DOI:
10.1057/s42984-020-00007w

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Stevens, T. (2020). Knowledge in the Grey Zone: AI and Cybersecurity. *Journal of Digital War*, *1*(1), 164-170. https://doi.org/10.1057/s42984-020-00007w

Digital War
https://doi.org/10.1057/s42984-020-00007-w

ORIGINAL ARTICLE

Knowledge in the grey zone: AI and cybersecurity

Tim Stevens[1]

Abstract
Cybersecurity protects citizens and society from harm perpetrated through computer networks. Its task is made ever more complex by the diversity of actors—criminals, spies, militaries, hacktivists, firms—operating in global information networks, so that cybersecurity is intimately entangled with the so-called grey zone of conflict between war and peace in the early twenty-first century. To counter cyber threats from this environment, cybersecurity is turning to artificial intelligence and machine learning (AI) to mitigate anomalous behaviours in cyberspace. This article argues that AI algorithms create new modes and sites of cybersecurity knowledge production through hybrid assemblages of humans and nonhumans. It concludes by looking beyond 'everyday' cybersecurity to military and intelligence use of AI and asks what is at stake in the automation of cybersecurity.

Keywords Algorithms · Artificial intelligence · Automation · Cybersecurity · Grey zone · Knowledge

Cybersecurity in the grey zone

Discussions of 'digital war' quickly resolve to an understanding that global digital connectivity has disturbed our orthodox understandings of war and peace. In the so-called grey zone (Wirtz 2017) below the threshold of armed conflict, a diverse and fluid cast of actors attempts to derive political, strategic and economic gain through the ever-proliferating assemblage of information communications technologies. Even if it is a step too far to label many of these activities—political activism, crime, espionage, subversion—as 'war' or 'war-like' (Levinson, this volume), it is not unreasonable to characterise the wider environment of cybersecurity as one of persistent grey-zone conflict (Corn 2019, 347–354). At the strategic level, for instance, suitably equipped states view 'cyber' as a coercive tool worth exploiting for national gain, be it in military operations, intelligence-gathering or commercial cyberespionage (Brantly 2018; Valeriano et al. 2018). As most of this occurs below the level of conventional conflict, it is seen as a way of limiting conflict escalation, albeit always with the potential to generate just the opposite (Brands 2016; Schneider 2019).

The twists and turns of operations in this digital domain may be surprising and their effects—inasmuch as causality can ever be demonstrated—remarkable, but the vectors of their propagation are familiar, omnipresent and quotidian, including the global internet and the myriad consumer electronics that share and shape our lives. So too the practices of cybersecurity, which, for all the justifiable interest in 'cyberwarfare' and 'cyberespionage', constitute the 'everyday' of digital conflict in the grey zone. In the daily exchanges of technical cybersecurity, there is a constant to-and-fro between network defenders and attackers. Cybersecurity practitioners tend to wring the most out of existing technologies to wrest the upper hand from their adversaries, whether the humblest hacker or the best-resourced state entity. Inevitably, however, they also seek to employ innovative technologies as the environment evolves. In tune with the zeitgeist, artificial intelligence and machine learning (hereafter, AI) are proposed as an essential 'solution' to all manner of cybersecurity threats.

The following discussion addresses the identification of cyber threats through AI and algorithmic means and demonstrates a shift from signature- to anomaly-based approaches to threat detection and mitigation. It then moves to the specific application of AI to anomaly detection, particularly in relation to the construction of 'normal' organisational 'patterns of life'. The epistemic implications of this technical orientation are then unpacked, specifically how AI creates

✉ Tim Stevens
tim.stevens@kcl.ac.uk

[1] King's College London, London, UK

Published online: 31 March 2020

new sites of knowledge production through new human–non-human subjectivities. This brief exploration concludes by looking beyond everyday cybersecurity to military and intelligence 'cyber' and asks what is at stake in the automation of cybersecurity.

From signatures to anomalies

Cybersecurity cannot proceed without data. Access to and protection of data are core practical considerations for cybersecurity professionals. As more computing devices become networked in the 'Internet of Things' and data volumes—at rest and in transit—continue to grow, the 'attack surface' of the global information environment expands continuously. Understood as the totality of weak points exploitable by nefarious actors, the attack surface is less a physical boundary to be defended than a logical membrane of potential vulnerability distributed in space and time. Preventing exploitation of flaws in this integument, through which malicious code can be inserted or valuable data extracted, is a near-impossible task, not least as vulnerabilities are inherent in information systems. As Libicki (2009, 14) notes, there is 'no forced entry in cyberspace. Whoever gets in enters through pathways produced by the system itself'. This ontological insecurity is compounded by the sheer diversity and growth of this environment, so that defence can never be perfect and attackers seem always to be one step ahead. Cybersecurity is about making visible those vulnerabilities and threats in order to counter them, an informational challenge that relies on timely and intelligible data for interpretation and effective use in prevention and remediation.

The collection and filtering of data about the status of information systems and threats to their intended functioning have long been largely automated. Humans do not have the cognitive or sensory capacity to cope with the enormous data volumes produced by software and hardware dedicated to alerting systems administrators to problems inside and outside their networks. Add to this a human capital shortfall in the cybersecurity industry (Shires 2018) and automation is a reasonable technical fix for some of these problems. Cybersecurity vendors, for example, offer thousands of different products for automated malware detection. These software packages inspect network traffic and match data packets to known signatures of malicious software (malware) that install themselves on users' machines and act in deleterious ways. Now much more than just 'anti-virus' protection, these software packages also detect and deny worms, Trojans, spyware, adware, ransomware, rootkits and other software entities searching for and exploiting digital vulnerabilities. Automated in this fashion, malware can be repelled, quarantined, destroyed or otherwise prevented from infecting a system and opening it up for exploitation. Effective as this has been in the past, new types of polymorphic and metamorphic malware, which change their 'appearance' every few seconds, can evade traditional techniques of 'signature-based' detection that rely on existing knowledge of malware species. You cannot defend against these rapidly mutating threats if they have no signature matches in databases that are inevitably always out of date.

Signature-based detection is still effective against known malware, but technical cybersecurity is developing 'anomaly-based' capabilities that compare network traffic to continuously revised baselines of 'normal' behaviour to detect intrusions, analyse malware and detect spam and phishing attacks. Traffic classified as legitimate against a given baseline is permitted across network boundaries, but unusual behaviours will be detected automatically and flagged for further investigation, including by human analysts. Yet, as the number of alerts triggered increases, many requiring human attention, we are thrown back again on the limits of organisational and cognitive capacity. This situation is exacerbated by the persistent creation of false negatives (bad behaviour treated as good) and false positives (good behaviour labelled as bad), which further drain resources (Libicki et al. 2015, 28–29). One response has been to add a new layer of automation to the parsing and analysis of these 'big data' sets, using emerging technologies of AI to process and interpret anomalous behaviour and feed it back into cybersecurity decision-making processes in ever-shortening cycles of action and reaction. Cybersecurity and AI have not always been obvious bedfellows. In the early days of each field, computer security researchers viewed with some distrust the relative unpredictability of AI system behaviours, their multiple possible outputs anathema to the order sought by software engineers (Landwehr 2007). Now, the possibility that AI can learn and adapt in dynamic ways makes it more attractive for countering similarly adaptive cybersecurity threats.

The promise of AI

The use of AI in cybersecurity is not entirely new. If we understand 'AI' as a placeholder for a set of algorithmic machine learning heuristics rather than a new form of 'intelligence', cybersecurity has been alive to its potential since at least the 1990s. Early applications include the filtering of spam email using Bayesian logic and the classification of spam by large neural networks (Brunton 2013). AI has increased significantly in sophistication since then and its development has partly been occasioned by adversaries' use of those same technologies. Learning algorithms fuel the evolution of the aforementioned polymorphic and metamorphic malware, for instance. They have also been created to detect vulnerabilities in target systems and assist in their

exploitation and subversion: the cybersecurity industry is deeply worried by the future deployment of 'adversarial' or 'offensive' AI, a concern shared with governments and military and intelligence agencies (Brundage et al. 2018). In this respect, AI deployments are part of an emerging 'arms race' between attackers and defenders (Scharre 2018, 229–230). This message is not lost on enterprises, 61% of which reported in 2019 that they could not detect network breaches without AI technologies and 69% believed AI would be necessary to respond to those breaches (Capgemini 2019).

One notable response to this market demand has been the development of cybersecurity AI platforms that claim to mimic natural 'immune systems', borrowing from a long analogical association between cybersecurity and notions of public health and epidemiology (Betz and Stevens 2013; Weber 2017). Companies promote anomaly-based AI as a way of learning what 'normal' means for an organisation and then begin to explore unusual behaviours. The tolerance of the AI engine to anomalous behaviours can then be adjusted, based on the risk appetite of the organisation and its resourcing levels (Wilkinson 2019). The 'immune system' metaphor plays out in a platform's putative capacity to detect, root out and inoculate against further suspicious behaviours. In one firm's account of its 'Enterprise Immune System' package, '[b]y detecting subtle deviations from the organisation's "pattern of life", it can distinguish friend from foe—and highlight true cyber threats or attacks that would otherwise go unnoticed' (Darktrace n.d.). These vendors claim another benefit for potential clients, in that they free up analysts' time to focus on more important tasks, instead of being mired in 'low-priority or benign events from the start' (Darktrace n.d.). This is a recognition of the frequency of false negatives and positives—although AI promises to reduce their incidence by learning what is 'normal' over time—but is also a legitimate comment on the exigencies of life in modern security operations centres. These are bombarded with security reports, whose importance at first glance can be difficult to discern. Automated AI systems can handle the bulk of low-level threat detection and response, allowing analysts to tackle high-level behaviours and incidents instead. This includes developing profiles of the human actors behind sophisticated cyber campaigns, such as so-called advanced persistent threats (APTs), in order to tailor deterrence and response regimes. A further promising avenue of research is to capture analysts' real-life behaviours and feed this data back into learning algorithms as a way of shaping AI responses to cybersecurity incidents (Bresniker et al. 2019).

AI systems learn about the threat environment within and without the notional perimeters of organisations. This includes employees and other people interacting with those networks: clients, customers and, in the case of government agencies, citizens and service-users. Monitoring these individuals allows AI to establish an organisation's internal rhythms and patterns, a normal baseline against which anomalies can be registered, such as 'insider threats' posed by disgruntled employees, leakers and whistle-blowers (Wall 2013). AI-powered behavioural analytics might detect, for instance, unusual employee activities at odd times of the night or unexpected spikes in transfers of sensitive commercial data or intellectual property. Combined with multiple other data streams, a picture is developed of anomalous behaviour within an organisation's immediate purview organisation, leading to further investigation and possible censure. Using AI to predict human rather than just machine behaviour is an important contributor to 'sociotechnical' framings of cybersecurity that address how technology, users and processes interact in complex organisational systems (Coles-Kemp and Hansen 2017). An uneasy tension exists in this field between designing computer networks that enhance the human security of end-users (Coles-Kemp et al. 2018) and using technology to discern anomalous behaviours, a perspective that tends to perceive all end-users as potential vectors of threat (Schneier 2016).

AI, anomalies and episteme

The search for anomalies shifts the emphasis of cybersecurity from known threats to the prolepsis of as-yet-unknown threats and into an anticipatory posture that has received much attention in the critical security literature (e.g. Amoore 2013). It also suggests a more inductive approach to the creation of useful cybersecurity knowledge from big data sets. Driven by AI, anomaly-based cybersecurity practices search for patterns, associations and correlations that cannot be identified solely by signature analysis, the latter understood as a form of deductive reasoning. In part, this channels the infamous claim made of big data analytics that they make the hypothetico-deductive method 'obsolete' (Anderson 2000). As critics of this position observe, all enquiry, big data analytics included, instantiates any number of theoretical assumptions and biases (Kitchin 2014). In AI-driven machine learning, parameters and data categories are often set in advance by human programmers. This 'supervised learning' is very different from 'unsupervised' learning techniques, in which algorithms do the labelling and categorisation themselves and might therefore be 'unbounded from the bias of specific localities' (Parisi 2019, 98). The quality of training data is therefore key to determining output quality, and supervised machine learning can give rise to outcomes that reproduce programmers' biases (McDonald 2019). In cybersecurity, the assembling of 'cyber threat intelligence' (CTI) from multiple data streams is prone to various biases, as analysts are swayed by groupthink, external social media

and reporting, seduction by the power of expertise, and their personal beliefs and prejudices; responsible CTI cannot therefore be the preserve of machines alone (Collier 2019). These everyday decisional acts constitute 'little security nothings' that have even greater import in the aggregrate than in their individual existence alone (Husymans 2011)

Furthermore, as Aradau and Blanke (2018) identify in their critique of 'algorithmic rationalities', the hunt for anomalies in big data security environments subtly alters the logics of security in three specific ways. Referring to intelligence and counter-terrorism, rather than cybersecurity, they note first how advanced analytics are, as outlined above, a response to the increased data volumes of interest to security practitioners. Anomaly-based analytics are geared, therefore, to the practical requirement for 'actionable information', rather than 'good or truthful information', which unravels claims to 'algorithmic objectivity' and shows 'how uncertainty is radically embedded within algorithmic reasoning' (Aradau and Blanke 2018, 20). This is compounded by the lack of standardised data collection and classification procedures, which complicates comparisons between data sets. Second, anomalies are constructed by their divergence from normality, which is defined as similarity between data points rather than predetermined categories. 'Otherness' is expressed as 'dots, spikes or nodes' in geometric or topological space that are anomalous in their dissimilarity to the 'normal' so constructed. Whilst this may seem to avoid the questions of bias previously noted, this unmooring from negative categorisations presents a political problem: how to 'reconnect techniques of producing dots, spikes and nodes with vocabularies of inequality and discrimination'? (Aradau and Blanke 2018, 20). Third, stable categories—as in statistical analyses—give way to 'continuous calculations of similarity and dissimilarity', thereby refusing direct political contestation, particularly as those calculations 'remain invisible, often even to the data analysts themselves' (Aradau and Blanke 2018, 21). Hidden within the archetypal 'black box' of technology, defined in terms of inputs and outputs only (Von Hilgers 2011), the computation of AI technologies is frequently unknowable and therefore radically opaque (Castelvecchi 2016).

The complexity, uncertainty and lack of transparency of anomaly-based AI technologies in cybersecurity and elsewhere raise questions of AI trust, safety and control. The postulated solution is either to double down on AI as an engineering solution to these problems, or to figure out precisely where humans should be located in respect of the 'loop' of AI decision-making. In the context of lethal autonomous weapons systems and the use of AI in war, the question has become: should humans be in, on or out of this loop? (Jensen et al. 2019; Scharre 2018) In cybersecurity, as in war, analysts and operators are enmeshed in hybrid human–machine assemblages that generate knowledge of particular utility in specific decision-making circumstances. The production of cyber threat intelligence is referred to in terms of hybrid configurations of people, data, machines and organisations, tasked with meeting specific operational and strategic demands (Jasper 2017). Srnicek (2014, 45) captures this dynamic when he analyses the 'delegation of thought' to machines in 'cognitive assemblages' of sociotechnical composition. In these epistemic loci, heterogeneous human and nonhuman actors, including algorithms, contribute to the creation of expert knowledges, which are 'collective and distributed rather than individual or solely social' (Srnicek 2014, 45; Hayles 2016). They establish 'collaborations of influence' (Kaufmann 2019) that have agency in how and what kind of knowledge is produced. Are we 'in the loop' if our choices are between paths of action relayed by machines? (Weber and Suchman 2016) If unsupervised learning systems develop their own questions and goals, what becomes of the human at all? The traditional mantra that cybersecurity is a 'team sport' is better read as a 'dance of agency' (Pickering 2010) once this relational process of delegation and complex agency in AI-rich environments takes hold.

Beyond the everyday

The foregoing review has addressed the everyday cybersecurity necessary to maintain the ordinary functioning of the information systems on which daily life depends. AI has an important role in this diverse landscape, automating the mundane, learning from its environment and promoting better protection for users, organisations and societies. However, as alluded to above, the role of algorithms in creating and sustaining specific socioeconomic and political arrangements is never neutral (Kitchin 2017; Berry 2019). AI-driven anomaly detection creates new subjectivities in the grey zone of modern informational conflict and risks sacrificing 'truth' for 'utility' and 'convenience' in the managerial practices of organisational cybersecurity. This radical contingency creates political problems as AI substitutes for human cognition as the arbiter of network decision-making and regulation of data flows. Anomalies may escape easy mapping to established groupings of people and behaviours, but there are multiple examples of the insidious effects of this computational logic. Tools of data manipulation similar to those in AI-fuelled cybersecurity are easily turned to economic and political projects of worrisome scale and impact. From the 'behavioural futures markets' of surveillance capitalism (Zuboff 2019) to the Chinese social credit system (Matsakis 2019), from the excesses of domestic surveillance (Bauman et al. 2014) to online information operations in a post-truth era (Woolley and Howard 2018), algorithms constitute new modalities and mediators of political influence

and control. Whilst we have few grounds for assuming that AI is inferior to human intelligence (Collins 2019), questions of knowledge, power and subjectivity are intimately bound up with these algorithmic rationalities (Aradau and Blanke 2018).

AI and cybersecurity are also far from neglected in the realms of secret intelligence and military operations. AI is a *sine qua non* for modern militaries (Lewis 2019) and a potent augmentation to strategic planning (Payne 2018). Without AI in cyberspace, the former head of US Cyber Command remarked in Senate testimony, 'you are always behind the power curve' (Congressional Research Service 2019, 11). Militaries and intelligence agencies are investing heavily in AI-enabled cyber, including the US Department of Defense's new Joint Artificial Intelligence Centre (Corrigan 2019) and numerous other initiatives leveraging AI to generate operational and strategic effect in and through cyberspace. The UK is developing 'human–machine teaming' in AI-rich environments including cyber, so as to promote 'the effective integration of humans, artificial intelligence (AI) and robotics into warfighting systems … that exploit the capabilities of people and technologies to outperform our opponents' (Ministry of Defence 2018, 39). Known in the USA as 'centaur warfighters', these hybrids will 'leverage the precision and reliability of automation without sacrificing the robustness and flexibility of human intelligence' (Scharre 2018, 321). This continues a long lineage of hybrid military subjectivities (Coker 2013), oriented now to predictive machine cognition providing decision advantage to the new military operations of 'cognitive manoeuvre' in global information space (Dear 2019). These hybrid entities are not just network defenders but agents of cyberwarfare and cyberespionage. As the US adopts new doctrines of 'persistent engagement' and 'defending forward' in cyberspace (Healey 2019), they will be on the front lines of digital warfare and intelligence, with as-yet unknown autonomy in decision-making and target discrimination. In this assemblage, 'we may never know if a decision *is* a decision … or if it has been "controlled by previous knowledge" and "programmed"' (Amoore and de Goede 2008, 180).

Militaries and intelligence agencies join firms, citizens, criminals, hacktivists and others using AI-enabled cyber tools in the complex battlespaces of the grey zone. Technical cybersecurity exploits AI to overcome constraints on human cognition, to speed decision-making and to provide better protection to organisations and end-users of information systems. In so doing, it prioritises operational efficiency over truth and creates new subjectivities and targets of intervention. Algorithms and automation also create new sites of knowledge production whilst simultaneously obfuscating the precise calculations that lead to particular outputs. This draws attention to problematic aspects of the delegation of thought to machines, in that it portends a decrease in opportunities for meaningful oversight and regulation. The stakes could not be higher in military cyber targeting or digital intelligence—strategic cybersecurity—by which standards the use of AI in everyday cybersecurity is a straightforward proposition. Nevertheless, the role of AI in cybersecurity generally, a core modality of offence–defence dynamics in the grey zone, remains open to contestation, modulating as it does the global flows of data in our nominal 'information age'. In considering AI in cybersecurity and elsewhere, perhaps we should revisit Lewis Mumford's (1964, 273) provocation: 'Unless you have the power to stop an automatic process, you had better not start it'. Where is agency in the new cybersecurity assemblage and who or what makes the decisions that matter?

Acknowledgements I am grateful to Joe Devanny, Lilly Pijnenburg Muller, Erik Reichborn-Kjennerud, Max Smeets and the editors for their insightful comments and suggestions.

References

Amoore, Louise. 2013. *The Politics of Possibility: Risk and Security Beyond Probability*. Durham, NC: Duke University Press.

Amoore, Louise, and Marieke de Goede. 2008. Transactions after 9/11: the banal face of the preemptive strike. *Transactions of the Institute of British Geographers* 33 (2): 173–185.

Anderson, Chris. 2000. The end of theory: The data deluge makes the scientific method obsolete. Wired, 23 June. https://www.wired.com/2008/06/pb-theory/.

Aradau, Claudia, and Tobias Blanke. 2018. Governing others: Anomaly and the algorithmic subject of security. *European Journal of International Security* 3 (1): 1–21.

Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and R.B.J. Walker. 2014. After Snowden: Rethinking the impact of surveillance. *International Political Sociology* 8 (2): 121–144.

Berry, David M. 2019. Against infrasomatization: Towards a critical theory of algorithms. In *Data Politics: Worlds, Subjects, Rights*, ed. Didier Bigo, Engen Isin, and Evelyn Ruppert, 43–63. London: Routledge.

Betz, David, and Tim Stevens. 2013. Analogical reasoning and cyber security. *Security Dialogue* 44 (2): 147–164.

Brands, Hal. 2016. Paradoxes of the gray zone. Foreign Policy Research Institute, 5 February. https://www.fpri.org/article/2016/02/paradoxes-gray-zone/.

Brantly, Aaron F. 2018. *The Decision to Attack: Military and Intelligence Decision-Making*. Athens: University of Georgia Press.
Bresniker, Kirk, Ada Gavrilovska, James Holt, Dejan Milojicic, and Trung Tran. 2019. Grand challenge: Applying artificial intelligence and machine learning to cybersecurity. *Computer* 52 (12): 45–52.
Brundage, Miles and 25 others. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*. Oxford: Future of Humanity Institute and 13 others. https://maliciousaireport.com/.
Brunton, Finn. 2013. *Spam: A Shadow History of the Internet*. Cambridge, MA: MIT Press.
Capgemini. 2019. Reinventing Cybersecurity with Artificial Intelligence: A New Frontier in Digital Security. https://www.capgemini.com/research/reinventing-cybersecurity-with-artificial-intelligence/. Accessed 8 Jan 2020.
Castelvecchi, Davide. 2016. Can we open the black box of AI? *Nature* 538: 20–23.
Coker, Christopher. 2013. *Warrior Geeks: How 21st Century Technology is Changing the Way We Fight and Think About War*. London: Hurst.
Coles-Kemp, Lizzie, Debi Ashenden, and Kieron O'Hara. 2018. Why should I? Cybersecurity, the security of the state and the insecurity of the citizen. *Politics and Governance* 6 (2): 41–48.
Coles-Kemp, Lizzie, and René Rydhof Hansen. 2017. Walking the line: The everyday security ties that bind. In *Human Aspects of Information Security, Privacy and Trust*, ed. Theo Tryfonas, 464–480. Cham: Springer.
Collier, Jamie. 2019. A threat intelligence analyst's guide to today's sources of bias. Digital Shadows, 5 December. https://www.digitalshadows.com/blog-and-research/a-threat-intelligence-analysts-guide-to-todays-sources-of-bias/.
Collins, Jason. 2019. Principles for the application of human intelligence. Behavioral Scientist, 30 September. https://behavioralscientist.org/principles-for-the-application-of-human-intelligence/.
Congressional Research Service. 2019. *Artificial Intelligence and National Security*. Washington, DC. https://crsreports.congress.gov/product/pdf/R/R45178. Accessed 8 Jan 2020.
Corn, Gary P. 2019. Cyber national security: Navigating gray-zone challenges in and through cyberspace. In *Complex Battlespaces: The Law of Armed Conflict and the Dynamics of Modern Warfare*, ed. Christopher M. Ford and Winston S. Williams, 345–428. New York: Oxford University Press.
Corrigan, Jack. 2019. Pentagon, NSA prepare to train AI-powered cyber defences. Defense One, 4 September. https://www.defenseone.com/technology/2019/09/pentagon-nsa-laying-groundwork-ai-powered-cyber-defenses/159650/.
Darktrace. n.d. Cyber AI platform. https://www.darktrace.com/en/technology/. Accessed 1 Jan 2020.
Dear, Keith. 2019. Artificial intelligence and decision-making. *RUSI Journal* 164 (5–6): 18–25.
Hayles, N. Katherine. 2016. Cognitive assemblages: Technical agency and human interactions. *Critical Inquiry* 43 (1): 32–55.
Healey, Jason. 2019. The implications of persistent (and permanent) engagement in cyberspace. *Journal of Cybersecurity*. https://doi.org/10.1093/cybsec/tyz008.
Husymans, Jef. 2011. What's in an act? On security speech acts and little security nothings. *Security Dialogue* 42 (4–5): 371–383.
Jasper, Scott E. 2017. US cyber threat intelligence sharing frameworks. *International Journal of Intelligence and Counterintelligence* 30 (1): 53–65.
Jensen, Benjamin M., Christopher Whyte, and Scott Cuomo. 2019. Algorithms at war: The promise, perils, and limits of artificial intelligence. *International Studies Review*. https://doi.org/10.1093/isr/viz025.
Kaufmann, Mareile. 2019. Who connects the dots? Agents and agency in predictive policing. In *Technologies and Agency in International Relations*, ed. Marijn Hoijtink and Matthias Leese, 141–163. London: Routledge.
Kitchin, Rob. 2014. Big data, new epistemologies and paradigm shifts. *Big Data and Society*. https://doi.org/10.1177/2053951714528481.
Kitchin, Rob. 2017. Thinking critically about and researching algorithms. *Information, Communication and Society* 20 (1): 14–29.
Landwehr, Carl E. 2007. Cybersecurity and artificial intelligence: From fixing the plumbing to smart water. *IEEE Security and Privacy* 6 (5): 3–4.
Lewis, Larry. 2019. Resolving the battle over artificial intelligence in war. *RUSI Journal* 164 (5–6): 62–71.
Libicki, Martin C. 2009. *Cyberdeterrence and Cyberwar*. Santa Monica: RAND Corporation.
Libicki, Martin C., Lillian Ablon, and Tim Webb. 2015. *The Defender's Dilemma: Charting a Course Toward Cybersecurity*. Santa Monica: RAND Corporation.
McDonald, Henry. 2019. AI expert calls for end to UK use of 'racially biased' algorithms. The Guardian, 12 December. https://www.theguardian.com/technology/2019/dec/12/ai-end-uk-use-racially-biased-algorithms-noel-sharkey.
Matsakis, Louise. 2019. How the West got China's social credit system wrong. Wired, 29 July. https://www.wired.com/story/china-social-credit-score-system/.
Ministry of Defence. 2018. *Human–Machine Teaming*. Joint Concept Note 1/18. Shrivenham: Development, Concepts and Doctrine Centre.
Mumford, Lewis. 1964. The automation of knowledge. *AV Communication Review* 12 (3): 261–276.
Parisi, Luciana. 2019. Critical computation: Digital automata and general artificial thinking. *Theory, Culture and Society* 36 (2): 89–121.
Payne, Kenneth. 2018. Artificial intelligence: A revolution in strategic affairs? *Survival* 60 (5): 7–32.
Pickering, Andrew. 2010. Material culture and the dance of agency. In *The Oxford Handbook of Material Culture Studies*, ed. Dan Hicks and Mary C. Beaudry, 191–208. Oxford: Oxford University Press.
Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York: W.W. Norton.
Schneider, Jacquelyn. 2019. The capability/vulnerability paradox and military revolutions: Implications for computing, cyber, and the onset of war. *Journal of Strategic Studies* 42 (6): 841–863.
Schneier, Bruce. 2016. Stop trying to fix the user. *IEEE Security and Privacy* 14 (5): 96.
Shires, James. 2018. Enacting expertise: Ritual and risk in cybersecurity. *Politics and Governance* 6 (2): 31–40.
Srnicek, Nick. 2014. Cognitive assemblages and the production of knowledge. In *Reassembling International Theory: Assemblage Thinking and International Relations*, ed. Michele Acuto and Simon Curtis, 40–47. Basingstoke: Palgrave Macmillan.
Valeriano, Brandon, Benjamin Jensen, and Ryan C. Maness. 2018. *Cyber Strategy: The Evolving Character of Power and Coercion*. New York: Oxford University Press.
Von Hilgers, Philipp. 2011. The history of the black box: The clash of a thing and its concept. *Cultural Politics* 7 (1): 41–58.
Wall, David S. 2013. Enemies within: Redefining the insider threat in organizational security policy. *Security Journal* 26 (2): 107–124.
Weber, Steven. 2017. Coercion in cybersecurity: What public health models reveal. *Journal of Cybersecurity* 3 (3): 173–183.
Weber, Jutta, and Lucy Suchman. 2016. Human-machine autonomies. In *Autonomous Weapons Systems: Law, Ethics, Policy*, ed. Nehal Bhuta, Susanne Beck, Robin Geiß, Hin-Yan Liu, and Claus Kreß, 75–102. Cambridge: Cambridge University Press.

Wilkinson, Dick. 2019. The future of AI in cybersecurity. CISO Mag, 29 October. https://www.cisomag.com/the-future-of-ai-in-cybersecurity/.

Wirtz, James J. 2017. Life in the 'gray zone': Observations for contemporary strategists. *Defense and Security Analysis* 33 (2): 106–114.

Woolley, Samuel C., and Philip N. Howard (eds.). 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. New York: Oxford University Press.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Dr. Tim Stevens is Senior Lecturer in Global Security at the Department of War Studies, King's College London. His research engages with the political and strategic dimensions of cybersecurity, cyberwarfare and cyberespionage. He is head of the KCL Cyber Security Research Group, Senior Fellow and Associate Researcher at Conservatoire national des arts et métiers, Paris, and a Fellow of the Royal Geographical Society. His latest book, with Nicholas Michelsen, is *Pessimism in International Relations: Provocations, Possibilities, Politics* (Palgrave Macmillan, 2019).