19. Deception in Cyberwarfare

Simon Henderson

Introduction

> …there has arisen the mistaken impression that the magician's art begins and ends in the devices he employs — whereas, in fact, those devices are merely his working tools. His art does not consist in the things he uses, nor in the trade secrets and technical processes he has at command, but in the employment of those facilities with adequate efficiency. It consists in what he does with the things he uses, not in those things themselves.
> (Maskelyne and Devant, 1911, p. viii)

Deception is ubiquitous across life, from the smallest microbial and unicellular organisms to the global machinations of geopolitical statecraft. Deception pervades all natural habitats, with diverse species of predators and prey exploiting deception to gain an advantage over their competition, enabling survival and reproduction. Deception is inherent in almost all human endeavours, where it enables one actor to creatively outthink, outmanoeuvre and outgun another. Increasingly, deception is enacted in cyberspace, where truth and falsehood are

exceedingly challenging to discern, and covert malicious activity is especially problematic to attribute.

Deception has been a foundational strategy of warfare since the beginning of human conflict. Its use has been articulated throughout history in the manuscripts of revered commanders and strategists from China (Tzu, 2002), ancient Greece (Herodotus, 1899), ancient India (Kautilya, 1992), the Roman Empire (Frontinus, 1925) and a variety of other regions and time periods (Wavell, 1946; Musashi, 1974; Jomini, 1992; Whaley, 1982, 2006; Machiavelli, 2003). In contemporary warfare, consideration of deception is doctrinally stipulated within the planning processes of most militaries.

The prevalence of cyberspace as a contemporary theatre of war raises important questions about the evolution of deception in warfare:

- What exactly constitutes deception in the present day?
- Is cyber deception comparable to deception in the physical world?
- What purpose does deception serve in cyberwarfare?
- How do humans and machines collaborate to enable cyber deception?
- And how will the relentless advancement of technology, particularly in machine learning, shape the role of deception in future cyberwarfare?

This chapter explores these questions. It begins by reviewing dictionary and military definitions of deception before proposing a more rigorous and utilitarian alternative. I then consider traditional military applications of deception and examine the different forms of deception in cyber operations. The chapter reviews various case studies to show how deception contributes to cyberwarfare. Challenges to using deception within current and future cyberwarfare are discussed, and I conclude by summarising key emergent issues.

What is deception?

> One must not let oneself be deceived by the word 'deception' (Kiekegaard, 2009, pp. 63-64).

Before considering the role and function of deception in cyberwarfare, it is first essential to address the thorny issue of what deception 'is'. Definitions of the term abound. For example, the Merriam Webster Dictionary (2023) defines 'deception' as, 'The act of causing someone to accept as true or valid what is false or invalid'. Dictionary definitions like this are often weak, incomplete, confusing, or wrong. Problems frequently arise from the naive and impoverished equation of deception with lying. In practice, deception does not require any false statement to be made or, indeed, any statement at all. Dictionary definitions tend to focus only on the communicative aspects of falsehood while failing to recognize that deception also involves intentionally *not* communicating things that are real (as is fundamental to all covert action). Notions of truth and falsehood also tend to be grossly oversimplified, as the distinction is rarely binary.

Military definitions fare slightly better, although they often lack rigour, precision, and utility. For example, a common definition used across a range of US publications, including *Joint Doctrine for Military Deception* (US Joint Staff, 1996), Joint Publication 3-13.4 - *Military Deception* (2012), and Army Field Manual 3-13.4, *Army Support to Military Deception* (Headquarters Department of the Army, 2019) suggests that deception comprises, 'actions executed to deliberately mislead adversary military, paramilitary, or violent extremist organization decision makers, thereby causing the adversary to take specific actions (or inactions) that will contribute to the accomplishment of the friendly mission'. While this definition helpfully incorporates the notion of intent, potential types of target, and a behavioural outcome, the definition relies on the notion of 'misleading' — in other words, deceiving — and is, therefore, tautological. Tautology also occurs in the more recent proposed

NATO definition contained in *Allied Joint Doctrine for Operations Security and Deception* (NATO, 2020, p. 4): 'Deliberate measures to mislead targeted decision-makers into behaving in a manner advantageous to the commander's objectives.'

In 2011, I proposed a new definition of deception that sought to address such concerns and set the notion of deception against a more contemporary, pragmatic and utilitarian psychological foundation. Over the intervening years, the definition has been subject to thorough examination, critique and adoption by multiple military, intelligence, cyber and law enforcement organisations and communities. Deception is therefore (re)defined as, 'Deliberate measures to induce erroneous sensemaking and subsequent behaviour within a target audience, to achieve and exploit an advantage' (Henderson, 2011). The definition is generic and not specific to the military, hence the specification of a 'target' and not 'the enemy'. The definition also seeks to accommodate *all* deception, including military deception, *and* the subset of deceptive communications that comprises lying. While the definition frames deception as a human activity, by identifying analogue processes for sensemaking and behaviour within other domains, it also accounts for non-human deception, including plant and animal deception, software deception, deceptive manipulation of machine learning, etc. The definition, therefore, also applies aptly to the field of cyber deception. Importantly, this definition is value-neutral, so as to embrace both malevolent and benevolent applications of deception.

The definition incorporates the following components:

- **Deception is deliberate, intentional, and motivated.** Deception does not occur spontaneously and without motivation.
- **Deception requires action.** Deception does not and cannot happen by itself.
- **Deception induces errors in sensemaking.** Sensemaking is the deliberate effort to understand one's environment and events. It involves drawing from experience to

recognize and make sense of patterns in the data we perceive from the environment. A deceiver intentionally causes the target's understanding of the world to be wrong or in error. Erroneous sensemaking thereby differentiates deception from other related concepts, such as influence, persuasion, or coercion.

- **Deception aims to change the target's behaviour.** If the target's behaviour does not change, the deceiver could and would have achieved the same outcome by doing nothing.
- **Deception targets specific individuals, groups, organisations, or systems** (including computer software, algorithms, hardware control systems, etc.).
- **All successful deception creates an advantage for the deceiver.** Benevolent forms of deception also benefit the target.

As deception seeks behaviour change in the target, all deception involves influencing the target. However, not all influence involves deceiving the target; for example, when the influencer has no need to invoke errors in the target's sensemaking to elicit a desired behaviour. To effect a change in the target's behaviour, a deceiver can manipulate the six core psychological processes that a target uses to make sense of the world and take action (Henderson, 2019). These processes comprise:

- **Attention.** Where a target deploys or orientates its sensory systems to collect information about its environment.
- **Perception.** What a target sees, hears, smells, tastes, feels, etc., based on the information it has collected.
- **Sensemaking.** What a target understands and believes, and what they decide to do as a result of this belief.

- **Expectations.** What a target anticipates about the future state of the world, both if they do nothing and if they take action.
- **Emotion.** How a target feels about the current and anticipated future situation.
- **Behaviour.** The action a target takes resulting from these preceding processes.

Dependencies between these processes mean that a deceiver cannot shape later processes (sensemaking, expectations, emotion and behaviour) directly. Rather, they can only be shaped by manipulating earlier processes that are 'accessible' to the deceiver: attention (where the target 'looks') and perception (what the target 'sees'). The two fundamental principles of all deception therefore comprise 'hiding the real' and 'showing the false'. (Under certain circumstances, it may alternatively incorporate 'showing the real' and 'hiding the false'.)

The six core psychological processes identified above are scalable from individual to organisation and, potentially, even higher levels, such as a nation-state. The processes serve the same functions, irrespective of their scale, as depicted in Table 1.

Table 1 - The psychological building blocks of deception (Henderson, 2019).

	Attention	**Perception**	**Sensemaking**	**Expectations**	**Emotion**	**Behaviour**
Individual	Orientation of sight, sound, taste, touch, and smell	Recognition of stimuli	Sensemaking	Mental simulation	Emotional state	Action and communication

Organisation	Direction and deployment of Intelligence, Surveillance, Target Acquisition and Reconnaissance sensor systems	Multi-source sensor reporting processes	Intelligence analysis and planning	Forecasts, predictions, and plans	Individual and collective emotional states	Kinetic and information activities

Military deception primarily seeks to manipulate these processes at the collective level, although it may also target an individual's sensemaking and behaviour, such as that of the enemy commander. The six psychological processes also have equivalent technological processes that fulfil the same functions, and cyber deceivers similarly target and manipulate these. Technological analogues of the psychological processes are depicted in Table 2.

Table 2 - Technological analogues of the psychological processes

Attention	Perception	Sensemaking	Expectations	Emotions	Behaviour
The channels, connections, networks, pipes and filters that a system uses to connect to and acquire data.	Critical indicators that the system detects and recognizes, such as flags, keywords, spikes, trends, deviations, anomalies, etc.	The learning, pattern-matching, logic and probabilistic reasoning used to establish 'meaning' and govern the logic, flow and 'deductions' of the system.	Calculated, hypothesized or probabilistic data that the system searches for in its current data set or incoming live data streams.	Normality and baseline states, windows, thresholds, violations, anomalies, inconsistencies, alarms, and warnings.	Communication of meaning to users, visualisations and presentations, communication with other systems, further data collection, activation and control of connected hardware.

Having defined deception and its psychological and technological building blocks, let us now turn to the ways in which cyber deception seeks to induce erroneous sensemaking and subsequent behaviour in adversaries.

What is cyber deception?

> When the West does become involved [in wars], it increasingly relies on its huge technological advantage. This is to its benefit only so long as it remembers that wars are

> fought not by machines, but by men; and the best soldiers have a seasoning of devilry. (Wavell, 1948, p. 47).

Military deception is a strategy used by one force to gain an advantage over another. The enemy's collection capabilities and intelligence staff are led to formulate an erroneous assessment of the state of the world. As a result, the enemy's planning staff and its commander are led to make poor decisions and take actions that benefit the deceiver. By targeting and manipulating the six processes identified earlier, the deceiving force can shape how the enemy force makes sense of the numbers, capabilities, locations, actions, timings and intentions of friendly forces.

Traditional deception in warfare includes:

- Camouflage to make forces harder to detect by blending them with their background.
- Decoys to amplify perceived force size, suggest non-existent capabilities, project false presence, or draw enemy fire away from real equipment.
- False displays and movements to suggest presence, capability, or intent, portraying activities like surveying, engineering, preparatory activity, repairs, etc.
- Portraying false temporal indicators, such as speed of movement or other rates of change, or releasing information containing temporal details (for example, news stories mentioning forthcoming dates of military deployments).
- The 'accidental' release of false plans due to apparent mistakes, leaks or losses, a strategy known as 'The Haversack Ruse', particularly noted for its use in Operation Mincemeat in World War II (Macintyre, 2010; Coyle and Wilson, 2013).
- Feeding rumours into enemy collection networks via agents or other channels monitored by the enemy's intelligence network.

These activities lead the enemy to make erroneous sense of a situation, formulate incorrect assumptions, experience uncertainty, ambiguity, or confusion, and develop high degrees of misplaced confidence in their erroneous understanding of the current and future situation.

With respect to the contemporary application of deception to warfare, militaries have long pioneered and exploited information technologies to fool their adversaries, from the earliest written signals and ciphers to the advent of the internet, artificial intelligence, and generative systems. Despite exponential advances in technological capabilities, and as foreshadowed in the quotation above from Wavell, wars remain fought by people and not by machines. It is people who create, operate, maintain, and exploit military technologies and it is people who constitute our adversaries. Deception in cyberspace takes a multitude of forms. It encompasses humans deceiving humans, software deceiving humans, humans deceiving software, and software deceiving other software. All cyber deception comprises a battle between human wills, mediated through cyberspace and sometimes enacted by proxy. The human-to-human nature of cyber deception is depicted in Figure 1.

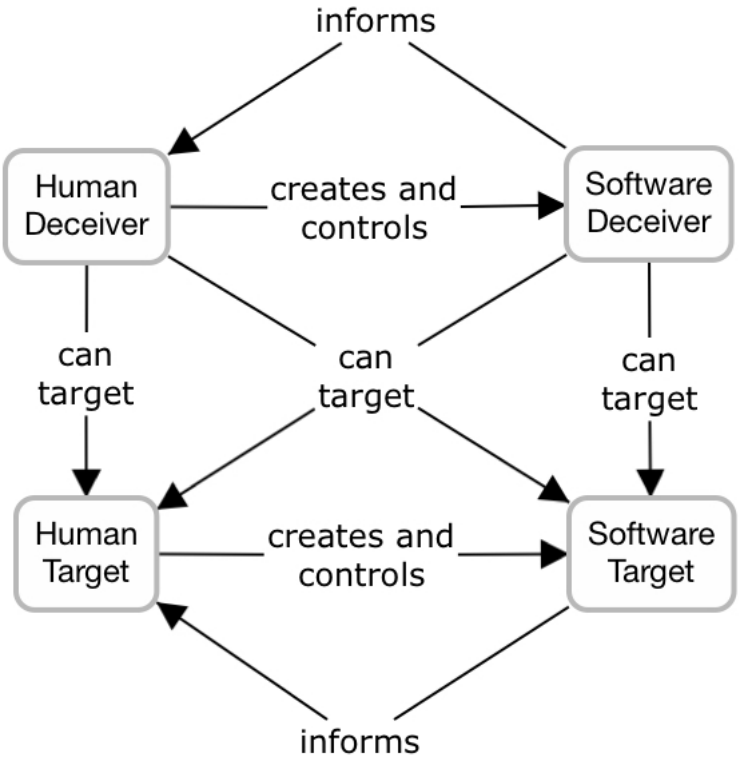

Figure 1 - Human to human deception by proxy (Henderson, 2019).

As a result, the target audience for deception within cyberspace may comprise not only an individual or collective human audience but software, such as an algorithm or machine learning system. In this context, inducing 'erroneous sensemaking' within the target can involve inducing errors in the operation and outputs of software, resulting in errors in decisions and actions, learning, classification and recognition, inferencing and deductions, predictions, or corrupted or biased data generation. This approach recognizes the 'sociotechnical' nature of cyberspace as an assemblage of diverse human and nonhuman actors (Collier, 2018).

Does cyberspace afford new forms of deception?

Despite the interplay between humans and machines that facilitates deception in cyberspace, a fundamental question remains: is deception in cyberspace the same as, or different from, deception in the real-world? The received wisdom is that core principles of deception persist irrespective of the technological environment in which it occurs. 'Deception is rooted in human nature,' write Michael Bennett and Edward Waltz (2007, p. 1), and 'humans continue to refine the means of deceit. While technology introduces new avenues and mechanism, the motives for deceit remain the same.' Barton Whaley (2006, p. ix) concurs, stating that '[b]ecause deception is a psychological mind-game, it doesn't change. However, the technology used to communicate disinformation does change.' Has traditional military deception therefore merely migrated into cyberspace, or does cyberspace afford and enable new forms of deception? The question is vital in adversarial settings as any novel, previously unseen approach to deception can sidestep active counterdeception measures, thereby creating a significant competitive advantage for the deceiver. In cyberspace, the fundamental strategies of deception (hiding the real and showing the false) remain unchanged. However, various characteristics of cyberspace *do* enable new means of deceiving that are quite different from those employed in the real-

world. Such approaches can be used in isolation or in concert to achieve synergistic effects. Let us consider some examples.

Modes of interaction. The enforced mediation of action in cyberspace via user and system interfaces, encoding, communications and decoding affords many new ways to deceive. For example, the design of web pages or emails can trick users into engaging in certain behaviours, such as clicking a malicious link (Potthast, et al., 2016); machine learning can recognize and respond to the distorted text in a CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) designed to validate human-only entry into a site (Geitge, 2017).

Global reach. Deception now extends beyond the boundaries imposed by the physical and geographical features of the real world, enabling the execution of deception to create effects at a global scale, all from the device carried in one's pocket (Jabbour, 2009).

Behavioural residue. People unknowingly leave behind traces of their online behaviour in the form of digital footprints or residue. Such data can be harvested and used to analyse, profile, target, manipulate, fool and impersonate any individual that has left such data (Sartonen et al., 2016).

Anonymity and impersonation. It is increasingly difficult to establish and validate the identity of online individuals and organisations. Cyberspace enables more rapid and voluminous manipulation of identity and presence than is achievable in the physical world. Online identities are now easy to obfuscate, spoof, simulate, generate, farm, network, replicate and modify (van der Walt et al., 2018).

Autonomy, repetition and permutation. Autonomy enables a host of new ways to cheat online through the navigation, mining and processing of mass data; network spidering, mapping, and exploitation; generation of new data and modification of old data; data

exfiltration, data posting and communication at scale; and the creation of false networks and other online activities (e.g., Spiliotopoulos et al., 2020).

Asymmetric effects. Small actors can achieve significant effects within cyberspace that would be almost impossible to create in the physical world, such as the 16-year-old schoolboy who, from his bedroom on the Shetland Islands, hacked the websites of the CIA, the UK Serious Organised Crime Agency, News International, Sony Ltd. and the Church of Scientology (Cadwalladr, 2012).

Generative systems. The generative capabilities of machine learning are growing exponentially, and the past couple of years have seen an explosion in the use and public recognition of such capabilities. It is now feasible to create machine-generated text, imagery, audio and video with sufficient credibility that humans can no longer differentiate it from real (natural) human-generated media. Indeed, some generative outputs, such as synthetic human faces, are now so good that people perceive them as 'more real' than real faces (Tucciarelli et al., 2022). Generative systems can render fake content more convincing (while eliminating tell-tale grammatical errors often made by influencers who create content in a non-native language), create a greater volume of content, or amplify real extreme content with fake comments and reactions. Generative systems can also produce credible scripts for deepfake videos and audio of leaders, politicians and military commanders (Oltermann, 2022).

Simulated mass movements. Simulated mass movements in cyberspace involve the creation of apparent large-scale, coordinated, but, in reality, non-existent online social movements. Such strategies are often called 'astroturfing' (Zhang et al., 2013), relating to the simulation of apparent grass-roots support. Such movements can include creating and distributing fake news or propaganda at scale, using bot networks to amplify messaging or artificially simulate mass sentiment or intent, and manipulating social media algorithms to increase the visibility and impact of accurate online content. Movements often use persona networks across multiple

platforms to simulate a credible presence, including media, postings, conversations, engineered influential debate (sometimes referred to as 'sock-puppeting'), histories and networks (Oleshchuk, 2020).

Hacking, espionage and sabotage. Circumventing cyber security systems relies upon humans fooling software and software fooling other software. Deceptive activities include: the covert exploration of software and hardware architectures to find vulnerabilities; presentation of false credentials; code modification; installing and executing malware; structured query language (SQL) injection attacks to gain access to databases; cross-site scripting to steal the credentials of site visitors; denial-of-service (DOS) attacks to overload a system with false requests and deny service provision; and man-in-the-middle attacks involving an attacker who sits between a user and a system to facilitate their interactions to their own benefit (Brar and Kumar, 2018). Gaining access to a system enables the exfiltration of critical information, holding the information to ransom (thereby creating the same effect as ransomware without the need for encryption or command and control), modification and potential posting of the (now false but credible) information, deletion of the information, or encryption of the information to extort a ransom.

Cyber social engineering. A wide variety of deceptive activities online involve the manipulation of individuals' curiosity and impulses to gain access to their credentials and information (Breda et al., 2017). Social engineering approaches include:

- Phishing attacks (provoking a user to activate a link).
- Scareware (for example, informing a user that they need to install anti-virus software to clean their infected system).
- Watering hole attacks that compromise a legitimate website to activate or install malware on the systems of its visitors.

- Pretexting (using a false identity, such as an IT manager) to interact with and manipulate a user to provide credentials or other information.
- Using cyberspace to help gain physical access to buildings (for example, by analysing Street View images of building entrances on Google Maps), as well as servers, systems, and information.

Misattribution and false-flagging. Misattribution involves planting false clues about the identity behind cyberspace actions. This includes using false personas; spoofing network and system identifiers and locations; manipulating residual digital footprints (for example, modifying or falsifying cookies); and falsifying other technical identifiers. For example, the malware used in an attack may have been coded on a system using a different language setting, have nationally specific browser settings and fonts, and be built using traceable national code libraries or libraries used in other previously attributed malware attacks. The malware may have been compiled at a time that corresponds with the working hours of another country. When executed, it communicates with servers that have been registered using a false identity, etc. (Pihelgas, 2015).

Poisoning the well. Poisoning the well involves manipulating the training data used for machine learning or the operational data the system subsequently processes (Baracaldo et al., 2017). As machine learning develops and becomes more widespread, there is a commensurate expansion in opportunities for its manipulation. For example, images can be manipulated so that they appear unaffected to humans but are interpreted differently by AI, either within their training and learning or in their operational data processing. This can be done by manipulating a few critical pixels or adding an invisible filter; both changes are imperceivable by humans. Consequently, an AI system may fail to identify the presence of real objects, identify non-existent objects, or misinterpret objects within a scene with a high degree of confidence (Nguyen et al., 2015; Ilyas et al., 2018; Wu et al., 2020). Machine learning can currently

manipulate text so that other machine learning text analysis systems formulate an incorrect interpretation while humans interpret the manipulated text as initially intended (Jin et al., 2020). Printable disruptive patches can be attached to real-world objects to render them undetectable or to support their misclassification by AI scene analysis, including the presence of people (Sharif et al., 2019; Wu et al., 2020), vehicles (Du et al., 2022), road signs (Sitawarin et al., 2018) and other objects (Athalye et al., 2018).

Defensive cyber deception. The past decade has seen a massive investment in deception-based cyber defence systems for protecting enterprise networks, and the market is expected to grow rapidly (Fortune Business Insights, 2022). Such systems aim to lead attackers to waste their time attacking false systems instead of real systems and to conduct their attacks within a safe and instrumented environment that can capture and analyse their behaviour, tools and strategies. Such approaches thereby enable the collection of threat intelligence. Some of the defensive systems that have been developed undoubtedly employ novel technologies to replicate real networks, computer systems, file structures, data, operating systems, security measures and system vulnerabilities. Some systems can adapt dynamically to the attacker's behaviour, like generating new false file systems as the attacker traverses the simulated network and reveals what they are looking for (Sajid, et al., 2020). Advanced monitoring systems enable users' behaviour to be tracked without detection.

However, the philosophy behind such systems and their practical effectiveness is open to question (Raina, 2023; Roncevich, n.d.). For example, if you were tasked to defend a castle, would you purchase and deploy another false castle beside the real castle to lure attackers so you can study them? Studying attackers and collecting threat intelligence is not the same as defending the network. And much of the functionality of deception-based cyber defence can be achieved without deception — for example, through routine data logging of file activity, service provision, user commands, application activation and activities, access attempts, etc.

Defensive cyber deception systems also tend to use an extremely limited subset of deceptive strategies (primarily, breadcrumbs and decoys) while omitting the vast range of other available deception strategies and they often struggle with false positives. They can also be challenging to justify in terms of their return on investment: why buy a threat analysis system when you could purchase contemporary threat knowledge from other experts, or invest that money in building better defences?

The use of deception in cyberwarfare

To paraphrase Sun Tzu's (2000, p. 3) famous quotation concerning warfare, 'All cyberwarfare is based on deception.'. Any online activity that involves hiding the real or showing the false is, by definition, deceptive. Online attackers and influencers fool defensive systems to gain access to protected information. They covertly exfiltrate information for intelligence, industrial espionage, ransom and extortion. They conceal their true identity and use false flags and cut-outs (witting or unwitting third parties) to lead analysts to misattribute their online activities to other actors. They seed, amplify and propagate falsehood to influence the thinking and behaviour of target audiences. They identify and exploit vulnerabilities to disrupt critical infrastructure. They execute campaigns of sabotage against government and commercial networks and they cause their targets significant economic damage. Deception is crucial to the success of these activities. The following examples illustrate the deployment of diverse forms of deception in cyberwarfare.

Cyber deception in support of conventional warfare

Russia's invasion of Ukraine in February 2022 exemplified effective state use of cyber deception to reinforce and support conventional military deception operations. In January 2022, Russia moved troops, artillery, and armour near the Ukrainian border, citing military exercises in response to a perceived threat from Ukraine. Troop movements extended into

Belarus, supposedly for joint exercises (Culbertson, 2022). However, various pieces of evidence strongly suggested these actions were preparations for an invasion. For example, Russia established large field hospitals in Belarus and Crimea capable of treating 1500 casualties (Devine, 2022). While such hospitals could be involved in a military exercise, the subsequent transportation of large amounts of blood to these hospitals raised suspicions (Stewart, 2022). On 15 February 2022, Russia claimed the exercises were complete and showcased tanks being moved onto rail transportation (Reevell, 2022). Nine days later, on 24 February, Russia invaded Ukraine, primarily using Belarus as a staging ground for a swift approach to Kyiv. Russia falsely justified the invasion as a special military operation to protect Donbas residents from genocide and to demilitarize and denazify Ukraine (Nikolskaya and Osborn, 2022; Rice-Oxley, 2022; Weber et al., 2022).

Before the invasion, Russia engaged in a comprehensive series of cyber activities to construct a supposed *casus belli* (provocation or justification for war). These actions included widespread information operations on platforms like Telegram and Twitter to portray Ukraine as the aggressor and Russia as the defender. This narrative was reinforced through traditional and social media channels, where Russia falsely accused Ukraine and the US of secretly producing biological weapons in clandestine laboratories (Ling, 2022).

An organisation associated with Russia's primary military intelligence organisation (the Main Intelligence Administration, formerly known as the Glavnoye Razvedyvatelnoye Upravlenie, or GRU) called Cadet Blizzard, used its destructive WhisperGate malware to target government agencies, nonprofits, IT organisations and emergency services. According to a Microsoft Threat Intelligence (2023) report, government and critical infrastructure websites were defaced, and a hack-and-leak operation conducted under the guise of hacktivist activity (a trademark GRU tactic) dumped data stolen from Ukrainian organisations onto the 'Free Civilian' Telegram channel.

In addition to this broader disinformation and disruption campaign, videos emerged online just days before the invasion, purporting to show Ukrainian sabotage of Russian targets and alleged shelling of a Russian kindergarten. Forensic analysis of these materials conducted by the open-source intelligence community (including Bellingcat and InformNapalm) soon discredited the veracity of the footage (Culverwell, 2022).

Cyber operations following the invasion targeted Ukraine's public, energy, media, financial, business and non-profit sectors, exfiltrating critical intelligence and disrupting critical national infrastructure and capabilities (Przetacznik and Tarpova, 2022). Often, surges of cyber activity support on-the-ground kinetic activity. For example, emergency response services that perform search and rescue, offer medical care, and distribute food, water, and medicine have been routinely targeted and disrupted by spikes in malicious traffic coinciding with Russian bombings (Cloudflare, 2023).

Russia's military and intelligence organisations have blended different forms of deception across the spectrum of real-world and cyber operations to support their activities in Ukraine. Preparations for the invasion were disguised as military exercises. Hacking activities for intelligence collection, disruption, and propaganda have been obfuscated, and vast amounts of false information have been promulgated to justify the invasion and support Russia's evolving narratives about the war.

Attack obfuscation

A 2023 Microsoft alert (Microsoft Threat Intelligence, 2023) stated that a People's Republic of China (PRC) state-sponsored cyber actor known as 'Volt Typhoon' had engaged in 'stealthy and targeted malicious activity focused on post-compromise credential access and network system discovery aimed at critical infrastructure organizations in the United States.' The alert suggested that the group was developing capabilities to disrupt critical communications

infrastructure between the United States and Asia region during future crises. One of the actor's primary deceptive tactics to support an attack is 'living off the land' (U.S. Cybersecurity and Infrastructure Security Agency, 2020). The strategy involves using built-in network administration tools to perform attack objectives, which enables an attacker to:

- Evade detection by merging and blending malicious activities into the stream of regular Windows system and network activities. Many behavioural indicators of an attack can also arise from legitimate, benign system administration commands, significantly complicating attack detection. These tactics resemble military camouflage techniques employed to conceal warfighters, vehicles, and structures by seamlessly blending them into their surroundings across the electromagnetic spectrum, including visual wavelengths.
- Avoid endpoint detection and response product alerts resulting from introducing third-party applications to the host.
- Limit the amount of activity that is captured in default logging configurations.

Volt Typhoon has also leveraged compromised small and home office network devices as intermediate infrastructure to obscure its malicious activities. It achieves this by having much of its command and control (C2) traffic emanate from local Internet Service Providers (ISPs) in the geographic area of the victim. The deceptive strategies used by Volt Typhoon include perceptual manipulation to repackage its malicious activities within a wrapper of real system activities and hiding indicators of its covert activity within the larger volume of legitimate system indicators — a form of perceptual 'dazzling' (Bell and Whaley, 2017, p. 50).

Deception in cyber espionage

In 2010, 34 companies, including major players such as Google, Yahoo, Symantec, Adobe, Northrop Grumman and Dow Chemical, were attacked and a trove of sensitive information and

intellectual property was exfiltrated (Cha and Nakashima, 2010). As part of the attack (named Operation Aurora), hackers also penetrated the Gmail accounts of Chinese political dissidents and human rights advocates in the United States, Europe and China. Data, including registry keys, IP addresses, runtime behaviour and other information derived from the attack indicated strongly that it was launched from China.

The attackers initially used phishing emails addressed to people in the target companies, which lured users to open a malware attachment. The malware then exploited zero-day vulnerabilities in Internet Explorer and other applications. Once activated, the malware could be controlled remotely, enabling access to the target's emails, file scanning and data exfiltration, and access to and recording of the user's webcam and microphone. In addition to stealing data from the initial target, the user's compromised system then proceeded to investigate the secure corporate intranet it belonged to, seeking out other weak systems and exfiltrating potential sources of intellectual property, including source code repositories. The connection used for exfiltrating user information mimicked a regular SSL connection (a standard security technology for establishing an encrypted link between a server and a client).

Deception strategies used in this attack included capturing the user's attention and provoking their curiosity to lure them to click a malware link, hiding the real through obfuscated network exploration and capture of data, and showing the false by disguising exfiltration activity as routine traffic within a regular SSL connection.

Cyber misattribution (false flagging)

Russia's Federal Security Service (Federalnaya Sluzhba Bezopasnosti, or FSB) has a group known as Turla, which uses various tools and techniques to target foreign government, military, technology, energy and commercial organisations for intelligence collection. A joint UK National Cyber Security Centre and US National Security Agency (2019) advisory

announced that Turla had taken over another hacker group's infrastructure to hijack and appropriate their spying operation. Turla used malware stolen from an Iranian hacking group known as APT34 (Oilrig) to misdirect attribution and sow confusion about their attacks. They also took control of Iranian command-and-control servers used by the malware. This enabled them to intercept data as the Iranian hackers were exfiltrating it and to send their commands to the target computers that had been hacked.

While false flagging has been a standard real-world ruse in military deception and espionage for centuries, if not millennia, it is increasingly used to obfuscate cyber attacks. The deception involved in Turla's activities primarily involved showing the false by mimicking or appropriating another organisation's source, methods, and tools (enabling the discovery of such indicators is akin to the Haversack Ruse indicated earlier). False flagging of this nature also seeks to exploit forensic analysts' emotional and sensemaking processes, as they are made to work hard to uncover and piece together clues as to the source of the attack and achieve satisfaction in coming confidently (but erroneously) to their conclusions.

Seed, amplify and propagate online falsehood

In 2020, several conservative news sources, including the Washington Examiner, RealClear Markets, American Thinker and The National Interest, released stories regarding the Middle East that were highly critical of Qatar while simultaneously advocating for stricter sanctions against Iran (Rawnsley, 2020; Vincent, 2020). The articles were authored by a network of at least 19 fake personas that had placed more than 90 opinion pieces in 46 different publications, all supporting a broader Middle East propaganda campaign. The fictitious authors' online profiles featured mirror-flipped facial imagery of actual journalists and synthetically generated faces lifted directly from the website thispersondoesnotexist.com.

The practice of using generative imagery to bolster false messaging is increasing rapidly. On 22 May 2023, a ‘verified’ Twitter account (with a blue tick, purchased for $8) called ‘Bloomberg Feed’ reported that there had been an explosion at the Pentagon (Marcelo, 2023). The tweet included an image of the Pentagon with black smoke billowing from it, captioned ‘Large Explosion near the Pentagon Complex in Washington D.C. - Initial Report’. The tweet soon began circulating on social media and was even shared by Russia Today (RT). The tweet was widely shared in investment circles, and the markets soon reacted (likely due to high-frequency algorithmic trading that accounts for news headlines when determining trades). As they opened at 9.30 AM, the Standard and Poor's 500 (S&P 500) dropped 0.3%. In addition, U.S. Treasury bonds and gold began to climb, suggesting that investors were seeking a more secure location to invest their funds. Analysis of the propagated image suggested that it had been generated by artificial intelligence software (as flaws and imperfections in features such as the building, fence, grass and concrete are typical of generative imagery).

The deception involved in these activities included mimicking the source of a legitimate news organisation, attracting attention through the promulgation of emotionally laden imagery and associated claims, showing the false by mimicking damage to the Pentagon and through the use of generative facial imagery to suggest account veracity.

Deception in the disruption of critical infrastructure

In June 2017, a variation of the ransomware Petya, known as NotPetya, targeted 300 Ukrainian companies and the country’s government, banking, and power grid systems (Greenberg, 2018). 30% of the nation's computers were paralysed, and 10% were erased completely, including those used for the Chernobyl nuclear facility cleanup. Various Western governments attributed the NotPetya attack to the Russian military (e.g., National Cyber Security Centre, 2018). The malware exploited a leaked US National Security Agency penetration tool that allows hackers to run their own code remotely on any unpatched machine. Once present on a device, NotPetya

recovered any user passwords persisting in temporary Random Access Memory (RAM) and used these to access and spread to other machines. Unlike ransomware, however, NotPetya was entirely destructive. The worm only simulated ransomware's look, feel and functionality. In reality, NotPetya's encryption was irreversible, and no decryption key existed. Any ransom paid by users was futile. In the first few hours following its release, the worm extended beyond the borders of Ukraine and began spreading indiscriminately around the world. It infected a raft of multinational commercial organisations such as Maersk, Merck and TNT Express and even spread back to Russia, impacting the state oil company Rosneft. The US White House assessed the total damages caused by NotPetya to be more than $10 billion (Greenberg, 2018).

The deception in the NotPetya attack primarily involved the simulation of common ransomware. In addition to hiding the real by obfuscating its infection, encryption, and transmission processes, the malware also showed the false by offering a false hope of recovery to its targets, leading them to waste precious time and (primarily financial) resources attempting to salvage their systems and operations.

Deception in the sabotage of government and commercial facilities

In June 2010, security researchers discovered a worm that had infected the control systems of at least 14 industrial sites in Iran, including a uranium-enrichment plant at Natanz (Kushner, 2013). Further research identified that the worm, known as Stuxnet, was created and deployed as part of a joint US and Israeli operation named Olympic Games (Nakashima and Warrick, 2012). Stuxnet was designed to infect industrial Programmable Logic Controllers (PLCs) to take control of gas centrifuges used for separating nuclear material. Deception enabled various stages of the attack. The worm was introduced into the control systems via a USB drive that could covertly execute malware merely due to a user browsing its file structure (Kushner, 2013). The operation recruited double agent cut-outs to deliver the drive to the plant, localising potential (false) attribution and distancing accurate attribution from the perpetrators (Sale,

2012). The worm was also signed with a stolen digital certificate that made it appear to originate from a reliable company, thereby enabling it to evade the automated security system (Bureau, 2010). Most striking, however, is that once installed on the target system, the worm passively recorded 21 seconds of regular operational centrifuge data. It later played this data on a loop to operators while it spun the centrifuges out of control, causing them to explode (Langner, 2013). Operators monitoring the control system were, therefore, unable to diagnose and rectify the cause of the problem.

The deception involved in this attack included hiding the real through the covert transmission, infection and execution of the malware, clandestine operational data collection, and hiding the genuine activity of spinning up the centrifuges. It also showed the false by using cut-outs for USB stick delivery, presenting a stolen digital certificate, and the simulation of routine centrifuge operations by presenting captured historical data as if it were real-time.

Deception in election interference

In its report entitled 'Russian Active Measures, Campaigns and Interference in the 2016 U.S. Election', the US Senate Select Committee on Intelligence (2020) reported that 'The Russian government directed extensive activity, beginning in at least 2014 and carrying into at least 2017, against U.S. election infrastructure at the state and local level.' The operation, named 'Project Lakhta', was conducted by the Internet Research Agency, a Russian company engaged in online propaganda and influence operations on behalf of Russian business and political interests. The operation aimed to harm the election campaign of Hillary Clinton while boosting the candidacy of Donald Trump. It also aimed to increase political and social discord in the United States by undermining the public's faith in democratic and electoral institutions. (Indeed, the discovery of Russia's interference activities only served to amplify such effects).

Activities conducted under the remit of Project Lakhta included:

- Covertly scanning election systems in all 50 states and collecting data on general election-related web pages, voter ID information, election system software, and election service companies.
- Scanning and exfiltrating the voter registration database from the Illinois Board of Elections website.
- Stealing Americans' identities and using them to open fraudulent bank and cryptocurrency accounts.
- Creating thousands of social media accounts across various platforms that falsely claimed to be operated by Americans supporting radical political groups. These accounts were used to organize and promote events favouring Trump and against Clinton. A Columbia University study (Timberg, 2017) revealed that Russians had created 470 Facebook accounts during the 2016 campaign. Six of those accounts had produced content that was shared over 340 million times.
- Computer hackers affiliated with the Russian GRU infiltrated the information systems of the Democratic National Committee (DNC), the Democratic Congressional Campaign Committee (DCCC), and the Clinton campaign.

Deception occurred throughout Project Lakhta. Scanning and exfiltration activities were hidden. False online profiles simulated the presence and activities of radical political groups, attracting attention and support and swaying voters' opinions. False yet strongly viral political information was created and promoted to capture attention and entrench recipients' thinking, leading them to further disseminate falsehoods.

These case studies, originally conceived as covert operations, have now emerged from the shadows and are exposed for the world to analyse. Their transformation into public knowledge arguably suggests their failure. Perhaps the best examples of cyber deception are the

(presumed) ongoing operations that have remained active and uninterrupted for years or even decades, evading detection by their unsuspecting targets who unknowingly continue to fuel the very activities that serve their deceivers' interests.

Challenges to conducting cyber deception

As cyber capabilities develop apace while costs continue to fall, practical challenges to planning, executing and learning from cyber deception remain. Talented and creative individuals and small groups can use cyberspace to wield asymmetric power, but state-sponsored or -sanctioned cyber deception usually is more of a team sport. When multiple groups and organisations need to work together to coordinate and synergize their actions in pursuit of national objectives, a variety of problems may arise. Examples of such problems are now discussed.

Interagency coordination and delineation of responsibility. Different parts of government, the military, and commercial and private organisations will likely have overlapping goals, target sets, intelligence assets, raw and refined intelligence, analytical capabilities, access to, and prior history and experience with their targets. Operational deconfliction, coordination, and sharing of capabilities and intelligence will continue to prove extremely challenging across organisations engaged in deceptive cyberwarfare.

The language of deception. The language for describing deception and deceptive operations is often local to a group or organisation. Such language is often incomplete, imprecise, and not reflective of the extensive history of deception research and practice and well-established language (Stech et al., 2011; Henderson, 2019). When planning and executing deceptive activities across departmental and organisational boundaries, differences in language can impose severe communications overheads and dramatically increase operational risk, frustrating and compromising the design, execution, and effectiveness of cyber operations.

Explicit and implicit delegation of cyber deception activities to outside organisations. According to a report by the Russian state-owned domestic news agency, RIA Novosti (2023), the Russian government is currently exploring the possibility of absolving Russian patriotic hackers from criminal liability for attacks carried out 'in the interests of the Russian Federation'. This immunity would apply to hackers located abroad and within Russia's borders. If the exemption is granted, it will embolden pro-Kremlin hackers by giving them unfettered legal freedom to conduct attacks. However, by sanctioning the use of cyber deception by others in this manner, states incur the risk that such beneficiaries may overstep their responsibilities, claim responsibility for the actions of others, make claims that are not backed up by evidence, and engage in adversarial and deleterious competition with other immune organisations (see also Maurer, 2017). Actions taken by unconstrained cyber actors could result in the state's loss of control, accurate or inaccurate attribution of attacks, compromise of technical capabilities, and the loss of plausible deniability. The risk of blowback and unintended national self-harm also increases (as was experienced by Russia resulting from its NotPetya attack).

Advances in counterdeception capabilities. Staying secret has always been challenging, but as adversarial sensing capabilities evolve and proliferate, covert online operations must now contend with a vast array of technologies that can expose their activities. Compromise of deceptive cyber operations may arise from cell phone and satellite tracking, ubiquitous CCTV, facial and number plate recognition, drones, real-time social media, a nexus of state-sponsored, military and criminal globally-connected databases, DNA analysis, and increasingly sophisticated machine learning that can infer and anticipate patterns, identity, tactics, and intent. Such developments give rise to a cyclical deception/counter-deception arms race, where advances in counterdeception capabilities spur on advances in deception capabilities, and *vice versa*, *ad infinitum* (Henderson, 2021).

Channel proliferation paradoxes. The development and proliferation of counterdeception capabilities, paradoxically, create new opportunities to increase the effectiveness of deception. More sensors increase the chances of a target uncovering deception by detecting discrepancies across information channels. However, more sensors also make it easier for an adversary to feed seemingly independent, corroborative, but false information to a target. More extensive sensor capabilities, therefore, both pose a threat to and create opportunities for a deceiver. As British physicist R.V. Jones noted in 1942, 'Deception becomes more difficult as the number of channels of information available to the victim increases. However, within limits, the greater the number of controlled channels, the greater the likelihood of the deception being believed.' **(**Jones, 1942).

Training and education for cyber deception. While World War II gave rise to a hard-won body of knowledge about the use of deception in warfare, its loss to future generations of deception practitioners was keenly anticipated by US President Eisenhower in 1947. In a memorandum to Lauris Norstad, Director of the Plans and Operations Division of the War Department, Eisenhower wrote: 'I consider it essential that the War Department should continue to take those steps that are necessary to keep alive the arts of psychological warfare and of cover and deception and that there should continue in being a nucleus of personnel capable of handling these arts in case an emergency arises' **(**Galambos, 1978, p. 1763).

The erosion and atrophy of military deceptive capability are now widely recognized (US Army, 1988; Nisbett, 2005; Sharpe, 2006; Baker, 2011). Military deception courses are rare in comparison to other forms of training. Military cyber deception-specific training is rarer still. Many technical aspects of cyber deception are well understood, but deception *operations* in cyberspace (beyond technical methods) are, arguably, significantly less well understood and practised. As noted by Whaley (2016, p. ix), 'The royal road to learning how to deceive in war has been paved with speed bumps. It is widely assumed that this learning process has been

incremental, a gradual accumulation of experience in combat, lessons learned in staff studies, scholarly analysis of historical cases, the passing of knowledge from master to apprentice, and practical experience in combat. In other words, the art of military deception is generally seen as improving slowly but steadily through a long chain of theory and practice. However, the reality is very different.' While this situation persists, cyber deception practitioners can only slowly and incrementally advance their deception skills through *ad hoc* on-the-job learning.

Unanticipated consequences. The ease with which disinformation and malware can be created, modified and proliferated across cyberspace means that the consequences of cyber deception can be incredibly difficult to control. This was exemplified in the indiscriminate proliferation and worldwide impact of NotPetya in the first few hours following its release, including its spread back to Russia. The complexity of the online environment also means that the consequences of deceiving in cyberspace may not be known for many years (for example, deceptive Russian cyber activities conducted to support its invasion of Ukraine are only just coming to light now).

Legal and ethical issues concerning cyber deception. While the legal aspects of real-world military deception are well established in national and international law, legalities pertaining to deception in the physical world do not necessarily translate to cyberspace. The *Tallinn Manual on the International Law Applicable to Cyberwarfare* describes illegal perfidious action within cyberspace as, 'acts inviting the confidence of an adversary to lead him to believe that he is entitled to, or is obliged to accord, protection under the rules of international law applicable in armed conflict, with the intent to betray that confidence' (Schmitt, 2013). However, the *Manual* does not address how such principles apply to cyberspace, for example, when falsified protective indicators are used to establish system-to-system trust. Further, 'If cyber deception operations become pervasive so that little or no reliance can be placed, say, on targeting data, what implications does this have for the ability of combatants to comply with distinction,

discrimination, proportionality and precautions rules, and does that matter?' (Chelioudakis, 2017). To better make sense of the legal issues surrounding the use of deception within cyberwarfare, perhaps it would be wise for legal scholars to start with a better foundational understanding of deception itself?

The ethical considerations in using deception as a component of cyberwarfare are perhaps clearer. The same deception strategies, methods and processes can be used to achieve malevolent *or* benevolent outcomes (Henderson, 2023). Deception is, therefore, value-neutral, and it is imperative that deceptiveness is not confounded with ethicality. Deception is like a surgeon's scalpel that can be used to kill or cure, depending only on *how* it is used. The scalpel itself, like deception, has no intrinsic ethical value. Therefore, the ethics of deceptive action in cyberwarfare should instead be assessed relative to the intent, execution and consequences of that deception.

While most of the case studies presented in this chapter have been non-Western-run deceptive cyber operations, the West engages in a raft of similar activities against what it perceives as hostile foreign states. Legal and ethical issues may arise when state-sanctioned deceptive cyber activities are made public, as exemplified in the leaks of Edward Snowden (Macaskill and Dance, 2013), and Graphika and Stanford Internet Observatory's (2022) revelations about the US government's use of fake social media accounts to spread pro-Western propaganda. It is essential that Western governments and militaries establish absolute clarity concerning the legal and ethical foundations underpinning their deceptive cyber operations when engaging (or when forced to engage) with the public concerning such activities (see also National Cyber Force, 2023; Joshi, 2023).

Conclusions

Deception in warfare has come a long way from its earliest manipulations of physical objects and information on the battlefield. While the primacy of inducing errors in human sensemaking pervades cyber deception, cyberspace affords new forms of deception that are not possible in the real-world. These new forms of deception extend considerably the tools available to the cyber warrior.

Cyber deception can be used as a force multiplier to enhance conventional kinetic warfare. It can augment real-world deception, providing apparent impetus, motivation, explanation and corroboration of real-world activities. Cyber deception can also occur entirely within the virtual world of cyberspace, although humans remain the sole initiators and targets of such deception, irrespective of the mediating technology and capabilities of any artificial intelligence involved. As in the real-world, cyber deception involves hiding the real and showing the false. It targets human psychological processes of attention, perception, sensemaking, expectations, emotion and behaviour. It also targets analogues of these processes operating within technological systems.

The expanding geography of cyberwarfare means that physical terrain, boundaries, resources, time, and the limitations of human psychology and physiology no longer serve to constrain deception. The traditional military deception target of the enemy commander and their staff is no longer guaranteed. Cyber deception may instead seek to induce erroneous sensemaking in technical, legal, political, governmental, diplomatic, financial, commercial, public and infrastructure targets. Or to induce error in the ‘sensemaking’ of software logic, algorithms, machine learning models, or hardware systems. To make sense of cyber deception across these varied domains, scholars and practitioners must adopt a multidisciplinary, socio-technical perspective.

Advances in machine learning, automated pattern discovery, generative media, cryptology, changes in media and news consumption, and state attitudes and policies towards risk will likely shape radically the future of cyber deception and counterdeception. The technological deception/counterdeception arms race will continue to evolve and accelerate, but the capabilities of the human mind will remain essentially unchanged. In conclusion, to quote from the words of magician Dariel Fitzkee, 'Ultimately it is the spectator's mind which must be deceived, or there is no deception whatever. All of the apparatus we use, all of the secret gimmicks we employ, all of the sleights and stratagems we invoke — everything which identifies magic as mystery — the whole is designed to deceive the mind, and the mind alone, of the spectator.' (Fitzkee, 1945, p.27).

References

Athalye, Anish, Logan Engstrom, Andrew Ilyas and Kevin Kwok (2018) Synthesizing robust adversarial examples. *Proceedings of Machine Learning Research* 80: 284-293.

Baker, Richard (2011) The lost and found art of deception. *US Army.* 17 November. http://www.army.mil/article/66819/Thelostandfoundartofdeception/.

Baracaldo, Nathalie, Bryant Chen, Heiko Ludwig and Jaehoon Amir Safavi (2017) Mitigating poisoning attacks on machine learning models: a data provenance based approach. *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, 103-110.

Bennett, Michael and Edward Waltz (2007) *Counterdeception: Principles and Applications for National Security*. Norwood, MA: Artech House.

Bell, J. B., & Whaley, B. (2017). *Cheating and Deception*. Routledge.

Brar, Harmandeep Singh and Gulshan Kumar (2018) Cybercrimes: a proposed taxonomy and challenges. *Journal of Computer Networks and Communications*, https://doi.org/10.1155/2018/1798659.

Breda, Filipe, Hugo Barbosa and Telmo Morais (2017). Social engineering and cyber security. *Proceedings of the 11th International Technology, Education and Development Conference*, 4204-4211.

Bureau, Pierre-Marc (2010). Win32/Stuxnet signed binaries. *We Live Security*. 19 July. https://www.welivesecurity.com/2010/07/19/win32stuxnet-signed-binaries/.

Cadwalladr, Carole (2012) "Look out for the pale kid that needs a haircut". *Irish Examiner*. 22 September. https://www.irishexaminer.com/lifestyle/features/humaninterest/look-out-for-the-pale-kid-that-needs-a-haircut-208421.html.

Cha, Ariana Eunjung and Ellen Nakashima (2010) Google China cyberattack part of vast espionage campaign, experts say. *Washington Post*. 14 January. https://www.washingtonpost.com/wp-dyn/content/article/2010/01/13/AR2010011300359.html.

Chelioudakis, Eleftherios (2017) Deceptive AI machines on the battlefield: Do they challenge the rules of the Law of Armed Conflict on military deception? https://ssrn.com/abstract=3158711.

Cloudflare (2023) *Project Galileo 9th Anniversary Report*. 5 June. https://radar.cloudflare.com/reports/project-galileo-9th-anniv.

Collier, Jamie (2018) Cyber security assemblages: a framework for understanding the dynamic and contested nature of security provision. *Politics and Governance* 6(2): 13-21.

Coyle, Gene A. and Alexander Wilson (2013) Haversack ruses: From leather to digital. *International Journal of Intelligence and CounterIntelligence* 27(1): 156-177.

Culbertson, Alix (2022, February 18) Ukraine crisis: Putin says military drills 'purely defensive' and 'not a threat' as Western leaders warn invasion imminent. *Sky News*. 18 February. https://news.sky.com/story/ukraine-crisis-putin-says-military-drills-purely-defensive-and-not-a-threat-as-western-leaders-warn-invasion-imminent-12545284.

Culverwell, Dominic (2022) Russian disinformation accuses Ukraine of kindergarten attack. *Intellinews.* 20 February. https://www.intellinews.com/russian-disinformation-accuses-ukraine-of-kindergarten-attack-235479/.

Devine, Kieran (2022) The final pieces: Three new signs of Russian invasion plans. *Sky News.* https://news.sky.com/story/the-final-pieces-three-new-signs-of-russian-invasion-plans-for-ukraine-12533199.

Du, Andrew, Bo Chen, Tat-Jun Chin et al. (2022) Physical adversarial attacks on an aerial imagery object detector. *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision*, 3798-3808.

Fitzkee, Dariel (1945) *Magic By Misdirection*. San Rafael, CA: San Rafael House.

Fortune Business Insights (2022) *Deception Technology Market Size, Share and COVID-19 Impact Analysis*. February. https://www.fortunebusinessinsights.com/deception-technology-market-102220.

Frontinus (1989) [1925] *Stratagems and Aqueducts* Trans. Charles E. Bennett. London and Cambridge, MA: Loeb/Harvard University Press.

Galambos, Louis, ed. (1978) *The Papers of Dwight David Eisenhower: The Chief of Staff*, vol. VIII. Baltimore, MD: Johns Hopkins University Press.

Geitge, Adam (2017) How to break a CAPTCHA system in 15 minutes with Machine Learning. *Medium.* 13 December. https://medium.com/@ageitgey/how-to-break-a-captcha-system-in-15-minutes-with-machine-learning-dbebb035a710.

Graphika and Stanford Internet Observatory (2022) *Unheard Voice: Evaluating Five Years of Pro-Western Covert Influence Operations*. 24 August. https://stacks.stanford.edu/file/druid:nj914nx9540/unheard-voice-tt.pdf.

Greenberg, Andy (2018) The untold story of NotPetya, the most devastating cyberattack in history. *Wired*. 22 August. https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/.

Headquarters Department of the Army. (2019). *FM3-13.4: Army Support to Military Deception*. Army Publishing Directorate. https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN15310-FM_3-13.4-000-WEB-2.pdf

Henderson, Simon (2011) Deceptive Thinking Workshop. Paper presented at the 1st MilDec Military Deception Symposium, 2-3 November, Defence Academy of the United Kingdom, Shrivenham. Manuscript in possession of author.

Henderson, Simon (2019) *The Artifice System Handbook*. National Cyber Deception Laboratory and Artifice Ltd. https://www.cyberdeception.org.uk/wp-content/themes/ncdl/im/NCDL-Abridged-Version.pdf.

Henderson, Simon (2021) How to win the deception/counter-deception arms race? April. https://deceptionbydesign.com/wp-content/uploads/2021/04/Henderson-Deception-Arms-Race-1.0.pdf.

Henderson, Simon (2023) Creativity and morality in deception. In Hanisha Kapoor and James C. Kaufman, eds. *Creativity and Morality*. Cambridge, MA: Academic Press, 101-124.

Herodotus (1899) *The Histories of Herodotus*. Trans. Henry Cary. New York: D. Appleton and Company.

Ilyas, Andrew, Logan Engstrom, Anish Athalye and Jessy Lin (2018). Black-box adversarial attacks with limited queries and information. In *Proceedings of the 35th International Conference on Machine Learning, PMLR* 80: 2137-2146.

Jabbour, Kamal (2009). The science and technology of cyber operations. *High Frontier* 5(3): 11-15.

Jin, Di, Zhijing Jin, Joey Tianyi Zhou and Peter Szolovits (2020) Is BERT really robust? A strong baseline for natural language attack on text classification and entailment. In *Proceedings of the AAAI Conference on Artificial Intelligence* 34(5): 8018-8025.

Jomini, Antoine Henry (1992) [1838] *The Art of War.* Trans. G.H. Mendell and W.P. Craighill. London: Greenhill Books.

Jones, R.V. (1942) *'Report No.13. D.T.: Beams/Radar'*. 10 January. London: National Archives, NCUACS 95.8.00/B.24.

Joshi, Shashank (2023, April 4th). Cyberwarfare is all in the mind, says Britain. *The Economist*. 4 April. https://www.economist.com/britain/2023/04/04/cyberwarfare-is-all-in-the-mind-says-britain.

Kautilya (1992) *The Arthashastra*. Ed. and trans. L. N. Rangarajan. New Delhi: Penguin Books India.

Kierkegaard, Søren (1998) *Kierkegaard's Writings XXII: The Point of View*. Eds. Howard V. Hong and Edna H. Hong. Princeton, NJ: Princeton University Press.

Kushner, David (2013) The real story of Stuxnet. *IEEE Spectrum* 50(3): 48-53.

Langner, Ralph (2013) *To Kill a Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve*. November. Arlington, VA: The Langner Group. https://www.langner.com/wp-content/uploads/2017/03/to-kill-a-centrifuge.pdf.

Ling, Justin (2022) How 'Ukrainian bioweapons labs' myth went from QAnon fringe to Fox News. *The Guardian*. 18 March. https://www.theguardian.com/media/2022/mar/18/ukrainian-bioweapons-labs-qanon-fox-news.

Macaskill, Ewan and Gabriel Dance (2013). NSA files: Decoded. *The Guardian*. 1 November. https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded.

Machiavelli, Niccolo (2003) *Art of War.* Ed. and trans. Christopher Lynch. Chicago, IL: University of Chicago Press.

Macintyre, Ben (2010) *Operation Mincemeat: The True Spy Story That Changed the Course of World War II*. London: Bloomsbury.

Marcelo, Philip (2023) Fact focus: Fake image of Pentagon explosion briefly sends jitters through stock market. *Associated Press*. 23 May. https://apnews.com/article/pentagon-explosion-misinformation-stock-market-ai-96f534c790872fde67012ee81b5ed6a4.

Maskelyne, Nevil and David Devant (1911) *Our Magic: The Art in Magic, the Theory of Magic, the Practice of Magic*. Boston, MA: E.P. Dutton

Maurer, Tim (2017) *Cyber Mercenaries*. Cambridge: Cambridge University Press.

Merriam Webster Dictionary (2023) Deception. https://www.merriam-webster.com/dictionary/deception.

Microsoft Threat Intelligence (2023) Cadet Blizzard emerges as a novel and distinct Russian threat actor. 14 June. https://www.microsoft.com/en-us/security/blog/2023/06/14/cadet-blizzard-emerges-as-a-novel-and-distinct-russian-threat-actor/.

Microsoft Threat Intelligence (2023) Volt Typhoon targets US critical infrastructure with living-off-the-land techniques. 24 May. https://www.microsoft.com/en-us/security/blog/2023/05/24/volt-typhoon-targets-us-critical-infrastructure-with-living-off-the-land-techniques.

Musashi, M. (1974). *The Book of Five Rings*. Trans. Victor Harris. London: Allison and Busby.

Nakashima, Ellen and Joby Warrick (2012) Stuxnet was work of U.S. and Israeli experts, officials say. *The Washington Post*. 2 June. https://www.washingtonpost.com/world/national-security/stuxnet-was-work-of-us-and-israeli-experts-officials-say/2012/06/01/gJQAlnEy6Ustory.html.

National Cyber Force (2023) *Responsible Cyber Power in Practice*. 4 April. https://www.gov.uk/government/publications/responsible-cyber-power-in-practice.

National Cyber Security Centre (2018). Russian military 'almost certainly' responsible for destructive 2017 cyber attack. 17 Septemember. https://www.ncsc.gov.uk/pdfs/news/russian-military-almost-certainly-responsible-destructive-2017-cyber-attack.pdf.

National Cyber Security Centre and National Security Agency (2019) Advisory: Turla group exploits Iranian APT to expand coverage of victims. 21 October. https://www.ncsc.gov.uk/news/turla-group-exploits-iran-apt-to-expand-coverage-of-victims.

Pihelgas, Mauno, ed. (2015) *Mitigating Risks Arising from False-Flag and No-Flag Cyber Attacks.* Tallinn: NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/uploads/2018/10/False-flag-and-no-flag-20052015.pdf.

NATO (2020) *Allied Joint Publication-3.10.2, Allied Joint Doctrine for Operations Security and Deception*. Edition A, Version 2. March. https://www.gov.uk/government/publications/allied-joint-doctrine-for-operations-security-and-deception-ajp-3102a.

Nguyen, Anh, Jason Yosinski and Jeff Clune (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR*) 2015, 427-436.

Nisbett, Thad (2005) Operational Deception: The Lost Art in Today's Operations. Newport, RI: Naval War College.

Oleshchuk, Petro (2020) The instruments of modern media lobbying. *Future Human Image* 14: 48-55.

Oltermann, Philip (2022) European politicians duped into deepfake video calls with mayor of Kyiv. *The Guardian*. 25 June. https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko.

Osborn, Andrew and Olina Nikolskaya, P. (2022) Russia's Putin authorises 'special military operation' against Ukraine. *Reuters*. 24 February. https://www.reuters.com/world/europe/russias-putin-authorises-military-operations-donbass-domestic-media-2022-02-24/.

Potthast, Martin, Sebastian Köpsel, Benno Stein and Matthias Hagen (2016) Clickbait detection. Advances in Information Retrieval. In Nicola Ferro, Fabia Crestani, Marie-Francine et al. eds., *Advances in Information Retrieval: 38th European Conference on IR Research.* Cham: Springer, 810-817.

Przetacznik, Jakub and Simona Tarpova (2022). Russia's war on Ukraine: Timeline of cyber-attacks. European Parliamentary Research Service. 21 June. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2022)733549.

Raina, Kapil (2023) How deception technology fools everyone with a false sense of security. *SC Magazine.* 30 March. https://www.scmagazine.com/perspective/how-deception-technology-fools-everyone-with-a-false-sense-of-security.

Rawnsley, Adam (2020) Right-wing media outlets duped by a Middle East propaganda campaign. *The Daily Beast*. 6 July. https://www.thedailybeast.com/right-wing-media-outlets-duped-by-a-middle-east-propaganda-campaign.

Reevell, Patrick (2022) Russia says some troops returning to base from Ukraine border. *ABC News*. 15 February. https://abcnews.go.com/International/russia-signals-troop-pullback-ukraine-border-exercises/story?id=82896967.

RIA Novosti (2023) В Госдуме предложили не наказывать хакеров, работающих в интересах России (The State Duma proposes not to punish hackers working in the interests of Russia). 10 February. https://ria.ru/20230210/khakery-1851213742.html.

Rice-Oxley, Mark (2022). Is there any justification for Putin's war? *The Guardian.* 13 March. https://www.theguardian.com/world/2022/mar/13/is-there-any-justification-for-putins-war.

Roncevich, Tim (n.d.) Deception technology: Useful tool or just more busywork? *Cyber Defense Magazine.* https://www.cyberdefensemagazine.com/deception-technology-useful-tool-or-just-more-busywork/

Sajid, Mohammed S.I., Jingping Wei, J., Mohammed Rabbi Alam et al. (2020) Dodgetron: Towards autonomous cyber deception using dynamic hybrid analysis of malware. *2020 IEEE*

Conference on Communications and Network Security (CNS). https://doi.org/10.1109/CNS48642.2020.9162202.

Sale, Richard (2012) Stuxnet loaded by Iran double agents. *ISS Source*. 11 April. https://www.isssource.com/stuxnet-loaded-by-iran-double-agents/.

Sartonen, Miika, Aki-Mauri Huhtinen and Martti Lehto (2016) Rhizomatic target audiences of the cyber domain. *Journal of Information Warfare* 15(4): 1-13.

Schmitt, Michael N., ed. (2013) *Tallinn Manual on the International Law Applicable to Cyber Warfare*. New York: Cambridge University Press.

Sharif, Mahmood, Sritu Bhagavatula, Lujo Bauer and Michael K. Reiter (2016) Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 1528-1540.

Sharpe, Richard R. (2006) Bluffing with a pair of deuces: the downside of successful deception. Master's thesis. Monterey, CA: Naval Postgraduate School.

Sitawarin, Chawin, Arjun Nitin Bhagoji, Arsalan Mosenia et al. (2018) Rogue signs: Deceiving traffic sign recognition with malicious ads and logos. https://doi.org/10.48550/arXiv.1801.02780.

Spiliotopoulos, Dimitris, Dionisi Margaris and Costas Vassilakis (2020) Data-assisted persona construction using social media data. *Big Data and Cognitive Computing* 4(3): 21-35.

Stech, Frank, Kristin Heckman, Phil Hilliard and Janice Ballo (2011) Scientometrics of deception, counter-deception, and deception detection in cyber-space. *PsychNology Journal* 9(2): 79-122.

Stewart, Phil (2022) Russia moves blood supplies near Ukraine, adding to U.S. concern, officials say. *Reuters*. 29 January. https://www.reuters.com/world/europe/exclusive-russia-moves-blood-supplies-near-ukraine-adding-us-concern-officials-2022-01-28/.

Timberg, Craig (2017) Russian propaganda may have been shared hundreds of millions of times, new research says. *Washington Pos*t. 5 October.

Tucciarelli, Raffaele, Neza Vehar, Shamil Chandaria and Manos Tsakiris (2022) On the realness of people who do not exist: the social processing of artificial faces. *iScience* 25(12): 105441.

Tzu, Sun (2002) *Sun Tzu on The Art of War*. Trans. Lionel Giles. London: Routledge.

US Army (1988) *Field Manual FM90-2: Battlefield Deception*. https://irp.fas.org/doddir/army/fm90-2/toc.htm.

U.S. Cybersecurity and Infrastructure Security Agency (2020) People's Republic of China state-sponsored cyber actor living off the land to evade detection. 24 May. https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-144a.

U.S. Joint Staff (1996) *Joint Pub 3-58. Joint Doctrine for Military Deception*. https://irp.fas.org/doddir/dod/jp3_58.pdf

United States Senate Select Committee on Intelligence (2020) *Russian Active Measures, Campaigns and Interference in the 2016 U.S. Election*, Vols. 1-5. https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures.

van der Walt, E., J.H.P. Eloff and Jacomine Grobler (2018) Cyber-security: Identity deception detection on social media platforms. *Computers and Security* 78: 76-89.

Vincent, James (2020) An online propaganda campaign used AI-generated headshots to create fake journalists. *The Verge*. 7 July. https://www.theverge.com/2020/7/7/21315861/ai-generated-headshots-profile-pictures-fake-journalists-daily-beast-investigation.

Wavell, Archibald P. (1948) *The Good Soldier*. London: Macmillan.

Wavell, Archibald P. (1946) *Speaking Generally: Broadcasts, orders and addresses in time of war (1939-43).* London: Macmillan.

Weber, Joscha, Andrea Grunau, Matthias von Hein and Eugen Theise (2022). Fact check: Do Vladimir Putin's justifications for going to war against Ukraine add up? *Deutsche Welle.* 25 February. https://www.dw.com/en/fact-check-do-vladimir-putins-justifications-for-going-to-war-against-ukraine-add-up/a-60917168.

Whaley, Barton (1982) Towards a General Theory of Deception. In John Gooch and Amos Perlmutter, eds. *Military Deception and Strategic Surprise*. New York: Frank Cass, 178-192.

Whaley, Barton (2006) *Detecting Deception: A Bibliography of Counterdeception Across Time, Cultures, and Disciplines*. Second edn. Washington, DC: US Foreign Denial and Deception Committee.

Whaley, Barton (2007) *Stratagem: Deception And Surprise In War*. Norwood, MA: Artech House.

Whaley, Barton (2016) *Practise to Deceive: Learning Curves of Military Deception Planners*. Annapolis, MD: Naval Institute Press.

Wu, Zuxuan, Ser-NamLim, Larry S. Davis and Tom Goldstein (2020) Making an invisibility cloak: Real world adversarial attacks on object detectors. In Andrea Veldadi, Horst Bischof, Thomas Brox and Jan-Michael Frahm, eds. *Computer Vision: ECV 2020.* Springer, Cham. Springer, 1-17.

Zhang, Jerry, Darrell Carpenter and Myung Ko (2013) Online astroturfing: a theoretical perspective. *Proceedings of the Nineteenth Americas Conference on Information Systems*, 1-7.