What works well? A safety-iI approach to cybersecurity

Steen, T. van; Real, C. del; Berg, B. van den; Schmorrow, D.D.; Fidopiastis, C.M.

What Works Well? A Safety-II Approach to Cybersecurity

Tommy van Steen(✉), Cristina Del-Real, and Bibi van den Berg

Leiden University, Turfmarkt 99, 2511 DP The Hague, The Netherlands
t.van.steen@fgga.leidenuniv.nl

Abstract. The field of cybersecurity is used to focusing on what goes wrong. Threats, incidents, and impact are factors that are widely investigated, and the solutions presented often lie in correcting errors and mistakes. However, in many organisations, cybersecurity incidents do not happen, or at least not as often as the focus on incidents would predict. We argue that a focus on what works well, instead of focusing only on the incidents and what went wrong, can provide unique insights into how to improve cybersecurity in organisations. This focus, known as Safety-II in the safety science literature, aims to investigate what end-users, teams and organisations do well and what factors lead to incidents being prevented, or dealt with more swiftly. In this paper, we argue for a Safety-II approach to cybersecurity, and outline various topics of interest along an incident timeline. Furthermore, we discuss a research agenda: Which avenues should be explored further to improve cybersecurity in organisations using a Safety-II approach?

Keywords: Safety-II · Cyber Incidents · Behavioural Cybersecurity · Organisational Cybersecurity · PPDRG-Model

1 The Way of Working in the Cybersecurity Domain

Policy makers, politicians and the media regularly warn about the risk of cybersecurity incidents. The sense of urgency is sometimes expressed by claiming that a cyber or a digital Pearl Harbor [1] or some other form of 'cyber doom' [2] may happen any day now. Interestingly, in everyday life the number of large(r) scale incidents, and especially truly debilitating incidents, is actually very low. In the past 15 years there have indeed been big and impactful incidents such as Stuxnet or NotPetya, but the number of events does not appear to match the level of urgency and the rhetoric of fear surrounding this domain. As a matter of fact, when we take a step back and look at the highly digitalised and interconnected world we live in, it is actually very surprising that there are so few incidents. In the vast majority of cases, citizens, consumers and employees go through their days using an endless array of networked, digital systems that seem to operate with few disturbances from digital attacks or outages. In this paper, we choose to look at cybersecurity incidents at the level of organisations, to get a better understanding of why we witness so few large-scale and debilitating cybersecurity incidents today. Our starting point is that cybersecurity depends on the conjunction of a wide variety of different

D. D. Schmorrow and C. M. Fidopiastis (Eds.): HCII 2024, LNAI 14695, pp. 250–262, 2024.
https://doi.org/10.1007/978-3-031-61572-6_17

elements within organisations, including the digital networked technologies deployed there, the people that work with and use them, the governance landscape in which the organisation finds itself, and the organisational environment and culture within it. To understand better why the number of disruptive cybersecurity incidents is lower than expected, we need to first understand which lens is normally applied to cybersecurity, often without us realising that we do so.

2 From 'What Goes Wrong' to 'What Goes Right': Lessons from Safety Science

Over the past century, safety science has made significant contributions to increasing a broad range of societal domains, including transportation, public health, automation in industry, construction, infrastructure management and disaster management, to name but a few. Risk management has become the dominant paradigm for dealing with risk in public and private organisations [3]. At the same time, it has also received criticism from different directions. For the purpose of this paper, the most important line of criticism we will focus on is the fact that risk management, and safety science in general, tends to focus exclusively on the prevention of incidents by establishing all the many things that might potentially go wrong. Note that in this perspective, the aim is to find oftentimes rare and highly irregular occurrences, digressions from the ordinary that lead to (severe) incidents, and reduce the odds of their materialization. Under normal circumstances, after all, incidents usually do not materialize. Hollnagel provides an example of how this works [4]. Let's say we look at a sample of 10,000 events within a system. In 9,999 cases no incident will arise and operations will continue smoothly but in one out of every 10,000 events something goes wrong and some incident will materialize. By studying the causes of what goes wrong, scholars point out, the focus thus is on finding the one time in which an incident arises, rather than the 9,999 times when it does not [4]. Moreover, when the goal of safety science is to find the root causes of incidents and then remedy these so that future incidents of the same kind will be prevented, then by implication the percentage of the set of 10,000 events that leads to an incident will become ever lower – ultimately nearing zero. Erik Hollnagel calls this approach within safety science 'Safety-I thinking' and explains that this kind of thinking is driven by a 'causality credo': there is a clear and direct causal relationship between a vulnerability and a potential incident, and incidents may be prevented (or at least their likelihood and impact may be reduced) by addressing the vulnerability. This credo is also known as 'find and fix' [4].

Three elements of Safety-I thinking stand out. First, systems can be taken apart and reduced to a set of steps or elements that each play a consecutive role in the working of the whole. Incidents entail that one of the elements of the system has broken down, or that one of the steps has led to a faulty outcome. Second, as a result of this linear approach, systems either function, or they do not. Functioning is a binary thing: a system is either operational or it fails. Third, in the Safety-I perspective, human beings are at best considered to be one of many components in the sequential process, simply another 'cog in the machine'. At worst, they are considered to be the main cause of incidents, because human beings make mistakes [5]. Human beings, in this perspective, are considered to be the ‘weakest link’ in the system [6, 7].

Research has revealed that this perspective does not do justice to how systems within organisations function. Systems, in fact, are far more complex in their workings, and the level of interconnectedness within systems is such that the ‘causality credo’ falls short: many different factors may contribute to the rise of incidents at the same time, as well as to their prevention [8, 9]. A linear interpretation of incident causation, therefore, does not do sufficient justice to the reality of highly complex and interconnected systems. Moreover, systems do not simply function or stop functioning, they do not work or fail in a binary sense. Instead, systems may sometimes fail partially or insignificantly, or they may drift into failure over time [10] or they may degrade gracefully [4]. Both operation and failure are far more complex than a simple ‘on’ or ‘off’. Finally, empirical studies show that human beings, in fact, oftentimes play a crucial role in preventing or stopping emerging incidents: because they are responsive, aware and creative, they may see when dangerous situations arise and step in to stop a chain of events from unfolding. By contrast, machines and devices lack this kind of flexibility: they simply keep going once a process is under way. Rather than seeing human beings as the weakest link in relation to incidents, therefore, one can also argue that human beings may act as the strongest link in the chain when critical situations arise [5].

These arguments show that the Safety-I approach falls short. Rather than focusing on what goes wrong, a more productive way of thinking about safety would be to focus on what goes right. As Hollnagel et al. argue: “the surprise is not that things occasionally go wrong but that they go right so often” [4]. Hollnagel calls this ‘Safety-II thinking’. This perspective starts from the assumption that systems are highly complex and interconnected, and that different elements in a system – including human beings – influence the workings of the whole in multiple ways and directions. Parts of a system may compensate for one another, or take over, or play a role in preventing incidents and thus creating a near miss. Moreover, there is much to be learned from the normal workings of a system: if only one in 10,000 instances of a process a failure arises, then the 9,999 cases when the expected outcome emerges give far more data to study on why and how safety is maintained.

While this paper uses a safety science approach, it is not the only field where learning from ‘what goes well’ is applied to improve understanding and designing effective solutions. For instance, scholars have similarly delved into understanding why, in many instances, individuals, communities, cities, and countries do *not* experience offline crime. The assumption that a motivated criminal is always present raises the question: why are they not invariably successful? Research across these fields has explored various facets of this question.

In a manner akin to the Safety-I principles, criminology often addresses the crime event itself, its execution, and the strategies for its reduction. However, similar to Safety-II advocates, other researchers have posited that some spaces exhibit specific design features that render them more defensible compared to others [11]. From this point of view, individuals, and their interrelations, are perceived as integral protective elements of the system, rather than as potential risks. Furthermore, certain attributes of the social system, such as the collective efficacy within neighbourhoods, have been identified as enhancing the protection and resilience of these areas [12]. Objects can also be designed to improve their resilience to crime. One example of this is the International Mobile Equipment Identity, a 17- or 15-digit code to uniquely identify individual mobile phones. In our discussion around Safety-II principles for cybersecurity, it is therefore important to not only adopt safety science methods to study this approach, but also learn from other disciplines in how to study success in organisational cybersecurity.

3 Putting Incidents Centre Stage: A Model for Cybersecurity in Organisations

When looking at the dominant set of activities with respect to cybersecurity in organisations today, it is striking to see that the majority of effort, whether it is focused on finding and remedying vulnerabilities or changing human behaviour, aims at preventing incidents from arising. Encrypting messages to ensure they cannot be accessed by unintended audiences, scanning networks for intruders and partitioning them so that intruders can only get into the outer shell of the organisation, patching software so that vulnerabilities can no longer be exploited, managing access so that information cannot be stolen – all of these activities are intended to lower, or ideally even fully eliminate, the likelihood of incidents. Safety-I thinking, expressed in risk management and the use of barriers for prevention, is at the heart of cybersecurity practices in organisations today. When viewed on a timeline, currently, one could argue, cybersecurity activities are predominantly focused on what in crisis management has been termed ‘left of bang’ [13] (See Fig. 1).

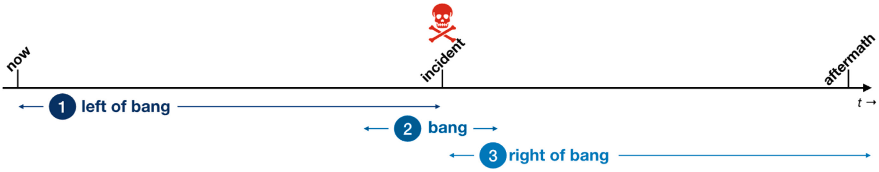

Fig. 1. Timeline of a cybersecurity incident, with subdivisions into left of bang (before the incident), bang (the incident) and right of bang (the aftermath).

However, in recent years there has been a growing awareness that due to the complexity and dynamic nature of cyberspace it is unwise to focus on prevention only as the dominant cybersecurity strategy.

Instead, the Prevent-Prepare-Detect-Respond-Governance (PPDRG) model states that in order to raise cybersecurity in organisations, the latter ought to focus on four different phases along a timeline, of which prevention is only the first [14]. While preventative strategies are certainly important, time and effort should also be invested in the detection of incidents, in preparing for incidents, in incident response and in the governance of all four. The PPDRG model posits these four activities on a timeline (See Fig. 2).

At the far left of the timeline, we find the notion of prevention, as discussed above. The following two phases are sometimes collectively called ‘resilience’. This term refers

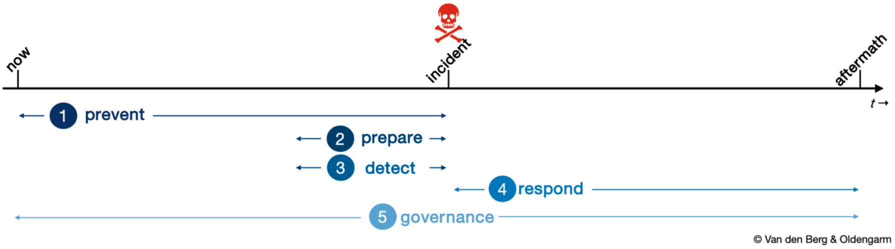

Fig. 2. Timeline of a cybersecurity incident with the various activities mapped on the 'left of bang' and 'right of bang' stages of the timeline.

to organisations' abilities to bounce back from incidents [15]. This entails for instance that organisations need to invest in redundancy in systems so that when one system fails another system can take over [16], or in the development of incident scenarios and large-scale crisis exercises to uncover weaknesses that may hamper quick recovery [17]. When looking more closely at the notion of resilience, it actually falls apart in two different elements. On the one hand, resilience refers to being prepared for incidents. This means for instance that organisations are aware of potential incident scenarios that may arise, and have equipment, manuals, procedures and roles and responsibilities in place that may be activated to deal with the incident as efficiently and quickly as possible. When plotted on a timeline, preparedness involves activities 'just left of bang'. On the other hand, resilience is about being able to recover quickly from incidents, so it involves what happens immediately after an incident materializes, i.e., 'right of bang'. Here, the focus is on incident response, which includes activities such as analysing the incident, containing it, eradicating it and recovering processes and data [18]. Depending on the impact and size of the incident, crisis management activities may also be part of this approach.

One challenging element of cybersecurity incidents is their detection. Intruders may access networks and systems and remain undetected for long periods of time, with a significant window of opportunity to wreak havoc. In 2023 the average time between intrusion and detection was 204 days [19]. This is partially due to the systems used for incident detection. These may provide security analysts with an overwhelming amount of information, often including a high number of false negatives. As a consequence, the skills required for detecting intrusions and emerging incidents are complex, and even for experienced security analysts detecting incidents is difficult [20]. Another reason why detection times are long is that organisations invest too little in this aspect of incident management. For many, the focus is on preventing incidents from happening, with fewer investments made into preparedness, detection and response capabilities. At the same time, incident detection, located immediately 'left of bang,' is where human beings may make crucial differences between (large-scale) incidents and near misses. 'Right of bang' lessons from incidents may lead to improvements in the prevention of future incidents, in preparedness for similar situations, in the detection of anomalies, and improvements on a general level with regard to processes, procedures, and behaviour.

Note that governance aspects are relevant for all stages on this timeline: for prevention, detection, preparedness, recovery and learning. Only with sufficient guidance, with adequate policies and procedures, with funds and available means can organisations implement interventions to increase their cybersecurity maturity. Proper governance facilitates and underpins all phases with regard to cybersecurity incidents.

4 Using the PPDRG Model to Get an Understanding of What Goes Right

The PPRDG model provides a clear-cut overview of the various phases related to incidents and is therefore a helpful starting point for organisation in increasing their ability to prevent and respond to incidents. It is also a useful tool in increasing our understanding of why incidents do not materialize, i.e., why things go right in organisations. This is so, because the insights from Safety-II thinking are not only applicable to the prevention of incidents, but also to detection, preparedness, recovery and governance. After all, human being may step in to detect incidents quickly or curb an emerging incident to prevent it from escalating, policies and procedures may increase resilience by providing guidance on secure and safe processes, and redundancy in (the elements of) systems may ensure that normal operations are solidified. The PPRDG model, therefore, provides a framework that can be used as investigative guidance for all phases that need to be addressed for cybersecurity maturity in organisations. We will explain how this works by discussing examples of research on cybersecurity from as Safety-II perspective in relation to prevention (Sect. 5), detection (Sect. 6), and preparedness and response (Sect. 7). We will end this article with a research agenda for future research into what goes right in cybersecurity using a Safety-II lens (Sect. 8).

5 Prevention from a Safety-II Perspective

5.1 Phishing

Traditionally, phishing is approached from a Safety-I perspective. Organisations are worried about people who click on links in emails and other forms of communication that can harm their organisation by sharing private information, installing malware and/or being a way in for criminals who want to perform a ransomware attack. Attempts to reduce the impact of phishing often rely on awareness campaigns and other means to reduce the 'click rate', the percentage of people who click on a link in a phishing test message [21–23]. In a typical test, 30–50% of employees click on the link in the test email and the organisation is then eager to take steps to reduce the click rate. A successful solution, for example through the use of behavioural change campaigns or training, might result in a click-rate of 20% or less. While the improvement is impressive from a behavioural science point of view, an attacker would on average only need to send out five emails to have one employee clicking on their malicious URL and divulge sensitive information. What changes when we instead focus on a Safety-II perspective in this case? As Safety-II focuses on why incidents do not happen, we would look at the employees who do not click on the URL, or, even more interestingly, the employees that report the phishing email to the relevant in-house expert. In a study on repeated clickers [24], Canham identified not only users who clicked on the link in a phishing test on every single occasion, but also a group of what he considers ‘protective stewards’, people who did not click on any phishing link and often reported these emails to the relevant security experts. Understanding why these people did not click on links from a behavioural science perspective, but perhaps also by looking at organisational and environmental factors can help shape solutions to improve reporting rates and speed, instead of focusing on reducing clicking behaviour.

5.2 Screen Locking and Clean Desk Policy

In addition to phishing, there are of course also other factors that end-users have control over that increase the security of their organisations. For instance, to combat insider threat or the impact of a successful site visit by a social engineer, adhering to a clean desk policy so that no confidential documents are lying around, and locking your screen when leaving your desk so that your system cannot be accessed and abused, are vital aspects of daily habits in the workplace. While attempts can be made to understand the causes of people not following these procedures, more interesting insights can be gained by focusing on the people who always lock their screen and follow the clean desk policy. These people might have taught themselves tricks to ensure that they will lock their screen, or perhaps have imprinted these behaviours into automatic ‘muscle memory’ responses when performing activities. However, it is also possible that they want to lead by example, are (overly) worried about potential risks, or see a personal danger instead of only a risk to the organisation. The latter could take the form of employees being afraid that colleagues might send messages to the board in their name, or can access their salary information when not locking their screen when leaving their desk. By understanding what drives the people who are exemplar employees in this matter, we can develop better cybersecurity solutions.

6 Detection from a Safety-II Perspective

Under the Safety-II paradigm, detection mechanisms are designed not only to identify threats as they hit the organisation, but also to understand the conditions under which its systems operate well. By proactively understanding and monitoring when systems are performing optimally, it should be possible to enable the early identification of deviations from these norms. On a technical level, scholars have examined the efficacy of machine learning and deep learning algorithms in detecting intrusions [25, 26]. The Safety II perspective would contribute to these developments by exploring tools that monitor systems based on evidence of a ‘normal’ situation, thereby allowing organisations to quickly respond to deviations that may indicate a security incident.

Beyond the technical aspects that have predominantly been the focus of incident detection literature, the Safety-II perspective underscores the crucial role of human actors in the detection process. A Safety-II approach involves examining organisations with shorter detection times to understand what sets them apart. Linking back to the phishing example in the previous section, the capability to detect phishing emails could be further supported by a culture of proactive reporting of actual clicking behaviour among employees. If employees are trained to identify threats and the organisation has effective, well-known mechanisms for incident reporting, the likelihood of accurately detecting incidents is significantly increased, and the lengthy average detection time of over 200 days can be reduced substantially. In this scenario, humans are seen not as a risk factor but as a safeguard for the organisation.

7 Preparedness and Response from a Safety-II Perspective

Finally, the PPDRG model advocates for a balanced approach to the preparation and response to cyber incidents, aligning it with the emphasis on prevention and detection. Current literature on the readiness and management of cyber incidents, including crises, is notably limited and predominantly concentrates on the technical facets of incident management such as the application of data analytics during an incident [27, 28]. Existing business continuity and disaster recovery frameworks echo the critique articulated in this discussion. They prioritize risk management during a cyber incident over delineating proven strategies that enhance response efficacy [29, 30]. Limited research in cyber incident management seeks to decipher the repercussions of data breaches on consumer behaviour or market dynamics [31], and to identify optimal strategies for stakeholder communication [32]. Yet, the exploration into organisational traits and employee behaviours that effectively reduce response times and limit organisational damage during a cyber incident remains largely uncharted. It would be advisable to explore, for example, whether it is possible to identify high reliability organisations in the context of incident response based on these characteristics [33]. These organisations, known for their high level of security and safety practices, are not only be better prepared and respond more effectively, but also show evidence of improvement with respect to the pre-incident situation due to effective learning and improved preparation for incidents. We propose to study what works during the preparation and response to cyber incidents, that is, what technical, human, and organisational factors are present when an organisation effectively responds to a cyber incident.

8 A Research Agenda for What Goes Right in Cybersecurity

Approaching cybersecurity from a Safety-II perspective allows us to not only view issues in a new light, but also helps in better understanding what mechanisms are underlying the various secure and risky behaviours and situations. Moreover, it provides insights that can be used to improve cybersecurity solutions across the board. In this section, we outline a research agenda that will help us describe, understand, and improve cybersecurity in organisations. We propose to investigate cybersecurity using a Safety-II perspective on three distinct levels: the end-user level, team level and organisation level. The offered research avenues are a first step towards more research into Safety-II solutions for cybersecurity and we hope that they can be a starting point for the community to further develop these, and other research ideas, from this point of view.

8.1 End-Users

Several steps can be taken to expand our understanding of end-users' cybersecurity behaviour in organisational settings. The earlier mentioned example of phishing is one area where a move towards Safety-II would play a large role in designing new solutions for behavioural cybersecurity topics. The protective stewards, or 'security champions' as others have named them [34, 35] are not only interesting from a behavioural point of view, but also in terms of cognitive processes, attitudes and values that these people hold. Understanding why they behave securely, and differentiating between aspects that are trainable (e.g., habits), and those that are not (e.g., personality) helps in finetuning solutions for both these aspects.

For instance, incorporating these scientific insights in cybersecurity training for end-users would greatly improve training effectivity, as currently little or no scientific underpinnings are present in this field [36]. In terms of aspects that cannot be trained, these can still be of importance when working towards a more secure organisation. Perhaps these skills and abilities can be detected through the use of psychometrical methods. This could improve the security of an organisation in two ways.

First, using these methods can provide input in hiring decisions, as well as decisions as to where to place an individual and which tasks and job role to allocate to them. If a future employee is seen as a security champion, they can be placed in parts of the organisation where security is of the utmost importance, such as in departments that work with the highest level of classified data, or that are working on financial aspects of the organisation. But they can also be put in the position of an exemplar employee that can be a role model for people who are performing less securely. Mechanisms such as social contagion [37] could then lead to higher levels of security within the organisation, not only at the prevention level, but also at for example the detection or response levels.

Second, insights into what makes an employee successful in doing their work securely can help in deciding which behavioural cybersecurity solutions are more likely to be needed to avoid or report (detect) incidents, such as nudging and affordances [38] and a wider focus on access management for instance. On a governance level, they can also lead to an overhaul of existing security policies. Perhaps some policies are too stringent, and employees are likely to be too restricted by them, while they would still perform their tasks in a secure fashion when not hindered by these policies. Knowing what defines a successful end-user when it comes to cybersecurity is vital in developing solutions to make this success sustainable over time, especially when individuals, or whole teams, leave the organisation.

8.2 Teams

Sometimes focusing on individuals will not result in a successful improvement in cybersecurity level within an organisation. However, by adopting a Safety-II perspective, we could also investigate why certain teams (or, on a higher level, whole departments) seem to adopt a higher level of cybersecurity than other parts of the organisation. For instance, perhaps the finance team is more focused on potential scams and fraud due to the nature or the work, or merely because of a focus on detail that might come with the job that acts as a support factor to improve cybersecurity at the same time. Investigating what makes secure teams successful allows us to improve other teams either by instilling the same values, skills and focus as the successful teams, or perhaps a successful team needs to be disbanded and the successful members spread over the other teams to share their insights and ways of working through social contagion principles mentioned earlier.

There are several aspects of teams that can result in high cybersecurity levels compared to other teams. For instance, it is possible that less successful teams suffer from higher levels of social loafing [39, 40], where team members expect that their lack of input or responsible behaviour will be covered by the team as a whole, and that a single person making a mistake is not disastrous. The successful teams might stand out, merely because they have the habit of double-checking decisions made by oth-ers, or feel comfortable asking a colleague for help while the lower performing teams might lack a team spirit or distrust others' perspectives. Furthermore, the successful teams might consist of a better mix of qualities within individual employees. These factors could include personality [41], demographic backgrounds, or past experiences within the same or other organisations. Teams that have a tight-knit community feel-ing might be more successful as asking fellow team members for help might come more naturally, or is encouraged through social processes. Understanding why these teams are successful not only helps in improving training and policies made by organ-isations, but can also be relevant when making hiring decisions by focusing on who would add an important security skill or mindset to an existing team, as explained in the section on the end-user level.

8.3 Organisations

While end-users individually or in teams can help to improve cybersecurity by preventing, detecting, preparing for and responding to incidents, on an organisational level the governance aspect of the PPDRG-model is key. High reliability organisations [42, 43] can be a starting point to improve understanding of what it is that makes organisations successful in dealing with security threats. By identifying these organisations, and studying the potentially unique aspects that make these organisations highly reliable, we can improve the standards being set and the ways of working of other organisations as well. To achieve this, three distinct aspects of successful organisations need to be addressed.

First, the structure of these organisations needs to be investigated. Perhaps the organisational structure allows for more (hierarchical) power for the security department over other departments, or the flow of information within the organisation is improved by a specific set of base rules regarding the organisational structure. Second, there might be specific cultural aspects that successful organisations have incorporated to achieve their level of security. This could be related to a 'just' culture [44], where people are not unduly punished for making mistakes, but where people are encouraged to speak up when they believe they might have made a mistake. But it could also be that employees are more confident in speaking up against their superiors, or feel the support of managers when bringing up doubts about new policies. Third and last, how organisations design, implement, and adhere to policies is of interest. While policies can be designed and implemented top-down, perhaps successful organisations are more likely to only adopt policies when there is wide support for these policies in the organisation. The implementation might also benefit from input from employees to decide how the policy is best implemented and, more importantly, what to do when the demands of the job clash with the policies at hand. To avoid shadow security practices [45], where people work around existing solutions to get the job done, a wider conversation about when to strictly adhere to a policy and when to find alternative solutions might be useful. Using

Safety-II principles to understand what successful organisations do differently is key in understanding which of these elements play a role in improving cybersecurity across the incident timeline.

9 Conclusion

In this paper we argued for a Safety-II approach to cybersecurity to tackle existing cybersecurity issues in organisations. While the need to understand what goes wrong is not diminished, the Safety-I toolbox should be expanded upon with tools using Safety-II principles. More research into understanding why the number of successful attacks is not so high as expected from a Safety-I perspective, and describing the underlying mechanisms of these protective factors will provide the cybersecurity community with valuable insights and, hopefully, tools to better protect the organisations of the future by focusing on the positive: What is going well and what can we learn from that? This question cannot simply be answered from a prevention focus only. Adopting a broad view including the preparation for, detecting of, responding to, and learning from incidents is key if we want to improve organisational cybersecurity across the board. To achieve this, we believe that we should not merely focus on policies, or only on end-user behaviour. Instead, we believe that it is likely to be the interplay between individuals, the teams they operate in, and the wider organisational structure, culture and way of working that is key in learning why some organisations do so well with regards to cybersecurity. Learning from organisational successes instead of failures is not only an exciting new way of approaching cybersecurity, but is also urgently needed to have a strong and effective response to future cyber threats.

References

1. Goldman, E.O., Warner, M.: Why a Digital Pearl Harbor Makes Sense. .. and Is Possible. Understanding Cyber Conflict: Fourteen Analogies (2017)
2. Lawson, S.T., Yeo, S.K., Yu, H., Greene, E.: The cyber-doom effect: the impact of fear appeals in the US cyber security debate. In: International Conference on Cyber Conflict, CYCON (2016). https://doi.org/10.1109/CYCON.2016.7529427
3. Dionne, G.: Risk management: history, definition, and critique. Risk Manag. Insur. Rev. (2013). https://doi.org/10.1111/rmir.12016
4. Hollnagel, E.: From Safety-I to Safety-II: A White Paper (2013)
5. Reason, J.: The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries. Routledge, London (2017). https://doi.org/10.1201/9781315239125
6. Kleinberg, H., Reinicke, B., Cummings, J.: Cyber security best practices: what to do? In 2014 Proceedings of the Conference for Information Systems Applied Research. Univeristy of North Carolina, Baltimore (2014)
7. Dunn Cavelty, M.: Breaking the cyber-security dilemma: aligning security needs and removing vulnerabilities. Sci. Eng. Ethics **20**, 701–715 (2014). https://doi.org/10.1007/s11948-014-9551-y
8. Dahlberg, R.: Resilience and complexity. J. Curr. Cult. Res. **7**, 541–557 (2015)
9. Woods, D.D.: Essential characteristics of resilience. In: Hollnagel, E., Woods, D.D., Leveson, N. (ed.) Resilience Engineering: Concepts and Precepts, pp. 21–34. Taylor and Francis Group (2006)

10. Dekker, S.: Drift into Failure. Taylor and Francis Group (2016). https://doi.org/10.1201/9781315257396
11. Cozens, P.: Crime prevention through environmental design. In: Environmental Criminology and Crime Analysis, pp. 175–199 (2013). https://doi.org/10.4324/9780203118214-19
12. Sampson, R.J., Raudenbush, S.W., Earls, F.: Neighborhoods and violent crime: a multilevel study of collective efficacy. Science **277**, 918–924 (1997)
13. Baskerville, R., Spagnoletti, P., Kim, J.: Incident-centered information security: managing a strategic balance between prevention and response. Inf. Manag. **51**, 138–151 (2014)
14. van den Berg, B., Oldengarm, P.: Een tijdlijn voor het denken over digitale incidenten. In: Handboek Digitale Veiligheid. Wolters Kluwer (2024)
15. Linkov, I., Eisenberg, D.A., Plourde, K., Seager, T.P., Allen, J., Kott, A.: Resilience metrics for cyber systems. Environ. Syst. Decis. **33**, 471–476 (2013)
16. Harms-Ringdahl, L.: Analysis of safety functions and barriers in accidents. Saf. Sci. **47**, 353–363 (2009)
17. Demchak, C.C.: Resilience and cyberspace: recognizing the challenges of a global socio-cyber infrastructure (GSCI). J. Comp. Policy Anal. Res. Pract. **14**, 254–269 (2012)
18. Schlette, D., Caselli, M., Pernul, G.: A comparative study on cyber threat intelligence: the security incident response perspective. IEEE Commun. Surv. Tutor. **23**, 2525–2556 (2021)
19. Petrosyan, A.: Global mean time to identify and contain data breaches 2017–2023. https://www.statista.com/statistics/1417455/worldwide-data-breaches-identify-and-contain/. Accessed 26 Feb 2024
20. Ben-Asher, N., Gonzalez, C.: Effects of cyber security knowledge on attack detection. Comput. Hum. Behav. **48**, 51–61 (2015)
21. Manoharan, S., Katuk, N., Hassan, S., Ahmad, R.: To click or not to click the link: the factors influencing internet banking users' intention in responding to phishing emails. Inf. Comput. Secur. **30**, 37–62 (2022)
22. Sutter, T., Bozkir, A.S., Gehring, B., Berlich, P.: Avoiding the hook: influential factors of phishing awareness training on click-rates and a data-driven approach to predict email difficulty perception. IEEE ACCESS. **10**, 100540–100565 (2022). https://doi.org/10.1109/ACCESS.2022.3207272
23. Quinkert, F., Degeling, M., Holz, T.: Spotlight on phishing: a longitudinal study on phishing awareness trainings. Presented at the Detection of Intrusions and Malware, and Vulnerability Assessment: 18th International Conference, DIMVA 2021, Virtual Event, 14–16 July 2021, Proceedings 18 (2021)
24. Canham, M.: Repeat clicking: a lack of awareness is not the problem. In: Degen, H., Ntoa, S., Moallem, A. (eds.) HCII 2023. LNCS, vol. 14059, pp. 325–342. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-48057-7_20
25. Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., Marchetti, M.: On the effectiveness of machine and deep learning for cyber security. Presented at the 2018 10th International Conference on Cyber Conflict (CyCon) (2018)
26. Geetha, R., Thilagam, T.: A review on the effectiveness of machine learning and deep learning algorithms for cyber security. Arch. Comput. Methods Eng. **28**, 2861–2879 (2021)
27. Naseer, A., Naseer, H., Ahmad, A., Maynard, S.B., Siddiqui, A.M.: Real-time analytics, incident response process agility and enterprise cybersecurity performance: a contingent resource-based analysis. Int. J. Inf. Manage. **59**, 102334 (2021)
28. Naseer, A., Naseer, H., Ahmad, A., Maynard, S.B., Siddiqui, A.M.: Moving towards agile cybersecurity incident response: a case study exploring the enabling role of big data analytics-embedded dynamic capabilities. Comput. Secur. **135**, 103525 (2023)
29. Zare, H., Wang, P., Zare, M.J., Azadi, M., Olsen, P.: Business continuity plan and risk assessment analysis in case of a cyber attack disaster in healthcare organizations. Presented at the

17th International Conference on Information Technology–New Generations (ITNG 2020) (2020)

30. Järveläinen, J.: IT incidents and business impacts: validating a framework for continuity management in information systems. Int. J. Inf. Manage. **33**, 583–590 (2013)
31. Algarni, A.M., Malaiya, Y.K.: A consolidated approach for estimation of data security breach costs. Presented at the 2016 2nd International Conference on Information Management (ICIM) (2016)
32. Kuipers, S., Schonheit, M.: Data breaches and effective crisis communication: a comparative analysis of corporate reputational crises. Corp. Reput. Rev. **25**, 176–197 (2022)
33. Weick, K.E., Sutcliffe, K.M.: Managing the Unexpected. Jossey-Bass, San Francisco (2001)
34. Beris, O., Beautement, A., Sasse, M.A.: Employee rule breakers, excuse makers and security champions: mapping the risk perceptions and emotions that drive security behaviors. Presented at the Proceedings of the 2015 New Security Paradigms Workshop (2015)
35. Gabriel, T., Furnell, S.: Selecting security champions. Comput. Fraud Secur. **2011**, 8–12 (2011)
36. Prümmer, J., van Steen, T., van den Berg, B.: A systematic review of current cybersecurity training methods. Comput. Secur. 103585 (2023)
37. Christakis, N.A., Fowler, J.H.: Social contagion theory: examining dynamic social networks and human behavior. Stat. Med. **32**, 556–577 (2013)
38. van Steen, T.: When choice is (not) an option: nudging and techno-regulation approaches to behavioural cybersecurity. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) HCII 2022. LNCS, vol. 13310, pp. 120–130. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-05457-0_10
39. Harkins, S.G.: Social loafing and social facilitation. J. Exp. Soc. Psychol. **23**, 1–18 (1987)
40. Simms, A., Nichols, T.: Social loafing: a review of the literature. J. Manag. Policy Pract. **15**, 58 (2014)
41. Sackett, P.R., Walmsley, P.T.: Which personality attributes are most important in the workplace? Perspect. Psychol. Sci. **9**, 538–551 (2014)
42. Sutcliffe, K.M.: High reliability organizations (HROs). Best Pract. Res. Clin. Anaesthesiol. **25**, 133–144 (2011)
43. Roberts, K.H., Bea, R.: Must accidents happen? Lessons from high-reliability organizations. Acad. Manag. Perspect. **15**, 70–78 (2001)
44. Dekker, S.W.: Just culture: who gets to draw the line? Cogn. Technol. Work **11**, 177–185 (2009)
45. Kirlappos, I., Parkin, S., Sasse, M.A.: "Shadow security" as a tool for the learning organization. ACM SIGCAS Comput. Soc. **45**, 29–37 (2015)