Philos. Technol. (2018) 31:331–338
DOI 10.1007/s13347-016-0251-1

RESEARCH ARTICLE

Five Kinds of Cyber Deterrence

N. J. Ryan[1]

Received: 30 September 2016 / Accepted: 27 December 2016 / Published online: 27 January 2017

Abstract There were five kinds of cyber deterrence presented at the workshop on *Landscaping strategic cyber deterrence*, hosted at the Oxford Internet Institute. They were the well-studied areas of deterrence by 'punishment' and 'denial', and the novel concepts of deterrence by 'association', 'norms and taboos', and finally, 'entanglement'. In the following workshop commentary, I present these five kinds of deterrence and explain them in light of recent developments in the academy and industry. I argue for analytical congruence between all three novel concepts, since they aim to alter the behaviour of actors by adding a social cost in response to breaking norms and conventions. Throughout, I argue that we are beginning to understand how cyber deterrence works, both in theory and practice, and when all concepts are taken together, they become more than the sum of their parts. Finally, I point out an omission of the workshop, where computational modelling and simulation could be added to the landscape of strategic cyber deterrence.

Keywords Cyber deterrence · Denial · Punishment · Association · Norms · Taboos · Entanglement

'Does deterrence work in cyberspace?' is one of the most vexing questions constantly asked of cybersecurity analysts. The question was posed at the beginning of the workshop on *Landscaping strategic cyber deterrence*, hosted by the Oxford Internet Institute (OII), similar to how it opens the current commentary. Up until now, many well-informed analysts have been quick to admonish deterrence theory for being just that—theory—which is ineffective in practice. I claim the prevailing

✉ N. J. Ryan
nryan@rand.org

[1] Defence, Security and Infrastructure, RAND Europe, Cambridge CB4 1YG, UK

sentiment in the academy has changed, which is based on the presentations at the OII and the recent threat report in the cybersecurity industry.[1] These recent developments have led some scholars in the academy to think harder about how cyber deterrence works, evidenced by the addition of three novel concepts to the debate. I continue the persuasive arguments made by others, and I argue we are beginning to understand how cyber deterrence works both in theory and practice.

There were five kinds of cyber deterrence presented at the workshop. All of the concepts begin with 'deterrence by' (a predictable part of the nomenclature of deterrence theory) and the first two concepts are fitted with the well-studied nouns of 'punishment' and 'denial'. The other three concepts involve the novel terms of 'association', 'norms and taboos' and 'entanglement', which I come to argue are analytically congruent. Individually, each concept captures a kernel of truth about deterrence theory, although any one concept by itself may be insufficient to preclude the actions of an adversary. Yet, as a collective, I posit that cyber deterrence actually works, since the concepts taken together become more than the sum of their parts. The primary actor class I deal with in the commentary are states, although we should in principle be able to apply the concepts to other actors like firms, civil society and individuals. My aim will be to present and critique the ideas of the original authors'[2] and then enter a discussion about the substantive content of cyber deterrence theory. Finally, I will point out an omission at the workshop, which I believe is a necessary feature of cyber deterrence strategy, and that is the role of computer simulation and modelling of game-theoretic problems in cyber strategy.

1 Deterrence by Punishment

The first concept is deterrence by punishment and it takes the general form of 'if you do X to me, then I will do Y in response'. The point is that Y punishment is larger than the perceived value of X, so as to preclude the adversary from doing X in the first instance. If the threat of retaliation is credible in the mind of the adversary, such that their behaviour is modified, then deterrence is said to work. The punishment served must be orientated towards a valued asset, and retaliatory options can be anything from a prison sentence, to sanctions, or kinetic force, depending on the actors involved and other exogenous features.

Although it is straightforward to comprehend, deterrence by punishment is one of the more problematic concepts to translate into the domain of cyberspace. Meting out punishment in cyberspace, especially between adversarial states in the international system, is made particularly difficult given the attribution problem. Aggressive actors

[1] Here, I cite the decrease in the volume of computer network operations (CNOs) and industrial espionage between the USA and China, as reported by FireEye's iSight intelligence unit. While the report highlights the reduction in the overall volume of CNOs between the USA and China, the authors are quick to comment on the increasing sophistication of attacks that are 'focused, calculated, and still successful in compromising corporate networks'. FireEye iSight Intelligence, 'Red Line Drawn: China Recacluates Its Use of Cyber Espionage', (Milpitas, CA 2016).

[2] Throughout the course of the day, presentations were given by leading professors, lecturers and thinkers in the disciplines of the humanities, such as philosophy and law, as well as the social sciences in the areas of sociology, political science and subfields of military studies and international relations. I hope to paraphrase each concept accurately and I accept all misinterpretations as my own.

can choose to hide their identity and claim the defence of plausible deniability, shifting the onus of proof onto the victim to correctly identify the attacker. Recent scholarship on attribution finds the problem is getting easier over time, given advances in passive network monitoring and threat detection.[3] Moreover, there are social and political factors affecting the attribution of a cyber adversary.[4] Successful attribution requires great technical expertise, supported by organisational coordination and political will to hold the aggressor accountable. To be clear, attribution is not a necessary condition for deterrence by punishment. States can practice indiscriminate retaliation or excessive punishment to make examples of others and strike fear into adversaries, thus aiming to deter them. Such a strategy of indiscriminate punishment is likely to be condemned as immoral, given the violation of the ethical principles of discrimination and proportionality. An indiscriminate and atrocious act of cyberwarfare errs into scenario design impossibility, let alone the otherwise unlikely 'cyber-9/11' or 'digital Pearl Harbour' black swan events.

'Hacking back' is seen as one of the more controversial options available to states and businesses who wish to deter adversaries. A number of risks are associated with hacking back, including escalatory attacks, mistaken attribution and the charge of digital vigilantism without proper authority. Also there is the complication and expense of timely digital forensic attribution before hacking back can begin. Firms often look to the state to provide security from foreign threats. States have the ability to threaten the use of kinetic force as punishment to virtual attacks in a way that firms or individuals cannot. Cross-domain conflict—only mentioned briefly by participants—is the ability for a state to strike back in another domain, be it land, sea, air or space, in the event of a cyberattack.

Deterrence by punishment in cyberspace can work, especially if there is cooperation between nations' law enforcement agencies. When there is a strong cooperative agreement between national law enforcement agencies, it is easy to see why threats of punishment are compelling deterrents, because the victim is likely to be caught if they are not highly skilled and technologically competent. More opaque are the limits of punishment, and indeed deterrence, which are on display when looking at instances where antagonistic states prosecute foreign state hackers. The February 2013 *APT1* report by Mandiant displayed the height of the prosecutorial impotence, when they merely linked a sophisticated and persistent group of hackers in Shanghai to the People's Liberation Army (PLA) Unit 61398.[5] The inability of the USA to nudge China towards a mutually beneficial relationship—out of the rut of malfeasance—led to a tougher stance by the USA. In the first case of its kind, the FBI indicted five members of the PLA in May 2014 for computer hacking, economic espionage and other offenses relating to US victim companies in the nuclear power, metals and solar industries.[6] More recently, in January 2016, the FBI indicted seven foreign nationals with links to the Iranian Government for their role in DDoS attacks against US companies.[7] The law's delay is unsurprising in these two cases,

[3] Thomas Rid and Ben Buchanan, 'Attributing Cyber Attacks', *Journal of Strategic Studies* 38, no. 1–2 (2015).

[4] Jon R Lindsay, 'Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack', *Journal of Cybersecurity* (2015).

[5] Mandiant, 'Apt1: Exposing One of China's Cyber Espionage Units', (Alexandria, VA 2013).

[6] US Department of Justice, 'U.S. Charges Five Chinese Military Hackers with Cyber Espionage against U.S. Corporations and a Labor Organization for Commercial Advantage', (Office of Public Affairs, 2014).

[7] Federal Bureau of Investigation, 'Iranian Ddos Attacks', FBI.gov, https://www.fbi.gov/wanted/cyber/iranian-ddos-attacks.

since it is highly unlikely that the Chinese or Iranian Governments do not intend to comply with US authorities and extradite their nationals to face charges. The ability to deter states from attacking private firms requires more nuance than mere threats of punishment, cross-domain conflict or jail time because it involves an understanding of international relations and the games that states play with one another.

2 Deterrence by Denial

Denial has long been a favoured strategy in cyber deterrence, especially given the perception that punishment and attribution was thought to be incredibly challenging down on a technical and up on strategic level. The history of warfare is punctuated by advances in denial techniques. The textbook example follows the increasing fortification of castles, which were built with higher walls, deeper moats and thicker walls, to protect the most precious Crown Jewels and keep people safe inside. Adversaries were meant to perceive the castles as too tough and too well-defended to warrant attacking. Whether or not a hypothetical assault on the castle is successful is a matter of defence, not deterrence. If an attack was prevented and never took place due to the high level of protection afforded by the castle, a game-theoretic position would ascribe a higher deterrent value to the castle, over the perceived gains of attacking, minus the cost of attacking. In essence, deterrence by denial is achieved when an adversary’s cost/benefit calculus is affected by defensive measures, such that their psychological position is to abstain from attacking because it is simply ‘not worth it’.

Directly appropriating the concept to cyberspace is as simple as it is misleading: in theory, building stronger network defences will keep information assets more secure. Unfortunately, the analogy breaks down since computer networks are not relevantly similar to castles in all respects. Indeed, the current landscape is changing, as the OII workshop highlighted. In the recent past, it has been conventional to consider the cyber domain as offence-dominant; it is easier to attack than defend, or so the assumption goes. Given the advances in machine learning, artificial intelligence (AI) and deep packet inspection, when combined with applied mathematics, the confluence of technologies has altered how information and network security is practiced. It is no longer about having a strong perimeter defence. Instead, the aim is to detect malicious actors once they are inside a network and ‘ban’ them before they can steal the digital ‘Crown Jewels’, so to speak. A more sophisticated, AI-driven, deterrence by denial is now in practice. Defensive measures are also strengthened by efforts to deceive attackers. If hackers actively avoid these kinds of newly defended networks because they psychologically believe them to be too strong or impenetrable, then these hackers are said to be deterred by denial.

3 Deterrence by Association

Deterrence by association, a term coined by Paul Cornish, can be characterised as a political mechanism to modify the behaviour of an adversary by linking their malfeasant cyber activities with their ‘real’ identity, whether it be states or other actors.[8] By making it possible to 'name and shame' those exhibiting poor behaviour, it re-installs the ability to dish out punishment, if only as a social cost. In terms of game-theoretic outcomes combined with international relations, the idea of deterrence by association tips the cost calculus towards *détente* rather than *premier coup*. The concept of deterrence by association is premised on the claim that cyberspace is inherently ambiguous. At every layer of cyberspace—from the technical below, to the middle social layer, and politico-strategic on top—all actors have incomplete information over the whole system. It is difficult to know with confidence who, what, where, when, how and at what layer threats should be deterred in cyberspace. While the lack of technical attribution gives rise to ambiguity on the upper layers of cyberspace, the concept is not a rally for technical attempts to 'solve' ambiguity. Rather, the concept is a political mechanism in order to 'call-out' poor behaviour and strongly condemn such actions publicly, by those with the right authority, because it acts as a clear signal to others in the community of actors what is right and wrong behaviour.

Recalling the instances of criminal prosecution brought against Chinese and Iranian hackers, actual punishment of these crimes is secondary to the deterrent value given by the act of publicly reprimanding foreign states for their poor behaviour. In the two cases mentioned in the last section, the USA sent a clear and direct message to both Beijing and Tehran, and an indirect signal to the rest of the international community, that private firms should not be the target of state-conducted, state-sponsored, even state-permitted cyber espionage. It is clear from these cases that associating an adversary with their poor behaviour inflicts a social cost onto them. It can take many forms, whether it is a loss of credibility in the international community for states, reputational damage for firms, or being ostracised from a community of like actors, either in civil society or as a private citizen.

4 Deterrence by Norms and Taboos

The idea that norms and taboos in cyberspace can act as a deterrent was explicitly introduced at the workshop by Joseph Nye and was successively picked up by other participants. Norms are thought of as non-binding conventions or a standard of appropriate behaviour about how a class of actors should act. Taboos are similar to norms, although they have a negative, inverted connotation since they refer to the inappropriate ways of acting or cultural mores that are 'off-limits'. Norms emerge over time and when they provide order, stability and security, they are often codified into law. We have already considered the law's ability to deter; the ability of norms to deter ambiguous aggressors is more complex than law, given their consensual nature and voluntary adherence. Norms, by their definition, are unenforceable since they rely on social actors observing certain ways of behaving by their own volition.

The prime examples of deterrence by norms and taboo are the Sino-US agreement, as well as the response to the Office of Personnel Management (OPM) hack. The norm

[8] Paul Cornish, 'Arms Control Tomorrow: The Challenge of Nuclear Weapons in the Twenty-First Century', *America and a changed world: A question of leadership/ed. by Robin Niblett. Chatham House.-Chichester…: Wiley-Blackwell* (2010).

to target legitimate intelligence assets was expressed by the US Director of National Intelligence, James Clapper, when he exhibited professional admiration for the Chinese hack of OPM. Whereas the taboo was established when both the USA and China agreed to refrain from state-sponsored industrial espionage or intellectual property theft. The words, handshakes and photo opportunities firming this agreement were more than a shared moment between Presidents Xi Jinping and Barack Obama in the Rose Garden of the White House in September 2015. It was a watershed moment of digital *détente* between arguably the world's two largest cyberpowers.

The construction of norms and their eventual agreement was part of a larger process involving policymakers from around the globe and the United Nations Group of Governmental Experts (UN GGE). The June 2015 consensus report from the UN GGE firmed the norms, rules and principles of the responsible behaviour of states in cyberspace. The consensus report agreed states should not exploit the attribution problem, in conjunction with the fact that states ought to observe international law in cyberspace.[9] Washington and Beijing had the necessary precedent, if only non-binding norms, to agree on an explicitly cooperative relationship going into the future.

From the example above, norms can deter through a number of mechanisms. Breaking norms can have a social cost amongst a group of actors. Much like deterrence by association, the 'calling out' of poor behaviour inflicts a kind of social punishment on the normative transgressor. A loss of face amongst the international community is the likely outcome of such a naming and shaming exercise. More severe forms include formal sanctions, being ostracised from the international community and suffering a domestic backlash. Finally, the Sino-US agreement sets a strong precedent for other states, particularly for great- and middle-powers, who would not want to be seen going against the convention set by a more powerful ally or adversary. By this measure, norms set by superpowers have the most influence and social influence, and every subsequent observation only strengthens their deterrent value.

5 Deterrence by Entanglement

Scott Jasper and Thomas Mahnken have recently used the term 'deterrence by entanglement' to explain how embedded actors behave cooperatively due to their mutual interest in cyberspace.[10] 'Mutual interest' is defined as a common reliance on the internet that could unite otherwise hostile actors, who refrain from attacking because they rely on internet connectivity for financial gain, for instance. Entanglement is predicated on the tight-coupling that occurs between actors, predominately states, due to the effects of globalisation. Mutual interest is at the core of the concept; however, it expands to include other methods for encouraging restraint between actors, such as encouraging responsible state behaviour through norms and principles. Closely coupled clusters of states with economic, diplomatic and strategic relationships must calculate the extent to which potential aggressive behaviour in cyberspace could potentially affect other aspects of their relations. It bears similarities to

[9] United Nations General Assembly, 'Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security', (A/70/174 2015).

[10] Thomas G Mahnken and Scott Jasper, *Conflict and Cooperation in the Global Commons: A Comprehensive Approach for International Security* (Georgetown University Press, 2012).

the classic international relations concept of interdependence and trade as a disincentive for conflict. The speech delivered by President Obama emphasises the importance of trade relations and economic interdependencies between the USA and China, which was a contributing factor behind both countries coming to the Rose Garden agreement.[11] It can therefore be seen in practice that by adding more factors into the deterrence cost calculus—economic, political and diplomatic, for instance—then an adversary can be entangled, and it is 'too costly' for them to act uncooperatively, since they would have to suffer the consequences in other areas of their relations.

6 Discussion

I have taken each kind of cyber deterrence discussed at the OII workshop in turn. This approach could be read as misleading, since it implicitly claims each form of deterrence to be different from one another. Indeed they are descriptively unique; however, I argue they are not all analytically unique. I find a conceptual congruence between cyber deterrence by 'association', 'norms and taboos', and 'entanglement'. The commonality between all three is the reinforcement of deterrence by punishment. Each occurs in a slightly different way, but all seek to punish and curb behaviour by adding a social cost. Further discussion is required in cases where reputational costs are only quite minor and how to deter actors who do not value their own standing in the world. Although all the terms can be applied to many kinds of actors, they are almost exclusively reserved for state relations. Social cost in diplomacy can be served in many forms, from raising a demarche, to recalling an ambassador, to officially breaking relations between countries. Outside of international relations, social cost can be considered as reputation damage and its downstream secondary effects. Ultimately, I see the congruence as adherence to norms of behaviour in cyberspace—which are no longer 'emerging', but they are now 'here'—and its influence on encouraging and discouraging certain norms or conventions of behaviour.

The landscaping workshop overlooked a key component in the survey of the contemporary understanding of deterrence in cyberspace. The omission of recent work in computational modelling and simulation of deterrence theory is one such oversight.[12] Promising work has been conducted by Jon Lindsay, published in the *Journal of Cybersecurity*, where he used a simple deterrence model to explain why there are many low-value anonymous cyberattacks but few high-value ones.[13] By building the attribution problem into a game-theoretic model that costs deterrence against the value received from attacking, it showed the offence-defence balance is skewed towards the ease of defence for higher-value targets. It is good news for strategists, since deterrence can effectively protect high-value targets where the defender's resolve is high and the attribution of the attacker is likely. The novel finding is

[11] The White House, 'Remarks by President Obama and President Xi of the People's Republic of China in Joint Press Conference', (Office of the Press Secretary, 2015).

[12] There are other sematic iterations of deterrence not presented at the workshop or discussed in detail here in the commentary. 'Deterrence by deception' is one such concept, developed by Erik Gartzke and John Lindsay, in Erik Gartzke and Jon R Lindsay, 'Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace', *Security Studies* 24, no. 2 (2015).

[13] Lindsay, 'Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack'.

that low-level aggression should be tolerated because it is a small price to pay for credible deterrence against large-scale attacks.

There is a growing understanding of how cyber deterrence works, both in theory and practice. Whether or not any single kind of deterrence theory works alone—like hardening network defences, or threatening punishment, or simply agreeing to observe responsible norms—is difficult to say with certainty. When all five kinds of deterrence are taken together, they become more than the sum of their parts, to produce an effective instrument and policy lever. Credible deterrence changes the behaviour of otherwise antagonistic adversaries to conform to responsible ways of behaving. The outcome of credible deterrence has arguably been observed in practice, with the Chinese military no longer preparing the digital battlefield by conducting industrial espionage and instead focusing on legitimate intelligence targets. With more confidence than in the past, cybersecurity analysts can assert that we are beginning to understand how cyber deterrence works both in theory and in practice.

Acknowledgements I would like to thank Oxford Internet Institute for hosting the workshop and acknowledge RAND Europe for covering the costs of my attendance. I would like to thank Dr. Mariarosaria Taddeo for her invitation and encouragement to write the workshop commentary. Thanks go to Prof. Paul Cornish and Erik Silfversten for their efforts in reading early drafts of the commentary, as well as the two anonymous reviewers.

References

Cornish P. (2010). *Arms control tomorrow: the challenge of nuclear weapons in the twenty-first century. America and a changed world: a question of leadership/ed. by Robin Niblett. Chatham House.-Chichester...:* Wiley-Blackwell 223–37.

Federal Bureau of Investigation. *Iranian Ddos attacks*. FBI.gov, https://www.fbi.gov/wanted/cyber/iranian-ddos-attacks.

FireEye iSight Intelligence (2016). *Red line drawn: China recalculates its use of cyber espionage*. Milpitas, CA

Gartzke, E., & Lindsay, J. R. (2015). Weaving tangled webs: offense, defense, and deception in cyberspace. *Security Studies, 24*(2), 316–348.

Lindsay J R. (2015). Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. *Journal of Cybersecurity:* tyv003.

Mahnken, T. G., & Scott, J. (2012). *Conflict and cooperation in the global commons: a comprehensive approach for International Security.* Washington D.C: Georgetown University Press.

Mandiant. (2013). *Apt1: exposing one of China's cyber espionage units*. 76. Alexandria.

Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. *Journal of Strategic Studies, 38*(1–2), 4–37.

The White House (2015). *Remarks by President Obama and President Xi of the People's Republic of China in Joint Press Conference*. Office of the Press Secretary.

United Nations General Assembly. (2015). *Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security.* A/70/174.

US Department of Justice. (2014). *U.S. Charges Five Chinese Military Hackers with Cyber Espionage against U.S. Corporations and a Labor Organization for Commercial Advantage*. Office of Public Affairs.