

# EVALUATING AUTONOMOUS WEAPONS SYSTEMS: A DICHOTOMIC LENS OF MILITARY VALUE AND ACCOUNTABILITY

Emily L. Drake\*

## Table of Contents

Introduction.....	299
I. The Legal Framework of AWS.....	303
A. Defining Autonomy on a Spectrum.....	303
1. Different Spectrums of Autonomy.....	305
B. The International Legal System Governing Weapons and Use of Force.....	308
1. Basics of IHL vs. IHRL.....	310
i. Principle of Proportionality.....	313
C. Accountability and Responsibility.....	314
1. Actus Reus.....	317
2. Mens Rea.....	318
II. The Accountability Gap: The Problem of AWSs in the Current Legal System.....	320
A. The Problem of the Accountability Gap in AWSs.....	320
1. The Overarching Problems of Establishing Responsibility and Accountability.....	322
2. A Bias Towards Agreeing with the AWS Exacerbates the Accountability Gap.....	325
i. Presumption that the Machine is Correct.....	325
ii. Even if the Machine Is Incorrect, the Personal Consequences of Action Outweigh Those of Inaction.....	326
B. The Accountability Gap Leads to Problems with Criminal Accountability.....	328

---

\* J.D. Candidate 2022, Columbia Law School; B.S. and B.A., 2018, University of North Carolina at Chapel Hill. The author would like to thank Professor Gabor Rona for his unwavering guidance and support throughout the writing and editing process. Special thanks to Adi Radhakrishnan, Sara Kaufhardt, and Yahel Kaplan for their encouragement and feedback on earlier drafts and research inquiries.

1. Problems with Attributing Mens Rea for Criminal Accountability .....	328
2. Lowering the Mens Rea Threshold Would Not Solve the Problem .....	329
C. The Problems Posed by SAWSs .....	330
1. The U.S. Patriot System as an Example of SAWSs' Flaws .....	331
i. Operation Desert Storm .....	331
ii. Operation Iraqi Freedom.....	333
2. AC-130 2015 Médecins Sans Frontières Attack.....	335
III. Implications for Legality of Use: Defining a Lens of Analysis.....	337
A. Problems Within the Criminal Justice System and Humans as Ineffective Stopgap Mechanisms.....	337
B. Alternative solutions to holding parties responsible are flawed .....	339
1. War Torts .....	339
2. Administrative Accountability.....	340
3. Military Justice.....	341
C. Exploring the Proposed Dichotomy: Possible Accountability and Potential Military Value .....	341
D. Applying the Dichotomy to AWS: Reaching the Point of No Return .....	342
Conclusion .....	343

## INTRODUCTION

On October 30, 2015, an AC-130 gunship, also known as the “Angel of Death,”<sup>1</sup> was dispatched ahead of schedule to fire on a Taliban insurgent compound.<sup>2</sup> Due to its early departure from Bagram Airfield, the gunship was not equipped with a no-strike list which identified protected buildings.<sup>3</sup> The video feed of the target failed, leaving the human operators<sup>4</sup> unable to see what they were supposed to be aiming at.<sup>5</sup> On the way to the compound, enemy missile fire forced the gunship off course;<sup>6</sup> once it was able to return to its course, the coordinates the gunship was given led it to an empty field.<sup>7</sup> After being ordered to continue the attack, the gunship’s operators began visually searching for the target.<sup>8</sup> Finding a similar-looking building, the operators relayed the new coordinates to their senior officials who, assuming the gunship was correct, confirmed the coordinates without checking their own no-strike list as was required by protocol.<sup>9</sup> Ground forces that had been engaged in heavy fighting for nearly five consecutive days and nights also confirmed the target’s location without actually seeing the target, again

---

1. *The AC-130 Gunship*, AIRMAN MAG. (Dec. 22, 2017), <https://www.airmanmagazine.af.mil/Features/Display/Article/2594229/airframe-the-ac-130-gunship/> [<https://perma.cc/Z7SM-H9N5>] [hereinafter *AC-130*].

2. Matthew Rosenberg, *Pentagon Details Chain of Errors in Strike on Afghan Hospital*, N.Y. TIMES (Apr. 29, 2016), <https://www.nytimes.com/2016/04/30/world/asia/afghanistan-doctors-without-borders-hospital-strike.html> (on file with the *Columbia Human Rights Law Review*) [hereinafter Rosenberg].

3. Phil Stewart & Yeganeh Torbati, *Human Error Led to Deadly U.S. Strike on Afghan Hospital: Military*, REUTERS (Nov. 25, 2015), <https://www.reuters.com/article/uk-afghanistan-attack-msf-investigation/human-error-led-to-deadly-u-s-strike-on-afghan-hospital-military-idUKKBN0TE1X020151125> [<https://perma.cc/238U-WJ3R>] [hereinafter Stewart & Torbati]. The vessel’s satellite radio on board, the aircraft’s only datalink, also failed. Phil Stewart, *U.S. Military Punishes 16 Over 2015 Afghan Hospital Bombing*, REUTERS (Apr. 28, 2016), <https://www.reuters.com/article/us-usa-afghanistan-msf-investigation-idUSKCN0XP2S7> [<https://perma.cc/653J-LSPB>].

4. “Operators” will be used throughout this Note, although many of these Autonomous Weapons Systems do not have operators in the conventional meaning. Once they reach a certain level of autonomy, the weapons are no longer under the direct control of the human being. For further discussion, see Jack M. Beard, *Autonomous Weapons and Human Responsibilities*, 45 GEO. J. INT’L L. 617, 652 (2014) (discussing the nature of “operators” of autonomous weapons) [hereinafter Beard].

5. *Id.*

6. Rosenberg, *supra* note 2.

7. *Id.*

8. *Id.*

9. *Id.*

assuming the gunship was correct.<sup>10</sup> The AC-130 then began a twenty-nine-minute-long strike at 2:08 AM.<sup>11</sup>

By 2:20 AM, the Bagram air base received notification that the target at which the gunship was firing was not, in fact, the insurgent compound—it was a Médecins Sans Frontières (“MSF”) hospital. Although the AC-130 did eventually identify the correct building via grid coordinates and there was no hostile enemy activity at the MSF building, the U.S. attack continued on the hospital.<sup>12</sup> It took another seventeen minutes before the mistake was realized, by which time the AC-130 had already departed.<sup>13</sup> The strike killed forty-two people, including staff members and patients, all of whom were civilians or *hors de combat*, and injured thirty-seven others.<sup>14</sup>

An internal investigation found that the military actions on MSF did not constitute war crimes and required no criminal prosecution, despite allegations that the strike had been intentional.<sup>15</sup> Instead, sixteen military personnel faced administrative punishment after the “military investigation determined the attack to be unintentional.”<sup>16</sup> The investigation found a chain of human errors, equipment failures, and procedural mistakes attributable to no one person in particular.<sup>17</sup>

The AC-130 is a semi-autonomous weapons system (“SAW”). Autonomous weapons systems (“AWSs”) have become more prevalent in recent years, both in popular culture<sup>18</sup> and on the battlefield.<sup>19</sup> SAWs are

---

10. Andrew Tilghman, *C-130 Crew Blamed for Kunduz Hospital Attack*, *Top General Says*, MIL. TIMES (Nov. 25, 2015), <https://www.militarytimes.com/2015/11/25/c-130-crew-blamed-for-kunduz-hospital-attack-top-general-says/> [https://perma.cc/Z5N3-T245]; Rosenberg, *supra* note 2.

11. Tilghman, *supra* note 10.

12. Stewart & Torbati, *supra* note 3.

13. Tilghman, *supra* note 10.

14. *On 3 October 2015, US Airstrikes Destroyed Our Trauma Hospital in Kunduz, Afghanistan, Killing 42 People*, MÉDECINS SANS FRONTIÈRES, <https://www.msf.org/kunduz-hospital-attack-depth> [https://perma.cc/Y8YX-4BJ6].

15. Gregor Aisch, Josh Keller & Sergio Peçanha, *How a Cascade of Errors Led to the U.S. Airstrike on an Afghan Hospital*, N.Y. TIMES (Apr. 29, 2016), <https://www.nytimes.com/interactive/2015/11/25/world/asia/errors-us-airstrike-afghan-kunduz-msf-hospital.html> (on file with the *Columbia Human Rights Law Review*) [hereinafter Aisch, Keller & Peçanha].

16. *Id.*

17. *Id.*

18. A prime example of an artificially intelligent killer robot is SkyNet, the Terminator’s main antagonist. *THE TERMINATOR* (Hemdale Film Corporation 1984); *see also* *THE MATRIX* (Warner Brothers 1999); *BLADERUNNER* (Warner Brothers 1982) (involving killer artificial intelligence or robots).

19. Nearly all major powers’ military forces now rely on autonomous guidance systems to improve the accuracy of, for example, long range missiles, particularly for anti-

regularly used by the U.S. military.<sup>20</sup> With the use of these weapons, though, comes complicated ethical and legal questions—fully-autonomous weapons systems (“FAWSs”), for example, are “able to make decisions that define their actions and adapt to their environment based on pre-programmed rules or boundaries.”<sup>21</sup> When a weapons system can make decisions and choices on its own, questions of accountability and responsibility become increasingly complex. The question, in brief, becomes: if one cannot punish a robot, how does one punish a human for a robot’s choice?<sup>22</sup>

As the U.S. Department of Defense Law of War Manual puts it: The law of war rules on conducting attacks . . . impose obligations on persons. These rules do not impose obligations on the weapons themselves; of course, an inanimate object could not assume an “obligation” in any event . . . . The law of war does not require weapons to make legal determinations, even if the weapon . . . may be characterized as capable of making factual determinations, such as whether to fire the weapon or to select and engage a target . . . . Rather, it is persons who must comply with the law of war.<sup>23</sup>

With the use of AWSs increasing, there has been a corresponding increase in concern about their use, partially due to the inability to hold a robot responsible for its actions or to justly hold a human responsible for the

---

ship and anti-aircraft attacks. *See, e.g.*, JAMES J. BUSUTTIL, NAVAL WEAPONS SYSTEMS AND THE CONTEMPORARY LAW 191 (1998) (discussing the legality of AWSs in anti-ship weapons).

20. OFF. OF THE SEC’Y OF DEF., UNMANNED SYSTEMS ROADMAP: 2007-2032 54 (2007), [https://www.globalsecurity.org/intell/library/reports/2007/dod-unmanned-systems-roadmap\\_2007-2032.pdf](https://www.globalsecurity.org/intell/library/reports/2007/dod-unmanned-systems-roadmap_2007-2032.pdf) [<https://perma.cc/M8XQ-ZBNQ>] [hereinafter Unmanned Systems Roadmap].

21. INT’L COMM. RED CROSS, AUTONOMOUS WEAPON SYSTEMS: TECHNICAL, MILITARY, LEGAL, AND HUMANITARIAN ASPECTS 62 (2014), <https://www.icrc.org/en/download/file/1707/4221-002-autonomous-weapons-systems-full-report.pdf> [<https://perma.cc/W823-QCMT>] [hereinafter AWS REPORT].

22. *See, e.g.*, Peter M. Asaro, *How Just Could a Robot War Be?*, in CURRENT ISSUES IN COMPUTING AND PHILOSOPHY 2 (Adam Briggie et al. eds., 2008), <https://peterasaro.org/writing/Asaro%20Just%20Robot%20War.pdf> [<https://perma.cc/P5PJ-3G3X>] (“[As] robotic technologies advance, it is possible that they will acquire moral capacities that imitate or replicate human moral capacities.”); WENDELL WALLACH & COLIN ALLEN, MORAL MACHINES: TEACHING ROBOTS RIGHT FROM WRONG 196 (2009) [hereinafter WALLACH & ALLEN] (introducing emotions into robots is “a virtual Pandora’s box filled with both benefits and ethical challenges”).

23. OFF. OF GEN. COUNS., DEP’T OF DEF., DEPARTMENT OF DEFENSE LAW OF WAR MANUAL § 6.5.9.3 (2016), <https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190> [<https://perma.cc/E7VR-EDW7>] [hereinafter LAW OF WAR MANUAL].

robot's actions.<sup>24</sup> There have been calls to create a legally binding treaty banning AWSs<sup>25</sup> of all kinds, as well as arguments that such a ban would not only be ineffective, but also would result in arbitrary deaths.<sup>26</sup> The technology to create and use AWSs exists—it is not a question of whether we can create these weapons systems, but a question of whether we should use them.

In Part I, this Note will explain that autonomy is a spectrum, not a categorical condition. AWSs exist at various points along that spectrum. These machines can—and do—make mistakes resulting in accidental or improper loss of life. It will also outline the principle of proportionality<sup>27</sup> and

---

24. Currently, there is an “enormous gap between any technologies that meet the conditions necessary for ‘moral personhood’ and any existing, planned, or even contemplated autonomous weapon system.” Beard, *supra* note 4, at 662; see Robert Sparrow, *Killer Robots*, 24 J. APPLIED PHIL. 62, 73 (2007) [hereinafter Sparrow].

25. Reasons for an AWS ban include the risk of high-tempo fratricide incidents occurring at speeds too great for human intervention, concerns over the ability of AWS to satisfy the international humanitarian law (“IHL”) principles of distinction and proportionality, the potential violation of fundamental human values in delegating the decision to kill a human to an algorithm in a machine, and the accountability gap that would be created by the human abdication of political decision-making. See PAUL SCHARRE, *ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR* 252 (2018) [hereinafter SCHARRE 2018] (noting that Steve Goose, Director of the Human Rights Watch’s Arms Division and leading figure in the Campaign to Stop Killer Robots, calls for all AWSs to be banned); Frank Sauer, *Stopping ‘Killer Robots’: Why Now Is the Time to Ban Autonomous Weapons Systems*, ARMS CONTROL ASS’N (Oct. 2016), <https://www.armscontrol.org/act/2016-09/features/stopping-%E2%80%98killer-robots%E2%80%99-why-now-time-ban-autonomous-weapons-systems> [<https://perma.cc/Z2N9-GFZX>].

26. Some concerns about an AWS ban include that some factions and countries would build AWS regardless, creating an imbalance in the battlefield; that it is not more unethical for an AWS to make a decision to engage a target with deadly force than an armed combatant; and that AWSs have the capacity to be more accurate and precise than human beings, meaning there could be fewer mistaken casualties should AWSs be used. There are also logistical problems with such a ban, including the difficulty in defining what exactly would be an AWS subject to the ban and the ease with which countries could “cheat” on such a ban, given the near impossibility of inspecting a country’s weapons’ software. Evan Ackerman, *We Should Not Ban ‘Killer Robots,’ and Here’s Why*, IEEE SPECTRUM (Jul. 29, 2015), <https://spectrum.ieee.org/autamaton/robotics/artificial-intelligence/we-should-not-ban-killer-robots> [<https://perma.cc/7H3N-4EC3>]; Noah Gibbs, *Why Banning Autonomous Weapons Is More Dangerous Than Developing Them Responsibly*, POLEMICS MAG. (Apr. 25, 2019), <https://www.polemics-magazine.com/opinion/banning-autonomous-weapons-more-dangerous-developing-responsibly> [<https://perma.cc/K8A8-JU9Y>].

27. The principle of proportionality prohibits launching an attack which “may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated . . . .” *Rule 14: Proportionality in Attack*, INT’L COMM.

its dichotomous structure within international humanitarian law (“IHL”). Finally, Part I will establish the importance of accountability and responsibility in the legal system.<sup>28</sup> In Part II, this Note will demonstrate how, due to the increasingly autonomous nature of these weapons, it becomes correspondingly more difficult to assign individual responsibility when these weapons malfunction. Finally, Part III will propose that in an analogical balance to proportionality, the legality assessment of AWSs should be to compare the increasingly autonomous nature of the weapons to the decreasing possibility of assigning responsibility for errors.<sup>29</sup> This Note will conclude that if a finite point cannot be determined where the decreasing possibility of assigning responsibility outweighs the likelihood of the machine functioning as intended,<sup>30</sup> then there must be a ban implemented against exploration into AWSs until that point of limitation can be adequately and clearly defined.

## I. The Legal Framework of AWS

Part I will define relevant terms, explain applicable international law concepts, and introduce AWSs. Part I.A will outline the different definitions used to describe autonomy. Part I.B will then provide an overview on relevant international law and IHL concepts. Part I.C will distinguish between accountability and responsibility.

### A. Defining Autonomy on a Spectrum

The International Committee of the Red Cross (“ICRC”)<sup>31</sup> has defined AWSs as weapons systems that “can independently select and attack targets,

---

RED CROSS IHL DATABASE, [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_rul\\_rule14](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule14) [<https://perma.cc/K9BR-72A6>] [hereinafter Rule 14].

28. International Humanitarian Law, also known as the law of war or the law of armed conflict, is a “set of rules which seek, for humanitarian reasons, to *limit the effects of armed conflict*. It protects persons who are not or are no longer participating in the hostilities and restricts the means and methods of warfare.” INT’L COMM. RED CROSS, WHAT IS HUMANITARIAN LAW 1 (July 2004) (emphasis in original), [https://www.icrc.org/en/doc/assets/files/other/what\\_is\\_ihl.pdf](https://www.icrc.org/en/doc/assets/files/other/what_is_ihl.pdf) [<https://perma.cc/72UP-C777>] [hereinafter *What is IHL*].

29. See *infra* Part III.C–D. This Note will argue that, just as proportionality requires a balancing between the collateral loss of civilian life and the anticipated military advantage, AWSs should be evaluated by balancing the difficulty of establishing proper accountability and responsibility for the weapons system with the military advantage anticipated of the weapons system.

30. In other words, in accordance with a legal assessment of a permissible use of force.

31. The ICRC was tasked to be the “guardian of international humanitarian law” and is the organization appointed to interpret IHL. *The International Committee of the Red*

with or without human oversight.”<sup>32</sup> In particular, AWSs are fitted with autonomous functions of “acquiring, tracking, selecting, and attacking targets.”<sup>33</sup> For the purposes of this Note, AWSs will be used to encompass both SAWs and FAWs. The point of demarcation between SAWs and FAWs is not fully established and, partially due to their differing definitions, there is debate as to whether FAWs will inevitably exist in the future,<sup>34</sup> whether FAWs already do exist,<sup>35</sup> or if FAWs are not likely to exist even in the long term.<sup>36</sup>

By one definition, FAWs are “highly sophisticated weapon systems with ‘artificial intelligence’ that are programmed to independently

---

*Cross as Guardian of International Humanitarian Law*, INT’L COMM. RED CROSS (Dec. 31, 1998), <https://www.icrc.org/en/doc/resources/documents/misc/about-the-icrc311298.htm> [<https://perma.cc/XUX7-DEMG>]. The ICJ has stated that the ICRC is given special competency and authority to interpret IHL. *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 I.C.J. 226, ¶ 97 (July 8, 1996).

32. AWS REPORT, *supra* note 21, at 57.

33. *Id.*

34. See, e.g., AWS REPORT, *supra* note 21, at 13 (claiming AWSs will exist in the near future); Mike Ryder, *Killer Robots Already Exist, and They’ve Been Here a Very Long Time*, THE CONVERSATION (Mar. 27, 2019), <https://theconversation.com/killer-robots-already-exist-and-theyve-been-here-a-very-long-time-113941> [<https://perma.cc/Z3HS-JARP>] (arguing that lethally AWSs will exist soon).

35. See, e.g., Rebecca Crotoof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837, 1837 (2015) (arguing AWS already exist); SCHARRE 2018, *supra* note 25, at 104 (“[A] handful of countries already possess the fully autonomous Harpy”); John K. Hawley, *Not by Widgets Alone*, ARMED FORCES J. (Feb. 1, 2011), <http://armedforcesjournal.com/not-by-widgets-alone/> [<https://perma.cc/2YH4-5YLR>] [hereinafter Hawley] (arguing that modern-day weaponry has reached a critical point in the human-machine spectrum such that there are weapons systems which could properly be described as fully autonomous). One of the more controversial weapons systems currently in use is the South Korean SGR-A1 Sentry Guard Robot, which has the ability to detect, give warning, and provide suppressive fire along the Korean DMZ. Designed to replace human guards, the robot is 1.2 meters high and is mobile. Although normally in a semi-autonomous state, the robot does have an automatic mode in which it is tasked with determining whether to fire on an individual. *Samsung Techwin SGR-A1 Sentry Guard Robot*, GLOBAL SEC. (Nov. 7, 2011), <https://www.globalsecurity.org/military/world/rok/sgr-a1.htm> [<https://perma.cc/B9G7-MT5A>].

36. See, e.g., AWS REPORT, *supra* note 21, at 7 (“Autonomous weapon systems that are highly sophisticated and programmed to independently determine their own actions, make complex decisions and adapt to their environment . . . do not yet exist . . . [T]here are different views on whether future technology might one day achieve such high levels of autonomy”); Jean-François Caron, *Defining Semi-Autonomous, Automated and Autonomous Weapon Systems in Order to Understand Their Ethical Challenges*, 1 DIGIT. WAR 173, 176 (2020), <https://link.springer.com/article/10.1057%2Fs42984-020-00028-5> [<https://perma.cc/ZDE4-FZ49>] [hereinafter Caron] (“Since there are currently no robots that are able to engage targets independently of a human’s will, there is thus no need at this point to entertain a conversation on these science fiction-esque weapons.”).



determine their own actions, make complex decisions and adapt to their environment.”<sup>37</sup> Meanwhile, the United Nations (“U.N.”) Special Rapporteur on extrajudicial, summary, or arbitrary executions, Christof Heyns, defines such machines as “robotic weapon systems that, once activated, can select and engage targets without further intervention by a human operator,”<sup>38</sup> with the distinguishing element being that FAWs have a choice in regards to the target and use of force *without* human input.<sup>39</sup> SAWs are weapons that “can act independently of external control but only according to a pre-defined set of programmed rules”<sup>40</sup> and with “little adaptability to their external environment.”<sup>41</sup>

### 1. Different Spectrums of Autonomy

Conventional literature differentiates between SAWs and FAWs as a dichotomy,<sup>42</sup> but recently scholars have increasingly emphasized the blurred nature of that distinction. Autonomy in machines “can be understood as the capacity . . . to operate without any external control in some or all areas of . . . operation for extended periods of time.”<sup>43</sup> Describing AWs along a spectrum of autonomy provides more flexibility than the FAWs/SAWs dichotomy.

One conception of an autonomy spectrum emphasizes the increasing levels of autonomy as having different characteristics with respect to the human engagement. Intermediate automation, for example, occurs when pre-programmed machines look for certain signs and then alert

---

37. AWS REPORT, *supra* note 21, at 13.

38. Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, ¶ 38, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013).

39. Caron, *supra* note 36, at 173–74.

40. AWS REPORT, *supra* note 21, at 62.

41. *Id.* Unmanned Systems Roadmap, *supra* note 20, at 66.

42. See, e.g., CONG. RSCH. SERV., IF11150, DEFENSE PRIMER: U.S. POLICY ON LETHAL AUTONOMOUS WEAPON SYSTEMS 1 (Dec. 1, 2020), <https://fas.org/sgp/crs/natsec/IF11150.pdf> [<https://perma.cc/37MB-ZYVW>] (“[FAWS are] ‘weapon system[s] that, once activated, can select and engage targets without further intervention by a human operator.’ . . . [SAWS] ‘only engage individual targets or specific target groups that have been selected by a human operator.’”); HUM. RTS. WATCH, LOSING HUMANITY: THE CASE AGAINST KILLER ROBOTS 2 (2012), [https://www.hrw.org/sites/default/files/reports/arms1112\\_ForUpload.pdf](https://www.hrw.org/sites/default/files/reports/arms1112_ForUpload.pdf) [<https://perma.cc/N9QJ-VBSW>] [hereinafter LOSING HUMANITY] (“Human-*in*-the-Loop Weapons: Robots that can select targets and deliver force only with a human command . . . Human-*out*-of-the-Loop Weapons: Robots that are capable of selecting targets and delivering force without any human input or interaction.”).

43. AWS REPORT, *supra* note 21, at 62.

a human operator who decides whether to engage.<sup>44</sup> From there, autonomy can increase by not requiring a human to give an affirmative command but rather only allowing the human to veto the machine-generated attack. Further towards complete automation, a machine could select and engage a target on its own, then wait and call for human authority if it determines there are civilians involved. Finally, the weapons system could assess the possible collateral damage from an attack and decide on its own only to engage if the collateral damage is not disproportionate, without needing to engage a human operator at all.<sup>45</sup>

Autonomy can also be described based on how humans factor into the weapons system's decision-making loop.<sup>46</sup> Human-in-the-loop autonomy occurs when a human selects a target but then the robot tracks and engages the target. Human-on-the-loop occurs when robots can select and engage with a target, but a human operator oversees the weapons system and can override its actions.<sup>47</sup> Finally, human-out-of-the-loop autonomy occurs with weapons that are capable of selecting and engaging in targets without any human input or interaction.<sup>48</sup> This categorical assessment, while more

---

44. Kenneth Anderson & Matthew C. Waxman, *Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation Under International Law*, Colum. L. Sch., Research Paper No. 14-553, 1097, 1101 (2017), [https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3038&context=faculty\\_scholarship](https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3038&context=faculty_scholarship) [<https://perma.cc/9ZXZ-R2NP>] [hereinafter Anderson & Waxman]. Machines like this are not limited to the weapons context—consider the Hewlett-Packard's Central Nervous System for the Earth ("CeNSE"), a project which aims to create a network of remote sensing systems to provide early warning for natural disasters, like tidal waves. CeNSE, though, is also planned to be equipped with some automatic response mechanisms not requiring human input. *CeNSE*, HP, <https://www8.hp.com/us/en/hpinformation/environment/cense.html> [<https://perma.cc/LE36-UYSC>].

45. Anderson & Waxman, *supra* note 44, at 1101.

46. Also known as the OODA loop, the decision-making process consists of four steps: Observe, Orient, Decide, and Act. SCHARRE 2018, *supra* note 25, at 142.

47. Adam Cook, *Taming Killer Robots: Giving Meaning to the "Meaningful Human Control" Standard for Lethal Autonomous Weapon Systems* 4 (JAG Paper No. 1, 2019), [https://media.defense.gov/2019/Jun/18/2002146749/-1/-1/0/JP\\_001\\_COOK\\_TAMING\\_KILLER\\_ROBOTS.PDF](https://media.defense.gov/2019/Jun/18/2002146749/-1/-1/0/JP_001_COOK_TAMING_KILLER_ROBOTS.PDF) [<https://perma.cc/GAB3-5PU3>]; Michael N. Schmitt & Jeffrey S. Thurnher, "Out of the Loop": *Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARV. NAT'L SEC. J. 231, 235 n.12 (2013), <https://harvardnsj.org/2013/05/out-of-the-loop-autonomous-weapon-systems-and-the-law-of-armed-conflict/> [<https://perma.cc/DEP6-UA27>].

48. LOSING HUMANITY, *supra* note 42, at 2–3. The U.S. Patriot system can operate on a "manual" mode, which is a semi-autonomous human-in-the-loop option in which a human operator is required to give approval for an engagement before launch, as well as an auto-fire supervised autonomous mode. The auto-fire mode has different settings available for different threats, but when engaged in auto-fire, the Patriot, once detecting an incoming threat within its parameters, will automatically engage on its own, demonstrating a human-out-of-the-loop capability. It can also be in a more traditional

nuanced than a simple “autonomous or not” distinction, is more definitive than the “levels of autonomy spectrum” but remains vulnerable to a similar criticism—the difference between a human-on-the-loop and a human-out-of-the-loop is not clear-cut.

On a continuum,<sup>49</sup> weapons become progressively more autonomous as technology advances.<sup>50</sup> Humans are given an increasingly “brief opportunity” in AWSs to accept or override the weapon system’s decision.<sup>51</sup> As the human role in the process shrinks, in particular the role in the selection of targets and the extent of force to be used, the weapon system by its very nature becomes more and more autonomous. An AWS thus can become “fully autonomous” when the role of the human is “vanishingly small.”<sup>52</sup> A similar outcome can arise when operators of AWSs are inadequately trained, which in combination with extensive automation can result in a “de facto fully automated system”<sup>53</sup> with no meaningful human supervision or control. As such, even if a weapon system is designed to be semi-autonomous, in practice, it can be fully autonomous.

While describing AWSs along a spectrum of categories rather than a dichotomy is a step in the right direction, it still does not grasp the full complexity of these devices. Autonomy remains a spectrum both *between* weapons systems and *within* a singular system. A single weapons system may not have a single general level of autonomy across its entire structure, but rather “different functions of [the] weapons system may have different levels of autonomy.”<sup>54</sup> The U.S. Navy MK-15 Phalanx C.I.W.S., for example,

---

supervised autonomous mode in which the human has the capacity to override the launch, demonstrating human-on-the-loop. SCHARRE 2018, *supra* note 25, at 141. For more on the Patriot system, see *infra* Part II.C.1.

49. Anderson & Waxman, *supra* note 44, at 1101.

50. See WALLACH & ALLEN, *supra* note 22, at 19 (“[I]n a world where operator error is a fact of life, and humans are unable to monitor the entire state of system software, the pressures for increased automation will continue to mount.”).

51. Laura Dickinson, *Lethal Autonomous Weapons System: The Overlooked Importance of Administrative Accountability*, GWU L. Fac. Publ’ns & Other Works, 2018, at 11. [https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2621&context=faculty\\_publications](https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=2621&context=faculty_publications) [<https://perma.cc/V66J-ADM9>] [hereinafter Dickinson].

52. Anderson & Waxman, *supra* note 44, at 1101.

53. Hawley, *supra* note 34.

54. 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, *Rep. of the 2018 Session of the Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons Systems*, Ann. III(6)(a)(3), U.N. Doc. No. CCW/GGE.1/2018/3 (Oct. 23, 2018). Different critical functions for weapons systems includes acquiring, tracking, selecting, and attacking targets. INT’L COMM. RED CROSS, AUTONOMOUS WEAPON SYSTEMS: TECHNICAL, MILITARY, LEGAL AND HUMANITARIAN ASPECTS 7

automatically “detects, evaluates, tracks, engages, and performs kill assessment against [anti-ship missiles] and high-speed aircraft threats,”<sup>55</sup> but it can also be integrated with other systems and maintain autonomy in only select functions.<sup>56</sup>

### B. The International Legal System Governing Weapons and Use of Force

Before examining how AWSs have been treated under the existing legal system, it is important to explain exactly what that legal system is. Legal regimes governing permissible uses of force, including weapons systems, are governed by international law as well as respective countries’ domestic laws. International law differs greatly from U.S. domestic law.<sup>57</sup> International law is not a set body of law emanating from a central authority.<sup>58</sup> The subjects of international law include states, international organizations, and, in some respects, individuals.<sup>59</sup> The sources of international law are varied and often disputed.<sup>60</sup> As such, international law tends to be more nebulous than the U.S. domestic legal system.

---

(2014), <https://www.icrc.org/en/document/report-icrc-meeting-autonomous-weapon-systems-26-28-march-2014> [<https://perma.cc/KSP3-KPCP>].

55. *MK 15- Phalanx Close-In Weapon System (CIWS)*, AMERICA’S NAVY (Jan. 15, 2019), <https://www.navy.mil/Resources/FactFiles/DisplayFactFiles/Article/2167831/mk-15-phalanx-close-in-weapon-system-ciws/> [<https://perma.cc/QM68-B6Z7>].

56. For example, the Phalanx can be integrated into ship combat control systems to provide sensor and fire-control support. *Id.* The Phalanx differs from other CIWS in that its search, detection, threat evaluation, acquisition, track, firing, target destruction, kill assessment, and cease fire are all incorporated into a single system rather than as a combination of several independent systems. *United States of America: 20 mm Phalanx Close-In Weapon System (CIWS)*, NAVWEAPS (Mar. 5, 2018), [http://www.navweaps.com/Weapons/WNUS\\_Phalanx.php](http://www.navweaps.com/Weapons/WNUS_Phalanx.php) [<https://perma.cc/9V3T-L83E>].

57. For example, for international courts, such as the International Court of Justice (“ICJ”) and the International Criminal Court (“ICC”), there is no required observation of the doctrine of precedent, though the courts do strive for judicial consistency. Statute of the International Court of Justice art. 59, *opened for signature* June 26, 1945, 33 U.N.T.S. 993 [hereinafter *ICJ Statute*]; Rome Statute of the International Criminal Court art. 21 § 2, *opened for signature* July 17, 1998, 2187 U.N.T.S. 90 [U.N.T.S. 104] (entered into force July 1, 2002) [hereinafter *Rome Statute*].

58. CHRISTOPHER JOYNER, *INTERNATIONAL LAW IN THE 21ST CENTURY* 24 (2005).

59. *Id.*

60. Primary rules lay down rights and obligations, whereas secondary rules are used to determine primary rules. André Nollkaemper, *The Power of Secondary Rules of International Law to Connect the International and the National Legal Orders* 5 (Amsterdam Ctr. for Int’l L., Working Paper, Dec. 2009) [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1515771](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1515771) [<https://perma.cc/D5BB-GYAU>].

While the exact sources themselves are often disputed, the *types* of sources of international law are more accepted. International law is governed primarily by treaty and customary law.<sup>61</sup> International conventions or treaties establish rules expressly recognized by the states. These treaties include, most relevantly for this inquiry, the Hague Conventions of 1899 and 1907, the Rome Statute of the International Criminal Court, and the Geneva Conventions of 1949, along with their corresponding additional protocols.<sup>62</sup> International law, as with domestic U.S. law, is comprised of many subsets, including IHL, International Human Rights Law (“IHRL”), and International Criminal Law. Treaties, along with customary international law,<sup>63</sup> form the traditional components of international law.<sup>64</sup> Customary international law is formed by the consent of nation states. Therefore, a state can persistently and vocally object to a custom during its formation in order to not be bound by the custom after it is formed.<sup>65</sup> In other words, customary international law is an almost organic

---

61. See, e.g., *ICJ Statute*, *supra* note 57, art. 38(1) (outlining sources of international law). Sources of international law also include general principles of law, judicial decisions, and teachings of highly qualified publicists, although these are not as relevant to use of force and weaponry law. JAMES CRAWFORD, *BROWNIE’S PRINCIPLES OF PUBLIC INTERNATIONAL LAW* 21–23 (8th ed. 2012); Rome Statute, *supra* note 57, art. 21. General principles of law recognized by civilized nations are derived from universal (or near universal) principles of domestic law. Antonio Cassese, *The Contribution of the International Criminal Tribunal for the Former Yugoslavia to the Ascertainment of General Principles of Law Recognized by the Community of Nations*, in *INTERNATIONAL LAW IN THE POST-COLD WAR WORLD: ESSAYS IN MEMORY OF LI HAOPEI* 43–45 (Sienho Yee & Wang Tieya eds., 2001). Judicial decisions and teachings of highly qualified publicists can be used as subsidiary means for determining the rules of law. These judicial decisions can include those of courts such as the PCIJ, ICJ, international criminal tribunals (such as the ICTY, ICTR, and ICC), specialized bodies like the WTO and ITLOS, human rights courts like the U.N. Human Rights Committee, and regional systems like the Inter-American Court of Human Rights. Highly qualified publicists include U.N. Rapporteurs, scholars, and the International Law Commission.

62. See *Treaties and Customary Law*, INT’L COMM. RED CROSS (Oct. 29, 2010), <https://www.icrc.org/en/document/treaties-and-customary-law> [<https://perma.cc/ZD3F-Q8Q4>].

63. International custom requires widespread and consistent state practice and a sense of legal obligation, or *opinio juris*, as evidence of a general practice accepted as law and binding on all states. Roozbeh B. Baker, *Customary International Law in the 21<sup>st</sup> Century: Old Challenges and New Debates*, 21 EUR. J. INT’L L. 173, 176 (2010), <https://academic.oup.com/ejil/article/21/1/173/363352> [<https://perma.cc/V5LR-VNSN>] [hereinafter Baker]. The Specially Affected States Doctrine articulated by the ICJ clarifies the requirement for widespread and consistent state practice. Under this doctrine, “very widespread and representative participation can be sufficient to create custom if it includes participation of ‘States whose interests were specially affected.’” *North Sea Continental Shelf (Ger. V. Neth.)*, Judgement, 1969 I.C.J. 3, ¶ 73 (Feb. 20, 1969).

64. Baker, *supra* note 63, at 176.

65. *Id.*

growth of international law found in a near-uniform consensus by states on what the law should be.

### 1. Basics of IHL vs. IHRL

The two most relevant subcomponents of international law are IHL and IHRL. IHL is a set of rules which seek, for humanitarian reasons, to limit the effects of armed conflict.<sup>66</sup> Because armed conflict can occur between states (international armed conflict), between a state and a non-state actor, or exclusively between non-state actors (non-international armed conflict),<sup>67</sup> each of which has different intensity threshold minimums, determining when exactly a clash rises to the level of an armed conflict can be problematic.<sup>68</sup> This distinction is important as IHL applies to armed conflicts exclusively, while IHRL applies during peacetime. IHRL also applies during armed conflict, but with IHL as *lex specialis*,<sup>69</sup> some IHRL provisions may be suspended during armed conflict or may hold to a lesser extent than in peacetime.<sup>70</sup> “Weapons law” outlines what requirements weapons must meet to be lawful and permissible. It is a subcategory of law under IHL.<sup>71</sup>

---

66. *What Is IHL*, *supra* note 28; Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶ 36 (July 8). IHL is rooted in the Geneva and Hague Conventions, the Additional Protocols to the Geneva Conventions, a series of treaties governing the means and methods of warfare (i.e., banning blinding laser weapons and landmines), and customary international law. *IHL and Human Rights Law*, INT’L COMM. RED CROSS (Oct. 29, 2010), <https://www.icrc.org/en/doc/war-and-law/ihl-other-legal-regimes/ihl-human-rights/overview-ihl-and-human-rights.htm> [<https://perma.cc/D9CF-X6W3>] [hereinafter *IHL and human rights*].

67. *How is the Term “Armed Conflict” Defined in International Humanitarian Law?* INT’L COMM. RED CROSS (Mar. 2008), <https://www.icrc.org/en/doc/assets/files/other/opinion-paper-armed-conflict.pdf> [<https://perma.cc/UC73-UUQM>].

68. Factors considered by international tribunals when determining if a conflict rises to the level of a NIAC include whether conflict is “protracted” and whether the non-state actor is “organized.” Prosecutor v. Dusko Tadic, Case No. IT-94-10-A, Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 70 (ICTY Oct. 2, 1995).

69. *Lex specialis derogat legi generali* is a legal maxim that states that, generally, specific rules prevail over more general rules. It can lead to either customary international law or treaty law overtaking the other, depending on which is more relevant and which rule is in line with *jus cogens* rules or obligations *erga omnes*. *Lex Specialis*, INT’L COMM. RED CROSS, <https://casebook.icrc.org/glossary/lex-specialis> [<https://perma.cc/U2RW-RW7V>].

70. *What Is IHL*, *supra* note 28. While non-fundamental tenants of IHRL can be suspended when IHL is *lex specialis*, a state cannot suspend or waive certain fundamental rights, including the right to life, the prohibition against torture and inhuman punishment, the outlawing of slavery, and the freedom of thoughts, conscience, and religion. *IHL and Human Rights*, *supra* note 66.

71. *See, e.g.*, Int’l Comm. Red Cross, A Guide to the Legal Review of New Weapons, Means and Methods of Warfare: Measures to Implement Article 36 of Additional Protocol I of 1977, 3-4 (Jan. 2006), <https://e-brief.icrc.org/wp-content/uploads/2016/09/12-A->

Articles 36 and 48 of the Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflict (“Additional Protocol I”) lay out some of the requirements<sup>72</sup> a weapon must meet in addition to numerous customary law restrictions.<sup>73</sup>

There are also more fundamental requirements weapons systems must meet. For AWSs, these require a higher degree of analysis and consideration than other weapons systems. Legal requirements necessitating human judgement require a separate inquiry from philosophical concepts of morality which may suggest a need for humanity in the form of human control in weapons systems. The Martens Clause, a clause in the Preamble of the 1899 Hague Convention, has come to embody both the role of conscience in formulating law and the need for humanity in

---

Guide-to-the-Legal-Review-of-New-Weapons.pdf [https://perma.cc/F3UT-Z4Y9] [hereinafter Legal Review Guide] (stating that states are prohibited from using certain weapons under any circumstances and must ensure new weapons comply with the weapons standards under international law). For particular weapons restrictions relevant to AWS, see *id.* § 1.2.2.3. The use of violative weapons can give rise to individual criminal responsibilities. For example, the ICTY Statute empowers the Tribunal to prosecute violations of IHL for the “employment of poisonous weapons or other weapons calculated to cause unnecessary suffering.” Statute of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the Former Yugoslavia since 1991, S.C. Res. 827, U.N. SCOR, 48th Sess., 3217th mtg., U.N. Doc. S/RES/827 (1993), implementing Report of the Secretary-General Pursuant to Paragraph 2 of the Security Council Resolution 903 (1993), U.N. GAOR, annex 37, art. 3(a) (1993) [hereinafter ICTY Statute] (empowering the Tribunal to punish individuals for the use of weapons that violate IHL). The ICC has jurisdiction over war crimes related to the use of “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.” Rome Statute, *supra* note 57, art. 8.2(b)(xx).

72. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts arts. 36, 48, June 8, 1977, 1125 U.N.T.S. 3 (entered into force Dec. 7, 1978) [hereinafter Additional Protocol I]. These include the basic principles of IHL, including that a weapon must be able to distinguish between civilians and combatants and not cause unnecessary suffering. For more, see *supra* note 36 and accompanying text.

73. See, e.g., Legal Review Guide, *supra* note 71, at 13–14 (stating that the use of chemical weapons and riot-control agents are prohibited); Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶¶ 78, 84 (July 8, 1996) (discussing the Martens Clause, which requires components of humanity); Rupert Ticehurst, *The Martens Clause and the Laws of Armed Conflict*, INT’L COMM. RED CROSS art. no. 317 (Apr. 30, 1997), <https://www.icrc.org/en/doc/resources/documents/article/other/57jnhy.htm> [https://perma.cc/Y3BE-FLJT] (providing background on the Martens Clause).

warfare.<sup>74</sup> IHL, as it stands, requires a “meaningful level of human control” over weapons systems in addition to the weapons requirements found in the Additional Protocol I<sup>75</sup> and under the Martens Clause.<sup>76</sup> What “meaningful human control” entails, however, is not as clear, and there is no single universally accepted interpretation of the Martens Clause.<sup>77</sup>

---

74. E van Trigt, *The Martens Clause: A New Research Guide*, PEACE PALACE LIBR. (Aug. 31, 2017), <https://peacepalacelibrary.nl/blog/2017/martens-clause-newresearch-guide> [<https://perma.cc/JC44-BE8K>].

75. Additional Protocol I, *supra* note 72, art. 1(2); INT’L COMM. RED CROSS, ETHICS AND AUTONOMOUS WEAPON SYSTEMS: AN ETHICAL BASIS FOR HUMAN CONTROL § 8 (Apr. 3, 2018), <https://www.icrc.org/en/document/ethics-and-autonomous-weapon-systems-ethical-basis-human-control> [<https://perma.cc/E5H8-A93H>]; LAW OF WAR MANUAL, *supra* note 23, § 6.5.9.3; Legal Review Guide, *supra* note 71, § 1.2.2.3. This requirement has been confirmed by various international tribunals, including the International Criminal Tribunal for the former Yugoslavia (“ICTY”) as well as state practice. Prosecutor v. Galić, Judgement and Opinion, 2003 ICTY Trial Chamber, Case No. IT-98-29-T, ¶ 58 (Dec. 5, 2003). The Court stated proportionality requires the exercise of judgement by a reasonably well-informed person. Since 2013, 30 countries have called for a ban on fully autonomous weapons systems, though Russia and the United States have opposed proposals at CCW (Convention on Conventional Weapons) meetings to negotiate a treaty banning or limiting AWS, calling the move premature. HUM. RTS. WATCH, STOPPING KILLER ROBOTS: COUNTRY POSITIONS ON BANNING FULLY AUTONOMOUS WEAPONS AND RETAINING HUMAN CONTROL 4–5 (Aug. 2020), [https://www.hrw.org/sites/default/files/media\\_2020/08/arms0820\\_web\\_0.pdf](https://www.hrw.org/sites/default/files/media_2020/08/arms0820_web_0.pdf) [<https://perma.cc/WDX3-R8TQ>]; *Killer Robots: Russia, US Oppose Treaty Negotiations*, HUM. RTS. WATCH (Aug. 19, 2019, 9:01 AM), <https://www.hrw.org/news/2019/08/19/killer-robots-russia-us-oppose-treaty-negotiations> [<https://perma.cc/MX3P-K4ZF>].

76. The Martens Clause states “[u]ntil a more complete code of the laws of war is issued, the High Contracting Parties think it right to declare that in cases not included in the Regulations adopted by them, populations and belligerents remain under the protection . . . of international law, as they result from the usages established between civilized nations, from the laws of humanity, and the requirements of the public conscience.” Convention (II) with Respect to the Laws and Customs of War on Land and Its Annex: Regulations Concerning the Laws and Customs of War on Land Preamble, 32 Stat. 1803 (July 29, 1899).

77. See, e.g., ARTICLE 36, KEY ELEMENTS OF MEANINGFUL HUMAN CONTROL 2 (Apr. 2016), <https://article36.org/wp-content/uploads/2016/04/MHC-2016-FINAL.pdf> [<https://perma.cc/K3E8-98AB>] (“[S]ome human control is required and it must be in some way substantial – we use the term ‘meaningful’ to express that threshold.”). The Martens Clause, to some, is an independent rule of IHL able to be used to ban violative weapons, such as AWS. To others, it is a codification of customary international law. Regardless, the standard that the Martens Clause crystallizes is unclear, especially with respect to what exactly “public conscience” actually is. SCHARRE 2018, *supra* note 25, at 263–64. In 2013, a study showed that 53% of respondents supported a ban for AWS, with 20% of those not directly opposed unsure on the use of AWS. Service members and veterans were in stronger opposition to the use of AWSs than civilians. Charli Carpenter, *Beware Killer Robots: Inside the Debate over Autonomous Weapons*, FOREIGN AFFS. (July 3,



### i. Principle of Proportionality

Proportionality is a fundamental principle of IHL<sup>78</sup> which states that it is prohibited to launch an attack which may be expected to cause an excess of “incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof” in relation to the “concrete and direct military advantage anticipated.”<sup>79</sup> Proportionality applies in both international armed conflicts, for which it is codified in the Rome Statute, and non-international armed conflicts.<sup>80</sup> Proportionality *strictu sensu* reflects a need to conduct a balancing test of “the effects of [the] measures chosen against the objective sought, taking into account whether those effects are excessive

---

2013), <https://www.foreignaffairs.com/articles/united-states/2013-07-03/beware-killer-robots> [<https://perma.cc/BJ25-KETV>]. In 2016, a different study found that when asked about their views on AWSs in a vacuum, the results were similar to the 2013 study. When the respondents were informed that AWSs were more effective and helped protect friendly troops, however, support rose to 61% compared to a baseline of 38%, showing that the context in which the question was asked was highly influential and there was no clear “public conscience” one way or another. Michael C. Horowitz, *Public Opinion and the Politics of the Killer Robots Debate*, 3 RSCH. & POL. 4 (Jan.–Mar. 2016). Both of these studies, however, fall victim to conflating public conscience with public opinion. See Peter Asaro, Jus Nascendi, *Robotic Weapons and the Martens Clause*, in *ROBOT LAW* 367, 373–74 (Ryan Calo et al. eds., 2016) (exploring the differences between “public conscience” and “public opinion”). This flexible reading of the Martens Clause is compounded by an imprecise definition of what constitutes an “attack” under IHL. See *infra* note 83 and accompanying text.

78. IHL fundamental principles include the principle of distinction, or being able to distinguish between combatants and civilians on the battlefield; the principle of avoiding unnecessary suffering prohibiting the use of weapons to cause superfluous injury; the principle of proportionality; and, arguably, the principle of humanity stemming from the Martens Clause. *Fundamental Principles of IHL*, INT’L COMM. RED CROSS, <https://casebook.icrc.org/glossary/fundamental-principles-ihl> [<https://perma.cc/C2TN-U4JH>].

79. Rule 14, *supra* note 27.

80. *Id.*; Rome Statute, *supra* note 57, art. 31(1)(c); see also 2.b. Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices as amended on 3 May 1996 (Protocol II, as amended on 3 May 1996) annexed 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, Geneva, 3 May 1996, Art. 3(8)(c), May 3, 1996, 2048 U.N.T.S. 93 (stating that indiscriminate use is placement of weapons “. . . which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.”); Additional Protocol I, *supra* note 72, art. 51(5)(b) (attacks are indiscriminate when “. . . an attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated.”).

according to those most affected.”<sup>81</sup> In other words, proportionality looks to find the nexus at which the increasing military advantage of the attack meets the decreasing likelihood of avoiding incidental noncombatant casualties at an acceptable level, and it only allows attacks below that level.

There is debate about the proper way to measure proportionality and the scale that must be considered when evaluating an attack for proportionality.<sup>82</sup> Several states have argued that the military advantage must be measured in relation to the military campaign as a whole, rather than to specific isolated aspects of single attacks.<sup>83</sup> While the Rome Statute refers to the civilian injuries being excessive in relation to the “overall military advantage anticipated,”<sup>84</sup> the ICRC has stated that the word “overall” does not change the existing law restricting the principle to each aspect of an attack.<sup>85</sup> As such, proportionality must be measured individually, not cumulatively.

### C. Accountability and Responsibility

Accountability and responsibility are critical aspects of criminal law, domestically and internationally.<sup>86</sup> When there is wrongdoing resulting in

---

81. Emily Crawford, *Proportionality*, MAX PLANCK ENCYCLOPEDIAS INT’L L., Part A (May 2011), <https://opil.ouplaw.com/view/10.1093/law:epil/9780199231690/law-9780199231690-e1459?rskey=n8UXc0&result=1&prd=MPIL> [<https://perma.cc/B495-QPS7>].

82. The International Committee of the Red Cross (“ICRC”), for instance, has argued that the definition of proportionality as codified under Additional Protocol I is binding customary international law, whereas other highly qualified publicists and the United States have not conceded to this view. Rule 14, *supra* note 27; *see* LAW OF WAR MANUAL, *supra* note 23, § 5.10 (discussing the “Proportionality in Conducting Attacks”); James Kilcup, *Proportionality in Customary International Law: An Argument Against Aspirational Laws of War*, 17 CHI. J. INT’L L. 244, 248–50 (2016) (discussing the ambiguity of proportionality).

83. *See, e.g.*, Rule 14, *supra* note 27, at n.27 (including Australia, Belgium, Canada, France, Germany, Italy, Netherlands, New Zealand, Nigeria, Spain, United Kingdom, and United States).

84. Rome Statute, *supra* note 57, art. 8(2)(b)(iv) (emphasis added).

85. JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW VOLUME I: RULES 49–50 (Int’l Comm. Red Cross ed., 3d ed. 2009). Interestingly, this would seem counterintuitive to one plain meaning reading of Additional Protocol I, which defines “attacks” as “acts of violence against the adversary, whether in offence or in defence.” Additional Protocol I, *supra* note 72, art. 49(1).

86. The Nuremburg Tribunal in the Trial of the Major War Criminals affirmed that “[c]rimes against international law are committed by men, not by abstract entities, and only by punishing individuals who commit such crimes can the provisions of international law be enforced.” SECRETARIAT OF THE INT’L MIL. TRIB. NUREMBURG, TRIAL OF THE MAJOR WAR CRIMINALS BEFORE THE INT’L MILITARY TRIBUNALS, NUREMBERG VOLUME I 223 (1947).

the suffering of an innocent party, the desire to hold someone responsible does not stem from a rule of law, IHL or otherwise, but rather from an innate sociological and philosophical desire for justice.<sup>87</sup> Accountability also helps deter future flawed action, encouraging individuals to make morally-sound decisions.<sup>88</sup>

Under both IHL and IHRL, as is the case in domestic legal systems, state responsibility can coexist with individual responsibility.<sup>89</sup> By attaching responsibility to individuals, IHL obligations are reinforced and encouraged with deterrent effects. Under IHL, the state parties to the Geneva Convention and the Additional Protocols are primarily responsible for enforcement and,

---

87. SCHARRE 2018, *supra* note 25, at 262; Andrew Oldenquist, *An Explanation of Retribution*, 85 J. PHIL. 464, 464–65 (1988). See also Jerry Louis Mashaw, *Accountability and Institutional Design: Some Thoughts on the Grammar of Governance* 115–18 (Yale L. Sch. Pub. L. Working Paper, Paper No. 116, 2006), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=924879](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=924879) (suggesting that in determining where accountability should lie, six factors should be considered: (1) who is liable, (2) to whom, (3) what they are liable to be called to account for, (4) through what processes accountability is to be assured, (5) by what standards the putative accountability is to be judged, and (6) what the potential effects are of finding that such standards have been breached). See generally, Wayne P. Pomerleau, *Western Theories of Justice*, INTERNET ENCYCLOPEDIA OF PHIL., <https://iep.utm.edu/justwest/> [<https://perma.cc/79TR-JULN>] (finding that “Western philosophers generally regard justice as the most fundamental of all virtues for ordering interpersonal relations and establishing and maintaining a stable political society”); *Atrocity Crimes*, U.N. OFF. ON GENOCIDE PREVENTION & THE RESP. TO PROTECT, <https://www.un.org/en/genocideprevention/accountability.shtml> [<https://perma.cc/W7CB-F4JC>] (finding that when “[p]roperly pursued, accountability for atrocity crimes can serve not only as a strong deterrent, it is also key to successful reconciliation processes and the consolidation of peace in post-conflict societies”); *First Annual Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the former Yugoslavia Since 1991*, U.N. GAOR, 49th Sess., Agenda Item 152, ¶¶ 11–15 U.N. Doc. A/49/342-S/1994/1007 (1994) (stating that ICTY will contribute to peace because “[t]he only civilised alternative to this desire for revenge is to render justice: to conduct a fair trial by a truly independent and impartial tribunal and to punish those found guilty.”); *Second Annual Report of the International Tribunal for the Prosecution of Persons Responsible for Serious Violations of International Humanitarian Law Committed in the Territory of the former Yugoslavia since 1991*, U.N. GAOR, 50th Sess., ¶ 199, U.N. Doc. A/50/365-S/1995/729 (1995) (“If the Tribunal can prove . . . it is possible to administer international criminal justice, that it is imperative for legal and moral reasons . . . [i]t will . . . send a message to the victims . . . that humanity will not turn its back on them.”).

88. SCHARRE 2018, *supra* note 25, at 263 (quoting Bonnie Docherty); Sheron Fraser-Burgess & Keri L. Rodgers, *Reconsidering the Moral Work of Teaching Framework: Weighing the Moral Hazards of Accountability*, 46 PHIL. STUD. IN EDUC. 62, 65 (2015), <https://files.eric.ed.gov/fulltext/EJ1076663.pdf> [<https://perma.cc/U9K6-A4CP>]; see *infra* Part II.A.2.

89. Elies van Sliedregt, *Individual Criminal Responsibility in International Law* 5–6 (1st ed. 2012).

consequently, any violations.<sup>90</sup> Individual criminal responsibility can attach to individuals who commit war crimes through the International Criminal Court (“ICC”) under the Rome Statute<sup>91</sup> or through domestic law.<sup>92</sup> Not all IHL violations, however, amount to war crimes. As such, not all IHL violations have the potential for criminal accountability under the Rome Statute or under domestic law.<sup>93</sup> To be criminal under the Rome Statute, the crime must

---

90. *Practice Relating to Rule 149. Responsibility for Violations of International Humanitarian Law*, INT’L COMM. RED CROSS IHL DATABASE, [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_rul\\_rule149](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule149) [<https://perma.cc/5E2C-YDVS>].

91. Additional Protocol I, *supra* note 72, art. 85; Rome Statute, *supra* note 57, art. 5, 25; *see also* Geneva Convention for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field art. 49, Aug. 12, 1949, 6 U.S.T. 3114, 75 U.N.T.S. 31 (entered into force Oct. 21, 1950) (“The High Contracting Parties undertake to enact any legislation necessary to provide effective penal sanctions for persons committing, or ordering to be committed, any of the grave breaches of the present Convention . . .”); Geneva Convention for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea art. 50, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 85 (entered into force Oct. 21, 1950) (“The High Contracting Parties undertake to enact any legislation necessary to provide effective penal sanctions for persons committing, or ordering to be committed, any of the grave breaches of the present Convention . . .”).

92. *See, e.g.*, War Crimes, 18 U.S.C. § 2441 (defining “war crime,” which includes murder, defined as “the act of a person who intentionally kills, or conspires or attempts to kill, or kills whether intentionally or unintentionally in the course of committing any other offense . . . one or more persons taking no active part in the hostilities . . .”) (emphasis added). The United States has jurisdiction over war crimes only if the offender or victim is a member of the U.S. Armed Forces or if the offender or victim is a U.S. national. U.S. DEP’T OF JUST. CRIM. DIV. HUM. RTS & SPECIAL PROSECUTIONS SECTION, GUIDE TO HUMAN RIGHTS STATUTES 3 (2017), <https://www.justice.gov/criminal-hrsp/file/1002896/download> [<https://perma.cc/NV6X-4M7N>].

93. Of note are the military war crimes tribunals at Guantanamo Bay, Cuba, which sought to try suspected “war crime” criminals through military tribunals outside the civilian legal system. In the 14 years of operation at Guantanamo tribunals, there were only eight convictions, and of those, three were overturned completely and one partially. This is in comparison to Federal courts having convicted over 660 individuals in terrorism-related crimes since 9/11. A number of the charges filed in the Guantanamo tribunals would not be considered war crimes under international law. ACLU, *Guantanamo by the Numbers*, (May 2018), <https://www.aclu.org/issues/national-security/detention/guantanamo-numbers> [<https://perma.cc/5TKT-WYZF>]; HUM. RTS. FIRST, *Dropped Charges, Overturned Convictions, and Delayed Trials in Guantanamo Military Commissions* (Feb. 16, 2018), <https://www.humanrightsfirst.org/resource/dropped-charges-overturned-convictions-and-delayed-trials-guantanamo-military-commissions> [<https://perma.cc/9HBC-L8VG>]; Johan Steyn, *Guantanamo Bay: The Legal Black Hole*, 53 INT’L & COMPAR. L.Q. 1, 9 (2004). In one decision by the D.C. Circuit Court of Appeals, the Court affirmed the decision of the U.S. Court of Military Commission Review that individuals can be convicted in a military tribunal for a domestic law violation regarding war crimes, even if the crime is not in violation of international law. Here, the

be listed in the Rome Statute, have the proper actus reus, and have the proper mens rea.<sup>94</sup> A single crime can require different mentes rea to be established for each objective element of the offense.<sup>95</sup> Culpability can be defined in the definition of the particular crime or stated in the elements of the particular crime.<sup>96</sup>

### 1. Actus Reus

Actus reus is the conduct element of criminal liability.<sup>97</sup> Actus reus can constitute an act, an omission of an act, or a combination thereof,<sup>98</sup> and includes involuntary and unwilled action.<sup>99</sup> It is contentious whether AWSs, including both Artificial Intelligence systems in charge of controlling physical objects (i.e. robots) and those without a physical presence (i.e. software agents and bots), are able to fulfill the conduct requirement of an actus

---

individual was charged with conspiracy to commit war crimes. *Bahlul v. United States*, 840 F.3d 757, 757–759 (D.C. Cir. 2016).

94. *Prosecutor v. Ongwen*, ICC-02/04-01/15, Defence Motion on Defects in the Confirmation of Charges Decision: Defects in the Modes of Liability (Part II), ¶ 6 (Feb. 1, 2019). Mens rea and actus reus are a “compound concept . . . generally constituted only from concurrence of an evil-meaning mind with an evil-doing hand.” *Morissette v. United States*, 342 U.S. 246, 251 (1952).

95. Mohamed Elewa Badar, *The Mental Element in the Rome Statute of the International Criminal Court: A Commentary from a Comparative Criminal Law Perspective*, 19 CRIM. L. FORUM 473, 4-5 (2008), [https://www.researchgate.net/publication/225475480\\_The\\_Mental\\_Element\\_In\\_The\\_Rome\\_Statute\\_Of\\_The\\_International\\_Criminal\\_Court\\_A\\_Commentary\\_From\\_A\\_Comparative\\_Criminal\\_Law\\_Perspective](https://www.researchgate.net/publication/225475480_The_Mental_Element_In_The_Rome_Statute_Of_The_International_Criminal_Court_A_Commentary_From_A_Comparative_Criminal_Law_Perspective) [<https://perma.cc/UR58-RA9Q>] [hereinafter Badar].

96. *Id.* at 476; Rome Statute, *supra* note 57, art. 30; War Crimes Rsch. Off., *Modes of Liability and the Mental Element Analyzing the Early Jurisprudence of the International Criminal Court* 7–8 (2010), <https://www.wcl.american.edu/impact/initiativesprograms/warcrimes/ourprojects/icclegalanalysisandeducationproject/reports/report-13-modes-of-liability-and-the-mental-element-analyzing-the-early-jurisprudence-of-the-international-criminal-court/> [<https://perma.cc/T6RA-FWVX>].

97. *Prosecutor v. Ongwen*, ICC-02/04-01/15, Defence Motion on Defects in the Confirmation of Charges Decision: Defects in the Modes of Liability (Part II), ¶ 6 (Feb. 1, 2019).

98. Rome Statute, *supra* note 57, art. 28. This was further confirmed by the ICC Lubanga Pre-Trial Chamber in that Article 28 covers both acts and omissions, as well as by the ICTY and ICTR statutes Articles 7(1) and 6(1), respectively. *Prosecutor v. Lubanga Dyilo*, ICC-01/04-01/06, Pre-Trial Chamber I, ¶ 351 (Jan. 29, 2007). Namely, *actus reus* “covers physically perpetrating a crime or engendering a culpable omission in violation of criminal law.” *Prosecutor v. Limaj*, Case No. IT-03-66, Judgement, ¶ 509 (ICTY Trial Chamber Nov. 30, 2005); *Prosecutor v. Galić*, Case No. IT-98-29-T, Judgement & Opinion, ¶ 58 (ICTY Trial Chamber Dec. 5, 2003).

99. Douglas Guilfoyle, *International Criminal Law* 183 (2016).

reus.<sup>100</sup> Whether a weapons system can commit the requisite actus reus is not the focus of this Note, and for the purposes of this paper it is assumed that an actus reus can be established for SAWs or, at the very least, that the question can be severed from an inquiry into establishing the mens rea requirement.

## 2. Mens Rea

In order to be found guilty of a crime, the perpetrator must be proven to have had the necessary state of mind, or mens rea. Mens rea is required to prove culpability and establish moral fault. Having intent, or guilt, is a key part of establishing an effective accountability method with deterrent effects.<sup>101</sup> Under the Rome Statute, the material elements of a crime, unless otherwise established by the law, must be committed with intent and knowledge.<sup>102</sup> The majority of crimes must be committed intentionally. Negligence and recklessness are exceptions that have to be explicitly provided for by law.<sup>103</sup> In this way, unwanted side effects can be equated with intended results, in particular for result crimes, which require either knowledge *or* intent, either being sufficient.<sup>104</sup> Some scholars view Article 30 of the Rome Statute as encompassing three separate categories of mens rea: *dolus directus* in the first degree,<sup>105</sup> *dolus directus* in the second

---

100. Francesca Lagioia & Giovanni Sartor, *AI Systems Under Criminal Law: A Legal Analysis and Regulatory Perspective*, 33 PHIL. & TECH. 433, 441 (2020); Thomas C. King et al., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, 26 SCI. & ENG'G ETHICS 89, 96 (2019), <https://link.springer.com/article/10.1007/s11948-018-00081-0> [<https://perma.cc/WQF6-6JE8>].

101. Intent, in relation to conduct crimes, requires the perpetrator to have meant to engage in the conduct. Rome Statute, *supra* note 57, art. 30(2)(a). In relation to consequence crimes, intent requires the person to know the result will occur in the "ordinary course of events." Badar, *supra* note 95, at 477.

102. Rome Statute, *supra* note 57, art. 30. This includes, for example, the crime of attacks on civilians, which requires that the perpetrator acted "willfully," meaning acting "consciously and with intent." Prosecutor v. Galić, Case No. IT-98-29-T, Judgement, ¶¶ 54, 57 (ICTY Trial Chamber Dec. 5, 2003).

103. Rome Statute, *supra* note 57, art. 30.

104. Badar, *supra* note 95, at 477.

105. *Dolus directus* corresponds to the U.S. Model Penal Code mens rea concept of "purposely." In relation to conduct crimes, *dolus directus* requires the perpetrator to have as his "conscious object" engaging "in conduct of [the requisite] nature or to cause such a result." MODEL PENAL CODE § 2.02(2)(a)(i) (AM. L. INST., PROPOSED OFFICIAL DRAFT 1962). In relation to attendant circumstances, the perpetrator must be "aware of such circumstances or he believes/hopes they exist." *Id.* §2.02(2)(a)(ii).

degree,<sup>106</sup> and *dolus eventualis*.<sup>107</sup> The ICC confirmed this view in the Lubanga Pre-Trial Chamber 1 decision, in which the court found the cumulative reference to intent and knowledge found in Article 30 to require a volitional element encompassed by all three degrees of *dolus*.<sup>108</sup>

For example, a disproportionate attack is, by its nature, a violation of IHL.<sup>109</sup> To be criminal under the Rome Statute, however, the attack must be done with “knowledge” of the disproportionality.<sup>110</sup> This knowledge is a precise *mens rea* element.<sup>111</sup> As such, a merely negligent or reckless disproportionate attack would not incur criminal liability under the Rome Statute. Furthermore, the *actus reus* must be linked to the *mens rea*, in that the *mens rea* must be in relation to the specific *actus reus* outlined in the element of the crime.<sup>112</sup> This requirement becomes especially important in relation to AWSs because AWSs are able to act on their own unprompted from their human operators, disconnecting the *actus reus* from the operators’ *mens rea*. As such, FAWs, by their very nature, make the prosecution of criminal perpetrators more difficult.<sup>113</sup> This concern is appropriate for SAWs currently on the battlefield as well. Because AWSs, including both FAWs and SAWs, can be causal agents that inflict harm but do not qualify as a “who” in terms of accountability and responsibility

106. Also known as *dolus indirectus*, this corresponds to the U.S. Model Penal Code requirement of “knowingly.” In relation to conduct crimes, this requires the perpetrator to be “aware that his conduct is and of that nature.” *Id.* § 2.02(2)(b)(i). In relation to result crimes, the perpetrator must be “aware that it is *practically certain* that his conduct will cause such a result.” *Id.* § 2.02(2)(b)(ii) (emphasis added).

107. *Dolus eventualis* does not have a corresponding element in the U.S. Model Penal Code. Rather, it requires the “knowledge of a possible (as distinct from inevitable) outcome of one’s actions *combined* with a positive mental or emotional disposition towards it.” G. Taylor, *Concept of Intention in German Criminal Law*, 24 OXFORD J. OF LEGAL STUD. 99, 102 (2004) (emphasis added), <https://www.jstor.org/stable/pdf/3600521.pdf> [<https://perma.cc/5RMJ-LL4X>]; Badar, *supra* note 95, at 477.

108. Prosecutor v. Lubanga Dyilo, ICC-01/04-01/06, Pre-Trial Chamber I, ¶ 352 (Jan. 29, 2007). Importantly, the ICTY has not included mere negligence under the category of “willfully,” only recklessness. Prosecutor v. Galić, Case No. IT-98-29-T, Judgment, ¶ 54 (ICTY Trial Chamber Dec. 5, 2003).

109. Ian Henderson & Kate Reece, *Proportionality Under International Humanitarian Law: The “Reasonable Military Commander” Standard and Reverberating Effects*, 51 VAND. J. OF TRANSNAT’L L. 835, 835 (2018), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3108324](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3108324) [<https://perma.cc/5U5C-XVD9>].

110. Rome Statute, *supra* note 57, art. 8(2)(b)(iv).

111. *See id.* (requiring intention to violate the statute).

112. *Id.* art. 30(2–3).

113. Thompson Chengeta, *Accountability Gap: Autonomous Weapon Systems and Modes of Responsibility in International Law*, 45 DENVER J. OF INT’L L. & POL’Y 1, 9 (2020), <https://digitalcommons.du.edu/cgi/viewcontent.cgi?article=1011&context=djilp> [<https://perma.cc/H4LX-C2LB>] [hereinafter Chengeta].

attribution, their existence poses problems for international law mechanisms.

## II. The Accountability Gap: The Problem of AWSs in the Current Legal System

Part II will explore more in-depth the problems posed by AWSs. First, it will examine how AWSs create an accountability gap within the existing criminal law system and how inherent human biases further this gap. Then, the importance of this accountability gap will be established as part of the larger role accountability plays in our legal system. Finally, Part II will discuss the performance issues of AWSs and provide examples of AWSs malfunctioning, demonstrating this accountability gap.

### A. The Problem of the Accountability Gap in AWSs

All weapons systems have the capacity to be misused or to fail, including AWSs.<sup>114</sup> With conventional weapons, the chain of steps leading to a mistake of choice is rather clear-cut. For the most part, the decision-makers involved are human beings, each of whom can be held accountable and responsible if so justified.<sup>115</sup> AWSs, however, can make unpredictable decisions outside the knowledge and control of the user.<sup>116</sup> The difficulty of determining which party should be held responsible, and then successfully holding that party accountable, creates an “accountability gap”<sup>117</sup> in the use of AWSs. The accountability gap problem does not come into play until the AWS does something unexpected or acts in a manner inconsistent with what the human operator intended.<sup>118</sup> Until then, the AWS is acting in accordance

---

114. UNIDIR, SAFETY, UNINTENTIONAL RISK AND ACCIDENTS IN THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES 4 (2016), <https://heinonline-org.prx.law.columbia.edu/HOL/Page?handle=hein.unl/sftunrsk0001&div=1&id=&page=&collection=unl> [https://perma.cc/C8ZM-Q9NX] [hereinafter UNIDIR]; see *infra* Part II.C.

115. SCHARRE 2018, *supra* note 25, at 262; Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 PENN. L. REV. 1347, 1377 (2016), [https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9528&context=penn\\_la\\_w\\_review&httpsredir=1&referer=https://perma.cc/35E8-XWKJ](https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=9528&context=penn_la_w_review&httpsredir=1&referer=https://perma.cc/35E8-XWKJ) [hereinafter Crootof].

116. AWS REPORT, *supra* note 21, at 45.

117. See, e.g., SCHARRE 2018, *supra* note 25, at 261 (quoting Bonnie Docherty); See Darren Stewart, *New Technology and the Law of Armed Conflict: Technological Meteorites and Legal Dinosaurs?*, 87 INT'L LEGAL STUD. 271, 290 (2011), <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1082&context=ils> [https://perma.cc/FV52-827U] (arguing that since inputting certain data and programming prior to operation will not necessarily result in a specific outcome, it is almost impossible to attribute the AWS' behavior directly to a particular human).

118. Crootof, *supra* note 115, at 1366. U.S. policy has been that if a given planned attack is expected to involve civilian casualties above a certain number, the attack must be



with the operator's intent, meaning the actus reus and mens rea are connected. It would neither be "fair nor legally viable" to punish either the operator or the commander due to the lack of mens rea if an AWS was to make a mistake and not correctly carry out its operator's intent resulting in the killing of civilians.<sup>119</sup> With complex machines the responsibility requirement becomes more diffuse regardless of the error rate of the machine; necessarily, there will be less capacity to assign responsibility and thus establish accountability.<sup>120</sup> The issue is not that AWSs are more inclined to make mistakes than conventional weapons, but rather that conventional weapons have greater potential for accountability than AWSs.

With AWSs, the problem of accountability is multifold. First, while the AWS is now an actor in the chain of responsibility with the capacity to make its own choices and to severely influence the choices of others, the AWS itself does not have the corresponding capacity to be held accountable.<sup>121</sup> As will be discussed, while having human supervision over the machine is designed to help account for this issue, in reality, this supervision is predisposed to being superficial at best and is not an adequate safeguard.<sup>122</sup> Second, the chain of accountability as a whole also grows increasingly more complicated and removed, creating more challenges to holding parties responsible.<sup>123</sup> For example, a programmer who creates a piece of software

---

approved at a high level of command and the exact threshold for such approval is changing. *See, e.g.*, Mark Benjamin, *When Is an Accidental Civilian Death Not an Accident?*, SALON.COM (July 30, 2007, 12:01 PM), [https://www.salon.com/2007/07/30/collateral\\_damage/](https://www.salon.com/2007/07/30/collateral_damage/) [<https://perma.cc/R3GN-FFN4>] (saying the threshold is no longer 30).

119. SCHARRE 2018, *supra* note 25, at 261 (quoting Bonnie Docherty); Crotoft, *supra* note 115, at 1385; HUM. RTS. WATCH, MIND THE GAP: THE LACK OF ACCOUNTABILITY FOR KILLER ROBOTS 2 (2015), [https://www.hrw.org/sites/default/files/reports/arms0415\\_ForUpload\\_0.pdf](https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_0.pdf) [<https://perma.cc/KD47-HLEZ>] [hereinafter MIND THE GAP]. Among the suggested parties who should be held accountable for an AWS' mistakes, both operators and commanders are among the most frequently cited. *See, e.g.*, Nikolas Sturchler & Michael Siegrist, *A "Compliance-Based" Approach to Autonomous Weapons Systems*, EJIL: TALK! (Dec. 1, 2017), <https://www.ejiltalk.org/a-compliance-based-approach-to-autonomous-weapon-systems/> [<https://perma.cc/AAW7-9NEK>] ("Given that autonomous weapon systems possess no agency or legal personality of their own, the question of individual criminal responsibility focuses entirely on the responsibility of humans that are involved as operators, commanding officers, programmers, engineers, technicians or in other relevant functions.").

120. *See, e.g.*, Sparrow, *supra* note 24, at 74-75 ("[T]he more autonomous these systems become, the less it will be possible to properly hold those who designed them or ordered their use responsible for their actions.").

121. Sturchler & Siegrist, *supra* note 119.

122. *See infra* Part II.A.1-2.

123. Paul Scharre, *Autonomous Weapons and Operational Risk*, Ethics Autonomy Project 12 (Feb. 2016) <https://s3.useast1.amazonaws.com/files.cnas.org/documents/>

that winds up failing could have made the program years before use or designed it for a purpose different from that for which it was used.<sup>124</sup> Not only does this create issues with holding the programmer responsible, but it also creates problems with holding the rest of the decisionmakers along the chain responsible, even if they made mistakes. As the chain grows increasingly complex, responsibility becomes more diffuse and the likelihood for a “bug” in the system increases.<sup>125</sup>

### 1. The Overarching Problems of Establishing Responsibility and Accountability

This accountability vacuum, or gap, stems from the difficulty in ascribing responsibility to machines once they reach a certain level of autonomous decision-making capacity.<sup>126</sup> Even if a theoretical basis for accountability could be found in law, “in practice those who activate autonomous weapon systems may find a lot of sympathy from judges and others who have to assess their conduct.”<sup>127</sup> When the accused has acted purposefully or with malice, establishing responsibility is much easier due to the clear link between the individual’s *mens rea* and the *actus reus*. In situations where errors go wrong in the normal course of authorized usage, however, the problems of accountability and responsibility remain. As a threshold matter, “[a]ccountability requires . . . knowledge of the facts by those applying the norms.”<sup>128</sup> Machines are becoming more autonomous and elaborate and are operating in unpredictable and complex environments. All the possible interactions which the system may have cannot be foreseen or predicted by the humans involved in its chain of responsibility. Thus, even when a human can be found in the chain,<sup>129</sup> establishing accountability is not straightforward.

---

CNAS\_Autonomous-weapons-operationalrisk.pdf?mtime=20160906080515&focal=none [https://perma.cc/4QYM-YFC8] [hereinafter Scharre 2016].

124. Furthermore, “[t]he most sophisticated autonomous machines may also make reproducible testing difficult, raising serious questions about holding a developer or designer legally responsible for the actions of a [decision-making technique] that cannot be meaningfully tested.” Beard, *supra* note 4, at 650.

125. Crootof, *supra* note 115, at 1374.

126. See generally MIND THE GAP, *supra* note 119, at 6–7 (discussing the lack of accountability in relation to autonomy); Chengeta, *supra* note 113, at 7, 11, 22 (detailing the problem of the accountability gap).

127. AWS REPORT, *supra* note 21, at 46.

128. *Id.* at 45.

129. A human can always be found in the chain if one opens the system’s loop wide enough. For example, there is always both a programmer and a human who turned the machine on. *Id.* at 46.

As an example, a common suggestion is to hold the individual operator in charge of deploying the system responsible.<sup>130</sup> Mens rea, however, would still be required to be established.<sup>131</sup> This could require proving that the operator had foreseen how the system would act. In practice, while the operator may be authorized to use the machine, establishing she had enough knowledge of the details of the systems to foresee the outcome herself, as opposed to relying on others to evaluate the associated risks, would be difficult.<sup>132</sup>

Command responsibility is another mechanism through which accountability and responsibility may be assigned. Under the doctrine of command responsibility, those in command are responsible for the actions of those below them,<sup>133</sup> which does not require direct contribution to the crime's actus reus.<sup>134</sup> However, it also has mens rea problems. Command responsibility is limited to holding superiors accountable only if they "knew

130. See, e.g., Jeffrey S. Thurner, *Examining Autonomous Weapon Systems from a Law of Armed Conflict Perspective*, in *NEW TECH. AND THE L. OF ARMED CONFLICT* 213, 225 (Hitoshi Nasu & Robert McLaughlin eds., 2014) (arguing that the operator should be held responsible based on a reasonableness standard).

131. The crime of conspiracy is not found in the Rome Statute, the Ad Hoc Tribunals' statutes (referring to the ICTY and ICTR), or the statutes of international criminal tribunals generally. The Ad Hoc Tribunals did hold accomplices responsible for their conduct under international criminal law, in that "[a] person who planned, instigated, ordered, committed or otherwise aided and abetted in the planning, preparation or execution of a crime . . . shall be individually responsible for the crime." *Statute of International Criminal Tribunal for the Prosecution of Persons Responsible for Genocide and Other Serious Violations of International Humanitarian Law Committed in the Territory of Rwanda and Rwandan Citizens Responsible for Genocide and Other Such Violations Committed in the Territory of Neighboring States, Between January 1, 1994 and December 31, 1994*, S.C. Res. 955, U.N. SCOR, 49th Sess., Annex, 3453d mtg. at 15, art. 6.1, U.N. Doc. S/RES/955 (1994), reprinted in 33 I.L.M. 1598, 1604 (1994) [hereinafter ICTR Statute]; ICTY Statute, *supra* note 71, art. 7.1; see also Rome Statute, *supra* note 57, art. 25.3 (declaring that anyone who "aids, abets or otherwise assists" in the commission of war crimes or crimes against humanity can be subject to criminal responsibility). Relatedly, there is an argument that customary law has been established recognizing the required mens rea for individuals who aid and abet to be having the knowledge that they assisted in criminal conduct. Beard, *supra* note 4, at 649; Doug Cassel, *Corporate Aiding and Abetting of Human Rights Violations: Confusion in the Courts*, 6 NW. J. INT'L HUM. RTS. 304, 314 (2008).

132. AWS REPORT, *supra* note 21, at 46.

133. *Command Responsibility*, EUR. CTR. FOR CONST. & HUM. RTS., <https://www.ecchr.eu/en/glossary/command-responsibility/> [<https://perma.cc/P95S-EFYD>].

134. Beard, *supra* note 4, at 655-656; see, e.g., PATRICK LIN ET AL., *AUTONOMOUS MILITARY ROBOTICS: RISK, ETHICS, AND DESIGN* 41 (2008) (suggesting it would be up to the military officers responsible for the AWS to ensure that the weapons system "only enters situations in which its character can function appropriately") [hereinafter LIN ET AL.]. An actus reus related to the command responsibility, however, does still need to be established. See *infra* note 134 and accompanying text.

or should have known of a subordinate's criminal act and failed to prevent or punish it."<sup>135</sup> It also deals exclusively with the circumstances in which a commander has explicit and understood responsibility of the humans over which she exercises control.<sup>136</sup> The problem of proving there was explicit and understood responsibility is exacerbated by the need to establish what would be sufficient mens rea.<sup>137</sup> Regardless, a commander would likely not have the requisite knowledge or sufficient warning time to be able to stop an AWS from committing an illegal act.<sup>138</sup> Further, once someone can be proven to have the proper mens rea, it also must be proven they had the proper actus reus.<sup>139</sup> Importantly for both these solutions, there is also the underlying question of fairness—should we hold someone accountable for an action which they neither intended nor could have foreseen? In other words, should we hold them accountable for an action over which they had no choice?<sup>140</sup>

---

135. MIND THE GAP, *supra* note 119, at 2.

136. AWS REPORT, *supra* note 21, at 46. An officer cannot be held “criminally responsible for acts or transactions in which he has not personally engaged or in some other way participated”. Prosecutor v. Tadic, Case No. IT-94-1-A, Judgment, ¶ 186 (ICTY Appeals Chamber July 15, 1999). *See also, e.g.*, Prosecutor v. Halilovic, Case No. IT-01-48-T, Judgment, ¶ 38 (ICTY Trial Chamber I Nov. 16, 2005) (stating that failure to act when there is a duty to do so is the essence of command responsibility); Beard, *supra* note 4, at 657 (explaining that the commander needs to have knowledge of the conduct, the power to prevent or punish it, and the duty to do so, assuming that the subordinate's conduct was in fact criminal). An awareness of the risk that crimes might be committed by subordinates is not enough to create sufficient mens rea. *See, e.g.*, Prosecutor v. Blaskic, Case No. IT-95-14-A, Judgment, ¶ 41 (ICTY Appeals Chamber July 29, 2004) (stating that “[t]he knowledge of any kind of risk, however low, does not suffice for the imposition of criminal responsibility for serious violations of international humanitarian law.”).

137. Beard, *supra* note 4, at 654–56. Command responsibility does not create strict liability for the crimes of subordinates. Rather, in cases of command responsibility, an “accused [person] is not charged with the crimes of his subordinates but [instead] with his failure to carry out his duty as a superior to exercise control.” Prosecutor v. Krnojelac, Case No. IT-97-25-A, Judgment, ¶ 171 (ICTY Appeals Chamber Sept. 17, 2003).

138. MIND THE GAP, *supra* note 119, at 3.

139. Beard, *supra* note 4, at 652. Assuming the commander had the requisite knowledge requirements under mens rea, if the commander “failed to take the necessary and reasonable measures to prevent such acts or to punish the perpetrators thereof,” the commander is not relieved of criminal responsibility for her subordinates' actions. ICTY Statute, *supra* note 71, art. 7.3; ICTR Statute, *supra* note 131, art. 6.3; *see also* Additional Protocol I, *supra* note 72, art. 86.2 (stating that, “[t]he fact that a breach of the Conventions or of this Protocol was committed by a subordinate does not absolve his superiors from . . . responsibility . . . if they knew, or [should have known] . . . that he was committing or was going to commit such a breach . . .”).

140. Importantly, command responsibility creating criminal liability is distinct from a “breach of duty”, which would ordinarily be handled through disciplinary action or administrative accountability within the commander's command structure. Jamie Allan Williamson, *Some Considerations on Command Responsibility and Criminal Liability*, 90 INT'L REV. RED CROSS 303, 303 (2008), <https://www.icrc.org/en/doc/>

## 2. A Bias Towards Agreeing with the AWS Exacerbates the Accountability Gap

While it would seem that a meaningful degree of human control would help prevent mistakes and errors from occurring, in reality, bias makes humans ineffective stopgaps for several reasons. First, humans can have an inherent bias towards machines.<sup>141</sup> Second, even if people were not biased towards the machines, there would still be a bias towards acquiescing to the machine's decision to avoid the potential personal liability from overruling the decision. In this way, human supervision can benefit from, and even exacerbate, the already existing accountability gap through avoiding responsibility for the human's actions.

### i. Presumption that the Machine is Correct

There is a presumption that the AWS is correct in its analysis.<sup>142</sup> This presumption is rooted in a variety of biases, including automation bias and confirmation bias.<sup>143</sup> Operators can over-rely on the machine, resulting in monitoring failures as well as unwarranted or uncritical trust in the automation process.<sup>144</sup> This is a particularly salient issue when the operators are poorly trained or unequipped to handle autonomous systems. Traditional U.S. Army training, for example, does not provide its U.S. Patriot crews with the requisite expertise for effective use of such technology. The Army's post-fratricide board of inquiry criticized the U.S. Patriot battery training for "emphasizing rote battle drills over critical thinking and problem solving."<sup>145</sup>

---

resources/documents/article/review/review-870-p303.htm [https://perma.cc/65PG-7RAL]. It also involves complicity, requiring a crime to have been committed by subordinates to which the commander acquiesced. Beard, *supra* note 4, at 657.

141. Christopher D. Wickens et. al., *Complacency and Automation Bias in the Use of Imperfect Automation*, 57 HUMAN FACTORS 728, 729 (2015).

142. *Id.*; see also, e.g., *infra* P.II.C (describing that in a number of different SAWSs failures, there was overreliance on the machines and a presumption the SAWS was correct).

143. See *infra* P.II.A.2.ii.

144. See, e.g., Peter M. Asaro, *Modeling the Moral User*, 28 IEEE TECH. & SOC'Y MAG. 20, 22 (2009), (defining automation bias as "the tendency to trust an automated system, in spite of evidence that the system is unreliable, or wrong in a particular case.").

145. Hawley, *supra* note 34; Robert R. Hoffman, John K. Hawley & Jeffrey M. Bradshaw, *Myths of Automation, Part 2: Some Very Human Consequences*, IEE INTELLIGENT SYSTEMS, Mar./Apr. 2014, at 82, 84, <http://jeffreymbradshaw.net/publications/Hoffman-Hawley-54.%20Myths%20of%20Automation%20Part%202.pdf> [https://perma.cc/8GCW-P87Z] [hereinafter Hoffman].

This is not limited to dealing with AWSs. Automation bias towards machines can be seen in our day-to-day lives, film and television, and the military more broadly.<sup>146</sup> This bias is only exacerbated by stress and time pressure,<sup>147</sup> both of which are inherent in the armed conflicts in which these systems are most used. One problem with this bias is that automated systems are inherently reliant on past analyses—they rely on data and algorithmic analyses of past situations to project the likelihood of further occurrences based on significant amounts of novel data, meaning they may generate unexpected reactions or situations that the AWS' operators could not anticipate.<sup>148</sup> This leaves a vulnerability in uncertain and novel environments, which is where such technology tends to be deployed. Further exacerbating the problem, people are not good predictors of risk, especially in unfamiliar situations.<sup>149</sup>

ii. Even if the Machine Is Incorrect, the Personal  
Consequences of Action Outweigh  
Those of Inaction.

In general, people are prone to confirmation bias—the tendency to look for evidence that confirms a statement and failing to look for evidence that would disprove it.<sup>150</sup> Due to the lethal nature of their use, any potential

---

146. See, e.g., Kate Goddard et al., *Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators*, 19 J. AM. MED. INFORM. ASSOC. 121, 121 (2011), (explaining different factors that exacerbate and mitigate automation bias); see, e.g., *The Office: Dunder Mifflin Infinity* (NBC television broadcast Oct. 4, 2007) (depicting Michael, one of the characters, who drives his car into a lake because of his GPS' instructions).

147. Mary Cummings, *Automation Bias in Intelligent Time Critical Decision Support Systems*, AM. INST. OF AERONAUTICS & ASTRONAUTICS 1, 5 (2004), (on file with the Columbia Human Rights Law Review).

148. Michael Pilling, *Issues Regarding the Future Application of Autonomous Systems to Command and Control (C2)*, AUSTRALIAN GOV'T DEP'T OF DEF. (2015), <https://apps.dtic.mil/sti/pdfs/ADA620307.pdf> [<https://perma.cc/H7E9-9XSC>]; Beard, *supra* note 4, at 650.

149. See, e.g., Caroline Beaton, *Humans Are Bad at Predicting Futures That Don't Benefit Them*, THE ATLANTIC (Nov. 2, 2017), <https://www.theatlantic.com/science/archive/2017/11/humans-are-bad-at-predicting-futures-that-dont-benefit-them/544709/> [<https://perma.cc/6278-FBSD>] (finding that “[p]sychology research indeed suggests that the more desirable a future event is, the more likely people think it is.”).

150. See, e.g., Iqra Noor, *Confirmation Bias*, SIMPLY PSYCH. (June 10, 2020), <https://www.simplypsychology.org/confirmation-bias.html> [<https://perma.cc/FT4L-CRW9>] (finding that this bias is not limited to autonomous weapons systems, but rather holds true more generally. In this respect, however, AWSs make this problem even more severe than other automated systems).

mistake when using an AWS would likely have dire consequences.<sup>151</sup> In this way, the diffuse responsibility inherent in AWSs production and use protects any wrongdoer when the wrong stems from an *absence* of action on their part; with so many people partially responsible, holding any of them accountable is nearly impossible. Besides, should any of them actually be held accountable, they would only be held accountable for the minute role they played in the overall chain.

If, on the other hand, a user was to believe that the machine was likely wrong in its analysis and take steps to prevent the machine from acting, they would become the singular figure responsible for the result. The ever-famous trolley problem, a thought experiment used to demonstrate ethical dilemmas, illustrates this problem.<sup>152</sup> A runaway trolley is about to hit a group of people unable to move from the tracks, but the operator can pull a lever to divert it onto another set of tracks where it would hit only one person. The question becomes, would the operator do nothing, letting the group of people die? Or would she pull the lever, meaning only one person would die, but she would have taken action leading to their death?<sup>153</sup> Interestingly, the trolley problem is unrepresentative of how people would behave in real life,<sup>154</sup> perhaps suggesting that people act differently in moral quandaries in real life than in training or hypothetical scenarios.

With respect to the supervision of AWSs, automation bias and confirmation bias combine to form an overall tendency towards inaction, partly because of an abdication of responsibility, not because inaction is the right decision. The personal consequences of agreeing with the machine are minimal due to the diffuse nature of responsibility, whereas the personal consequences of disagreeing with the machine are maximal. Compounded with the rapidity with which AWSs can operate, human operators can be not only redundant, but also disadvantageous to the AWSs' functioning.<sup>155</sup> Thus, relying on human supervision to close the accountability gap can be

---

151. See *infra* Part II.C.

152. Lauren Cassani, *Would You Pull the Trolley Switch? Does It Matter?*, THE ATLANTIC (Oct. 9, 2015), <https://www.theatlantic.com/technology/archive/2015/10/trolley-problem-history-psychology-morality-driverless-cars/409732/> [https://perma.cc/TT5G-8EGK].

153. *Id.*

154. See, e.g., Daniel Engber, *Does the Trolley Problem Have a Problem?*, SLATE (June 18, 2018), <https://slate.com/technology/2018/06/psychologys-trolley-problem-might-have-a-problem.html> [https://perma.cc/J8PC-SLEQ] (detailing an experiment in which people were tasked with either allowing five mice to receive a painful electric shock or choosing to have just one mouse be shocked and comparing the results of that experiment with the hypothetical trolley problem).

155. Sparrow, *supra* note 24, at 68; Beard, *supra* note 4, at 671.

described as a fool's errand at best and a fatal flaw with equally fatal consequences at worst.

## B. The Accountability Gap Leads to Problems with Criminal Accountability

### 1. Problems with Attributing Mens Rea for Criminal Accountability

Protection of the right to life is impossible absent accountability for violations.<sup>156</sup> Mens rea requires proof the perpetrator *specifically* had the requisite mens rea, not that any reasonable person in the same situation would.<sup>157</sup> In many cases, humans interacting with AWSs would not satisfy this mens rea requirement.<sup>158</sup> For many crimes under the Rome Statute, the required mens rea is willfulness.<sup>159</sup> War crimes, for example, require the perpetrator to act with intent to commit the violation or to act recklessly.<sup>160</sup>

---

156. AWS REPORT, *supra* note 21, at 45. The principle of individual responsibility and punishment for crimes under international law is the “cornerstone of international criminal law.” Prosecutor v. Tadic, Case No. IT-94-1-T, Judgement, ¶¶ 664–66 (ICTY May 7, 1997). Likewise, the Rome Statute itself recognizes that “the most serious crimes of concern to the international community as a whole must not go unpunished.” Rome Statute, *supra* note 57, at 1; Cf. Jae Egeland, *Humanitarian accountability: putting principles into practice*, HUMANITARIAN PRACTICE NETWORK, at 53 (2005), <https://odihpn.org/magazine/humanitarian-accountability-putting-principles-into-practice/> [<https://perma.cc/NBY8-UWVH>] (arguing that accountability is a current discussion in international spaces and ultimate accountability strengthens capacity to save lives).

157. Prosecutor v. Lubanga, ICC-01/04-01/06-803-tEN, Decision on the Confirmation of Charges, ¶¶ 351–52 (Jan. 29, 2007).

158. Dickinson, *supra* note 51, at 15.

159. Crotoof, *supra* note 115, at 1350–51, 1375; *see, e.g.*, War Crimes Rsch. Off., *Modes of Liability and the Mental Element: Analyzing the Early Jurisprudence of the International Criminal Court 9* (2010) (explaining the requisite mens rea under the Rome Statute).

160. Dickinson, *supra* note 51, at 20. *But cf.* Prosecutor v. Blaskic, Case No. IT-95-14-T, Judgement, ¶ 152 (ICTY Mar. 3, 2000) (stating that both guilty intent and recklessness, which can be likened to serious criminal negligence, are included in the mens rea set out in Article 2 of the Rome Statute). This idea of recklessness being enough to satisfy the necessary intent is reflected in a variety of cases before international tribunals. *See, e.g.*, Prosecutor v. Mucic et al., Case No. IT-96-21, Judgement, ¶¶ 437–39 (ICTY Nov. 16, 1998) (analyzing relevant international law and finding that “the necessary intent . . . required to establish the crimes of wilful [*sic*] killing and murder . . . is present where there is demonstrated an intention on the part of the accused to kill, or inflict serious injury in reckless disregard of human life”); Prosecutor v. Stakic, Case No. IT-97-24, Judgement, ¶ 587 (ICTY Jul. 31, 2003) (“[B]oth a *dolus directus* and a *dolus eventualis* are sufficient to establish the crime of murder . . .”).



For operators of AWSs, however, the concern generally is not direct intent, but rather risk acceptance with respect to civilian deaths. Further, not only is there an issue establishing accountability along the horizontal line of production, development, and use of the AWS itself should it malfunction, but also the military decision-making model itself makes establishing vertical accountability vastly more difficult.<sup>161</sup> Having a structured layering of command in which each person must obey all *lawful* orders given to them by the higher rank distributes responsibility in its own right, making it even harder to impute criminal responsibility to just one actor.

## 2. Lowering the Mens Rea Threshold Would Not Solve the Problem

When an AWS malfunctions in a way that could not have been obvious to the user, the questions of who *should* be held accountable and who *can* be held accountable become even more complicated. Often, the answer to one is not the same as the answer to the other. Further, establishing a hard and fast rule on who would be accountable should such an event occur, thereby providing proper deterrence, is also not an easy feat.

One proposed solution has been to impose criminal punishment based on strict liability or mere negligence, lowering the bar to make it easier to include operators of AWSs,<sup>162</sup> or to allow for Joint Criminal Enterprise (“JCE”) liability.<sup>163</sup> This solution is contrary to the fundamental tenets of

---

161. See generally Markus Wagner, *The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems*, 47 *VAND. J. TRANSNAT'L L.* 1371, 1402 (2014) (discussing the difficulties that AWSs cause for assigning individual responsibility) [hereinafter Wagner]; Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, ¶ 77, U.N. Doc. A/HRC/23/47 (Apr. 9, 2013) (“The composite nature of [lethal autonomous robot] technology and the many levels likely to be involved in decisions about deployment result in a potential accountability gap or vacuum.”). For this Note, “horizontal” accountability is used to describe accountability along the line of production of the weapon itself, including designing, building, programing, testing, and operating the weapon, whereas “vertical” accountability is analogous to the traditional chain of command structure, with the operator and subsequent commanders. Wagner, *supra* note 161, at 1371.

162. Crotoof, *supra* note 115, at 1394.

163. JCE liability, while not in the Ad Hoc Tribunals’ statutes, was developed mostly by judges and prosecutors before the ICTY to impose “criminal responsibility on individuals responsible for furthering the commission of mass murder . . . and other atrocities in the former Yugoslavia.” JCE liability has been nicknamed “Just Convict Everyone”. The three forms of JCE liability all require the prosecution to show a “plurality of persons” “the existence of a common plan, design, or purpose which amounts to or involves the commission of a crime provided for in the Statute”, and “participation of the accused in the common design”. Beard, *supra* note 4, at 661; Prosecutor v. Tadic, Case No. IT-94-1-A, Judgement, ¶ 227 (ICTY July 15, 1999). The three forms include when (1) the

individual criminal responsibility and it has been criticized as greatly diminishing the requirements of individual criminal responsibility.<sup>164</sup> JCE liability requires a common criminal purpose present and “an intention to systematically target civilians . . . established.”<sup>165</sup> Further, if every individual were to be “held criminally liable for negligent actions in war and . . . her commander . . . liable for negligence, [then] every commander would be a war criminal.”<sup>166</sup> The problem is not that the mens rea itself is too high, but rather that the manufacturing and use of AWSs make it too difficult to reach to achieve any meaningful accountability.

### C. The Problems Posed by SAWs

Concerns with AWSs’ accountability are relevant because an AWS can malfunction or perform in unexpected manners. SAWs already used in the battlefield have demonstrated the risks posed by these machines. One concern with an AWS is that the weapons system might not perform in the exact manner the human operator intended.<sup>167</sup> This could cause civilian casualties, friendly fire incidents, or unintended escalations in severity and degree of warfare. There are a number of risks posed by SAWs that can cause these unintended results. Increased autonomy can make the human operator’s job more complicated, potentially reducing the operator’s situational awareness.<sup>168</sup> Such undisciplined automation is caused by designers and builders of SAWs not fully considering or accounting for “the downstream consequences for human performance.”<sup>169</sup> General criticisms of programmer bias and prejudice in automation reflect these performance issues,<sup>170</sup> but with an AWS, the concern is further attenuated because the risk

---

perpetrator acted with the intent to perpetrate a certain crime, (2) the perpetrator holds a position of authority in a military or administrative unit and has personal knowledge of the system of ill-treatment and “inten[ds] to further this common concerted system of ill-treatment”, and (3) the perpetrator has a “criminal intention to participate in a common criminal design” while “criminal acts other than those envisaged in the common criminal design are likely to be committed by other participants in the common design.” Prosecutor v. Dusko Tadic, Case No. IT-94-1-A, ¶¶ 196, 204, 206, 223; Beard, *supra* note 4, at 661–62.

164. Dickinson, *supra* note 51, at 21; Mark A. Drumbl, Atrocity, Punishment, and International Law 39 (2007).

165. Beard, *supra* note 4, at 662.

166. Crootof, *supra* note 115, at 1384.

167. Scharre 2016, *supra* note 123, at 8–10.

168. UNIDIR, *supra* note 114, at 3.

169. Hoffman, *supra* note 145, at 82.

170. See Nicol Turner Lee et al., *Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms*, BROOKINGS INST. (May 22, 2019), <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/> [https://perma.cc/XJ83-DDUA] (explaining programming bias found in hiring software and facial recognition which leads

of improper automation includes the potential loss of innocent lives.<sup>171</sup> Consequently, the operators' roles within the SAWSs are defined oftentimes as byproducts of the automation.

## 1. The U.S. Patriot System as an Example of SAWSs' Flaws

### i. Operation Desert Storm

Throughout Operation Desert Storm, the Patriot system<sup>172</sup> was instructed to intercept tactical ballistic missiles, including Scud missiles.<sup>173</sup> The Patriot system suffered from a software bug—after very long run times, the system could experience a shift of the range gate.<sup>174</sup> Notably, U.S.

to disparate impact); Craig S. Smith, *Dealing with Bias in Artificial Intelligence*, N.Y. TIMES (Nov. 19, 2019), <https://www.nytimes.com/2019/11/19/technology/artificial-intelligence-bias.html> [https://perma.cc/3CE2-A7PD] (on file with the *Columbia Human Rights Law Review*).

171. See U.N. INST. FOR DISARMAMENT RSCH., ALGORITHMIC BIAS AND THE WEAPONIZATION OF INCREASINGLY AUTONOMOUS TECHNOLOGIES: A PRIMER 3–9 (2018), <https://unidir.org/sites/default/files/publication/pdfs/algorithmic-bias-and-the-weaponization-of-increasingly-autonomous-technologies-en-720.pdf> [https://perma.cc/GL2T-J5V6] (explaining the difference biases algorithms may have and the possible discriminatory effect); Noel Sharkey, *The Impact of Gender and Race Bias in AI*, ICRC HUMANITARIAN L. & POL'Y (Aug. 28, 2018), <https://blogs.icrc.org/law-and-policy/2018/08/28/impact-gender-race-bias-ai/> [https://perma.cc/Y4HF-UU7X]; Emily Addison, *Autonomous Weapon Systems: Understanding Learning Algorithms and Bias*, U.N. OFF. FOR DISARMAMENT AFFS. (Oct. 13, 2017), <https://www.un.org/disarmament/update/auto-weapon-systems-understanding-learning-algorithms-and-bias/> [https://perma.cc/HU2R-7FVX].

172. The Patriot system operated as part of a battalion of six separate Patriot batteries, each one consisting of a ground-based radar unit, an Engagement Control Station, eight missile launchers, and a Communications Relay Group. The weapons control computer of the Patriot system during Operation Desert Storm performed the system's major functions for tracking, interception, battle management, and command and control functions. U.S. GOV'T ACCOUNTABILITY OFF., GAO/IMTEC-92-26, PATRIOT MISSILE DEFENSE: SOFTWARE PROBLEM LED TO SYSTEM FAILURE AT DHAHRAN, SAUDI ARABIA 1, 3–4 (1992) [hereinafter GAO Report].

173. BERNARD ROSTKER, IRAQ'S SCUD BALLISTIC MISSILES, DEP'T OF DEF. (July 25, 2000) (Information Paper) (on file with the *Columbia Human Rights Law Review*). A Scud missile is a family of missiles designed to be short-range and liquid-fueled. Originally Soviet, Scuds generally carry conventional explosives. *SS-1 "Scud"*, MISSILETHREAT CSIS MISSILE DEF. PROJECT, <https://missilethreat.csis.org/missile/scud/> [https://perma.cc/5XYM-KF8S].

174. GAO Report, *supra* note 172, at 7–8. A range gate is “an electronic device in the radar [which] calculates an area in the air space for where the system should look next for the incoming missile.” Andrew Lum, *Patriot Missile Software Problem*, UNIV. OF SYDNEY: ENG'G 2 (2012), [http://course1.winona.edu/ppaulson/Spring2013/MIS362/Assignments/Assignment04/patriot\\_bug.pdf](http://course1.winona.edu/ppaulson/Spring2013/MIS362/Assignments/Assignment04/patriot_bug.pdf) [https://perma.cc/N5VE-MQDA].

commanders declined to use available external data recorders to retain performance information.<sup>175</sup> The U.S. Army had received notice from Israeli allies detailing the loss in accuracy in the range gate resulting from the system running continuously for just eight consecutive hours. This error increased in severity the longer the system ran.<sup>176</sup> When the U.S. Army warned the Patriot operators of the bug, officials provided no clarification as to what qualified as a “very long run time,” leaving it to each operator’s interpretation.<sup>177</sup> The U.S. Army had never used the Patriot system to defend against Scud missiles and had never prepared for the Patriot to run continuously for prolonged periods of time.<sup>178</sup>

In February 1991, in Dhahran, Saudi Arabia, a Patriot system deployed by the United States failed to track and intercept a Scud missile.<sup>179</sup> The missile subsequently hit an Army barracks still intact, killing twenty-eight soldiers.<sup>180</sup> The mistake was attributed to the problem with the inaccurate tracking calculation reported by the Israeli forces.<sup>181</sup> The U.S. Army had received the data from Israeli officials outlining the problem with the Patriot two weeks prior to the Dhahran attack.<sup>182</sup> This particular Patriot system had been operating continuously for over one hundred hours, making the error severe enough that the system was looking in the wrong place for the incoming missile.<sup>183</sup>

---

175. GAO Report, *supra* note 172, at 6.

176. The data provided by the Israeli forces suggested that after about 20 hours, the error would become large enough to cause the radar to look in the wrong place for the target. The Israeli data showed that there was a 20 percent shift in the system’s radar range gate after eight hours. To have the highest probability of tracking the target, the target needed to be in the center of the range gate. The Patriot system would not be able to track a Scud when the range gate shifted 50% or more. Extrapolating from the Israeli data, it was determined that the range gate would shift 50% after approximately 20 hours of continuous use. *Id.* at 7–8.

177. *Id.* at 9. Army officials assumed that Patriot users were not running their systems for eight or more continuous hours and did not use the data to determine how long the Patriot could operate before the inaccuracy would render the system ineffective.

178. *Id.* at 1.

179. *Id.*

180. At the time, a second artillery battery system that was also placed to protect the barracks was shut down for repairs. *Id.*; Eric Schmitt, *U.S. Details Flaw in Patriot Missile*, N.Y. TIMES (June 6, 1991), <https://www.nytimes.com/1991/06/06/world/us-details-flaw-in-patriot-missile.html> [<https://perma.cc/4MS9-7ZXXN>] (on file with the *Columbia Human Rights Law Review*) [hereinafter Schmitt June].

181. GAO Report, *supra* note 172, at 7.

182. *Id.* at 1.

183. *Id.*

Initially, however, Army officials explained this Patriot failure only as a “freak” accident<sup>184</sup> and said that the Scud missile broke up prior to impact rather than hitting the barracks intact.<sup>185</sup> The family members of the victims were unable to get more information about the Dhahran attack, and the Army attributed the attack solely to a computer error, writing off operator error and focusing on the computer software program.<sup>186</sup> Senior Army officials blamed the failure on the “unforeseen combination of ‘dozens’ of variables”, calling it an “anomaly that never showed up in thousands of hours of testing.”<sup>187</sup>

## ii. Operation Iraqi Freedom

During Operation Iraqi Freedom, the U.S. Armed Forces had eleven total Patriot engagements, nine of which resulted in successful tactical ballistic missile engagements.<sup>188</sup> In 2003, there were two instances (18% of total Patriot engagements) of mistaken fires resulting in fratricides.<sup>189</sup> In the first, a U.S. Patriot shot down a British FR-4 Tornado, destroying the aircraft and killing the crew onboard. There, the Patriot system misidentified the British aircraft as an anti-radiation missile.<sup>190</sup> At the time, the Patriot’s equipment was outdated, meaning it could not connect to other radars to confirm the missile.<sup>191</sup> The human operator, assuming the Patriot was correct, mistakenly accepted the Patriot’s incorrect identification. The Lieutenant who made the call to fire stayed at her post and was cleared of all

---

184. Schmitt June, *supra* note 180.

185. Eric Schmitt, *AFTER THE WAR; Army Is Blaming Patriot’s Computer for Failure to Stop Dhahran Scud*, N.Y. TIMES (May 20, 1991), <https://www.nytimes.com/1991/05/20/world/after-war-army-blaming-patriot-s-computer-for-failure-stop-dhahran-scud.html> [<https://perma.cc/4VTV-TN43>] [hereinafter Schmitt May]; MICHAEL BARR, *CASE STUDY: LETHAL SOFTWARE DEFECTS - PATRIOT MISSILE FAILURE*, BARR GROUP, 3 (Sept. 22, 2021), <https://barrgroup.com/sites/default/files/case-study-patriot-missile-defects.pdf> [<https://perma.cc/NBZ6-EEBG>] [hereinafter BARR].

186. Schmitt May, *supra* note 185; BARR, *supra* note 185, at 1.

187. Schmitt May, *supra* note 185; BARR, *supra* note 185, at 3.

188. Hawley, *supra* note 34.

189. *Id.*

190. The Tornado’s Identification Friend or Foe (IFF) signal, which broadcasted to friendly ground and air forces to let them know not to fire, was not on at the time. Even if it had been working, the IFF had not been loaded into the Patriot’s computers. As the aircraft descended to land at Ali Al Salem air base, the Patriot system read the radar signal as an incoming missile. While the Patriot’s job was to shoot down ballistic missiles, which had a distinct trajectory from either landing aircrafts or anti-radiation missiles, the Patriot was authorized to engage if the missile was homing in on the radar. SCHARRE 2018, *supra* note 25, at 138–39.

191. *Id.* at 139.

wrongdoing.<sup>192</sup> Despite changes made to avoid similar mistakes in the future,<sup>193</sup> a different Patriot unit had its own misfire less than two weeks later.

In the second instance, the Patriot system falsely identified an incoming track of a tactical ballistic missile,<sup>194</sup> the missile for which the Patriot was designed to track and destroy. The human operators set the missile in ready status while it was in auto-fire mode for ballistic missiles.<sup>195</sup> As a result, the Patriot system fired without further human interaction, even though the operators had not intended it to fire. When the Patriot found no ballistic missile, the Patriot's trackers instead locked onto a U.S. Navy F/A-18C Hornet nearby and shot it down, killing the operator instantly.<sup>196</sup>

A report written by the Defense Science Board Task Force on the performance of the Patriot System throughout Operation Iraqi Freedom found that Patriot operators were trained to trust the system's software and tended to assume that "data are routinely communicated from one system to the other, that targets are correlated, and target information is shared and assimilated by all."<sup>197</sup> Flawed performance,<sup>198</sup> however, had been seen in

---

192. This Patriot would have two more successful missile shootdowns before the end of the war, and that battery system as a whole would be responsible for 45% of the successful ballistic shootdowns in the war. *Id.* at 140–1.

193. After the first accidental shootdown, the U.S. Army ordered all Patriots to be kept in manual mode for anti-radiation missiles and for the systems to be kept in "standby" mode. In standby mode, the systems could track targets, but could not fire without a human putting the system back into operating mode. This two-step process, bringing the system to operation status and authorizing the system to fire, was hoped to be able to prevent another fratricide incident. *Id.* at 141.

194. There actually was no missile—the Patriot picked up a "ghost track" from electromagnetic interference. Once the two missiles were launched by the Patriot, however, the Patriot missiles' seekers were activated to "lock onto" the supposed incoming ballistic missile. When no ballistic missile was found, the Patriot missiles' seekers locked onto whatever else they could find. *Id.* at 143.

195. The requirement to keep the Patriot in manual mode instituted after the last fratricidal incident was only for anti-radiation missiles, not for ballistic missiles. *See supra* note 193 and accompanying text.

196. The Hornet's Identification Friend or Foe System ("IFF") signal was operating and was picked up on the Patriot radar as an aircraft, to no avail. SCHARRE 2018, *supra* note 25, at 143.

197. Off. of the Under Sec'y of Def. for Acquisition, Tech., and Logistics, Report of the Def. Sci. Board Task Force on Patriot System Performance: Report Summary 2 (2005), <https://dsb.cto.mil/reports/2000s/ADA435837.pdf> [https://perma.cc/E2J2-EZVC] [hereinafter Task Force Report].

198. The report found that one aspect of the Patriot, the Mode IV IFF system, performed very poorly. *Id.* at 2.

many training exercises prior to and throughout its deployment.<sup>199</sup> Because the possible Patriot-friendly aircraft observations during testing were in the millions, even with a very low probability of failures it was still known that fratricide incidents could result<sup>200</sup> and were even likely to occur. The report concluded that the Patriot system needed “a protocol that allows more operator oversight and control of major system actions.”<sup>201</sup>

In none of the fratricide instances discussed did the human operators *intend* harm. The post-incident assessment traced the Patriot crews’ shortcomings to decisions made by concept developers, software engineers, procedure developers, testers, trainers, and unit commanders years earlier.<sup>202</sup> Investigations found that no one involved had acted criminally, negligently, or recklessly.<sup>203</sup> In the three-week span over which the U.S. Forces approached Baghdad, Iraq fired at least twenty-three ballistics and cruise missiles. Of the fourteen missiles not engaged by Patriots, four were outside the range of any Patriot system and one exploded shortly after launch.<sup>204</sup> No official explanation was provided for why the other nine were not fired upon, though for at least three it was suggested the failure was because the targets were incorrectly judged to be non-threatening.<sup>205</sup> The U.S. Army announced in December 2004 that it found the Patriot Missile System to be successful in performing its mission in Operation Iraqi Freedom.<sup>206</sup>

## 2. AC-130 2015 Médecins Sans Frontières Attack

As discussed in the Introduction,<sup>207</sup> in October 2015, a U.S. Army AC-130 aircraft fired on and destroyed a hospital.<sup>208</sup> The intended target, a different building in the city believed to be occupied by insurgents, was

---

199. In addition to the two fratricide instances discussed above, there was also an instance which, while not resulting in actual fratricide, came remarkably close. A U.S. Air Force F-16 fired upon a Patriot battery, though no injuries or fatalities occurred. SCHARRE 2018, *supra* note 25, at 143.

200. Task Force Report, *supra* note 197, at 2.

201. *Id.* at 3.

202. Hawley, *supra* note 34; Hoffman, *supra* note 145, at 83.

203. Dickinson, *supra* note 51, at 31.

204. Wade Boese, *Army’s Report Details Patriot Record in Iraq War*, ARMS CONTROL ASS’N. <https://www.armscontrol.org/act/2003-11/press-releases/army-report-details-patriot-record-iraq-war> [<https://perma.cc/36XX-H3UA>].

205. *Id.*; Task Force Report, *supra* note 197, at 1.

206. *Army Announces Patriot Missile System’s Performance in Operation Iraqi Freedom*, DEF. AEROSPACE (Dec. 10, 2004), [http://www.defense-aerospace.com/articles-view/release/3/50401/us-army-clears-patriot-in-iraq-\(dec.-14\).html](http://www.defense-aerospace.com/articles-view/release/3/50401/us-army-clears-patriot-in-iraq-(dec.-14).html).

207. *See supra* Introduction.

208. Stewart & Torbati, *supra* note 3.

located nearby.<sup>209</sup> The AC-130 was able to track and engage multiple targets and relied on visual spotting at low altitudes.<sup>210</sup> The twenty-nine-minute-long strike began at 2:08 AM. The AC-130 was deployed sixty-nine minutes earlier than scheduled, before it was able to receive the no-strike list that would have identified the hospital properly as a protected building.<sup>211</sup> When the crew then entered the coordinates of the target, the gunship's system directed the aircraft to an empty field. Senior officials also confirmed the target's coordinates without checking the no-strike list, as the protocol required, because they assumed the AC-130 was correct. Although the grid location identifier of the AC-130 did eventually identify the correct building, and there was no hostile enemy activity at the MSF building, the U.S. attack continued.<sup>212</sup>

An investigation conducted in accordance with the Uniform Code of Military Justice ("UCMJ") found that the attack did not constitute a war crime. Instead, sixteen military personnel faced administrative punishment after the "military investigation determined the attack to be unintentional."<sup>213</sup> Echoing concerns about administrative punishment more generally, the Asia policy director for Human Rights Watch, John Sifton, expressed concern that "any decision-making about any possible criminal charges, if they are made, remains within the chain of command responsible for the military operations in Afghanistan."<sup>214</sup>

In each of these instances, no criminal liability was found. The failures were attributed to a combination of technological glitches, human error, and overreliance on the AWSs. Of the two fratricidal incidents during Operation Desert Storm, there was a human operator required to give the command to fire or bring the missile launchers into operation.<sup>215</sup> The Patriot itself did not break—it operated in accordance with its design.<sup>216</sup> The operators themselves were not negligent and no one was found at fault in either incident.<sup>217</sup> Instead, the relevant personnel, and even the military

---

209. *Id.*

210. *AC-130*, *supra* note 1.

211. Rosenberg, *supra* note 2.

212. Stewart & Torbati, *supra* note 3.

213. Aisch, Keller & Peçanha, *supra* note 15.

214. Rod Nordland, *U.S. General Says Kunduz Hospital Strike was 'Avoidable'*, N.Y. TIMES (Nov. 25, 2015), <https://www.nytimes.com/2015/11/26/world/asia/afghanistan-kunduz-hospital-airstrike.html> (on file with the *Columbia Human Rights Law Review*).

215. *See supra* P.II.C.1.

216. Despite knowing, for example, that the friendly forces' identifying signal was inadequate to prevent fratricides and the potential for the Patriot to confuse incoming aircraft with anti-radiation missiles, the Patriot was still put in force, and no one could be held accountable for its mistakes. SCHARRE 2018, *supra* note 25, at 144.

217. Dickinson, *supra* note 51, at 31; *see supra* Part II.C.1.



community at large, succumbed to automation and confirmation biases; they had a “culture of ‘trusting the system without question.’”<sup>218</sup> Simply put, the operators trusted the automated Patriot System, but unfortunately the AWS was wrong.

In both U.S. Patriot fratricides during Operation Iraqi Freedom,<sup>219</sup> the machine fired without regard to the operator’s intent. The mens rea in each situation was disconnected from the actus reus or was found not to be satisfactorily willful, in part due to the unpredictable manner in which the AWS performed. In the AC-130 attack, it was nearly universally acknowledged that mistakes were made up and down the chain of command.<sup>220</sup> Here, a spread of responsibility resulted in the accountability becoming diffuse up and down the chain of command and across the production and implementation line of the AC-130. As a result, because so many people could have been held accountable, no one could be held accountable in a criminal court.

### III. Implications for Legality of Use: Defining a Lens of Analysis

The importance of the accountability gap—and the weaknesses it poses for criminal accountability and for increasing the bias towards machine reliance—is critical to understand when evaluating the use of SAWs. In Part III, this Note will explore several proposed solutions to the accountability gap and explain why they are unsatisfactory. Next, it will propose a suggested dichotomic lens through which to evaluate the use of AWSs. Finally, using that lens, the Note will conclude that SAWs in their present state are unable to be reconciled with the current legal accountability options.

#### A. Problems Within the Criminal Justice System and Humans as Ineffective Stopgap Mechanisms

The accountability gap creates problems ensuring mens rea can be established when the weapons system goes wrong<sup>221</sup> and creates an incentive for supervising operators to follow the AWS’ recommendation even when the weapons system might be wrong.<sup>222</sup> There is a propensity in favor of letting the machine do what it wants despite existing reservations as to its

---

218. SCHARRE 2018, *supra* note 25, at 144.

219. *See supra* Part II.C.1.

220. *See supra* note 215-218 and accompanying text.

221. Chengeta, *supra* note 113, at 3.

222. *See supra* Part II.A.2.

accuracy.<sup>223</sup> The personal consequences of agreeing with the machine are minimal thanks to the diffuse responsibility along both the chain of command and the chain of production, whereas the personal consequences of disagreeing with the AWS can be great.<sup>224</sup>

These consequences of disagreeing with the AWS can be great even when the operator is correct. In September 1983, a Soviet early warning satellite reported with the highest confidence that an intercontinental ballistic missile was heading towards the Soviet Union from the United States.<sup>225</sup> Stanislav Petrov, the Soviet officer on duty, thought the detection was a computer error.<sup>226</sup> According to protocol, he should have notified his superiors of the impending attack, which would have triggered a sequence of steps to initiate a counterattack with the Soviet nuclear missiles.<sup>227</sup> Despite being uncertain, Petrov<sup>228</sup> dismissed the missile, as well as the warning of four more incoming, as an error. While Petrov turned out to be correct, he later attributed his choice to his civilian training, saying that his full-time professional soldier colleagues would have followed orders and protocol and reported the incident to their higher chain of command.<sup>229</sup> If Petrov had been removed from the decision-making loop, and left leaving the decision up to the machine, the weapons system would have fired causing a nuclear war.<sup>230</sup> Instead, Petrov overrode a complex system, as well as his military protocol, making him solely responsible for the decision.<sup>231</sup> Even though he was correct, Petrov received no award, was moved to a “more discreet”

---

223. See *supra* text accompanying note 218.

224. *Id.*

225. SCHARRE 2018, *supra* note 25, at 1.

226. *Id.*

227. While not an AWS, this automatic series of steps that would have followed subsequent to Petrov’s notification shows that even without a fully SAWS, the pressure and accountability problems exigent in the military structure and the development of these weapons systems are present.

228. In fact, Petrov put the odds at about 50/50. SCHARRE 2018, *supra* note 25, at 2.

229. Pavel Aksenov, *Stanislav Petrov: The Man Who May Have Saved the World*, BBC NEWS (Sept. 26, 2013), <https://www.bbc.com/news/world-europe-24280831> [<https://perma.cc/D4P2-4CJ8>]; Grant Schreiber, *A Lesson from the Man Who Saved the World*, REAL LEADERS (Dec. 5, 2017), <https://real-leaders.com/a-lesson-from-the-man-who-saved-the-world/> [<https://perma.cc/S9WR-TQQL>].

230. SCHARRE 2018, *supra* note 25, at 2.

231. Carl Schreck & Mike Eckel, *‘Just Doing His Job’: Son Recalls Life of Soviet Colonel Heralded for Averting Nuclear War*, RADIO FREE EUROPE/RADIO LIBERTY (Sept. 19, 2017), <https://www.rferl.org/a/russia-petrov-son-recalls-soviet-officer-who-averted-nuclear-war/28745011.html> [<https://perma.cc/37WD-759L>] [hereinafter Schreck & Eckel]; Allan Little, *‘How I Stopped Nuclear War’*, BBC NEWS (Oct. 21, 1998), <http://news.bbc.co.uk/2/hi/europe/198173.stm> [<http://news.bbc.co.uk/2/hi/europe/198173.stm>] [hereinafter Little].

assignment in an effort to cover up the machine's mistake, and left the army early.<sup>232</sup>

## B. Alternative solutions to holding parties responsible are flawed

This Section analyzes different suggested solutions to closing the accountability gap presented by SAWs: civil liability, otherwise known as "war torts"; administrative accountability; and the military justice system.

### 1. War Torts

Civil regimes have a much lower intent threshold, providing liability for both intentional wrongs and acts of negligence, and would seem able to provide redress to an aggrieved party.<sup>233</sup> They allow for both individual and State or organizational liability.<sup>234</sup> Unlike criminal liability, which is generally more concerned with moral wrongs and guilt, tort law focuses on injurious wrongs, fault, and regulation.<sup>235</sup>

War torts, however, have their own share of problems. Most of international law focuses on state responsibility and States are already responsible for respecting IHL.<sup>236</sup> Logistically speaking, there are few international venues that can adjudicate war torts; the ICJ, for instance, only admits causes of action initiated by States.<sup>237</sup> While States may be better positioned to pay reparations to the aggrieved party, such a remedy would not have the same individualistic deterring effects that criminal accountability from wrongful conduct would establish.<sup>238</sup> Furthermore, most

---

232. Later in life, Petrov suffered a nervous breakdown. Little, *supra* note 231; Schreck & Eckel, *supra* note 232.

233. Crootof, *supra* note 115, at 1351.

234. *Id.* at 1401.

235. *Id.* at 1353.

236. See, e.g., *State Responsibility*, INT'L COMM. RED CROSS, <https://casebook.icrc.org/glossary/state-responsibility> [<http://perma.cc/A7VV-JM9A>] (outlining state responsibility); *How the Court Works*, INT'L COURT OF JUST., <https://www.icj-cij.org/en/how-the-court-works> [<http://perma.cc/B5FW-CYJ3>] [hereinafter *How the Court Works*] (only States can be parties before the ICJ).

237. *How the Court Works*, *supra* note 236. Even if a court was available, tort liability requires victims to pursue legal action, which can be unrealistic. The victims in most of these cases would be civilians being forced to sue for relief in a foreign court while suffering from the other effects of war (poverty, displacement, etc.). LOSING HUMANITY, *supra* note 42, at 44.

238. See generally Dickinson, *supra* note 51, at 24 ("[I]nternational law could in theory permit such individual liability for torts, but most of the existing substantive

military and defense weapons manufacturing contractors in war are shielded from civil liability.<sup>239</sup>

## 2. Administrative Accountability

Administrative accountability, such as demotions, firing, reassignments, and financial penalties,<sup>240</sup> could be seen as a possible avenue of redress. While criminal liability focuses on potential prosecution, administrative inquiries are primarily focused on pure fact-finding.<sup>241</sup> This does allow for some flexibility, but also for bias—instances of administrative accountability usually occur within the same organization and are less transparent to outsiders.<sup>242</sup> Similar to war torts, administrative accountability is limited. It provides a forward-looking prospective relief without holding the individual party properly accountable for the harms

---

international law norms and international judicial or quasi-judicial venues provide for state responsibility”).

239. SCHARRE 2018, *supra* note 25, at 262; Vivian S. Chu & Kate M. Manuel, *Tort Suits Against Federal Contractors: An Overview of the Legal Issues*, CONG. RSCH. SERV. 12 (Apr. 7, 2011), <https://fas.org/sgp/crs/misc/R41755.pdf> [<https://perma.cc/2VY5-AERK>] (“Courts have crafted several ‘federal common law’ defenses that shield government contractors from liability . . . the most prominent of these are the government contractor defense and the combatant activities exception . . . some contractors have also asserted ‘derivative sovereign immunity’”). Private weapons manufacturers in particular are not generally punished for how their weapons are subsequently used, especially since most disclose any risks of malfunctions prior to purchase. LOSING HUMANITY, *supra* note 42, at 44. Holding weapons manufacturers liable would create a great disincentive to produce weapons, which are by nature inherently dangerous. Beard, *supra* note 4, at 647–648. A counterpoint could be seen through the individual criminal responsibility imposed on corporate executives in the post-World War II trials of Nazi industrialists, including the top officials for the firm that manufactured and sold Zyklon B for use in the Nazi gas chambers being convicted by a British military court as accessories to war crimes. Trial of Bruno Tesch and Two Others (The Zyklon B Case), 1 L. REPS. OF TRIALS OF WAR CRIM. 93, 93 (Brit. Mil. Ct., Hamburg, Mar. 1-8, 1946); Beard, *supra* note 4, at 648. The Nuremberg Charter allowed for this possibility through imposing individual responsibility on “accomplices participating in the formulation or execution of a common plan or conspiracy to commit” a crime enumerated within the Charter. Nuremberg Charter, Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis, and Charter of the International Military Tribunal, Charter II, art. 6, Aug. 8, 1945, 82 U.N.T.S. 280; Beard, *supra* note 4, at 649.

240. Dickinson, *supra* note 51, at 27.

241. Schmitt May, *supra* note 185.

242. See, e.g., *id.* (stating that family members of victims of the Dhahran attack during the Gulf War were unable to get information about the attack. Many outside organizations were dissatisfied with the results of the investigation and called for an in-depth, independent investigation separate from the military).

already caused,<sup>243</sup> thus likely failing to provide the proper sense of justice and accountability.

### 3. Military Justice

In the United States, the military operates its own separate justice system distinct from the civilian criminal system for members of the military.<sup>244</sup> A reportable incident within the Department of Defense includes possible, suspected, and alleged violations of the law of war, for which there is credible information, or conduct during military operations that would otherwise constitute a violation of the law of war if it had occurred during an armed conflict.<sup>245</sup> Within the military, a commander can pursue criminal punishment or administrative punishment under the UCMJ.<sup>246</sup> This parallel system falls prey to the same weaknesses as the civilian system, however: a lack of ability to establish the appropriate level of mens rea for the criminal system with respect to AWSs and a lack of suitable alternatives otherwise.<sup>247</sup>

#### C. Exploring the Proposed Dichotomy: Possible Accountability and Potential Military Value

Considering these problems, this Note suggests that the analysis of AWSs should be structured as a dichotomous balance between different interests — namely, the need to establish accountability and responsibility with the military value these machines have. Akin to the analysis inherent in the principle of proportionality,<sup>248</sup> the chance of the weapons system malfunctioning should be compared to the possibility of assigning

243. Dickinson, *supra* note 51, at 45. In regards to the airstrike on the MSF hospital discussed in the introduction, the punishments allotted were “unlikely to satisfy” *Medicins sans Frontières* and many called for criminal punishment. *See supra* P.II.C.2.; Aisch, Keller & Peçanha, *supra* note 15.

244. CONG. RSCH. SERV., MILITARY COURTS-MARTIAL UNDER THE MILITARY JUSTICE ACT OF 2016 3 (Aug. 28, 2020), <https://fas.org/sgp/crs/natsec/R46503.pdf> [<https://perma.cc/6C96-A8ZS>].

245. OFF. OF THE GEN. COUNS. OF THE DEP’T OF DEF., DOD DIRECTIVE 2311.01E, DOD LAW OF WAR PROGRAM 15 (2006), <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/231101p.pdf?ver=2020-07-02-143157-007> [<https://perma.cc/3T7B-ZMNG>].

246. In most instances, administrative punishment is the result. In a normal course of action, most incidents begin with an inquiry by the commander or an administrative investigation. LT. GEN. JEFFREY A. ROCKWELL ET. AL, *THE MILITARY COMMANDER AND THE LAW* 133 (Major Micah C. McMillan et al. eds., 15th ed. 2019).

247. See Aiden Warren & Alek Hillas, *Lethal Autonomous Weapons Systems: Adapting to the Future Unmanned Warfare and Unaccountable Robots*, 12 *YALE J. INT’L AFF.* 71, 71–73 (2017).

248. *See supra* P.I.B.2.

responsibility for these errors. As these weapons systems become more autonomous, and thus more complicated, the possibility of assigning responsibility shrinks.<sup>249</sup>

Instead of weighing a specific attack like in the proportionality analysis,<sup>250</sup> the specific weapons system's use should be weighed. The circumstances for which it is designed and in which it will be used (as well as the disparity between the two) should be especially considered. As the assignment of responsibility becomes more difficult, in accordance with the proportionality dichotomy, there will be a limit reached. At some point along the autonomy spectrum, the AWS will have created too great of a risk of civilian harm or mistaken fire such that it will outweigh, as a general matter, the military benefit of using it. The question becomes then, at what point is this limit reached, and what happens if it cannot be determined.

#### D. Applying the Dichotomy to AWS: Reaching the Point of No Return

Turning then to an application of this dichotomy, if a finite point on the spectrum in which the decreasing possibility of assigning responsibility outweighs the likelihood of the machine functioning properly cannot be determined, then the exploration into and use of AWSs needs to be stopped until that point of limitation can be adequately and clearly defined. The difficulties with determining that precise point, as may already be clear, are immense.<sup>251</sup>

With machines, the responsibility requirement becomes more diffuse as the chain of accountability along the creation, implementation, and use of the machine becomes more complex.<sup>252</sup> The problem does not necessarily arise because a human is more inclined to make mistakes than the AWSs or vice versa, but rather because one has greater accountability than the other.<sup>253</sup> While much of the literature focuses on the smaller error rate that use of SAWs may have as compared to pure human operators,<sup>254</sup> the important issue, and reason to be skeptical of increasing reliance on these

---

249. Beard, *supra* note 4, at 649–650.

250. Jean-Marie Jenckaerts & Louise Doswald-Beck, Customary International Humanitarian Law Volume I: Rules 50 (Int'l Comm. of the Red Cross 3rd ed. 2009).

251. See Anderson & Waxman, *supra* note 44, at 16 (arguing that “agreeing on what constitutes a prohibited autonomous weapon will be unattainable”).

252. See *supra* P.II.A.1.

253. See *supra* P.II.B.

254. See, e.g., Scharre 2016, *supra* note 123 (arguing AWSs are less likely to make mistakes than humans).

machines, is the decreasing availability of accountability unrelated to error rate.

Not only is there an issue with decreasing individual responsibility when things go wrong as a result of increasing autonomy, but also it becomes increasingly difficult to assign metrics for what exactly is the mens rea to be found along the process, who should have the mens rea (let alone who actually does have it), and how to establish the mens rea even if it could be defined and found.<sup>255</sup> As seen through the AC-130 example at MSF, with all different groups and individuals making errors and mistakes, thus having actus reus and perhaps mens rea at least to some degree, how can one attribute a sufficient actus reus or mens rea to any of them?

#### CONCLUSION

The current development of AWSs shows no signs of slowing, and as these weapons systems become more prolific and more complex, the question of how they will integrate into our legal framework becomes ever more pressing. Problems addressing the accountability gap posed by AWSs must be resolved before the implementation of the systems on the battlefield to assure the existence of an effective mechanism for protecting human rights. Using the dichotomic perspective in an analysis similar to the principle of proportionality will allow for a proper balancing of the need for attributable accountability with the potential military advantage.

Using this method shows that there is a point at which military advantage is outweighed by accountability problems. As new AWSs are developed at an increasingly rapid pace, they risk outstripping the ability of programmers to ensure their safety and compliance with IHL obligations.<sup>256</sup> Because the already-existing problems of enforcing proper accountability due to the layered, vertical military decision-making process<sup>257</sup> are exacerbated by the long and complex horizontal chain of production,<sup>258</sup> the

---

255. Many countries that have stated they would not develop FAWs have likewise failed to clarify a SAWs limit. For example, the U.K. Minister for Counter Proliferation stated that the “United Kingdom does not possess fully autonomous weapon systems and has no intention of developing them . . . the operation of our weapons will always be under human control as an absolute guarantee of human oversight and authority and of accountability for weapons usage.” HC Deb (17 June 2013) (564) col. 734 (UK) (<https://hansard.parliament.uk/Commons/2013-06-17/debates/1306174400002/LethalAutonomousRobotics?highlight=lethal%20autonomous%20weapons#content-13061744000214>) [<https://perma.cc/HK86-NM6K>].

256. LIN ET. AL, *supra* note 134, at 40; Beard, *supra* note 4, at 673.

257. See *supra* Part II.A.

258. See *supra id.*

implementation and use of the AWSs, even currently existing SAWs used on the battlefield, causes a lack of enforcement and punishment. As such, until either AWSs can be redesigned to better fit into our legal system or our legal system can be redesigned to account for the problems with enforcement when AWSs are used, these weapons systems must be removed from usage and development of FAWs must be ceased.