Autonomous cyber capabilities and the international law of sovereignty and intervention

Article

Published Version

INTERNATIONAL LAW STUDIES

Published Since 1895

Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention

Michael N. Schmitt

96 INT'L L. STUD. 549 (2020)

Volume 96

2020

Published by the Stockton Center for International Law

ISSN 2375-2831

Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention

Michael N. Schmitt[*]

CONTENTS

* Professor of International Law, University of Reading; Francis Lieber Distinguished Scholar, Lieber Institute, U.S. Military Academy at West Point; Charles H. Stockton Distinguished Scholar-in-Residence, U.S. Naval War College; Distinguished Scholar, Strauss Center, University of Texas; Senior Fellow, NATO Cooperative Cyber Defence Centre of Excellence; Director of Legal Affairs, Cyber Law International. The author is grateful for the assistance of, and comments by, Christopher Greulich.

This article originated from a NATO Cooperative Cyber Defence Centre of Excellence project examining autonomous cyber capabilities. It and other papers produced during the project will appear in AUTONOMOUS CYBER CAPABILITIES UNDER INTERNATIONAL LAW (Rain Liivoja & Ann Väljataga eds., forthcoming 2021).

The thoughts and opinions expressed are those of the author and not necessarily those of the U.S. government, the U.S. Department of the Navy, or the U.S. Naval War College.

I. INTRODUCTION

The issue of how international law can respond to the advent of autonomous systems and autonomous cyber capabilities is fraught and emotive, especially in the context of warfare, with images of "killer robots" on one side and claims that autonomy will further humanitarian ends on the other.[1] This article explores the intersection of autonomous cyber capabilities and two primary rules of international law—that requiring respect for the sovereignty of other States and the prohibition on coercive intervention into their internal or external affairs. Of all of the rules of international law, these are the likeliest to be violated through employment of cyber capabilities, whether autonomous or not. This raises the question of whether a cyber operation that involves autonomous capabilities presents unique issues with respect to the application of the two rules. Are they up to the task of governing autonomy in cyberspace?

II. INTERNATIONALLY WRONGFUL ACTS

To address this question, it is first necessary to understand the concept of unlawfulness. The legal term for a violation of international law is "internationally wrongful act." According to Article 2 of the Articles on State Responsibility, a reliable restatement of the customary law of State responsibility prepared by the International Law Commission, "There is an *internationally wrongful act* of a State when conduct consisting of an action or omission: (a) Is *attributable* to the State under international law; and (b) Constitutes a *breach* of an international obligation of the State."[2] Both criteria must be satisfied for any cyber operation to be unlawful.

---

1. *Compare* HUMAN RIGHTS WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS 1 (2012), https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf (arguing that autonomous capabilities will result in "killer robots"), *with* Michael N. Schmitt & Jeffrey C. Thurner, *"Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231, 233 (2013) (concluding that banning autonomous weapons would be premature and that this technology may help minimize civilian harm).

2. International Law Commission, Report on the Work of its Fifty-Third Session, U.N. Doc. A/56/10, at 43 (2001), *reprinted in* [2001] 2 YEARBOOK OF THE INTERNATIONAL LAW COMMISSION 32, U.N. Doc. A/CN.4/SER.A/2001/Add.1 (Part 2) [hereinafter Articles on State Responsibility] (emphasis added). On customary international law, see International Law Commission, Report on the Work of Its Seventieth Session, U.N. Doc. A/73/10 (2018); *see also* G.A. Res. 73/203, ¶¶ 1, 7 (Jan. 11, 2019) (noting that the General Assembly welcomed the International Law Commission's report, took note of its recommendations,

As to the first, there are a number of bases for attributing a cyber operation to a State. The clearest example is when an "organ" of the State, such as the armed forces, a security service, an intelligence agency, or the State's cyber agency, conducted the cyber operation in question.[3] Another example of when a cyber operation is attributable to a State under law is when an individual or non-State group, such as a hacktivist, terrorist group, or private cyber security firm, acts on "the instructions of, or under the direction or control of, that State in carrying out" the operation.[4] Of course, these are just two of the several bases of attribution recognized by the International Law Commission.[5]

In the absence of attribution, a cyber operation will generally not violate international law (although there are limited exceptions, such as violations of international criminal law by individuals). For instance, operations mounted by patriotic hackers or cyber criminals who are not acting at the behest of a State do not qualify as internationally wrongful acts. Even beyond this key limitation, the attribution rules can prove challenging. To take one example, the type of relationship between a State and a non-State group that qualifies as "instructions or direction or control" is somewhat ambiguous legally, aside from the fact that evidence of that nexus may not be ironclad. In that regard, claims of attribution to a State often provoke debates over the requisite standard of evidential sufficiency.

The fact that a cyber operation involves autonomous capabilities can complicate factual attribution, but it does not make attribution more difficult as a matter of law. It is the nature of the relationship between the State and the individual or group conducting the operation that determines whether the attribution criterion for an internationally wrongful act has been satisfied. Taking the most straightforward example, a military unit's cyber operation that employs an autonomous capability is attributable to the unit's State irrespective of the consequences of the operation, including whether the unit anticipated, or could have reasonably anticipated, those consequences. Those are instead issues that bear on the second criterion of an internationally wrongful act, breach of a legal obligation owed another State.

---

and encouraged the "widest possible dissemination" of the report's findings on the identification of customary international law).

3. Articles on State Responsibility, *supra* note 2, art. 4.

4. *Id.* art. 8.

5. Other attributable cyber operations could include those conducted by persons or entities exercising elements of governmental authority (*id.* art. 5), organs placed at the disposal of a State by another State (*id.* at art. 6), operations carried out in the absence or default of the official authorities (*id.* art. 9), an insurrectional or other movement that becomes the new government (*id.* art. 10), and conduct that is acknowledged and adopted by a State (*id.* art. 11).

For the sake of this article's analysis, it will be assumed that the use of the autonomous cyber capabilities under consideration is attributable to a State. Therefore, the remaining analysis will focus on breach of the international law obligations requiring respect for the sovereignty of other States and prohibiting coercive intervention.

III. AUTONOMY

Before proceeding to those two obligations and the question of whether autonomy presents unique challenges to their application in the cyber context, it is first necessary to lay the groundwork for analysis by considering the concept of autonomy. Unfortunately, discussions of autonomous systems are plagued by a cacophony of definitions. For the purposes of this article, however, the definitional framework provided by Rain Liivoja, Maarja Naagel, and Ann Valjataga works well.

> [W]e consider autonomous operation in its simplest sense to refer to the ability of a system to perform some task without requiring real-time interaction with a human operator. Thus, the way a system performs is not decided, in each instance, by a person, but is the result of the design and programming of the system and the stimuli that it receives from its operational environment.
>
> . . . .
>
> [T]his broad definition of autonomy does not mean that an autonomous system is by definition one that is completely beyond human control. Rather, it means that the manner in which a human interacts with the system and exercises control over it differs from a system that is operated manually in real time.
>
> . . . .
>
> . . . Thus, when we speak in this paper of an autonomous cyber capability, we mean a capability that involves the performance of some significant function with a significant degree of autonomy. What constitutes significant would, however, vary from capability to capability.[6]

By this approach, different capabilities have different degrees of autonomy, ranging from so-called automated to those that are highly autonomous, with the common feature being the lack of real-time human

6. RAIN LIIVOJA, MAARJA NAAGEL & ANN VALJATAGA, NATO COOPERATIVE CYBER DEFENCE CENTRE OF EXCELLENCE, AUTONOMOUS CYBER CAPABILITIES UNDER INTERNATIONAL LAW 10–11 (2019), https://ccdcoe.org/uploads/2019/07/Autonomy-in-Cyber-Capabilities-under-International-Law_260619-002.pdf.

direction.[7] Thus, using common terminology, the autonomous systems referred to in this article include most "on the loop" (human monitoring and, if necessary, control) and "out of the loop systems" (system operating without human involvement), but not those in which the human is "in the loop" (human involved in operation of the system).

In the context of the law surrounding autonomous cyber capabilities, it also is useful to distinguish cyber operations that are offensive from ones that are defensive. As discussed in this article, the former category comprises cyber operations employing autonomous capabilities that are attributable to a State, whereas the latter are operations that are a direct response to the ongoing or imminent hostile cyber operations of another State. For instance, an autonomous capability designed to disable cyber infrastructure that is being used to carry out a hostile operation falls into the defensive category, whereas the operation to which it responds is offensive in character. A borderline case is one in which an autonomous cyber capability is employed in response to another State's hostile cyber operation but the responsive action targets cyber infrastructure other than that used to conduct the hostile operation. As examined herein, such a response would be encompassed in the offensive category, even though its motivation was defensive.

Defensive cyber operations employing autonomy may be further divided into passive and active operations. A passive capability operates within the targeted system. Examples are most firewalls and intrusion detection or prevention systems. Active defensive measures, by contrast, operate beyond the targeted systems, the paradigmatic example being a "hack back." As will be apparent, both the offensive-defensive and passive-active distinctions are relevant when assessing whether the use of an autonomous cyber capability

---

7. For a survey of this issue, see TIM MCFARLAND, AUTONOMOUS WEAPON SYSTEMS AND THE LAW OF ARMED CONFLICT 29–51 (2020). For U.S. specific military terminology and definitions related to autonomy in weapons systems, see U.S. Department of Defense, Directive 3000.09, Autonomy in Weapons Systems 13 (2012, Incorporating Change 1, May 8, 2017), https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf. Directive 3000.09 defines autonomous weapon system as:

> A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation.

Id. at 13–14. Further, the Directive defines a human-supervised autonomous weapon system as "[a]n autonomous weapon system that is designed to provide human operators with the ability to intervene and terminate engagements, including in the event of a weapon system failure, before unacceptable levels of damage occur" and a semi-autonomous weapon system as "[a] weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator." *Id.* at 14.

amounts to an internationally wrongful act in violation of the rules governing sovereignty and intervention. It is to those rules that this analysis turns.

IV. SOVEREIGNTY

The existence of a rule of sovereignty in international law was questioned in a 2018 speech by then-U.K. Attorney General, Jeremy Wright.

> Some have sought to argue for the existence of a cyber specific rule of a "violation of territorial sovereignty" in relation to interference in the computer networks of another state without its consent.
>
> Sovereignty is of course fundamental to the international rules-based system. But I am not persuaded that we can currently extrapolate from that general principle a specific rule or additional prohibition for cyber activity beyond that of a prohibited intervention. The UK Government's position is therefore that there is no such rule as a matter of current international law.[8]

Under this approach, cyber operations, whether involving autonomous capabilities or not, never violate the sovereignty of the State into which they are conducted.[9] For the United Kingdom, therefore, analysis typically begins with an assessment of whether a hostile cyber operation constitutes unlawful intervention,[10] or even a use of force in violation of U.N. Charter Article 2(4) and its customary analogue.

No other State has publicly taken the same position, although during the 2020 U.S. Cyber Command conference, the U.S. Department of Defense General Counsel expressed a degree of sympathy with elements of the

8. Jeremy Wright, U.K. Attorney General, Cyber and International Law in the 21st Century (May 23, 2018), https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century [hereinafter Wright Address].

9. *See, e.g.*, Gary P. Corn & Robert Taylor, *Sovereignty in the Age of Cyber*, 111 AMERICAN JOURNAL OF INTERNATIONAL LAW UNBOUND 207 (2017). *But see* Michael N. Schmitt & Liis Vihul, *Sovereignty in Cyberspace: Lex Lata Vel Non*?, 111 AMERICAN JOURNAL OF INTERNATIONAL LAW UNBOUND 213 (2017) (arguing that actions that reach a threshold degree of infringement on the territorial integrity of another State, as well as those that interfere with or usurp inherently governmental functions, necessarily violate the rule of sovereignty and are internationally wrongful acts).

10. *See infra* Part V.

position.[11] A number of States, including France,[12] the Netherlands,[13] Czech Republic,[14] Austria,[15] and Switzerland,[16] have taken the opposite position. In its 2020 *Allied Joint Doctrine for Cyberspace Operations*, NATO States did so as well, although the United Kingdom issued a reservation on that particular element of the doctrine.[17]

That sovereignty is a rule of international law applicable in the cyber context is the more defensible position, one well-founded in treaty law, State practice, and *opinio juris*, as well as the subsidiary sources of international law, decisions of tribunals, and the work of scholars.[18] Sovereignty is the rule of

---

11. Paul C. Ney, Jr., U.S. Department of Defense General Counsel, DOD General Counsel Remarks at U.S. Cyber Command Legal Conference (Mar. 2, 2020), https://www.defense.gov/Newsroom/Speeches/Speech/Article/2099378/dod-general-counsel-remarks-at-us-cyber-command-legal-conference/ [hereinafter Ney Address]. For a fuller discussion, see Michael Schmitt, *The Defense Department's Measured Take on International Law in Cyberspace*, JUST SECURITY (Mar. 11, 2020), https://www.justsecurity.org/69119/the-defense-departments-measured-take-on-international-law-in-cyberspace/.

12. MINISTÈRE DES ARMÉES, DROIT INTERNATIONAL APPLIQUÉ AUX OPÉRATIONS DANS LE CYBERSPACE [MINISTRY OF THE ARMIES, INTERNATIONAL LAW APPLIED TO CYBERSPACE] 6–7 (2019) (Fr.) [hereinafter FRANCE, MINISTÈRE DES ARMÉES]. For an analysis of the document, see Michael Schmitt, *France's Major Statement on International Law and Cyber: An Assessment*, JUST SECURITY (Sept. 16, 2019), https://www.justsecurity.org/66194/frances-major-statement-on-international-law-and-cyber-an-assessment/.

13. Ministry of Foreign Affairs, Government of the Netherlands, Letter to the Parliament on the International Legal Order in Cyberspace, Appendix: International Law in Cyberspace, at 1–2 (2019), https://www.government.nl/ministries/ministry-of-foreign-affairs/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace [hereinafter Netherlands, Ministry of Foreign Affairs]. For an analysis of the letter, see Michael Schmitt, *The Netherlands Releases a Tour de Force on International Law in Cyberspace: Analysis*, JUST SECURITY (Oct. 14, 2019), https://www.justsecurity.org/66562/the-netherlands-releases-a-tour-de-force-on-international-law-in-cyberspace-analysis/.

14. Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security: Second Substantive Session (10–14 February 2020), U.N. WEB TV (Feb. 11, 2020), http://webtv.un.org/search/3rd-meeting-open-ended-working-group-on-developments-in-the-field-of-information-and-telecommunications-in-the-context-of-international-security-second-substantive-session-10–14-february-2020/6131646836001/.

15. *Id.*

16. *Id.*

17. NATO, ALLIED JOINT PUBLICATION-3.20, ALLIED JOINT DOCTRINE FOR CYBERSPACE OPERATIONS 20 (ed. A, v.1 2020). For an analysis of the publication's legal significance, see Michael N. Schmitt, *Noteworthy Releases of International Cyber Law Positions – Part I: NATO*, ARTICLES OF WAR (Aug. 27, 2020), https://lieber.westpoint.edu/nato-release-international-cyber-law-positions-part-i/.

18. *See, e.g.*, Michael N. Schmitt & Liis Vihul, *Respect for Sovereignty in Cyberspace*, 95 TEXAS LAW REVIEW 1638 (2017).

international law most likely to be violated by hostile cyber operations attributable to States. The aspect of autonomy changes nothing in this regard.

Sovereignty can be violated based on either territoriality or on interference or usurpation of inherently governmental functions. For there to be a territorial violation, a cyber operation attributable to a State must cause some effect on another State's territory; it makes no difference whether that effect manifests on government or private cyberinfrastructure. More to the point, it makes no legal difference whether the requisite effect is caused by a system with autonomous capabilities. It is the nature of the effect that matters.[19]

The unresolved issue is the type of effects that qualify an operation as a sovereignty violation. It seems clear that non-*de minimis* physical damage or injury caused in another State's territory by the use of an autonomous cyber capability would do so. Below the threshold of physical damage or injury, however, consensus is elusive. The prevailing view appears to be that at least a cyber operation resulting in a permanent loss of functionality of the targeted cyber infrastructure, or systems that rely upon it, qualifies.[20] Similarly, an operation necessitating either replacement or physical repair of that system, as in the case of replacing components, violates sovereignty.[21]

Unfortunately, States have been reticent to set forth their legal positions as to the threshold for a cyber violation of sovereignty. To date, only France has done so with any degree of granularity. In a document issued by its Ministry of the Armies, that State took the position that "Any cyberattack against French digital systems or any effects produced on French territory by digital means by a State organ, a person or an entity exercising elements of governmental authority or by a person or persons acting on the instructions of or under the direction or control of a State constitutes a breach of sovereignty."[22] Although the precise parameters of France's approach remain to be determined, it is an extremely broad approach to qualifying cyber operations as violations of sovereignty, one that other States may feel uncomfortable adopting, lest it bar their own cyber operations.

---

19. *See* TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS, r. 4, at 17, and accompanying commentary, at 17–27 (Michael N. Schmitt ed. 2017) [hereinafter TALLINN MANUAL 2.0].

20. *Id.* at 20–21; *see also* FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 7.

21. For instance, a 2012 hostile cyber operation targeting Saudi Aramco affected 35,000 computers, necessitating the replacement of affected hard drives. *See* Jose Pagliery, *The Inside Story of the Biggest Hack in History*, CNN BUSINESS (Aug. 5, 2015), https://money.cnn.com/2015/08/05/technology/aramco-hack/.

22. FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 7.

Returning to the operational typology, a passive cyber defensive measure employing autonomous capability will not violate the sovereignty of other States since it takes place on the territory of the State conducting it. However, both active defensive measures and offensive cyber operations involving autonomy raise the prospect of a sovereignty violation. Whether sovereignty is violated is a question of law (the threshold for violation) and one of fact (the scale and nature of the effects). Autonomy does not alter the application of either of these determinations.

Sovereignty can also be violated when a cyber operation by one State interferes with, or usurps, an inherently governmental function of another State. Whether this violation can take place outside the territory of the State against which the hostile cyber operation is directed remains unsettled in international law.[23] For instance, it is unclear whether a cyber operation that leverages autonomous capabilities to target the Estonian government data stored at a data center in Luxembourg, thereby impeding Estonian's ability to carry out its inherently governmental functions, violates Estonian sovereignty on this basis.

In most cases, hostile operations are directed against cyberinfrastructure located in a State's territory. There is a key distinction between violations based on interference with or usurpation of an inherently governmental function and those based on territorial effects. The former, unlike the latter, does not require any type of harm. The determinative factor is simply whether interference or usurpation occurred. This distinction opens the door to non-destructive and non-injurious cyber operations employing autonomous capabilities, or those that otherwise do not reach the threshold of territorial violation, amounting to a sovereignty violation.

An inherently governmental function may best be understood as a function that States alone have the authority to perform (or authorize other entities to perform on their behalf). Classic examples include collecting taxes, conducting elections, and enforcing laws. For instance, take the case of an autonomous cyber capability that searches for systems being used by a particular candidate's campaign and disrupts their use. Irrespective of whether the effects on those systems qualify the operation as a breach on the basis of territoriality, the fact that the candidate's campaign has been disrupted would amount to interference in the conduct of the election by the State concerned.

Similarly, consider an autonomous cyber capability used by law enforcement that activates when it senses criminal activity. Such a cyber capability might then attempt to penetrate the criminal infrastructure to disable it or gather evidence against the perpetrator. Deploying such law

23. TALLINN MANUAL 2.0, *supra* note 19, at 23.

enforcement cyber capabilities into another State's criminal jurisdiction without consent would constitute a violation of the territorial State's sovereignty and would usurp an inherently governmental function. This is because only the State from which the purported criminal activity emanated enjoys the competency under international law to exercise, or consent to another State's exercise of, law enforcement authority on its territory. That the intrusion relied on autonomous capabilities has no bearing on the lawfulness of the law enforcement activity.

As with territoriality, the use of an autonomous passive defense capability is unlikely to trigger a violation of another State's sovereignty on the basis of interference with or usurpation of another State's inherently governmental functions because States seldom have a right under international law to engage in those functions abroad (except in the commons). And as with a violation of sovereignty on the basis of territoriality, active cyber defense capabilities and offensive operations, even if being employed autonomously, risk violating the law should they interfere with or usurp another State's exclusive right to engage such functions on its own territory.

V. INTERVENTION

Unlike sovereignty, the existence of a rule of non-intervention in the cyber context is uncontroversial, as illustrated by the U.N. Group of Governmental Experts' confirmation in its 2015 report,[24] a position subsequently endorsed by the General Assembly.[25] Intervention into the internal or external affairs of another State is an internationally wrongful act in both customary international law and under certain treaties, such as the Charter of the Organization of American States.[26] The parameters of a treaty violation of the rule are to be found in the text of the instruments themselves, as well as through interpretation consistent with precepts in the

24. Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (2015), transmitted by Letter Dated 26 June 2015 from the Chair of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security Established Pursuant to Resolution 68/243 (2014) ¶¶ 26, 27(b), U.N. Doc. A/70/174 (July 22, 2015).

25. G.A. Res. 70/237, at 2 (Dec. 30, 2015).

26. Charter of the Organization of American States, art. 15, Apr. 30, 1948, 2 U.S.T. 2394, 119 U.N.T.S. 3.

Vienna Convention on the Law of Treaties,[27] while the following analysis of intervention by autonomous cyber means is limited to the customary international law rule of non-intervention.[28]

In its *Nicaragua* judgment, the International Court of Justice (ICJ), applying customary international law, observed that intervention consists of two elements, both of which must be satisfied for a violation to occur. First, the object of the cyber operation must be another State's internal or external affairs, known as the *domaine réservé*. As, the Court explained,

> [T]he principle forbids all States or groups of States to intervene directly or indirectly in internal or external affairs of other States. A prohibited intervention must accordingly be one bearing on matters in which each State is permitted, by the principle of State sovereignty, to decide freely. One of these is the choice of a political, economic, social and cultural system, and the formulation of foreign policy.[29]

In other words, *domaine réservé* is an area of activity that international law leaves for States to regulate, thereby recognizing their discretion to make their own choices about such activities. Although the precise contours of the *domaine réservé* are indistinct, certain activities unambiguously fall within its ambit. For example, language policy, elections, crisis management, the structure of government, and diplomatic activities clearly qualify, thereby opening the door to the possibility that using autonomous cyber capabilities to affect them, as in the case of disrupting the functioning of a nation's response to a pandemic,[30] might run afoul of the non-intervention rule. By contrast, matters committed to international law, such as the international human rights to expression and privacy online, do not qualify. For instance, using autonomous cyber capabilities to disrupt another State's efforts to block lawful online expression would not qualify as a violation of the non-intervention rule; it might, however, violate the sovereignty of the State concerned.

---

27. Vienna Convention on the Law of Treaties, arts. 31–33, May 23, 1969, 1155 U.N.T.S. 331.

28. *See* TALLINN MANUAL 2.0, *supra* note 19, r. 66, at 312, and accompanying commentary, at 312–25.

29. Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Judgement, 1986 I.C.J. Rep. 14, ¶ 205 (June 27) [hereinafter Nicaragua].

30. *See, e.g.*, Marko Milanovic & Michael N. Schmitt, *Cyber Attacks and Cyber (Mis)information Operations During a Pandemic*, JOURNAL OF NATIONAL SECURITY LAW AND POLICY (forthcoming 2020) (discussing sovereignty and intervention in the context of a global pandemic).

Although there is significant overlap with the concept of inherently governmental functions in the law of sovereignty, *domaine réservé* is a broader notion.[31] Most inherently governmental functions qualify as a *domaine réservé*, but certain *domaine réservés* are not inherently governmental. An example is the provision of tertiary education, which in many States is provided by the private sector and thus not inherently governmental. However, it is a *domaine réservé* since international law generally leaves States free to regulate such education. Accordingly, an offensive cyber operation involving autonomous capabilities that disrupts the functioning of tertiary education would likely not violate sovereignty unless it caused the requisite territorial effects but could constitute prohibited intervention so long as the second element of intervention, coercion, is satisfied.[32]

The ICJ discussed coercion in its *Nicaragua* judgment, stating:

> Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention which uses force, either in the direct form of military action, or in the indirect form of support for subversive or terrorist armed activities within another State.[33]

Applying this standard by analogy, using an offensive autonomous cyber capability to support insurgents fighting their government would amount to a clear case of intervention. The question, though, is in what other circumstances is use of an autonomous cyber capability against a *domaine réservé* prohibited by the rule?

In a 2019 letter to the Parliament on the "International Legal Order in Cyberspace," the Netherlands Ministry of Foreign Affairs noted the imprecise definition of coercion, before characterizing the concept of coercion as follows:

---

31. On the relationship between sovereignty and the non-intervention principle, see HARRIET MOYNIHAN, CHATHAM HOUSE, THE APPLICATION OF INTERNATIONAL LAW TO STATE CYBERATTACKS: SOVEREIGNTY AND NON-INTERVENTION 48–51 (2019), https://www.chathamhouse.org/sites/default/files/publications/research/2019-11-29-Intl-Law-Cyberattacks.pdf.

32. However, the analysis must be precise. If universities are engaged in developing responses to a pandemic at the behest of or in cooperation with the government, use of an autonomous cyber capability could be a violation of sovereignty on the basis that dealing with a pandemic is an inherently governmental function.

33. Nicaragua, *supra* note 29, ¶ 205.

> The precise definition of coercion, and thus of unauthorised intervention, has not yet fully crystallised in international law. In essence it means compelling a state to take a course of action (whether an act or an omission) that it would not otherwise voluntarily pursue. The goal of the intervention must be to effect change in the behaviour of the target state.[34]

Restated, an act of coercion is one that deprives another State of choice by either causing that State to behave in a way it otherwise would not or to refrain from acting in a manner in which it otherwise would act.[35] Merely influencing the other State's choice does not suffice; the choice to act or not has to effectively be taken off the table in the sense that a reasonable State in the same or similar circumstances would no longer consider it to be a viable option.

To illustrate, using autonomous cyber capabilities to spread disinformation during an election is a noxious form of influence, but it is not necessarily coercive, for voters (the State) retain their ability to decide for whom to vote. But using autonomous cyber capabilities to disrupt the operation of voting machinery or alter vote counts would certainly be coercive because the very ability of members of the electorate to exercise political choice has been denied.[36]

An often-misunderstood dynamic of the prohibition involves the relationship between the coercion and the *domaine réservé*. The *domaine réservé* is not the physical target of the operation. Rather, it is that area of activity that the cyber operation is meant to coerce. Consider a State's covert cyber operation that employs autonomous capabilities in a ransomware attack against the sole international port facility of another State. To assess whether

---

34. Netherlands, Ministry of Foreign Affairs, *supra* note 13, at 3; Wright Address, *supra* note 8. In his speech, Attorney General Wright observed, "The precise boundaries of this principle are the subject of ongoing debate between states, and not just in the context of cyber space." *Id.*; *see also* Brian J. Egan, *International Law and Stability in Cyberspace*, 35 BERKELEY JOURNAL OF INTERNATIONAL LAW 169, 174–75 (2017) (noting the challenges of applying the rule of non-intervention, and sovereignty more generally, to cyberspace while serving as U.S. Department of State Legal Adviser).

35. *See* DEPARTMENT OF FOREIGN AFFAIRS AND TRADE, AUSTRALIA'S INTERNATIONAL CYBER ENGAGEMENT STRATEGY: 2019 INTERNATIONAL LAW SUPPLEMENT (2019), https://www.dfat.gov.au/publications/international-relations/international-cyber-engagement-strategy/aices/chapters/2019_international_law_supplement.html (defining a prohibited intervention as "one that interferes by coercive means (in the sense that they effectively deprive another state of the ability to control, decide, or govern matters of an inherently sovereign nature), either directly or indirectly, in matters that a state is permitted by the principle of state sovereignty to decide freely").

36. *See* Michael N. Schmitt, *"Virtual" Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law*, 19 CHICAGO JOURNAL OF INTERNATIONAL LAW 30 (2018) (discussing cyber election meddling that would be considered coercive).

the operation constitutes unlawful intervention, it is necessary to determine why the former is conducting that hostile activity. If it is merely a criminal attempt to acquire funds, it is not coercive vis-a-vis any *domaine réservé*. However, if designed to force the State to, for instance, alter its trade practice by creating a situation in which there is no choice but to transship through the attacker's logistics network, the relationship between the coercive operation and a *domaine réservé* exists.

As to the typology of operations, passive defensive cyber operations enabled by autonomy will not violate this rule because there is no *domaine réservé* to coerce; States do not enjoy control over a *domaine réservé* on the territory of other States. In most cases, the same is true with regard to active defensive cyber operations that employ autonomous capabilities. This is because there must be an attempt to deprive the State concerned of its exercise of choice over an area of activity that is not committed to international law. Since the State conducting the initial hostile cyber operation to which the defensive action responds is operating extraterritorially, that operation is committed to international law rules ranging from the requirement to respect the sovereignty of other States to the prohibition on the use of force. It may be that the specific operation does not violate any particular rule, but that extraterritorial cyber operations into another State's territory are governed by the general rules of international law, a position long accepted by the international community.[37] Of course, offensive cyber operations are subject to the rule of non-intervention, whether conducted using autonomous capabilities or not. Beyond attribution, the only question is whether the elements necessary for breach of that primary rule have been satisfied.

VI. INTENT AND MISTAKE OF FACT

The fact that autonomous cyber capabilities operate without human involvement, and sometimes without immediate human oversight, raises issues of intent and mistake of fact. In this regard, it is necessary to dispense with one red herring at the outset. Just because a cyber capability operates autonomously does not mean that the State that employs it lacks the intent to cause the requisite consequences. Autonomous systems are not independent actors in the legal system. Rather, autonomous capabilities are

37. *See, e.g.*, Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (2013), transmitted by Letter Dated 7 June 2013 from the Chair of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security Established Pursuant to Resolution 66/24 (2012), ¶¶ 19–20, U.N. Doc. A/68/98 (June 24, 2013).

programmed by humans and, more importantly, humans decide to use them. So long as that decision is attributable to a State as described above, the use of an autonomous cyber capability in no way takes the operation beyond the reach of the rules regarding sovereignty and intervention.

However, that a human may not entirely understand how a system with autonomous capabilities might operate, or at least be able to predict the consequences of its use, raises an interesting question. If the individual or entity deciding to use the capability did not intend the effect that occurred, but that effect would otherwise qualify the operation as a violation of either the sovereignty or intervention rules, have those rules nevertheless been violated?

Consider a cyber operation that uses autonomous capabilities to map a targeted system in another country. The State conducting the operation harbors no intention of causing any physical effects that would violate sovereignty, and mere cyber espionage is generally not considered to be an internationally wrongful act.[38] However, during the course of the operation, some damage unexpectedly results to the targeted system. Has the State conducting the operation breached its obligation to respect the target State's sovereignty?

Or consider a State's covert cyber operation employing autonomous capabilities to engage in the theft of intellectual property related to the development of a vaccine vital to combating an ongoing pandemic. The State does not seek to impede the process, but after discovering the breach the affected laboratories have to shut down temporarily to assess the integrity of research data. As a result, development of the critical vaccine is slowed. Did the operation violate the rule of nonintervention because (1) a nation's pandemic response falls within its *domaine réservé* and (2) the laboratories were forced to temporarily interrupt vaccine development? Of course, such situations could arise in the case of a cyber operation not employing autonomous capabilities, but they would seem more likely to surface should autonomy be relied upon.

The International Law Commission addressed the issues of intent and knowledge in its commentary to the Articles on State Responsibility.

> Whether there has been a breach of a rule may depend on the intention or knowledge of relevant State organs or agents and in that sense may be

38. TALLINN MANUAL 2.0, *supra* note 19, r. 32, at 168; *see also* Ashley Deeks, *An International Legal Framework for Surveillance*, 55 VIRGINIA JOURNAL OF INTERNATIONAL LAW 291, 300–13 (2015). *But see* Inaki Navarrete & Russell Buchan, *Out of the Legal Wilderness: Peacetime Espionage, International Law and the Existence of Customary Exceptions*, 51 CORNELL INTERNATIONAL LAW JOURNAL 897 (2019); RUSSELL BUCHAN, CYBER ESPIONAGE AND INTERNATIONAL LAW (2018).

> "subjective". For example, article II of the Convention on the Prevention and Punishment of the Crime of Genocide states that: "In the present Convention, genocide means any of the following acts committed with intent to destroy, in whole or in part, a national, ethnical, racial or religious group, as such …" In other cases, the standard for breach of an obligation may be "objective", in the sense that the advertence or otherwise of relevant State organs or agents may be irrelevant. Whether responsibility is "objective" or "subjective" in this sense depends on the circumstances, including the content of the primary obligation in question. The articles lay down no general rule in that regard. The same is true of other standards, whether they involve some degree of fault, culpability, negligence or want of due diligence. Such standards vary from one context to another for reasons which essentially relate to the object and purpose of the treaty provision or other rule giving rise to the primary obligation. Nor do the articles lay down any presumption in this regard as between the different possible standards. Establishing these is a matter for the interpretation and application of the primary rules engaged in the given case.[39]

In other words, the role of intent turns on whether it is an element of the breach in question. On the one hand, if it is, as is textually the case with genocide and other rules of international criminal law, the absence of intent will preclude a cyber operation that involves autonomous cyber capabilities from amounting to either an internationally wrongful act by the State concerned or an act generating individual criminal responsibility. Importantly though, the commentary acknowledges that intent can be a condition precedent to the breach of a primary rule in which the requirement is not clear on its face. Thus, in cases of an implicit intent requirement, no breach will lie absent intent.

On the other hand, the absence of an express or implied intent requirement raises the possibility of breach even if the consequences that manifested were unforeseen and unforeseeable. Accordingly, the role of intent in assessing whether a cyber operation employing autonomous capabilities violates international law depends on the presence or absence of a *mens rea* element in the individual primary rules.

However, a degree of caution is merited. As Marko Milanovic has noted, certain rules and regimes of international law have developed bespoke standards with respect to mistakes of fact. For instance, he notes that in international human rights law and international humanitarian law an "honest and reasonable" mistake as to the facts can exonerate the State

39. Articles on State Responsibility, *supra* note 2, art. 2, commentary ¶ 3, at 34–35.

concerned.[40] This begs the question of whether a similar mistake of fact standard should apply in the case of other rules of international law like sovereignty and intervention.

To illustrate, consider a State A cyber countermeasure[41] involving autonomous capabilities mounted against State B that bleeds over into State C. The result is a permanent loss of functionality of affected cyber infrastructure in State C, a violation of that State's sovereignty. If State A should have known (constructive knowledge) that bleed over would occur, it has violated State C's sovereignty even though the operation's qualification as a countermeasure precluded its wrongfulness as to State B. The belief that there would be no bleed over was not reasonable. But if the belief was reasonable, should that fact excuse the violation of State C's sovereignty?

The experts who drafted *Tallinn Manual 2.0* concluded that a reasonable mistake of fact as to the need to use force in self-defense against another State would excuse that use of force.[42] As Milanovic notes, there is a degree of State practice supporting this position.[43] Yet the ICJ seemed to come to a contrary conclusion in its *Oil Platforms* judgment.[44] And in the context of countermeasures, the International Law Commission, in its commentary to the Articles on State Responsibility, opined that,

> A State taking countermeasures acts at its peril, if its view of the question of wrongfulness turns out not to be well founded. A State which resorts to countermeasures based on its unilateral assessment of the situation does so at its own risk and may incur responsibility for its own wrongful conduct in the event of an incorrect assessment. In this respect, there is no difference between countermeasures and other circumstances precluding wrongfulness.[45]

A majority of the experts who authored *Tallinn Manual 2.0* took the same position. In doing so, they "emphasised the desirability of preventing a proliferation of countermeasures and the fact that countermeasures, despite being designed to resume lawful relations between the States concerned,

40. Marko Milanovic, *Mistakes of Fact When Using Lethal Force in International Law: Part I*, EJIL:TALK! (Jan. 14, 2020), https://www.ejiltalk.org/mistakes-of-fact-when-using-lethal-force-in-international-law-part-i/.

41. *See infra* Section VII.A.

42. TALLINN MANUAL 2.0, *supra* note 19, at 347.

43. Marko Milanovic, *Mistakes of Fact When Using Lethal Force in International Law: Part II*, EJIL:TALK! (Jan. 15, 2020), https://www.ejiltalk.org/mistakes-of-fact-when-using-lethal-force-in-international-law-part-ii/.

44. Oil Platforms (Iran v. U.S.), Judgment, 2003 I.C.J. Rep. 161, ¶ 73 (Nov. 6).

45. Articles on State Responsibility, *supra* note 2, commentary to art. 49, at 130–31.

nevertheless present a risk of escalation."[46] The experts distinguished this position from their view with respect to a mistake of fact in the context of self-defense on the basis that States should be afforded a wide degree of discretion to act when the consequences of a failure to do so can be extremely serious, as is the case with respect to a failure to respond to an armed attack.

But that conclusion was not unanimous. Some experts contended that an honest and reasonable mistake of fact should operate to leave the countermeasure's preclusion of wrongfulness intact.[47] In their view, States must be empowered to defend themselves against hostile cyber operations, whether those operations are at the level of an armed attack entitling the victim State to act in self-defense or an internationally wrongful act below that level that opens the door to countermeasures.

As is apparent, the law surrounding the mistake of fact doctrine, beyond discreet bodies of law in which such a doctrine clearly applies, remains unsettled. This is certainly the case with respect to both sovereignty and intervention. The sounder legal position is that it does not excuse a violation of international law unless it negates intent with regard to a primary rule of international law requiring intent as a condition of violation. Otherwise, the State that was the victim of the mistake of fact would have to suffer the consequences of that mistake without the possibility of securing reparations, which are only due in the face of an internationally wrongful act.[48] By rejecting the applicability of a mistake of fact doctrine, the costs of a mistake of fact are appropriately shouldered by a State making it, not the victim of that mistake.

Since intent is not a required element of the breach of the obligation to respect the sovereignty of another State, a cyber operation using autonomous capability that causes unintended qualifying effects would violate international law. As to the unsettled question of whether a mistake of fact doctrine might excuse a sovereignty violation, States are likely to reject its applicability for the aforementioned reason, particularly as autonomous, and especially artificial intelligence, cyber capabilities become common. After all, the less control a State exercises over the conduct of an operation, the more logical it is that the State bears the risk of its mistake and the less appropriate it is that victim States should be left less than whole.

By contrast, intent is an implied requirement for the internationally wrongful act of intervention into the internal or external affairs of another State. Recall that there must be a relationship between coercion and the

46. TALLINN MANUAL 2.0, *supra* note 19, at 116.

47. *Id.*

48. Articles on State Responsibility, *supra* note 2, art. 31.

domaine réservé; the State conducting the operation has to seek to deprive the target State of choice with respect to its behavior or policies involving a *domaine réservé*. Therefore, absent intent to do so, there would be no violation of this prohibition if an autonomous cyber capability caused unexpected harm that in fact deprived the affected State of choice.

To take a simple example, consider a case in which a State uses autonomous passive cyber defences to enhance the security of cyber systems on its territory. An insurgent group in another State has been using a social media platform operated from the former for command, control, and communications (C^3) in hostilities with the government. The autonomous passive defensive measures significantly improve the security of social media, thereby contributing to the security of the insurgent group's C^3. In that there was no intent to enhance the insurgent group's operational capabilities, there is no intervention.

VII. CIRCUMSTANCES PRECLUDING WRONGFULNESS

Even though certain cyber operations employing autonomous capabilities might breach either the obligation to respect the sovereignty of other States or the prohibition on intervention into the internal or external affairs of those States, international law sets forth a number of circumstances in which international law nevertheless would not be violated. These so-called "circumstances precluding wrongfulness" include consent, self-defense, qualification of the action as a countermeasure, *force majeure*, distress, and necessity.[49] The most significant in the context of autonomy are countermeasures, necessity, and self-defense; the analysis that follows focuses on these three circumstances.

A. Countermeasures

A countermeasure is an "act" (either an action or an omission) that would be unlawful but for the fact that it is designed to put an end to another State's (the "responsible State") operation that is breaching an obligation owed the former (the "injured State").[50] Nothing bars application of this circumstance precluding wrongfulness to cyber operations that involve autonomous capabilities.

---

49. *Id.* arts. 20–25.

50. *See* TALLINN MANUAL 2.0, *supra* note 19, rr. 20–25, at 111–34; *see also* Michael N. Schmitt, *"Below the Threshold" Cyber Operations: The Countermeasures Response Option and International Law*, 54 VIRGINIA JOURNAL OF INTERNATIONAL LAW 697 (2014).

As an example, this basis for precluding wrongfulness of an act could allow for active defense, such as an autonomously conducted hack-back or a human launched hack-back involving autonomous capabilities. It could also take the form of an offensive operation employing autonomous capabilities against systems other than those used to conduct the unlawful cyber operation if the objective is to compel the responsible State to desist. This is because a countermeasure need not be directed at the entity conducting the unlawful cyber operation or the cyberinfrastructure from which it originated. For instance, a cyber countermeasure might leverage autonomous capabilities to target vulnerable government or private cyberinfrastructure having nothing to do with the cyber operation to which the injured State is responding. A countermeasure need not even be in-kind; a cyber operation involving autonomous capability may be used in response to a non-cyber internationally wrongful act, as in the case of providing funding or arms to an insurgent group fighting the government.[51] The key limitation on countermeasures is instead that they may only be intended to either put an end to an ongoing unlawful action or to secure reparations for one that has been completed, or both; countermeasures may not, however, be motivated by a desire to punish or retaliate.[52]

The prospect of employing an autonomous capability as a countermeasure raises three issues. First, countermeasures must be proportionate. Proportionality is understood in the countermeasures context as meaning "commensurate with the injury suffered, taking into account the gravity of the internationally wrongful act and the rights in question."[53] In practical terms, the negative effects of the countermeasure for the responsible State may not be excessive relative to the harm the injured State is suffering. If the autonomous capability causes excessive harm, the State taking the purported countermeasure will have itself violated international law. In this regard, recall that the absence of intent or a mistake of fact often will not excuse the injured State's violation even if the nature and extent of harm caused were unforeseen and unforeseeable. In most cases, a disproportionate countermeasure will violate the responsible State's sovereignty, but other violations might also occur.

Second, the Articles on State Responsibility provide that "[b]efore taking countermeasures, an injured State shall call upon the responsible State . . . to fulfil its obligations [to cease the operation and offer any appropriate

51. Such actions qualify as intervention. *See e.g.*, Nicaragua, *supra* note 29, ¶ 242.
52. Articles on State Responsibility, *supra* note 2, art. 49(1).
53. *Id.* art. 51.

assurances, guarantees and reparations.[54] Further, the injured State must "notify the responsible State of any decision to take countermeasures and offer to negotiate with that State."[55] An absolute notification requirement would not necessarily preclude the post-notice launch of a cyber countermeasure involving autonomous capabilities, but it would bar using autonomous capabilities to launch an automatic response to an incoming hostile cyber operation.

The commentary to the Articles on State Responsibility acknowledges that there may be certain situations requiring "urgent countermeasures" to preserve an injured State's rights.[56] States that have spoken to the issue have taken a strong stance against a notice requirement in situations in which notice might diminish the countermeasure's likelihood of success, for instance by allowing the responsible State to take measures in anticipation of the action[57] or because providing notice could reveal sensitive capabilities.[58] This does not necessarily mean that an automatic hack-back relying upon autonomous capabilities or a no-notice countermeasure involving autonomy would never run afoul of the purported notice requirement. But it does open the door to no-notice countermeasures so long as the State employing the autonomous capability can make a cogent argument that it was necessary to act without notice, as might be the case with hostile operations against critical infrastructure that can only be defeated by exploiting a zero day vulnerability in the responsible State's systems.

Third, countermeasures are only available in response to internationally wrongful acts that are attributable to States.[59] Therefore, to be lawful there would have to be a relatively high degree of certainty that a particular State was behind the hostile cyber operation if autonomous capabilities were used to determine whether to launch the countermeasure response or the countermeasure response itself involved autonomous capabilities. This is an important limitation in light of the view expressed above that a mistake of

54. *Id.* arts. 30–31.

55. *Id.* art. 52.

56. *Id.* art. 52, commentary ¶ 1, at 135.

57. FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 8; Netherlands, Ministry of Foreign Affairs, *supra* note 13, at 7; Ney Address, *supra* note 11.

58. Wright Address, *supra* note 8.

59. Articles on State Responsibility, *supra* note 2, art. 22. Note that a countermeasure directed at a non-State actor conducting hostile cyber operations might be appropriate if the State from which the operation being mounted is in breach of its due diligence obligation. *See* Michael N. Schmitt, *In Defense of Due Diligence*, 124 YALE LAW JOURNAL FORUM 68, 79–80 (2015), https://www.yalelawjournal.org/forum/in-defense-of-due-diligence-in-cyberspace.

fact does not excuse an internationally wrongful act unless provided for in the body of law or primary rule in question, which is not the case with sovereignty or intervention. Indeed, recall that both the International Law Commission and a majority of the *Tallinn Manual 2.0* experts were of the view that countermeasures are taken at the injured State's risk.[60]

B. Necessity

A second basis upon which the wrongfulness of a cyber operation utilizing autonomous capability is precluded is in a circumstance of necessity. As noted in the Articles on State Responsibility, a cyber operation is "necessary" when it is "the only way for the State to safeguard an essential interest against a grave and imminent peril" and the act "does not seriously impair an essential interest of the State or States towards which the obligation exists, or of the international community as a whole."[61]

This circumstance precluding wrongfulness is especially important, for there is no requirement that the hostile cyber operation to which the cyber operation responds be attributable to a State, or even that the initiator of the operation be known. Moreover, the hostile cyber operation to which the State responds in necessity need not be an internationally wrongful act. Most importantly, a State's cyber operation conducted on the basis of necessity is lawful even though it may breach an obligation such as sovereignty that is owed another State that bears no responsibility whatsoever for the situation, as long as doing so does not seriously affect the latter's essential interests. This makes the possibility of bleed over caused by an autonomous capability less likely to result in a violation of international law. Thus, necessity fills key gaps left by these requirements in the context of countermeasures.[62]

As with countermeasures, there may be practical issues with respect to using autonomous capabilities in situations of necessity, both when they contribute to determining whether to launch a response (perhaps without human involvement), and as to those that form part of the cyber response. With respect to the former, the autonomous capability would have to discern if an essential interest of the State is at stake and determine whether the negative impact on that interest is grave. Part of the challenge is that neither "essential interest" nor "grave and imminent peril" are well-defined in international law.

---

60. *See supra* notes 45–46 and accompanying text.

61. Articles on State Responsibility, *supra* note 2, art. 25(1); *see generally* TALLINN MANUAL 2.0, *supra* note 19, r. 26, at 135, and accompanying commentary, at 135–42.

62. Articles on State Responsibility, *supra* note 2, commentary to art. 25.

In this regard, policymakers and scholars often speak in terms of hostile cyber operations against critical infrastructure as triggering necessity. However, it is not the infrastructure that must be essential, but rather the interest that an operation against the infrastructure will affect that must qualify as essential. Moreover, the notion of critical infrastructure is relative; one State's critical infrastructure may not be another's because State's have differing needs. And even if it can be agreed that certain cyber infrastructure is of a nature that an operation conducted against it will always affect an essential interest, as in the case of nuclear facilities, a cyber operation targeting that infrastructure might not gravely affect the interest. Thus, while there could be circumstances in which the employment of autonomous capabilities on the basis of necessity is lawful, the capability would have to be programmed very carefully to ensure it comports with necessity's demanding criteria.

Finally, the requirement that a cyber operation mounted on the basis of necessity not place the essential interests of other States in grave and imminent peril presents a significant obstacle if autonomous capabilities are used. Should the response cause an effect at that level, the fact that the State did not anticipate those consequences, a possibility that is likely exacerbated by autonomous capabilities, would not shield it from responsibility for violations of international law, in particular sovereignty, involving those effects.

C. Self-Defense

A third circumstance precluding wrongfulness is self-defense pursuant to Article 51 of the U.N. Charter and customary international law.[63] That article provides, in relevant part, "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security."[64] Although self-defense as a circumstance precluding wrongfulness is usually discussed in the context of the prohibition on the use of force found in Article 2(4) of the U.N. Charter and customary international law, most uses of force also violate the sovereignty of the State into which they are conducted and, as noted by the ICJ in its *Nicaragua* judgment, the rule of

63. *Id.* art. 21; *see generally* TALLINN MANUAL 2.0, *supra* note 19, rr. 71–75 and accompanying commentary, at 339–56.

64. U.N. Charter art. 51.

non-intervention.[65] Thus, if a cyber operation involving autonomous capability qualifies as an act of self-defense, neither of those rules is violated.

In that preclusion of wrongfulness under self-defense envisions a use of force, strict criteria govern its applicability. Most important, self-defense is only available when the operation to which it responds is at the level of an "armed attack."[66] That threshold is somewhat ambiguous in the non-cyber context but much more so with respect to hostile cyber operations.[67] Cyber operations involving autonomous capabilities that result in significant human injury or physical damage clearly qualify, but below that kinetic threshold there is a lack of international consensus.[68]

The most robust position taken to date is that of the French Ministry of the Armies, which announced in 2019 that:

> [a] cyberattack could be categorised as an armed attack if it caused substantial loss of life or considerable physical or economic damage. That would be the case of an operation in cyberspace that caused a failure of critical infrastructure with significant consequences or consequences liable to paralyse whole swathes of the country's activity, trigger technological or ecological disasters and claim numerous victims.[69]

---

65. Nicaragua, *supra* note 29, ¶ 205.

66. *See* U.N. Charter art. 51.

67. Michael N. Schmitt, *The Use of Cyber Force and International Law*, *in* THE OXFORD HANDBOOK OF THE USE OF FORCE IN INTERNATIONAL LAW 1110, 1119–29 (Marc Weller ed. 2015).

68. *See* Netherlands, Ministry of Foreign Affairs, *supra* note 13, at 9 ("At present there is no international consensus on qualifying a cyberattack as an armed attack if it does not cause fatalities, physical damage or destruction yet nevertheless has very serious non-material consequences."). However, there is a growing sense that the assessment should be contextual, as recommended by the *Tallinn Manual* Experts. *See* TALLINN MANUAL 2.0, *supra* note 19, at 333–37; *see also* Netherlands, Ministry of Foreign Affairs, *supra* note 13, at 4

> It is necessary, when assessing the scale and effects of a cyber operation, to examine both qualitative and quantitative factors. The Tallinn Manual 2.0 refers to a number of factors that could play a role in this regard, including how serious and far-reaching the cyber operation's consequences are, whether the operation is military in nature and whether it is carried out by a state.

DEPARTMENT OF FOREIGN AFFAIRS AND TRADE, AUSTRALIA'S INTERNATIONAL CYBER ENGAGEMENT STRATEGY: ANNEX A: AUSTRALIA'S POSITION ON HOW INTERNATIONAL LAW APPLIES TO STATE CONDUCT IN CYBERSPACE (2017), https://www.dfat.gov.au/publications/international-relations/international-cyber-engagement-strategy/aices/chapters/annexes.html ("In determining whether a cyber attack, or any other cyber activity, constitutes a use of force, states should consider whether the activity's scale and effects are comparable to traditional kinetic operations that rise to the level of use of force under international law.").

69. FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 8.

Since the French position has not yet been publicly embraced by other States, most of whom have remained silent on the matter, the threshold at which self-defense will preclude the wrongfulness of a cyber operation involving autonomous capabilities remains uncertain.

This being so, States resorting to autonomous capabilities must be alert lest they inadvertently respond in self-defense to a cyber operation that falls short of the armed attack threshold, wherever it might lie. This prospect is particularly problematic because while it is uncertain whether a mistake of fact excuses a mistaken use of cyber force in self-defense, there is no question that it does not excuse a mistake of the law, such as an error regarding the threshold for breach. And even though the threshold of harm necessary to trigger the right of self-defense is ambiguous, a State operating in the grey zone of normative uncertainty always risks the condemnation of other States. That autonomous capabilities might generate results that are somewhat less predictable than cyber operations not employing such capabilities only increases this risk.

Two additional uncertainties in the law of self-defense further complicate cyber operations involving autonomous capabilities. First, there is a longstanding debate as to whether States are entitled to resort to self-defense in the face of hostile operations at the armed attack level that were neither mounted by another State nor, in the words of the ICJ in the *Nicaragua* judgment, conducted "by or on behalf," or with the "substantial involvement" of, another State.[70] Although the better view is that the right of self-defense applies to armed attacks by non-State actors,[71] the ICJ has on two occasions confirmed the restrictive position it took in *Nicaragua*.[72] Should that approach prevail as a matter of law, those employing an autonomous capability, or the autonomous capability itself, would need to have the capacity to distinguish operations satisfying the conditions set forth by the Court from those that do not.

Second, this uncertainty relates directly to the "unwilling-unable" debate.[73] Assuming for the sake of analysis that self-defense is available against non-State actors, consider a case in which non-State actors are

---

70. Nicaragua, *supra* note 29, ¶ 195.

71. *Compare* Netherlands, Ministry of Foreign Affairs, *supra* note 13, at 9 (stating the right to self-defense applies to the actions of non-State actors), *with* FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 8 (stating that self-defense is only available in response to actions conducted "directly or indirectly" by a State).

72. Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion, 2004 I.C.J. Rep. 136, ¶ 139 (July 9); Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), Judgment, 2005 I.C.J. Rep. 168, ¶¶ 146–47 (Dec. 19).

73. TALLINN MANUAL 2.0, *supra* note 19, at 347–48.

operating from the territory of another State without the involvement of that State. May the victim State conduct cyber operations involving autonomous capabilities into the territorial State against the non-State actor without violating the territorial State's sovereignty or the rule of non-intervention?

It may not do so on the basis of countermeasures because they are unavailable in response to the operations of non-State actors, cyber or otherwise, that are not attributable to a State. Should the non-State actor's operations not affect an essential interest of the victim State in a grave and imminent manner neither would there be any basis to conduct the operation pursuant to necessity. And if cyber operations involving autonomous capability at the use of force level are needed to address the situation, neither countermeasures nor necessity allow for the use of force.[74] This leaves only self-defense as a possible circumstance precluding the wrongfulness of the cyber response to the non-State actor attacks.

There is substantial disagreement over whether self-defense may preclude the wrongfulness of the violation of sovereignty that would occur should the operation involving autonomy be launched on that basis into a State to which the operation cannot be attributed. Some are of the view that it cannot—that sovereignty is a veil pierceable only when the State concerned is considered under international law to have directly or indirectly launched the armed attack.[75] However, numerous States hold a less restrictive view, espousing the right of self-defense against a non-State actor in the territory of another State when the territorial State is either "unable or unwilling" to put an end to the hostile operations from its territory.[76] In light of this debate, States employing autonomous cyber capabilities into other States against non-State actors under a theory of self-defense run the risk of some States and scholars characterizing their operations as breaches of sovereignty, unlawful, intervention, and, perhaps, unlawful uses of force.

Finally, any use of an autonomous cyber capability on the basis of self-defense must comply with the requirements of necessity and proportionality that have been recognized by the ICJ and are uniformly accepted across the international community.[77] In the context of self-defense, necessity denotes

74. The possibility is expressly ruled out in the Articles on State Responsibility. *See* Articles on State Responsibility, *supra* note 2, art. 50(1)(a).

75. *See, e.g.*, FRANCE, MINISTÈRE DES ARMÉES, *supra* note 12, at 8.

76. The United States, for instance, has long held this position in the non-cyber context. *See, e.g.*, President Barack Obama, Remarks by the President at the National Defense University (May 23, 2013), https://obamawhitehouse.archives.gov/the-press-office/2013/05/23/remarks-president-national-defense-university.

77. Nicaragua, *supra* note 29, ¶¶ 176, 194; Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. Rep. 226, ¶ 41 (July 8); Oil Platforms, *supra* note

the requirement that there be no non-forcible means of dealing with the situation effectively, while proportionality refers to the requirement that no more force, cyber or non-cyber, be used than that which is required to end the armed attack. Defensive responses at the use of force level that employ autonomous capabilities, and the autonomous capabilities themselves, must be capable of making such calculations if self-defense is to operate as a circumstance precluding wrongfulness.

VIII. CONCLUSION

It seems to be *de rigueur* in international law circles to approach new technologies with grave concern. The rebuttable presumption seems to be that international law will fall short in adequately governing them. That was certainly the case with cyber operations. At the time the original Tallinn Manual project was launched in 2009, claims that cyberspace was a normative Wild West were frequent, and very much in vogue.[78] Yet, by the time of its publication in 2017, *Tallinn Manual 2.0*, drawing upon a diverse group of international law experts from around the world, had identified 154 consensus rules and agreed upon nearly 600 pages of commentary.[79]

This does not mean that there are no remaining challenges in the interpretation and application of the extant international law in the cyber context. Nevertheless, States are making significant progress in assessing how international law governs cyberspace, as illustrated by the work of the multiple U.N. Groups of Governmental Experts, the proceedings of the U.N. Open-Ended Working Group, and the number of statements on the subject that have been issued in the last two years.[80]

To some extent, autonomy and international law suffer the same dynamic. Initially, attention centered on lethal autonomous weapons systems, with battle lines drawn between those who would outlaw the systems and those who argued international humanitarian law suffices to govern them, primarily through the interpretive process that occurs with all

---

44, ¶¶ 43, 73–74, 76; *see also* TALLINN MANUAL 2.0, *supra* note 19, r. 72, at 348, and accompanying commentary, at 348–50.

78. Unfortunately, such claims continue to reappear. *See, e.g.*, Michael Schmitt, *Norm-Skepticism in Cyberspace? Counter-Factual and Counterproductive*, JUST SECURITY (Feb. 28, 2020), https://www.justsecurity.org/68892/norm-skepticism-in-cyberspace-counter-factual-and-counterproductive/.

79. *See* TALLINN MANUAL 2.0, *supra* note 19.

80. For a fuller discussion, see Michael N. Schmitt, *Taming the Lawless Void: Tracking the Evolution of International Law Rules for Cyberspace*, 3 TEXAS NATIONAL SECURITY REVIEW, at 32, 32 (Summer 2020).

new technologies of war.[81] Discussion of autonomy and international law beyond the topic of warfare has only just begun.

It would appear, however, that as with many other nascent technologies, at least with respect to the international law rules requiring respect for the sovereignty of other States and prohibiting intervention into their internal or external affairs, autonomy presents few challenges. Indeed, the normative architecture appears quite sound. While there are numerous unsettled issues surrounding application of these two primary rules to cyber operations, the fact that a cyber operation employs autonomous capability has little legal bearing on their resolution. Rather, autonomy simply makes it more difficult, at least at times, to confidently apply the rules because of uncertainty as to the consequences. Yet, these are dilemmas of fact, not law, and must be understood and acknowledged as such.

81. Michael N. Schmitt, *War, Technology, and International Humanitarian Law*, 82 INTERNATIONAL LAW STUDIES 137 (2006).