The technopolitics of cybersecurity

Sociotechnical configurations, epistemic devices, and the production of knowledge

Aanonsen, C.E.

Publication date
2025

5

Chapter five: Risk assessments in cybersecurity and counterterrorism[20]

Box 5.1 Risk assessments as sociotechnical configurations

Whereas Chapter 4 was focused on how automation and machine learning actively (re)configures cybersecurity knowledge and practice, this chapter shows how this comes to light also in risk assessments. I do so through a comparative analysis of assessments in cybersecurity and counterterrorism. The chapter does not focus the comparison on similarities and dissimilarities per se, but on how difference comes to show that sociotechnical configurations are situated and contextual. When analysing risk assessments as configured across different systems – exemplified here through the U.S. intrusion detection system EINSTEIN and parts of the UK counterterrorism programme Prevent – the chapter shows the following: Whereas Prevent more clearly presupposes normative subjects and standards of (in)security, EINSTEIN's anomaly detection engenders threat not as a binary or normative distinction but a separate category of risk. The analysis underscores what situated and contextual assessments *do* for the production of categories of risk, vulnerability, and threat. It shows how 'meaning-making' and 'sense-making' are processes that shape both security responses and timelines through which risk is conceptualised in different ways. Moreover, it reveals that rather than being fixed, risk and (in)security are perpetually co-produced with the tools used for assessment. This highlights that production of knowledge about cyber(in)security hence has little to do with real or imagined risk but rather emerges from a particular configuration of social, political, and technological relations.

[20] With some minor formatting (e.g., editing section titles) and grammatical edits (e.g., 'article' has been replaced by 'chapter'), this chapter was previously published as: Aanonsen, C. E. and Augestad Knudsen, R. (2025) 'Riskification and the production of threat: a comparison of risk assessments in cybersecurity and counter-terrorism', *Critical Studies on Security*, pp. 1–26. Available at: https://doi.org/10.1080/21624887.2025.2509343.

5.1 Introduction

It is well established that security practices have become increasingly oriented towards governing risk (Aradau and van Munster, 2007; Beck, 1992, 2009; Foucault, 2007; Giddens, 1990). Developing into practices of 'pre-emptive security', this seems to entail a focus shift away from only seeking to neutralise imminent threats or building resilience so as to enable recovery from attacks to embeddedness within anticipatory security regimes that seek to neutralise potential threats before they materialise (Aradau, Lobo-Guerrero and van Munster, 2008; de Goede, 2012, 2008; Massumi, 2007). Across Critical Security Studies (CSS) and Science and Technology Studies (STS), scholars have examined this in the context of classifying and targeting individuals and societies (Amoore and de Goede, 2008; Chamayou, 2014; de Goede, 2012), counterinsurgency (Anderson, 2011) and the (re)configuration of military and intelligence practices (Reichborn-Kjennerud, 2025; Suchman, 2023). Despite scholarly differences, it is generally agreed that security is now more framed in terms of risk, with a broadening logic driving a precautionary security agenda (Friis and Reichborn-Kjennerud, 2016). The historical starting point for this has been located in the post-Cold War era and, above all, the 9/11 attacks and the subsequent 'war on terror', prompting reconceptualisations of security as a matter of managing catastrophic possibility rather than the threat itself (Aradau, Lobo-Guerrero and van Munster, 2008).

The turn from pinpointing threats to handling the emergence of risk has meaningfully been theorised as riskification (Corry, 2012). Rooted in critical risk literature and securitisation theory, scholarship on surveillance and terrorism has explored riskification as a matter of arranging security practices towards methods of precaution and prevention (Amoore, 2013; Hintz and Milan, 2018; Lyon, 2015). Digitalisation and the emergence of cybersecurity has also been studied in terms of riskification (e.g., Backman and Stevens, 2024; Friis and Reichborn-Kjennerud, 2016; Stevens, 2022). Within these contributions, riskification characterises a way of engaging with cyber 'problems' that are considered as inherent security risks due to the complex, uncertain, unpredictable, and invisible 'nature' of cyberspace (Friis and Reichborn-Kjennerud, 2016, p. 33; see also Brantly, 2021). With the target of riskification being risk rather than threat, it is more elusive and presents different opportunities for counteraction than the better-known theory of securitisation. Riskification does not mandate a state of emergency (Corry, 2012). As

already argued by earlier risk literature, such reorientations entail a continuous co-production of material-discursive practices of both responses and their targets (Aradau, 2010), with threat conceptualisations and ideas of how they may materialise co-produced by the measures initiated against them.

Illustrations of these theoretical reorientations include how, in the early 2000s, the field of counterterrorism moved towards risk and pre-emption, sometimes through profiling 'risky individuals' as part of the war on terror (Amoore, 2013). In a similar manner, information security was also reimagined from something perimeter-based – keeping threats outside the 'firewall' – to one of 'zero-trust' architectures, with apparatuses assuming danger and seeking its elimination through constant monitoring and encryption-enabled verification (Spencer and Pizio, 2024). As described in the previous Chapter 4, systems for intrusion detection – hardware and software solutions designed to monitor networks or systems for malicious activity – have embraced the detection of 'anomalies' alongside what is known as signature-based detection. Signature-based detection resembles counterterrorism's pre-emptive profiling, in that it identifies threats based on predefined profiles or signatures, creating a threshold for distinguishing 'normal' from 'abnormal' behaviour. Anomaly detection powered by machine learning, however, goes beyond this logic by notifying irregularity through singling out deviations from the data pattern, continuously updating the system's differentiation between normal/abnormal (Aradau and Blanke, 2018). While operating different riskification logics, these developments have both entailed expanding attention beyond imminent attacks to a logic of governing (potential) risk that is assumed to be constant, infinite, and open-ended (see e.g., Collier and Lakoff, 2008). However, significant differences exist in the systems and practices that have emerged in each field in response to riskification. The computational and increasingly embodied automation in which riskification has played out in specific systems for cybersecurity, such as intrusion detection, and how these have both mirrored and contrasted with the operationalisation across other areas of security have so far not received much attention.

This chapter explores how these logics of riskification have manifested in systems within cybersecurity and counterterrorism, focusing on the U.S. government intrusion detection system EINSTEIN – a sensor-based system for detecting and alerting potentially malicious behaviour posed against government agency networks – and comparing it to the more 'conventionally' riskified area of counterterrorism, exemplified by practices of risk

assessment within the UK's Prevent programme. While cybersecurity and counterterrorism share some features, these fields' conditions of possibility for rethinking security in terms of risk and pre-emption have systemically evolved in dissimilar directions. This includes how sociotechnical configurations are contextual and therefore situated in different 'ways of knowing' threat, risk and (in)security (Aradau and Blanke, 2018, 2022; Pettinger, 2020; Suchman, 2006). What is at stake here is that this results in diverging conceptualisations of risk, with implications for the production of threat and (in)security. Rather than being fixed, risk and (in)security are perpetually co-produced with the tools used for assessment. I argue that the production of (in)security hence has little to do with real or imagined risk, but rather emerges from particular sociotechnical configurations. At the core of this difference is cybersecurity intrusion detection method's move from dichotomies of normal/abnormal and friend/enemy towards detecting 'anomalies' or 'unknown unknowns' as a third category (Aradau and Blanke, 2018, p. 4), manifesting a different security epistemology than that of counterterrorism. Whereas risk assessments within counterterrorism are based on a normative reading of threat and search for essentialised 'risky subjects', the vantage point of anomaly detection departs from a normative stance, instead giving precedence to dynamic behavioural patterns of irregularity as a separate analytical category of risk. As such, threat is here empirically – albeit not normatively – produced (Reichborn-Kjennerud, 2025). The 'risk of threat' is identified as irregular data rather than alerts matching predefined normative categories of potential risk.

This chapter continues in Section 5.2 with a brief outline of theorisations of risk and riskification relating to both fields of cybersecurity and counterterrorism. In Section 5.3, I then turn to the case of specialised risk assessments within counterterrorism as a distinct form of riskification, specifically the UK counterterrorism programme Prevent's use of risk/vulnerability/susceptibility assessment tools, which embodies a long-established form of riskification by focusing on subjects' ('internal') characteristics and pinpointing individuals 'at risk'. Following on from this, Section 5.4 explores riskification within cybersecurity, as a backdrop for contextualising the EINSTEIN intrusion detection system and its juxtaposition with Prevent. EINSTEIN encapsulates well the riskification of cybersecurity: Intrusion detection systems (IDS) have become core to critical infrastructure protection, attempting to mitigate threats and reduce system and network vulnerability. Since its inception in 2003, EINSTEIN has evolved through three versions until its apparent discontinuation in 2024 (Department of Homeland Security, 2023). The historical evolution

of EINSTEIN reveals a transition towards embracing anomaly detection that continuously produces the 'risk of threat' with no presupposition of a 'normal'.

After juxtaposing these two systems, Section 5.5 conceptualises the different riskification logics in play in practices of meaning-making and sense-making. While counterterrorism systems for risk assessment encapsulate meaning-making, seeking to capture known threats rooted in a normative reading of risk through interpretive categorisation, cybersecurity practices of anomaly detection as a distinct system increasingly embodies sense-making, processes that continuously adapt and evolve the meaning of risk in response to renderings of more data. Ultimately, while riskification practices aim to know and neutralise threats, their own anatomy and operation have largely gone below the radar. This chapter contributes to this thesis' overarching aim, as well as critical risk/security scholarship by taking into consideration how STS work regarding sociotechnical configurations can strengthen empirical research on increasingly computational security/risk domains that render what 'counts' as threat and (in)security in new and different ways. By doing so, I highlight the need for sensitivity not only to the adoption of computational/automated riskification practices and their consequences, but how those riskification practices adapt, evolve, and possibly expand over time, even beyond intention: Future-oriented risk logics emerge when different practices and methods are applied to manage 'new' and 'old' problems.

5.2 Risk and riskification

International Relations (IR) and security studies have for decades theorised risk and its relationship to threats (Aradau, Lobo-Guerrero and van Munster, 2008; Buzan, Wæver and de Wilde, 1998; Corry, 2012). This became a topic of particular interest to IR scholarship in the 1990s post-Cold War environment, when risk was conceptualised as a way of broadening ideas of 'security' and threat by including the logics of areas such as terrorism, crises, and war (e.g., Collier and Lakoff, 2008). Anthony Giddens and Ulrich Beck's influential sociological theorisations of 'risk society' argued that, as the bipolar world faded, dangers were increasingly framed as risks, defined by their probability and potential consequences. In their view, risks are both real and socially constructed, emerging through assessments of future dangers and decisions on whether to act. They saw contemporary societies as governed through incalculable risk (Beck, 1992, 2009; Giddens, 1990). In

contrast, Foucauldian governmentality perspectives pay attention to everyday practices, such as technologies of risk management and corporate logics of routine (Aradau and van Munster, 2017; Backman and Stevens, 2024). They view risk as a tool of governance to expand neoliberal control mechanisms, by legitimising regulatory regimes that portray the future as calculable and manageable (see also Dean, 1999). Both approaches express concern over the expanding security agenda driven by precautionary principles and pre-emption and are ultimately concerned with the centrality of risk to contemporary security practices (Friis and Reichborn-Kjennerud, 2016).

Following on from these seminal contributions, others have concentrated on the role of risk conceptualisations to states of exception, legitimising expansions of what could be done in the name of security and minimising risk (Agamben, 2018; Buzan, Wæver and de Wilde, 1998; Huysmans, 2006). Somewhat relatedly, scholars referencing Foucault's ideas of biopolitics have criticised the production of vague yet urgent risks that enable authorities to exert control over ever-increasing societal domains. As per a suggested distinction between 'manufactured uncertainties' and 'calculable risk' (Beck, 2009), the former has been especially posited as particularly prone to instrumentalisation by authorities and private sector actors (e.g., Krahmann, 2011).

Investigating how risk dynamics have shaped security within contemporary societies, many scholars consider 9/11 and the subsequent war on terror as key to political action (Dunn Cavelty and Kristensen, 2008). Historically oriented CSS literature has zoomed in on how 9/11 prompted a refocusing of security efforts that target the timeline of assumed paths towards materialisation of threat. De Goede (2008), for instance, looks at the 9/11 Commission's report, which insists that a failure of imagination played into the authorities' inability to stop the attacks. This conclusion led authorities worldwide to rethink security and threat in terms of imagined worst-case scenarios and adjust their focus towards low-probability, high-consequence events and still-to-emerge threats (de Goede, 2008). This again encouraged acting on risks that were only just emerging, as if they were a danger about to materialise (e.g., Clarke, 2006; Posner, 2004). Some have also emphasised that networked, interconnected societies and the continuous sophistication of interdependent infrastructures has rendered them increasingly vulnerable (Barry, 2001; Beck, 1992). This has aided the "zeitgeist of the risk society" (Dunn Cavelty and Kristensen, 2008, p. 2) whereby governance focuses on the fear of future scenarios (Bigo, 2002; de Goede, 2008).

Collier and Lakoff (2008) historicise the security paradigm shift in the wake of the Cold War. They show how existing security frameworks aimed at deterrence and military strategy were considered inadequate to address emerging threats, including terrorism, energy crises, and technological failures (see also Aradau, Lobo-Guerrero and van Munster, 2008; Aradau and van Munster, 2007). Security efforts increasingly prioritised the identification and mitigation of systemic vulnerabilities, leading to the adoption of pre-emptive and preventive measures that integrated risk assessment techniques like vulnerability mapping and systems analysis to safeguard vital networks from potential disruptions. By the late 20th century, this approach had evolved into what they call 'vital systems security', redefining national security as the continuous management of infrastructural resilience through anticipatory governance, contingency planning, and crisis management (Collier and Lakoff, 2008).

These lines of analysis have since spawned further scholarship on the anatomy of the conceptual shift from threats to risks. The reorientation in question seems paradigmatic in conceptualising security action, from halting plots and imminent attacks to prevention, and then onto ever earlier stages of pre-emption and precaution (Bracken, Bremmer and Gordon, 2008). Within this framework, the target for intervention is no longer the calculated probability of a specific attack or manifest danger, but rather its mere imagined possibility, as theorised by Louise Amoore (2013). It is hence the risk calculus, rather than the risk itself, that has changed with this shift. While conventional risk management would seek information to reduce uncertainty, the pre-emptive logic of possibility is oriented towards the capacity to act on the basis of the unknown and the would-be suspicious (Amoore, 2013, pp. 57–58). Significantly, this shifts the burden of proof, from those arguing to act because something is dangerous to those arguing in favour of not acting because the situation is too unclear, with the unknown becoming an indicator of risk, and hence impetus for action.

To capture the practice of risk as security governance, Olaf Corry (2012, p. 238) highlights how 'riskification' differs from securitisation. He conceptualises it as a "social process of constructing something politically in terms of risks". As security governance, risk does not focus on imminent threat but rather on preventing threats from materialising. Corresponding to Foucauldian governmentality perspectives on regulatory regimes, institutionalised practices such as surveillance and profiling are discernible examples of increasingly normalised customs of data-based populational governance (Amoore and de

Goede, 2008; Aradau and van Munster, 2007; Dillon, 1996; Dillon and Reid, 2001, 2009; Hintz and Milan, 2018).

As will be shown, both counterterrorism and cybersecurity represent contemporary security 'programmes' that have manifested risk-security governance thinking in the 21st century. National and international counterterrorism strategies have in part been built on principles of profiling patterns of characteristics to determine 'types'. These types are not only based on binary friend/enemy dichotomies but an amalgamation of risky and dangerous features (Amoore and de Goede, 2008; Corry, 2012). With the motivation to 'discover' normative threats, a preordained causal path is laid out, labelling conditions of possibility for danger. Cybersecurity has been subject to a transition towards precautionary governance in a similar manner. As shown in Chapter 6, unlike counterterrorism's 9/11 watershed moment, cybersecurity's transition into a core security domain has been gradual, shaped by sociotechnical imaginaries of existential threat scenarios (Gjesvik and Szulecki, 2022; Jasanoff and Kim, 2015). Since WWII, imaginaries of destructive attacks on critical infrastructure have informed pre-emptive and preventative measures. These measures have increasingly become oriented towards identifying risks on the cusp of emergence rather than responding to concrete plots (Collier and Lakoff, 2008).

Resembling the style of counterterrorism strategies, cybersecurity measures are also "premised on the belief that risks can be classified, quantified and thereby predict possible futures to be managed" (Friis and Reichborn-Kjennerud, 2016, p. 12). Overall, responses and countermeasures to digitalisation and threats to network and system vulnerabilities have been based on mitigating potential scenarios. However, as will be argued in the case of EINSTEIN, systems and practices for network intrusion detection have evolved from signature-based risk assessments to accounting for threat and risk based also on seeking the 'unknown' – or the anomaly. Handling potential danger in this domain does not presuppose risk as something calculable but co-produces security practices and danger simultaneously, enabled and constrained by the materiality of system configurations and networks interplaying with how the uncertain, unpredictable, and invisibility of digital infrastructures is experienced (Aradau, 2010). The following Sections 5.3-4 tease out how riskification has come to characterise modalities of systems for counterterrorism and cybersecurity – first by turning the focus towards the sociotechnical underpinning of risk assessment practices that highlight ways of 'knowing' danger, then through a study of

EINSTEIN as emblematic of a transition in risk assessment practices from signature-based intrusion detection to anomaly detection.

5.3 Risk, counterterrorism and Prevent

The field of counterterrorism – domestic and international – represents a continuing embodiment of the logics of risk and riskification in ways that are relevant to understanding contemporary cybersecurity (see e.g., Amoore and de Goede, 2008, p. 9; Amoore and Raley, 2017; de Goede, 2008; Rasmussen, 2006, pp. 38–39). Since the early 2000s, counterterrorism has become the paradigmatic field in which these developments have played out. One obvious example is the way in which the terrorism legislation of many countries increasingly targets earlier stages of the timeline assumed to lead to a violent terrorist attack by including actions conceived as prior and precursor acts, based on the idea of certain kinds of risk materialising into 'hard' threats. This is similar to how 'de-perimeterisation' of information security has come to embrace the 'zero-trust' architecture, based on the assumption that inherent threats exist both inside and outside networks, enforcing a 'default-deny' approach (Spencer and Pizio, 2024). The UK's 2006 Terrorism Act (TACT), for instance, expanded the definition of terrorism offences to include various preparatory acts, including those not involving violence or plans for attack, such as disseminating propaganda, collection of materials, and/or providing training or material support assumed to be terrorism-related (Terrorism Act, 2006; see also Gelev, 2011). Several other European states have criminalised similar offences, including ones relating to foreign fighting around the time of ISIS declaring its 'caliphate' in 2014 (e.g., Sweden in 2016).

Whereas these legislative developments in themselves could be seen to encapsulate pre-emptive and anticipatory logics in ways that could be summed up as riskification, this trend has been even more pronounced in the field of preventing terrorism. Indeed, by expanding interventionist attention beyond stopping looming attacks and halting immediate danger – and towards thwarting the conditions in which terrorism risk (and vulnerability) might emerge – preventive counterterrorism could in itself be framed as a practice of riskification. While counterterrorism or terrorism prevention have thus far not been discussed in the explicit language of riskification, numerous critical scholars have looked to the UK and its Prevent programme as a particularly clear example of how the logics of

pre-emption and precaution have been enacted in practice (e.g., Heath-Kelly 2017a; Sian, 2017; Thomas, 2024; Younis, 2021). Since 2015, the 'Prevent Duty' – which obliges educational, health, and other state institutions to report suspicions of 'radicalisation' to the authorities – has been harshly criticised, as it introduces a national security mandate into sectors that has been historically more centred on individual welfare and care (see Thomas, 2024).[21] Prevent is the part of UK counterterrorism strategy that seeks to address and combat the ideological causes of terrorism, intervene early to stop radicalisation, and help convicted terrorists rehabilitate (HM Government, 2023a, p. 8). Within the Prevent programme, individuals are followed up within a 'safeguarding' framework that seeks to steer them away from ideological radicalisation,[22] after they have been referred to authorities over such concerns.[23]

Clearly, for Prevent – and elsewhere within counterterrorism – ideas of 'radicalisation' and its causal role in terrorism has been instrumental in both 'mainstreaming' (Augestad Knudsen, 2023) counterterrorism into a vast array of societal domain, and in perpetuating and operationalising the riskification logic theorised by the scholarship cited above. Radicalisation is still an unclear and disputed notion that is nonetheless associated with significant risk warranting decisive state intervention. Countries and organisations tend to define radicalisation in terms of "the idea of a process through which an individual adopts an increasingly extremist set of beliefs and aspirations", without much clarity on how this should be identified or measured (Augestad Knudsen, 2018; UNODC, 2018). Nonetheless, the mainstreaming of counterterrorism through counter-radicalisation has entailed an explicit combating of 'pre-crime' (McCulloch and Wilson, 2016), with the increasing enlistment of non-security actors not only within education and health, but also social care and border control, in reporting and acting on (possible) radicalisation-related concerns (Aked 2020, Heath-Kelly 2024; HM Government, 2023b; Shawcross 2023). Much of this has happened within established frameworks and terminologies for 'safeguarding' and vulnerability, although the UK's

[21] For the most recent Prevent duty guidance, see HM Government (2023a).
[22] Of the nearly 7,000 individuals referred in the year to 2023, around 63% were below the age of 20, and 39% below 16. See UK Home Office (2023a).
[23] For the most recent edition of the UK counterterrorism strategy CONTEST, see HM Government (2023b).

CONTEST strategy also explicitly cites the 'risk of radicalisation' as a target, as an assumed precursor to terrorism (see e.g., HM Government, 2023b, p. 30).

While both Prevent and counterterrorism's overall evolving risk conceptualisations have been studied and discussed vigorously, less attention has been paid to how the logic of riskification through the idea of radicalisation has been materially operationalised with the practice of terrorism-related risk assessments. Emerging in the mid 2000s, specialised counterterrorism risk assessment tools became less concerned with acute and clearly identifiable threats. Instead, these tools have become focused on conditions enabling the emergence of risk based on individual characteristics and vulnerabilities, in part embodied in the idea of radicalisation (see Corry, 2012). In other words, many of the tools – including those used within Prevent – targeted not primarily emerging terrorist behaviours or, for instance, propensity for interpersonal violence, but rather (cognitive and attitudinal) radicalisation as both relevant to capturing risk, and as a proper locus for intervention. Examples discussed below include the Extremism Risk Guidelines 22+ and the Vulnerability Assessment Framework. The centrality of these tools to post-9/11 counterterrorism was based on the assumptions that, first, terrorism risk is ontologically and epistemologically distinct from other forms of risk, and second, that 'radicalisation' is central to seeing, assessing and knowing such risk. The 'Unique Selling Point' of such specialised tools appeared to be their presumed ability to pinpoint extremist thought and mindset, rather than just the likelihood of someone committing a terrorist offence (e.g., Lloyd and Dean, 2015; Pressman, 2009; RAN Centre of Excellence, 2017; Singh, Bjørkly and Fazel, 2016).

In this way, specialised counterterrorism risk assessment tools work as systems embodying the post-9/11 reconceptualisation of risk thinking and practice, from pinning down the probability of threat materialisation to enabling actions on the basis of uncertainty and vague possibilities (Amoore, 2013, p. 58). In their de facto focus on radicalisation and enabling pre- (or rather non-)criminal action – to neutralise the conditions from which risk was expected to grow more concrete – such tools also epitomise the logic and practice of riskification. Encouraged by the promise of such tools to order, measure, and quantify risk often considered as too intangible to enable action, countries around the world came to rely on a variety of such tools in their counterterrorism efforts.

To further illustrate this point, it is useful to look a bit closer at the Channel programme, which was officially established as a cornerstone of Prevent in 2012 (UK Home Office, 2023b). Within Prevent, Channel ensures follow-up and typically face-to-face interventions with individual Prevent referrals. According to the UK Home Office, Channel "is about safeguarding children and adults from being drawn into terrorist-related activity. It is about early intervention to protect and divert people away from the risk they face before illegality occurs" (HM Government, 2012, p. 2). In order to identify which individuals should receive "support to address their vulnerability to radicalisation", and to determine what sort of intervention would be most appropriate (ibid.; see also Heath-Kelly, 2017b), Channel previously relied heavily on the assessment tool called Vulnerability Assessment Framework (VAF).

While originally framed in terms of vulnerability rather than risk, Channel remains focused on supporting and safeguarding individuals to stop them becoming involved in terrorism. VAF is a direct transfer of the 22 explicitly named risk indicators from the specialised risk assessment tool Extremism Risk Guidelines 22+ (ERG22+). The ERG22+ was developed by the UK prison and probation service after 2006, based on – and in order to assess – convicted terrorism offenders ('true positives') (Lloyd and Dean, 2015). When rolled out in UK prisons, it was also used to assess other imprisoned offenders considered at risk of committing a terrorism offence in the future (ibid.). In the VAF version, the same 22 indicators originating in the ERG22+ are identically grouped in the same three categories of 'engagement', 'intent', and 'capability' (see Figure 1). Transferring these indicators directly from a risk assessment tool designed for convicted terrorism offenders to a context involving non-criminal, often very young, individuals deemed vulnerable rather than dangerous would seem to present significant challenges. This shift risks generating both false negatives and false positives, leading to misidentification and unwarranted interventions.

Vulnerability Assessment Framework (VAF)		
1. *Engagement* with a group, cause or ideology Engagement factors are sometimes referred to as "psychological hooks". They include needs, susceptibilities, motivations and contextual influences and together map the individual pathway into terrorism). They can include: • Feelings of grievance and injustice • Feeling under threat • A need for identity, meaning and belonging • A desire for status • A desire for excitement and adventure • A need to dominate and control others • Susceptibility to indoctrination • A desire for political or moral change • Opportunistic involvement • Family or friends involvement in extremism • Being at a transitional time of life • Being influenced or controlled by a group • Relevant mental health issues	**2.** *Intent* to cause harm Not all those who become engaged by a group, cause or ideology go on to develop an intention to cause harm, so this dimension is considered separately. Intent factors describe the mindset that is associated with a readiness to use violence and address what the individual would do and to what end). They can include: • Over-identification with a group or ideology • 'Them and Us' thinking • Dehumanisation of the enemy • Attitudes that justify offending • Harmful means to an end • Harmful objectives	3. *Capability* to cause harm Not all those who have a wish to cause harm on behalf of a group, cause or ideology are capable of doing so, and plots to cause widespread damage take a high level of personal capability, resources and networking to be successful. What the individual is capable of is therefore a key consideration when assessing risk of harm to the public). Factors can include: • Individual knowledge, skills and competencies • Access to networks, funding or equipment • Criminal Capability

Table 5.1. VAF indicators. Drawn from "Channel: Vulnerability assessment framework" by HM Government (2012).

Prevent's operationalisation of the 'ERG22+ as VAF' also reveals a perpetuation of a form of riskification where the target of security attention (i.e., threat) is not only conceived of as individual attitudes and thinking as direct motivational precursors for terrorist offences, but also through psychological makeups and 'profiles' as the potential, still-to-emerge condition that may give rise to such motivations in the future (see also The Metropolitan Police Service, n.d.). The VAF's original foregrounding of vulnerability rather than risk in enabling interventive action is worth underlining here. Notably, the language of 'vulnerability' has recently been removed from the official vocabulary of UK counterterrorism after a review of Prevent in 2023 that criticised the programme's "excessive focus on vulnerability and insufficient attention paid to individual agency" (Shawcross, 2023, p. 116). The language of vulnerability has since been replaced with 'susceptibility'. VAF has been amended and renamed PAF (Prevent Assessment Framework), however, with uncertain practical implications – and with Prevent's underpinning references to 'risk of radicalisation' remaining (see also Macdonald, Whiting and Jarvis, 2024; Walker and Skoczylis, 2023).

Since their inception more than a decade ago, several of these specialised counterterrorism risk assessment tools have become the subject of critique and controversy. Nonetheless, reliance on them persists. Practitioners report finding them useful, with relatively high internal scores on both user satisfaction and interrater reliability (Elliot, Randhawa-Horne and Hambly, 2023; Lid, 2023; Powis et al., 2019). Since data associated with their use are generally inaccessible to independent researchers, and so-called base rates for their target groups remain low, proper validation on par with more established risk assessment tools (for instance for violence) is difficult (e.g., Sarma 2017). More fundamentally, the categories that these risk assessment practices aim to capture are vague, contested, and appear ill-suited for the near-total risk quantification that these tools claim to deliver. In light of terrorism's sociopolitical salience, such tools' pretence of pinning down whom among a subset of 'unknowns' pose the greatest 'known' (terrorism) risk based on assumed vulnerability profiles might in fact be their greatest appeal to authorities set on counterterrorism action. Although indictors from tools such as VAF are numbered and calculable, such tools encourage the use of the professional judgement[24] and systematised

[24] They are Structured Professional Judgement (SPJ) tools, relying on assessors' competence and experience, rather than being purely numerical/summative.

instincts of assessors, and offer them the opportunity to make risk seeable, assessable, and knowable (see also Pettinger, 2020).

In terms of analytical frameworks, the riskification of specialised terrorism-related assessment tools, driven by the concept of 'radicalisation', mirrors the way signatures for intrusion detection are made (such as with EINSTEIN's early versions) based on a threshold or 'rules' distinguishing normal from abnormal network behaviours. As with signature-based intrusion detection, terrorism-related assessment tools embody riskification by concentrating on the conditions enabling risk to emerge, including notions of vulnerability/susceptibility, based on calculated dichotomies of normality/abnormality, friend/enemy, as well as essentialised identity, quantification, typologisation, and 'profiling'. There are also other dimensions of difference. First, in technical and operational integration, IDS can trigger automated defensive responses, such as blocking malicious traffic, while human risk assessments require careful human oversight due to the high stakes of acting on individuals. Second, ethical and legal constraints are far greater for profiling individuals – as actions directly affect civil liberties and principles of transparency, fairness, and accountability – whereas IDS primarily focus on system protection with fewer ethical or legal hurdles. Third, feedback loops in IDS are rapid, allowing continuous system calibration, while human risk assessments face ambiguous outcomes, making improvement slow and uncertain. Finally, their data inputs differ fundamentally: IDS use structured, machine-generated technical data, while individual risk assessments rely on frequently sparse, qualitative human behavioural and social data, introducing subjectivity and interpretive complexity.

With these differences in mind, the sections below will zoom in on differences in analytical frameworks, that is, the contrasting sociotechnical configurations underpinning later versions of EINSTEIN (3 and 3A). It shows a shift in cybersecurity methods in which anomaly detection (identifying irregularities) precedes the development of thresholds for risk assessments (signatures). Thus 'meaning making' (knowing the conditions enabling risk to emerge) in the context of risk assessments for cybersecurity happens at a later stage than the initial risk assessment. After first outlining the consolidation of risk and riskification within cybersecurity in general, I then show how EINSTEIN embodies cybersecurity's modulation away from dichotomies (normality/abnormality) for risk assessments. Thus, EINSTEIN represents a shift in IDS methods towards a different ontology than that of risk assessments in counterterrorism, relying on distinct

sociotechnical configurations that do not presuppose a normative or calculated danger. Instead of targeting previously known threats, it focuses on identifying unknown risks linked to unknown system vulnerabilities, seeking to identify threats through autonomous and adaptive risk management techniques.

5.4 Risk, cybersecurity and EINSTEIN

Once treated by scholarship, industry, and practitioners alike as a mere technical challenge with technical solutions, the conceptual and sociotechnical evolution of cybersecurity has emerged as an all-encompassing notion, transforming into a vast industry and a transnational and global defence policy area, as well as its own scholarly field (Dunn Cavelty, 2018; Hansen and Nissenbaum, 2009; Liebetrau and Monsees, 2024). Across industries, governments, and multilateral organisations, cybersecurity has expanded defence budgets and become essential for securing the digitalised infrastructures that underpin contemporary societies (Dunn Cavelty, 2018). National security strategies worldwide treat cybersecurity as a matter of conventional defence, but also a means to economic resilience, seeking to balance between the open global market for trade and innovation as well as preservation of national control amid market competition.

Clearly, cybersecurity practices and technologies, but also discourses, have become central to (inter)national security and profoundly shape, and are shaped by, politics from the local to the global. Digital connectivity and information security are typically referred to by practitioners as the networks, technologies, and services that enable the collection, storage, and analysis of data provisional to maintaining the confidence, integrity, and availability of information. They are understood as essential to maintaining communication and digital governance such as surveillance, financial flows, and means of personal identification. However, while cybersecurity encompasses both the physical protection of digital infrastructure – ranging from cables to control systems, networks, and data servers – and computational security for maintaining data integrity through software and algorithms, it is not a purely a technical ‘matter’ (Kaufmann, 2024). Rather, it is inherently sociotechnical, emerging from the interplay of discourses, practices, and technologies (Musiani, 2022). While digital infrastructures are configured in a way that is tied to how materiality both enables and constrains what can be done to secure and maintain

it (Aradau, 2010), their configuration is also informed by what and how humans come to understand and deal with the limits and possibilities of a given technology.

Riskification has been integral to cybersecurity (Stevens, 2022). Signatures produced from particular risk assessments render how (and when) (in)security becomes known (Aradau and Blanke, 2017, 2022; Bousquet, 2018; Bueger, 2015; Edwards, 1996; Klimburg-Witjes, Poechhacker and Bowker, 2021). These practices have involved conceptualisations of the possibility of harm both from external threats and from the uncertainty of inherent system vulnerabilities (Backman and Stevens, 2024). There are notable similarities and differences between cybersecurity intrusion detection and the specialised and quantified terrorism-related assessment tools discussed earlier. These practices and technologies, along with their material foundations, shape and enact the seemingly intangible, providing a basis for speculative security to be both implemented and, in practical terms, theorised. For example, as Engin Isin and Evelyn Ruppert (2020, p. 2) argue, computational technologies have facilitated 'sensory power' as a new form of power – "a governing logic through which power produces effects".

Sensory power can be understood as how different technological devices make humans 'sense-able' through their day-to-day conduct (Isin and Ruppert, 2020; see also Amicelle, Aradau and Jeandesboz, 2015; Markussen, 2022). Equivalent to the risk assessment and profiling practices discussed above, data-driven surveillance enables the mapping of patterns of behaviour though monitoring of individuals' movements, activities, and interactions (see e.g., Kaufmann, 2024). Drawing on Foucault's analysis of power and its historical evolution from sovereign to disciplinary and regulatory forms, Isin and Ruppert (2020) argue that sensory power has developed alongside digital technologies since the 1980s, becoming particularly visible to the public during the Covid-19 pandemic. Situating sensory power within the history of governance, it emerges as an additional layer to other forms of power by manipulating real-time data to manage so-called clusters of people and their relations. The role of risk within cybersecurity also extends to how sensing infrastructures shape specific narratives about threats. Sensors actively produce conceptions of risk by determining what is made visible and measurable, often privileging certain phenomena over others (Klimburg-Witjes, Poechhacker and Bowker, 2021).

As digital infrastructures throughout the world become increasingly digitalised and interconnected – through networks, databases, and control systems for critical

infrastructure protection – IDS have become a standard way of safeguarding them. IDS have become vital components of securing critical infrastructures to prevent malicious activity from compromising the functioning of systems and databases. States typically rely on them for surveillance of traffic that could damage or obstruct the day-to-day operations of, for instance, electrical grids, water supply systems, or other societal functions. IDS are software that, through sensors, aims to detect and flag unwanted activity in networks or systems. Signatures for authenticating, for instance, users and devices are determined by code and set the parameters for 'normal' network traffic in a given organisation. Known or unknown indicators of compromise (IOCs), such as byte sequences, files, or IP addresses, alert IDS in the case of unauthorised attempts to access networks.

In 2003, the EINSTEIN programme was developed by the U.S. Cybersecurity and Infrastructure Security Agency (CISA). As part of the National Cybersecurity Protection System (NCPS), the task was to create and maintain a system to support federal civilian agencies combating potential attacks on internal systems and networks (Department of Homeland Security, 2004; CISA, n.d.). EINSTEIN's first version (E1) included monitoring sensors for networks that oversaw the flow of data traffic to support analysts in identifying and notifying potentially malicious patterns of behaviour entering or exiting federal U.S. networks. In 2008–09, E1 was replaced by EINSTEIN 2 (E2), which used signature-based intrusion detection in addition to monitoring sensors on government networks to augment threat detection capabilities. This meant that instead of merely monitoring networks, E2 could capture 'abnormal' behaviour entering or exiting the network. Signatures would automatically notify analysts of breaches that could indicate potential malicious behaviour, requiring less human labour. Government agencies were eventually required to share information in real time to increase the efficacy of the programme (Bellovin et al., 2011). According to CISA:

> On a typical day, E2 sensors generate approximately 30,000 alerts about potential cyberattacks. These alerts are each evaluated by CISA cybersecurity personnel to determine whether the alert represents a compromise and if further remediation is needed. (CISA, n.d.).

Thus, E2 was premised on developing signatures to determine 'normal' behaviour based on mass data, using this to pinpoint breaches within federal networks. At the time of the establishment of the Bush administration's Comprehensive National Cybersecurity

Initiative (CNCI) in 2008, most of EINSTEIN remained confidential (Bellovin et al., 2011, p. 4).

Around 2009, the Department of Homeland Security (DHS) piloted EINSTEIN 3 (E3) to test a modification that would include an intrusion prevention system (IPS). The idea was to enable EINSTEIN not only to observe and detect, but also to automatically stop attacks in real time (Department of Homeland Security, 2010). The main difference from E2 was that E3 was supposed to be based on a 'bigger and better' threat map to pre-empt harm with support from the National Security Agency (NSA) with "commercial technology and specialized government technology" (Department of Homeland Security, 2010, pp. 2–3). E3 was to,

> identify and characterize malicious network traffic to enhance cybersecurity analysis, situational awareness and security response… automatically detect and respond appropriately to cyber threats before harm is done, [and ensure the] ability to automate alerting of detected network intrusion attempts". (Department of Homeland Security, 2010, p. 3).

Further, DHS claimed they would warrant,

> substantial and long-term investments to increase national intelligence capabilities to discover critical information about foreign cyber threats and use this insight to inform EINSTEIN 3 systems in real time. DHS will be able to adapt threat signatures determined by NSA in the course of its foreign intelligence and DoD information assurance missions for use in the EINSTEIN 3 system in support of DHS's federal system security mission. (Department of Homeland Security, 2010, p. 3).

The references to "foreign intelligence" and "information assurance missions" suggest that the E3 pilot would allow for new and more extensive methods for data collection and analysis for determining signatures. According to some public officials, E3 was aimed at monitoring public networks and government data traffic through *private* corporations (Radack, 2009). A reading of the DHS privacy assessment for E3 reveals that instead of monitoring incoming traffic to government servers, the E3 programme would effectively monitor data traffic exchange points (Department of Homeland Security, 2010). That is, rather than monitoring perimeter traffic, meaning traffic entering and exiting networks coming from the outside of a given network, the exchange points would now monitor the inside of the network itself (Bellovin et al., 2011). The method of the E3 therefore assumes that there is a constant and *inherent* risk of threat to be measured and calculated to gain

'situational awareness' within the networks, rather than merely concentrating on obstructing danger.[25]

Controversies accompanied the proposal of E3. In a 2009 opinion piece, national security and human rights attorney Jesselyn Radack (2009) called the type of scheme used for E3 "unconstitutional". E3 would, according to Radack, monitor, collect, and analyse internal communication data: "Einstein 3 is based primarily on covert technologies developed by the NSA for the purposes of wiretapping". Allegedly, the predetermined original signatures of E2, designed to observe and notify known patterns of network traffic, would be supplied with sensors for content monitoring of Internet traffic, including email communications. Radack further stated that E3 would "include signatures based on personally identifiable information. The privacy implications are great. Any citizen logging on to a '.gov' website would trigger this" (Radack, 2009).

Some of the reactions to the E3 pilot's proposition to expand to private sector networks also critically referenced the NSA contract with the former U.S. weapons manufacturer company Raytheon (Bellovin et al., 2011). The plan was originally "to study the value of sensors in recognizing impending cyberattacks in critical infrastructure cyber networks", with the goal of extending "EINSTEIN-like protections to the private sector" (ibid, pp. 4–5). Apprehensions also originated from within the DHS, allegedly because of the history of the NSA and the Bush administration's controversial use of wiretapping (Nakashima, 2009). Likely due to both legal and privacy concerns, as well as the negative reactions to the E3 pilot, the version never became operational.

Around 2013, however, a different version was announced: EINSTEIN 3 Accelerated, or E3A (Department of Homeland Security, 2013). According to unclassified DHS documents, the scope of E3A was narrower than the E3 pilot, E3A being limited to two intrusion prevention services, namely: i) DNS (domain name server) 'sinkholing' (containing threats by blocking malicious traffic on a server), and ii) email filtering (Department of Homeland Security, 2013). Unlike E1 and E2, E3A was based on Internet Service Providers (ISPs), amplifying anomaly detection:

---

[25] For a recent discussion on the implications of 'situational awareness' for practices of security in managing uncertainty and anticipation, see Hentschel, Krasmann and Zebrowski (2025).

> Leveraging existing investments in the ISP infrastructure, our non-signature based [E3A] pilot efforts to move beyond current reliance on signatures are yielding positive results in the discovery of previously unidentified malicious activity. DHS is demonstrating the ability to capture data that can be rapidly analyzed for anomalous activity using technologies from commercial, government, and open sources. The pilot efforts are also defining the future operational needs for tactics, techniques, and procedures as well as the skill sets and personnel required to operationalize the non-signature based approach to cybersecurity. (Department of Homeland Security, 2017).

To enable EINSTEIN to automatically detect and alert potential or previously unknown cyberattacks or malicious behaviour, E3A incorporated anomaly detection. Anomaly detection seeks to identify previously unknown and 'risky' activity. Anomalies are, as mentioned, not unknown threats but rather *patterns of irregularity*. That is, anomalies are not abnormal, as per a binary friend/enemy distinction. Rather, they represent discrepancies that do not conform to previously known or predefined categories of, for instance, risk. Unlike traditional statistics, which eliminate outliers, modern security techniques – driven by machine learning and Big Data – actively seek anomalies to identify unknown threats (Aradau and Blanke, 2018). Should anomalies occur among incoming network traffic on government servers, US-CERT (the United States Computer Emergency Readiness Team) would be notified.

Overview of the National Cybersecurity Protection System (NCPS)			
Operational Name	**Deployment Year**	**NCPS Objective**	**Description**
EINSTEIN 1	2003	Intrusion detection	Provides an automated process for collecting, correlating, and analysing agencies' computer network traffic information from sensors installed at their Internet connections.
EINSTEIN 2	2009	Intrusion detection	Monitors federal agencies' Internet connections for specific predefined signatures of known malicious activity and alerts DHS's U.S. Computer Emergency Readiness Team (US-CERT) when specific network activity matching the predetermined signatures is detected.
EINSTEIN 3 Accelerated	2013	Intrusion detection Intrusion prevention	Automatically blocks malicious traffic from entering or leaving federal civilian agency networks. This capability is managed by Internet Service Providers, who administer intrusion prevention and threat-based decision making using DHS-developed indicators of malicious cyber activity to develop signatures.

Table 5.2. Iterations of the National Cybersecurity Protection System. Based on U.S. Government Accountability Office (2018, p. 15).

In 2018, a U.S. Government Accountability Office (GAO) report found that EINSTEIN's monitoring capabilities were flawed, and that EINSTEIN "was limited in its capabilities to identify potential threats using anomaly-based detection" (U.S. Government Accountability Office, 2018, p. 15). Notably, the U.S. federal government's discovery of the SolarWinds data breach in 2020 was a wakeup call for government agencies to improve present-day capabilities to detect and ward off cyberattacks internally. Apparently, however, EINSTEIN was not able to detect the attack because of CISA's lack of resources to protect federal networks (Timberg and Nakashima, 2020). Soon after, in 2021, the Colonial Pipeline attack that caused a short-lived but intense fuel crisis in several U.S. states was yet another reminder of the weaknesses of internal network security. According

to Thomas Bossert, a former Homeland Security Advisor to U.S. President Donald Trump and longtime top cybersecurity official, "Einstein wasn't designed properly" (Timberg and Nakashima, 2020).

Thus, E2 formally ceased operations in September 2024, and according to the DHS, E3A's email filtering services were to be decommissioned that same year. CISAs budget overview for the 2024 fiscal year did not mention E3A (Department of Homeland Security, 2023), and although CISA will temporarily continue to operate EINSTEIN, this is soon to be replaced with the Cyber Analytics and Data System (CADS):

> CADS will provide a robust and scalable analytic environment capable of integrating mission visibility data sets and providing visualization tools and advanced analytic capabilities to CISA's cyber operators, allowing more rapid analyses to inform more rapid actions, and, in turn, reducing the prevalence of intrusions and exploitable conditions across Federal and critical infrastructure networks. (U.S. House of Representatives Committee on Homeland Security, 2023).

In other words, the DHS states that CADS will amplify CISA's abilities to reduce cybersecurity *vulnerabilities* ('exploitable conditions') by creating a more resilient environment through better 'situational awareness'. CADS will be a so-called "system of systems" that will integrate data from both private and public partners and centralise all data for CISA's cyber analyses, meaning that there will no longer be a need for manually comparing information across databases (Doubleday, 2023).

U.S. security officials characterised efforts to build EINSTEIN over more than two decades as something of a failure (Timberg and Nakashima, 2020). However, it serves as a telling example of the logics of riskification that increasingly characterise the systems for cybersecurity, which are both similar and different to systems for risk assessment in counterterrorism. The evolution from E1 to E3A (and then to CADS) highlights the situated logics of cybersecurity practices and technologies and how this sociotechnical configuration shapes the governance of risk and the dynamics of riskification. Notably, as E3A and now CADS are configured with the ambition to employ anomaly detection, this entails the foci shifting away from originally aiming to fight 'enemy' intruders – by pre-empting intruders from materialising (as with 'profiling') – to (re)configuring the signatures based on irregularities. Although the goal of detecting threat remains, the initial method of pre-emption has been replaced by the goal of reducing vulnerability to limit (the risk of) damage. Thus, the story of EINSTEIN is illustrative of a shifting sociotechnical

configuration that departs from preconfigured distinctions of good/bad behaviour and is instead oriented towards sensing technologies for continuous adaptability to create a more 'secure' environment.

This difference is key to understanding the consequences of actions that accompany sociotechnical systems and the practices that inform conceptions of risk for mitigating threat. Risk assessments are normative by nature. That is, risk assessments presuppose a normal in order to define the abnormal for profiling risky individuals. It describes a quest of meaning-making: Knowing the objective threat and the causal presuppositions of it. For EINSTEIN, the risk assessment for creating 'risky profiles', i.e., signature-based intrusion detection, has been the status quo for IDS and counterterrorism programmes alike. But as computationally automated threat detection practices and technologies have embraced machine learning techniques seemingly to increase efficiency, the orientation towards pre-emption has gradually shifted towards possibility and precaution instead. Based on vast amounts of real and synthetically generated data that enable a method of pinpointing irregularities as IOCs, the preconfigured process of defining the 'normal' has become a combined effort, where the 'risk of threat' is never constant but always co-produced based on what is unknown. As discussed in Section 5.5, it is, rather than solely a meaning-making process, better understood as a sense-making process: Unremittingly co-producing threat and (in)security on and through mutable ways of sensing it.

This progressive development in ways of detecting and knowing threat – from sensors that alert deviations from the 'normal' through signatures to detecting anomalies for 'sensing' what constitutes a threat – illustrates an emerging foundational ambiguity in sensors/sensing powered by automation through machine learning. Sensors are for measuring and quantifying information, making it both intelligible and actionable. Sensing, by contrast, is not about the function of sensors but encompasses the ways in which data (from sensors) is interpreted, acted upon, and embedded in security thinking and practices (Amoore, 2009). Effectively a sociotechnical practice, sensing is highly subjective, selective, and dynamic. While the latter (sensors) resonate with the modes of risk assessments and tools in the UK's Prevent programme and EINSTEIN in its earlier versions (1 and 2), the former (sensing) represents the shift towards empirically sensing threat, albeit not having a seemingly objective or calculated idea about the threat as precursor to risk

assessment. It brings attention to when and how meaning (i.e., the threat) is produced in the timeline of assessing risk, and how this changes by virtue of computational automation.

5.5 Discussion: Riskification, meaning-making and sense-making

At the technical level of risk assessments, both counterterrorism and cybersecurity operate within logics of risk and riskification, emphasising anticipation as a key strategy for threat mitigation. As has been shown, riskification has transformed how threats are identified and dealt with within both systems. Profiling risk as a precursor to threat has become the default approach (see also Dunn Cavelty and Wenger, 2022; Hansen and Nissenbaum, 2009). This 'de-perimeterisation' of security dissolves traditional delineations in favour of proactive interventions based on assumed inherent risks (Aradau and van Munster, 2007). Within both counterterrorism and cybersecurity, anticipation of danger still relies on some pre-emptive categorisation through sociotechnical practices. Both fields share the objective of 'knowing' threats in advance and transform abstract dangers into actionable information, driving a chronological shift towards addressing risk at ever-earlier stages, seeking to pre-empt conditions of possibility for danger and damage. In counterterrorism systems of practice, such as Prevent's risk and vulnerability assessment, the emphasis is effectively on 'profiling' individuals deemed 'at risk', through a process rooted in normative assumptions. 'Normality' is here predefined, enabling the identification and isolation of individual deviations as risks, through sets of indicators. Signature-based cybersecurity programs like EINSTEIN's early versions are based on the same logic, identifying threat based on known IOCs. As with the Prevent programme, the 'normal' here is to be found through empirical observation of 'known' indicators and behaviour.

Through these dynamics, both counterterrorism risk assessments and signature-based intrusion detection rely on 'meaning-making' with predefined and normative threat categorisations. Meaning-making is primarily addressed in constructivist and poststructuralist discourse-based work (e.g., securitisation theory), focusing on how actors represent security (Buzan, Wæver and de Wilde, 1998; Hansen, 2006). In this context, meaning-making operates through a discursive and categorical framework for risk assessment, where risk is understood in terms of predefined identities, behaviours, and sociopolitical markers. Instead of dynamically sensing deviations, meaning-making relies

on established semiotic and interpretative frameworks that categorise individuals based on assumptions about radicalisation, susceptibility, and behavioural predispositions. The risk and vulnerability assessment tools used in counterterrorism also extend their targets beyond physical security concerns into psychological and ideological domains, often reifying threat categories rather than discovering them. Indeed, STS scholarly critiques have already been attentive to how counterterrorism infrastructures can function as semiotic systems, embedding security threats within narratives of deviance, extremism, and radicalisation rather than within emergent, data-driven anomalies (Aradau and Blanke, 2018).

By contrast, cybersecurity's anomaly-based detection – such as EINSTEIN's later iterations – represents a different, 'sense-making' paradigm. Sense-making – or 'sensing' as it has been explored across STS, as well as Foucauldian and practice-oriented approaches – turns the focus towards how security is enacted, embodied, and materialised in everyday routines and technologies (Amoore, 2009; Aradau and van Munster, 2007; de Goede, Simon and Hoijtink, 2014; Isin and Ruppert, 2020). The sense-making of anomaly detection entails continuous adaptation by identifying irregularities, shifting from fixed risk assessment practices that rely on prior knowledge of threats to open-ended, dynamic processes, where normalcy evolves in response to new or 'more' data. Unlike meaning-making, the dynamics and practice of sense-making produces 'meaning' (i.e., conceptions of threat) after detection, rather than before. Sense-making in cybersecurity, particularly practices of anomaly detection, is computational and autonomous, relying on sensor-based infrastructures, algorithmic processing, and machine learning to identify deviations from statistical norms. As seen in the move from EINSTEIN 2 to E3A – instead of looking for predefined categories of threat, sense-making operates through continuous search for outliers (Aradau and Blanke, 2018) to anticipate 'the situation' (Krasmann and Hentschel, 2019). Security practitioners here do not necessarily know in advance what constitutes a threat but assume it will emerge from irregularities or deviations from established 'normal' patterns. Thus, anomaly detection's sense-making epistemology differs from that of pre-emptive security (de Goede, 2012), since sensemaking's goal is not only to detect existing risks but to generate knowledge about potential threats through real-time monitoring, data fusion, and correlation across heterogeneous datasets. This evokes parallels to current intelligence and targeting practices – less directed towards fixed enemies than constantly producing and updating enemy models through iterative surveillance and digital sensing (Reichborn-Kjennerud, 2025).

This shows that meaning-making and sense-making represent fundamentally different risk conceptualisations: Meaning-making approaches, as seen in counterterrorism and signature-based instruction detection for cybersecurity, produce risk based on known threats, while sense-making approaches, such as anomaly detection, derive risk from 'unknown' data. Meaning-making is categorical and interpretive, whereas sense-making is probabilistic and emergent. As cybersecurity increasingly embraces anomaly detection, the focus moves further away from pre-emptive elimination of threats and towards resilience-building aimed at minimising damage. This transformation highlights the further shift in security practices from a probability-based logic, where threats are calculable, to a possibility-based logic, where risk remains constant, fluid, and adaptive (Dillon, 2008). In this manner, cybersecurity becomes a practice of continuous vigilance through self-adjusting, automated surveillance systems with pervasive and invasive monitoring. These shifts underscore the dynamic nature of sensory power (Isin and Ruppert, 2020), where clusters are generated through real-time data analytics and visualisations that continuously recalibrate irregularity as risk and danger.

At the heart of the different logics of meaning-making and sense-making are their situated and contextual sociotechnical configurations. Specifically, counterterrorism relies on interpretive models of (cognitive) radicalisation, working to operationalise security through predefined categories. This kind of meaning-making presupposes that certain pre-existing social, political, and ideological factors correlate with risk, which leads to profiling, classification, and intervention based on perceived susceptibility rather than empirical emergence. Meanwhile cybersecurity uses anomaly detection *prior* to meaning-making (knowing the threat), reflecting an unsolidified epistemology, where threat is co-produced through sociotechnical configurations of algorithms and data. This form of sense-making assumes that threats exist within networked systems but are not immediately visible; placing emphasis on adaptive sensing processes where the 'normal' is continuously (re)discovered. These different logics enact security differently, shape and are shaped by the sociotechnical frameworks through which risk is understood and managed (Backman and Stevens, 2024; Spencer and Pizio, 2024), and demonstrate how riskification dynamics may materialise differently in practice. I do not argue that cybersecurity exclusively comprises practices of sense-making (particularly when considering meaning-making processes such as cyber threat intelligence (CTI) and other 'strategically oriented' risk practices), nor that counterterrorism is exclusively about meaning-making. Riskification

logics differ less by field than by operational level or contextual application. Therefore, I have not compared the 'fields' as such but juxtaposed them in terms of how particular systems generate actionable information in different ways, continuously co-producing countermeasures that rely on ostensibly objective knowledge – imaginaries of omniscience (Suchman, 2023).

In the end, sense-making and meaning-making highlight diverging conceptualisations of risk, with implications for the production of threat and (in)security. Thus, while risk and (in)security are perpetually co-produced with the tools used for assessment, it shows that the production of (in)security has little to do with real or imagined risk but rather emerges in particular configurations. Sensory power emerges through interaction between devices, algorithms, and data that produces clusters as objects of governance. In certain cybersecurity practices, sensory power operates through systems like EINSTEIN, with sensory assemblages continuously tracing, tracking, and modulating patterns of irregularity powered by machine learning. Contrary to imposing predefined classifications – as with, for instance, mapping of threat ontologies for CTI – anomaly detection for IDS generates knowledge through recursive feedback loops, enabling adaptive responses to evolving risks. Counterterrorism's reliance on normative typologies for risk assessment systems has so far contrasted with the mutability of sensory power in practices of cybersecurity anomaly detection. It does, however, remain an open question what the implications for counterterrorism may be if computational and increasingly automated risk assessment techniques come to rely on machine learning and AI. As shown in this chapter, anomaly detection only recalibrates the threat based on what fails to fit the predetermined signatures. Risk assessment practices, therefore, are not grounded in some objective truth, nor 'reality', except that of what is considered actionable information that continuously produces countermeasures.

5.6 Conclusion

The comparative outlines of this chapter have made visible some of the range and reach of riskification within contemporary security practices through the different situated logics of meaning-making and sense-making. The chapter has highlighted how the sociotechnical configurations underpinning speculative security practices in cybersecurity and counterterrorism are foundational to their enactment. Whether through the predefined

categories of Prevent's specialised risk assessment or the dynamic sensing systems of EINSTEIN 3 and 3A, the processes, technologies, and sociotechnical practices that transform intangible risks into actionable insights play a central role in shaping security practices and ways of knowing threat through risk assessment.

By offering an STS vocabulary to explore how sociotechnical (re)configurations condition possibility for knowing the 'risk of threat', this chapter has shown how speculative security systems function and how they might be theorised in practical terms. In particular, meaning-making and sense-making represent two distinct risk logics and thus modes of security knowledge production: Meaning-making produces threat knowledge through discourse, classification, and ideological profiling, embedding security risks within pre-existing sociopolitical frameworks, whereas sense-making produces threat knowledge through iterative, machine-assisted anomaly detection, rendering security an ongoing, computationally mediated process. While both are used to justify pre-emptive security measures, their epistemological, technological, and political implications differ significantly. Threat, in the end, is not 'fixed', but rather perpetually co-produced with the tools used to see, interpret, and ultimately act upon it. Hence, this chapter has also shown that sensitivity is warranted across studies on risk/security not only to the adoption of riskification practices and their consequences, but also to how riskification logics adapt, evolve, and expand over time and beyond intention.