The compatibility of the use of autonomous weapons with the principle of precaution in the law of armed conflict

Elliot Winter*

Lecturer at Newcastle University Law School

The notion of 'precautions in attack' is an important concept in the law of armed conflict. It requires attackers to verify their targets, to ensure that only proportionate strikes are launched, to mitigate the potential for collateral damage in the selection of means and methods of warfare and, where possible, to issue warnings in advance of attacks. The emergence of 'autonomous weapons' requires us to review how these precautions might apply in a new era of delegated warfare in which the fulcrum around which the regime revolves – the human combatant – would be replaced by artificial intelligence-powered machines. This paper contends that technology is fast approaching the point at which the use of autonomous weapons will be compatible with the requirements of precaution. The author pushes further than most commentators in demonstrating just how capable the technology is becoming; in explaining how close it is to supplanting humans on the battlefield and in showing the potential for machines to discharge precautionary obligations better than humans in at least some areas. However, the article recognizes that technological innovations are still required before this transition comes to pass and so it also illuminates the way forward by identifying where remaining deficiencies linger.

Les «précautions dans l'attaque» constituent un concept important du droit des conflits armés. Ces précautions imposent aux attaquants de vérifier leurs cibles, de s'assurer de ne lancer que des frappes proportionnelles, de sélectionner les moyens et méthodes de guerre de sorte de mitiger le potentiel de dommages collatéraux, et, si possible, de donner des avertissements avant d'attaquer. Avec l'émergence des «armes autonomes», il nous faut toutefois examiner comment ces précautions pourraient s'appliquer dans une nouvelle ère de guerre déléguée où le combattant humain axe central régime serait remplacé par des machines dotées d'une intelligence artificielle. Cet article soutient que la technologie approche rapidement le stade où l'utilisation d'armes autonomes sera compatible avec les exigences de précaution. L'auteur va plus loin que la plupart des commentateurs en faisant la démonstration des nouvelles capacités de la technologie; en expliquant combien elle est proche de supplanter les humains sur le champ de bataille; et en montrant le potentiel des machines de s'acquitter des obligations de précaution mieux que les humains, dans certains domaines du moins. Toutefois, l'article reconnaît que des innovations technologiques sont encore nécessaires avant que cette transition ne puisse s'effectuer, et éclaire la voie à suivre en identifiant les points où des lacunes résiduelles subsistent.

* The author is a Lecturer at Newcastle University Law School (elliott.winter@newcastle.ac.uk). A debt is owed to Dr Elena Katselli and Dr Conall Mallory for their support.

Het begrip ‘voorzorg bij aanvallen’ is een belangrijk concept in het recht der gewapende conflicten. Van aanvallers wordt verwacht dat ze hun doelwitten controleren, ervoor zorgen dat ze enkel proportionele aanvallen inzetten, de mogelijke collateral schade beperken wanneer ze middelen en methoden van oorlogsvoering kiezen en, waar mogelijk, waarschuwingen geven vooraleer aan te vallen. De opkomst van ‘autonome wapens’ dwingt ons te onderzoeken hoe deze voorzorgsmaatregelen kunnen worden toegepast in een nieuw tijdperk van gedelegeerde oorlogsvoering waarbij de spil waarrond het regime draait – de menselijke strijder – zou worden vervangen door kunstmatige, door intelligentie aangedreven machines. In deze paper wordt betoogd dat de technologie snel het punt aan het naderen is waarop het gebruik van autonome wapens verenigbaar zal zijn met de eisen van het voorzorgsbeginsel. De auteur gaat verder dan de meeste commentatoren als hij aangeeft hoe bekwaam de technologie wel aan het worden is, als hij uitlegt hoe dicht we het punt naderen dat mensen op het slagveld worden vervangen en als hij het potentieel aantoont van machines om zich, minstens op sommige gebieden, beter dan mensen van voorzorgsverplichtingen te kwijten. Niettemin wordt in het artikel ook erkend dat er nog technologische innovaties noodzakelijk zijn vooraleer deze transitie definitief realiteit wordt en het geeft daarom ook de te volgen richting aan door te wijzen op de gebreken die de voortgang hinderen.

La noción de “precaución en el ataque” es un concepto importante en el Derecho de los Conflictos Armados. Ello requiere que los atacantes verifiquen los objetivos para asegurarse que tan solo se producen ataques proporcionados, mitigando así el potencial daño colateral en la selección de los medios y métodos de combate, además de, en la medida de lo posible, efectuar advertencias antes de los ataques. La emergencia de “armas autónomas” nos obliga a revisar el modo en que estas precauciones resultan de aplicación en una nueva era de la guerra por delegación, en la cual el punto de apoyo sobre el que descansa la normativa – el humano combatiente – sería reemplazado por artefactos impulsados mediante inteligencia artificial. Este artículo sostiene que la tecnología está alcanzando rápidamente un punto tal que el uso de armas autónomas será compatible con las exigencias del principio de precaución. El autor va mucho más allá que otros comentaristas en demostrar cuán capaz se está volviendo la tecnología; en explicar lo cerca que está de suplantar a los humanos en el campo de batalla y en exponer el potencial de estos artefactos para cumplir las obligaciones derivadas de la precaución mejor que los humanos, al menos en algunas áreas. Sin embargo, el artículo reconoce que aun hacen falta innovaciones tecnológicas antes de que se produzca la transición y, por ello mismo, ilumina el camino a seguir identificando aquellas deficiencias que aún persisten.

La nozione di “precauzione nell’attacco” è un importante concetto della Legge sui Conflitti Armati. Tale Legge richiede agli attaccanti l’obbligo di verificare gli obiettivi, di garantire che vengano lanciati solo attacchi proporzionati, di mitigare il potenziale dei danni collaterali nella scelta dei mezzi e dei metodi di guerra e, dove possibile, di inviare avvertimenti prima degli attacchi. L’emergere delle “armi autonome” ci impone di rivedere come queste precauzioni potrebbero essere applicate in una nuova era di guerra delegata, nella quale il fulcro intorno al quale ruota il regime – il combattente umano – verrebbe sostituito da macchine alimentate dall’intelligenza artificiale. Questo documento sostiene che la tecnologia si stia rapidamente avvicinando al punto in cui l’uso delle armi autonome sarà compatibile con i requisiti di precauzione. L’autore si spinge più in là della maggior parte dei commentatori nel dimostrare quanto la tecnologia stia diventando capace; nello spiegare quanto sia vicina a soppiantare gli esseri umani sul campo di

battaglia e nel dimostrare il potenziale delle macchine nell'adempire agli obblighi di precauzione meglio degli esseri umani almeno in alcune aree. Tuttavia, l'articolo riconosce come siano necessarie ancora delle innovazioni tecnologiche prima che avvenga questa transizione così come identifica la strada da seguire dove persistono carenze residue.

Der Begriff 'Vorsichtsmaßnahmen beim Angriff' ist ein wichtiges Konzept im Recht der bewaffneten Konflikte. Dieses Konzept fordert von Angreifern, ihre Ziele zu prüfen, dafür zu sorgen, dass nur verhältnismäßige Angriffe durchgeführt werden, den möglichen Kollateralschaden bei der Auswahl der Mittel und Methoden der Kriegsführung zu beschränken und, wenn möglich, im Voraus Warnungen vor Angriffen zu geben. Das Aufkommen 'autonomer Waffen' nötigt uns, zu überprüfen, wie diese Vorsichtsmaßnahmen angewandt werden sollen in einer neuen Ära delegierter Kriegsführung, in der der Angelpunkt, um den das System dreht – der menschliche Kombattant – ersetzt würde durch Maschinen, die durch künstliche Intelligenz angetrieben werden. Der Autor behauptet, dass sich die Technologie rasch dem Punkt nähert, an dem die Verwendung autonomer Waffen mit den Vorsichtsanforderungen vereinbar sein wird. Der Autor geht weiter als die meisten anderen Kommentatoren, wenn er verdeutlicht, wie tauglich die Technologie wird, wenn er erklärt, wie nahe sie daran ist, Menschen auf dem Schlachtfeld zu ersetzen und wenn er darlegt, dass Maschinen die Vorsichtspflichten, jedenfalls in gewissen Bereichen, besser erfüllen können als Menschen. Der Autor erkennt allerdings, dass noch weitere technologische Innovationen erforderlich sind, bevor sich dieser Wandel vollzieht, und beleuchtet somit auch die künftige Vorgehensweise, indem er bestimmt, wo es noch Unvollkommenheiten gibt.

Keywords: law of armed conflict, precautions in attack, autonomous weapons, artificial intelligence

INTRODUCTION

The law of armed conflict (LOAC) regulates the conduct of hostilities and protects people in situations of both international and intranational violence. The principal legal instruments are Geneva Conventions I,¹ II,² III³ and IV⁴ (and their Additional Protocols I⁵ and II⁶). At present, there is much debate on whether the use of 'autonomous weapons' – machines that are capable of waging war independently

1. Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 31.

2. Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 85.

3. Convention (III) Relative to the Treatment of Prisoners of War (adopted 12 August 1949, entered into force 21 October 1950) 75 UNTS 135.

4. Convention (IV) Relative to the Protection of Civilian Persons in Time of War (adopted 12 August 1949, entered into force 21 October 1950) 78 UNTS 287.

5. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 3.

6. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the

from humans after deployment – would be compatible with various aspects of LOAC. In fact, autonomous weapons have been a hot topic in LOAC for some years now. Much has been written on the ‘humanity’ of their use,⁷ or on who may be held accountable when things go wrong.⁸ The author has written on matters such as how ‘killer robots’ might interact with the principle of proportionality⁹ and on the relevance of the principles of humanity and military necessity¹⁰ to the debate. The primary focus of the present article will be the less-discussed matter of the interaction of autonomous weapons with ‘precautions in attack’ as found in Additional Protocol I.¹¹ There are related rules on ‘precautions *against the effects* of attacks’ (i.e. ‘precautions in defence’).¹² However, these will not be considered here as the focus is on machines with offensive capabilities rather than on the behaviour of the people they may target and because the issue has been explored elsewhere.¹³ For the remainder of the article, ‘precautions’ (or sometimes ‘precaution’ for expediency) will be used as shorthand to mean precautions in attack only. In terms of substance, precaution is designed to facilitate the more fundamental principles of LOAC such as distinction and proportionality and so it shares the underlying logic that, while death and destruction are inevitable in war, they should be minimized as much as possible. To this end, as will be explained more fully below, precaution comprises discrete components on verification, proportionality, mitigation and warnings.

The topic of precaution in the context of autonomous weapons is worthy of study because, as Boothby put it, ‘novel weapon technologies ... render it appropriate to consider taking particular precautions that are additional to those prescribed by existing law [or mean] that certain precautions take on a special importance [or] require entirely new kinds of precaution to be considered’.¹⁴ In other words, precaution is a dynamic concept that can mould around and accommodate technological advances provided we give thought to the matter beforehand. Given that autonomous weapons represent a ‘paradigm shift’¹⁵ in military hardware (because unlike any existing offensive weapon they would not require direction from a human after deployment) the degree of dynamism necessary to keep up with them is likely to be

Protection of Victims of Non-International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1978) 1125 UNTS 609.

7. O Ulgén, ‘Human Dignity in an Age of Autonomous Weapons: Are We in Danger of Losing an “Elementary Consideration of Humanity”?’ (2017) 17 *Baltic Yearbook of International Law* 167.

8. D Amoroso and B Giordano, ‘Who Is to Blame for Autonomous Weapons Systems’ Misdoings?’ in E Carpanelli and N Lazzarini (eds), *Use and Misuse of New Technologies: Contemporary Challenges in International and European Law* (Springer, Cham 2019) 211.

9. E Winter, ‘Autonomous Weapons in Humanitarian Law: Understanding the Technology, Its Compliance with the Principle of Proportionality and the Role of Utilitarianism’ (2018) 6(1) *Groningen Journal of International Law* 183.

10. E Winter, ‘Pillars not Principles: The Status of Humanity and Military Necessity in the Law of Armed Conflict’ (2020) 25(1) *Journal of Conflict and Security Law* 1.

11. Additional Protocol I (n 5) at Art 57.

12. *Ibid* at Art 58.

13. E Talbot Jensen, ‘Precautions Against the Effects of Attacks in Urban Areas’ (2017) 98(901) *International Review of the Red Cross* 147.

14. WH Boothby, ‘The Law on the Conduct of Hostilities’ in WH Boothby (ed), *New Technologies and the Law in War and Peace* (Cambridge University Press, Cambridge 2019) 54.

15. International Committee of the Red Cross, ‘Autonomous weapon systems – Q&A’ (*ICRC*, 12 November 2014), available at <<http://bit.ly/2ixib2p>> (accessed 3 October 2020).

especially high and as much participation in the debate as possible is to be encouraged. The thrust of the argument in this article is somewhat nuanced. Both autonomous weapons and precaution are complex matters so to say that the use of the former would or would not be compatible with the latter would be an oversimplification. Rather, it will be argued that the use of autonomous weapons is likely to be compatible with precaution as a whole in the near future, but that compatibility will be achieved with some components sooner than others. Thus, with sufficient advances in artificial intelligence and sufficient clarification of certain areas of the law, we are likely to see the emergence of autonomous systems that demonstrate better compatibility with all aspects of precaution than human combatants employing other means of warfare. Inevitably, the piece is somewhat speculative and numerous caveats are made along the way. It is also recognized that this argument, while limited to precaution alone, is controversial. Nonetheless, the author believes this to be the direction of travel.

In terms of structure, the article has six numbered sections. First, there is an overview of precaution dealing with preliminary matters such as its origins, components, legal status and ambit. Second, there is an overview of autonomous weapons which deals with the definition of the technology, the viability of its deployment to the battlefield and its current treatment in international fora. Third, there is treatment of the extent to which autonomous weapons could handle the verification component of precaution, both in the planning and execution stages of an attack. Fourth, there is consideration of how machines could discharge the proportionality component of precaution, again both in the planning and execution stages of a strike. Fifth, the article will consider the ability of autonomous weapons to work with the mitigation limb of precaution in the sense of determining, based on their programming, appropriate means and methods of attack and which targets would render the least collateral damage. Sixth, and finally, there is consideration of how autonomous weapons might discharge the requirement to deliver, where possible, advanced warning of an attack.

1 AN OVERVIEW OF PRECAUTIONS IN ATTACK

It seems sensible to begin with a detailed overview of the duty to take precaution before considering autonomous weapons and then how the two phenomena interact. The underlying objective of precaution is, according to Additional Protocol I, to ensure that ‘constant care [is] taken to spare the civilian population, civilians and civilian objects’.¹⁶ In this sense, it is a corollary of the high-level principles of distinction and proportionality (the latter, as well as a principle, is also a rule forming part of precaution) which set out the basic obligations to discern civilians from combatants¹⁷ and to minimize collateral damage¹⁸ respectively. The duty to take precautions renders distinction and proportionality more concrete by fleshing them out with additional procedural rules, thus facilitating their real-world implementation.

These procedural rules are relatively complex but precaution can generally be distilled down to four principal components. First, there is ‘verification’ in the sense of checking that targets are legitimate and do not benefit from any form of special

16. Additional Protocol I (n 5) at Art 57(1).

17. *Ibid* at Art 48.

18. *Ibid* at Arts 51(5)(b) and 57(2)(a)(iii).

protection.¹⁹ Verification applies during the planning stage of an attack²⁰ and also during the execution stage such that it is necessary to be able to react to changing circumstances by cancelling or suspending an attack if necessary.²¹ In both of these stages, attackers are expected to do ‘everything feasible’ to ensure verification determinations are accurate.²² Second, there is ‘proportionality’ in the sense of ensuring that the military advantage expected from an attack will outweigh the collateral damage it is anticipated to cause. Again, this applies during both the planning stage of an attack²³ and also during execution with the result that an attack must be cancelled or suspended if circumstances change rendering this necessary.²⁴ Third, there is ‘mitigation’ which demands that choices of means and methods of warfare are designed to avoid or at least minimize collateral damage.²⁵ Mitigation also requires that, when different target options are available of broadly equal military value, the option carrying the least collateral damage must be selected.²⁶ Fourth, there is ‘warning’ which simply requires giving advanced notice of an impending attack, where possible, to allow civilians and other protected persons to evacuate the area.²⁷ Note that this four-part breakdown is favoured by the author for research and teaching purposes and is not necessarily reflective of the approach taken elsewhere. Dinstein, for example, organizes his exposition around specific practical contexts such as ‘high-altitude bombing raids’ and the use of ‘precision-guided’ or ‘directed energy’ munitions.²⁸

In terms of status, precaution is typically regarded as a customary rule of international law. The International Committee of the Red Cross (ICRC) confirmed this in its customary law study and notes that ‘State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts’.²⁹ Similarly, according to Boothby, ‘many of the precautionary rules ... are generally accepted by States in the terms in which they are drafted, thus have customary status and are binding on all States’.³⁰ There are some minor exceptions when it comes to the finer detail of how precaution is to be interpreted and those exceptions and ‘clarifications’ will be considered below in the context of specific component rules of precaution. More generally, Kalshoven argues that precaution is ‘essentially relative in nature, as situations may arise where civilians simply cannot be spared’.³¹ Notwithstanding these caveats, it can still be assumed that precaution is indeed a rule of customary international law and that the content of the customary law rule matches the content of the rule as set out in Additional Protocol I.

19. Ibid at Art 57(2)(a)(i).

20. Ibid at Art 57(2)(a)(i).

21. Ibid at Art 57(2)(b).

22. Ibid at Art 57(2)(a)(i).

23. Ibid at Art 57(2)(a)(iii).

24. Ibid at Art 57(2)(b).

25. Ibid at Art 57(2)(a)(ii).

26. Ibid at Art 57(3).

27. Ibid at Art 57(2)(c).

28. Y Dinstein, *The Conduct of Hostilities under the Law of International Armed Conflict* (3rd edn, Cambridge University Press, Cambridge 2016) 164–73.

29. J-M Henckaerts and L Doswald-Beck, *Customary International Humanitarian Law, Volume I: Rules* (Cambridge University Press, Cambridge 2005) 51.

30. Boothby (n 14) at 54.

31. F Kalshoven, *Reflections on the Law of War: Collected Essays* (Martinus Nijhoff, Leiden 2007) 546.

In terms of geographical ambit, precaution applies (along with LOAC more generally) to the '(i) land areas; (ii) internal waters; (iii) archipelagic waters; (iv) territorial sea; (v) subsoil and submarine areas underneath these expanses of land and water, as well as the continental shelf; and (vi) the superjacent airspace'³² of the relevant states in international armed conflicts and to 'the whole territory under the control of a party' in non-international armed conflicts.³³ Moreover, it is generally accepted by experts today that precaution applies no matter whether the combat arena happens to fall on land, in the sea or the air. The Harvard Program on Humanitarian Policy and Conflict Research Manual (HPCR Manual) opines that 'the same legal regime applies equally in all domains of warfare'.³⁴ When it comes to drones (remotely piloted weapon platforms), the HPCR Manual notes that LOAC 'has no provisions that are specific to UAV/UCAV [and so] the general requirements to take feasible precautions apply'.³⁵ The same point applies, *a fortiori*, in the context of the use of autonomous weapons – there is no specific provision for precaution in the context of such weapons, so the general regime applies. As Schmitt put it, 'each of these [precautionary] obligations applies fully to the use of autonomous weapon systems'.³⁶ In short then, the entire precaution in attack regime would apply to autonomous weapons no matter what forms of hardware they might adopt or in which contexts they might serve.

2 AN OVERVIEW OF AUTONOMOUS WEAPONS

Having considered an overview of precaution in the abstract, all that remains to set the scene is to supply an overview of autonomous weapons. In 2012, Human Rights Watch (HRW) defined autonomous weapons as 'robots that are capable of selecting targets and delivering force without any human input or interaction'.³⁷ In the same year, the United States (US) adopted a similar working definition providing that an autonomous weapon is a 'weapon system that, once activated, can select and engage targets without further intervention by a human operator'.³⁸ Since those early days, international effort has been directed towards finding a suitable definition for autonomous weapons, principally under the auspices of the Convention on Conventional Weapons (CCW).³⁹ A Group of Governmental Experts (GGE) was formed to consider the matter but it has so far failed to produce an agreed definition. This may be

32. Y Dinstein, *War, Aggression and Self-Defense* (3rd edn, Cambridge University Press, Cambridge 2001) 19.

33. *Prosecutor v Dusko Tadic* (Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction), ICTY-94-1-A (02 October 1995) at para 70.

34. Program on Humanitarian Policy and Conflict Research at Harvard University, *HPCR Manual on International Law Applicable to Air and Missile Warfare* (Cambridge University Press, Cambridge 2013) 142.

35. *Ibid* at 156.

36. MN Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) *Harvard National Security Journal* 23, available at <<https://bit.ly/2WCnpwb>> (accessed 3 October 2020).

37. Human Rights Watch, 'Losing Humanity: The Case Against Killer Robots' (*Human Rights Watch* 2012) 2, available at <<https://bit.ly/2z17EFm>> (accessed 3 October 2020).

38. US Department of Defense, 'Autonomy in Weapons Systems' (2012) Directive 3000.09, Glossary Part II, available at <<https://bit.ly/2UCP4fc>> (accessed 3 October 2020).

39. Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons

understandable given that, as McFarland notes, ‘autonomous capability is a continuously varying phenomenon’.⁴⁰ The lack of a definition is unfortunate, however it is not overly concerning for present purposes as autonomous weapons are envisaged here as something along the lines suggested by both HRW and the US.

Of course, those definitions set out the theory of what an autonomous weapon might be, but they do not address whether this sort of technology actually exists or might exist. We are unlikely to see the deployment of autonomous war robots imminently. As yet, there is no known example of an autonomous, offensive, weapon owing to technological limitations. So far, states have made it clear that they intend to keep humans ‘in’ or ‘on’ the loop in the sense that humans would always be required to authorize a critical strike or to supervise critical operations with the ability to step in and countermand strikes. Consequently, for the immediate future, it seems that development will be limited to weapons with limited autonomous functions such as navigation while in practice humans will take over when it comes to the ‘critical’ (i.e. potentially life and death) decisions such as those made in the context of precaution. As Boothby said ‘in the nearer future, personnel controlling the systems ... will have to be able to take sufficient precautionary steps ... if, as is being assumed, the weapon system itself is not capable of applying the prescribed precautions’.⁴¹ This sort of limited, dependent, (semi) autonomous weapon is envisaged by Suchman who writes that these machines will possess only ‘autonomous characteristics’ and not full autonomy.⁴² This is also what Van Rompaey has in mind when he criticizes the tendency to perceive autonomous weapons as a ‘replacement for human soldiers’⁴³ and argues instead that they will merely form part of broader ‘network-centric sociotechnical systems’.⁴⁴ Similarly, McFarland foresees that autonomous weapons are likely to operate ‘at more than one “level” of autonomy simultaneously’ with different amounts of independence conferred for matters such as ‘planning, navigation or weapon release’.⁴⁵

While it is conceded that autonomous weapons with *offensive* capabilities are not a likely prospect in the immediate term, it should be highlighted that autonomous systems with a purely *defensive* role already exist such as Israel’s missile defence shield, ‘Iron Dome’.⁴⁶ Furthermore, a number of offensive autonomous systems are in development such as ‘Taranis’ – an aerial combat vehicle being developed by BAE

Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects (adopted 10 October 1980, entered into force 02 December 1983) 1342 UNTS 137.

40. T McFarland, ‘Factors Shaping the Legal Implications of Increasingly Autonomous Military Systems’ (2015) 97 *International Review of the Red Cross* 1313, 1321.

41. WH Boothby, ‘Highly Automated and Autonomous Technologies’ in WH Boothby (ed), *New Technologies and the Law in War and Peace* (Cambridge University Press, Cambridge 2019) 147.

42. L Suchman and J Weber, ‘Human-Machine Autonomies’ in N Bhuta and others (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, Cambridge 2016).

43. L Van Rompaey, ‘Shifting from Autonomous Weapons to Military Networks’ (2019) 10 *Journal of International Humanitarian Legal Studies* 111, 115.

44. *Ibid* at 111.

45. McFarland (n 40) at 1322.

46. Raytheon Missiles & Defense, ‘Iron Dome System and SkyHunter Missile’ (*Raytheon Missiles & Defense*), available at <<https://bit.ly/3dTYTNz>> (accessed 3 October 2020).

Systems (a UK-based aerospace manufacturer) with certain autonomous capabilities.⁴⁷ The viability of autonomous weapons increases further still as we look to the long term. Ekelhof argues that ‘intelligence is of the greatest value when humans and technology join forces’.⁴⁸ However, there are huge tactical incentives for states to automate warfare as far as possible. A chain is only as strong as its weakest link and, for as long as humans remain in the chain, weapons systems will be slow. Also, for as long as humans control autonomous weapons, they will be susceptible to jamming and to hacking (whereby communications between the weapon and its handlers are blocked or hijacked). If autonomous weapons can operate independently, their isolation from external stimuli will reduce, if not eliminate, these vulnerabilities. As a result of these consequences, Baker observed that ‘major powers appear reluctant to impose any limits on these systems beyond what international law already prescribes’.⁴⁹ In parallel with the interests of states, technology companies also have huge incentives to develop technologies that would lend themselves to autonomous warfare. The author has previously argued that three core capabilities must be acquired for autonomous weapons to function in an independent manner: machine observation, machine recognition and machine judgement.⁵⁰ As will be seen below, great strides have been made in each of these areas. When combined, these ‘push’ factors behind autonomous weapons mean that, as Schmitt summarized, ‘at some point in the future, [they] will find their way into the battlespace’.⁵¹ Similarly, while of the view that autonomous weapons will remain under the control of humans for the medium to long term, Ekelhof nevertheless concedes that ‘it could very well be that, if a major conflict arises, all bets will be off, with states feeling forced into more reliance on autonomous systems’.⁵² The consequence of this is the inevitability of autonomous weapons. With that in mind, academic focus is best directed towards how LOAC might be interpreted to ensure the robust protection of civilians and other persons once deployments begin to occur.

The rise of autonomous weapons need not be a negative development for LOAC. Organizations such as HRW argue that ‘killer robots’ ‘should be banned and [that] governments should urgently pursue that end’⁵³ and that ‘such robots would appear to be incapable of abiding by key principles of international humanitarian law’.⁵⁴ However, others have a more positive outlook. Boothby stated that autonomous weapons ‘with further development ... may ... meet or even exceed [applicable] legal criteria’.⁵⁵ Similarly, Anderson and Waxman take the view that the technology might prove to be more precise and less harmful than existing approaches to warfare.⁵⁶ The

47. BAE Systems, ‘Taranis’ (*BAE Systems*), available at <<http://bit.ly/2uamk2j>> (accessed 3 October 2020).

48. M Ekelhof, ‘Lifting the Fog of Targeting’ (2018) 71(3) *Naval War College Review* 61, 77.

49. D-P Baker and others, ‘Introducing Guiding Principles for the Development and Use of Lethal Autonomous Weapon Systems’ (2020) *E-International Relations*, available at <<https://bit.ly/3bngewB>> (accessed 3 October 2020).

50. E Winter, ‘The Compatibility of Autonomous Weapons with the Principle of Distinction in the Law of Armed Conflict’ (2020) 69(4) *International and Comparative Law Quarterly* 845–46.

51. Schmitt (n 36) at 5.

52. Ekelhof (n 48) at 84.

53. Human Rights Watch (n 37) at 2.

54. *Ibid* at 30.

55. Boothby (n 41) at 159.

56. K Anderson and MC Waxman, ‘Law and Ethics for Autonomous Weapon Systems: Why a Ban Won’t Work and How the Laws of War Can’ (2013) *Stanford University The Hoover*

remainder of the piece will explore these areas of difficulty, and of potential, in the context of the various components of precaution in attack.

3 THE INTERSECTION OF PRECAUTION AND AUTONOMOUS WEAPONS – VERIFICATION

Now that the two main ingredients of this debate (precaution and autonomous weapons) have been considered in abstract, it is possible to consider how they interact. As explained, there are four components to precaution and the first is ‘verification’, so it is prudent to begin there. According to Additional Protocol I, those who plan or decide upon attacks must ‘do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection but are military objectives’.⁵⁷ Given that verification is so firmly grounded in the requirement to distinguish civilian persons and objects from military persons and objects, it is to some extent a reiteration of the distinction principle narrated elsewhere in LOAC.⁵⁸ However, verification builds on distinction with tighter procedural guidance on how it is to be implemented and, in this sense, it adds important texture to the business of targeting. Specifically, verification splits into two substantive limbs: checking that targets are legitimate objectives in the *planning* stage and checking that targets remain legitimate objectives in the *execution* stage. It is necessary to do ‘everything feasible’ during the commission of these checks to ensure that they are carried out correctly. These elements of verification will now be considered in turn.

Regarding verification at the ‘planning stage’, in the context of ‘normal’ means of warfare (guns, missiles, mortars and so on) we tend to think of this as occurring in a discrete space somewhere, say a room or inside a military vehicle, with a decision being taken to perhaps launch a missile – whereas we imagine that the ‘execution stage’ begins thereafter with ongoing assessments made to determine if the missile should continue or be aborted or redirected. However, when it comes to autonomous weapons, we need to rethink this approach and recognize that two sub-components to planning will emerge. First, there will be the ‘human-centric’ phase of the planning stage which will look familiar and which will continue to involve military personnel, albeit in a new role of setting policy objectives and feeding machines with information, weapons and fuel. Here, people will make decisions and upload decisions made at higher levels by others (with some adjustment for local factors) about what is to be targeted and what the priorities should be. Second, after their release onto the battlefield and when in autonomous mode, there will be the ‘machine-centric’ phase of the planning stage where autonomous weapons will scout for targets, find them and proceed to plan (by making determinations based on their programming from the first stage) which targets to attack, when, how, using which vector, in what order and so on. In this piece, the focus is on the unfamiliar machine-centric aspect of planning.

The real question here is what verification in the ‘machine centric’ phase of planning might involve. In 2016 the North Atlantic Treaty Organization (NATO) set

Institution (Jean Perkins Task Force on National Security and Law Essay Series), available at <bit.ly/2pBFnO9> (accessed 3 October 2020).

57. Additional Protocol I (n 5) at Art 57(2)(a)(i).

58. Ibid at Art 48.

out its approach on targeting in its Joint Doctrine for Joint Targeting document.⁵⁹ It described the ‘joint targeting cycle’ as comprising ‘determining the effects necessary to achieve the commander’s objectives; identifying the actions necessary to create them based on the means available; selecting and prioritizing targets; synchronizing capabilities; and then assessing their cumulative effectiveness, taking remedial action if necessary’.⁶⁰ The NATO doctrine was not designed with autonomous weapons in mind of course – indeed it does not mention the phenomenon at all – but it was written at a point by which war machines were firmly on the radar. Perhaps then it should be unsurprising that the targeting cycle might lend itself to machine computation in the sense that it has been carefully dissected into its constituent elements. At the very least, it begs the question of whether the cycle could be hived across into autonomous weapons for the purposes of verification-related precaution. The extent to which that is possible depends on the detail of NATO’s targeting cycle.

The principles of NATO’s process are listed as ‘objective-based’, ‘effects-driven’, ‘multidisciplinary’, ‘timeliness’, ‘centrally controlled and coordinated’, ‘information-accessibility’ and ‘security’.⁶¹ A few of these are worth unpacking. ‘Centrally-controlled’ might seem to preclude autonomous weapons as they are, by definition, not centrally controlled. However, that is not the case. This is because it is explained that although targeting *policy* is to be set at a central level, authority for the implementation of policy, including planning for specific targets, ‘is delegated to the lowest practicable level’.⁶² It is quite right of course that targeting policy is determined at the human-centric phase due to its importance, complexity, political sensitivity, the need to avoid friendly fire and duplication and so on. However, when it comes to *implementing* targeting policies, it is submitted that the autonomous weapons would sit at the ‘lowest practicable level’ and therefore would be best placed to apply the policy on the ground. ‘Effects-driven’ in the NATO cycle means ‘contributing to creating synchronized, measurable physical and psychological effects intended to achieve the [mission] objectives’.⁶³ Synchronicity and measurements of this sort would seem to align well with computer guided warfare where co-ordination and precise calibrations for attacks can be achieved with high accuracy and in minute detail. Finally, ‘timeliness’ is required by the cycle and is key because ‘targeting is often time critical [hence] it is ... fundamental that transferring information from source to user is as direct and as fast as possible’.⁶⁴ Speed is an edge that machines have had over humans for decades, regardless of whether they are acting on information they themselves have gathered or on data that have been relayed to them by an external source.⁶⁵ In short, it seems that the use of autonomous weapons could theoretically be compatible with the verification in planning component of precaution by building upon a targeting doctrine such as that employed by NATO.

Of course, it is accepted that moving from the theory of machine-centric verification and towards its practice will present many difficulties. For example, as Thurnher notes, the environment in which the machine might operate plays a major role and so

59. North Atlantic Treaty Organization, ‘Allied Joint Doctrine for Joint Targeting’ (2016) AJP-3.9, available at <<https://bit.ly/3bD23Us>> (accessed 3 October 2020).

60. Ibid at para 0104.

61. Ibid at para 0115.

62. Ibid.

63. Ibid.

64. Ibid.

65. S Raviv, ‘The Secret History of Facial Recognition’ (*Wired*, 21 January 2020), available at <<https://bit.ly/2U4jYxa>> (accessed 3 October 2020) at para 49.

systems ‘deployed to a remote area generally devoid of civilians, such as underwater or a desert environment’ would have a much easier time of complying with verification in the planning stage than one deployed ‘in urban or other populated areas’.⁶⁶ This is undoubtedly correct and such machines would need to have highly sophisticated capabilities to complete the task. However, as the author has noted elsewhere, these capabilities are already being developed.⁶⁷ In terms of the raw ability to ‘see’, the US defence manufacturer Raytheon has developed ‘mapping autonomous drones’ that are able to fully perceive their surroundings without access to Global Positioning System (GPS) or mapping data.⁶⁸ More specifically, it has developed ‘a fully autonomous aerial robot, that ... can operate in GPS-denied environments to map dense urban environments in 3-D’⁶⁹ using ‘a combination of sensors, including cameras and lidar [which is] similar to radar, but using pulsed, infrared laser light’.⁷⁰ Beyond simply ‘seeing’, there would be a requirement for machines to integrate and operationalize the huge amounts of data they collect. The HPCR Manual notes that the current requirement is that, to enable proper target verification, ‘command echelons must utilize all technical assets (such as intelligence, reconnaissance and surveillance systems) at their disposal, to the extent that these assets are reasonably available’.⁷¹ Similarly, it notes that ‘verification must be based on sources of information that are sufficiently up-to-date and reliable, according to sound military practice’.⁷² Computers are already very good at doing this sort of task – taking huge amounts of information from many sources in a short space of time and knitting it all together to build a better picture of an environment is their forte. Security systems such as the PatScan Platform – essentially an automated security guard – can already carry out this task in real-time using a combination of sensors from visual and magnetic to particle detection.⁷³ Similarly, Meiya Pico has developed a system that can detect languages and symbols embedded in pictures that might one day be useful in war for the purpose of identifying an ‘enemy’ language on uniforms or military equipment based on reconnaissance imagery.⁷⁴ This task also involves complex work in the sense of analysing all the visual cues in an environment and comparing them to a vast database of information to find matches, all in limited time. These capabilities, with adaption and refinement, could be transposed to the battlefield.⁷⁵

In summary then, autonomous weapons seem to have much practical (as well as theoretical) potential in terms of verifying targets at the planning stage of an attack. More than that, autonomous weapons are likely to become essential to make sense of the huge amounts of data now gathered on the battlefield because, as Ekelhof notes,

66. JS Thurnher, ‘Feasible Precautions in Attack and Autonomous Weapons’ in W Heintschel von Heinegg, R Frau and T Singer (eds), *Dehumanization of Warfare: Legal Implications of New Weapon Technologies* (Springer, Cham 2018) 108.

67. Winter (n 50).

68. Airsoc, ‘Where Hazards Lurk’ (*Airsoc*, January 2020) para 5, available at <<https://bit.ly/2Tzvcb5>> (accessed 3 October 2020).

69. Ibid at para 3.

70. Ibid at para 6.

71. HPCR (n 34) at 144.

72. Ibid.

73. Patriot One Technologies, ‘About’ (*Patriot One Technologies*), available at <<https://bit.ly/3dT3UWH>> (accessed 3 October 2020).

74. P Li and C Cadell, ‘At Beijing Security Fair: An Arms Race for Surveillance Tech’ (*Reuters*, 28 May 2018), available at <<https://reut.rs/2RCJPuJ>> (accessed 3 October 2020).

75. Winter (n 50).

‘although the massive increase in data available might seem a positive development, the positive results remain limited if the data cannot be processed for use’.⁷⁶

The precautionary obligation of verification applies not only in the planning stage of an attack but also in the execution stage, with Additional Protocol I requiring that an attack shall be cancelled or suspended ‘if it becomes apparent that the objective is not a military one or is subject to special protection’.⁷⁷ This additional obligation exists because, as the HPCR Manual notes, ‘the information available ... on the scene may be different from the information on which command echelons at a distance from the scene are basing their decisions’ and because ‘real-time observations [may] negate the information on which the decision to attack was based’.⁷⁸ In other words, the human or machine at the scene is simply closer to events and thus is generally in possession of better information than any other actor involved in the operation and so is better placed to tell if the facts have changed in such a way that the attack is no longer legitimate or desirable.

The question for present purposes is whether autonomous weapons would be able to handle verification in the execution stage. Many are of the clear view that they could not. HRW argues that ‘fully autonomous weapons would not have the ability to sense or interpret the difference between soldiers and civilians, especially in contemporary combat environments’.⁷⁹ The organization gives a couple of reasons for this. For example, it cites the fact that autonomous weapons would not be able to verify targets in situations where ‘combatants are not identifiable by physical markings’.⁸⁰ Similarly, it argues that individuals’ intentions are often relevant to their status (such as in the context of individuals who directly participate in hostilities) whereas ‘fully autonomous weapons would not possess human qualities necessary to assess an individual’s intentions’.⁸¹ Sharkey affirms this point, opining that ‘humans understand one another in a way that machines cannot ... cues can be very subtle, and there are an infinite number of circumstances where lethal force is inappropriate’.⁸² Guarini and Bello paint a vivid example when they give a scenario where, on the approach of an autonomous weapon, a scared mother runs to her children and screams at them to stop playing with pretend guns. They contend that a machine would simply see a person running towards it as well as two armed individuals and so attack, whereas a human would realise there is no threat and hold fire.⁸³ McFarland sums up the point by noting that autonomous weapons ‘may not be trusted to release a weapon autonomously’.⁸⁴ These concerns are valid, but there is doubt as to whether they would materialize in practice. In fact, machines might be able to perform verification in the execution stage of an attack quite satisfactorily and perhaps better than people. There are four principal reasons for suggesting that this may be the case.

The first reason for arguing that machines have the potential to surpass humans in the execution stage of verification is that – as Ekelhof noted – intelligence will always

76. Ekelhof (n 52) at 77.

77. Additional Protocol I (n 5) at Art 57(2)(b).

78. HPCR (n 34) at 144.

79. Human Rights Watch (n 37) at 30.

80. Ibid at 31.

81. Ibid.

82. N Sharkey, ‘Killing Made Easy’ in P Lin, K Abney, and GA Bekey (eds), *Robot Ethics: The Ethical and Social Implications of Robotics* (MIT Press, Cambridge, MA 2012) 118.

83. M Guarini and P Bello, ‘Robotic Warfare: Some Challenges in Moving from Noncivilian to Civilian Theaters’ in Lin, Abney, and Bekey (eds) (n 82) at 130.

84. McFarland (n 40) at 1322.

‘play a role in each phase of the targeting process’⁸⁵ and artificial intelligence could afford autonomous weapons a high degree of the necessary contextual sensitivity (with the effect that they could recognize the predicament of the mother and children above). According to Boothby, advances in artificial intelligence may ‘enable an autonomous system to react appropriately, for example, to the absence [of], or presence of, particular kinds of target at a specific location scheduled for an attack by refraining from the planned attack’.⁸⁶ This is probably true given that, even decades ago, computers were beginning to demonstrate the ability to judge context better than humans (albeit within limited parameters). An early example of this in a military context arose in 1988 when a US warship downed a civilian Iranian plane. The American ship’s systems correctly identified the plane as ascending away from the area, but human error resulted in a flawed assessment that the plane was descending in an attack vector and so it was fired upon. If the verification assessment had been left to the computer systems alone, the tragedy would have been avoided.⁸⁷ This lends credence to Schmitt’s observation that ‘such tragedies demonstrate that a man in the loop is not a panacea’.⁸⁸ In short, as Ekelhof summarized, autonomous weapons would simply need to be ‘taught what “normal” looks like’ in order to spot what is abnormal.⁸⁹ The second reason is that machines would not be encumbered by fear in the heat of battle and make the ‘fog of war’ induced errors that humans sometimes do. As Thurnher noted, ‘whereas human judgement is often clouded by anger, revenge, fear, and fatigue, autonomous machines may offer the promise of removing those worries from the battlefields of the future’.⁹⁰ Similarly, for Boothby, ‘robotic technologies will not be distorted by fear, anger, vengeance, amnesia, tiredness or other peculiarly human fallibilities’.⁹¹ Of course, organizations such as HRW would argue the flip side – that machines are not elevated by compassion either. However, on balance, it seems more likely that fear is the prevalent emotion in the heat of battle and that its removal offers a net benefit when weighed against the loss of compassion (assuming machines could not somehow replicate compassion). In any event, we already live in an age of aerial and drone warfare where the chance of any emotional connection between the attacker and ‘attackee’ is, arguably, precluded such that the removal of compassion from the battlefield would hardly be a novelty.⁹²

The third reason for arguing that machines have the potential to surpass humans in the execution stage of verification is that autonomous weapons would certainly have the capacity to deal with much more information much more quickly. Upon ratifying Additional Protocol I, the United Kingdom (UK) noted that ‘military commanders and others responsible for ... executing attacks necessarily have to reach decisions on the basis of their assessment of the information from all sources ... *reasonably available* to them at the relevant time [emphasis added]’.⁹³ Presumably this caveat was seen to be necessary due to imperfections of battlefield information, the potential for gaps in communications and, crucially, human limitations in the sense of our

85. Ekelhof (n 48) at 75.

86. Boothby (n 41) at 150.

87. Schmitt (n 36) at 13.

88. Ibid.

89. Ekelhof (n 48) at 79.

90. Thurnher (n 66) at 114.

91. Boothby (n 41) at 160.

92. Schmitt (n 36) at 12.

93. Letter from C Hulse (HM Ambassador) to the Swiss Government (28 January 1998) para (c).

ability to receive and digest large amounts of information against the backdrop of an armed conflict. After all, there is a limit to the volume of material a human commander can gather themselves and, equally, a limit to how much can feasibly be communicated to them by others without the subject becoming overwhelmed. On the contrary, the amount of information autonomous weapons could gather on an environment would be vast with machines such as Exyn Technologies drones already able to scan entire mines and chart them with effectively unlimited memory reserves.⁹⁴ Furthermore, the time they take to achieve all of this is vanishingly brief.⁹⁵ The fourth and final reason that autonomous weapons might handle verification in the execution stage well is that they would be more able to abort and retreat than humans. The UK, upon ratifying Additional Protocol I, noted that the obligation to cancel or suspend an attack ‘only extends to those who have the ... practical possibility to cancel or suspend the attack’.⁹⁶ In terms of ‘practical possibility’, an autonomous weapon could retreat or perhaps even self-destruct rather than continue with a unwarranted attack. For example, a bipedal autonomous weapon might storm a building believed to contain enemy combatants who turn out to be merely untargetable armed criminals (who are not ‘direct participants in hostilities’ on account of the lack of any ‘belligerent nexus’⁹⁷). Assuming it was armour plated, the machine might simply be able to leave even if the individuals were shooting at it as it departed. If leaving was not viable, perhaps because the criminals have barricaded the machine inside, it could simply destroy itself. In such scenarios, in contrast, human soldiers might have to complete the attack to secure their own safe exit. In short, there is ample reason to believe that, in time, autonomous weapons would be able to handle verification in the execution stage of an attack competently.

Finally, it must be borne in mind that the verification limb of precaution – at both the planning *and* execution stages – requires that ‘everything feasible’ is done to sort legitimate targets from protected persons and objects. Doing ‘everything feasible’ has proved to be a controversial test. Although ‘feasible’ is not as high a standard as ‘possible’ would have been, it remains quite a high threshold. It is more demanding than, for example, doing everything merely ‘convenient’ or ‘expedient’. Consequently, certain states have moved to clarify the meaning of feasibility in this context. The UK states in its military manual that “‘feasible’ means that which is *practicable or practically possible*, taking into account all circumstances ruling at the time, including humanitarian and military considerations [emphasis added]’.⁹⁸ A number of other states have taken the opportunity to define ‘feasible’ in a similarly narrow manner: Argentina; Australia; Canada; Germany; India; Ireland; Italy; the Netherlands; New Zealand; Turkey and the US.⁹⁹ Consequently, in many cases, this would be the standard that autonomous weapons would need to satisfy. This is all very

94. S Crowe, ‘Exyn Drone Maps Inactive Mine on the Fly’ (*The Robot Report*, 19 November 2019) para 3, available at <<https://bit.ly/35R3YRo>> (accessed 3 October 2020).

95. Winter (n 50).

96. Letter from C Hulse (n 93) at para (o).

97. N Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (ICRC 2009) 58–64.

98. UK Ministry of Defence, *The Manual of the Law of Armed Conflict* (Oxford University Press, Oxford 2004) 5.32 (fn 191).

99. International Committee of the Red Cross, ‘Practice Relating to Rule 15: The Principle of Precautions in Attack’ (*International Committee of the Red Cross*), available at <<https://bit.ly/3c8XCQR>> (accessed 3 October 2020).

well, but it leaves us with the issue of determining what is 'feasible', or 'practicable', when it comes to target verification in the context of autonomous weapons.

So, what can we 'feasibly' expect machines to do in the course of verifying targets as legitimate military objectives? According to Schmitt, the obligation would 'require full use of onboard ... sensors that could boost the reliability of target identification'.¹⁰⁰ However, he goes further and adds that the sensors of an autonomous weapon could not be used in isolation if *external means* of target verification would 'measurably improve identification and their use was militarily feasible in the circumstances'.¹⁰¹ Thurnher agrees but caveats this by adding that the need to use external systems is only required 'if the degree of operational risk to the external systems is reasonable' and that relevant considerations here include 'the survivability of military personnel involved and the [other putative] weapon system'.¹⁰² This seems to be a sensible compromise. Many of the benefits of autonomous weapons stem from the fact that they are designed to operate entirely independently so blanket requirements on external input may be unrealistic. It might, in some cases, be possible to allow them to interact with external actors, perhaps human scouts, to corroborate their target assessments. However, if doing so would present a threat to those human units that a reasonable commander would not take, then such communication would probably not be deemed 'practicable or practically possible' by states such as the UK and others who view feasibility in these terms. For example, ground troops could scout buildings in a town from a hillside to determine which harbour an enemy presence before deploying an autonomous weapon to that site. However, if the degree of benefit from this close quarters scoping was low (perhaps at night it might not clarify much) and if the risk to the troops of being spotted was high, then the measure would probably not be feasible. Similarly, if the external actor happened to be another autonomous weapon that was urgently needed elsewhere then it would probably not be practicable to call it away from that matter to double-check the assessment of its colleague machine. Notwithstanding the above limitations, it is submitted that in most cases it will indeed be feasible for the assessments made by autonomous weapons to be corroborated by external actors where there is no (or minimal) risk to humans, there are no competing priorities for other machines or, indeed, because the external information comes from impervious sources such as satellites. Indeed, autonomous weapons could theoretically triangulate all of these sources at higher speed and accuracy than humans with the result that their assessments would incorporate all available sources of information thus setting a new, higher, standard for what is feasible.

The addition of autonomous weapons to the battlefield may augment what is considered feasible in verification in other ways too. As was alluded to above, and as Thurnher has noted, 'an attacker is not required to assume greater risk to protect the civilian populace if a reasonable attacker in a similar situation would not take such actions'.¹⁰³ This point is accepted because a *human* in such a scenario would be weighing the increased risk to one human life (his own) against the diminution in risk to another human life (a civilian). The attacker might, depending on the situation, reasonably conclude that it would not be feasible to put their own life at significant risk for, say, a small reduction in the risk to the life of a civilian. Importantly though,

100. Schmitt (n 36) at 23.

101. Ibid.

102. Thurnher (n 66) at 109.

103. Ibid.

for present purposes, if the attacker of the future turned out to be an *autonomous weapon*, the calculus in such situations would change significantly. These machines would have some inherent monetary and tactical value, but they would be expendable in the way that human lives are not. Thus, the range of what is ‘feasible’ expands when autonomous weapons are involved as machines could put themselves at increased danger by getting much closer to their targets, and scrutinizing them for far longer, than a human soldier could before exceeding what would be considered feasible. They could even be required to take this to the point of their own destruction in a manner that would, of course, be wholly wrong to expect of human soldiers.

In summary then, machines at least have the potential to allow their users to achieve greater compliance with LOAC when it comes to verifying targets. This is true at the planning stage where they could process more information to ensure quality decisions are taken; at the execution stage where they can react more rapidly than humans and without fear and, finally, in terms of doing everything ‘feasible’ to ensure their targets are legitimate – getting up close and personal to them to ensure their assessment was correct and even precipitating their own destruction if need be.

4 THE INTERSECTION OF PRECAUTION AND AUTONOMOUS WEAPONS – PROPORTIONALITY

In addition to supplying procedural verification-based reinforcements in support of the principle of distinction, the rules on precaution in attack shore up the principle of proportionality (the International Court of Justice considers proportionality to be a ‘cardinal principle’ of LOAC¹⁰⁴) by explaining when and how that concept operates in practice. Thus, proportionality is the second limb of precaution that we will consider. Proportionality has many permutations in LOAC but, in particular, it prohibits attacks which would cause collateral damage ‘excessive in relation to the concrete and direct military advantage anticipated’.¹⁰⁵ The ICRC has confirmed proportionality to be a customary rule of LOAC¹⁰⁶ and abiding by proportionality is important because breaches not only carry the potential for state and individual responsibility but, as NATO put it, they ‘could subject ... leadership to strategic consequences that may have an adverse impact on the military mission’.¹⁰⁷ In other words, acting disproportionately can undermine international support for a state and its strategic objectives in war. A full account of what proportionality means at a theoretical level, and how it might intersect with autonomous weapons, has been provided by the author elsewhere.¹⁰⁸ However, precaution raises additional, procedural, considerations in terms of how proportionality ought to be applied in practice. As with verification, there are two limbs: one to deal with the planning stage of an attack and one to deal with the execution stage. These will be considered in turn.

In terms of proportionality at the planning stage, as with verification, there are two phases of ‘planning’, and the one considered in this piece is the ‘machine-centric’ phase that might occur after an autonomous weapon has been released onto the

104. *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion) [1996] ICJ Rep 226 at para 78.

105. Additional Protocol I (n 5), Arts 51(5)(b) and 57(2)(a)(iii).

106. Henckaerts and Doswald-Beck (n 29) at 46–48.

107. NATO (n 59) at para 0205.

108. Winter (n 9).

battlefield and is planning what to target and how. Given that proportionality is concerned with ‘excessive’ harm, naturally a key requirement would be for an autonomous weapon to be able to grapple with the thorny issue of weighing military gain against collateral damage to make some sort of *quantitative* estimate around, for example, the number of enemies and civilians that may be killed or injured by any contemplated strike. As Thurnher put it, proportionality is ‘essentially a quantitative determination’.¹⁰⁹ Solving this issue would, according to the HPCR Manual, necessitate having a grasp of a range of factors such as ‘the known effects of the weapons that are to be used; the vulnerability of any civilian buildings in the target area; the number of civilians that are likely to be present in the target area at the particular time; and whether they have any possibility to take cover before the attack takes place’.¹¹⁰ Numerous difficulties arise when trying to discern how this might apply in the context of autonomous weapons and the author has previously argued that a utilitarian approach would need to provide part of the solution here.¹¹¹

However, in addition to the quantitative aspect of proportionality, there is also a *qualitative* angle that would need to be overcome if there was to be robust application at the machine-centric planning stage of precaution. For example, account must be taken of factors such as the physical pain that forms part of an injury (and potentially even the mental harm bound up therewith¹¹²) or the relative difference in destroying a dilapidated outbuilding compared to destroying a vacant, but otherwise habitable, family home. As Asaro notes, many aspects of proportionality are ‘abstract, not easily quantified, and [are] highly relative to specific contexts and subjective estimates of value’.¹¹³ HRW made this point forcefully when it stated that proportionality demands more than a mere balancing of quantitative data and that ‘a robot could not be programmed to duplicate the psychological processes in human judgment that are necessary to assess proportionality’.¹¹⁴ Consequently, the organization has taken the position that judgements on proportionality should be left solely within the purview of human actors.¹¹⁵ While it is acknowledged that enabling machines to make sound proportionality assessments while planning attacks is a particularly tricky problem, it is submitted that the issue is not necessarily insurmountable and that a number of avenues with potential solutions are available for exploration.

The first and most straightforward option for proportionality-compatible machine planning is simply to limit the operational ambit of autonomous weapons. Thurnher suggests, for example, that ‘in circumstances where civilians are not expected to be harmed, this principle [of proportionality] would not be in effect’.¹¹⁶ The corollary of this is that, provided autonomous weapons are kept to areas where only military objectives are present, they would be able to operate freely and without the need for detailed proportionality-based capabilities. This is not really a solution per se but rather an effort to circumvent the problem. It is certainly possible to imagine scenarios where autonomous weapons might be deployed with no expectation of civilian casualties. To take an extreme example, an autonomous submarine tasked with a

109. Thurnher (n 66) at 109–10.

110. HPCR (n 34) at 146.

111. Winter (n 9).

112. Ibid at 197–98.

113. PM Asaro, ‘Modeling the Moral User’ (2009) 28(1) IEEE Technology and Society Magazine 20, 21.

114. Human Rights Watch (n 37) at 33.

115. Ibid at 34.

116. Thurnher (n 66) at 110.

hunter-killer role in respect of enemy submarines in Arctic waters could perhaps be used without the need for any awareness of the notions of ‘military gain’ or ‘collateral damage’ as there are unlikely to be any civilians submarines sailing around in that area. However, the range of scenarios in which the presence of civilians can be ruled out entirely is, realistically, extremely limited. Even in the context of a remote mountain-top military base for example, there is the risk that civilians may frequent the area to deliver food, medical supplies, ammunition and so on (though some activities have the potential to amount to ‘direct participation in hostilities’ which results in the loss of protected status¹¹⁷). Thus, if machines were programmed on the assumption that no civilians will ever be in their area of operation – and without the necessary awareness of proportionality – this would be a recipe for disaster. Further, from a military point of view, such machines would have such a limited range of operation, and would lack adaptability to such an extent, that their cost would likely be prohibitive in comparison to their utility. For these reasons, it is submitted that a ‘blinkered’ approach could never be honed into a model that would be fit for purpose in terms of satisfying this limb of precaution and thus it should be entirely discounted.

In fairness, while Thurnher probably put too much stock into the above solution, he goes on to offer an alternative, more nuanced, solution for situations in which there is an expectation of collateral damage and it does have more likelihood of discharging the proportionality component of precaution when it comes to taking account of quantitative and qualitative issues. In particular, he argues that ‘nations might be able to comply by ... preprogramming the acceptable values for the various situations that an autonomous weapon might encounter’ such that ‘the operator has pre-determined for any specific target what the commander considers as an excessive amount of collateral damage’.¹¹⁸ While this proposal is an improvement on simply blinkering autonomous weapons to potential civilian harm, it comes with problems too. Most obviously, it requires commanders to foresee any situation that an autonomous weapon might encounter. Of course, the world is a dynamic place – perhaps at its most dynamic in wartime – so to expect commanders to be able to forecast an almost unlimited range of potential permutations would be untenable. Following on from this is the issue of whether or not it would indeed still be military commanders who would set the level of what is proportionate, with McFarland noting that certain decisions traditionally made by such individuals ‘would be transferred away from traditional decision-makers and would instead be made, in effect, by the people who define the behaviour of the autonomous system’.¹¹⁹ Similarly, Ekelhof notes that the adoption of autonomous weapons ‘might generate an increase in responsibilities for certain superiors or the developers of systems’.¹²⁰ In addition to the problem of commanders’ (or software developers’) foresight, any systems reliant solely on human pre-programming would be unable to engage ‘targets of opportunity’ which had not been foreseen. This would undermine the military benefits of autonomous weaponry. Finally, Thurnher suggests that this pre-programmed awareness of collateral damage would only be needed where there was an ‘expectation’ of collateral damage. Again, for the reasons outlined above, this is not appropriate as civilian presence can seldom be ruled out entirely and, in any event, enemy combatants must not be subjected to excessive harm. Thus, the

117. Melzer (n 97) at 47–64.

118. Thurnher (n 66) at 110.

119. McFarland (n 42) at 1333.

120. Ekelhof (n 48) at 83.

pre-programming route does not offer a viable path to discharging the proportionality limb of precaution any more than the blinkered approach does.

With two solutions discounted, the conundrum of implementing precautionary proportionality in the machine-centric phase of planning remains unresolved. To answer it, we need a much more ambitious solution – one that counts on technological innovations that have yet to be achieved but that seem to grow closer every day with leaps in artificial intelligence and machine learning technologies. Over a decade ago, in the earlier days of the autonomous weapons debate, Arkin argued that the implementation of proportionality by machines would require the development of an ‘ethical governor’.¹²¹ The ethical governor would evaluate both anticipated military gain and collateral harm based on technical calculations – with the results being fed into an algorithm that makes a proportionality determination assessment.¹²² At the time, Arkin’s approach seemed far-fetched. However, in the intervening years, the approaches taken by states to matters that have a direct bearing on proportionality assessments have become increasingly systematic and detailed to the point where it is now possible to see how machine-made assessments of what is proportionate might eventually work. For example, nowadays, collateral damage estimation methodologies (CDEMs) are employed by states such as the US¹²³ and supranational bodies such as the EU¹²⁴ to facilitate the assessment of collateral damage by placing reliance on scientific data, objective standards and structured workflow models that stand in stark contrast to the traditional approach of leaving proportionality to be decided *ad hoc* based on the subjective and sometimes inconsistent judgements of military commanders.

More detail on CDEMs will be given shortly in the context of precautionary proportionality in the execution stage of an attack. However, as briefly summarized by Schmitt, for planning purposes they require consideration of such matters as ‘the precision of a weapon, its blast effect, attack tactics, the probability of civilian presence in structures near the target, and the composition of structures’ in an effort to assess likely collateral damage.¹²⁵ It should be noted that, at present, proportionality assessments are not intended to be left to CDEMs alone. Rather, they are ‘policy-related instruments used to determine the level of command at which an attack causing collateral damage must be authorized’.¹²⁶ In other words, they are currently conceived as tools designed to structure and streamline human decision-making rather than to replace it with machine-made decisions. However, that could change in the future as machine capabilities evolve and, indeed, machines would seem to lend themselves to the implementation of this sort of detailed framework. Schmitt is certainly confident on this point, arguing that ‘there is no question that autonomous weapon systems could be programmed to perform CDEM-like analyses’.¹²⁷ This may

121. RC Arkin, *Governing Lethal Behavior in Autonomous Robots* (CRC Press, Boca Raton, FL 2009).

122. *Ibid* at 187.

123. US Department of Defense, ‘No-Strike and the Collateral Damage Estimation Methodology’ (2012) CJCSI 3160.01A, available at <<https://bit.ly/2Zzzggj>> (accessed 3 October 2020).

124. European Union Military Committee, ‘Avoiding and Minimizing Collateral Damage in EU-led Military Operations Concept’ (2015) EEAS 772 REV 8, available at <<https://bit.ly/2LWq0L7>> (accessed 3 October 2020).

125. Schmitt (n 36) at 19.

126. *Ibid*.

127. *Ibid* at 20.

be correct, but there is a question as to when autonomous weapons with this sort of capability will become technologically feasible.

At present, the big stumbling block to fully machine-implemented CDEMs is context. Context remains key to proportionality assessments. To make an assessment on proportionality, it is necessary first to categorize the various persons and objects on the battlefield as civilian or military while recognizing that nothing has a wholly fixed status. For example, a (targetable) soldier may be actively scouting an enemy area one minute but then become wounded and thus *hors de combat* (i.e. out of action and protected) the next. A building may go from housing refugees one day (thus bearing protected status) to quartering troops the next (thus becoming targetable). As a result of these contextual problems, Schmitt concludes that ‘it is unlikely in the near future that, despite impressive advances in artificial intelligence, “machines” will be programmable to perform robust assessments of [proportionality]’.¹²⁸ Still, as Ekelhof notes, the complexities of modern battle and the high volumes of data are moving beyond *human* capacity for processing: ‘States cannot ... solve the multiple challenges ... simply by throwing more people at the problem [but] autonomous technology ... could speed up the process, processing large amounts of data so as to discern interrelationships and dependencies that human beings would fail to recognize’.¹²⁹ Moreover, while human capabilities diminish in real terms, machines are getting better at making context-sensitive judgement calls. For example, UK-based developer DeepMind Technologies has developed an artificial intelligence system called ‘AlphaGo’ which has been able to beat human champions at the ancient Chinese strategy game ‘Go’.¹³⁰ Go is vastly complex, with more potential moves than there are atoms in the universe, yet AlphaGo was able to win by applying deep learning in neural networks to make sound decisions even against a dynamic backdrop of endless possible permutations and contextual shifts caused by the opponent’s actions. According to Gibney, this success was possible by developing ‘brain-inspired programs in which connections between layers of simulated neurons are strengthened through examples and experience’.¹³¹ An important caveat here is that AlphaGo is still limited in the sense of having a sole objective: to surround a larger total area of the board with its stones than the opponent.¹³² The system was not required to consider nuanced competing objectives in the manner that is essential in armed conflict where, for example, humanitarian considerations need to be balanced against military necessity.¹³³

These concepts are difficult to balance and, efforts to square this circle remain in an embryonic stage. The author has previously argued¹³⁴ that the key to unlocking this problem is first to measure various harms and gains using a utilitarian model drawing on the work of Bentham and, in particular, his ‘dimensions of value’ – intensity; duration; certainty or uncertainty; propinquity or remoteness; fecundity (the chance of a sensation generating a later sensation of the same kind); purity (the chance of a sensation not generating a later sensation of the opposite kind); and extent (the

128. Ibid.

129. Ekelhof (n 48) at 79 and 83.

130. E Gibney, ‘Google AI Algorithm Masters Ancient Game of Go’ (2016) 529 Nature 445.

131. Ibid at 445.

132. Ibid at 446.

133. Winter (n 50).

134. Winter (n 9).

number of people affected).¹³⁵ Second, the results must then be weighed against each other by making seemingly incommensurable considerations measurable under the same index. This can be achieved if, borrowing from Newton and May, we recognize that ‘the value of the military objective [must] be couched in terms of lives to be protected or saved so that the costs of such an operation, also often drawn in stark terms of the risk of loss of non-combatant lives, can be assessed more straightforwardly’.¹³⁶ Notwithstanding the complications, systems such as AlphaGo which exhibit high-level artificial intelligence capability are moving from the realms of science fiction to science fact and bring with them the potential for much more sophisticated machine thinking. They might just be able to resolve the difficulties mentioned above – especially in the context of military technology where funding and expertise abound. Of course, many argue that this sort of advancement would not involve the emergence of intelligence per se. McFarland emphasizes that, ‘while such systems may not be programmed with precise predetermined responses to every situation they encounter, they are programmed with some system for developing a response and are thereby operating in accordance with their programming’.¹³⁷ Even if that is the case, ultimately, it is the quality of the result arrived at – rather than how it is achieved – that matters in this context. Moreover, while McFarland worries that the deployment of autonomous weapons would result in the replacement of ‘on the spot’ human-made decisions with ‘broader policy-like choices’ which might be less well tailored to specific scenarios,¹³⁸ one could equally argue that this would be a positive development as it would obviate reliance on subjective and impressionistic human efforts at balancing collateral harm and military advantage – efforts often clouded by the fog of war.

In short then, any machines deployed to act independently in combat must be able to make proportionality assessments in the planning stage of an attack if they are to properly discharge precaution. Their assessments should be based on rigorous and capable onboard systems and not on blinkering machines to the risk of collateral harm or making them dependent on unrealistic levels of human ‘pre-programming’. Admittedly, machines have not been demonstrated to achieve this level of contextual decision making beyond the confines of a board game, but technological innovations might mean that this soon changes.

Of course, precaution requires not only that proportionality is considered at the *planning* stage of an attack but also throughout the *execution* stage of a strike as it unfolds. This is essential given that, as the HPCR Manual notes, ‘the reality on the ground may appear ... different from the one [imagined] during the pre-mission [stage]’.¹³⁹ Consequently, there is a second limb to precautionary proportionality to the effect that an attack ‘shall be cancelled or suspended if it becomes apparent that [it] may be expected to cause ... excessive [collateral damage]’.¹⁴⁰ This mirrors the provision made in the context of verification, discussed above, requiring ongoing review of any situation. A question therefore arises as to how well autonomous

135. J Bentham, *The Principles of Morals and Legislation* (Prometheus Books, New York 1988) 30.

136. M Newton and L May, *Proportionality in International Law* (Oxford University Press, Oxford 2014) 285.

137. McFarland (n 40) at 1328.

138. Ibid at 1335.

139. HPCR (n 34) at 146.

140. Additional Protocol I (n 5) at Art 57(2)(b).

weapons might be able to satisfy this requirement and, again, there is much scepticism surrounding the ability of machines to make robust assessments on such matters – especially in the midst of an engagement. For example, according to HRW, ‘it is highly unlikely that a robot could ... handle the infinite number of scenarios it might face’.¹⁴¹ Similarly, for Sharkey, the number of permutations that can arise after an attack has begun ‘is vast and could cause chaotic robot behavior with deadly consequences’.¹⁴² Krishnan also speculates that the inability of an autonomous weapon to keep pace with evolving situations might lead to ‘the autonomous robot’s incomplete understanding of its external environment resulting from software limitations’ and that this would inevitably lead to ‘faulty behavior’.¹⁴³

The reticence about the ability of autonomous weapons to act proportionately in the course of executing an attack is understandable. However, whether these fears are borne out in practice depends on how clearly states define their parameters for assessing matters such as collateral damage. This is where it becomes instructive to consider in more detail systems such as the CDEM used by the US which was alluded to above.¹⁴⁴ The US CDEM is a complex process that employs a five stage analytical framework to assess collateral damage using ‘a mix of empirical data, probability, historical observations, and complex modeling for [collateral damage estimation] analysis’.¹⁴⁵ Each stage has its own considerations and results in an output of ‘low’ or ‘high’. If the risk of collateral damage in any given stage is low the attack is permitted under the CDEM whereas, if it is high, one must move on to the next stage and take account of its considerations. These considerations become progressively more robust as one moves up the hierarchy.¹⁴⁶

The detail in the various stages of the US CDEM is useful as it gives a sense of how a broadly framed concept such as proportionality can, in fact, be refined into discrete steps that lend themselves more readily to machine application. Stage 1 is ‘target validation and initial [CDEM] analysis’.¹⁴⁷ This stage involves drawing a radius around the target within which persons or objects may be affected by most of the conventional weapons used by the US military. It ‘identifies all collateral concerns’ and extends to matters such as ‘potential [chemical, biological and radiological] plume and environmental hazards’ which may magnify the effects of an attack.¹⁴⁸ It attempts to capture the difficulties posed by, for example, dual-use sites and sets the groundwork for an ‘intelligence analysis to estimate the population density’ in such cases.¹⁴⁹ Stage 2 is ‘general and target size analysis’.¹⁵⁰ This stage involves looking at the size of the target and comparing it to the area of effect for various weapon options with a view to finding the least destructive means of warfare that will still allow that target to be engaged effectively.¹⁵¹ Two analyses are employed in Stage 2. First, there

141. Human Rights Watch (n 37) at 32.

142. N Sharkey, ‘Automated Killers and the Computing Profession’ (2007) 40(11) Computer 122, 123.

143. A Krishnan, *Killer Robots: Legality and Ethicality of Autonomous Weapons* (Ashgate, Farnham 2009) 98–99.

144. US Department of Defense (n 123).

145. Ibid at D-1.

146. Ibid at D-A-7 to D-A-30.

147. Ibid at D-A-7.

148. Ibid.

149. Ibid at D-A-8.

150. Ibid at D-A-12.

151. Ibid at D-A-13.

is a 'minimum target size analysis' based on the principle that 'weapon systems/delivery platforms with [an area of effect] less than or equal to the target area have a higher probability of [being proportionate]'.¹⁵² This is fleshed out in considerable detail with a number of reference tables and diagrams giving more information on how these areas are to be calculated using 'rectangular', 'circle' or 'segment' methods.¹⁵³ Second, there is a 'precision-guided munition analysis' which estimates the risk of collateral damage in cases where this sort of weaponry is used based on either the 'aimpoint method' (which estimates collateral damage if targeting a particular enemy combatant or object) or the 'facility method' (which estimates collateral damage if targeting a broader structure). Again, further reference tables are given in support of these to capture the minutia.

Stages 3 and 4 of the US CDEM relate to mitigation in the sense of weapons selection and deployment and so these will be parked for the moment and returned to below. Stage 5 is 'casualty analysis' which is effectively the last resort and is used when all other techniques have been exhausted and 'some level of collateral damage appears unavoidable'.¹⁵⁴ Here, focus is on estimating 'noncombatant demographics' – i.e. how many civilians are likely to be in a given area (usually a building or complex of some sort) – based on a range of factors such as 'unique characteristics and cultural behaviors ... as well as the population distributions, customs, and cultural practices, as well as particular habits unique to a region'.¹⁵⁵ There are various data, tables and procedures in place to assist with computing casualty estimates.¹⁵⁶ This includes an appendix setting out further details on population density calculations.¹⁵⁷ Factors such as time are accounted for here with day and night time acknowledged to have an impact on the degree of collateral damage and with 'twilight' hours being set as the default time for strikes.¹⁵⁸

The US CDEM may sound rather dry, technical and detailed. However, that is the point. It is often said that the requirements of LOAC in general, and proportionality in particular, are too vague.¹⁵⁹ CDEMs give a sense of how a level of specificity can be brought to bear in a way that might one day enable machine implementation. However, it is important to point out that the existence of CDEMs is not a silver bullet when it comes to having autonomous weapons apply precautionary proportionality in the execution of an attack. As currently constituted, they come with many – often significant – limitations. To begin with, the US itself acknowledges that the application of its CDEM 'is not an exact science'.¹⁶⁰ It is accepted that the 'operational environment, weapon reliability, and fidelity of intelligence data' can impact on reliability of the results.¹⁶¹ It is also conceded that 'the science is inherently limited by the quantity and reliability of collected and analysed weapons effects data, weapon delivery uncertainties, and target information'.¹⁶² Furthermore, certain matters are excluded altogether from the ambit of the CDEM. For example, a range of means of

152. *Ibid.*

153. *Ibid* at D-A-15.

154. *Ibid* at D-A-30.

155. *Ibid* at D-A-31.

156. *Ibid.*

157. *Ibid* at D-D-1.

158. *Ibid* at D-A-32.

159. Newton and May (n 136) at 285.

160. US Department of Defense (n 123) at D-2.

161. *Ibid.*

162. *Ibid* at D-1.

warfare are expressly omitted including nuclear, non-kinetic and surface-to-surface direct fire weapons.¹⁶³ Similarly, the CDEM does not account for weapon malfunctions.¹⁶⁴ Nor does it account for ‘unknown transient civilian or noncombatant personnel ... in the vicinity of a target’.¹⁶⁵ The list goes on. The consequence of all of the above is that, at present, the CDEM ‘is an estimative process to *help* inform a commander’s decision making [emphasis added]’ rather than to replace it.¹⁶⁶ Another limitation comes not from the CDEM itself, but from the absence of its hypothetical counterpart. In other words, there is no corresponding model for military advantage, no ‘Military Advantage Estimation Methodology’ or ‘MAEM’ (an admittedly unfortunate acronym in this context). This is something that could, and would need to, be developed by states serious about devising strong proportionality modelling for use by autonomous systems and they ought actively to be considering such matters now. Again, as noted above, the author is of the view that utilitarianism and the work of Bentham offers a solid theoretical starting point for the work on missing pieces of the jigsaw such as the putative MAEM.

Notwithstanding the limitations and gaps, it is submitted that CDEM-style modelling holds the key to ensuring that the use by militaries of autonomous weapons on the battlefield would be compatible with precautionary proportionality in the execution stage of attacks. Of course, for this to be realized, further efforts are required in terms of articulating the rules as concrete processes and also in expanding the rules to cover as-yet unrefined considerations (such as military advantage). There is also the small matter of fielding autonomous weapons with sufficient independent judgement-making capability to implement this new beefed-up regime. On that point though, as we saw above, artificial intelligence is growing inexorably better at handling complex matters such as contextual decision making. In the near term, it is likely that such capabilities, through the magic of machine learning, could be carried from the board game to the battlefield. In the longer term, the majority of scientific opinion estimates that 2040 will be the year in which ‘artificial general intelligence’ arrives enabling machines to function to the same intellectual standard as humans and thus deal with any remaining ambiguities or gaps in LOAC.¹⁶⁷ If (or when) that day arrives, machines might in fact be better positioned than humans to cope with the shifting proportionality calculus as they could be fitted with more sensors to capture the full picture of the battlefield and their higher processing speeds would render them able to deal with the high volumes of data that attend complex battlespaces – all, as Arkin notes, without the normal human burdens of fear or panic that often blight the execution stage of an attack.¹⁶⁸ Ultimately, as Thurnher put it, ‘if States can find a reasonable solution to these proportionality and precautions requirements, then their use of autonomous weapons should not be deemed unlawful’.¹⁶⁹

163. Ibid at D-5.

164. Ibid.

165. Ibid.

166. Ibid at D-2.

167. V Müller and N Bostrom, ‘Future Progress in Artificial Intelligence: A Survey of Expert Opinion’ in V Müller (ed), *Fundamental Issues of Artificial Intelligence* (Springer, Cham 2016).

168. Arkin (n 121) at 29–30.

169. Thurnher (n 66) at 110.

5 THE INTERSECTION OF PRECAUTION AND AUTONOMOUS WEAPONS – MITIGATION

Thus far, we have considered the extent to which autonomous weapons might one day be able to implement the verification and proportionality related components of precaution in attack. The next component is ‘mitigation’ which is an umbrella term adopted in this article to cover two related sub-rules of precaution: mitigation of the effects of means (i.e. weapons of war) on one hand and mitigation of the effects of methods (i.e. tactics of war) on the other. The term ‘mitigation’ itself is not found in the primary legal sources, but it captures the essence of the provisions. According to Additional Protocol I, those deciding on attacks must ‘take all feasible precautions in the choice of means and methods of attack with a view to avoiding, and in any event to minimizing [collateral damage]’.¹⁷⁰ In the context of an autonomous weapon, without a human ‘in the loop’, these obligations would need to be satisfied independently by artificial intelligence based systems. The question that is to be addressed in this part is whether this is viable.

Boothby does not believe that autonomous weapons will be able to perform the mitigation in precaution task adequately, stating that: ‘currently available autonomous and highly automated technology does not enable the evaluative precautionary decisions’ required to determine appropriate means and methods of warfare.¹⁷¹ However, he caveats this by saying that ‘it is likely that there will with the passage of time be some improvement in sophistication of these technologies’ and thus leaves open the possibility of their compatibility with this task in the future.¹⁷² Thurnher is bolder and argues that the requisite level of artificial intelligence ‘seems conceivable in the future given expected technological advances [and] will need to be embedded into any future autonomous weapon, which might encounter civilians on the battlefield’.¹⁷³ The author agrees with Thurnher here, with the exception of his last clause which should be excised on the basis that the absence of civilians can never be assumed. Furthermore, following from what was said above about the US CDEM, we can see that states have already begun to put such LOAC considerations into terms that machines might soon be able to operationalize. For example, according to the UK:

In considering the means or methods of attack to be used, a commander should have regard to the following factors: (a) the importance of the target and the urgency of the situation; (b) intelligence about the proposed target – what it is being, or will be, used for and when; (c) the characteristics of the target itself, for example, whether it houses dangerous forces; (d) what weapons are available, their range, accuracy and radius of effect; (e) conditions affecting the accuracy of targeting, such as terrain, weather, and time of day; (f) factors affecting incidental loss or damage, such as the proximity of civilians or civilian objects in the vicinity of the target or other protected objects or zones and whether they are inhabited, or the possible release of hazardous substances as a result of the attack [and] (g) the risks to his own troops of the various options open to him.¹⁷⁴

In other words, again, we have seen a trend towards taking nebulous concepts that would be difficult for machines to apply and honing them into a series of refined,

170. Additional Protocol I (n 5) at Art 57(2)(a)(ii).

171. Boothby (n 41) at 149.

172. Ibid.

173. Thurnher (n 66) at 114.

174. UK Ministry of Defence (n 98) at para 5.32.5.

discrete steps that are more amenable to application by artificial intelligence. To assess whether this general trend will be sufficient to allow autonomous weapons to apply mitigatory precaution, both limbs – means and methods – will now be considered in turn.

Means are ‘weapons, weapons systems and munitions’.¹⁷⁵ They range from non-lethal devices such as canisters of incapacitating gases and guns armed with rubber bullets to lethal devices such as rifles, machine guns, grenades, missile launchers and nuclear bombs. Autonomous weapons could theoretically be outfitted with any weapons that might be desirable on the battlefield. Consequently, they will need to be programmed in such a way that they are able to determine which particular weapon to select in any given situation in order to maximize the chance of achieving their objectives while minimizing the chance, and extent, of collateral damage. In order to do this, as Thurnher observed, machines would need to be able ‘to calculate which warheads or weapons can accomplish the mission ... to determine the amount of harm to the civilian populace each of those weapons would be expected to produce [and] select the one that is expected to cause the least harm’.¹⁷⁶

There are, of course, several concerns surrounding how realistic it is for machines to undertake complex tasks such as determining the most appropriate weaponry. As noted by Ulgen, groups opposed to the deployment of autonomous weapons cite their inability to ‘think’ like people, the potential for a malfunction in the technology, the inherent inhumanity of delegating such decisions to machines and so on.¹⁷⁷ One concern that specifically relates to the present issue is what particular factors autonomous weapons might be programmed to consider when making weaponry selections. The UK expects commanders to consider, *inter alia*, the availability of weapons.¹⁷⁸ The US elaborates by noting that ‘the decision of which weapon to use will be subject to many practical considerations including effectiveness, cost and the need to preserve capabilities for other engagements’.¹⁷⁹ The HPCR Manual similarly explains that munitions capacity must be borne in mind and, for example, ‘when the number of precision-guided munitions is limited, it would be imprudent for the attacker to expend them early in the conflict without considering possible future needs and capabilities’.¹⁸⁰ It is perhaps unsurprising that states would seek to leave their commanders scope to consider matters such as availability and cost of weapons in this way. Nonetheless, these factors are controversial when taken from a humanitarian point of view as, by some, it is seen as the prioritization of firepower and money over human life. This controversy is likely to deepen if the call is being made by a machine. However, such concerns might be overstated as autonomous weapons have the scope to discharge the precautionary obligation here better than human combatants.

There are a couple of justifications to support this assertion. First, unlike human soldiers with encumbrance limitations, autonomous weapons could be outfitted with a vast array of different weapon types during their operations. This would offer them a wider palette of means of warfare to use and this is important because, as the HPCR Manual put it, ‘an attacker ought to choose a weapon with greater precision or lesser

175. HPCR (n 34) at 145.

176. Thurnher (n 66) at 113.

177. Ulgen (n 7).

178. UK Ministry of Defence (n 98) at para 5.32.5.

179. Office of General Counsel, *Department of Defense Law of War Manual* (US Department of Defense 2015) at para 5.11.3.

180. HPCR (n 34) at 148 (fn 269).

explosive force if doing so would minimize the likelihood of collateral damage, *assuming the selection is militarily feasible* [emphasis added].¹⁸¹ Naturally, the broader the range of options available, the more likely it will be that a lower impact weapon could be selected in any given scenario. In this way, the use of machines has the potential to allow militaries to better comply with precaution in attack when it comes to mitigation in the context of means. Second, the specifications for the various weapons one might expect on the modern battlefield are often highly complex and detailed. They cover matters such as the penetrative proficiency of their ammunition, the blast radius of their missiles, the effective duration of the crowd dispersal substances they carry and so on – as well as full modelling of how all of these behave when they are used in different environments and interact with different materials. It takes humans a lot of time and training to become familiar with these specifications and, even then, it is probably fair to say that not all of this detail is necessarily retained perfectly – especially in the fog of war. Machines, on the other hand, can be uploaded with meticulous details on each weapon system they carry and can retain them perfectly throughout even the longest and most demanding deployments. Therefore, in addition to having a wider arsenal to use, it might soon be the case that machines would be able to make more fully considered determinations – albeit based on prior human programming – about precisely which weapon to employ for a particular purpose than a human combatant could (even one who happened to have all the same weaponry at his or her disposal).

Finally, in terms of mitigation of the effects of means, it is important to reiterate that states already have in place detailed frameworks for protecting humanitarian considerations in the form of CDEMs and that these could be built upon and utilized by autonomous weapons performing precautionary evaluations. It was noted above that Stages 3 and 4 of the US CDEM relate to means of warfare – and discussion of those stages was omitted so that it could be included here instead. Stage 3 is the ‘weaponizing analysis’ which involves weapon selection.¹⁸² It is at this stage that ‘appropriate delivery systems, warhead, and fuze combinations’ are selected to achieve the mission objective while mitigating the risk of collateral damage.¹⁸³ In support of this stage, there are detailed tables on matters such as warheads which set out critical information on their respective ranges of effect and so on. There are also details on how these ranges change depending on the ‘delivery platforms’ used to launch them (for example F-16 or F-18 fighter jet or B-52 bomber).¹⁸⁴ Further specifics are given about ‘fuzing’ warheads (i.e. building in a delay between strike and detonation).¹⁸⁵ There are also step-by-step guides on how personnel are to implement the weaponizing analysis in the field.¹⁸⁶ Stage 4 of the US CDEM is ‘refined analysis’.¹⁸⁷ This is a continuation of the weaponizing process but moves beyond matters such as weapon selection and fusing and into how the weapons selected will impact with the relevant environment. It considers matters such as how the weapons will behave – and how individuals in the vicinity will be affected – when they strike targets such as buildings or vehicles. Analysts are required, for example, to consider

181. Ibid at 145.

182. US Department of Defense (n 123) at D-A-18.

183. Ibid.

184. Ibid at D-A-19.

185. Ibid at D-A-23.

186. Ibid at D-A-20 to D-A-22.

187. Ibid at D-A-23.

the construction types of surrounding buildings to assess what impact destruction of the target will have on them. The starting point is that buildings near to each other will have similar tolerances because ‘in most cases ... structures located in the vicinity of a target are similarly constructed’.¹⁸⁸ However, the analysis endeavours to be more nuanced than that and acknowledges that ‘an analyst may need to consider the weakest, not the nearest, collateral structure’ in cases where local architecture varies significantly.¹⁸⁹ The point here is that the complexities presented in the context of selecting means of warfare are not new or unexplored. States and militaries have been dealing with them for a long time and they have detailed processes in place for the purpose. These processes (admittedly, with a lot of work and in combination with more advanced artificial intelligence than we have now) could be adapted and built upon in the future to allow autonomous weapons to make robust means of warfare choices in a manner that would allow their use by militaries to be compatible with the mitigation of the effects of means aspect of precaution.

Selecting the appropriate means of warfare is only the first limb of the precautionary requirement of mitigation. There is a second limb – mitigation of the effects of *methods* of warfare employed. Methods are ‘tactics ... such as [in an aerial context] timing, angle and altitude of attack’.¹⁹⁰ In terms of timing, the US notes that ‘attacking a military objective when civilians are less likely to be present may be appropriate [and that] it may be appropriate to wait until enemy forces have departed from populated areas before attacking’.¹⁹¹ In terms of the angle of attack, the HPCR Manual notes that this is ‘one of the factors that determine where a bomb may land if it falls short of, or beyond, the target [and so] to spare a building located, e.g., to the west of a target, it may be advisable to attack from the north or the south’.¹⁹² Thus, the precautionary rule in the context of mitigation of the effects of methods is largely about timing and angles. At present, according to the US Handbook of Naval Operations (HNO), it is up to the (human) commander to ‘decide ... whether to adopt an alternative method of attack ... to reduce civilian casualties and damage’.¹⁹³

Could autonomous weapons make appropriate methods of warfare selections? The answer is a qualified ‘yes’. Calculations about matters such as timing and trigonometry lend themselves well to application by machines. Indeed, machines are likely to be better at this task than humans as these issues are less nuanced than some of the matters above such as how to balance qualitative considerations about military gain and civilian harm. The HNO, for example, notes that it is incumbent on the commander to make decisions about methods ‘in light of all the facts known or reasonably available to him’.¹⁹⁴ The amount of information available to an autonomous weapon is likely to be greater than the amount available to, or processable by, a human. This could be in the form of live data from their own sensors, live information shared by other military assets in the area or indeed archived material held in storage. All of this information could then be synthesized quickly to decide on attack timing

188. Ibid at D-A-24.

189. Ibid at D-A-25.

190. HPCR (n 34) at 145.

191. US Department of Defense (n 179) at para 5.11.2.

192. HPCR (n 34) at 145.

193. US Navy, US Marine Corps and US Coastguard, ‘The Commander’s Handbook on the Law of Naval Operations’ (2007) NWP 1–14M Para 8.3.1, available at <[https://www.marines.mil/Portals/1/Publications/MCTP%2011-10B%20\(%20Formerly%20MCWP%205-12.1\).pdf?ver=2017-07-11-151548-683](https://www.marines.mil/Portals/1/Publications/MCTP%2011-10B%20(%20Formerly%20MCWP%205-12.1).pdf?ver=2017-07-11-151548-683)> (accessed 3 October 2020).

194. Ibid.

and vectors that would suppress to the greatest possible extent the level of collateral damage from an attack. Indeed, the level of ability and rapidity required in this area has only increased over the last century given that, as Melzer notes, conflicts continue to spill from discrete battlefields into urban areas with the result that the likelihood of civilian harm increases while the margin for error vanishes.¹⁹⁵ Machines could leverage their unique talents and their immunity to the fog of war to achieve optimal outcomes in such environments.

A final requirement which can perhaps be subsumed into discussion of mitigation of the effects of methods is that, according to Additional Protocol I, when there is a choice between targets whose destruction would yield comparable military advantage, 'the objective to be selected shall be that ... which may be expected to cause the least danger to civilian lives and to civilian objects'.¹⁹⁶ According to the ICRC, this requirement (which might be thought of as 'target mitigation') is established as a rule of customary international law.¹⁹⁷ However it has proven controversial in the US which – though not directly challenging the customary status of the rule – favours a much more limited interpretation. It argues that 'the provision ... is not mandatory [and that] ... an attacker may comply with it if it is possible to do so, subject to mission accomplishment and allowable risk'.¹⁹⁸ In essence, the US position is based on the view that the broader interpretation of target mitigation is too focussed on collateral harm and that it does not afford adequate consideration to the potential for greater allied losses or the decreased likelihood of achieving a successful outcome. Thurnher supports this view, arguing that attackers should be free to 'consider military factors such as how significant the risk is to their own or friendly forces [such that] if one of the targets is more heavily defended, the attacker need not choose that target even if less collateral damage may result [from its selection]'.¹⁹⁹ The HPCR Manual shares this sentiment.²⁰⁰ It is submitted that US concerns here are overstated as target mitigation already factors in the balance of collateral harm with military advantage and thus allows for consideration of factors such as increased risk of mission failure. In particular, the obligation only kicks in when two or more targets offer 'similar military advantage'.²⁰¹ If attacking the alternate target came with a significantly greater risk of mission failure, then that attack would not carry similar military advantage and so there would be no obligation to launch it instead. In any event, the US is an outlier here. NATO more broadly recognizes that 'targets are prioritized based on ... intent to maximize effective use of joint force capabilities while minimizing the likelihood of unintended and potentially undesired consequences'.²⁰²

Proceeding, then, on the basis that target mitigation is indeed an obligation in all conflicts and incumbent on all combatants, what would it mean in practice for autonomous weapons? Schmitt gave an example, stating that 'an autonomous weapon system could not be employed to attack electrical substations if attacking transmission lines was militarily feasible, would achieve the same military objective ... and placed

195. N Melzer, 'The Principle of Distinction between Civilians and Combatants' in A Clapham and P Gaeta (eds), *The Oxford Handbook of International Law in Armed Conflict* (Oxford University Press, Oxford 2014) 298.

196. Additional Protocol I (n 5) at Art 57(3).

197. Henckaerts and Doswald-Beck (n 29) at 65.

198. US Department of Defense (n 179) at para 5.11.5 (fn 303).

199. Thurnher (n 66) at 112.

200. HPCR (n 34) at 148.

201. Additional Protocol I (n 5) at Art 57(3).

202. NATO (n 59) at para 0205(e).

civilians and civilian objects at less risk'.²⁰³ One can imagine similar examples where, for example, it might be possible to destroy a tunnel that would prevent an enemy's advance rather than destroying the enemy column itself. This sounds relatively straightforward, however Thurnher is correct to caution that 'selecting between different military objectives may present challenges for autonomous weapons'.²⁰⁴ The HPCR Manual cites the fact that destroying tunnels or bridges in lieu of a direct attack will not always be viable because it might be the case that 'several bridges [would] have to be destroyed in order to break an axis of communications [and that the] destruction of only one of them will serve little or no practical purpose'.²⁰⁵ Similarly, complexities are posed by the fact that, in LOAC, the term 'attack' is framed broadly and an attack should not be disaggregated into narrow components when assessing whether it may proceed but, rather, it should be taken 'as a whole'.²⁰⁶ The consequence is that, for example, if the three bridges referred to above needed to be destroyed to prevent an enemy from advancing across a river, this would amount to one broader attack against all three bridges as opposed three separate attacks. The mitigation assessment would therefore need to be made against this wider backdrop. Consequently, teaching machines to understand the proper ambit of an 'attack' would be critical if they were to be expected to compare two targets for the purposes of target mitigation.

To resolve this problem of target selection, again, autonomous weapons would need to be able to handle contextual considerations. This has the potential to be a technically difficult feat because, as Thurnher put it, 'these decisions involve complex, contextual judgments [and it] may be challenging to embed such qualitative decision making capabilities into an autonomous weapon system'.²⁰⁷ However, it is submitted that Thurnher may have overstated the problem because for him, as we saw previously, the solution revolves around the ability of human commanders to 'predict [the] different scenarios and situations the autonomous weapon may face and proactively provide input to the weapon system's direction control mechanisms'.²⁰⁸ Once more, it is the opinion of the author that this sort of 'pre-programming' approach would be unlikely. Rather, states will probably rely on machine learning techniques to 'train' their autonomous weapons systems on how to respond to issues such as these with the effect that the need for foresight on the part of commanders responsible for their deployment will be much reduced. As we saw above, artificial intelligence has been able to defeat humans in various games which require many levels of strategy in situations involving effectively infinite permutations. The systems learn not by human input but, as Gibney explains, by playing against themselves countless times to analyse how and why various approaches do (and do not) work – almost in the way that humans learn from play, but many orders of magnitude faster.²⁰⁹ If this could be translated across into strategic 'thinking' in the context of mitigation and target selection in warfare, then machines could prove to be most adept at the task.

203. Schmitt (n 36) at 31.

204. Thurnher (n 66) at 112.

205. HPCR (n 34) at 147.

206. L. Doswald-Beck (ed), *San Remo Manual on International Law Applicable to Armed Conflicts at Sea* (Cambridge University Press, Cambridge 1995) Art 46(d).

207. Thurnher (n 66) at 113.

208. Ibid.

209. Gibney (n 130) at 446.

In short then, if autonomous weapons could carry more weapon varieties than humans, could make all-encompassing calculations about the impact of different weapons and could rely on adapted CDEMs to accurately fuse warheads and so on, then they would seem to be well placed to make appropriate assessments for the purposes of precautionary decisions in the context of means of warfare selection. Similarly, as such machines already have the potential to make timing and vectoring decisions – and have the clear potential to make contextual determinations about the ambit of attacks taken ‘as a whole’ – they also seem to be well placed to make reliable decisions in terms of precautionary methods of warfare choice. Consequently, the mitigation component of precaution is unlikely to be a barrier to the deployment of autonomous weapons systems in the near future.

6 THE INTERSECTION OF PRECAUTION AND AUTONOMOUS WEAPONS – WARNING

The final requirement of precaution in attack – after verification, proportionality and mitigation – concerns warnings. According to Additional Protocol I, ‘effective advance warning shall be given of attacks which may affect the civilian population, unless circumstances do not permit’.²¹⁰ The emphasis here is on ‘civilian population’ and thus, as Greenspan notes, there is no obligation to give notice where no civilians remain, and only the military will come under fire’.²¹¹ Nonetheless, it is clear that ‘the warning obligation can be considered customary law’.²¹²

In terms of the form of warnings, by way of a historical example, one might consider the bombing runs conducted by the Allies against German targets in the French town of Annecy in 1944. For nearly half an hour prior to the attack, a plane outfitted with a siren was deployed to fly around the area and warn the inhabitants of the impending destruction and, consequently, most were able to flee and survive.²¹³ In terms of modern forms, the HPCR Manual notes that warnings may be given ‘for instance, through dropping leaflets or broadcasting’.²¹⁴ Of course, those are just indicative examples. As the US notes, ‘there is no set form for warnings [and they] may be general, communicated to the national leadership of the enemy state, or delivered to the civilian population through military information support operations ... advising the civilian population of risk of injury if they remain near military objectives’.²¹⁵ However, despite the flexibility here, it is clear that some forms of warning are not acceptable. For example, the ‘knock on the roof’ technique used by Israel in Palestinian territory – whereby a small yield explosive strikes the roof of the target a few minutes before the main detonation – has been rejected by the United Nations Human Rights Council. This was on the basis that the first explosions amount to attacks in themselves, that the delays are not sufficient to allow evacuation and that most civilians were not aware of their significance.²¹⁶ Clearly then, some minimum

210. Additional Protocol I (n 5) at Art 57(2)(c).

211. M Greenspan, *The Modern Law of Land Warfare* (University of California Press, Berkeley 1959) 338–39.

212. HPCR (n 34) at 153.

213. JM Spaight, *Air Power and War Rights* (3rd edn, Longmans, London 1947) 243.

214. HPCR (n 34) at 153.

215. US Department of Defense (n 179) at para 5.11.1.1.

216. United Nations, ‘Report of the Detailed Findings of the Independent Commission of

thresholds are in place when it comes to warnings. According to the HPCR Manual, ‘warnings ought to be as specific as circumstances permit’.²¹⁷ Further, the degree of specificity required will depend on factors such as ‘the length of time prior to the attack in relation to which the warning is issued’ and ‘the most effective mode in which the effective warning can be issued to civilians’.²¹⁸ However, those behind the HPCR Manual were not able to determine how widely, in terms of geography, warnings would need to be issued from the site of the attack.²¹⁹ For Thurnher, the warning must be sufficient to provide the civilian population ‘the opportunity to take measures to avoid the dangers’.²²⁰

The question is then whether the use of autonomous weapons by militaries would be compatible with the obligation to issue warnings before they launched an attack. According to Boothby, ‘where the rules ... require ... warnings to be given before any attack may take place ... difficult issues will obviously arise’.²²¹ However, Boothby does not articulate what these ‘difficult issues’ are and, in fact, it is submitted here that the matter is unlikely to be all that difficult for two main reasons. First, the issuance of warnings is a relatively simple affair. It concerns only the perfunctory ability to take action, such as sounding a siren to advise targets of an imminent attack (as occurred in Anncy). Indeed, Thurnher noted that ‘it seems reasonable to assume that nations could design autonomous systems to have broadcast or loudspeaker-type capabilities [or] other types of system’.²²² One could even go further and imagine autonomous weapons being outfitted with an array of different warning systems – potentially more than would be viable for human-operated systems to carry – to give them a wider palette to use in a similar way to that discussed above in the context of the panoply of weaponry they may carry. Second, the deployment of autonomous weapons could make the whole business of warnings less risky, and thus more viable, for attackers. For human combatants, just as increasing accuracy in terms of ‘verification’ tends to result in more risk to oneself; so too does the issuance of warnings. Both can involve greater exposure to the enemy (in terms of physical proximity or duration) and both increase the risk that enemies will survive given that the very notion of ‘warning’ allows people, both civilian and military, a chance to escape. However, when autonomous weapons are used, there is little or no additional attendant risk other than to the machine itself. Of course, by virtue of issuing a warning, a machine may give away its position and be destroyed and so there would be a potential financial and operational risk to the attacker of warnings. However, civilians will have been given their legally due warning and no friendly combatants will have been lost. This seems to be an improvement on the current position.

In short, there is little to suggest that there would be any dilution of warnings in the course of conflict in the event that autonomous weapons were to be deployed. On the contrary, they could be equipped with sirens, radio transmitters and leaflets in the same way that manned systems are often equipped. Their flexible hardware and the space freed up by the absence of needing to accommodate an on-board human

Inquiry Established Pursuant to Human Rights Council Resolution S-21/1’ A/HRC/29/CRP.4, 24 June 2015, paras 235–39.

217. HPCR (n 34) at 153.

218. *Ibid* at 154.

219. *Ibid*.

220. Thurnher (n 66) at 111.

221. Boothby (n 41) at 149.

222. Thurnher (n 66) at 112.

combatant may even allow them to carry more warning devices than a manned system. Furthermore, autonomous weapons could put themselves at more risk when delivering warnings in terms of proximity than would be reasonable to expect of a human in combat.

CONCLUSION

Precaution is a multi-faceted concept with limbs concerning verification, proportionality, mitigation and warnings. Whether the use of autonomous weapons would be compatible with precaution requires a consideration of these different facets. A systematic investigation has been conducted in this article to provide a detailed exposition of each of these components and to consider whether the use of ‘killer robots’ might be compatible with their respective requirements. The result of the analysis is that autonomous weapons will soon have the capacity to satisfy each component of precaution once the remaining technological gaps are bridged. While the closure of these gaps was once deeply speculative, we have seen technological advancement in this field progress at an exponential rate even in the last couple of decades with the result that it is now a question of if, not when, they will disappear. When they finally do, precaution will no longer act as a barrier to the deployment of autonomous weapons on the battlefield. Indeed, in some areas, machines are likely to be able to discharge the precautionary obligations better than humans with the effect that they ought to be used. It should be noted that this article is not intended to amount to an exercise in advocacy for the use of autonomous weapons in war. It is acknowledged that there are many other aspects of LOAC – and other regimes such as international human rights law – that have a bearing on the future use of these machines. Rules flowing from these sources might preclude the fielding of autonomous weapons in some or all cases. A ban or other regulatory instrument, should such an agreement ever materialize, would have the same effect. The argument here is merely that precaution, taken in isolation, is unlikely to inhibit the deployment of autonomous weapons in armed conflict for much longer.