.D., 2000. Towards a model of safety culture. Saf. Sci. 36 (2000), 111–136. https://doi.org/10.1016/S0925-7535(00)00035-7, 2.
Cranor, Lorrie Faith, Garfinkel, Simson, 2005. Security and Usability: Designing Secure Systems That People Can Use. O'Reilly Media, Inc.
Cyber Safety Review Board, 2022. Review of the December 2021Log4j Event. Cyber Safety Review Board. Retrieved from. https://www.cisa.gov/sites/default/files/publications/CSRB-Report-on-Log4-July-11-2022_508.pdf.
Da Veiga, A., Eloff, J.H.P., 2007. An information security governance framework. Infor. Syst. Manage. 24 (October 2007), 361–372. https://doi.org/10.1080/10580530701586136, 4.
Da Veiga, A., Eloff, J.H.P., 2010. A framework and assessment instrument for information security culture. Comput. Security 29 (March 2010), 196–207. https://doi.org/10.1016/j.cose.2009.09.002, 2.
Dalal, Reeshad S., Howard, David J., Bennett, Rebecca J., Posey, Clay, Zaccaro, Stephen J., Brummel, Bradley J., 2022. Organizational science and cybersecurity: abundant opportunities for research at the interface. J. Bus. Psychol. 37 (February 2022), 1–29. https://doi.org/10.1007/s10869-021-09732-9, 1.
Dekker, Sidney, Cilliers, Paul, Hofmeyr, Jan-Hendrik, 2011. The complexity of failure: implications of complexity theory for safety investigations. Saf. Sci. 49 (July 2011), 939–945. https://doi.org/10.1016/j.ssci.2011.01.008, 6.
Dekker, Sidney W.A., 2004. The hindsight bias is not a bias and not about history. Hum. factors aerosp. safety 4 (2004), 87–99, 2.
Dekker, Sidney, 2019. Foundations of safety science: A century of understanding accidents and disasters. Routledge.
Department for Digital, Culture, Media & Sport, 2022. Exploring Organisational Experiences of Cyber Security Breaches. Retrieved April 1, 2023 from. https://www.gov.uk/government/publications/exploring-organisational-experiences-of-cyber-security-breaches.
Department for Digital, Culture, Media & Sport, 2022. Cyber Security Breaches Survey 2022. Retrieved April 1, 2023 from. https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2022/cyber-security-breaches-survey-2022.
Dhillon, Gurpreet, Backhouse, James, 2000. Technical opinion: information system security management in the new millennium. Commun. ACM 43 (July 2000), 125–128. https://doi.org/10.1145/341852.341877, 7.
Dolezel, Diane, McLeod, Alexander, 2019. Managing security risk: modeling the root causes of data breaches. Health Care Manag. (Frederick) 38 (2019), 322–330, 4.
Drogkaris, P., Bourka, A., 2019. Cybersecurity culture guidelines: Behavioural aspects of cybersecurity. Eur. Union Agency for Netw. Inform. Security (ENISA) (2019).
Duncan, Rory, 2020. What does 'secure by design' actually mean? Netw. Security 2020 (August 2020), 18–19. https://doi.org/10.1016/S1353-4858(20)30095-7, 8.
Dupont, Benoît, 2019. The cyber-resilience of financial institutions: significance and applicability. J. Cybersecur. 5 (January 2019), tyz013. https://doi.org/10.1093/cybsec/tyz013, 1.
Fovino, Igor Nai, Masera, Marcelo, De Cian, Alessio, 2009. Integrating cyber attacks within fault trees. Reliab. Eng. Syst. Saf. 94 (September 2009), 1394–1402. https://doi.org/10.1016/j.ress.2009.02.020, 9.
Fu, Gui, Xie, Xuecai, Jia, Qingsong, Li, Zonghan, Chen, Ping, Ge, Ying, 2020. The development history of accident causation models in the past 100 years: 24Model, a more modern accident causation model. Process Saf. Environ. Prot. 134 (February 2020), 47–82. https://doi.org/10.1016/j.psep.2019.11.027.
Goo, Jahyun, Yim, Myung-Seong, Kim, Dan J., 2014. A path to successful management of employee security compliance: an empirical study of information security climate. IEEE Trans. Prof. Commun. 57 (December 2014), 286–308. https://doi.org/10.1109/TPC.2014.2374011, 4.
Grant, Eryn, Salmon, Paul M., Stevens, Nicholas J., Goode, Natassia, Read, Gemma J., 2018. Back to the future: what do accident causation models tell us about accident prediction? Saf. Sci. 104 (2018), 99–109. https://doi.org/10.1016/j.ssci.2017.12.018.
Greitzer, Frank L., Strozer, Jeremy R., Cohen, Sholom, Moore, Andrew P., Mundie, David, Cowley, Jennifer, 2014. Analysis of unintentional insider threats deriving from social engineering exploits. In: 2014 IEEE Security and Privacy Workshops, pp. 236–250. https://doi.org/10.1109/SPW.2014.39.
Guldenmund, F.W., 2000. The nature of safety culture: a review of theory and research. Saf. Sci. 34 (2000), 215–257. https://doi.org/10.1016/S0925-7535(00)00014-X, 1.
He, Ying, Johnson, Chris, Renaud, Karen, Lu, Yu, Jebriel, Salem, 2014. An empirical study on the use of the generic security template for structuring the lessons from information security incidents. In: 2014 6th International Conference on Computer Science and Information Technology (CSIT), pp. 178–188. https://doi.org/10.1109/CSIT.2014.6805998.
Heartfield, Ryan, Loukas, George, 2018. Protection against semantic social engineering attacks. In: Conti, Mauro, Somani, Gaurav, Poovendran, Radha (Eds.), Versatile Cybersecurity. Springer International Publishing, Cham, pp. 99–140. https://doi.org/10.1007/978-3-319-97643-3_4.
Heartfield, Ryan, Loukas, George, Gan, Diane, 2016. You are probably not the weakest link: towards practical prediction of susceptibility to semantic social engineering attacks. IEEE Access 4 (2016), 6910–6928. https://doi.org/10.1109/ACCESS.2016.2616285.
HW Heinrich. 1931. Industrial accident Prevention.(with D. Petersen & N. Roos, 1980).
Heraghty, Derek, Dekker, Sidney, Rae, Andrew, 2018. Accident report interpretation. safety 4 (December 2018), 46. https://doi.org/10.3390/safety4040046, 4.
Herrmann, Dominik, Pridöhl, Henning, 2020. Basic concepts and models of cybersecurity. In: Christen, Markus, Gordijn, Bert, Loi, Michele (Eds.), The ethics of cybersecurity. Springer International Publishing, Cham, pp. 11–44. https://doi.org/10.1007/978-3-030-29053-5_2.
Hielscher, Jonas, Menges, Uta, Parkin, Simon, Kluge, Annette, Sasse, M Angela, 2023. Employees who don't accept the time security takes are not aware enough": the CISO view of human-centred security. In: 32st USENIX security symposium (USENIX security 23), boston, MA.
Hlaing, Swe Zin, Ochimizu, Koichiro, 2018. An integrated cost-effective security requirement engineering process in SDLC using FRAM. In: 2018 International Conference on Computational Science and Computational Intelligence (CSCI), pp. 852–857. https://doi.org/10.1109/CSCI46756.2018.00170.
Hollnagel E., Speziali J. Study on Developments in Accident Investigation Methods: A Survey of the "State-of-the-Art. 2008, 45 p. ffhal-00569424f https://minesparis-psl.hal.science/hal-00569424/file/SKI-Report2008_50.pdf.
Hollnagel, Erik, 2001. Anticipating Failures: What Should Predictions be About? Linkoeping Univ (Sweden) Graduate School for Human-Machine Interaction.
Hollnagel, Erik, 2009. The ETTO Principle: Efficiency-Thoroughness Trade-Off: Efficiency-Thoroughness Trade-Off: Why Things That Go Right Sometimes Go Wrong, New ed Edition ed. Routledge, Burlington, VT.
Hollnagel, Erik, 2011. Epilogue: RAG – the resilience analysis grid. Resilience Engineering in Practice. CRC Press.
Hollnagel, Erik, 2014. Is safety a subject for science? Saf. Sci. 67 (August 2014), 21–24. https://doi.org/10.1016/j.ssci.2013.07.025.
Hollnagel, Erik, 2016. FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems. CRC Press, London. https://doi.org/10.1201/9781315255071.
Hollnagel, Erik, 2017. Safety-I and Safety-II: The Past and Future of Safety Management. CRC Press, London. https://doi.org/10.1201/9781315607511.
Hopkins, A., 1999. The limits of normal accident theory. Saf. Sci. 32 (August 1999), 93–102, 2–3.
Hove, Cathrine, Tårnes, Marte, Line, Maria B., Bernsmed, Karin, 2014. Information security incident management: identified practice in large organizations. In: 2014 Eighth International Conference on IT Security Incident Management & IT Forensics, pp. 27–46. https://doi.org/10.1109/IMF.2014.9.
Hu, Qing, Dinev, Tamara, Hart, Paul, Cooke, Donna, 2012. Managing employee compliance with information security policies: the critical role of top management and organizational culture*. Decision Sci. 43 (2012), 615–660. https://doi.org/10.1111/j.1540-5915.2012.00361.x, 4.
ISO 2023. ISO - ISO/IEC 27001 and related standards — Information security management. Retrieved April 1, 2023 from https://www.iso.org/isoiec-27001-information-security.html.
James Reason, 1990. Human error. Cambridge university press.
Jensen, Matthew L., Wright, Ryan T., Durcikova, Alexandra, Karumbaiah, Shamya, 2022. Improving phishing reporting using security gamification. J. Manage. Inform. Syst. 39 (July 2022), 793–823. https://doi.org/10.1080/07421222.2022.2096551, 3.
Johnston, Allen C., Warkentin, Merrill, McBride, Maranda, Carter, Lemuria, 2016. Dispositional and situational factors: influences on information security policy violations. Eur. J. Inf. Syst. 25 (May 2016), 231–251. https://doi.org/10.1057/ejis.2015.15, 3.
Katsakiori, Panagiota, Sakellaropoulos, George, Manatakis, Emmanuel, 2009. Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. Saf. Sci. 47 (2009), 1007–1015. https://doi.org/10.1016/j.ssci.2008.11.002, 7.
Kazaras, Konstantinos, Kirytopoulos, Konstantinos, Rentizelas, Athanasios, 2012. Introducing the STAMP method in road tunnel safety assessment. Saf. Sci. 50 (November 2012), 1806–1817. https://doi.org/10.1016/j.ssci.2012.04.013, 9.
Kessler, Stacey R., Pindek, Shani, Kleinman, Gary, Andel, Stephanie A., Spector, Paul E., 2020. Information security climate and the assessment of information security risk among healthcare employees. Health Informatics J. 26 (2020), 461–473, 1.
Khan, Shaharyar, Madnick, Stuart, 2022. Cybersafety: a system-theoretic approach to identify cyber-vulnerabilities & mitigation requirements in industrial control systems. IEEE Trans. Dependable Secure Comput. 19 (September 2022), 3312–3328. https://doi.org/10.1109/TDSC.2021.3093214, 5.
Khan, Shaharyar, Kabanov, Ilya, Hua, Yunke, Madnick, Stuart, 2023. A systematic analysis of the capital one data breach: critical lessons learned. ACM Trans. Priv. Secur. 26 (February 2023), 1–29. https://doi.org/10.1145/3546068, 1.
Khanzode, Vivek V., Maiti, J., Ray, P.K., 2012. Occupational injury and accident research: a comprehensive review. Saf. Sci. 50 (2012), 1355–1367. https://doi.org/10.1016/j.ssci.2011.12.015, 5.
Kjellén, Urban, Albrechtsen, Eirik, 2017. Prevention of Accidents and Unwanted Occurrences: Theory, Methods, and Tools in Safety Management. CRC Press.
Kleij, Rick van der, Leukfeldt, Rutger, 2020. Cyber resilient behavior: integrating human behavioral models and resilience engineering capabilities into cyber security. *Advances in Human Factors in Cybersecurity* (Advances in Intelligent Systems and

Computing). Springer International Publishing, Cham, pp. 16–27. https://doi.org/10.1007/978-3-030-20488-4_2.
Kraemer, Sara, Carayon, Pascale, Clem, John, 2009. Human and organizational factors in computer and information security: Pathways to vulnerabilities. Comput. Security 28 (October 2009), 509–520. https://doi.org/10.1016/j.cose.2009.04.006, 7.
Lain, Daniele, Kostiainen, Kari, Capkun, Srdjan, 2022. Phishing in organizations: findings from a large-scale and long-term study. In: 2022 IEEE Symposium on Security and Privacy (SP), pp. 842–859. https://doi.org/10.1109/SP46214.2022.9833766.
Landwehr, Carl E., Bull, Alan R., McDermott, John P., Choi, William S., 1994. A taxonomy of computer program security flaws. ACM Comput Surv 26 (September 1994), 211–254. https://doi.org/10.1145/185403.185412, 3.
Lawrence Pfleeger, Shari, Sasse, M.Angela, Furnham, Adrian, 2014. From weakest link to security hero: transforming staff security behavior. J. Homel. Secur. Emerg. Manag. 11 (December 2014), 489–510. https://doi.org/10.1515/jhsem-2014-0035, 4.
Lehto, Mark, Salvendy, Gavriel, 1991. Models of accident causation and their application: review and reappraisal. J. Eng. Tech. Manage. 8 (1991), 173–205. https://doi.org/10.1016/0923-4748(91)90028-P, 2.
Leveson, Nancy G., 2011. Applying systems thinking to analyze and learn from events. Saf. Sci. 49 (January 2011), 55–64. https://doi.org/10.1016/j.ssci.2009.12.021, 1.
Leveson, Nancy G., 2016. Engineering a Safer World: Systems Thinking Applied to Safety, Reprint Edition ed. The MIT Press, Cambridge, Massachusetts London, Englang.
Levin, Kate Ann, 2005. Study design I. Evid Based Dent 6 (September 2005), 78–79. https://doi.org/10.1038/sj.ebd.6400355, 3.
Li, Yuling, Guldenmund, Frank W., 2018. Safety management systems: a broad overview of the literature. Saf. Sci. 103 (March 2018), 94–123. https://doi.org/10.1016/j.ssci.2017.11.016.
Li, Wilson Weixun, Man Leung, Alvin Chung, Yue, Wei Thoo, 2023. Where is it in information security? The interrelationship among it investment, security awareness, and data breaches. MIS Q. 47 (March 2023), 317–342. https://doi.org/10.25300/MISQ/2022/15713, 1.
Liginlal, Divakaran, Sim, Inkook, Khansa, Lara, 2009. How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. Comput. Security 28 (May 2009), 215–228. https://doi.org/10.1016/j.cose.2008.11.003, 3.
Lipner, Steve, Pescatore, John, 2023. Updates, threats, and risk management. Commun. ACM 66 (April 2023), 21–23. https://doi.org/10.1145/3587826, 5.
Littlewood, Bev, Strigini, Lorenzo, 2004. Redundancy and diversity in security. *Computer Security – ESORICS 2004* (Lecture Notes in Computer Science). Springer, Berlin, Heidelberg, pp. 423–438. https://doi.org/10.1007/978-3-540-30108-0_26.
Lundberg, Jonas, Rollenhagen, Carl, Hollnagel, Erik, 2009. What-you-look-for-is-what-you-find – the consequences of underlying accident models in eight accident investigation manuals. Saf. Sci. 47 (December 2009), 1297–1311. https://doi.org/10.1016/j.ssci.2009.01.004, 10.
Maschmeyer, Lennart, Deibert, Ronald J., Lindsay, Jon R., 2021. A tale of two cybers - how threat reporting by cybersecurity firms systematically underrepresents threats to civil society. J. Inform. Technol. Politics 18 (January 2021), 1–20. https://doi.org/10.1080/19331681.2020.1776658, 1.
Mirtsch, Mona, Kinne, Jan, Blind, Knut, 2021. Exploring the adoption of the international information security management system standard ISO/IEC 27001: a web mining-based analysis. IEEE Trans. Eng. Manage. 68 (February 2021), 87–100. https://doi.org/10.1109/TEM.2020.2977815, 1.
Mitnick Security. How to avoid holiday scams & social engineering tricks at work. Retrieved May 11, 2023 from https://www.mitnicksecurity.com/blog/how-to-avoid-holiday-scams-social-engineering-tricks-at-work.
Moody, Gregory D., Siponen, Mikko, Pahnila, Seppo, 2018. Toward a unified model of information security policy compliance. MIS q. 42 (2018), 1.
Mosteiro-Sanchez, Aintzane, Barcelo, Marc, Astorga, Jasone, Urbieta, Aitor, 2020. Securing IIoT using defence-in-depth: towards an end-to-end secure industry 4.0. J. Manuf. Syst. 57 (October 2020), 367–378. https://doi.org/10.1016/j.jmsy.2020.10.011.
NASA | Columbia Accident Investigation Board. Retrieved April 1, 2023 from https://history.nasa.gov/columbia/CAIB.html.
NCSC, 2023. Developing a Positive Cyber Security Culture. Retrieved May 11,from. https://www.ncsc.gov.uk/collection/board-toolkit/developing-a-positive-cyber-security-culture.
Neal, Andrew, Griffin, Mark A., 2006. A study of the lagged relationships among safety climate, safety motivation, safety behavior, and accidents at the individual and group levels. J. Appl. Psychol. 91 (2006), 946–953. https://doi.org/10.1037/0021-9010.91.4.946.
Ofte, Håvard Jakobsen, Katsikas, Sokratis, 2023. Understanding situation awareness in SOCs, a systematic literature review. Comput. Security 126 (March 2023), 103069. https://doi.org/10.1016/j.cose.2022.103069.
Parnell, Katie J., Stanton, Neville A., Banks, Victoria A., Plant, Katherine L., 2023. Resilience engineering on the road: using operator event sequence diagrams and system failure analysis to enhance cyclist and vehicle interactions. Appl. Ergon. 106 (January 2023), 103870 https://doi.org/10.1016/j.apergo.2022.103870.
Parrish Jr, James L., Bailey, Janet L., Courtney, James F., 2009. A Personality Based Model for Determining Susceptibility to Phishing Attacks. University of Arkansas, Little Rock, pp. 285–296 (2009).
Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P., Hollnagel, E., 2020. Framing the FRAM: a literature review on the functional resonance analysis method. Saf. Sci. 129 (September 2020), 104827 https://doi.org/10.1016/j.ssci.2020.104827.
Patterson, Clare M., Nurse, Jason R.C., Franqueira, Virginia N.L., 2023. Learning from cyber security incidents: a systematic review and future research agenda. Comput. Security 132 (September 2023), 103309. https://doi.org/10.1016/j.cose.2023.103309.
Petitta, Laura, Probst, Tahira M., Barbaranelli, Claudio, Ghezzi, Valerio, 2017. Disentangling the roles of safety climate and safety culture: multi-level effects on the relationship between supervisor enforcement and safety compliance. Accident Anal. Prevent. 99 (February 2017), 77–89. https://doi.org/10.1016/j.aap.2016.11.012.
Petrič, Gregor, Roer, Kai, 2022. The impact of formal and informal organizational norms on susceptibility to phishing: combining survey and field experiment data. Telemat. Inform. 67 (February 2022), 101766 https://doi.org/10.1016/j.tele.2021.101766.
Pidgeon, N., O'Leary, M., 2000. Man-made disasters: why technology and organizations (sometimes) fail. Saf. Sci. 34 (February 2000), 15–30. https://doi.org/10.1016/S0925-7535(00)00004-7, 1.
Pollini, Alessandro, Callari, Tiziana C., Tedeschi, Alessandra, Ruscio, Daniele, Save, Luca, Chiarugi, Franco, Guerri, Davide, 2022. Leveraging human factors in cybersecurity: an integrated methodological approach. Cogn. Tech. Work 24 (May 2022), 371–390. https://doi.org/10.1007/s10111-021-00683-y, 2.
Qureshi, Z.H., 2008. A review of accident modelling approaches for complex critical sociotechnical systems. (2008).
Ranney, T.A., 1994. Models of driving behavior: a review of their evolution. Accid. Anal. Prev. 26 (December 1994), 733–750. https://doi.org/10.1016/0001-4575(94)90051-5, 6.
Rasmussen, Jens, 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27 (November 1997), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0, 2.
Renaud, Karen, Searle, Rosalind, Dupuis, Marc, 2021. Shame in cyber security: effective behavior modification tool or counterproductive foil? New Security Paradigms Workshop. ACM, Virtual Event USA, pp. 70–87. https://doi.org/10.1145/3498891.3498896.
Renaud, Karen, 2011. Blaming noncompliance is too convenient: what really causes information breaches? IEEE Secur. Privacy 10 (2011), 57–63. https://doi.org/10.1109/MSP.2011.157, 3.
Ross, Ron, Pillitteri, Victoria, Graubart, Richard, Bodeau, Deborah, McQuaid, Rosalie, 2021. Developing Cyber-Resilient Systems: A Systems Security Engineering Approach. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-160v2r1.
Safety thinking and safety methods. *safetysynthesis.com*. Retrieved April 1, 2023 from https://safetysynthesis.com/safetysynthesis-facets/safety-thinking-and-safety-methods.
Salim, Hamid M., 2014. Cyber safety: a systems thinking and systems theory approach to managing cyber security risks. Thesis.
Sasse, M.A., Brostoff, S., Weirich, D., 2002. Transforming the "Weakest Link" - a human-computer interaction approach to usable and effective security. In: Temple, Robert, Regnault, John (Eds.), Internet and Wireless Security. Institution of Engineering and Technology, pp. 243–262. https://doi.org/10.1049/PBBT004E_ch15.
Schmittner, Christoph, Ma, Zhendong, Puschner, Peter, 2016. Limitation and improvement of STPA-Sec for safety and security co-analysis. In: Computer safety, reliability, and security: SAFECOMP 2016 workshops, ASSURE, DECSoS, SASSUR, and TIPS, trondheim, norway, september 20, 2016, proceedings 35, pp. 195–209.
Schneier, Bruce, 2015. Secrets and Lies: Digital Security in a Networked World. John Wiley & Sons.
Siponen, Mikko, Vance, Anthony, 2010. Neutralization: New insights into the problem of employee information systems security policy violations. MIS q. (2010), 487–502.
Sklet, Snorre, 2004. Comparison of some selected methods for accident investigation. J. Hazard. Mater. 111 (2004), 29–37, 1–3.
Stembert, Nathalie, Padmos, Arne, Bargh, Mortaza S., Choenni, Sunil, Jansen, Frans, 2015. A study of preventing email (spear) phishing by enabling human intelligence. In: 2015 European Intelligence and Security Informatics Conference, pp. 113–120. https://doi.org/10.1109/EISIC.2015.38.
Steven, Yule, 2003. Safety culture and safety climate:a review of the literature. Industr. Psychol. Res. Centre (January 2003), 1–26.
Sundaramurthy, Sathya Chandran, Case, Jacob, Truong, Tony, Zomlot, Loai, Hoffmann, Marcel, 2014. A tale of three security operation centers. In: *Proceedings of the 2014 ACM Workshop on Security Information Workers* (SIW '14). Association for Computing Machinery, New York, NY, USA, pp. 43–50. https://doi.org/10.1145/2663887.2663904.
Sathya Chandran Sundaramurthy, John McHugh, Xinming Ou, Michael Wesch, Alexandru G. Bardas, and S. Raj Rajagopalan. 2016. Turning Contradictions into Innovations or: How We Learned to Stop Whining and Improve Security Operations. 237–251. Retrieved April 2, 2023 from https://www.usenix.org/conference/soups2016/technical-sessions/presentation/sundaramurthy.
Tøndel, Inger Anne, Line, Maria B., Jaatun, Martin Gilje, 2014. Information security incident management: current practice as reported in the literature. Comput. Security 45 (2014), 42–57. https://doi.org/10.1016/j.cose.2014.05.003.
Uchendu, Betsy, Nurse, Jason R.C., Bada, Maria, Furnell, Steven, 2021. Developing a cyber security culture: current practices and future needs. Comput. Security 109 (October 2021), 102387. https://doi.org/10.1016/j.cose.2021.102387.
Uebelacker, Sven, Quiel, Susanne, 2014. The social engineering personality framework. In: 2014 Workshop on Socio-Technical Aspects in Security and Trust, pp. 24–30. https://doi.org/10.1109/STAST.2014.12.
Underwood, Peter, Waterson, Patrick, 2013. Accident analysis models and methods: guidance for safety professionals. Loughborough University (2013).
Van Niekerk, J.F., Von Solms, R., 2010. Information security culture: a management perspective. Comput. Security 29 (June 2010), 476–486. https://doi.org/10.1016/j.cose.2009.10.005, 4.
von Solms, Basie, 2000. Information security — the third wave? Comput. Security 19 (November 2000), 615–620. https://doi.org/10.1016/S0167-4048(00)07021-8, 7.

von Solms, Basie, 2006. Information security – the fourth wave. Comput. Security 25 (May 2006), 165–168. https://doi.org/10.1016/j.cose.2006.03.004, 3.
Vroom, Cheryl, von Solms, Rossouw, 2004. Towards information security behavioural compliance. Comput. Security 23 (May 2004), 191–198. https://doi.org/10.1016/j.cose.2004.01.012, 3.
Waller, Julian A., 1977. Epidemiologic approaches to injury research. Rare Event/Accident Res. Methodol. (1977), 29.
Wangen, Gaute, Hellesen, Niclas, Torres, Henrik, Brækken, Erlend, 2017. An empirical study of root-cause analysis in information security management. In: Proceedings of the SECURWARE (2017).
WEF, 2016. The fourth industrial revolution: what it means and how to respond. The Fourth Industr. Revol. Retrieved May 11, 2023 from https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/.
Wood, Charles Cresson, Banks, William W., 1993. Human error: an overlooked but significant information security problem. Comput. Security 12 (February 1993), 51–60. https://doi.org/10.1016/0167-4048(93)90012-T, 1.
Woods, David D., Hollnagel, Erik, 2017. Prologue: resilience engineering concepts. Resilience Engineering. CRC Press, pp. 1–6.
Woods, David D., Johannesen, Leila J., Cook, Richard I., Sarter, Nadine B., 1994. Behind Human Error: Cognitive Systems, Computers and Hindsight. Dayton Univ Research Inst.
Woods, David D., 2003. Creating foresight: how resilience engineering can transform NASA's approach to risky decision making. Work 4 (2003), 137–144, 2.
Wright, Ryan T., Marett, Kent, 2010. The influence of experiential and dispositional factors in phishing: an empirical investigation of the deceived. J. Manage. Infor. Syst. 27 (2010), 273–303, 1.
Yadav, Tarun, Rao, Arvind Mallari, 2015. Technical aspects of cyber kill chain. *Security in Computing and Communications* (Communications in Computer and Information Science). Springer International Publishing, Cham, pp. 438–452. https://doi.org/10.1007/978-3-319-22915-7_40.
Yan, Zheng, Robertson, Thomas, Yan, River, Park, Sung Yong, Bordoff, Samantha, Chen, Quan, Sprissler, Ethan, 2018. Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment? Comput. Hum. Behav. 84 (July 2018), 375–382. https://doi.org/10.1016/j.chb.2018.02.019.
Young, William, Leveson, Nancy, 2013. Systems thinking for safety and security. In: Proceedings of the 29th annual computer security applications conference, pp. 1–8.
Zimmermann Verena. 2023. Moving from a "human-as-problem" to a "human-as-solution" cybersecurity mindset. doi:10.1016/j.ijhcs.2019.05.005.



Nico Ebert is a Senior Researcher and Lecturer at the Zurich University of Applied Sciences in Human Factors in Security and Privacy. He has a Ph.D. in Information Systems from the University of St. Gallen and several years of practical experience in the IT industry in consulting companies and large corporations. His research interests include organizational and behavioral aspects of security and privacy, security and privacy by design, and human-computer interaction.

Thierry Schaltegger is a doctoral researcher at the Zurich University of Applied Sciences and at the University of Zurich. He received his MSc in Psychology from the University of Zurich in 2022. His research interests include risk and decision-making in the context of cybersecurity.

Benjamin Ambuehl is a Posdoctoral researcher at the Zurich University of Applied Sciences. He received the Ph.D. degree in Health & Environmental Psychology from the University of Bern (Switzerland) in 2022. His research interest includes behavior change of individuals and groups towards sustainable adoption, use and maintenance of technology.

Lorin Schöni received a B.Sc. and M.Sc. in Psychology from the University of Zurich. He is currently a doctoral student in the Security, Privacy and Society group at ETH Zürich. Coming from a background in cognitive neuroscience and human-computer interactions, he now focusses on combining psychological concepts and technology to address challenges arising out of digitalization. To that end, he investigates human-centered interventions in a cybersecurity context.

Verena Zimmermann is an Assistant Professor for Security, Privacy, and Society at ETH Zürich. Before that, she studied psychology, focusing on human-technology interaction, and completed her Ph.D. at the intersection of psychology and computer science at the Technical University Darmstadt, Germany. She was also part of the German National Research Center for Applied Cybersecurity Research and contributed to several interdisciplinary projects there.

Melanie Knieps is a researcher at the Digital Society Initiative at the University of Zurich. She received her PhD in legal psychology from the University of Gothenburg (Sweden) where she studied how people plan and lie about their intentions. Today, she focuses on the impact of motivation and trust on knowledge exchange and collaboration in cybersecurity.