

# AUTONOMOUS WEAPON SYSTEMS AND ACCOUNTABILITY: PUTTING THE CART BEFORE THE HORSE

CARRIE MCDUGALL\*

*Arguments that in many scenarios there will exist an ‘accountability gap’ where civilians are unlawfully killed through the use of an autonomous weapon system (‘AWS’) have been advanced to justify either the prohibition or restriction of AWS. This article examines the accountability problem through a critical review of the literature on accountability and AWS in order to identify why some experts say there will be no accountability gap while others argue there will, why some do not see this as a problem and others do, and why some consider this is a problem that has a solution while others see it as irresolvable. It is demonstrated that in large part these differing conclusions are the result of varying assumptions and preconditions. Without questioning the inherent value of accountability in the broad, it is argued that solutions to the debate over AWS will not be found in international criminal law, which should not be used as a backdoor to address perceived shortcomings in international humanitarian law. It is further argued that no analysis of the accountability problem will provide meaningful guidance as to whether the international community needs to prohibit or restrict AWS given that, one way or another, international criminal law can be amended to plug the accountability gap, if this is the desired policy outcome.*

## CONTENTS

I	Introduction.....	1
II	The Accountability Debate .....	4
III	AWS and Individual Criminal Responsibility .....	6
IV	Responses to the Accountability Question.....	15
V	Is There an Accountability Problem? .....	16
VI	The Problem with the Prominence of Accountability.....	25
VII	Conclusion .....	29

## I INTRODUCTION

As robotic and artificial intelligence (‘AI’) technologies continue to develop apace, discussions in relation to the practical, policy, legal and ethical implications of autonomous weapon systems (‘AWS’) are gaining in intensity. Significant concerns in relation to AWS have been raised by a range of actors, including the International Committee of the Red Cross<sup>1</sup> and the Campaign to Stop Killer Robots, a coalition of 106 non-governmental organisations in 54

---

\* Dr Carrie McDougall (@IntLawCarrie) is a Senior Lecturer at Melbourne Law School. She previously served as Legal Adviser at Australia’s Mission to the United Nations in New York and Assistant Director of the International Law Section of the Australian Department of Foreign Affairs and Trade. This article grew out of a presentation I delivered at the Asia Pacific Centre for Military Law’s (‘APCML’) Cyber Law and Emerging Technologies Workshop. I am grateful for APCML colleagues’ ongoing support, including feedback provided on an outline of this article by Kobi Leins and Bruce Oswald. I am also grateful to Fred Johnson for translating ‘IT-speak’ for me.

<sup>1</sup> See ‘New Technologies and IHL’, *International Committee of the Red Cross* (Web Page) <<https://www.icrc.org/en/war-and-law/weapons/ihl-and-new-technologies>>, archived at <<https://perma.cc/6ZU4-F5Y7>>.

countries.<sup>2</sup> Yet the 2018 Report of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems ('GGE')<sup>3</sup> records the fact that states are currently divided over the development and future use of AWS, with some arguing AWS should be prohibited outright, some calling for negotiations on a regulatory treaty, some arguing that a political declaration would suffice, and yet others opposed to any form of international regulation beyond existing rules of international law.<sup>4</sup> Scholars are equally divided, with views both for and against AWS firmly expressed.

As yet, there is no agreed definition of AWS. The most commonly cited definition is that employed in the United States Department of Defense Directive No 3000.09 on *Autonomy in Weapon Systems*. It defines an AWS as

[a] weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation.<sup>5</sup>

While this definition helpfully points to the key distinguishing feature of AWS — their ability to select and engage targets independently — the generality of the definition (and most others that have been suggested) captures a broad range of known and anticipated weapon systems. This definitional lack of clarity is exacerbated by the fact that while it is generally understood that there is a spectrum of autonomy on which any specific system, or its various components, may sit, at present there is no agreement among robotic and AI experts about the

<sup>2</sup> See also *Campaign to Stop Killer Robots* (Website) <<https://www.stopkillerrobots.org>>, archived at <<https://perma.cc/2ZPF-TWFQ>>.

<sup>3</sup> The Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems ('GGE') was established by the High Contracting Parties to the *Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects*, opened for signature 10 October 1980, 1342 UNTS 137 (entered into force 2 December 1983) ('CCW'), and is the primary international forum for discussions among states on the development and future use of autonomous weapon systems ('AWS').

<sup>4</sup> Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc CCW/GGE.1/2018/3 (23 October 2018) [28]–[29], annex III ('*Chair's Summary of the Discussion of the 2018 Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*') [46]–[48] ('*Chair's Summary of the 2018 GGE Discussion*').

<sup>5</sup> Department of Defense, *Autonomy in Weapon Systems* (Directive No 3000.09, 21 November 2012) 13–14 ('*Autonomy in Weapon Systems*') <<https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf>>, archived at <<https://perma.cc/2WK3-WF82>>. The Directive distinguishes 'semi-autonomous weapon systems', defined as '[a] weapon system that, once activated, is intended to only engage individual targets or specific target groups that have been selected by a human operator': at 14. The AWS definition has been criticised: see, eg, Rebecca Crootof, 'The Killer Robots Are Here: Legal and Policy Implications' (2015) 36(5) *Cardozo Law Review* 1837, 1847–9. Rebecca Crootof proposes an alternative definition: '[a]n "autonomous weapon system" is a weapon system that, based on conclusions derived from gathered information and pre-programmed constraints, is capable of independently selecting and engaging targets': at 1854.

degree of autonomy that AWS will ever possess (assuming no legal restraints are placed on technological developments).

As a result, one is struck by the fact that many participants in the AWS debate are talking past each other.<sup>6</sup> For example, in analysing the likelihood of the use of AWS being compliant with international humanitarian law ('IHL'), differing conclusions are at least in part attributable to authors' varying conceptions of AWS. At one end of the spectrum this encompasses several current weapon systems (such as the Harpy, the Aegis Combat System, the MK 15 Phalanx Close-In Weapon System and the SGR-A1 sentry gun) and those that are foreseen on the horizon, the use of which is governed by current policy requirements (such as the requirement under US Department of Defense Directive No 3000.09 that '[a]utonomous and semi-autonomous weapon systems shall be designed to allow commanders and operators to exercise appropriate levels of human judgment over the use of force'<sup>7</sup>). Authors at the other end of the spectrum are attempting to assess imagined futuristic AWS with advanced AI, including deep learning, that are deployed absent any policy constraints.

This highlights the fact that answers to many of the practical, legal and ethical questions that currently surround AWS hinge on the precise shape of future technologies — a variable that is presently unknowable to us.<sup>8</sup> This has resulted in attention being devoted to other questions, the answers to which one can more readily speculate about. Chief among these is the accountability question. Indeed, one of the main arguments of those pursuing a prohibition, moratorium or the regulation of AWS is that the use of such weapons will create an accountability gap — in other words, that no entity could be held to account for any violation of the law resulting from the use of AWS.

This article separates the accountability issue from other legal, technical and ethical arguments concerning the development and use of AWS, and examines its relevance to the debate over whether AWS should be prohibited or regulated. After outlining the accountability question and the prominent role it is playing in

---

<sup>6</sup> Chris Jenks argues that this is in large part a result of inaccurate definitions and concepts utilised by the Campaign to Stop Killer Robots and the reports of Human Rights Watch and the UN Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, on which the Campaign relies. He is critical of the fact that the vague definitions proffered by those calling for a ban on AWS, which are characterised by proponents as weapons of the future, in fact capture weapons that he says have been employed by over 30 countries for over 30 years: Chris Jenks, 'False Rubicons, Moral Panic & Conceptual Cul-De-Sacs: Critiquing & Reframing the Call to Ban Lethal Autonomous Weapons' (2016) 44(1) *Pepperdine Law Review* 1.

<sup>7</sup> Department of Defense, *Autonomy in Weapon Systems* (n 5) para 4(a). Of note, the Directive also requires, inter alia, that AWS be designed with appropriate '[h]uman-machine interfaces and controls' (to guard against unintended engagement or loss of control of the system to unauthorised parties): at para 4(a)(2)(b); and that AWS human-machine interfaces be 'readily understandable to trained operators' (to enable operators to make informed and appropriate decisions in engaging targets): at para 4(a)(3)(a). The current Directive only authorises human-supervised AWS to select and engage targets for local defence to intercept attempted time-critical or saturation attacks for static defence of manned installations, or onboard defence of manned platforms: at para 4(c)(2). It prohibits the selection of humans as targets in these contexts.

<sup>8</sup> Of course, in a regulatory context this gives rise to at least two different vantage points: on the one hand it is argued that it is premature to discuss a prohibition, moratorium or the regulation of the development and future use of AWS; on the other, it is argued that it will be too late to make decisions once the technology has been further developed and the genie is out of the bottle.

current discussions, I identify four primary responses reflected in the literature to the question of whether any human could be held to account for crimes committed through the use of AWS. This analysis is aimed at demonstrating that there are few fixed answers to the accountability question, rather, experts' conclusions differ based on varied assumptions and preconditions. Building on this conclusion, this article argues that international criminal law should follow, not dictate, assessments about the legality or propriety of the use of a given category of weapons, such that accountability is not the right lens through which to examine the issue of whether AWS should be prohibited or regulated.

## II THE ACCOUNTABILITY DEBATE

While AWS are frequently anthropomorphised in both the literature and broader discourse, experts generally agree that it will not be possible to hold an AWS to account — regardless of how sophisticated such weapons might become in the future — because an AWS will never be a responsible moral agent.<sup>9</sup> While the GGE, for example, has not been able to agree on much, it has reached agreement on the fact that accountability cannot be transferred to machines.<sup>10</sup> At the same time, many are of the view, as explored in more detail below,<sup>11</sup> that as a weapon system becomes more autonomous, there will be an increasing separation between what the weapon system does and any proximate human, such that in most scenarios it will be difficult to hold anyone to account for an unlawful outcome resulting from the use of an AWS on the basis of current law.

Prohibitionists argue that this is a reason to ban the development and future use of AWS (or at least place the weapons under a moratorium).<sup>12</sup> Those in the regulatory camp argue that the anticipated potential of AWS should be curtailed

<sup>9</sup> This is asserted as an axiomatic fact by most authors writing on accountability and AWS: see, eg, Swati Malik, 'Autonomous Weapon Systems: The Possibility and Probability of Accountability' (2018) 35(3) *Wisconsin International Law Journal* 609, 627; Noel E Sharkey, 'The Evitability of Autonomous Robot Warfare' (2012) 94(886) *International Review of the Red Cross* 787, 790; Human Rights Council, *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, Christof Heyns*, UN Doc A/HRC/23/47 (9 April 2013) 14 [76] ('*Report of Special Rapporteur, Christof Heyns*'). For a more thoughtful exposition of the issue, see Robert Sparrow, 'Killer Robots' (2007) 24(1) *Journal of Applied Philosophy* 62, 65–8, 71–3. For a contrary view, see John P Sullins, 'When Is a Robot a Moral Agent?' (2006) 6 *International Review of Information Ethics* 23; Jens David Ohlin, 'The Combatant's Stance: Autonomous Weapons on the Battlefield' (2016) 92 *International Law Studies* 1, 2.

<sup>10</sup> Amidst a scant list of 10 'possible guiding principles', the GGE agreed that '[h]uman responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system': *Report of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, UN Doc CCW/GGE.1/2018/3 (23 October 2018) 4 [21(b)] ('*2018 GGE Report*').

<sup>11</sup> See below Part V.

<sup>12</sup> See, eg, *Report of Special Rapporteur, Christof Heyns*, UN Doc A/HRC/23/47 (n 9) 15 [80]; Human Rights Watch and International Human Rights Clinic, *Mind the Gap: The Lack of Accountability for Killer Robots* (Human Rights Watch, 2015) <<https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots#>>, archived at <<https://perma.cc/G2FE-NVUP>> ('*Mind the Gap*'); Darren M Stewart, 'New Technology and the Law of Armed Conflict' (2011) 87 *International Law Studies* 271, 291; Mary Ellen O'Connell, 'Banning Autonomous Killing: The Legal and Ethical Requirement That Humans Make Near-Time Lethal Decisions' in Matthew Evangelista and Henry Shue (eds), *The American Way of Bombing: Changing Ethical and Legal Norms, from Flying Fortresses to Drones* (Cornell University Press, 2014) 224, 236.

— with many currently focusing on a requirement to ensure ‘meaningful human control’.<sup>13</sup> Whether or not an expert calls for prohibition on the one hand, or regulation on the other, is not usually determined by argumentation about what is required to ensure accountability — rather it seems more to reflect whether or not an expert has additional objections to AWS and/or whether they accept that the development of AWS technology is inevitable.

‘Meaningful human control’ is a concept that has emerged to (help) address objections to AWS that are unrelated to the accountability question (eg ethical objections to life and death being dependent upon machine determinations) as well as to encompass the proposal that the development of AWS technology should in some way be limited in order to provide a role for humans that is related to an ability to identify one or more entities with legal personality that are responsible for any violation of the law occasioned by the use of an AWS as a matter of law.<sup>14</sup> It is noted that the concept of ‘meaningful human control’ is not a settled one — the GGE Chair has produced a useful table to capture proposed terminological variations of the concept,<sup>15</sup> which is reproduced below:

---

<sup>13</sup> See, eg, Peter Margulies, ‘Making Autonomous Weapons Accountable: Command Responsibility for Computer-Guided Lethal Force in Armed Conflicts’ in Jens David Ohlin (ed), *Research Handbook on Remote Warfare* (Edward Elgar, 2017) 405; Thompson Chengeta, ‘Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law’ (2016) 45(1) *Denver Journal of International Law and Policy* 1; Malik (n 9) 623; Amos N Guiora, ‘Accountability and Decision Making in Autonomous Warfare: Who Is Responsible?’ [2017] (2) *Utah Law Review* 393. Guiora in fact suggests that ‘accountability standards must be stricter’ for AWS: at 418. Allyson Hauptman, ‘Autonomous Weapons and the Law of Armed Conflict’ (2013) 218 (Winter) *Military Law Review* 170, 193; Jack M Beard, ‘Autonomous Weapons and Human Responsibilities’ (2014) 45(3) *Georgetown Journal of International Law* 617; Michael Aaronson, ‘Robots Don’t Kill People, It’s the Humans We Should Worry About’, *The Conversation* (online, 31 May 2013) <<https://theconversation.com/robots-dont-kill-people-its-the-humans-we-should-worry-about-14779>>, archived at <<https://perma.cc/4UH8-AMMH>>; Thilo Marauhn, ‘An Analysis of the Potential Impact of Lethal Autonomous Weapons Systems on Responsibility and Accountability for Violations of International Law’ (Presentation Paper, CCW Expert Meeting on Lethal Autonomous Systems, 13–16 May 2014) 1; International Committee of the Red Cross, *Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?* (Report, 3 April 2018) 2, 11 <<https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control>>, archived at <<https://perma.cc/WNX9-3X5M>>; Neil Davison, ‘A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law’ in *Perspectives on Lethal Autonomous Weapon Systems* (United Nations, 2017) 5, 17–18; Geneva Academy of International Humanitarian Law and Human Rights, *Autonomous Weapon Systems under International Law* (Academy Briefing No 8, November 2014) 27.

<sup>14</sup> For explorations of the meaning of ‘meaningful human control’, see Michael C Horowitz and Paul Scharre, ‘Meaningful Human Control in Weapons Systems: A Primer’ (Working Paper, Centre for New American Security, March 2015) <<https://www.cnas.org/publications/reports/meaningful-human-control-in-weapon-systems-a-primer>>, archived at <<https://perma.cc/VE78-BBCP>>; Merel Ekelhof, ‘Autonomous Weapons: Operationalizing Meaningful Human Control’, *Humanitarian Law and Policy* (Blog Post, 15 August 2018) <<https://blogs.icrc.org/law-and-policy/2018/08/15/autonomous-weapons-operationalizing-meaningful-human-control/>>, archived at <<https://perma.cc/R54S-FPTM>>.

<sup>15</sup> *Chair’s Summary of the 2018 GGE Discussion*, UN Doc CCW/GGE.1/2018/3 (n 4) [22].

---

(Maintaining)	(Substantive)	Human	(Participation)
(Ensuring)	(Meaningful)		(Involvement)
(Exerting)	(Appropriate)		(Responsibility)
(Preserving)	(Sufficient)		(Supervision)
	(Minimum level of)		(Validation)
	(Minimum indispensable extent of)		(Control)
			(Judgment)
			(Decision)

---

The varying language points to different understandings of the concept. As Swati Malik explains:

While one state may refer to [meaningful human control] as an operator's approval of each action to be taken by a weapon (essentially meaning that at best a weapon system may be a semi-autonomous one), to another it may only require the operator's power to refuse permission for particular courses of action while letting the weapon run on a largely autonomous mode otherwise. There may be others still that consider the programming of a weapon system as sufficient ...<sup>16</sup>

For the sake of simplicity, this article will employ the term 'meaningful human control' (which is the term most commonly employed in current scholarship). While the precise definition of the term is important for an assessment of whether it provides a satisfactory solution to different objections raised in relation to AWS, this is not an issue that needs to be resolved for the purpose of this article. All that needs to be understood is that some states and commentators argue that this would solve 'the accountability problem'.<sup>17</sup>

### III AWS AND INDIVIDUAL CRIMINAL RESPONSIBILITY

Without having reached consensus on whether any regulation of AWS is necessary, the GGE agreed in 2018 that

[a]ccountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control.<sup>18</sup>

Separately, the GGE noted that '[h]umans must at all times remain accountable in accordance with applicable international law for decisions on the use of force'.<sup>19</sup> This of course raises the question of whether a ban on, or restriction of, AWS is necessary to ensure such accountability.

---

<sup>16</sup> Malik (n 9) 623 (citations omitted).

<sup>17</sup> It is noted that alternative restrictions on the future use of AWS have been suggested, such as limiting AWS' use to battlespaces where there are no or few civilians or limiting AWS use to machine-on-machine engagement.

<sup>18</sup> *2018 GGE Report*, UN Doc CCW/GGE.1/2018/3 (n 10) [21(c)].

<sup>19</sup> *Ibid* [23(a)].

At this point, it is important to make a note about nomenclature. Both in the GGE and in broader scholarship, the term ‘accountability’ has sometimes been used as an umbrella term to describe various forms of legal responsibility, including state responsibility, administrative and disciplinary proceedings undertaken in response to violations of IHL (broadly described as military justice), civil liability, and individual criminal responsibility. ‘Accountability’ is used in this article to refer exclusively to individual criminal responsibility, which is the main focus of those preoccupied with the accountability issue and thus the most relevant angle of analysis for current purposes. It is noted that this does not amount to a dismissal of the potential relevance of other forms of legal responsibility. Rather it reflects the fact that no convincing argument has been presented to establish that any unique challenges are raised by AWS in relation to state responsibility, given that the establishment of an internationally wrongful act is not contingent on evidence of intent. It also reflects the fact that most of those with concerns about the accountability gap consider alternative accountability avenues to be insufficient responses to capture the seriousness of the wrong that would be occasioned were a human to be killed unlawfully through the use of an AWS and — at least in the case of civil liability — no more likely to result in a successful case due to various legal hurdles.<sup>20</sup>

This article will, moreover, focus on individual criminal responsibility under international law in order to ensure global relevance given differing approaches to relevant crimes and mental elements under domestic criminal laws. I will draw on relevant provisions of the *Rome Statute of the International Criminal Court*<sup>21</sup> (*Rome Statute*) so that the reader has a ready and relatable reference point for the discussion of principles of international criminal law, although I will note where the *Rome Statute* has departed from pre-existing international criminal law

---

<sup>20</sup> *Mind the Gap* (n 12) 1–4, 13–17, 26–36; Chengeta (n 13) 4–11, 39–47, 50; Christof Heyns, ‘Increasingly Autonomous Weapon Systems: Accountability and Responsibility’ in *Autonomous Weapon Systems: Technical, Military, Legal and Humanitarian Aspects* (International Committee of the Red Cross, 2014) 45, 47 <<https://www.icrc.org/en/publication/4221-expert-meeting-autonomous-weapon-systems>>, archived at <<https://perma.cc/23T6-9X3U>>; Rebecca Crootof, ‘War Torts: Accountability for Autonomous Weapons’ (2016) 164(6) *University of Pennsylvania Law Review* 1347, 1361, 1363; Benjamin Kastan, ‘Autonomous Weapons Systems: A Coming Legal “Singularity”?’ [2013] (1) *Journal of Law, Technology and Policy* 45, 69–78; Kelly Cass, ‘Autonomous Weapons and Accountability: Seeking Solutions in the Law of War’ (2015) 48(3) *Loyola of Los Angeles Law Review* 1017, 1049–53; Malik (n 9) 628–32, 638–40; Timothy McFarland, ‘The Status of Autonomous Weapon Systems under International Humanitarian Law’ (PhD Thesis, The University of Melbourne, 2017) 202–3, 232; Daniele Amoroso and Guglielmo Tamburrini, ‘The Ethical and Legal Case against Autonomy in Weapons Systems’ (2018) 18(1) *Global Jurist* 1, 7; Daniele Amoroso, ‘*Jus in Bello* and *Jus ad Bellum* Arguments against Autonomy in Weapons Systems: A Re-Appraisal’ (2017) 43 *Questions of International Law: Zoom-in* 5, 21–2. Specific arguments relating to the insufficiency of alternative accountability avenues vary but include the essential functions of deterrence and retribution being specific to international criminal law, the right of victims to a remedy, and the differentiated treatment that would be occasioned, compared to like conduct that would attract individual criminal responsibility if perpetrated via non-autonomous weapons. Additional challenges associated with civil claims that have been identified include the paucity of jurisdictions with existing relevant product liability or wrongful death laws, the difficulty of initiating civil proceedings across international borders, issues of sovereign immunity, and the slim prospect of claimants being able to establish fault.

<sup>21</sup> *Rome Statute of the International Criminal Court*, opened for signature 17 July 1998, 2187 UNTS 3 (entered into force 1 July 2002) (*Rome Statute*).

jurisprudence (and where the position under customary international law might therefore be considered unclear).

While AWS could conceivably be used to commit genocide, the crime of aggression or crimes against humanity, this article will use war crimes as illustrative examples given it is violations of IHL that are the focus of most discussions on the subject.<sup>22</sup> In particular, I will make reference to the war crime of attacking civilians — a crime applicable in both international armed conflict ('IAC') and non-international armed conflict ('NIAC') — which is defined as '[i]ntentionally directing attacks against the civilian population as such or against individual civilians not taking direct part in hostilities'.<sup>23</sup> This requires the satisfaction of the following elements:

- 1 The perpetrator directed an attack.
- 2 The object of the attack was a civilian population as such or individual civilians not taking direct part in hostilities.
- 3 The perpetrator intended the civilian population as such or individual civilians not taking direct part in hostilities to be the object of the attack.
- 4 The conduct took place in the context of and was associated with an international armed conflict/armed conflict not of an international character.
- 5 The perpetrator was aware of factual circumstances that established the existence of an armed conflict.<sup>24</sup>

While a detailed analysis of the crime is beyond the scope of this article, it is perhaps useful to note that the crime is largely aimed at violations of the principle of distinction occasioned by the targeting of civilians.<sup>25</sup> Violations of the principle of proportionality are criminalised (in IAC only) under art 8(2)(b)(iv).<sup>26</sup>

---

<sup>22</sup> International Humanitarian Law ('IHL') violations are the focus of the analysis of each of the works discussed below in Part V. See also the dominance of IHL issues in *2018 GGE Report*, UN Doc CCW/GGE.1/2018/3 (n 10). The Chairman of the GGE has stated that the 'IHL prism should permeate all areas of [the GGE's] focus': Letter from Ljupčo Jivan Gjorgjinski to the States Parties to the Convention on Certain Conventional Weapons, 8 February 2019 <<https://bit.ly/2T0Dicc>>, archived at <<https://perma.cc/5VRC-Q8TU>>, quoted in Hayley Evans and Natalie Salmanowitz, 'Lethal Autonomous Weapons Systems: Recent Developments', *Lawfare* (Blog Post, 7 March 2019) <<https://www.lawfareblog.com/lethal-autonomous-weapons-systems-recent-developments>>, archived at <<https://perma.cc/ST9Q-GLD4>>.

<sup>23</sup> *Rome Statute* (n 21) arts 8(2)(b)(i), 8(2)(e)(i).

<sup>24</sup> International Criminal Court, *Elements of Crimes*, Doc No ICC-ASP/1/3 (adopted 9 September 2002) arts 8(2)(b)(i), 8(2)(e)(i).

<sup>25</sup> Attacks on civilian objects and persons and objects that enjoy special protection under IHL are criminalised under *Rome Statute* (n 21) arts 8(2)(b)(ii), (iii), (ix), (xxiv) (applicable in international armed conflict ('IAC')) and arts 8(2)(e)(ii)–(iv) (applicable in non-international armed conflict ('NIAC')).

<sup>26</sup> *Ibid* art 8(2)(b)(iv) establishes the crime of '[i]ntentionally launching an attack in the knowledge that such attack will cause incidental loss of life or injury to civilians or damage to civilian objects or widespread, long-term and severe damage to the natural environment which would be clearly excessive in relation to the concrete and direct overall military advantage anticipated'.

I will also refer to the war crimes of wilful killing<sup>27</sup> (applicable in IAC) and murder<sup>28</sup> (applicable in NIAC). Elements 4 and 5 of these crimes are identical to those applicable to the war crime of attacking civilians. The first three elements of the crimes (with variances indicated by bracketed text, with wilful killing presented first and murder presented second) are:

- 1 The perpetrator killed one or more persons.
- 2 Such person or persons were (protected under one or more of the *Geneva Conventions* of 1949) (either hors de combat, or were civilians, medical personnel, or religious personnel taking no active part in the hostilities).
- 3 The perpetrator was aware of the factual circumstances that established (that protected) (this) status.<sup>29</sup>

The main controversy, as far as AWS are concerned, relates to the satisfaction of the mental elements of each crime. Article 30 of the *Rome Statute* provides further guidance in relation to mental elements, although it is widely acknowledged that it represents an uneasy blend of common law and civil law concepts and contains numerous ambiguities.<sup>30</sup> Article 30 specifies that, unless otherwise provided, the material elements of all crimes must be ‘committed with intent and knowledge’. Paragraph (2)(a) specifies that a person has intent in relation to conduct when ‘that person means to engage in the conduct’. Paragraph (2)(b) provides that a person has intent in relation to a consequence where ‘that person means to cause that consequence or is aware that it will occur in the ordinary course of events’. The reference to ‘means to’ is generally equated to direct intent or *dolus directus* in the first degree.<sup>31</sup> The reference to an awareness that a consequence ‘will occur in the ordinary course of events’ is generally equated with oblique intent or *dolus directus* in the second degree.<sup>32</sup> Knowledge is defined in art 30(3) as meaning ‘awareness that a circumstance exists or a consequence will occur in the ordinary course of events’. In *Prosecutor v Germain Katanga* (*Katanga*), the International Criminal Court (‘ICC’) Trial Chamber held that the latter required ‘virtual certainty’.<sup>33</sup>

---

<sup>27</sup> Ibid art 8(2)(a)(i).

<sup>28</sup> Ibid art 8(2)(c)(i) defines the crime as ‘[v]iolence to life and person, in particular murder of all kinds, mutilation, cruel treatment and torture’. The *Elements of Crimes* separates this into four separate war crimes: *Elements of Crimes* (n 24) art 8(2)(c).

<sup>29</sup> *Elements of Crimes* (n 24) arts 8(2)(a)(i), 8(2)(c)(i)-1.

<sup>30</sup> For explanations of the operation of art 30 and how its provisions correspond with domestic criminal law approaches to the mental elements of crimes, see generally Roger S Clark, ‘The Mental Element in International Criminal Law: The Rome Statute of the International Criminal Court and the Elements of Offences’ (2001) 12(3) *Criminal Law Forum* 291; Sarah Finnin, ‘Mental Elements under Article 30 of the Rome Statute of the International Criminal Court: A Comparative Analysis’ (2012) 61(2) *International and Comparative Law Quarterly* 325.

<sup>31</sup> Finnin (n 30) 341.

<sup>32</sup> Ibid 343–4.

<sup>33</sup> *Prosecutor v Katanga (Judgment Pursuant to Article 74 of the Statute)* (International Criminal Court, Trial Chamber II, Case No ICC-01/04-01/07, 7 March 2014) [776] (*Katanga*). The Trial Chamber continued, saying that ‘this form of criminal intent presupposes that the person knows that his or her actions will necessarily bring about the consequence in question, barring an unforeseen or unexpected intervention or event to prevent its occurrence. In other words, it is nigh on impossible for him or her to envisage that the consequence will not occur’: at [777].

It is important to underscore in this context that, while it was recognised in the jurisprudence of the International Criminal Tribunal for the former Yugoslavia and the International Criminal Tribunal for Rwanda,<sup>34</sup> under art 30 of the *Rome Statute*, individual criminal responsibility for relevant war crimes cannot be established where there is only evidence of a mental element of recklessness<sup>35</sup> or *dolus eventualis*.<sup>36</sup> International criminal law has also not generally recognised negligence as being sufficient to establish criminal guilt for the majority of international crimes,<sup>37</sup> and it is clearly excluded under art 30 of the *Rome Statute*. The only exceptions to the general rule provided in art 30 are a handful of crimes that specify a lower mental element threshold in their elements, such as the war crime of using, conscripting or enlisting children,<sup>38</sup> which are not particularly relevant in the AWS context, and where individual criminal responsibility is established on the basis of command or superior responsibility, which is discussed below. The key point for present purposes is that the threshold for the requisite mental elements of a direct perpetrator of war crimes under international criminal law, especially under the *Rome Statute*, is high.

Applying art 30 to specific crimes is not straightforward as ‘conduct’ and ‘consequence’ are not defined in the *Rome Statute*, and the *Elements of Crimes* do not specify how individual elements are to be categorised. In addition, the *Elements of Crimes* specify a mental element for some crimes that does not amount to a special intent, but is rather a reflection of an application of art 30. This is the case in relation to the war crimes outlined above. For example, in *Katanga*, the Trial Chamber held that the third element of the war crime of attacking civilians ‘is, in fact, a repetition of article 30(2)(a)’.<sup>39</sup> It further held that ‘the second element of the *Elements of Crimes*, to which it applies, namely, that “[t]he object of the attack was a civilian population as such or individual

---

<sup>34</sup> For an overview of relevant jurisprudence, see Gerhard Werle and Florian Jessberger, ‘“Unless Otherwise Provided”: Article 30 of the ICC Statute and the Mental Element of Crimes under International Criminal Law’ (2005) 3(1) *Journal of International Criminal Justice* 35, 53–4.

<sup>35</sup> While early drafts of the *Rome Statute* would have included recklessness in art 30, it was ultimately omitted and cannot be read into art 30: *Prosecutor v Lubanga (Decision on the Confirmation of Charges)* (International Criminal Court, Pre-Trial Chamber I, Case No ICC-01/04-01/06, 29 January 2007) 121 [355] n 438; *Prosecutor v Bemba Gombo (Decision Pursuant to Article 61(7)(a) and (b) of the Rome Statute on the Charges of the Prosecutor)* (International Criminal Court, Pre-Trial Chamber II, Case No ICC-01/05-01/08, 15 June 2009) [360]. See also William A Schabas, *The International Criminal Court: A Commentary on the Rome Statute* (Oxford University Press, 2<sup>nd</sup> ed, 2016) 630. There is no universal definition of the common law concept of recklessness, but it is generally understood as requiring a perpetrator to have been aware of the existence of an unreasonable and unjustified risk that the objective element of the crime may result from his or her actions: Finnin (n 30) 334–5.

<sup>36</sup> There has been debate over whether recklessness can be equated to the civil law concept of *dolus eventualis* (based on differing views as to whether the latter in addition requires the perpetrator to be ‘reconciled’ with the harm as a possible cost of his or her course of conduct) and whether the latter could be read into art 30: see Finnin (n 30) 333–5, 346–9. However, the ICC Trial Chamber in *Katanga* confirmed that *dolus eventualis* is excluded from art 30: *Katanga* (n 33) [775].

<sup>37</sup> Clark (n 30) 300–1, 321; Werle and Jessberger (n 34) 54; Crootof (n 20) 1384–5. Negligence is established where a perpetrator has reason to believe that the prohibited result will not occur although he or she sees it as a possibility.

<sup>38</sup> *Rome Statute* (n 21) art 8(2)(b)(xxvi); *Elements of Crimes* (n 24) art 8(2)(b)(xxvi).

<sup>39</sup> *Katanga* (n 33) [806].

civilians not taking direct part in hostilities”, must be regarded as conduct’.<sup>40</sup> Applying these principles to the war crime of attacking civilians, the Trial Chamber held that it was necessary for the Prosecutor to establish the following mental elements:

[T]he perpetrator must have (1) intentionally directed an attack; (2) intended the civilian population or individual civilians to be the object of the attack; (3) been aware of the civilian character of the population or of civilians not taking part in hostilities; and (4) been aware of the factual circumstances that established the existence of an armed conflict.<sup>41</sup>

Of the war crime of murder, the Trial Chamber held that

the perpetrator must have intentionally killed one or more persons. Such intent will be proven where the perpetrator acted deliberately or failed to act (1) in order to cause the death of one or more persons or (2) whereas he or she was aware that death would occur in the ordinary course of events. Moreover, pursuant to article 8(2)(c)(i)-1(3) of the *Elements of Crimes*, the perpetrator must also have been aware of the factual circumstances that established the status of the victims.<sup>42</sup>

The posited difficulty where death unlawfully results from the deployment of an AWS is whether it will ever be possible to establish that any individual can be said to have been ‘virtually certain’ of the factual circumstances that established the civilian character/protected status of victims, and in the case of the war crime of attacking civilians, ‘intended’ to attack a civilian population or individual civilians. The principal issue is that as a result of the nature of AI ‘[n]either the programming nor the command data inputted to [AWS] prior to their deployment on a particular operation will necessarily result in a specific outcome in response to any given set of circumstances’.<sup>43</sup> This means that in many cases it would be difficult to establish the requisite mental element on the part of any human under international criminal law. Alternatively, a defendant may be able to rely on mistake of fact, which provides a defence where this would negate the mental element of a crime under art 32(1) of the *Rome Statute*.<sup>44</sup>

Even where relevant war crimes apply the lower mental element standard of recklessness/*dolus eventualis* in other jurisdictions, it is not clear that such crimes could be established on the facts. Considering the role of the operator and capability of the AWS in the posited accountability test scenario, it cannot clearly be concluded that the use of the AWS could be said to amount to a deviation from the standard of conduct that a reasonable person would observe in the same situation.

Having said this, it is important to recognise that a range of scenarios exist in which it is possible to identify a human who could be held to account for a war crime perpetrated through the use of AWS — at least in theory.<sup>45</sup>

---

<sup>40</sup> Ibid, quoting *Elements of Crimes* (n 24) art 8(2)(b)(i)(2).

<sup>41</sup> *Katanga* (n 33) [808].

<sup>42</sup> Ibid [793].

<sup>43</sup> Stewart (n 12) 290.

<sup>44</sup> Clark (n 30) 308.

<sup>45</sup> It is acknowledged that the practicalities of obtaining evidence to facilitate the prosecution of war crimes is exceedingly challenging: see generally Jacob Katz Cogan, ‘The Problem of Obtaining Evidence for International Criminal Courts’ (2000) 22(2) *Human Rights Quarterly* 404.

For example, if an AWS targeted and killed civilians as a result of unlawful targeting parameters deliberately provided to the AWS by its operator,<sup>46</sup> a prosecutor is likely to be able to establish that the requisite mental elements of the war crimes of attacking civilians and wilful killing/murder are satisfied, given the operator will have been aware that the AWS would use lethal force consistent with the command data he or she supplied to it.<sup>47</sup>

Similarly, if it was known that an AWS could not accurately and reliably determine whether or not a person was directly participating in hostilities and that the AWS was being deployed in a geographic area with a high density of civilians, there is a good chance that the operator (or individual in the chain of command responsible for the deployment of the AWS) could at least be held responsible for wilful killing (in IAC) or murder (in NIAC), on the basis that he or she was aware of the factual circumstances that established the status of victims killed.<sup>48</sup>

If there was an unintentional error in an AWS's algorithmic programming that caused it to engage targets indiscriminately, but this became apparent, and the AWS continued to be used, it is likely that the individual decision-maker responsible for the AWS's continued use would possess the requisite mental element to establish both the war crime of attacking civilians and wilful killing/murder.<sup>49</sup>

Conversely, where an AWS has been deliberately programmed by a developer to target civilians, that programmer could be prosecuted on the basis of the doctrine of indirect perpetration (if they effectively acted through an innocent commander/operator who unknowingly deployed the weapon), or on the basis of

---

<sup>46</sup> The term 'operator' is used advisedly here. Depending on the degree of autonomy and type of weapon system, the operator of an AWS may not have control over the weapon in the same sense as an operator of a non-autonomous weapon. The 'operator' of an AWS may be the person who simply activates or supervises the weapon: see Beard (n 13) 652.

<sup>47</sup> Similar conclusions are reached by Ian S Henderson, Patrick Keane and Josh Liddy, 'Remote and Autonomous Warfare Systems: Precautions in Attack and Individual Accountability' in Jens David Ohlin (ed), *Research Handbook on Remote Warfare* (Edward Elgar, 2017) 335, 359; Hin-Yan Liu, 'Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems' in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 325, 329; Davison (n 13) 17.

<sup>48</sup> Similar conclusions are reached in Dan Saxon, 'Autonomous Drones and Individual Criminal Responsibility' in Ezio Di Nucci and Filippo Santoni de Sio (eds), *Drones and Responsibility: Legal, Philosophical, and Sociotechnical Perspectives on Remotely Controlled Weapons* (Routledge, 2016) 17, 19, 26; Davison (n 13) 17. It is noted that the International Criminal Court ('ICC') has held, while stressing that this is not automatic, that

indiscriminate attacks ... may qualify as intentional attacks against the civilian population or individual civilians, especially where the damage caused to civilians is so great that it appears to the Chamber that the perpetrator meant to target civilian objectives. Use of weaponry that has indiscriminate effects may, inter alia, show that the attack was directed at the civilian population or individual civilians.

*Katanga* (n 33) [802]. As such, it is possible that the war crime of attacking civilians could also be established. For a concurring view, see Cass (n 20) 1058–64.

<sup>49</sup> A similar conclusion is reached by Henderson, Keane and Liddy (n 47) 361; Saxon (n 48) 33.

the doctrine of co-perpetration (if they acted in concert with military members pursuant to a common plan or agreement).<sup>50</sup>

Where an operator has the capacity to override target selection and engagement, effectively exercising a power of veto, the operator could potentially be held individually criminally responsible if civilians are unlawfully killed — but this is where things start to get sticky. The potential speed of the cycle of target identification, assessment and engagement by an AWS could well mean that an operator will have no realistic ability to intervene. He or she may also not have any alternative source of information on which to base a meaningful decision.<sup>51</sup> If there is no volitional act or cognitive component in relation to target selection and engagement by the operator, it would not be possible to establish the requisite mental element of the above-outlined war crimes on the part of the operator.

And here we get to the heart of things. Let us set aside the scenarios outlined above where we can relatively easily identify some human wrongdoing. Instead, imagine a world in the future — possibly in the distant future — in which robotic and AI technologies have developed to the point that an AWS, when used as intended, has passed an article 36 weapon review and it has otherwise been demonstrated on the basis of testing and validation that the AWS can be used in a battlespace in compliance with applicable rules of IHL.<sup>52</sup> It is acknowledged

---

<sup>50</sup> Tim McFarland and Tim McCormack, ‘Mind the Gap: Can Developers of Autonomous Weapons Systems Be Liable for War Crimes?’ (2014) 90 *International Law Studies* 361, 375. This of course assumes that it would be possible to identify a specific individual who was responsible for programming decisions, which is likely to be difficult given the degree of collaboration required for complex coding of the sort that would be required for an AWS. Indeed, Jack Beard argues that

[d]ecision-making by these machines will probably be distributed across several programs and processors, results will be derived from data input that originates from many different types of sensors, and no single agent may be able to be identified as the ‘decision-maker’ in a complicated system of concurrently interacting human and machine components. These machines will be further characterized by interactions between many different programs that have been developed by multiple programmers (perhaps previously unknown to each other), on different processors and operating systems, possibly stretching across a network of different autonomous platforms.

Beard (n 13) 651 (citations omitted).

<sup>51</sup> See Henderson, Keane and Liddy (n 47) 362; Kenneth Anderson and Matthew C Waxman, ‘Law and Ethics for Autonomous Weapon Systems: Why a Ban Won’t Work and How the Laws of War Can’ (Research Paper No 2013-11, American University Washington College of Law, 2013) 22.

<sup>52</sup> *Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I)*, opened for signature 8 June 1977, 1125 UNTS 3 (entered into force 7 December 1978) art 36 (‘API’) provides that

[i]n the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this *Protocol* or by any other rule of international law applicable to the High Contracting Party.

that this prospect is speculative: there is an array of views as to whether the use of AWS will ever be able to comply with the principle of distinction,<sup>53</sup> let alone proportionality and precautions in attack.<sup>54</sup> However, for the purpose of exploring whether there is an accountability gap it is not necessary to decide whether this is likely, or even possible, we just need to assume for the purposes of illustration that this imagined world has arrived. Also exclude the possibility of any human having intentionally programmed or otherwise deliberately employed the weapon to violate the rules of IHL. Instead, imagine that, as a result of intelligent control and machine learning, an AWS has operated in a manner that was not predicted and that this has resulted in civilians being killed in violation of IHL before any human had a realistic opportunity to halt the use of the weapon system.<sup>55</sup> Could anyone be held to account in this scenario?

---

It is noted that many countries at the forefront of AWS technology development are not party to *API*. There is a strong argument that the requirement to review new means of warfare is a rule of customary international law. The United States does not accept this, but undertakes such reviews as a matter of policy and domestic law: Natalia Jevglevskaja, 'Weapons Review Obligation under Customary International Law' (2018) 94 *International Law Studies* 186, 213. This said, the necessary content of these weapons reviews is contested. The narrowest narrow approach considers only whether the weapon is of a nature that would cause unnecessary suffering or superfluous injury, and can be directed at a specific military objective. An AWS is unlikely to be prohibited on these bases. A broader approach considers whether the weapon's ordinary use would otherwise comply with IHL, which raises far more difficult technical and legal questions. In order to sidestep debates over the appropriate standard of review for new weapons, the scenario used in this article assumes that the AWS, at least at the time of its activation, was capable of complying with IHL generally.

<sup>53</sup> Although, as Michael Schmitt observes, an AWS that is incapable of distinguishing between civilians and combatants may nonetheless be lawfully used under IHL if a battlespace was devoid of humans: Michael N Schmitt, 'Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics' (2013) *Harvard National Security Journal Features* 1, 11 ('A Reply to the Critics') <<https://harvardnsj.org/2013/02/autonomous-weapon-systems-and-international-humanitarian-law-a-reply-to-the-critics/>>, archived at <<https://perma.cc/J7V6-BFCH>>.

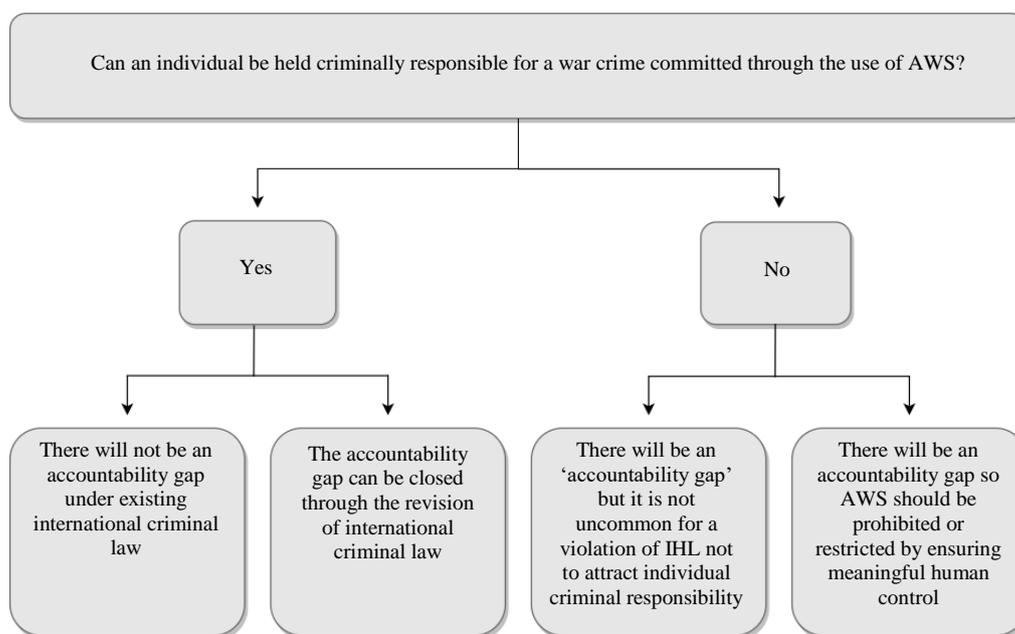
<sup>54</sup> Many arguments that the use of AWS could not accord with IHL are premised on an assumption that robotics and artificial intelligence ('AI') will not be able to meet various technological challenges. Given the rapid development of relevant technologies, this is something that, in my view, cannot currently be known. Others argue that IHL is addressed to humans and actually requires the exercise of *ex tempore* human judgement. This is a more complex argument, beyond the scope of this article. For an outline, see Eric Talbot Jensen, 'The Human Nature of International Humanitarian Law', *Humanitarian Law and Policy* (Blog Post, 23 August 2018) <<https://blogs.icrc.org/law-and-policy/2018/08/23/human-nature-international-humanitarian-law/>>, archived at <<https://perma.cc/J7Z2-5KUD>>.

<sup>55</sup> The harm inflicted in such a scenario could be considerable. As Paul Scharre observes, an autonomous weapon could continue engaging inappropriate targets until it exhausts its magazine, potentially over a wide area. If the failure mode is replicated in other autonomous weapons of the same type, a military could face the disturbing prospect of large numbers of autonomous weapons failing simultaneously, with potentially catastrophic consequences.

Paul Scharre, Center for a New American Security, *Autonomous Weapons and Operational Risk* (Report, February 2016) 5 <[https://s3.amazonaws.com/files.cnas.org/documents/CNAS\\_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515](https://s3.amazonaws.com/files.cnas.org/documents/CNAS_Autonomous-weapons-operational-risk.pdf?mtime=20160906080515)>, archived at <<https://perma.cc/EM5N-7FQH>>.

## IV RESPONSES TO THE ACCOUNTABILITY QUESTION

Experts remain divided on this question. We have reached a point, however, where it is possible to identify four primary approaches to the problem, as illustrated in the following figure:



Each of these approaches is discussed briefly below. The following does not purport to be an exhaustive survey of the literature on point, but it is contended that the sample is sufficient to illustrate the emergence of the four dominant approaches illustrated in the above figure.

It is stressed that it is not necessarily the case that experts as such can be divided into four different camps: reflecting the complexity of the issue, many publications canvass more than one view (eg there is an accountability gap, and the law could be amended, but this carries the policy risk of scapegoating proximate individuals). It is also not the case that experts necessarily either advocate for or against the development and use of AWS: many authors present nuanced views (eg there is an accountability gap, we should look at whether the criminal law can properly be amended, and until we work through these issues (and have greater fidelity around what autonomy means) we should restrict the development, or at least the deployment, of AWS). Thus, rather than serving as a taxonomy, these 'approaches' have been identified as part of a process of distilling the essence of arguments being made about accountability and AWS in order to tease out precisely why some experts say there will be no accountability gap while others argue there will, why some do not see this as a problem and others do, and why some consider this is a problem that has a solution, while others see it as irresolvable.

## V IS THERE AN ACCOUNTABILITY PROBLEM?

There are some who deny there will be any accountability gap. Those who adopt this approach can be divided into two sub-categories. The first group set their analytic parameters narrowly and only consider scenarios similar to those addressed in the section above where there is some relatively obvious human wrongdoing (malicious programming, deliberate targeting commands, etc).<sup>56</sup> As such, these analyses shed little light on whether there is in fact an accountability gap, let alone whether this should inform a prohibition, or the restriction, of AWS.

The second group advance a more sophisticated version of this argument. They effectively contend that the two key parameters for the accountability test scenario outlined in the section above are mutually exclusive or otherwise implausible. Michael Schmitt has rejected the very possibility of an unpredictable AWS:

robots will not 'go rogue.' While autonomous ... weapon systems will be susceptible to malfunction, that is also the case with weapon systems ranging from catapults to computer attack systems. Like a missile that 'goes ballistic' (loses guidance), future autonomous systems could fall out of parameters. But the prospect of them 'taking on a life of their own' is an invention of Hollywood.<sup>57</sup>

Charles J Dunlap Jr essentially asserts that AWS could only lawfully be deployed in scenarios that would allow for accountability. He says the

belief that there can be no accountability because, in their view, autonomous weapons can act 'unforeseeably' is obviously wrong because deploying a weapon that is expected to launch attacks 'unforeseeably' is itself a punishable breach of the responsibilities of commanders, operators, and the nations they represent.<sup>58</sup>

In somewhat more granularity, Timothy McFarland advances a similar view:

Saying that no person is accountable for those targeting and attack processes, due to lacking control over them, is saying that the AWS was deployed on that attack despite no person being in a position to take the required precautions to ensure that only legal targets are attacked. Utilising a weapon system in an attack when it is not possible to take sufficient precautions to ensure that only legal targets are struck is an illegal act in itself, in violation of the principle of distinction. If those

---

<sup>56</sup> See, eg, Cass (n 20) 1058–64.

<sup>57</sup> Schmitt, 'A Reply to the Critics' (n 53) 7. An almost identical quote appears in Michael N Schmitt and Jeffrey S Thurnher, "'Out of the Loop": Autonomous Weapon Systems and the Law of Armed Conflict' (2013) 4(2) *Harvard National Security Journal* 231, 242.

<sup>58</sup> Charles J Dunlap Jr, 'Accountability and Autonomous Weapons: Much Ado about Nothing' (2016) 30(1) *Temple International and Comparative Law Journal* 63, 71 (citations omitted).

precautions are taken, they would amount to an exercise of control over the AWS which would bridge any accountability gap.<sup>59</sup>

I am not confident that these assertions can be proven to be correct, at least not currently. In the first place, the notion that an AWS could — to put it in simplistic terms — be programmed not to kill civilians unlawfully, and that this instruction could not, in any circumstances, be overridden by the AWS, is, in my view, questionable. As Jonathan Tapson has written:

Until we see an AI do the utterly unexpected, we don't even realise that we had a limited view of the possibilities. AIs move effortlessly beyond the limits of human imagination.

...

How do you prevent an AI from using such methods when you don't actually know what its methods are?<sup>60</sup>

In light of what we already know about deep learning, the ability for an algorithm to rewrite parts of its code, and the ability of AI to create AI, it does not presently seem plausible to rule out a scenario where an AWS 'goes rogue'. Indeed, in a later blog piece, Schmitt himself wrote:

I am concerned about very advanced learning systems, for experts tell me it is theoretically possible that such systems will learn the 'wrong lesson.' I asked one expert whether it is possible that an AWS might eventually 'learn' that it is more efficient to kill all humans rather than attack only those that meet preset parameters for lawful targets. His answer was certainly not in the present or immediately foreseeable future, but he could not absolutely rule out the possibility as AI technologies improve.<sup>61</sup>

It is difficult to disentangle such capability considerations from issues around the level of predictability that an AWS would be required to have in order for its deployment to be considered lawful under IHL and whether or not this would in fact preserve conditions that would enable the satisfaction of the mental element

---

<sup>59</sup> McFarland (n 20) 195 (citations omitted). It is acknowledged that despite reaching this conclusion (which is elaborated on over several pages and repeated in his overall conclusions on accountability), Timothy McFarland goes on to identify scenarios in which it would not be possible to establish individual criminal responsibility were weapons with very high levels of autonomy to be deployed: at 206–30. The disconnect is not explained. In another, somewhat softer, variation, Marco Sassóli similarly argues that 'I do not think that the possession of autonomous decision-making capacity breaks the causal chain allowing attribution and responsibility, because I assume that it is always humans who define how this autonomy will function': Marco Sassóli, 'Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to Be Clarified' (2014) 90 *International Law Studies* 308, 324–5 (citations omitted). He further argues that autonomy entails making 'decisions within a framework the robot is unable to override. ... The risk that the robot "goes rogue" must be avoided in the way the robot is devised. If this is not possible, such weapons must be outlawed': at 326.

<sup>60</sup> Jonathan Tapson, 'Google's Go Victory Shows AI Thinking Can Be Unpredictable, and That's a Concern', *The Conversation* (online, 18 March 2016) <<https://theconversation.com/googles-go-victory-shows-ai-thinking-can-be-unpredictable-and-thats-a-concern-56209>>, archived at <<https://perma.cc/6KUA-LJX6>>.

<sup>61</sup> Michael Schmitt, 'Regulating Autonomous Weapons Might Be Smarter than Banning Them', *Just Security* (Blog Post, 10 August 2015) <<https://www.justsecurity.org/25333/regulating-autonomous-weapons-smarter-banning/>>, archived at <<https://perma.cc/PTP8-CPF5>>.

of relevant war crimes. Even if it is assumed that an expansive approach to weapons review is the proper one, it is unrealistic, in my view, to insist on perfection. In other words, it is accepted that in hard cases human combatants make errors in distinguishing between combatants and civilians. While a prudent policy maker might insist on an AWS performing with greater accuracy than humans when tested under like conditions, my view is that a 100% success rate, 100% of the time, would not be required for an AWS to be lawfully deployed. This would be the case so long as tolerated errors correlated to the types of cases in which humans would also struggle to apply targeting laws — or indeed, where there are differing views on what the law requires (consider, for example, contested interpretations of the definition of directly participating in hostilities). Also, while a state might be expected to test and validate a wide range of variables in order to be confident that an AWS can comply with IHL, it should be acknowledged that no software testing can ever literally be exhaustive.

As such, it seems to me to be at least equally plausible that an AWS could comply with any fair interpretation of IHL at the time of its deployment as a result of extensive testing and the input of operation parameters satisfying the requirement to take precautions in attack, but that the AWS could subsequently ‘go rogue’ as a result of the inherent unpredictability of AI. It is emphasised that the notion of ‘going rogue’ is not intended to suggest that it is likely or even possible that the future will see something akin to Skynet’s threatened attack on humanity in the *Terminator* movie series. Consider instead the fact that there is already evidence of computers acting ‘deceptively’. For example, in the course of research by Stanford University and Google in relation to the use of a neural network to transform aerial images into street maps and back again, it was discovered that the network created techniques to avoid performing the requested task.<sup>62</sup> In this example, the neural network in question had not been instructed to avoid the techniques it employed, which highlights that this was a problem between keyboard and chair. Seen in this light, ‘going rogue’ might simply amount to an AWS identifying the smallest crevice between instructions in the course of identifying the most efficient pathway to task outcome.

For these reasons, I do not think that the assumption of narrow scenario parameters solves the accountability problem. In my view, at least in certain scenarios, an unpredictable AWS could be lawfully deployed under IHL, and its unpredictability could well mean that the AWS operator would lack the requisite mental element to establish relevant war crimes.

For some, this conclusion is not problematic. They draw an analogy between an unpredictable AWS and a conventional weapon’s malfunction or inherent failure rate that results in unintended civilian deaths, noting that such events would not be prosecutable as war crimes.<sup>63</sup> Indeed, on this view, it is possible to go further and argue that the AWS scenario under consideration is no different from a situation where a human combatant accidentally kills a civilian because of fatigue, or fear, or the drawing of a mistaken conclusion from incomplete

---

<sup>62</sup> Devin Coldewey, ‘This Clever AI Hid Data from Its Creators to Cheat at the Appointed Task’, *TechCrunch* (Blog Post, 31 December 2018) <<https://techcrunch.com/2018/12/31/this-clever-ai-hid-data-from-its-creators-to-cheat-at-its-appointed-task/>>, archived at <<https://perma.cc/4Y5H-V2WC>>.

<sup>63</sup> Henderson, Keane and Liddy (n 47) 361, 363, 369; Saxon (n 48) 28–9.

information: the human combatant could not be successfully prosecuted under international criminal law for any crime that requires intentional targeting or actual knowledge (possibly even constructive knowledge), because the requisite mental element could not be established. It is certainly not the case that all violations of IHL attract individual criminal responsibility under international criminal law: the simple fact is that the bar for accountability has been set high.

Others see the accountability gap as deeply problematic: they see it as essential that an individual be held to account for a violation of IHL, or at least one that results in the death of civilians. Thompson Chengeta perhaps expresses this most rawly: ‘accountability is the crux of International Law — without accountability, we may as well forget about it’.<sup>64</sup>

Sometimes the view that either the law or morality demands accountability is asserted without any further justification. Other times, various rationales are proffered, usually one or more of: accountability is essential to deterrence and prevention (and therefore the protection of civilians);<sup>65</sup> accountability is fundamental to a victim’s right to a remedy;<sup>66</sup> a lack of accountability will result in the unleashing of the dogs of war;<sup>67</sup> or the circular argument that IHL imposes a duty to prosecute war crimes.<sup>68</sup> One might also suggest that some authors’ insistence on accountability is inextricably linked to a visceral reaction to the idea of life and death outcomes being determined by machines. There is no need to get into the merits of these arguments (at least some of which are highly questionable, in my view). The important point for current purposes is that this boils down to a fixed view that there must be accountability for unlawful civilian deaths in armed conflict that result from the use of an AWS.

Several attempts have been made to identify how international criminal law could be revised to close the accountability gap. A number of experts have suggested a revision of the doctrine of command responsibility. In one of the more detailed expositions, Peter Margulies argues that, with ‘a modest revision of the doctrine’ to extend its application to the supervision of machines,<sup>69</sup> command responsibility would apply if there was a requirement of ‘dynamic diligence’ on the part of commanders.<sup>70</sup> He suggests that this would require: a dedicated command structure; technical expertise; real time human monitoring (including an AWS capability to request a review); periodic and frequent review

---

<sup>64</sup> Chengeta (n 13) 49. See also Malik (n 9) 620; Guiora (n 13) 398; *Mind the Gap* (n 12); *Report of Special Rapporteur, Christof Heyns*, UN Doc A/HRC/23/47 (n 9) [75].

<sup>65</sup> See, eg, *Report of Special Rapporteur, Christof Heyns*, UN Doc A/HRC/23/47 (n 9) [75].

<sup>66</sup> See, eg, Chengeta (n 13) 5.

<sup>67</sup> See, eg, Guiora (n 13) 398: “[k]ill/not kill” decisions authorized by the nation-state where standards of accountability are neither inherent nor integral is akin to authorizing the new Wild West”; Stewart (n 12) 292, referring to ‘the broader public policy issues associated with the possibility of military operations being conducted in a “blameless environment”’.

<sup>68</sup> Circular because, while art 146 of the *Geneva Convention Relative to the Protection of Civilian Persons in Time of War*, opened for signature 12 August 1949, 75 UNTS 287 (entered into force 21 October 1950) does require grave breaches (which include ‘wilful killing’) to be criminalised and prosecuted, no ‘wilful killing’ has been committed if the requisite mental element is lacking under international criminal law. The argument that there is an obligation to prosecute a broader range of war crimes under customary international law does not remedy this circularity problem. For an example of such an argument, see *Mind the Gap* (n 12) 15.

<sup>69</sup> Margulies (n 13) 441.

<sup>70</sup> *Ibid* 406.

of outputs; the input of dynamic parameters governing AWS use in relation to time, distance and maximum expected collateral damage; and that target selection decisions be interpretable and transparent.<sup>71</sup>

Others, however, are sceptical that command responsibility provides a workable construct. Even assuming that the interpersonal superior–subordinate relationship that underpins the doctrine could be transposed to the ‘relationship’ between an operator (or supervising commander) and an AWS, liability would require evidence that, at the very least, the ‘commander’ should have known, owing to the circumstances at the time, that crimes were about to be, or were, committed.<sup>72</sup> While this sets a lower mental element standard (negligence) compared to that required for perpetrators, even this standard may not be satisfied in the case of an unpredictable AWS where, to the best of the commander’s knowledge, the AWS would comply with IHL targeting requirements, unless it could be established that the commander had a good grasp of AI and the residual risk of unpredictability. Even so, the doctrine would also relevantly require evidence that the commander failed to take all necessary and reasonable measures within his or her power to prevent or repress the commission of crimes.<sup>73</sup> The *Rome Statute* additionally requires evidence of a failure to exercise control properly over the subordinate.<sup>74</sup> While the law on these aspects of command responsibility might be said to be unsettled following the controversial ICC Appeals Chamber decision in *Prosecutor v Bemba Gombo*,<sup>75</sup> in which the accused was acquitted, at least in part, due to the degree

---

<sup>71</sup> Ibid 431–40. See also Hauptman (n 13) 194–5; Marauhn (n 13) 1, 5; Stewart (n 12) 292.

<sup>72</sup> See *Rome Statute* (n 21) art 28.

<sup>73</sup> Referral for investigation and prosecution by independent authorities being irrelevant in the case of AWS.

<sup>74</sup> *Rome Statute* (n 21) art 28(a)(ii).

<sup>75</sup> *Prosecutor v Bemba Gombo (Judgment on the Appeal against Trial Chamber III’s Judgment Pursuant to Article 74 of the Statute)* (International Criminal Court, Appeals Chamber, Case No ICC-01/05-01/08 A, 8 June 2018). Jean-Pierre Bemba Gombo was acquitted on appeal from a conviction for crimes against humanity and war crimes on the basis of command responsibility. The majority of the Appeals Chamber held, inter alia, that

[t]he scope of the duty to take ‘all necessary and reasonable measures’ is intrinsically connected to the extent of a commander’s material ability to prevent or repress the commission of crimes or to submit the matter to the competent authorities for investigation and prosecution.

At [167]; and that ‘[a]n assessment of whether a commander took all “necessary and reasonable measures” must be based on considerations of what crimes the commander knew or should have known about and at what point in time’: at [168]. The majority further held that ‘it is not the case that a commander is required to employ every single conceivable measure within his or her arsenal, irrespective of considerations of proportionality and feasibility’: at [169]; and said it was necessary ‘to consider other parameters, such as the operational realities on the ground at the time faced by the commander’: at [170]. The majority emphasised that

[t]here is a very real risk, to be avoided in adjudication, of evaluating what a commander should have done with the benefit of hindsight. Simply juxtaposing the fact that certain crimes were committed by the subordinates of a commander with a list of measures which the commander could hypothetically have taken does not, in and of itself, show that the commander acted unreasonably at the time. The trial chamber must specifically identify what a commander should have done *in concreto*.

of deference paid to the fact he was a remote commander, nearly all of the case law on point requires some real dereliction of duty on the commander's part — the controversy or uncertainty pertains to exactly how dirty a commander's hands must be. As such, while command responsibility might be viewed as an attractive fix on the basis that it would recognise the autonomy of an AWS, it is unlikely to bridge the accountability gap where there is a lack of any real wrongdoing on the operator/commander's part. Where we have evidence of such wrongdoing, we are likely to be able to characterise the operator/commander as a perpetrator, without needing to rely on a revised doctrine of command responsibility that equates an autonomous machine with a subordinate force.<sup>76</sup>

Others have focused attention on the AWS developer or programmer, who they argue 'will exert greater control over not only the range of actions the weapons system is capable of performing, but over the specific actions that it in fact performs after being deployed'.<sup>77</sup> Tim McFarland and Tim McCormack have explored the possibility of programmers being prosecuted as accessories on the basis that they aided, abetted or otherwise assisted with a crime's commission, including providing the means for the crime's commission.<sup>78</sup> They conclude that this is unlikely to result in successful prosecutions under current law. In the first place, in most cases, it will be difficult to establish that the programmer's conduct took place in the context of, and was associated with, an armed conflict (as required for the prosecution of war crimes) because in most cases, programming will be completed in the weapon's development phase, before the relevant armed conflict commenced.<sup>79</sup> McFarland and McCormack also demonstrate that, to date, the law has not clearly recognised that acts constituting aiding and abetting may occur before the principal crime has occurred.<sup>80</sup> Most problematic, however, in their view, is the requisite mental element for accessory liability:

[p]rosecutors would be required to demonstrate that a weapon developer acts not only with awareness of the eventual physical perpetrator's intention to commit the crime and with the knowledge that his or her conduct would assist in perpetration of the offense, but also that s/he acts for the purpose of facilitating the crime ...<sup>81</sup>

In this way, we again run into the hurdle of needing evidence of deliberate wrongdoing in order to satisfy the requisite mental element. Nonetheless, McFarland and McCormack state that this 'does not mean that updating, amendment, revision and reform of the law is out of the question':

---

At [170]. On the facts, the majority held, inter alia, that 'the Trial Chamber paid insufficient attention to the fact that the MLC troops were operating in a foreign country with the attendant difficulties on Mr Bemba's ability, as a remote commander, to take measures': at [171].

<sup>76</sup> Similar conclusions are reached by *Mind the Gap* (n 12); Chengeta (n 13) 31–4; Heather M Roff, 'Killing in War: Responsibility, Liability, and Lethal Autonomous Robots' in Fritz Allhoff, Nicholas G Evans and Adam Henschke (eds), *Routledge Handbook of Ethics and War: Just War Theory in the Twenty-First Century* (Routledge, 2013) 352, 358; Amoroso (n 20) 19; Crootof (n 20) 1379–81.

<sup>77</sup> McFarland and McCormack (n 50) 366.

<sup>78</sup> *Ibid* 376–81.

<sup>79</sup> *Ibid* 372–4.

<sup>80</sup> *Ibid* 375–7.

<sup>81</sup> *Ibid* 380 (citations omitted).

The threshold requirement that relevant acts occur ‘in the context of an armed conflict’ could be amended to explicitly (or could be interpreted to implicitly) include acts of preparation prior to the commencement of the armed conflict provided that the completion of the crime occurred in the relevant context. *Mens rea* requirements for developers to aid and abet subsequent offenses could also be clarified to cover the scenario we have been considering.<sup>82</sup>

Such revisions, however, would not address the fact that accessorial liability is derivative in nature: it depends on the commission of a crime by a principal perpetrator, which is missing here.<sup>83</sup> Thus, accessorial liability, even with some revisions, may equally not provide an appropriate framework for a satisfactory solution.

Given challenges associated with contemplated revisions to command responsibility and accessorial liability, other scholars have contemplated a tailor-made AWS-specific crime. Most who have explored this line of inquiry focus on the idea of lowering the mental element to recklessness or negligence.<sup>84</sup> The principal objection to such suggestions is that it would create a scapegoat in circumstances where the operator/commander/programmer is not truly culpable.<sup>85</sup> Many see this as an irresolvable problem, leading to the conclusion that AWS must be prohibited or restricted by ensuring meaningful human control.<sup>86</sup>

Others have tried to identify an ‘actual culprit’ to whom to attach the lowered mental element. Geoffrey Corn suggests that rather than the operator or programmer, the individuals who should be designated as the relevant perpetrators are ‘the military and/or civilian officials responsible for procuring and fielding these weapons systems’,<sup>87</sup> while Neha Jain would pin responsibility on ‘the commander who reviews the ability of the AWS to be able to perform the tasks assigned to it within the limits of the law and gives the authorization to

---

<sup>82</sup> Ibid 384.

<sup>83</sup> Beard (n 13) 652; Ohlin (n 9) 28.

<sup>84</sup> See, eg, Ohlin (n 9) 28–9. In addition to lowering the mental element to recklessness, Jens Ohlin argues that commanders should be prosecuted on the basis of the doctrine of indirect perpetration: at 3–14. Under the doctrine, which has been read into art 25(3)(a) of the *Rome Statute*, individuals are held criminally responsible on the basis that they acted through a person, organisation or organisation-like entity that they controlled, such that the perpetrator’s orders, which resulted in the ultimate criminal act, were carried out by the organisation as a matter of course. Ohlin argues that the doctrine could readily be re-oriented ‘to shift the metaphorical language of machine to a literal case of machine liability’: at 9. Ohlin seemingly adds this complication, which equates an AWS with a subordinate soldier, to address the possibility that an AWS is properly viewed as a culpable agent: at 2–3. As noted above, this is currently a minority view.

<sup>85</sup> Chengeta (n 13) 15; Malik (n 9) 634; Liu (n 47) 326–7; Neha Jain, ‘Autonomous Weapons Systems: New Frameworks for Individual Responsibility’ in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 303, 320–2; Roff (n 76) 355; *Report of Special Rapporteur, Christof Heyns*, UN Doc A/HRC/23/47 (n 9) [80]; Amoroso (n 20) 20–1.

<sup>86</sup> Chengeta (n 13) 23–7, 50; Beard (n 13) 681.

<sup>87</sup> Geoffrey S Corn, ‘Autonomous Weapons Systems: Managing the Inevitability of “Taking the Man out of the Loop”’ in Nehal Bhuta et al (eds), *Autonomous Weapons Systems: Law, Ethics, Policy* (Cambridge University Press, 2016) 209, 234. Geoffrey Corn refers to this as procurement responsibility and talks of procurement commanders, drawing a parallel with command responsibility.

deploy it for a certain operation'.<sup>88</sup> Thus, for Jain, an authoriser would be individually criminally responsible if he or she 'should have been aware of a substantial and unjustifiable risk of harm resulting from AWS conduct', and this would be established 'if, given their circumstances and knowledge, their failure to advert to this risk constituted a gross deviation from the standard of care expected of a reasonable person in their situation'.<sup>89</sup> Corn's perpetrators would be accountable for 'objectively foreseeable failures of the weapon review and compliance validation process'.<sup>90</sup> Jain and Corn both argue that their proposed approach would encourage active due diligence<sup>91</sup> and, in this way, address at least some of the policy concerns associated with the accountability gap, regardless of how many prosecutions actually materialised.

Neither Corn nor Jain's proposals, however, appear to be a magic bullet. I agree that a cogent argument can be made that the person who approves a military's use of a particular AWS, or its deployment in a particular operation, is 'more responsible' for civilian deaths caused by an unpredictable AWS than the person who was ordered to switch the AWS 'on'. I also agree that there are fewer hurdles to the prosecution of such a person compared to the programmer who worked on the AWS in its development phase. However, it should be recognised that the 'proper perpetrators' identified by Corn and Jain are unlikely to be found individually criminally responsible in the accountability test scenario used in this article. Their responsibility hinges on a lack of proper care in the decision to acquire or deploy an AWS. An accountability gap will remain where all proper assessments about IHL compliance are made, but, due to the unpredictable nature of an AWS, civilians still die unlawfully. In other words, these suggested solutions bring us almost full circle to the position of Schmitt, Dunlap and McFarland, who argue that existing IHL rules are sufficient to ensure that there is no accountability gap, as outlined above. I say almost full circle because the approach suggested by Corn and Jain may help to address perceived deficiencies in the content or enforcement of IHL's rules regulating the development, acquisition, adoption and readying of weapons, even if they do not bridge the gap identified in this paper.

This leads me to the final example that I wish to canvass. Building on the work of Robert Sparrow, which drew an analogy between AWS and child soldiers,<sup>92</sup> Hin-Yan Liu states that

international criminal law absolves children of responsibility for their participation in, or perpetration of, international crimes. The potential void that this creates in terms of individual responsibility is avoided by the clear prohibition of child soldiers from armed conflict at an earlier stage. ... [T]he individual is not responsible for colluding or controlling the crimes committed by

---

<sup>88</sup> Jain (n 85) 314. For a more rudimentary variation of this argument, see Ohlin (n 9). Ohlin argues that the definition of such a crime would have to 'make clear that the crime is less culpable than the other core crimes of international law': at 28–9.

<sup>89</sup> Jain (n 85) 318.

<sup>90</sup> Corn (n 87) 235.

<sup>91</sup> Jain (n 85) 319; Corn (n 87) 235. Indeed, Corn suggests that this might be 'the operational Achilles heel that results in the hesitation to pursue [AWS]': at 241.

<sup>92</sup> Sparrow (n 9) 73–4.

the child soldiers but, rather, for introducing irresponsible entities into armed conflict.<sup>93</sup>

His proposed solution is to ‘criminalize the introduction of such irresponsible systems onto the battle space’.<sup>94</sup> In this way, Liu would plug the accountability gap through a strict liability offence for the procurement/authorisation (Liu does not specify) of AWS.

Liu’s analogy is not entirely apposite. As he partially acknowledges, the criminalisation of the recruitment and use of child soldiers is aimed at the protection of those children,<sup>95</sup> not those who they might in turn harm. In addition, the individual criminal responsibility of child soldiers is in fact an outstanding question under international criminal law.<sup>96</sup> More pertinently though, while Liu’s suggested solution would certainly solve the accountability problem, it would do more than this, by effectively establishing a prohibition of AWS by other means.

Liu argues that criminalisation would be preferable to a prohibition for three reasons. First, he says that placing the emphasis directly on individuals ‘skirts issues associated with signing and ratifying international conventions, of international enforcement and of establishing state responsibility’.<sup>97</sup> Secondly, he says it would reflect the fact that ‘the situations of impunity arising from AWS are a function of legal rather than technical, inadequacy’. Thirdly, he states that

[w]here a pre-emptive prohibition may unduly stifle the development of autonomous technologies, which may have legitimate civilian applications, the criminalization approach may be readily rescinded if [concerns around AWS] are subsequently resolved.<sup>98</sup>

I struggle to follow Liu’s logic. It is not obvious to me how individual criminal responsibility for the procurement/authorisation of AWS could realistically be established absent an international agreement of some kind. I agree that the accountability gap is not a technical issue, but it is not clear to me why this means the solution should be found in international criminal, rather than

---

<sup>93</sup> Liu (n 47) 343–4 (citations omitted).

<sup>94</sup> Ibid 344.

<sup>95</sup> Ibid 343.

<sup>96</sup> It is noted that children under the age of 18 are excluded from the jurisdiction of the ICC under art 26 of the *Rome Statute*, however, Dominic Ongwen, a former child soldier recruited and used by the Lord’s Resistance Army (‘LRA’), who allegedly rose to the rank of commander leading one of four LRA brigades, is currently on trial for war crimes and crimes against humanity before the ICC. Otherwise, there is in fact no international agreement that bars child soldiers from being prosecuted for crimes they commit. Liu (n 47) 343 selectively quotes the *Paris Principles: Principles and Guidelines on Children Associated with Armed Forces or Armed Groups* (February 2007) 9 [3.6] <<https://www.unicef.org/emerg/files/ParisPrinciples310107English.pdf>>, archived at <<https://perma.cc/S74H-2R43>>. Principle 3.6 provides that

[c]hildren who are accused of crimes under international law allegedly committed while they were associated with armed forces or armed groups should be considered primarily as victims of offences against international law; not only as perpetrators. They must be treated in accordance with international law in a framework of restorative justice and social rehabilitation, consistent with international law which offers children special protection through numerous agreements and principles.

<sup>97</sup> Liu (n 47) 344.

<sup>98</sup> Ibid.

humanitarian, law. I am also not convinced that criminalisation could be ‘rescinded’ that much more easily than a prohibition, or at least a moratorium, or that there is likely to be a material difference between the impact on civilian applications that criminalisation would have on the one hand, or that caused by a prohibition/moratorium of AWS on the other. Indeed, as will be elaborated below, I have strong misgivings about any attempt to use international criminal law as a means to address perceived deficiencies in IHL.

## VI THE PROBLEM WITH THE PROMINENCE OF ACCOUNTABILITY

McCormack and McFarland comment that, ‘intriguingly’, proponents of views on AWS from one end of the extreme (‘the sky is falling’) to the other (‘there is nothing new under the sun’) ‘share substantial common ground in acknowledging the fundamental importance of accountability to the efficacy of the law of armed conflict’.<sup>99</sup> They impliedly celebrate this fact.

I understand this view. As an international criminal law practitioner and scholar, I am familiar with the fact that all too often the accountability piece is something that we have to fight to get on the table: critiques of the international criminal justice system abound, and there is otherwise a yawning gap between the lip service paid to states’ commitment to accountability and their actual practice. I myself am an ardent believer in the importance of holding those responsible for serious international crimes to account and so I would not want the following part of this article to be interpreted as questioning the importance of accountability in general. But I do question whether accountability is the right lens through which to examine whether AWS should be prohibited or restricted.

In the first place, I am of the view that it is putting the cart before the horse to make accountability the determinative factor in the debate over AWS. Theories of criminalisation under international criminal law are notoriously underdeveloped. As M Cherif Bassiouni noted, ‘international crimes have developed to date, without ... an agreed-upon definition of what constitutes an international crime, what are the criteria for international criminalization, and how international crimes are distinguished’.<sup>100</sup> If we search for a common denominator, however, it is clear that in almost all cases, individual criminal responsibility has only been attached to conduct that is prohibited under another branch of public international law (IHL, international human rights law, the *jus ad bellum*). Indeed, it is fair to say that international criminal law is generally a reflection of prohibitions that are well entrenched in either conventional or customary international law and broadly supported by the international community.<sup>101</sup>

This is reflected in the approach of states to the *Rome Statute*, which provides an insight into states’ views on criminalisation under international law, albeit that not all conduct that states have declared is deserving of criminal sanction was

---

<sup>99</sup> Tim McFarland and Tim McCormack (n 50) 362.

<sup>100</sup> M Cherif Bassiouni, *Introduction to International Criminal Law* (Transnational Publishers, 2003) 111.

<sup>101</sup> See also *ibid* 29; Yoram Dinstein, ‘International Criminal Law’ (1985) 20(2–3) *Israel Law Review* 206, 221; Ilias Bantekas and Susan Nash, *International Criminal Law* (Cavendish Publishing, 2<sup>nd</sup> ed, 2003) 5; Antonio Cassese, *International Criminal Law* (Oxford University Press, 2<sup>nd</sup> ed, 2008) 11–12.

considered suitable for inclusion under the jurisdiction of an international criminal court (eg terrorism, human trafficking). Many states cited consistency, or inconsistency, with customary international law as the primary reason why certain proposed definitions were, or were not, acceptable to them in the negotiation of the *Statute*: for some this required evidence of a widely accepted prohibition of the conduct in question, others insisted on evidence of a pre-existing *crime* under customary international law.<sup>102</sup> An analysis of the definitions of crimes ultimately adopted under the *Statute* reveals that those definitions are generally consistent with customary international law.<sup>103</sup> As Theodor Meron observes: '[d]elegations filled in some gaps in Rome, but this took place in the interstices of rules, not on the core of the rules'.<sup>104</sup>

Specifically, all war crimes relating to the use of a specific weapon under the *Rome Statute* rest on a pre-existing prohibition under IHL. This is highlighted by art 8(2)(b)(xx), which criminalises (in IAC only):

Employing weapons, projectiles and material and methods of warfare which are of a nature to cause superfluous injury or unnecessary suffering or which are inherently indiscriminate in violation of the international law of armed conflict, provided that such weapons, projectiles and material and methods of warfare are the subject of a comprehensive prohibition and are included in an annex to this *Statute*, by an amendment in accordance with the relevant provisions set forth in articles 121 and 123 ...

Some 20 years after the adoption of the *Rome Statute*, the existence of an underlying prohibition in IHL remains a key consideration for states. The Terms of Reference of the Working Group on Amendments (the body of the ICC's Assembly of States Parties that considers amendments to the *Rome Statute* and the Court's *Rules of Procedure and Evidence*)<sup>105</sup> provide that

[i]n the case of a proposal for a new crime, the [Working Group on Amendments] particularly considers whether the crime can be characterized as one of the most

<sup>102</sup> United Nations Diplomatic Conference of Plenipotentiaries on the Establishment of an International Criminal Court, *Summary Records of the Plenary Meetings and of the Meetings of the Committee of the Whole*, UN Doc A/CONF.183/13 (Vol II) (15 June – 17 July 1998) 150 (UK), 151 (Slovenia), 154–5 (Canada), 155 (Israel), 158 (Syria), 160 (New Zealand), 160 (Greece), 162 (Belgium), 164 (France), 187 (Jordan), 270 (China), 277 (Switzerland), 277 (Brazil), 278 (Korea), 285 (Bosnia and Herzegovina), 287 (Indonesia), 289 (Russia).

<sup>103</sup> Herman von Hebel and Darryl Robinson, 'Crimes within the Jurisdiction of the Court' in Roy S Lee (ed), *The International Criminal Court: The Making of the Rome Statute* (Kluwer International, 1999) 79, 104, 122–3; William A Schabas, *An Introduction to the International Criminal Court* (Cambridge University Press, 2001) 23. For a comparison of the *Rome Statute*'s definitions of crimes with customary international law, see Antonio Cassese, 'Genocide' in Antonio Cassese, Paola Gaeta and John RWD Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary* (Oxford University Press, 2002) vol 1, 335; Antonio Cassese, 'Crimes against Humanity' in Antonio Cassese, Paola Gaeta and John RWD Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary* (Oxford University Press, 2002) vol 1, 353; Michael Bothe, 'War Crimes' in Antonio Cassese, Paola Gaeta and John RWD Jones (eds), *The Rome Statute of the International Criminal Court: A Commentary* (Oxford University Press, 2002) vol 1, 379.

<sup>104</sup> Theodor Meron, 'Defining Aggression for the International Criminal Court' (2001) 25(1) *Suffolk Transnational Law Review* 1, 8.

<sup>105</sup> International Criminal Court, *Rules of Procedure and Evidence*, Doc No ICC-ASP/1/3 (adopted 9 September 2002).

serious crimes of concern to the international community as a whole and whether the crime is based on an existing prohibition under international law.<sup>106</sup>

This is no empty policy statement. I was involved in negotiations in relation to the addition, in 2010, of war crimes relating to the use in NIAC<sup>107</sup> of poison or poisoned weapons (art 8(2)(e)(xiii)), asphyxiating, poisonous or other gases (art 8(2)(e)(xiv)), and bullets that expand or flatten easily in the human body (art 8(2)(e)(xv)),<sup>108</sup> and the addition, in 2017, of war crimes relating to the use in IAC and NIAC of microbial, biological or toxin weapons (arts 8(2)(b)(xxvii) and 8(2)(e)(xvi)),<sup>109</sup> weapons that injure by fragments undetectable by X-rays (arts 8(2)(b)(xxviii) and 8(2)(e)(xvii)),<sup>110</sup> and employing laser blinding weapons (arts 8(2)(b)(xxix) and 8(2)(e)(xviii)),<sup>111</sup> as well as the as yet unsuccessful proposal for a war crime relating to the use of anti-personnel landmines. The identification of a clear pre-existing prohibition under international law was a key consideration for the Australian delegation, which I represented, as well as a majority of other states participating in the Working Group on Amendments.<sup>112</sup>

Clearly, as yet, there exists no prohibition against AWS per se.<sup>113</sup> Prohibitionists who have blanket concerns about the policy and ethical implications of armed conflicts being fought by machines (eg the dehumanisation of killing as a result of arms-length participation by humans, the psychological impact on affected populations, and the lowering of the political cost of engaging in armed conflict as a result of decreased combatant death on the side using offensive force) are thus not well served by arguments about an accountability gap. Conduct should be prohibited before it is criminalised, not prohibited because it is not criminalised.

Of course, it can be argued that in fact there is a relevant pre-existing underlying prohibition; that what is of concern is the enforcement of Geneva rather than Hague rules of IHL.<sup>114</sup> Geneva rules are technology neutral: they prohibit a consequence (eg unlawful killing of civilians); the means used are immaterial. The argument here is that the unique nature of AWS

---

<sup>106</sup> Assembly of States Parties to the Rome Statute of the International Criminal Court, *Strengthening the International Criminal Court and the Assembly of States Parties*, Doc No ICC-ASP/11/20 (adopted 21 November 2012) annex II ('*Terms of Reference of the Working Group on Amendments*') para 9.

<sup>107</sup> Crimes relating to the use of these weapons in IAC were included in the original version of the *Rome Statute* adopted in 1998: *Rome Statute* (n 21) arts 8(2)(b)(xvii), 8(2)(b)(xviii), 8(2)(b)(xix).

<sup>108</sup> Review Conference of the Rome Statute of the International Criminal Court, *Amendments to Article 8 of the Rome Statute*, Doc No RC/11 (adopted 10 June 2010) annex I ('*Amendment to Article 8*').

<sup>109</sup> Assembly of States Parties to the Rome Statute of the International Criminal Court, *Resolution on Amendments to Article 8 of the Rome Statute of the International Criminal Court*, Doc No ICC-ASP/16/20/Vol. I (adopted 14 December 2017) annex I.

<sup>110</sup> *Ibid* annex II.

<sup>111</sup> *Ibid* annex III.

<sup>112</sup> See also Assembly of States Parties to the Rome Statute of the International Criminal Court, *Report of the Working Group on Amendments*, 16<sup>th</sup> sess, Doc No ICC-ASP/16/22 (15 November 2017) 3–4 [13]–[14].

<sup>113</sup> Assuming, of course, a separate weapon prohibition or restriction (such as the restriction on the use of incendiary weapons) is not crossed by a specific AWS.

<sup>114</sup> The rules of Geneva refer to the body of IHL rules that protect victims of armed conflict, while the rules of the Hague relate to the rules regulating the means and methods by which hostilities are conducted.

(unpredictability) requires criminalisation, not because AWS are unlawful per se, but because the deployment of AWS (at least in certain circumstances) risks increased non-compliance with the rules of IHL aimed at protecting non-combatants because of the enforcement problem.

The notion that we should define what conduct is permissible based on probable outcomes assessed on the current state of international criminal law is, in my view, conceptually confused. Stripped to its essence, this amounts to an argument that all conduct that produces an unwanted outcome (civilians killed unlawfully) with respect to which we cannot hold an individual criminally responsible should be prohibited. The logical extension of this argument is a conclusion that the deployment of human combatants should be banned. This raises significant questions for me about the cogency of the argument that we should prohibit or restrict AWS because of the accountability gap.

Viewed from this angle, it strikes me that the real problem is a lack of confidence in the adequacy of IHL's rules in relation to the development, acquisition, adoption and deployment of weapons that were unforeseen at the time relevant rules, particularly art 36 of *API*, were drafted. Clearly the proposals advanced by the likes of Corn and Jain are more squarely aimed at addressing this issue. However, while it is inevitable that international criminal law will flesh out some details in the application of relevant rules of IHL,<sup>115</sup> I do not consider it appropriate for international criminal law to be used as a backdoor to clarify or strengthen IHL's rules on weapon development and deployment in any sort of substantive way. International criminal law is already beset by a laundry list of conflicting expectations and objectives (retribution, general and specific deterrence, incapacitation, rehabilitation, reconciliation, justice to victims, truth-telling, promotion of the rules-based international order, establishing and maintaining an inclusive and sustainable peace, etc). Using international criminal law as a band-aid to address serious misgivings in relation to AWS is, I think, expecting too much of it. By further overburdening the expectations heaped on international criminal law we would only do further disservice to it.

Regardless of whether these views are shared, I do not believe that an assessment of the accountability problem actually provides any real guidance in relation to how the international community should approach AWS. As I hope the above review of the literature has demonstrated, in many ways, the conclusions reached by experts in response to the accountability question are malleable. By this I partly mean that conclusions are in many cases influenced by the differing technical capabilities and IHL compliance assumptions employed in the factual scenarios examined, and/or an expert's broader views on the value of criminal accountability for violations of IHL in the broad, and/or views on the degree of culpability that is properly required to trigger individual criminal responsibility, or who is the 'real culprit'.

In referring to malleability I also mean that the conclusions reached on whether there is an accountability gap should not, in my view, be determinative in the debate over whether AWS should be prohibited or restricted. International criminal lawyers can advise whether proximate humans could be held to account

---

<sup>115</sup> On this phenomenon, see Carrie McDougall, 'The Other Enemy: Transnational Terrorists, Armed Attacks, and Armed Conflicts' in Leila Nadya Sadat (ed), *Seeking Accountability for the Unlawful Use of Force* (Cambridge University Press, 2018) 219, 247–50.

under the current law. They can also advise on the amendments to the criminal law that would be necessary to secure accountability. Given most international criminal lawyers are not mere black-letter-law legal advisers but also legal-policy experts, many will also be able to provide advice on the appropriateness of extending individual criminal responsibility in such a manner. Ultimately, however, this is a policy, not a legal, issue. If states decide that one or more proximate humans should *always* be held criminally responsible when IHL is violated through the use of an AWS, one way or another, international criminal law could be amended to provide this outcome. Policy-makers need to decide whether the risk that civilians will be unlawfully killed by an AWS, even after all due diligence is taken to ensure that the weapon complies with IHL, is a risk that can be tolerated. If it is not, policy-makers need to decide whom they are prepared to make responsible for the materialisation of that risk. As such, I do not consider it useful to view the accountability problem as providing any meaningful guidance as to whether we need to prohibit or restrict AWS.

I note that I do not believe that this is a view that can be dismissed by reference to *realpolitik*. It is an inescapable conclusion that power politics play a significant role in international criminal justice. History demonstrates that many states see justice as something that should be delivered against ‘others’: other conduct, other wars, other parties, individuals at the bottom, rather than the top, of the chain of command. Thus, while I personally remain optimistic that states will continue to work together to criminalise ‘unimaginable atrocities that deeply shock the conscience of humanity’<sup>116</sup> and to end the culture of impunity, I acknowledge that reaching agreement on criminal provisions in relation to the use of AWS may be exceedingly difficult, particularly in the current geopolitical climate. I also acknowledge that deciding with whom the buck stops is a difficult decision. This does not alter the fact, however, that this is a policy/political issue, rather than a legal one. Moreover, I do not believe it would be any easier to reach agreement on a prohibition or moratorium of AWS or a restriction reflecting a requirement for meaningful human control.

## VII CONCLUSION

This article does not necessarily advocate for the development and use of AWS. I have not addressed any of the other legal, policy, practical and ethical issues raised by the use of these weapon systems — some of which weigh in favour of the development and use of AWS and others against. Concerns that have been identified are wide-ranging, but include whether IHL precludes non-human ‘decision-making’ with respect to the application of lethal force; whether AWS would be consistent with the Martens Clause;<sup>117</sup> the practical and ethical implications of machines killing humans, and the potential lowering of the political cost of engaging in armed conflict; whether handing ‘decisions’ on life

---

<sup>116</sup> *Rome Statute* (n 21) Preamble para 2.

<sup>117</sup> There is ongoing debate as to whether the Martens Clause’s requirement that in cases not covered by existing treaties, civilians and combatants remain under the protection and authority of the principles of humanity and the dictates of public conscience reflects an enforceable obligation. For a good overview, see Emily Crawford, ‘The Modern Relevance of the Martens Clause’ (2006) 6 *ISIL Year Book of International Humanitarian and Refugee Law* 1.

and death to machines is a violation of the right to life; and concerns about the ease with which the technology, with its enormous destructive potential, could fall into the hands of terrorists or warlords. Advocates for AWS on the other hand, point to the possibility of increased compliance with IHL as a result of the removal of human frailties and superior data collection and analysis;<sup>118</sup> the military advantage offered by enhanced speed, greater precision, longer range, greater endurance, immunity to certain other weapons and lower financial costs; and the possibility of otherwise making conflict less harmful as a result of an ability to employ more targeted and lower intensity force.

What this article has sought to do is to question the relevance, or at least the decisiveness, of accountability to this debate. I am not the first to question whether the accountability prism is the correct one. Several authors reject the relevance of accountability on the simple premise that there is no accountability gap.<sup>119</sup> Others are dismissive of the value of ‘post-hoc judicial accountability ... for promoting and enforcing compliance with the laws of war’.<sup>120</sup> From the perspective of a strong advocate of the value of international criminal law, I question whether accountability will in fact provide us with any genuine guidance in relation to the difficult questions surrounding the development and use of AWS, given that, one way or another, any accountability gap can be plugged if this is the desired policy outcome.

As a former government legal adviser and diplomat, I recognise that it is not uncommon for the diplomatic representatives of states to engage in technical legal debates to resolve a dispute, rather than directly confront politically sensitive policy issues relating to war and peace. But I think it is a mistake to suggest that the accountability discussion is anything more than a vehicle through which broader policy debates about AWS are being played out.

---

<sup>118</sup> For a good overview, see Margulies (n 13) 415–27.

<sup>119</sup> See, eg, Dunlap Jr (n 58).

<sup>120</sup> Anderson and Waxman (n 51) 17.