Norms and Strategies for Stability in Cyberspace

Mariarosaria Taddeo
Oxford Internet Institute, University of Oxford
Alan Turing Institute
mariarosaria.taddeo@oii.ox.ac.uk

Abstract

Cyber attacks are escalating in frequency, impact, and sophistication. For this reason, it is crucial to identify and define *regulations* for state behaviour and *strategies* to deploy countering measures that would avoid escalation and disproportionate use of cyber means, while protecting and fostering the stability of our societies. To this end, strategies to deter cyber attacks and norms regulating state behaviour in cyberspace are both necessary; unfortunately neither is available at the moment. In this chapter, I offer a theory of cyber deterrence and a set of policy recommendations to fill this vacuum.

Key words: Artificial Intelligence; Cyber Attacks; Cyber Conflicts; Deterrence; Ethics of AI; Regulation; Stability; State.

1. Introduction

Cyber attacks are becoming more frequent and impactful. Each day in 2017, the United States suffered, on average, more than 4,000 ransomware attacks, which encrypt computer files until the owner pays to release them. In 2015, the daily average was just 1,000. In May 2017, when the WannaCry virus crippled hundreds of IT systems across the UK National Health Service, more than 19,000 appointments were cancelled. A month later, the NotPetya ransomware cost pharmaceutical giant Merck, shipping firm Maersk, and logistics company FedEx around US$300 million each. Estimates show that global damages from cyber attacks may reach $6 trillion a year by 2021 (Mariarosaria Taddeo, McCutcheon, and Floridi 2019).[1]

The fast-pace escalation of cyber attacks occurred during the past decade has prompted a mounting concern about international stability and the security of our

[1] https://cybersecurityventures.com/2015-wp/wp-content/uploads/2017/10/2017-Cybercrime-Report.pdf

societies. To address this concern, in April 2017, the foreign ministers of the G7 countries approved a 'Declaration on Responsible States Behaviour in Cyberspace' (G7 Declaration 2017). In the opening statement, the G7 ministers stress their concern

> […] about the risk of escalation and retaliation in cyberspace […]. Such activities could have a destabilizing effect on international peace and security. We stress that the risk of interstate conflict as a result of ICT incidents has emerged as a pressing issue for consideration. […], (G7 Declaration 2017, 1).

Paradoxically, state actors often play a central role in the escalation of cyber attacks. State-run cyber attacks have been launched for espionage and sabotage purposes since 2003. Well-known examples include Titan Rain (2003), the Russian attack against Estonia (2006) and Georgia (2008), Red October targeting mostly Russia and Eastern European Countries (2007), Stuxnet and Operation Olympic Game against Iran (2006-2012). In 2016, a new wave of state-run (or state-sponsored) cyber attacks ranged from the Russian attack against Ukraine power plant,[2] to the Chinese and Russian infiltrations US Federal Offices,[3] to the Shamoon/Greenbag attacks on government infrastructures in Saudi Arabia.[4] WannaCry has been attributed to North Korea and NotPetya to Russia in 21017. Russia has also been linked to a series of cyber attacks targeting US critical national infrastructures disclosed in 2018.

This trend will continue. The relatively low entry-cost and the high chances of success mean that states will keep developing, relying on, and deploying cyber attacks. At the same time, the Artificial Intelligence (AI) leap of cyber capabilities—the use of AI technologies for cyber offence and defence—indicates that cyber attacks will escalate in frequency, impact, and sophistication (Mariarosaria Taddeo and Floridi 2018a; King et al. 2019).

Cyber attacks contribute to shape political relations, national, and international equilibria of our societies and are becoming a structural element of their power dynamics. For this reason, it is crucial to identify and define *regulations* for state behaviour and *strategies* to deploy countering measures that would avoid escalation and disproportionate use of cyber means, while protecting and fostering the stability of our societies.

Regulations and strategies will only be effective insofar as they will rest on a deep understanding of the nature of these attacks, of their differences from violent (kinetic)

[2] https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/
[3] https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html?_r=0
[4] https://www.symantec.com/connect/blogs/greenbug-cyberespionage-group-targeting-middle-east-possible-links-shamoon

ones, as well as on a clear understanding of the moral principles that should shape state behaviour in cyberspace. In the first part of this chapter, I will analyse existing approaches to the regulation of state behaviour in cyberspace and to the specification of deterrence strategies as countering strategies. This analysis will provide the groundwork for the theory of cyber deterrence and for the policy recommendations that I offer in the second part of the chapter.

2. Analogies and Regulation

Efforts to regulate state-run (or sponsored) cyber attacks—and cyber conflicts understood as attack-and-response dynamics—rose to prominence almost a decade ago, when the risks for national and international security and stability arising from the cyber domain became clear.[5] As I argued elsewhere (Taddeo 2014), these efforts often rely on an *analogy-based approach*, according to which the regulatory problems concerning cyber attacks are only apparent, insofar as these are not radically different from other kinetic of attacks. Those endorsing this approach claim that the existing legal framework governing inter-state, kinetic attacks is sufficient to regulate cyber attacks, and by extension cyber conflicts. All that is needed is an in-depth analysis of such laws and an adequate interpretation of the phenomena, as there is

> "a thick web of international law norms suffuses cyber-space. These norms both outlaw many malevolent cyber-operations and allow states to mount robust responses" (Schmitt 2013, 177).

According to this view, interpretations often highlight that existing norms raise substantial barriers to the use of cyber weapons and to the use of force to defend cyberspace; and international law contains coercive means of permitting lawful responses to cyber provocations and threats of any kind. The legal framework that is referred to encompasses the four Geneva Conventions and their first two Additional Protocols, the international customary law and general principle of law, the Convention restricting or prohibiting the use of certain conventional weapons, and judicial decisions. Arms control treaties, such as the Nuclear Non-Proliferation Treaty and the Chemical Weapons Convention, are often mentioned as providing guidance for action in the case of kinetic cyber attacks (Schmitt 2013). At the same time, coercive measures addressing economic violations are generally

[5] http://www.nato.int/docu/review/2013/cyber/timeline/EN/index.htm

considered legitimate in the case of cyber attacks that do not cause physical damage (Lin 2012; O'Connell 2012).

Others maintain that the problem at stake is not whether cyber attacks and cyber conflicts can be interpreted in such a way as to fit the parameters of kinetic conflicts, economic transgressions, and conventional warfare, and hence whether they fall within the domain of international humanitarian law, as we know it. The problem rests at a deeper level and questions the very normative and conceptual framework of international humanitarian law and its ability to address *satisfactorily* and *fairly* the changes prompted by cyber conflicts (Dipert 2010; Floridi and Taddeo 2014; Taddeo 2014a).

Consider for example inter-state cyber conflicts. Regulation of these conflicts need to be developed consistently to (a) Just War Theory, (b) human rights, and (c) international humanitarian laws. However, applying (a)-(c) to the case of cyber conflicts proves to be problematic given the changes in military affairs that they prompted (Dipert 2010; Taddeo 2012a; Floridi and Taddeo 2014). When compared to kinetic ones, cyber conflicts show fundamental differences: their domain ranges from the virtual to the physical; the nature of their actors and targets involves artificial and virtual entities alongside human beings and physical objects; and their level of violence may range from non-violent to potentially highly violent phenomena. These differences are redefining our understanding of key concepts such as harm, violence, target, combatants, weapons, and attack, and pose serious challenges to any attempt to regulate conflicts in cyberspace (Dipert 2010; Taddeo 2012b; Taddeo 2014a; Floridi and Taddeo 2014; Taddeo 2014b).

Things are not less problematic when considering ethical issues. Cyber conflicts bring about three sets of problems, concerning risks, rights, and responsibilities (3R problems) (Taddeo 2012). The more contemporary societies are dependent on digital technologies, the more the 3R problems become pressing and undermine ethically blind attempts to regulate cyber conflicts. Consider the risks of escalation. Estimates indicate that the cyber security market will grow from US$106 billion in 2015 to US$170 billion by 2020, posing the risk of a progressive weaponization and militarisation of cyberspace (Taddeo and Floridi 2018). At the same time, the reliance on malware for state-run cyber operations (like Titan Rain, Red October, and Stuxnet) risks sparking a cyber arms race and competition for digital supremacy, hence increasing the possibility of escalation and conflicts (MarketsandMarkets 2015). Regulations of cyber conflicts need to address and reduce this risk by encompassing principles to foster cyber stability, trust, and transparency among states (Arquilla and Borer 2007; Steinhoff 2007; European Union 2015; Taddeo Forthcoming).

At the same time, cyber threats are pervasive. They can target, but can also be launched through, civilian infrastructures, e.g. civilian computers and websites. This may (and in some cases already has) initiate policies of higher levels of control, enforced by governments in order to detect and deter possible threats. In these circumstances, individual rights, such as privacy and anonymity may come under sharp, devaluating pressure (Arquilla 1999; Denning 2007; Taddeo 2013).

Ascribing responsibilities also prove to be problematic when considering cyber attacks. Cyberspace affords a certain level of anonymity, often exploited by states or state-sponsored groups and non-state actors. Difficulties in attributing attacks allow perpetrators to deny responsibility, and pose an escalatory risk in cases of erroneous attribution. The international community faced this risk in 2014, when malware initially assessed as capable of destroying the content of the entire stock exchange was discovered on Nasdaq's central servers and allegations were made of a Russian origin for the software.[6]

In the medium- and long-term, regulations need to be defined so to ensure security and stability of societies, and avoid risks of escalation. To achieve this end, efforts to regulate state-run cyber attacks will have to rely on an in-depth understanding of this new phenomenon; identify the changes brought about by cyber warfare and the information revolution (Floridi 2014; Taddeo and Buchanan 2015; Floridi and Taddeo 2016); and define a set of shared values that will guide the different actors operating in the international arena. The alternative is developing unsatisfactory, short-sighted approaches and facing the risk of a cyber backlash: a deceleration of the digitization process imposed by governments and international institutions to prevent this kind of conflicts to erode both the trust in economy and in political institutions. For this reason, it is necessary to seize the limits of the analogy-based approach, and to move past it. As Betz and Stevens (Betz and Stevens 2013) put it:

> "It is little wonder that we attempt to classify […] the unfamiliar present and unknowable future in terms of a more familiar past, but we should remain mindful of the limitations of analogical reasoning in cyber security".

Analogies can be powerful, for they inform the way in which we think and constrain ideas and reasoning within a conceptual space (Wittgenstein 2009). However, if the conceptual space is not the right one, analogies become misleading and detrimental for any attempt to

[6] http://arstechnica.com/security/2014/07/how-elite-hackers-almost-stole-the-nasdaq/

develop innovative and in-depth understanding of new phenomena, and they should be abandoned altogether. When the conceptual space is the right one, analogies are at best a step on Wittgenstein's ladder and need to be disregarded once they have taken us to the next level of the analysis. This is the case of the analogies between kinetic and cyber conflicts.

Cyberspace and cyber conflicts are now *relatively* new phenomena. Over the past two decades, possible uses, misuses, risks, and affordances of both have become clearer. As societies, we now know the successes, the failures, and the lessons learned necessary to start analysing and understanding the nature of cyberspace and cyber conflicts and to regulate appropriately both the environment and the actions in it to avoid risks of escalation and instability.

3. The Strategic Nature of Cyberspace

Escalation follows from the nature of cyber attacks and the dynamics of cyberspace (Floridi and Taddeo 2014; Taddeo 2014a, 2016, 2017). Non-kinetic cyber attacks—aggressive uses of information and communications technologies that do not cause destruction or casualties, e.g. deploy zero-day exploits or DDoS attacks—cost little in terms of resources and risks to the attackers, while having high chances to be successful. At the same time, cyber defence is porous by its own nature (Morgan 2012): every system has bugs in the program (vulnerabilities), identifying and exploiting them is just a matter of time, means, and determination. This makes even the most sophisticate cyber defence mechanisms ephemeral and, thus, limits their potential to deter new attacks.

Even when successful, cyber defence does not lead to strategic advantages, insofar as dismounting a cyber attack, may bring tactical success, but very rarely leads to the ultimate defeating of an adversary (Taddeo 2017). This creates an environment of *persistent offence* (Harknett and Goldman 2016), where attacking is tactically and strategically more advantageous than defending. As Haknett and Goldman argue, in an offence-persistent environment, defence can achieve tactical and operational success in the short term if it can adjust constantly to the means of attack, but it cannot win strategically. Offence will persist and the interactions with the enemy will remain constant. This is why inter-state cyber defence have shifted from reactive (defending) towards an *active* (countering) defence strategies.

In this scenario, state actors make policy decisions to protect their abilities to launch cyber attacks. *Strategic ambiguity* is one of these decisions. According to this policy, states decide neither to define and nor inform the international community about their *red lines*—thresholds that once crossed would trigger state response—for non-kinetic cyber attacks (Mariarosaria Taddeo 2011). This approach leaves *de facto* unregulated cyber attacks that remain below the threshold of an armed attack.

Strategic ambiguity has often been presented as a way to confuse the opponents about the consequences of their cyber attacks. As the US National Intelligence Officer for Cyber Issues officer put it:

> Currently most countries, including ours, don't want to be incredibly specific about the red lines for two reasons: You don't want to invite people to do anything they want below that red line thinking they'll be able to do it with impunity, and secondly, you don't want to back yourself into a strategic corner where you have to respond if they do something above that red line or else lose credibility in a geopolitical sense.[7]

However, by fostering ambiguity, state actors also leave open for themselves a wider room for manoeuvring. Strategic ambiguity allows state actors to deploy cyber attacks for military, espionage, sabotage, and surveillance purposes without being constrained by their own policies or international red lines. This makes ambiguity a dangerous choice, one that is strategically risky and politically misleading.

The risks come with the cascade effect following the absence of clear thresholds for cyber attacks. The lack of thresholds facilitates a proliferation of offensive strategies. This, in turn, favours an international cyber arms race and the weaponization of cyberspace, which ultimately spurs the escalation of cyber attacks. This is why strategic ambiguity is a policy hazard that fuels, rather than arrest, escalation of interstate cyber attacks. Cyber attacks would be deterred more effectively by a regime of international norms that makes attacks politically costly to the point of being disadvantageous for the state actors who launch them.

As I mention in section 1, stability of cyberspace hinges on both regulations and strategies. Having considered the limits of the existing approaches to the regulation of state behaviour in cyberspace, I shall now focus on existing view for the designing deterrence strategies for cyber attacks.

4. Conventional Deterrence Theory

Concerned by the risks of escalation, international organisations such as NATO, the UN Institute for Disarmament Research (UNIDIR), and national governments, like the UK

[7] http://www.c4isrnet.com/articles/cyber-red-lines-ambiguous-by-necessity

and US have started to consider whether, and how to, deploy deterrence to foster stability of cyberspace.

However, deploying cyber deterrence strategies is challenging. For conventional deterrence theory (hereafter: deterrence theory) does not work in cyberspace, as it does not address the global reach, anonymity, the distributed, and interconnected nature of this domain. Deterrence theory has three core elements: attribution of attacks; defence and retaliation as types of deterring strategies; and the capability of the defender to signal credible threats (see Figure 1). None of these elements is attainable in cyberspace.

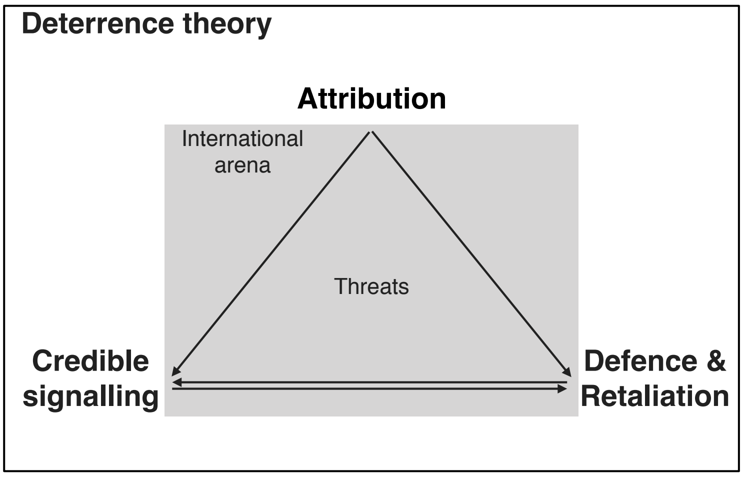

Figure 1. The core elements of deterrence theory and their dependences. This figure was published in M. Taddeo. "The Limits of Deterrence Theory in Cyberspace." *Philosophy & Technology*, 2017.

Consider attribution first. Prompt, positive attribution is crucial to deterrence: the less immediate is attribution, the less severe will be the defender's response. The less positive the attribution, the more time will be necessary to respond. In cyberspace, attribution is at best problematic, if not impossible. Cyber attacks are often launched in different stages and involve globally distributed networks of machines, as well as pieces of code that combine different elements provided (or stolen) by a number of actors. In this scenario, identifying the malware, the network of infected machines, or even the country of origin of the attack is not sufficient for attribution, as attackers can design and route their operations through third-party machines and countries with the goal of obscuring or misdirecting attribution. The limits of attribution in cyberspace pose serious obstacles to the deployment of effective deterrence. Recalling Figure 1, without attribution defence and retaliation, as well as signalling, are left without a target and are undermined by the inability of the defender to identify the attacker.

Signalling credible threats is also problematic in cyberspace. This element hinges on state's reputation. In kinetic scenarios, reputation is gained by showcasing military capabilities and by showing ability to resolve (to deter or defeat the opponent) over time. To some extent, the same also holds true in cyberspace, where a state's reputation also refers to a state's past interactions in this domain, its known cyber capabilities to defend and offend, as well as its overall reputation in resolving conflicts. However, state's reputation in cyberspace may not necessarily correspond to actual capabilities in this domain, as states are reluctant to circulate information about the attacks that they receive, especially those that they could not avert. This makes signalling less credible and, thus, more problematic than in other domains of warfare.

Also conventional deterrence strategies, defence and retaliation, are problematic in cyberspace. Every system has its security vulnerabilities and identifying and exploiting them is simply a matter of time, means, and determination. This makes vulnerable even the most sophisticated defence mechanisms, thus limiting their potential to deter new attacks by defence. Unlikely deterrence by defence, deterrence by retaliation may be effective in cyberspace. However, this strategy is coupled with serious risk of escalation. This is because the means to retaliate, i.e. cyber weapons, are *malleable* and difficult to control. Cyber weapons can be accessed, stored, combined, repurposed, and redeployed much more easily than it was ever possible with other kinds of military capability. This was the case for example of Stuxnet. Despite being designed to target specific configuration requirements of Siemens software installed on Iranian nuclear centrifuges, the worm was eventually released on the Internet and infected systems in Azerbaijan, Indonesia, India, Pakistan, and the US.

Clearly, classic deterrence theory faces severe limitations when applied in cyberspace. But it would be a mistake to conclude that as classic deterrence theory does not work in cyberspace, then deterrence is unattainable in this domain. As USN Commander Bebber stated:

> "History suggests that applying the wrong operational framework to an emerging strategic environment is a recipe for failure. During the World War I, both sides failed to realize that large scale artillery barrages followed by massed infantry assaults were hopeless on a battlefield that strongly favored well-entrenched defense supported by machine gun technology. […] The failure to adapt had disastrous consequences".[8]

[8] https://www.thecipherbrief.com/column_article/no-thing-cyber-deterrence-please-stop

We need to adapt. And adapting will be successful only if it rests on an in-depth understanding of cyberspace, cyber conflicts, their nature, and their dynamics. This understanding will allow us to forge a new theory of deterrence, one able to address the specificities of cyberspace and cyber conflicts. The alternative—developing cyber deterrence in analogy with conventional deterrence—is recipe for failure. It is equivalent to force the proverbial square peg in the round whole, we are more likely to smash the toy than to win the game.

5. Cyber Deterrence Theory

Cyber attacks and defence evolve with digital technology. As the latter becomes more autonomous and smart, leveraging the potential of AI, so do cyber attacks and cyber defence strategies. Both the public and private sectors are already testing AI systems in autonomous war games. The 2016, DARPA Cyber Grand Challenge was a landmark in this respect. The Challenge was the first, fully autonomous competition in which AI capabilities for defence were successfully tested. Seven AI systems, developed by teams from the United States and Switzerland, fought against each other to identify and patch their own vulnerabilities, while probing and exploiting those of other systems. The Challenge showed that AI will have a major impact on the waging of cyber conflicts, it will provide new capabilities for defence, shape new strategies, but also pose new risks. The latters are of particular concern. The autonomy AI systems, their capacity to improve their own strategies and launch increasingly aggressive counter-attacks with each iteration may lead to proportionality breaches and escalation of responses, which could, in turn, trigger kinetic conflicts. In this scenario, cyber deterrence is ever more necessary.

Elsewhere (Taddeo 2018) I argued that cyber deterrence rests on three core elements: target identification, retaliation, and demonstration (Figure 2).

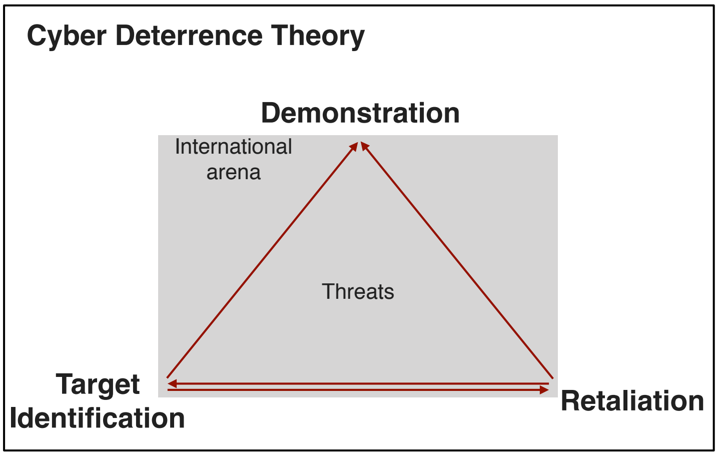

Figure 2. The three elements of Cyber Deterrence Theory and their dependencies. "How to Deter in Cyberspace", (Taddeo 2018).

Target identification is essential for deterrence. It allows the defendant to isolate (and counter-attack) enemy systems independently from the identification of the actors behind them, thus side-stepping the attribution problem, while identifying a justifiable target for retaliation. Identifying the attacking system and retaliate is feasible task, one which AI systems for defence can already achieve. As shown in Figure 2, cyber deterrence does not encompass defence among its possible strategies. This is due to the offence persistent nature of cyberspace, which makes retaliation more effective than defence both tactically and strategically.

Cyber deterrence uses target identification and retaliation for demonstrative purposes. According to this theory, deterrence in cyberspace works if it can demonstrate the defendant's capability to retaliate a current attack by harming the source system. While not being able to deter an incoming cyber attack, retaliation will deter the *next* round of attacks coming from the same opponent. This is because the mere threat of retaliation will not be sufficient to change the opponent's intentions to attacks. The chances of success and the likelihood that the attack will remain unattributed remain too high for any proportionate threat to be effective. Thus, to be successful, cyber deterrence need to shift from threatening to prevailing.

6. A Regime of Norms

Cyber deterrence alone is not a cure for all problems. Indeed, it is insufficient to ensure stability of cyberspace. This is true especially when considering how the rising distribution and automation, multiple interactions, and fast-pace performance of cyber attacks make control progressively less effective, while increasing the risks for unforeseen consequences, proportionality breaches, and escalation of responses (Yang et al. 2018). An international regime of norms regulating state behaviour in cyberspace is necessary to complement cyber deterrence strategies and foster stability.

Over the past twenty years, the UN, the Organisation for Cyber Security and Co-operation in Europe (OSCE), and the ASEAN Regional Forum (ARF) and several national governments (G7 and G20) have convened consensus to define such a regime. The G7 Declaration is the latest of a series of successful transnational initiatives made in this direction before the failure of the UN Group of Government Experts (UN GGE) on

‘Developments in the field of information and telecommunications in the context of international security’.[9]

The G7 Declaration identifies two main instruments: confidence building measures (CBMs) and voluntary norms. CBMs foster trust and transparency among states. In doing so, they favour co-operations and measures to limit the risk of escalation. CBMs range from establishing contact points, shared definitions of cyber-related phenomena, and communication channels to reduce the risk of misperception, and foster multi-stakeholder approach.

Voluntary norms identify non-binding principles that shape state conduct in cyberspace. *De facto*, voluntary norms identify red lines for state-run, non-kinetic cyber attacks and, thus, fill the void created by strategic ambiguity. They stress that states should not target critical infrastructures and critical information infrastructures of the opponent (norms 6, 8, and 11 of the G7 Declaration); should avoid using cyber attacks to violate intellectual property (norm 12 of the G7 Declaration); and remark the responsibility of state actors to disclose cyber vulnerabilities (norms 9 and 10 of the G7 Declaration).

CBMs and (in part) voluntary norms have been then included in the 2017 cyber security framework launched by the European Commission. The framework is one of the most comprehensive regulatory frameworks for state conduct in cyberspace so far. Yet it does not go far enough. The EU treats cyber defence as a case of cybersecurity, to be improved passively by making member states’ information systems more resilient. It disregards active uses of cyber defence and does not include AI.

This was a missed opportunity. The EU could have begun defining red lines and proportionate responses in its latest rethink. For example, the 2016 EU directive on ‘Security of Network and Information Systems’ provides criteria for identifying crucial national infrastructures, such as health systems or key energy and water supplies that should be protected. The same criteria could be used to define illegitimate targets of state-sponsored cyber attacks.

The EU cyber security framework remains a step in the right direction, but more work needs to be done. After the failure of the UN GGE, it is crucial that discussion on the regulation of state behaviour resume. Regional forums, such as NATO and the EU, may be a good starting point for more fruitful discussions. When considering state-run cyber defence, it is crucial that the following three steps are taken into consideration to

[9] https://www.justsecurity.org/42768/international-cyber-law-politicized-gges-failure-advance-cyber-norms/

avoid serious imminent attacks on state infrastructures, and to maintain international stability. These are:

- **Define 'red lines'** distinguishing legitimate and illegitimate targets and definitions of proportionate responses for cyber defence strategies.
- **Building alliances** by mandating 'sparring' exercises between allies to test AI-based defence capabilities and the disclosure of fatal vulnerabilities of key systems and crucial infrastructures among allies.
- **Monitor and enforce rules at international level** by defining procedures to audit and oversee AI-based state cyber defence operations, alerting and remedy mechanisms to address mistakes and unintended consequences. A third-party authority with teeth, such as the UN Security Council, should rule on whether red lines, proportionality, responsible deployment or disclosure norms have been breached.

7. Conclusions

"Those who live by the digit may die by the digit" (Floridi 2014a). Indeed, if the threats coming or targeting cyberspace pose serious risks to the stability and security of our societies is because we live in societies that are increasingly more dependent on digital technologies. As Ericcson and Giacomiello put it:

> "In 1962, Arnold Wolfers wrote that national security is the absence of threat to a society's core values. If modern, economically developed countries are increasingly becoming information societies, then, following Wolfers' argument, threats to information can be seen as threats to the core of these societies", (Eriksson and Giacomello 2006, 222).

A relation of mutual influence exists between the way conflicts are waged and the societies waging them. As Clausewitz remarked, more than an art or a science, conflicts are a social activity. And much like other social activities, conflicts mirror the values of societies while relying on their technological and scientific developments. In turn, the principles endorsed to regulate conflicts play a crucial role in shaping societies.

Think about the design, deployment, and regulation of weapons of mass destruction (WMDs). During World War II, WMDs were made possible by scientific breakthroughs in nuclear physics, which was a central area of research in the years leading to the War. Yet, their deployment proved to be destructive and violent beyond what the post-war world was willing to accept. The Cold War that followed, and the nuclear treaties that ended it, defined the modes in which nuclear technologies and WMDs could be used, drawing a line between conflicts and atrocities. In doing so, treaties and regulations for the use of WMDs contributed to shape contemporary societies as societies rejecting the belligerent rhetoric of the early twentieth century and to striving for peace and stability.

The same mutual relation exists between information societies and cyber conflicts, making the regulation of the latter a crucial aspect, which does and will contribute to shape current and future societies. In the short term, regulations are needed to avoid a digital wild west, as remarked by Harold Hongju Koh, the former Legal Advisor U.S. Department of State. In the long term, regulations are needed to ensure that cyber conflicts will not threat the development of open, pluralistic, and tolerant information societies (Taddeo and Floridi 2018b).

The only way to ensure this outcome is to develop new domain-specific, conceptual, normative, and strategic framework. Analogies with kinetic conflicts, strategies to deter them, and existing normative frameworks should be abandoned altogether, as they are misleading and detrimental for any attempt to develop innovative and in-depth understanding of cyberspace, cyber conflicts, deterrence, and ensure stability. The effort is complex, but also necessary.

References

Arquilla. 1999. 'Ethics and Information Warfare'. In *Strategic Appraisal: The Changing Role of Information in Warfare*, edited by Zalmay Khalilzad and John Patrick White, 379–401. Santa Monica, CA: RAND.

Arquilla, J., and Douglas A Borer. 2007. *Information Strategy and Warfare : A Guide to Theory and Practice*. New York: Routledge.

Betz, D. J., and T. Stevens. 2013. 'Analogical Reasoning and Cyber Security'. *Security Dialogue* 44 (2): 147–64. https://doi.org/10.1177/0967010613478323.

Denning. 2007. 'The Ethics of Cyber Conflict'. In *Information and Computer Ethics*. Hoboken, USA: Wiley.

Dipert, R. 2010. 'The Ethics of Cyberwarfare'. *Journal of Military Ethics* 9 (4): 384–410.

Eriksson, Johan, and Giampiero Giacomello. 2006. 'The Information Revolution, Security, and International Relations: (IR)Relevant Theory?' *International Political Science Review* 27 (3): 221–44. https://doi.org/10.1177/0192512106064462.

European Union. 2015. 'Cyber Diplomacy: Confidence-Building Measures - Think Tank'. Brussels. http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2015)571302.

Floridi, L. 2014. *The Fourth Revolution, How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press.

Floridi, L., and M. Taddeo, eds. 2014. *The Ethics of Information Warfare*. New York: Springer.

Floridi, Luciano, and Mariarosaria Taddeo, eds. 2014. *The Ethics of Information Warfare*. Law, Governance and Technology Series, volume 14. Heidelberg: Springer.

———. 2016. 'What Is Data Ethics?' *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2083): 20160360. https://doi.org/10.1098/rsta.2016.0360.

G7 Declaration. 2017. 'G7 Declaration on Responsible State Behavior in Cyberspace'. Lucca. http://www.mofa.go.jp/files/000246367.pdf.

Harknett, Richard, J., and Emily Goldman O. 2016. 'The Search for Cyber Fundamental'. *Journal of Information Warfare* 15 (2): 81–88.

King, Thomas C., Nikita Aggarwal, Mariarosaria Taddeo, and Luciano Floridi. 2019. 'Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions'. *Science and Engineering Ethics*, February. https://doi.org/10.1007/s11948-018-00081-0.

Lin, Herbert. 2012. 'Cyber Conflict and International Humanitarian Law'. *International Review of the Red Cross* 94 (886): 515–31. https://doi.org/10.1017/S1816383112000811.

MarketsandMarkets. 2015. 'Cyber Security Market by Solutions & Services - 2020'. http://www.marketsandmarkets.com/Market-Reports/cyber-security-market-505.html?gclid=CNb6w7mt8MgCFQoEwwodZVQD-g.

Morgan, Patrick M. 2012. 'The State of Deterrence in International Politics Today'. *Contemporary Security Policy* 33 (1): 85–107. https://doi.org/10.1080/13523260.2012.659589.

O'Connell, M. E. 2012. 'Cyber Security without Cyber War'. *Journal of Conflict and Security Law* 17 (2): 187–209. https://doi.org/10.1093/jcsl/krs017.

Schmitt, M. 2013. 'Cyberspace and International Law: The Penumbral Mist of Uncertainty'. *Harvard* 126 (176): 176–80.

Steinhoff, Uwe. 2007. *On the Ethics of War and Terrorism*. Oxford; New York: Oxford University Press.

Taddeo, M. 2012a. 'An Analysis for a Just Cyber Warfare'. In *2012 4th International Conference on Cyber Conflict (CYCON 2012)*, 1–10.

Taddeo, Mariarosaria. 2012b. 'Information Warfare: A Philosophical Perspective'. *Philosophy and Technology* 25 (1): 105–20.

Taddeo, Mariarosaria. 2011. 'Information Warfare: A Philosophical Perspective'. *Philosophy & Technology* 25 (1): 105–20. https://doi.org/10.1007/s13347-011-0040-9.

Taddeo, Mariarosaria. 2012. 'An Analysis For A Just Cyber Warfare'. In *Fourth International Conference of Cyber Conflict*. NATO CCD COE and IEEE Publication.

Taddeo, Mariarosaria. 2013. 'Cyber Security and Individual Rights, Striking the Right Balance'. *Philosophy & Technology* 26 (4): 353–56. https://doi.org/10.1007/s13347-013-0140-9.

Taddeo, Mariarosaria. 2014a. 'Just Information Warfare'. *Topoi*, April, 1–12. https://doi.org/10.1007/s11245-014-9245-8.

Taddeo, Mariarosaria. 2014b. 'The Struggle Between Liberties and Authorities in the Information Age'. *Science and Engineering Ethics*, September, 1–14. https://doi.org/10.1007/s11948-014-9586-0.

Taddeo, Mariarosaria. 2016. ‘On the Risks of Relying on Analogies to Understand Cyber Conflicts’. *Minds and Machines* 26 (4): 317–21. https://doi.org/10.1007/s11023-016-9408-z.

Taddeo, Mariarosaria. 2017. ‘Cyber Conflicts and Political Power in Information Societies’. *Minds and Machines* 27 (2): 265–68. https://doi.org/10.1007/s11023-017-9436-3.

Taddeo, Mariarosaria. 2018. ‘How to Deter in Cyberspace’. *The European Centre of Excellence for Countering Hybrid Threats* 2018 (6): 1–10.

Taddeo, Mariarosaria, and Elizabeth Buchanan. 2015. ‘Information Societies, Ethical Enquiries’. *Philosophy & Technology* 28 (1): 5–10. https://doi.org/10.1007/s13347-015-0193-z.

Taddeo, Mariarosaria, and Luciano Floridi. 2018a. ‘Regulate Artificial Intelligence to Avert Cyber Arms Race’. *Nature* 556 (7701): 296–98. https://doi.org/10.1038/d41586-018-04602-6.

Taddeo, Mariarosaria. 2018b. ‘How AI Can Be a Force for Good’: *Science* 361 (6404): 751–52. https://doi.org/10.1126/science.aat5991.

Taddeo, Mariarosaria, Tom McCutcheon, and Luciano Floridi. 2019. ‘Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword’. *Nature Machine Intelligence* 1 (12): 557–60. https://doi.org/10.1038/s42256-019-0109-1.

Wittgenstein, Ludwig. 2009. *Philosophical investigations*. Rev. 4th ed. Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell.

Yang, Guang-Zhong, Jim Bellingham, Pierre E. Dupont, Peer Fischer, Luciano Floridi, Robert Full, Neil Jacobstein, et al. 2018. ‘The Grand Challenges of *Science Robotics*’. *Science Robotics* 3 (14): eaar7650. https://doi.org/10.1126/scirobotics.aar7650.