

## ARTICLE

# THE DUTY TO TAKE PRECAUTIONS IN HOSTILITIES, AND THE DISOBEYING OF ORDERS: SHOULD ROBOTS REFUSE?

*Francis Grimal and Michael J. Pollard\**

“In war the first principle is to disobey orders. Any fool can obey an order.”<sup>1</sup>

### ABSTRACT

*This Article not only questions whether an embodied artificial intelligence (“EAI”) could give an order to a human combatant, but controversially, examines whether it should also refuse one. A future EAI may be capable of refusing to follow an order, for example, where an order appeared to be manifestly unlawful, was otherwise in breach of International Humanitarian Law (“IHL”), national Rules of Engagement (“ROE”) or, even, where they appeared to be immoral or unethical. Such an argument has traction in the strategic realm in terms of “system of systems”—the premise that more advanced technology can*

---

\* Francis Grimal is a Reader in Public International Law, University of Buckingham, UK, and Michael J. Pollard is a PhD Candidate in Public International Law, University of Buckingham, UK. The Authors would like to extend their sincerest thanks to Professor Christopher Waters, Dean of Law, University of Windsor, Ontario for all his considerable advice and invaluable feedback throughout the preparation of this Article. The Authors would also like to acknowledge the substantial work undertaken by the ICRC with regards to customary international law in general. Further, and very sincerest thanks too, to Casey Morin, Senior Articles Editor at FILJ for all the considerable help and patience shown during the submission process of this Article. And our deepest gratitude to the Editorial Board, to Michael Campbell Jr (Editor-in-Chief), to both Kelsey Halloran (Executive Notes & Articles Editor) & Noah Parson (Managing Editor), and to all at FILJ for working with us on this Article—it is a great honor and privilege to feature in FILJ.

1. Michael Symanski, *Any Fool Can Obey an Order*, MODERN WAR INST. (Mar. 17, 2017), <https://mwi.usma.edu/fool-can-obey-order/> (“so wrote First Sea Lord Jacky Fisher in angry Critique of Captain H.M. Pelly, a cruiser Captain under Admiral Beatty at the battle of Dogger Bank in 1915”); see also Ryan Scott, *Willful Disobedience: Character traits of Independent Thinkers in the Military*, MODERN WAR INST. (Feb. 23, 2017), <https://mwi.usma.edu/willful-disobedience-character-traits-independent-thinkers-military/> [<https://perma.cc/9MBP-268G>].

*potentially help overcome Clausewitzian “friction” or “fog of war.” An aircraft’s anti-stall mechanism, which takes over, and corrects human error, is seen as nothing less than “positive.”*

*As part of opening this much-needed discussion, the Authors examine the legal parameters, and by way of a solution provide a framework for overriding and disobeying. Central to this discussion, are state specific ROEs within the concept of “duty to take precautions.” At present, the guidelines relating to a human combatant’s right to disobey orders are contained within such doctrine, but vary widely. For example, in the United States, a soldier may disobey an order but only when the act in question is clearly unlawful. In direct contrast, however, Germany’s “state practice” requires orders to be compatible with the much wider concept of human dignity, and to be of “use for service.”*

*By way of a solution, the Authors propose the crafting of a test referred to as “robot rules of engagement” (“RROE”) with specific regard to the disobeying of orders. These RROE ensure (via a multi-stage verification process) that an EAI can discount human “traits” and minimize errors that lead to breaches of IHL. In the broader sense, the Authors question whether warfare should remain an utterly human preserve—where human error is an unintended but unfortunate consequence—or, whether the duty to take all feasible precautions in attack in fact require a human commander to utilize available AI systems to routinely question human decision-making, and where applicable, prevent mistakes. In short, the Article examines whether human error can be corrected and overridden, but for the better, rather than for the worse.*

ABSTRACT .....	671
I. INTRODUCTION .....	673
II. DUTY TO TAKE PRECAUTIONS DURING ARMED	
CONFLICT .....	677
A. How Does the Duty Take Precautions Fit Within	
IHL? .....	678
B. State Practice .....	686
III. RULES OF ENGAGEMENT AND DISOBEYING OF	
ORDERS .....	694
A. How does the Duty to Disobey an Order fit within	
DTP? .....	694
B. Disobeying of Orders .....	697
IV. DISOBEDIENCE AND NUCLEAR LAUNCH .....	708

A. Additional Test for Safeguarding Decision-making.	
.....	709
B. Scenario 1 .....	716
C. Scenario 2 .....	717
D. Scenario 3 .....	720
E. Scenario 4 .....	723
V. WIDER IMPLICATIONS OF ROBOT REFUSAL.....	729
VI. CONCLUSION.....	730

### I. INTRODUCTION

This Article goes significantly beyond traditional boundaries of debate surrounding a combatant’s legal obligation to disobey unlawful orders,<sup>2</sup> and by way of overall solution, presents both a legal and strategic “system of systems.” Grounded at the very heart or “Schwerpunkt” (to use Clausewitz’s terminology)<sup>3</sup> of strategic thinking is the concept of “friction” within warfare.<sup>4</sup> Friction, simply-put is more colloquially known as “the fog of war”—the concept that within warfare, things will go wrong, and mistakes will happen.<sup>5</sup> In essence, the Clausewitzian concept of

---

2. Indeed, to date, the primary focus in this area has been upon the acceptability (or not) of the “Nuremberg defense,” in respect of the commission of war crimes. See James B. Insko, *Defense of Superior Orders before Military Commissions*, 13 DUKE J. COMP. & INT’L L. 389 (2003); Mark J. Osiel, *Obeying Orders: Atrocity, Military Discipline, and the Law of War*, 86 CALIF. L. REV. 939 (1998); Robert E. Murdough, *I Won’t Participate in an Illegal War: Military Objectors, the Nuremberg Defense, and the Obligation to Refuse Illegal Orders*, 2010 ARMY L. 4 (2010).

3. Clausewitz refers to “Schwerpunkt” a number of times in the original German language edition of his *magnus opus*, *Vom Kreige* (On War). The authoritative English language edition of On War translates Schwerpunkt to mean “centers of gravity,” the notion that in warfare one can identify a “hub of all power and movement, on which everything depends . . . the point against which all . . . energies should be directed.” See CARL VON CLAUSEWITZ, ON WAR 720 (Michael Howard & Peter Paret eds., 1976, rev. ed. 1984); see also COLIN S. GRAY, MODERN STRATEGY, 75-112 (1999); see, e.g., John E. Tashjean, *Pious Arms: Clausewitz and the Right of War*, 44 MIL. AFF. 82, 83 (1980) (providing general works on Clausewitz); see generally CARL VON CLAUSEWITZ, CLAUSEWITZ ON STRATEGY: INSPIRATION AND INSIGHT FROM A MASTER STRATEGIST (Christopher Bassford et al. eds., 2001).

4. CLAUSEWITZ, ON WAR, *supra* note 3, at 138-40 (at page 138 Clausewitz notes that “Friction is the only concept that more or less corresponds to the factors that distinguish real war from war on paper.”).

5. Strategic literature is replete with examples. See, e.g., COLIN S. GRAY, AIRPOWER FOR STRATEGIC EFFECT 214 (2012). Here, for example, the author discusses sorties being

“friction,” is what distinguishes “real war” from war on paper, something (despite advances of technology) for which a strategist cannot always account.<sup>6</sup> Strategic doctrines that failed to account for “friction” were deemed as inherently flawed,<sup>7</sup> and considerable ink was spilt in the 1990s and 2000s to consider the means by which to overcome friction in warfare.<sup>8</sup> Fully-conscious that previous efforts to simply “throw” advanced technology at the problem would not resolve the issue,<sup>9</sup> strategists formulated the idea of “system of systems.”<sup>10</sup> Much as with a mechanical watch, technology/cogs would be assigned individual tasks, but those smaller “cogs” would then form part of a bigger and more advanced “system of systems.”<sup>11</sup>

As presented above, the Authors propose to transfer this strategic concept into the legal realm as a means of providing an overall solution. By doing so, they are able to consider how, and more importantly when, various “*individuals*” within the chain of command (*e.g., human to Autonomous Weapon System (“AWS”); AWS*

---

abandoned during the Kosovo conflict due to foggy conditions near the bases in Terano, Italy; the negative impact of bad weather (in the form of rain) in Vietnam; and even how relatively modern operations in Afghanistan and Iraq were negatively affected.

6. *Id.* at 44-45.

7. *See, e.g.,* BARRY D. WATTS, *THE FOUNDATIONS OF U.S. AIR DOCTRINE: THE PROBLEM OF FRICTION IN WAR 47-54* (2012). Watts discusses the doctrine of Strategic Air Power (“SAP”), a relatively simple notion that airpower alone is sufficient (in the military sense) to bring about victory. According to Watts, SAP failed to live up to its promise due to lack of “frictional” considerations. Despite technological advances, rain, fog, and the climate as a whole prevented SAP from operating in certain instances.

8. *Id.* Furthermore, as noted and reinforced by Watts, failure to account for friction prevents *any* strategic doctrine from fulfilling its potential. However, as with all doctrines, there is not just one overriding factor (although the omission of “frictional considerations”) is probably the most significant.

9. *Id.*

10. *See* GRAY, *supra* note 5, at 106.

11. In short, a system-of-systems is a collection of independent systems/elements that are each allotted a specific task. They each complete this task without necessarily having to consult other systems/ elements—in a vehicle, this could, for example, be power steering. However, sometimes a group of such systems can be considered holistically, with the result of the overarching system being much more advanced, and capable, than any of the individual elements could be alone—a self-driving car, for example, would need many individual components, perhaps including GPS, radar, lidar, and other sensors, automatic braking, lane recognition, self-parking, cameras, and, not least, a central processor. None of these systems alone, however, could ever ensure a vehicle getting safely from A to B. *See, e.g., Automated Vehicles for Safety*, NAT’L HIGHWAY TRAFFIC SAFETY ADMIN. (NHSTA), <https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety> [<https://perma.cc/8PKB-GY7W>] (last visited Feb. 18, 2021).

to human; EAI to AWS, human to EAI) would be able to correctly reject and overturn an erroneous and unlawful order.<sup>12</sup> Throughout, the Authors argue that simply transposing a combatant's right and duty to disobey an unlawful order into the EAI realm without further scrutiny is unwise, for it would close the door on a much-needed discussion in terms of how EAIs could potentially remove human error or "friction" in warfare.<sup>13</sup>

In terms of a unique and novel solution, the Authors envisage the following: the EAI (when considering *any* order) will not only run a continuous/dynamic proportionality assessment (an implicit factor of which is a perpetual target verification mechanism), but will also identify whether the human decision is motivated by human character "flaws" such as revenge or self-preservation. Where the EAI is satisfied that the order is IHL compliant but that human error is in evidence, the EAI should reject that human command. While there is undoubtedly a concern that EAIs might increase "friction" (due to their behavior) the Authors counter such a possibility by implementing the following: when it comes to a human to EAI order in the chain of command the Authors suggest several cogs and levels within this "system of systems" with each one independently making a specific determination. Consequently, each independent system would be tasked with only one computation. This is set this out comprehensively, and considered alongside a number of scenarios in Part IV. Here, the Authors consider the system-of-systems in its most extreme setting—"robot refusal" of nuclear launch.

An additional, but natural (and indeed desirable) implication of this discussion is whether the threshold is raised—the extra layers of "protection" proposed by the Authors imply that IHL obligations are set much higher. One could readily argue that an EAI's calculations place it under a higher burden in terms of assessment and accountability of information when

---

12. As the analysis that follows demonstrates, there is presently some deviation in state practice with respect to exactly which orders a soldier should disobey.

13. Note robot refusal in this realm is a controversial discussion, not least because some commentators hold that in order for a soldier to disobey an unlawful order (and therefore to any machine equivalent) "they must exercise 'agency' and engage in nuanced reasoning." See Tetyana Krupiy, *Unravelling Power Dynamics in Organizations: An Accountability Framework for Crimes Triggered by Lethal Autonomous Weapons Systems*, 15 LOY. U. CHI. INT'L L. REV. 1, 15 (2017).

taking a particular decision. Ostensibly, the Authors' analysis is distilled into three distinct areas: first, the understanding and scrutiny of existing guidelines regarding the duty to take precautions: wide (United States); narrow (Germany). Second, the way in which the guidelines are interpreted with regard to the right to refusal: wide (Germany); narrow (United States). Finally, the Authors propose a solution to the existing variations in application by introducing an additional test. This provides a method for determining precisely how an EAI may refuse erroneous human command. The Authors believe this discussion is necessary due to the fact that as long as there is human involvement, there will always be human error.

While one could argue that there are more things that could “go wrong” (due to technological limitations of 21st century technology), to suggest that this will remain the *status quo*, is dismissive of advances in machine learning.<sup>14</sup> While the absolute and complete removal of error is difficult to envisage, one could analogize such a step with an anti-stall mechanism in an aircraft—a simple and existing technology which innocuously overrides human error for the “better.”<sup>15</sup> In other words, while the Authors' proposed “system of systems” regarding refusal may not

---

14. See George Anadiotis, *The state of AI in 2019: Breakthroughs in machine learning, natural language processing, games, and knowledge graphs*, ZDNET (July 8, 2019), <https://www.zdnet.com/article/the-state-of-ai-in-2019-breakthroughs-in-machine-learning-natural-language-processing-games-and-knowledge-graphs/> [<https://perma.cc/V82E-3AV9>] (providing a useful and relaxed discussion about recent breakthroughs in machine learning and possible future developmental trajectories).

15. A stall is defined as “a sudden reduction in the lift generated by an aerofoil when the critical angle of attack is reached or exceeded.” *Stall*, SKYBRARY, <https://www.skybrary.aero/index.php/Stall> [<https://perma.cc/NL3P-BAJF>] (last visited Jan. 27, 2021). Anti-stall systems automatically reduce the pitch of the aircraft nose where, for example, the angle of attack is exceeded due to pilot error. They are typically considered so effective that they are a requirement on all transport aircrafts. *Id.* However, the system is not a complete failsafe, and following a spate of incidents in 2019, a malfunctioning anti-stall was responsible for grounding Boeing's entire fleet of 737 Max aircraft. See Gwyn Topham, *Anti-stall system was 'in play' on Ethiopian's Boeing 737 Max*, GUARDIAN (Mar. 25, 2019), <https://www.theguardian.com/world/2019/mar/25/anti-stall-system-was-in-play-on-ethiopians-boeing-737-max> [<https://perma.cc/9XA2-EKDD>]; David Shepardson & Jamie Freed, *FAA failed to properly review 737 MAX jet's anti-stall system - JATR findings*, *Business News*, REUTERS (Oct. 11, 2019), <https://uk.reuters.com/article/uk-usa-boeing-airplane-faa/faa-failed-to-properly-review-737-max-jets-anti-stall-system-jatr-findings-idUKKBN1WQ0PS> [<https://perma.cc/C9TQ-Y49E>].

fully eradicate “friction,” its negative effects will have been significantly reduced and constrained.<sup>16</sup>

The heart of this Article is the proposition that future EAI will possess sufficient decision-making capabilities in order for them to be considered, *prima facie*, “conscious.”<sup>17</sup> By way of overview, the Article is structured as follows: Part II of the Article examines the way in which the Duty to Take Precautions operates as a whole within the IHL realm. Specific focus is placed upon state practice so as to make the necessary link between the Duty to Take Precautions and the disobeying of orders. Part III examines the specific interface between the legal obligations contained in IHL, and their translation into state-based ROEs. Part IV introduces the Authors’ “stress test,” and the logistics of “robot refusal” in terms of providing a concrete legal test for determining when an erroneous command (whatever its designation: human, autonomous or EAI) should be rejected and refused. Finally, Part V considers some of the wider implications of EAI refusal, specifically with regard to robot spies and robot private military contractors.

## II. DUTY TO TAKE PRECAUTIONS DURING ARMED CONFLICT

Prevailing discussion within the existing scholarship regarding AWS and EAI systems falling under the scope of IHL has understandably centered on whether such systems can satisfy the necessary distinction and proportionality requirements.<sup>18</sup>

---

16. Existing military systems including satellite and thermal imaging, 3D mapping, and real-time battlefield updates received from various robots located in and around the battlespace already operate at speeds far in excess of human capabilities. However, such technologies *will* inevitably be surpassed, and overtaken in the not too distant future.

17. Indeed, the Authors acknowledge that this examination of robot refusal sits alongside a much wider philosophical, ethical and legal debate regarding robot sentience and the concept of robot refusal. Should a “sex”-robot, for example, be entitled to say no, and, should there be consequences for it, and/or for the “client” who chooses to ignore such an instruction? See Robert Sparrow, *Robots, Rape, and Representation*, 9 INT. J. SOC. ROBOTICS 465, 465-77 (2017); Neda Atanasoski & Kalindi Vora, *Why the Sex Robot Becomes the Killer Robot – Reproduction, Care, and the Limits of Refusal*, SPHERES J. 1, 3 (2020); Lily Frank & Sven Nyholm, *Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable?*, 25 ARTIFICIAL INTELLIGENCE & L. 305, 305-23 (2017).

18. See, e.g., Neil Davidson, *A legal perspective: Autonomous weapon systems under international humanitarian law*, in UNODA OCCASIONAL PAPERS NO. 30: PERSPECTIVES ON LETHAL AUTONOMOUS WEAPON SYSTEMS 5, 7 (2017); Marco Sassoli, *Autonomous Weapons*

Equally, the general corpus of literature regarding the duty to take precautions during armed conflict is also relatively well-established.<sup>19</sup> Consequently, the Authors navigate the key contours of this landscape without over elaboration. Rather, the purpose and focus of Part II is to identify how the duty to take precautions (“DTP”) fits within the larger corpus of IHL. At its core, the following section confirms the Authors’ firm assertion that future commanders will have to use EAs as part of their DTP, and ultimately, to “respect” their decision *not* to follow orders.

*A. How Does the Duty Take Precautions Fit Within IHL?*

The purpose of this Section is twofold. First, the Authors identify a combatant’s lawful obligation to take precautions in attack that is contained within Article 57 of Additional Protocol I (“API”).<sup>20</sup> Secondly, the Authors demonstrate how the *customary* duty to disobey a manifestly unlawful order is applicable to increasingly autonomous technologies such as EAs.

IHL strives to “achieve” the thankless task of reconciling two fundamental values operating at opposing ends of a “spectrum.”<sup>21</sup> These are: 1) the protection of civilian life and civilian objects; and 2) the concept of military necessity.<sup>22</sup> Considerable attention

---

*and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT’L L. STUD. 308, 338-40 (2014); Chantal Grut, *The Challenge of Autonomous Lethal Robotics to International Humanitarian Law*, 18 J. CONFLICT & SEC. L. 5, 5 (2013); Michael N. Schmitt, *Autonomous weapon systems and international humanitarian law: a reply to the critics*, 4 HARV. NAT’L SEC. J. 1, 8 (2013).

19. See, e.g., Thurnher J.S., *Feasible Precautions in Attack and Autonomous Weapons*, in DEHUMANIZATION OF WARFARE (Heintschel von Heinegg et al. eds., 2018); see Michael N. Schmitt, *Precision attack and international humanitarian law*, 87 INT’L REV. RED CROSS 445 (2005); Geoffrey Corn & James A. Schoettler, *Targeting and Civilian Risk Mitigation: The Essential Role of Precautionary Measures*, 223 MIL. L. REV. 785 (2015).

20. Protocol (I) Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, International Committee of the Red Cross, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter API].

21. The ICRC identifies that “International humanitarian law is a set of rules which seek, for humanitarian reasons, to limit the effects of armed conflict.” See *What is International Humanitarian Law?*, ICRC, [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/MQ5U-NVUL] (last visited Jan. 27, 2021). However, while IHL clearly does provide a number of obvious protections for both civilians and combatants, the point is it does not prevent or outlaw war.

22. According to the International Committee of the Red Cross (“ICRC”): “The ‘principle of military necessity’ permits measures which are actually necessary to accomplish a legitimate military purpose and are not otherwise prohibited by



has been dedicated to the “achievement” (or not) of this precarious balance, and the Authors do not seek to further that particular discussion.<sup>23</sup> Nevertheless, the legal barometers of distinction and proportionality act as safeguards to ensure that reconciliation between the two fundamental values takes place in practice. The first of these two principles is captured by Article 48 API which provides the “basic rule.”<sup>24</sup> This ensures that civilians—and civilian objects—are distinguished from military objectives. “Distinction” is clearly intended to reflect the protectionist ambit of the IHL balancing act. Once the distinction has been made, Article 51 (2) API ensures that civilians and civilian objects must not be made the object of a direct attack,<sup>25</sup> while also prohibiting threats of violence where their use is intended to spread terror.<sup>26</sup>

Article 57 API refers specifically to precautions in attack. Article 57(1) states, “[i]n the conduct of military operations, constant care shall be taken to spare the civilian population,

---

international humanitarian law. In the case of an armed conflict the only legitimate military purpose is to weaken the military capacity of the other parties to the conflict.” *Military Necessity*, ICRC, <https://casebook.icrc.org/glossary/military-necessity> [https://perma.cc/89DZ-J3CQ] (last visited Jan. 27, 2021).

23. See YORAM DINSTEIN, *THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT* 8-12 (3d ed. 2016) (the author identifies the need to balance these two “driving forces” of IHL); Nobuo Hayashi, *Requirements of Military Necessity in International Humanitarian Law and International Criminal Law*, 28 B.U. INT’L L.J. 39, 41 (2010); see generally Yishai Beer, *Humanity Considerations Cannot Reduce War’s Hazards Alone: Revitalizing the Concept of Military Necessity*, 26 EUR. J. INT’L L. 801, 801 (2015).

24. API, *supra* note 20, art. 48, at 36 (“In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives.”).

25. *Id.* art. 52, at 38. See also *id.* art. 49, at 36 (“‘Attacks’ means acts of violence against an adversary, whether in offence or defence.”).

26. The full text Article 51 (2) API reads: “The civilian population as such, as well as individual civilians, shall not be the object of attack. Acts or threats of violence the primary purpose of which is to spread terror among the civilian population are prohibited.” *Id.* art. 51(2). Note also that Article 51(3) API states: “Civilians shall enjoy the protection offered by this section, unless and for such time as they take a direct part in hostilities” (“DPH”). *Id.* art. 51(3), at 37. Exactly which actions should qualify as playing a DPH is the subject of some conjecture. For an in-depth analysis of the two leading schools of thought, see Francis Grimal & Michael J. Pollard, “*Embodied AI*” and *The Direct Participation in Hostilities*, 51 GEO. J. INT’L L. 513 (2020) (the Authors’ previous piece discussing EAI and DPH).

civilians and civilian objects.”<sup>27</sup> And, Article 57(2) provides further detail supplying that:

. . . a) those who plan or decide upon an attack shall: (i) do everything feasible to verify that the objectives to be attacked are neither civilians nor civilian objects . . . [and] . . . (ii) take all feasible precautions in the choice of means and method of attack with a view to avoiding, and in any event minimizing, incidental loss of life, injury to civilians and damage to civilian objects.<sup>28</sup>

Although it is not referred to explicitly, the principle of proportionality is contained within Article 57(2)(a)(iii) API. This ensures that a combatant must refrain from launching an attack where the damage caused to civilians or civilian objects “would be excessive in relation to the concrete and direct military advantage anticipated.”<sup>29</sup> This provision acknowledges the concept of military necessity and, in doing so, that “situations may arise where civilians simply cannot be spared.”<sup>30</sup> Article 8(2)(b)(iv) of the Rome Statute identifies that carrying out a disproportionate attack is a war crime.<sup>31</sup> According to the International Committee of the Red Cross (“ICRC”), proportionality is customary in

---

27. API, *supra* note 20, art. 57(1), at 41 (This is known as the “constant care” obligation, to which more attention is given in the discussion which follows.).

28. API, *supra* note 20, art. 57(2)(a)(i)-(ii).

29. Though note this provision does not refer to proportionality directly. Nevertheless, as the ICJ identified in the *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 I.C.J. 226 (July 8) (Higgins, J., dissenting) [hereinafter *Nuclear Weapons, Advisory Opinion*], API ensures that, “a legitimate target may not be attacked if the collateral civilian casualties would be disproportionate to the specific military gain . . . .” See also DINSTEIN, *supra*, note 23, at 152, 158 (noting that “military advantage and civilian casualties/damage are incomparable in a quantifiable manner . . . [and that] . . . [t]here is little prospect of agreement between opposing Belligerent Parties as to the rival values of . . . [these concepts].”).

30. DINSTEIN, *supra* note 23, at 164 (citing FRITS KALSHOVEN, *REFLECTIONS ON THE LAW OF WAR: COLLECTED ESSAYS* 346, 546 (2007)).

31. Article 8(2)(b)(iv) Rome Statute of the International Criminal Court (1998) states: “For the purpose of this statute, ‘war crimes’ means: Intentionally launching an attack in the knowledge that such attack will cause incidental loss of life or injury to civilians or damage to civilian objects or widespread, long-term and severe damage to the natural environment which would be clearly excessive in relation to the concrete and direct overall military advantage anticipated.”

nature,<sup>32</sup> as is the separate duty to take precautions in attack.<sup>33</sup> These obligations are therefore placed upon all states regardless of whether or not they are party to API—which enables due consideration of US practice.<sup>34</sup>

Article 35(1) API is relevant to DTP, stating, “. . . the right of the Parties to choose methods or means of warfare is not unlimited.”<sup>35</sup> Here, the term “*means*” refers to the choice of weapon—obligating the attacker to select “weapons and munitions,” that prevent “overkill.”<sup>36</sup> “[*M*]ethods” of attack include, “angles of attack,” “time on target,” and “similar tactical choices.”<sup>37</sup> If an attacker has a choice of means and method to achieve a legitimate military objective, they “should” choose the one that “is likely to cause the least collateral damage or incidental injury.”<sup>38</sup>

32. See JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, INTERNATIONAL COMMITTEE OF THE RED CROSS: CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, VOL. I: RULES (2005) [hereinafter ICRC Customary Rules] (ICRC Customary Rule 14, states: “[I]aunching 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, is prohibited . . . State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts.”).

33. See *id.* (“[i]n the conduct of military operations, constant care must be taken to spare the civilian population, civilians and civilian objects. All feasible precautions must be taken to avoid, and in any event to minimize, incidental loss of civilian life, injury to civilians and damage to civilian objects . . . State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts.”)

34. The United States is not currently a party to API.

35. See Regulation 22 annexed to the Hague Convention on Laws and Customs of War on Land (IV), Oct. 18, 1907 (“The right of belligerents to adopt means of injuring the enemy is not unlimited”); *Nuclear Weapons, Advisory Opinion*, *supra* note 29, at 10-11 (identifying that nations “do not have unlimited freedom of choice of means of weapons.”); see generally DINSTEIN, *supra* note 23; see generally STUART CASEY-MASLEN & STEVEN HAINES, HAGUE LAW INTERPRETED: THE CONDUCT OF HOSTILITIES UNDER THE LAW OF ARMED CONFLICT (2018) (the authors, for example, argue that the ICJ “strangely,” and incorrectly, limited the principle to weapons that caused unnecessary suffering to combatants).

36. See DINSTEIN, *supra* note 23, at 164.

37. See *id.*

38. Michael N. Schmitt, *The Principle of Discrimination in 21 Century Warfare*, 2 YALE HUM. RTS. & DEV. L.J. 143, 152 (1999); see ICRC Customary Rules, *supra* note 32 (ICRC Customary Rule 17 restates art. 57 (2)(a)(ii): “Each party to the conflict must take all feasible precautions in the choice of means and methods of warfare with a view to avoiding, and in any event to minimizing, incidental loss of civilian life, injury to civilians and damage to civilian objects.”).

This final statement is *central* to the discussion relating to the concept of robot refusal. And, it is the cord that binds the present discussion with that in Part III. As the Authors have argued elsewhere, the introduction of increasingly advanced AI technology into warfare is inevitable.<sup>39</sup> For example, an EAI could be an advanced theatre ballistic missile that was capable of conducting an additional distinction and proportionality assessment before striking its target. If the “on-board” tech identified, for example, that the intended target was positioned next to a hospital, it could choose to withdraw or self-destruct before carrying out the attack. However, an EAI might equally be an unarmed reconnaissance platform, or even a (unarmed) humanoid member of a special ops group (for the sake of the present article an EAI is distinguished from an AWS by way of the fact that while an EAI can refuse to follow an order to apply force, an AWS (once activated) can make targeting decisions and chose to apply force independently of human supervision and/or instruction). Regardless, such EAI systems will be capable of monitoring live feeds of battlefield conditions outside of the immediate area, and as a result, determine the “relative” value of a target.<sup>40</sup> These systems would then direct human decision-makers accordingly. Furthermore, an EAI, or AI software could be utilized in command and control to support high-end operational decision-making regarding the deployment and movement of strategic assets. Systems such as this may encourage decision-making that changes the very operational fabric of an armed conflict.

No matter which type of AI is at the operator’s disposal, however, (leaving aside the matter of whether it offers an improvement in the means or methods of warfare), there will be occasions where a decision-maker *must* utilize it. To not do so, and to instead use an alternative means that causes greater civilian harms, would breach the DTP obligation.<sup>41</sup> This is a natural

---

39. See Grimal & Pollard, *supra* note 26, at 520.

40. See generally Sassoli, *supra* note 18, at 339.

41. It is important to note that an attacker does not have to use a particular weapon, or tactic, just because it will lessen the amount of collateral damage. A commander who is in possession of a limited number of platforms, may for example, decide that there would be more efficiently deployed elsewhere, or, with only a limited number of munitions, decide that it would be more effective not to use one in the present

continuation of the well-versed discussion of Schmitt and Widmar, who note the loitering capabilities of UAVs' (unmanned aerial vehicles) significantly enhance target verification.<sup>42</sup> Central to the current discussion, is the postulation that where tech is capable of conducting additional distinction or proportionality assessments, it must also be capable of saying "no." This is analogous to existing operations whereby the tactical team on the ground can see that the actual circumstances are manifestly different to those that were imagined when the attack was ordered.<sup>43</sup>

In all scenarios the decision regarding the most appropriate means and methods are framed by the concept of "feasibility." This means, in planning an attack, a combatant must (as a minimum) do only what is "feasible" to verify civilians or civilian objects (hereinafter "feasible varication"), and that a means or method should avoid, or at least minimize, collateral damage (hereinafter "feasible precautions"). This means the tactical team may decide to consummate the attack because of a high target value. Though equally, command and control may choose to extract the team and authorize the use of an over-the-horizon precision strike instead. What is feasible in the circumstances however, is inherently contextual.<sup>44</sup> Moreover, the term "feasible" is itself subject to various interpretations. Notably, there is no supplementary clarification or definition of the term within either API, the wider Geneva Conventions, or the Additional Protocols.<sup>45</sup> Instead, one must look to an alternate treaty to locate

---

circumstances. See Michael N. Schmitt & Eric W. Widmar, "On Target": Precision and Balance in the Contemporary Law of Targeting, 7 NAT'L SEC. L. & POL'Y 379, 400-04 (2014).

42. See *id.* at 401.

43. But see *id.* at 401 (noting that sometimes a tactical situation can change after a targeting decision has been made, but beyond the point at which an attack can be called off.).

44. Corn & Schoettler, *supra* note 19, at 802; see, e.g., Schmitt & Widmar, *supra* note 41.

45. See, e.g., Geneva Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field, Aug. 12, 1949, 75 U.N.T.S. 31 [hereinafter *Geneva Convention I*]; Geneva Convention (II) for the Amelioration of the Condition of Wounded, Sick and Shipwrecked Members of Armed Forces at Sea, Aug. 12, 1949, 75 U.N.T.S. 85 [hereinafter *Geneva Convention II*]; Geneva Convention (III) Relative to the Treatment of Prisoners of War, Aug. 12, 1949, 75 U.N.T.S. 135 [hereinafter *Geneva Convention III*]; Geneva Convention (IV) Relative to the Protection of Civilian Persons in Times of War, Aug. 12, 1949, 75 U.N.T.S. 287 [hereinafter *Geneva Convention IV*]; Protocol (I) Additional to the Geneva Conventions of 12 August 1949,

a codified definition.<sup>46</sup> A regularly cited example of this can be found in Article 3(10) of the Protocol II Annexed to the Convention on Certain Conventional Weapons (“CCCW”), which states:

[f]easible precautions are those precautions which are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations.<sup>47</sup>

This is, therefore, generally understood to be the standard by which feasibility is assessed.<sup>48</sup> And, clearly, it acknowledges the delicate balance between the need to protect the civilian population and the concept of military necessity.<sup>49</sup> Two authors note that the ICRC identifies that “practicable or practically possible entails ‘common sense and good faith,’”<sup>50</sup> and that the feasibility standard is based upon the concept of “reasonableness.”<sup>51</sup> The brief examination that follows cross-references a number of statements or codifications of feasibility within past or present military doctrines in order to identify if they offer something more tangible. Previously, US doctrine has stated

---

and Relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter *Additional Protocol I*]; Protocol (II) Additional to the Geneva Conventions of 12 August 1949 and Relating to the Protection of Victims of Non-International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 609 [hereinafter *Additional Protocol II*]; Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Adoption of an Additional Distinctive Emblem (Protocol III), Dec. 8, 2005, T.I.A.S. No. 07-908 [hereinafter *Additional Protocol III*].

46. See DINSTEIN, *supra* note 23, at 165; Schmitt & Widmar, *supra* note 41, at 400 (directing the reader’s attention to art. 3(4) CCCW); see generally ICRC Customary Rules, *supra* note 32; Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125, U.N.T.S. 3, UK Reservation; see, e.g., Article 3(10) of Amended Protocol on Prohibitions or Restrictions on the use of Mines, Booby Traps, and other Devices, May 3, 1996, 2048 U.N.T.S. 93 [hereinafter *Amended CCW Protocol II*]; ICRC Customary Rules, *supra* note 32, Customary Rule 15.

47. Protocol on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices art. 3, Oct. 10, 1980, 19 I.L.M. 1523 [hereinafter *CCCW*].

48. Schmitt & Widmar, *supra* note 41, at 400.

49. *Id.* at 402 (“attackers need only select a less harmful means or methods that do not involve sacrificing military advantage and that are feasible.”).

50. *Id.* at 400; INT’L COMM. OF THE RED CROSS, COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949 681 (Yves Sandoz et al. eds., 1987).

51. Schmitt & Widmar, *supra* note 41, at 400-01.

that its position reflects Article 3(10) CCCW.<sup>52</sup> Arguably, however, the this places greater emphasis upon military operations than it does humanitarian concerns, as it removes any reference to the latter.<sup>53</sup> UK, German, Australian and Canadian doctrine reflects Article 3(10) CCCW more precisely—the text appearing in the manuals of each being virtually identical.<sup>54</sup> Similarly, though perhaps more liberally, French doctrine notes feasibility is “that which can be realized or which is possible in practice, taking into account all circumstances ruling at the time, including humanitarian and military considerations.”<sup>55</sup>

There is slight variation between these alternative interpretations, though each example does merely refer in some way to “what is possible” in the circumstances. And, given the

---

52. See IV OFFICIAL RECORDS OF THE DIPLOMATIC CONFERENCE ON THE REAFFIRMATION AND DEVELOPMENT OF INTERNATIONAL HUMANITARIAN LAW APPLICABLE IN ARMED CONFLICTS GENEVA (1974-1977) 241 (1977); INT’L COMM. OF THE RED CROSS, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW 361 (Jean-Marie Henckaerts et al. eds., 2005) [hereinafter ICRC Practice Relating to Customary Rules]. The United States referred to “that which is practicable or practically possible, taking into account all circumstances at the time, including those relevant to the success of military operations,” rather than mirroring art. 3(4) where the final sentence reads: “including humanitarian and military considerations.” *Id.* This was very similar to the German stance before the ratification of API. See IV OFFICIAL RECORDS OF THE DIPLOMATIC CONFERENCE ON THE REAFFIRMATION AND DEVELOPMENT OF INTERNATIONAL HUMANITARIAN LAW APPLICABLE IN ARMED CONFLICTS GENEVA (1974-1977) 226 (1977) (stating that for the sake of Art. 57 API, feasible should mean: what “is practicable or practically possible, taking into account all circumstances at the time, including those relevant to the success of military operations.”).

53. See *supra* note 52 and accompanying text.

54. For the United Kingdom doctrine, see United Kingdom, Declarations and Reservations Made upon Ratification of the 1977 Additional Protocol I, 28 Jan. 1998, § b (UK); for the German doctrine, see Germany, Declarations made upon Ratification of the 1977 Additional Protocol I, 14 Feb. 1991, § 2 (Ger.); for the Australian doctrine, see LAW OF ARMED CONFLICT 3 (Defence Publ’g Serv., 2006) (the glossary states that feasible precautions are those “which are practicable or practically possible taking into account all circumstances ruling at the time including humanitarian and military considerations”); for the Canada doctrine, see Canada, Reservations Made upon Ratification of the 1977 Additional Protocol I, 20 Nov. 1990 (Can.); see also CHIEF OF THE DEFENCE STAFF, USE OF FORCE FOR CF OPERATIONS §112.6 (2008) (“‘Feasible’ is understood as that which is practicable or practicably possible, taking into account all circumstances ruling at the time, including humanitarian and military considerations.”). See also Int’l Comm. of the Red Cross, *Practice Relating to Rule 15: The Principle of Precautions in Attack*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_rul\\_rule15#top](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule15#top) [<https://perma.cc/43ZW-77MP>] [hereinafter ICRC Database Rule 15].

55. ICRC Database Rule 15, *supra* note 54 (citing Reservations and declarations made upon ratification of the of the 1977 Additional Protocol, §3 (2003)).

inherently contextual nature of warfare, it is perhaps unsurprising that these examples do little to add much in the way of extrinsic context. Consequently, when this brief examination is considered in isolation, any difference in interpretation is arguably only a matter of semantics.

In summary, this Section has demonstrated that DTP is intertwined with the IHL principles of distinction and proportionality. DTP therefore, plays a pivotal role in maintaining the critical balance between the protection of the civilian population and the concept of military necessity. There is, perhaps, a slight variation in state interpretation of the concept of feasibility. Nevertheless, in order to adhere with DTP obligations, there will undoubtedly be instances in the future where—in order to ensure that everything practicably possible has been done to verify a target, or to minimize collateral damage—a commander will have to utilize an EAI/ AWS.<sup>56</sup>

### *B. State Practice*

With reference to the six military powers that were considered in the previous Section,<sup>57</sup> the following discussion underlines state practice (with specific regard to the interpretation of the duty to take precautions within IHL) that goes beyond the concept of “feasibility.” N.B. The Authors do not envisage consideration of “state practice” in terms of the formation process of a customary international rule<sup>58</sup> though of course, one could suggest that if there is a suitable variance

---

56. A question that may be raised here is whether a commander may discharge his/her duty by deploying a weapon which is independently capable of doing everything “feasible” in terms of making real-time CDEM assessments. In other words, could deploying an EAI lessen the burden upon the human commander? A full investigation is beyond the scope of the current discussion; however, due to the fluidity of armed conflict, the “fog” and/or unpredictability of war, this is perhaps unlikely.

57. For the sake of this Article, leading nations are considered to be those that have a (global firepower) power index that ranks them number twenty-five or above. *See 2020 Military Strength Ranking*, GLOBAL FIRE POWER (2020), <https://www.globalfirepower.com/countries-listing.asp> [https://perma.cc/5AGF-225D]; *see also* Ellen Ioanes, *This is How the US and Iran Rank Among the World’s 25 Most Powerful Militaries*, BUS. INSIDER (Jan. 7, 2020, 4:48 PM), <https://www.businessinsider.com/most-powerful-militaries-in-the-world-ranked-2019-9?r=US&IR=T> [https://perma.cc/5ANP-85YC].

58. For a useful discussion regarding the state practice and opinion juris, the elements that need to be satisfied in this sense, see DINSTEIN, *supra* note 23, at 16-17.



present between states it may undermine the idea that there is overall consistency in this respect.<sup>59</sup> The following examination is conducted with a considerable focus upon Article 57 API. Ultimately, however, it is during the processes of “feasible verification” and “feasible precaution” as per Article 57(2)(a)(i) and Article 57(2)(a)(ii) API respectively<sup>60</sup> that the Authors wish to ground their discussion in terms of rejecting an order.

In the first instance, the doctrine of all six states references the wider “constant care” obligation. While in the majority of cases the language utilized broadly reflects Article 57 (1),<sup>61</sup> one US Naval manual, once again, refers specifically to the concept of “reasonableness.”<sup>62</sup> Arguably, “all reasonable precautions” implies a lesser obligation than that of say, Germany, whose manuals identify “all necessary precautions.” A similar discrepancy appears with regard to feasible verification.<sup>63</sup> Here, a

59. DINSTEIN, *supra* note 23, at 16-17.

60. API, *supra* note 20, art. 57(2)(a)(i)-(ii).

61. *See generally* ICRC Practice Relating to Customary Rules, *supra* note 52, at 337-38 (citing AUSTRALIAN DEFENCE FORCE, MANUAL ON LAW OF ARMED CONFLICT, OPERATION SERIES, ADFT 37 – INTERIM EDITION § 556 (1994) (Australia’s Defence Force Manual states: “[i]n the conduct of military operations, constant care must be taken to spare the civilian population and civilian objects to the maximum extent possible.”)); OFFICE OF THE JUDGE ADVOCATE GEN., LAW OF ARMED CONFLICT AT THE OPERATIONAL AND TACTICAL LEVELS §15 (1999) (“civilians are entitled to protection from the dangers arising from military operations. In conducting operations care should always be taken to spare civilians and civilian objects.”); DIRECTIVE OF THE MINISTRY OF DEFENCE, HANDBOOK ON THE LAW OF ARMED CONFLICT 98 (2000) (“In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.”); GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 404 (“The civilian population as such as well as individual civilians . . . shall be spared as far as possible.”); *see also* ICRC Database Rule 15, *supra* note 54 (citing CTR. FOR INNER LEADERSHIP, HUMANITARIAN INTERNATIONAL LAW IN ARMED CONFLICTS – PRINCIPLES 4 (2006) (“When attacking a military objective, all necessary precautions shall be taken to spare as far as possible the civilian population located in the area or in the immediate vicinity of the object.”)); MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT § 5.23 (2004) (“In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects.”)).

62. ICRC Database Rule 15, *supra* note 54 (citing DEP’T OF THE NAVY ET AL., THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 8.1 (2007) (“[t]he law of targeting, therefore, requires that all reasonable precautions must be taken to ensure that only military objectives are targeted so that noncombatants, civilians, and civilian objects are spared as much as possible from the ravages of war.”)).

63. *See* ICRC Practice Relating to Customary Rules, *supra* note 52, at 367-70 (“[e]ach party to the conflict must do everything feasible to verify that targets are military objectives . . . [and that] . . . state practice establishes this rule as a norm of customary

German commander is required to positively verify every target,<sup>64</sup> where in contrast US doctrine, once again requires reasonable precautions to be taken.<sup>65</sup> UK doctrine states a decision-maker can only be expected to do everything feasible.<sup>66</sup> As do Australian,<sup>67</sup> and Canadian doctrines.<sup>68</sup> French doctrine, however, mirrors the German requirement to “verify that the objectives to be attacked are neither civilians nor civilian objects.”<sup>69</sup> In practice, it may prove difficult to identify the existence of a large disparity between what might be considered reasonable, and what is considered necessary. However, on paper at least, “individual” practice with regard to target verification can be established. As a result, while French and German armed forces may be “required” to deploy future EAI as often as is possible, the United States is likely to give their decision-makers more maneuverability when faced with the option.

The obligation to avoid, or at least minimize incidental damage (ICRC Rule 15) represents *one* of two elements of the

---

international law applicable in both international and non-international armed conflicts.”).

64. See Int’l Comm. of the Red Cross, *Practice Relating to Rule 16: Targeting Verification*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_rul\\_rule15#top](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule15#top) [<https://perma.cc/63NC-LGGZ>] [hereinafter ICRC Database Rule 16]; GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 457 (“Before engaging an objective, every responsible military leader shall verify the military nature of the objective to be attacked.”).

65. See DEP’T OF THE NAVY ET AL., THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 8.1 (2007).

66. *Id.*; MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT § 13.32 (2004) (“(a) those who plan, decide upon or execute an attack must take all feasible measures to gather information which will assist in determining whether or not objects which are not military objectives are present in an area of attack, (b) in the light of the information available to them, those who plan, decide upon or execute an attack shall do everything feasible to ensure that attacks are limited to military objectives.”).

67. AUSTRALIAN DEFENCE DOCTRINE PUBL’N 06.4: LAW OF ARMED CONFLICT § 5.61 (2006), <https://www.defence.gov.au/adfwc/documents/doctrinelibrary/addp/addp06.4-lawofarmedconflict.pdf> [<https://perma.cc/XY8L-4DFJ>].

68. *Id.*; CHIEF OF THE DEFENCE STAFF, USE OF FORCE FOR CF OPERATIONS § 112.2 (2008) (“All feasible precautions must be taken to verify that the target is a military objective, and not a civilian or a civilian object, and that it is not subject to any of the specialized regimes of protection which prohibit, or severely restrict, attacks on certain persons and objects.”).

69. DIRECTIVE OF THE MINISTRY OF DEFENCE, MANUAL OF ARMED CONFLICT 98 (2001).

concept of feasible precautions.<sup>70</sup> As noted, the other refers to the means and method of attack (ICRC Rule 17).<sup>71</sup> Given that the United States is not a party to API, these “customary” rules are also considered individually for the sake of the following examination. With regards to Rule 15, US,<sup>72</sup> and Australian doctrines refer, once again, to the concept of reasonableness.<sup>73</sup> This is repeated by Canadian doctrine,<sup>74</sup> though further guidance also states military operations should be “conducted in such a way that damage to civilians and their property is minimized . . . [and] . . . only the necessary force that causes the least amount of collateral civilian damage” is used.<sup>75</sup> French Doctrine requires “all precautions,”<sup>76</sup> while German returns to the concept of “all *feasible* precautions.”<sup>77</sup> Finally, UK guidance states that “[p]recautions must be taken . . . to avoid civilian death or injury and damage to civilian objects”,<sup>78</sup> and perhaps a little more ambiguously, “[c]are must be taken . . . .”<sup>79</sup>

---

70. Int’l Comm. of the Red Cross, *Rule 15. Principle of Precautions on Attack*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_rul\\_rule15](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule15) [<https://perma.cc/3LG6-FZ9M>] [hereinafter ICRC Rule 15].

71. Int’l Comm. of the Red Cross, *Rule 17. Choice of Means and Methods of Warfare*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1\\_rul\\_rule17](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule17) [<https://perma.cc/E7H4-379T>] [hereinafter ICRC Rule 17].

72. ICRC Database Rule 15, *supra* note 54 (citing DEP’T OF THE NAVY ET AL., THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 8.3.1(2007)) (“commanders must take all reasonable precautions, taking into account military and humanitarian considerations, to keep civilian casualties and damage to the minimum consistent with mission accomplishment . . .”).

73. AUSTRALIAN DEFENCE DOCTRINE PUBL’N 06.4: LAW OF ARMED CONFLICT § 5.53 (2006) (“all reasonable precautions must be taken to avoid injury, loss or damage to civilians and civilian objects and locations. It is therefore important to obtain accurate intelligence before mounting an attack.”).

74. ICRC Practice Relating to Customary Rules, *supra* note 52, at 345; *see also* OFFICE OF THE JUDGE ADVOCATE GEN., LAW OF ARMED CONFLICT AT THE OPERATIONAL AND TACTICAL LEVELS §711.3 (2001).

75. ICRC Database Rule 15, *supra* note 54 (citing § 4, Rule 2 of Canada’s Code of Conduct (2005)); *see also* Code of Conduct for CF Personnel, § 4, Rule 2 (Can.).

76. ICRC Practice Relating to Customary Rules, *supra* note 52, at 346 (citing DIRECTIVE OF THE MINISTRY OF DEFENCE, LOAC TEACHING NOTE 2 (2000)).

77. *Id.* (citing GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 510).

78. ICRC Database Rule 15, *supra* note 54 (citing MINISTRY OF DEFENCE ET AL., THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT 317, § 12.26 (2004)).

79. ICRC Database Rule 15, *supra* note 54 (citing MINISTRY OF DEFENCE, THE LAW OF ARMED CONFLICT, D/DAT/13/35/66, ARMY CODE 71130 § 4(b) (1981)).

Consequently, although “feasibility” is a central focus of the six states’ national doctrines, each arguably exhibits a small degree of divergence in terms of interpretation of Rule 15. Moreover, with regard to Rule 17, Australian doctrine<sup>80</sup> identifies that while military decision-makers must do everything feasible to ensure collateral damage is minimized, “the existence of precision guided weapons munitions . . . does not mean they must necessarily be used.”<sup>81</sup> This is a stance that is reflected by US doctrine, which states:

The commander must decide, in light of all the facts known or reasonably available to him, including the need to conserve resources and complete the mission successfully, whether to adopt an alternative method of attack, if reasonably available, to reduce civilian casualties and damage Under the heading of “proportionality.”<sup>82</sup>

Canadian doctrine refers back to Article 57(2),<sup>83</sup> and similarly, a French manual states that decision-makers shall “take all precautions which are practically possible in the choice of means and methods of attack . . . .”<sup>84</sup> German doctrine is, however, more stringent requiring that “before engaging an objective, every responsible military leader shall . . . choose means and methods minimizing incidental injury and damage to civilian life and objects.”<sup>85</sup> *Prima facie*, this narrow obligation is also a

---

80. See Int’l Comm. of the Red Cross, *Practice Relating to Rule 17. Choice of Means and Methods of Warfare*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_rul\\_rule17](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_rule17) [<https://perma.cc/P88W-65ZN>] [hereinafter ICRC Database Rule 17] (citing AUSTRALIAN DEFENCE FORCE, MANUAL ON LAW OF ARMED CONFLICT, OPERATION SERIES, ADFT 37 – INTERIM EDITION § 556(e) (1994)).

81. ICRC Practice Relating to Customary Rules, *supra* note 52, at 376 (citing AUSTRALIAN DEFENCE FORCE, MANUAL ON LAW OF ARMED CONFLICT, OPERATION SERIES, ADFT 37 – INTERIM EDITION § 834 (1994)).

82. ICRC Database Rule 17, *supra* note 80 (DEP’T OF THE NAVY ET AL., THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 8.3.1 (2007)).

83. ICRC Database Rule 17, *supra* note 80 (citing CANADIAN FORCES JOINT PUBL’N, USE OF FORCE FOR CF OPERATIONS § 112.4 (2008) (“[a]l feasible precautions must be taken in the choice of means and methods of attack to avoid, and in any event to minimize, incidental civilian loss and damage (i.e., collateral damage.”)).

84. ICRC Practice Relating to Customary Rules, *supra* note 52, at 377 (citing DIRECTIVE OF THE MINISTRY OF DEFENCE, MANUAL OF ARMED CONFLICT 2 § 5.2 (2001)).

85. ICRC Practice Relating to Customary Rules, *supra* note 52, at 377 (citing GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 457).

requirement of UK armed forces.<sup>86</sup> However, the UK Manual states this is only where feasible.<sup>87</sup> With the Part II analysis nearing completion, three primary interpretive tracks have come to the fore; (i) the narrow interpretation, (ii) the treaty-based interpretation, and (iii) the wide interpretation. These are represented as follows:

---

86. See MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT 82 §5.32.4 (2004) (“There is the obligation to select the means (that is, weapons) or methods of attack (that is, tactics) which will cause the least incidental damage commensurate with military success.”).

87. See MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT 83 (2004).

	Obligation to take Constant Care	Feasible Verification (ICRC Rule 16)	Feasible Precautions (A): Minimize Civilian Harms (ICRC Rule 15)	Feasible Precautions (B): Choice of Means and Methods (ICRC Rule 17)
<b>Wide Interpretation</b>	<b>US</b> “ . . . take all reasonable precautions . . . ”	<b>US</b> “ . . . ensure reasonable precautions are taken . . . ”	<b>US</b> “ . . . where reasonable . . . ”	<b>US</b> “ . . . conservation of resources . . . [and] . . . reasonable availability . . . ”
			<b>Aus.</b> “ . . . where reasonable . . . ”	<b>Aus.</b> “ . . . consider wider mission objectives . . . ”
			<b>Can.</b> “ . . . where reasonable . . . ”	
<b>Central/Treaty Based Approach</b>	<b>Aus.</b> “ . . . maximum extent possible . . . ”	<b>Aus.</b> “ . . . feasible precautions . . . ”		
	<b>Can.</b> “ . . . care should always be taken . . . ”	<b>Can.</b> “ . . . feasible precautions (as per Art. 3(4) CCCW) . . . ”		<b>Can.</b> “ . . . all feasible precautions . . . ”
	<b>UK</b> “ . . . constant care shall be taken . . . ”	<b>UK</b> “ . . . feasible precautions (as per Art. 3(4) CCCW) . . . ”	<b>UK</b> “ . . . where feasible (as per Art. 3(4) CCCW) . . . ”	<b>UK</b> “ . . . least collateral damage commensurate with military success . . . ”
	<b>Fra.</b> “ . . . constant care shall be taken . . . ”			<b>Fra.</b> “ . . . all precautions which are practically possible . . . ”
			<b>Ger.</b> “ . . . where feasible (as per Art. 3(4) CCCW) . . . ”	
<b>Narrow Interpretation</b>		<b>Fra.</b> “ . . . Must verify objects to be attacked . . . ”	<b>Fra.</b> “ . . . take all precautions . . . ”	
	<b>Ger.</b> “ . . . all necessary precautions (to spare as far as possible) . . . ”	<b>Ger.</b> “ . . . shall verify the military status of target . . . ”		<b>Ger.</b> “ . . . shall choose means and methods to minimize civilian harms . . . ”

Figure 1. *The IHL Duty to Take Precautions in Attack: State Practice.*

Figure 1 identifies that while the language used in national doctrine is comparable, some states nevertheless appear to place a greater burden upon their troops to ensure DTP compliance.

And, once the doctrine which largely repeats the text of the treaty is removed, this practice can be distilled one step further—presented as follows:

State Practice/ Interpretation	Measures taken to ensure target verification . . . ;	All precautions to minimize civilian harms where . . . ;	Use Means or Method . . . ;
<b>Wide</b>	Must be reasonable in the circumstances.	Reasonable	Where it is reasonable to do so, [ <i>and in particular where it is consistent with the concepts of resource conservation, or reasonable availability</i> ].
<b>Narrow</b>	Must be all available measures in all circumstances.	Feasible	Whenever/ wherever possible.

*Figure 2. The Narrow and Wide interpretations of the Duty to Take Precautions*

To summarize, DTP is irrevocably intertwined with distinction and proportionality. Together, the three concepts help maintain the delicate balance between humanitarian considerations and military necessity. The discussion in Part II has centered specifically upon “feasible verification,” and “feasible precautions.” To this, one may also add the obligation for the targeteer to take constant care, though this also remains subject to feasibility.<sup>88</sup> Having considered a small sample of state practice, the analysis has identified that certain states appear to sway towards the military necessity extreme of the IHL spectrum as compared to others. This is perhaps most noticeable when comparing US practice to that of Germany—for example, a German soldier *must* assess each target before choosing the means and method that *will* minimize civilian harm. The US targeteer is actively encouraged to consider the wider mission brief. However, regardless of such variations, it is undeniable that there will be circumstances where those that possess EAI and AWS technology will be legally obligated to use them. A key question

---

88. See Schmitt & Widmar, *supra* note 41, at 400-01.

that is posed by the current thesis, and one which the following section addresses, is what are the conditions under which an EAI should *refuse* an order?

### III. RULES OF ENGAGEMENT AND DISOBEYING OF ORDERS

The discussion in Part II identified that the DTP obligation can be subdivided into numerous interconnected paths.<sup>89</sup> The analysis concluded that while a narrow and wide interpretation of “feasibility” can be crafted, neither path detracts from the certainty that where EAI is a feasible option, armed-forces will be obligated to deploy them. The purpose of Part III is to forensically examine the interface between DTP and the state-based Rules of Engagement (“ROE”).<sup>90</sup> Here, specific reference is naturally centered on the disobeying of orders. This examination is necessary because if nations are to be required to utilize EAI, a key question arises as to whether robots can, and should, refuse an order to attack? Similar to the Part II examination, Part III first identifies how the obligation to disobey orders fits within the corpus of IHL, and in particular within DTP.<sup>91</sup> This is followed, once again, with an examination of state practice in this area.

#### A. How does the Duty to Disobey an Order fit within DTP?

No matter whether AWS as a “means” are realized, unarmed AI/EAI will become a “method” utilized by armed-forces. The introduction of increasingly autonomous tech can be seen

---

89. Indeed, the art. 57 obligations extend further than is considered for the current analysis. This includes, for example, a duty to provide an advance warning of an attack, where possible [i.e. where feasible] and, to choose the target that leads to lesser civilian harms when presented with two or more targets that offer a similar military advantage. See API, *supra* note 20, art. 57 (2)(c)-(3).

90. For the sake of the present article, the terms “ROE” and “military doctrine” are used interchangeably. The DoD Dictionary of Military and Associated Terms, for example states: “ROE [are] directives issued by competent military authority that delineate the circumstances and limitation under which US forces will initiate and/or continue combat engagement with other forces encountered. (“Doctrine” is undefined).” See *DOD Dictionary of Military and Associated Terms*, DEP’T DEFENSE (June 2020), <https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf?ver=2020-06-18-073638-727> [<https://perma.cc/D9BZ-WXRD>].

91. How DTP fits within the wider body of IHL has already been examined.



throughout industry, and there is no reason to suspect the military domain will be any different. For example, a human may program an autonomous car with a destination address, but the vehicles' software is responsible for making decisions about which turn to either make, or not.<sup>92</sup> In the military domain however, the task will not always be as straightforward as getting from "A" to "B." But, where deployed, an EAI may have to refuse to order to ensure a positive military advantage is balanced with humanitarian considerations—but in which circumstances?

This Section identifies the concept of manifestly unlawful orders and the customary duty to disobey one. This duty itself must be viewed alongside the fact that a combatant also has a duty to follow orders. Indeed, as noted by one commentator,<sup>93</sup> from 1863, Article 44 of Lieber code permitted a commanding officer to shoot and kill a subordinate on the spot, should they disobey their order to cease committing a crime.<sup>94</sup> Today's soldier may not face quite the same fate. However, they may still be subject to court martial or imprisonment.<sup>95</sup> As a consequence, in the vast majority of circumstances (while seemingly at odds with the ICJ's rejection of the "Nuremberg defense"),<sup>96</sup> a combatant *may* rely

---

92. See PAUL SCHARRE, *ARMY OF NONE* 31-32 (W. W. Norton & Co. eds., 1st ed. 2018) (identifying the automatic/automated/autonomous spectrum, the author notes: "[a]s machines become more sophisticated, they become more capable and able to accomplish more complex tasks in more open-ended environments.").

93. GURGEN PETROSIAN, *Elements of Superior Responsibility for sexual violence by subordinates*, 42 *MAN. L.J.* 123, 127 (2019).

94. General Orders No. 100: Instructions for the Government of Armies of the United States in the Field art. 44 (Apr. 24, 1863) [hereinafter Lieber Code] ("a soldier, officer or private, in the act of committing such violence, and disobeying a superior ordering him to abstain from it, may be lawfully killed on the spot by such superior.").

95. See, e.g., Armed Forces Act, (2006), art. 12(1) (UK) ("A person subject to service law commits an offence if— (a) he disobeys a lawful command; and (b) he intends to disobey, or is reckless as to whether he disobeys, the command. (2) A person guilty of an offence under this section is liable to any punishment mentioned in the Table in section 164, but any sentence of imprisonment imposed in respect of the offence must not exceed ten years.").

96. Charter of the International Military Tribunal, annexed to Agreement for the Prosecution and Punishment of Major War Criminals of the European Axis ("London Agreement") Aug. 8, 1945, 59 *Stat.* 1544, 82 *U.N.T.S.* 279 ("The fact that the defendant acted pursuant to order of his Government or of a superior shall not free him from responsibility, but may be considered in mitigation of punishment if the Tribunal determines that justice so requires.").

upon the defense that they were acting in accordance with orders.<sup>97</sup>

The majority of domestic courts recognize one of three variations of this defense: (i) that following orders is always a defense; (ii) that following orders is a defense where the combatant believed it was a lawful order, and; (iii) that following orders is a defense where it would have been *reasonable* to believe that the order was lawful.<sup>98</sup> The United States adheres to the latter of these three variants, with the passage previously identified continuing, “. . . unless the accused knew the orders to be unlawful or a person of ordinary sense and understanding would have known the orders to be unlawful.”<sup>99</sup> According to the second and third variants of the “following orders” defense therefore, there are circumstances where a soldier either “will” recognize an unlawful order, or where they “should” have recognized an unlawful order. And this is a customary duty, thus binding upon all parties.<sup>100</sup> Significantly, this not only means that there are limited circumstances under which a combatant *can* refuse to follow an order, but, that in such circumstances “[e]very combatant has a *duty* to disobey a manifestly unlawful order.”<sup>101</sup> Indeed, where the combatant does not fulfil this latter duty, but instead carries out a manifestly unlawful order, no “Nuremberg defense” can be relied upon, and the subordinate will face individual criminal responsibility. This is considered an additional customary provision.<sup>102</sup>

“Manifestly unlawful,” or “manifestly criminal,” generally refers to war crimes. This includes breaches of the IHL principles of distinction and proportionality.<sup>103</sup> In addition, crimes against

97. See, e.g., MANUAL FOR COURTS-MARTIAL, UNITED STATES, §916(d) (2016) [hereinafter US Courts-Martial Manual] (“It is a defense to any offense that the accused was acting pursuant to orders . . .”); see also Keith A. Petty, *Duty and Disobedience: The Conflict of Conscience and Compliance in the Trump Era*, 45 PEPP. L. REV. 55, 103-04 (2018) (citing Osiel, *supra* note 2, at 21 (“Virtually all military codes include some provision on due obedience . . .”)).

98. See Osiel, *supra* note 2, at 950; see also Petty, *supra* note 97, at 103-04.

99. US Courts-Martial Manual, *supra* note 97, II-116.

100. See ICRC Customary Rules, *supra* note 32, at 563-64.

101. See *id.*

102. See *id.* at 565 (“Obeying a superior order does not relieve a subordinate of criminal responsibility if the subordinate knew that the act ordered was unlawful or should have known because of the manifestly unlawful nature of the act ordered.”).

103. The legal basis of which was considered in Section II.A *supra*.

humanity and crimes against the peace would, in the majority of cases, also be considered such.<sup>104</sup> In short, a combatant is therefore only presumed to know a subset of crimes that are “immediately recognizable as manifestly criminal by a person of ordinary understanding.”<sup>105</sup> Such a limitation may be *just*, for it is questionable whether the law should require armed services personnel to carry out their non-civilian duties while simultaneously having to conduct a full legal assessment of each and every instruction that is handed down to them by a superior. Most soldiers are not lawyers and nor should they be required to be. Instead, the obligation to *obey* their superiors is what sets combatants apart from civilians.<sup>106</sup>

### B. Disobeying of Orders

The following Section examines existing state practice in order to identify whether it can help to determine the circumstances under which EAI should be programmed to refuse an order. It examines the practice of the six states previously considered, and extends the analysis to the practice of three further nations not widely considered to be leading military powers. These additional states are the Philippines, the Republic of the Congo (“Congo”), and the Democratic Republic of the Congo (“DRC”).

Science fiction writer Isaac Asimov’s second law of robotics states that, “a robot must obey orders given to it by a human

---

104. See Petty, *supra* note 97, at 97 (citing DEP’T OF THE ARMY, DEPARTMENT OF THE ARMY FIELD MANUAL: THE LAW OF LAND WARFARE FM 27-10 178 (1956) (“[a]lthough this manual recognizes the criminal responsibility of individuals for [crimes against peace, crimes against humanity, and war crimes], members of the armed forces will normally be concerned only with those offenses constituting ‘war crimes.’”)).

105. See Osiel, *supra* note 2, at 961. Note that the author supports the notion that the duty upon a US soldier is only to refuse to follow a manifestly unlawful order. Thus the US soldier does not need to exercise “situational judgement.” See Osiel, *supra* note 2, at 971.

106. Note for example, that military doctrine accepts that the “obligation” to follow a superior’s orders is what separates members of the armed forces from their civilian counterparts. See, e.g., IHL Database – New Zealand, *Practice Relating to Rule 154*, ICRC, [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_cou\\_nz\\_rule154](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_cou_nz_rule154) [https://perma.cc/VK9D-DYPD] (last visited Feb. 18, 2021) (New Zealand’s Armed Conflict Manual states: “[An] obligation, and the one which clearly sets a member of a military force apart from his civilian counterparts, is the obligation to obey lawful commands of a superior officer.”).

except where such orders conflict with the first law.”<sup>107</sup> His first law is “a robot may not injure a human being, or, through inaction, allow a human being to come to harm.”<sup>108</sup> As the Authors have noted elsewhere,<sup>109</sup> while these rules allow Asimov to cleverly write stories that toy with human-EAI relationships, they are utterly inadequate for robots that are programmed specifically to apply “deliberate . . . often lethal capabilities in order to produce maximum effect upon an enemy.”<sup>110</sup>

Instead, the discussion in the previous section identified that a soldier has a strict obligation to follow an order, except where they have a customary duty to refuse to follow an order. Presumably, when advanced EAI's are introduced, this obligation would simply be transposed and implemented. However, in practice, the types of orders and circumstances under which a combatant is expected to refuse an order differ significantly. These variations exist because they are contained within municipal military doctrine, or Rules of Engagement (“ROE”), as opposed to being codified within IHL.<sup>111</sup> Somewhat regrettably,

---

107. See generally ISAAC ASIMOV, I, ROBOT 40 (1950).

108. *Id.* (Asimov's 3 laws of robotics state that: “[o]ne, a robot may not injure a human being, or, through inaction, allow a human being to come to harm . . . [t]wo . . . a robot must obey the orders given it by human beings except where such orders would conflict with the First Law . . . [a]nd three . . . a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.”)

109. Grimal & Pollard, *supra* note 26, at 523.

110. Corn & Schoettler, *supra*, note 19, at 795. Indeed, as noted in a previous piece, it is not unusual for Asimov's rules to appear in the debate surrounding AWS. Corn & Schoettler, *supra*, note 19, at 795 (citing Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. PA. L. REV. 1347, 1372 n.135 (2016)). See also Andrew Figueroa, *License to Kill: An Analysis of the Legality of Fully Autonomous Drones in the Context of International Use of Force Law*, 31 PACE INT'L L. REV. 145, 156 n.71 (2018).

111. Article 77 of a draft version of API did state that “no person shall be punished for refusing to obey an order of his government or of a superior which, if carried out, would constitute a grave breach of the provisions of the Conventions or of the present Protocol.” See CDDH, Official Records, Vol. I, Part Three, Draft Additional Protocols, June 1973, at 25. However, as noted by the ICRC, the provision failed to gain the required two thirds majority that it needs to pass as law. See ICRC Practice Relating to Customary Rules, *supra* note 52, at 3799-3800; see also Rome Statute of the International Criminal Court (“Rome Statute”) art. 33(1), July 17, 1988, 2187 U.N.T.S. 3 (relating to Superior Orders and Prescription of Law, which states: “the fact that a crime within the jurisdiction of the court has been committed by a person pursuant to an order of a Government or of a superior, whether military or civilian, shall not relieve that person of a criminal responsibility unless: (a) The person was under a legal obligation to obey orders of the Government or the superior in question; (b) The person did not know that the order was unlawful; and (c) The order was not manifestly unlawful.”).

this means that while the members of certain armed forces may make a subjective determination as to which orders should be disobeyed, others may do so only objectively.<sup>112</sup> This is concerning because the Authors support the view that EAI, and AWS can increase adherence to international law. For example, EAIs will operate while simultaneously processing vast amounts of preprogrammed information, including unabridged IHL databases. Moreover, they will do so while remaining unaffected by the threats and pressures that are typically associated with armed conflict. However, while autonomous technology in general can offer an opportunity to ensure the duty to disobey a superior order is applied equally, the current variation in state practice is likely to be reflected as somewhat of a national bias when programming EAIs and other such technologies.

In the United States, ROE generally identify that a combatant must distinguish only between lawful and unlawful orders (the latter of which must be disobeyed).<sup>113</sup> In this instance US doctrine appears to support a “narrow” interpretation. This is consistent with the concept of “manifest unlawfulness” that is recognized, for example, by the ICRC.<sup>114</sup> UK doctrine adopts a similar stance, stating “[m]ilitary personnel are required to obey lawful commands but must not obey unlawful commands.”<sup>115</sup> And similarly, Philippine doctrine states “[a]nyone who shall refuse or fail to carry out a lawful order from the military chain of command shall be subject to military discipline.”<sup>116</sup>

---

112. Having identified that following an instruction would lead to the commission of a “manifestly” criminal act.

113. See, e.g., ICRC Practice Relating to Customary Rules, *supra* note 52, at 3804 (“[m]embers of the armed forces are bound to obey only lawful orders”) (citing U.S. DEP’T OF THE AIRFORCE, INTERNATIONAL LAW – THE CONDUCT OF ARMED CONFLICT AND AIR OPERATIONS, PAMPHLET 110-31 §15-4(d) (1976)); see also HEADQUARTERS DEP’T OF THE ARMY, YOUR CONDUCT IN COMBAT UNDER THE LAW OF WAR, PUBLICATION FIELD MANUAL NO. 27-2 26 (1984) (“although you are responsible for promptly obeying all legal orders issued by your leader, you are obligated to disobey an order to commit a crime.”).

114. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3799-3814. See also Rome Statute, *supra* note 104, art. 33(1). The existing standard is one of “manifest” unlawfulness. *Id.*

115. ICRC Practice Relating to Customary Rules, *supra* note 52 (citing MINISTRY OF DEFENCE, THE LAW OF ARMED CONFLICT, D/DAT/13/35/66, ARMY CODE 71130 46, Annex A § 2).

116. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3803 (citing ARMED FORCES OF THE PHILIPPINES CODE OF ETHICS 16-17 (1991)).

Though only subtly different, some states adopt a slightly wider approach. French ROE, for example, identifies the sources of law that should inform a decision whether to disobey an order, these being, “the customs of war . . . [*and*] . . . the rules of international law applicable in armed conflicts, or duly ratified or approved international treaties.”<sup>117</sup> Analogously, and offering guidance as what to do where an order is ambiguous, Australian doctrine supplies, “clarification should be sought. If clarification is unavailable, any action taken must comply with LOAC [Law of Armed Conflict].”<sup>118</sup> And, similarly, Congolese doctrine states “the subordinate must not execute an order to commit an act manifestly . . . contrary to the customs of war and to the international conventions.”<sup>119</sup>

The obligations contained within German, DRC, and Canadian doctrine must, however, be considered wide. This is because while each appears to be aligned with the standard of manifest unlawfulness,<sup>120</sup> the guidance that each state supplies to its combatants is significantly more conceptual than that which is offered to the armed forces of states such as the United States, and the United Kingdom. For example, the DRC Constitution provides,

No one is required to execute a manifestly illegal order. Every individual, every State agent is relieved from the duty to obey, when an order received constitutes a manifest infringement

---

117. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3802 (citing MINISTRY OF DEFENCE, DISCIPLINARY REGULATIONS AS AMENDED art. 8(3) (1975)).

118. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3800 (citing AUSTRALIAN DEFENCE FORCE, ADFT 37 – INTERIM EDITION § 1306); *see also* AUSTRALIAN DEFENCE DOCTRINE PUBLICATION 06.4, *supra* note 70, at § 13.7.

119. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3802 (Décret No. 86/057 du 14 janvier 1986 portant Règlement du Service dans l’Armée Populaire Nationale, 1986).

120. *See* Code of Conduct for CF Personnel (Can.), *supra* note 75, Rule 11, § 4 (“in accordance with military custom you should . . . obey and execute the order—unless—the order is manifestly unlawful”); *see also* Congo (Democratic Republic of the)’s Constitution of 2005 with Amendments through 2011, CONSTITUTE PROJECT, art. 28, [https://www.constituteproject.org/constitution/Democratic\\_Republic\\_of\\_the\\_Congo\\_2011.pdf?lang=en](https://www.constituteproject.org/constitution/Democratic_Republic_of_the_Congo_2011.pdf?lang=en) [https://perma.cc/6DTL-9KF6] (last visited Feb. 18, 2021) [hereinafter DRC Constitution of the Transition] (“No one is required to execute a manifestly illegal order.”); ICRC Practice Relating to Customary Rules, *supra* note 50, at 3802 (citing GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS §§ 144-145). It is expressly prohibited to obey orders whose execution would be a crime.

of the respect of the rights of man and of the public freedoms and of morality.<sup>121</sup>

This clearly goes further than identifying breaches of IHL, LOAC, or International Human Rights Law (“IHRL”).<sup>122</sup> Moreover, the obligation to apply a moral perspective when considering manifest unlawfulness casts the net much farther than the seemingly straightforward lawful or unlawful assessment that the narrow interpretation requires. In addition, while Canadian doctrine mirrors the Australian position regarding clarification,<sup>123</sup> it also identifies a manifestly unlawful act as one which “shocks the conscience of every reasonable, right-thinking person.”<sup>124</sup> While this is perhaps not quite as abstract as the DRC guidance, it nevertheless provides a wide scope of assessment.

The final doctrine considered in this area is that of Germany. Germany’s interpretation of the customary duty to disobey an unlawful order is arguably wider than that of its peers due to its historical requirement for unconditional obedience to military orders.<sup>125</sup> As a result, current German doctrine identifies that its soldiers should not execute an order if:

. . . it violates the human dignity of the third party concerned or the recipient of the order; it is not of any use for service;

---

121. See DRC Constitution of the Transition, *supra* note 120, art. 28.

122. See 1996 I.C.J. 226, *supra* note 27, ¶ 25. While International Human Rights Law is applicable in armed conflict, IHL must take precedence where there is a conflict due to the concept of *lex specialis*. See also DINSTEIN, *supra* note 27, ¶¶ 89-92; Marko Milanovic, *The Lost Origins of Lex Specialis: Rethinking the Relationship Between Human Rights and International Humanitarian Law*, in THEORETICAL BOUNDARIES OF ARMED CONFLICT & HUMAN RIGHTS (Jens. D. Ohlin ed., 2014).

123. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3800 (Code of Conduct for CF Personnel, Rule 11, § 4 (Can.)).

124. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3801-02 (citing Code of Conduct for CF Personnel, Rule 11, § 5 (Can.), which continues with the examples that: mistreating someone who has surrendered or beating a detainee is manifestly unlawful, though this is not, of course, intended to be exhaustive).

125. For a useful discussion, see Erin Blakemore, *Why German Soldiers Don’t Have to Obey Orders*, HISTORY (Nov. 17, 2017), <https://www.history.com/news/why-german-soldiers-dont-have-to-obey-orders> [<https://perma.cc/JGJ2-3BD2>]; Blakemore notes, for example, that from August 1934, German soldiers swore an oath not to the state, but the Fuehrer himself. See *Adolf Hitler: The Fuehrer Oath*, JEWISH VIRTUAL LIBR., <https://www.jewishvirtuallibrary.org/the-fuehrer-oath> [<https://perma.cc/YCA6-Y9H3>] (last visited Feb. 18, 2021) (providing the translation: “I swear by almighty God this sacred oath: I will render unconditional obedience to the Fuehrer of the German Reich and people, Adolf Hitler, Supreme Commander of the Wehrmacht, and, as a brave soldier, I will be ready at any time to stake my life for this oath.”).

or in a definite situation, the soldier cannot reasonably be expected to execute it.<sup>126</sup>

This is significantly wider than the obligation to not carry out an unlawful, or manifestly unlawful, order. Indeed, a German combatant must consider not only the lawfulness of the order, but their own “dignity,” and that of the targeted.<sup>127</sup> If that were not considerably “wide” enough, a German combatant must also consider whether, in the circumstances, there is a reasonable chance of mission success, or whether the order serves the “defense of Germany . . . [or] . . . the pursuit or achievement of its political or economic aims.”<sup>128</sup> Consequently, it appears a German soldier should evaluate the reasoning behind each command. Clearly there is a significant variation in state practice in this area. Once again, three primary tracks come to the fore. These are presented in the flowing graphic:

---

126. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3802 (citing GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 142).

127. Blakemore, *supra* note 125.

128. Ilja Baudisch, *Freedom of Conscience and Right to Conscientious Objection - Refusal to Obey to Military Orders - Legal Ban on the Use of Force (Article 2(4) UN Charter) - Neutrality of States in Armed Conflicts*, 100 AM. J. INT'L L. 911, 912 (2006). For a similar, but perhaps a more convivial discussion, see John Ford, *When Can a Soldier Disobey an Order?*, WAR ON ROCKS (July 24, 2017), <https://warontherocks.com/2017/07/when-can-a-soldier-disobey-an-order/> [<https://perma.cc/B9WZ-9TNX>].



	<p style="text-align: center;"><b>Disobeying Orders</b></p> <p>A soldier has a strict obligation to follow an order, and to do so is a defense (three variants):</p> <p>(i) Always;</p> <p>(ii) Where they believed it was a lawful order, or</p> <p>(iii) Where they should have known it was an unlawful order . . . (see ICRC Rule 155)</p> <p>Therefore, there is a duty to follow an order, except where there is a customary duty to refuse (ICRC Rule 154).</p> <p>State practice relating to the circumstances in which a soldier should therefore refuse to follow an order includes:</p>
<b>Narrow Interpretation</b>	<b>US</b> “. . . must distinguish between lawful and unlawful . . .”
	<b>UK</b> “. . . obey only lawful commands . . .”
	<b>Phi.</b> “. . . must carry out a lawful order . . .”
<b>Central/Treaty Based Approach</b>	<b>Aus.</b> “. . . Seek clarification if ambiguous, or follow LOAC . . .”
	<b>Can.</b> “. . . manifestly unlawful means it would shock the conscience of every right-thinking person . . .”
	<b>Fra.</b> “. . . the customs of war . . . LOAC . . . duly ratified or approved international treaties . . .”
	<b>Con.</b> “. . . manifestly contrary to customs of war and international conventions . . .”
<b>Wide Interpretation</b>	<b>Ger.</b> “. . . if it violates human dignity (of first and third party), if the order is of no use to service, of if soldier cannot reasonably be expected to execute it . . .”
	<b>DRC.</b> “. . . can disobey an order if it is a manifest violation of human rights, of public liberties and morals . . .”

*Figure 3. The Duty to Disobey an Order: State Practice.*

While some may posit that the discrepancies noted in Part II are arguably a matter of semantics, the result of the disparities between the narrow and wide variants of the duty to disobey an order are often stark. For instance, US military courts have consistently refused to recognize that the obligation to refuse to follow a unlawful order is applicable to the decision to wage war.<sup>129</sup> As a result, First Lieutenant Watada, who refused to deploy to Iraq in 2006 because he believed the war was unlawful—and therefore, that any order he would receive there would also be—failed to convince the court to that effect.<sup>130</sup> While he escaped a charge for insubordination, ultimately, he left military service with “an ‘other than honorable’ characterization of service—the worst administrative discharge . . . an officer can receive.”<sup>131</sup> This was the case even though applicable US doctrine stated “an essential foundation for Army leaders is a character ‘comprised of a person’s moral and ethical qualities [which] helps to determine what is right . . . regardless of circumstances or consequences.’”<sup>132</sup>

In contrast, in 2005, the German Federal Administrative court (Germany has no military courts) had to consider very similar facts. In the *Limits of Obedience to Superior Orders* case the court considered a German Major who had refused to take part in a NATO information technology project that would potentially

---

129. See *United States v. Huet-Vaughn*, 43 M.J. 105, 114 (C.A.A.F.1995); *Watada v. Head*, No. C07-5549BHS, 2008 U.S. Dist. Lexis 88489, at 8-10 (W.D. Wash. Oct. 21, 2008); ICRC Practice Relating to Customary Rules, *supra* note 50. Petty, *supra* note 97, at 100-01.

130. See Ford, *supra* note 128 (“A few servicemembers have tried unsuccessfully to disobey orders to deploy in support of these operations. In 2006, 1st Lt. Ehren Watanda refused to deploy to Iraq because he believed the war was illegal. His arguments fell on unsympathetic ears. In fact, Watanda was not even permitted to present his preferred defense because [t]he order to deploy soldiers is a non-justiciable political question . . . an accused may not excuse his disobedience of an order to proceed to foreign duty on the ground that our presence there does not conform to his notions of legality.”); see also Murdoch, *supra* note 2, at 6-8; Petty, *supra* note 97, at 100-01.

131. Petty, *supra* note 97, at 101 (citing U.S. Dep’t of the Army, *Regulation 600-8-24, Officer Transfers and Discharges*, 6 (Sept. 13, 2011), [https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/pdf/web/ARN3140\\_AR600-8-24\\_FINAL.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN3140_AR600-8-24_FINAL.pdf) [<https://perma.cc/UGM4-MQEL>]).

132. See Petty, *supra* note 97, at 129 (citing DEP’T OF THE ARMY, ARMY DOCTRINE REFERENCE PUBLICATION 6-22, ARMY LEADERSHIP ¶ 3-1 (Aug. 1, 2012)).

support Operation Iraqi Freedom.<sup>133</sup> Similar to Watada, the Major believed that the conflict in Iraq was unlawful.<sup>134</sup> The decision in the German court however, is the antithesis of Watada—the court upheld the claim of the Major, and acquitted him of the charge.<sup>135</sup> It did so because having considered that the lawfulness of the war was indeed questionable, the court held the Major could rely upon his fundamental right of freedom of conscience under German Basic law.<sup>136</sup> Consequently, he was permitted to seek, and to be assigned, an alternative task.<sup>137</sup> This is an important decision because it reaffirmed that under German law the duty to obey orders “does not demand blind or unconditional devotion to superiors.”<sup>138</sup> However, perhaps most notably, the decision also noted that a German combatant should not be made to act against their moral or ethical convictions.<sup>139</sup>

These two similar but contrasting cases provide a fundamental example of the differences between what the

133. See Bundesverwaltungsgericht [German Federal Administrative Court] June 21, 2005, 120 Deutsches Verwaltungsblatt 1455 (Ger.) (known also as, *Limits of Obedience to Superior Orders Case* (2005)). For a useful English Language translation of the relevant paragraphs of the judgement, see *Germany: Practice Relating to Rule 154; Obedience to Superior Orders*, ICRC, [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_cou\\_de\\_rule154](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_cou_de_rule154) [<https://perma.cc/VHP4-A7Q9>] (last visited Feb. 18, 2021).

134. See Baudisch, *supra* note 128, at 911.

135. Baudisch, *supra* note 128, at 911.

136. Note that at ¶ 4.1.2.1 the German court states that “human dignity” also includes the protection of the freedom of conscience according to Article 4, ¶ 1 of the Basic Law. In any case, it does not reduce that protection. Translating ICRC Practice Relating to Customary Rules, *supra* note 52, at 154. See also Baudisch, *supra* note 128, at 911-12 n.2, 5 (citing Article 4 of The German Basic Law (*Grundgesetz*): “freedom of faith and conscience, and freedom to profess a religious or philosophical creed, shall be inviolable . . . (3) No person shall be compelled against his conscience to render military service involving the use of arms. Details shall be regulated by a federal law.”).

137. See Baudisch, *supra* note 128, at 914; see also Int’l Comm. of the Red Cross, *Practice Relating to Rule 154. Obedience to Superior Orders*, ICRC CUSTOMARY IHL DATABASE (2005), [https://ihl-databases.icrc.org/applic/ihl/ihlweb.nsf/DocIndex/v2\\_cou\\_de\\_rule154?OpenDocument&Click=](https://ihl-databases.icrc.org/applic/ihl/ihlweb.nsf/DocIndex/v2_cou_de_rule154?OpenDocument&Click=) [<https://perma.cc/6XMT-PVAK>] [hereinafter ICRC Database Rule 154].

138. See ICRC Database Rule 154, *supra* note 137; see also Baudisch, *supra* note 128, at 911.

139. ICRC Database Rule 154, *supra* note 137, ¶ 4.1.3.1.1 (“a soldier . . . has to act with all the diligence and responsibility possible to him and has to act accordingly. An “unconditioned” or “unconditional” obedience is not compatible with this normative imperative. Requested is rather a “thinking” obedience, an obedience “reflecting” the consequences of carrying out the order—especially also with regard to the limits of the applicable law and the ethical “yardsticks” of the personal conscience.”).

Authors identify as the narrow and wide obligations. However, it is not necessarily a novel or a ground-breaking discussion.<sup>140</sup> What it does highlight for the sake of the present Article, is that in all but the most incontrovertible of circumstances,<sup>141</sup> a combatant can never be sure that their decision to disobey an order will be supported. While the reasons to disobey an order may include self-preservation, preventing breaches of IHL, or the commission of war-crimes,<sup>142</sup> other factors, including political<sup>143</sup> and financial pressures,<sup>144</sup> warrior training, a sense of patriotism, group solidarity, the fear of being branded a coward by one's peers,<sup>145</sup> and the potential for court-martial or criminal charges, might cause any reasonable person to hesitate before doing so.

Many of those persuasive factors could be removed if an EAI was tasked with making an objective determination. For example, if three primary reasons for refusing an order are: (i) deterring a perceived wrong; (ii) exercising a freedom of conscience, and; (iii) avoiding self-harm,<sup>146</sup> it is arguable that only the first is applicable to a machine. However, a machine can still support a human in respect of the other two. For example, an EAI could analyze an order and guide a human decision-maker accordingly,

---

140. See generally Osiel, *supra* note 2; Paola Gaeta, *The Defence of Superior Orders: The Statute of the International Criminal Court versus Customary International Law*, EJIL 172, 172-91 (1999); see also YORAM DINSTEIN, *THE DEFENCE OF 'OBEDIENCE TO SUPERIOR ORDERS' IN INTERNATIONAL LAW* (2012). The Authors also note that the latter two discussions are focused more upon the *jus ad bellum*, than they are in the *jus in bello*, which is the intended subject matter of the present article. Nevertheless, they do provide an obvious and clear example of a narrow and wide obligation.

141. For example, an order to commit genocide, or torture.

142. For example, torture, sexual violence and mistreatment of POWs.

143. Higher ranking members of the armed forces may be particularly aware of a need to consider the current political environment. Consider, for example, the Captain of the USS Theodore Roosevelt, Brett Crozier, who was recently removed from his position, and looks set to face charges for failing to respect the chain of command. See Bradley Peniston, *Navy to Punish Fired Captain of the USS Theodore Roosevelt*, DEF. ONE (June 19, 2020), <https://www.defenseone.com/threats/2020/06/navy-punish-fired-uss-roosevelt-captain/166300/> [<https://perma.cc/RC7B-7R5Z>].

144. There may also be financial implications if a combatant was to have a criminal conviction imposed upon them, as this may prevent them from gaining further employment. See *id.*

145. See Petty, *supra* note 97, at 128 (citing Carl Ficarrotta, *Selective Conscientious Objection: Some Guidelines for Implementation*, in *WHEN SOLDIERS SAY NO: SELECTIVE CONSCIENTIOUS OBJECTION IN THE MODERN MILITARY* 197, 199 (Andrea Ellner et al. eds., 2016); H.L.A. HART, *THE CONCEPT OF LAW* 55-57 (Peter Crane et al. eds., 2d ed. 1994)).

146. Petty, *supra* note 97, at 109.

or even take action to prevent an order that was given mistakenly or maliciously.

For the sake of the examination in the following section, the state practice that has been considered thus far can now be distilled into three tracks. These are:

	<b>Adherence to DTP, and Influencing Factors for Disobeying Superior Orders</b>
<b>Track 1</b>	<i>Every soldier should take all reasonable precautions to verify targets as military objects, and should take all reasonable steps to minimize civilian harms. In choosing the means and methods of attack, the decision-maker should consider their reasonable availability, with a specific regard for the wider mission objectives. A soldier must only refuse to follow an order where they know, or should know, that it is a manifestly unlawful order.</i>
<b>Track 2</b>	<i>Every soldier should take all steps to ensure target verification and the minimization of civilian harms, where they are practicable or practically possible. The soldier should take into account the circumstances ruling at the time, including humanitarian and military considerations. In doing so, the decision-maker must consider all of the customs of war and applicable international treaties, and refuse to follow any order which would be in breach of them.</i>
<b>Track 3</b>	<i>Every soldier must take all necessary precautions to ensure a target is verified, and in each case, must choose the means and methods that ensures civilian harms are minimized. The factors which should influence the decision to refuse to follow an order include, but are not necessarily limited to: the lawfulness of the order; whether following the order would violate the human dignity of the first or third party (including personal moral and ethical standards); whether there is a reasonable chance of mission success; and whether the order is consistent with national policy/strategy.</i>

Figure 4. The Three Interpretive Tracks.

Naturally, each of these are themselves subject to some level of fluidity with regard to interpretation. This is particularly true of the central/ treaty-based track, which could either be heavily weighted toward the narrow or the wide variant. Nevertheless, for the sake of examination in Part IV, these provide an adequate point of reference.

To summarize, Part III has shown that the minimum standard (reflected by the narrow obligation) requires that a soldier has an obligation to follow an order, except where the customary duty requires for them to refuse. Nevertheless, a gray area exists because certain state practice suggests that the customary obligation to refuse an order extends further than merely distinguishing between lawful and unlawful orders. Indeed, for some states, a soldier must exhibit high levels of “situational awareness” and determine whether the order should be followed in light of many varied considerations.<sup>147</sup> The question posed in the following Section however, is which of the three tracks identified in *Figure 4* is the most suitable for programming EAI’s?

#### *IV. DISOBEDIENCE AND NUCLEAR LAUNCH*

This Part begins by introducing the Authors’ proposed “test,” or, the Robot Rules of Engagement (“RROE”). As previously noted, in this instance the RROE are tailored to address the concept of “refusal.” The test is based upon a “system of systems” approach, where each system represents an individual task undertaken by an EAI. Once these RROE are established, Part IV goes on to consider a number of tangible scenarios to determine how the RROE might operate in practice. The fourth and final of the envisaged scenarios is the instance of an EAI questioning the human decision of nuclear launch. Currently, an order for nuclear launch goes simultaneously to both central command and the team in the field, making recall difficult.<sup>148</sup>

---

147. Whether the order is consistent with the concept of human dignity, human rights, public liberties, morals, the conscience of reasonable people, and even whether the order is of service to the state (which might be referred to as being consistent with policy), and finally, whether the soldier has a reasonable chance of executing the order successfully (reflected by the wide obligation).

148. See GRAY, *supra* note 3. At least strategically speaking, a nuclear strike with a single warhead would in many cases be unwise. The point is, that if an ambiguous, or

Indeed, there could be the curious case of one part of the launch team wishing to disobey the order to launch, but because that original message goes to both, recall and override is no longer possible.

*A. Additional Test for Safeguarding Decision-making.*

Before moving on to consider the scenarios, this Section introduces the RROE. In short, this “test” calculates human error in order to prevent certain orders from being carried out. This has the effect of reducing Clauzewitizian friction, though the Authors acknowledge that it cannot do so all of the time. The EAI examination of orders is seen as an “additional” way for ensuring DTP obligations are met. The RROE are grounded in the discussion in Parts II and III of this analysis, and in particular to the three obligations identified in *Figure 4*. For the reasons in the following paragraphs, the RROE take elements from all three tracks. This is for a variety of reasons, but not least because, as previously noted, the central premise of IHL is to provide a system of “checks and balances . . . aimed at minimizing human suffering without undermining the effectiveness of military operations.”<sup>149</sup> IHL is therefore, a compromise,<sup>150</sup> and any usable test cannot be weighted too heavily upon one concept or the other.<sup>151</sup>

With that in mind, Track 1 is considered to be too strict for the following reasons. First, excessive weight should not be given to the wider mission objectives, and certainly not at the sacrifice of humanitarian considerations. Although the Authors agree that a soldier should be aware of the means and methods at their disposal, and the likelihood of needing to utilize them in the future to give them better effect (humanitarianly and militarily), each analysis should nevertheless attempt to minimize civilian harms where practicable or practically possible. This is arguably a

---

erroneous order was sent to a large number of launch sites, it would clearly be more difficult to prevent launch at several sites (if that was the desired course of action), than if an order that was sent to a single location.

149. DINSTEIN, *supra*, note 23, at 9.

150. *Id.* at 10.

151. *Id.* ¶ 9 (“[i]f military necessity were the sole beacon to guide the path of armed forces in wartime, no meaningful constraints would have been imposed on the freedom of action of Belligerent Parties . . . [but] . . . [i]f benevolent humanitarianism were the only factor to be weighed in hostilities, war would have entailed no bloodshed, no human suffering and no destruction of property; in short, war would not be war.”).

stricter application than to do so only where there is a reasonable availability, and with specific regard for the wider mission objectives.<sup>152</sup> EAI will be capable of effectively determining whether the order is lawful, but also whether it is consistent with the political and strategic doctrine that is contained within ROEs. These often impose greater restrictions than IHL alone requires,<sup>153</sup> and may help to ascertain whether an order is of “*use for service*” as is required, for example, by German doctrine.<sup>154</sup> If these more “precise” guidelines are programmed into an EAI, they help to identify when an order should be refused, for example, because it was provided by an individual acting out of revenge.<sup>155</sup>

Nevertheless, an entirely Track 3-based approach is also unsuitable. In the first instance, this is because a soldier must have at least *some* regard to the wider mission objectives. For example, an infantryman might carry two grenades—one non-autonomous and the other a form of AWS capable of independently carrying out proportionality assessments before detonating. Given this scenario, a combatant may wish to reserve the latter weapon with the knowledge that they were about to enter a more densely populated locality where the advanced tech might help to save civilian lives. However, strict adherence to Track 3 would prevent

---

152. In addition, due to the fact that when programmed to carry out a specific task, EAI will not “forget.” However, this is not always the case when an EAI or AI is “repurposed.” As is discussed in the following Section, for example, one of the difficulties that AI experts currently face, is that when a self-learning system is given a new task, i.e., learning to play the ancient Chinese strategy game “Go,” instead of the one it has been playing, i.e., Chess, the system, placing all of its “attention” upon the former, forgets how to play the latter. The concept is referred to as catastrophic forgetting. See Anthony Robins, *Catastrophic Forgetting, Rehearsal, and Pseudorehearsal*, 7 CONNECTION SCI.: J. NEURAL COMPUTING, A.I. & COGNITIVE RSCH. 123, 123-46 (1995).

153. Corn & Schoettler, *supra* note 19, at 821-22. Indeed, future ROE are likely to contain vital information on joint-force tactics and operation procedures regarding the introduction of AI, and how it will affect all services, across all domains. See also Modern War Inst. Podcast, *Competition, Conflict, and the Future of Irregular Warfare*, MODERN WAR INST. (July 22, 2020), <https://mwi.usma.edu/mwi-podcast-competition-conflict-and-the-future-of-irregular-warfare/> [<https://perma.cc/G989-CCN9>].

154. ICRC Practice Relating to Customary Rules, *supra* note 52, at 3802 (citing GERMAN MINISTRY OF DEFENCE ET AL., HUMANITARIAN LAW IN ARMED CONFLICTS MANUAL § 142).

155. This might be, for example, where an EAI identified that an order was legal, but was otherwise inconsistent with ROE regarding certain targets. See Corn & Schoettler, *supra*, note 19, at 822 (noting, for example, that ROEs commonly reflect policy that imposes “restrictions on combat power above and beyond LOAC . . . [and that] . . . are often adopted in hopes of avoiding alienation of the civilian population.”).



this. Track 3 also fails to satisfy, because the obligations it carries are excessively conceptual.<sup>156</sup> For example, the ICJ provides that Human Dignity is “[t]he essence of the whole corpus of IHL as well as human rights law . . . .”<sup>157</sup> Nonetheless, as an independent concept, human dignity is undefined, and unquantifiable. The German court cited the right to freedom of conscience under German Basic law as a dignity “enabler.”<sup>158</sup> Nonetheless, the concept of human dignity must go much deeper than that. Indeed, as posited by one author, a good deal of “contemporary ideas about the role of international law are grounded on a very misplaced notion of what human dignity is.”<sup>159</sup>

An in-depth investigation into this matter is well beyond the scope of the present Article. However, the Authors believe that it would be particularly problematic to include this element of the EAI obligation for two primary reasons. Although somewhat controversial, the first of these is that both objects and people can ultimately be assigned values. They can, therefore, be expressed or measured as a quantity and transferred into programmable code. Indeed, this would *have* to be the case if machines were ever to successfully carry out distinction and proportionality assessments.<sup>160</sup> This would be very difficult to achieve with a conceptual provision such as human dignity. Nonetheless, the overriding issue here is not that it would be difficult to “upload” the concept, but that “[w]ar itself takes a toll on human dignity through the intentional sacrificing of lives to achieve military objectives.”<sup>161</sup> In other words, war is undignified.

---

156. Though one might argue that if a machine was able to display the “nuanced reasoning” that is discussed, for example by Krupiy, *supra* note 13, it could potentially apply Track 3. See *infra* note 162 (discussing the possibility of introducing an ethical governor).

157. ICTY Prosecutor v. Furundzija, Case No. IT-95-17/1-T, ¶ 183 (Int’l Crim. Trib. for the Former Yugoslavia Dec. 10, 1998).

158. See Limits of Obedience to Superior Orders Case (2005).

159. Jeremy Rabkin, *What We Can Learn about Human Dignity from International Law*, 27 HARV. J.L. & PUB. POL’Y 145, 146 (2003).

160. See, e.g., Sassoli, *supra* note 18, at 327 (“A robot must be able to sense all the necessary information in order to distinguish between targets in the same manner as a person.”).

161. Adam Saxton, *(Un)Dignified Killer Robots?: The Problem with the Human Dignity Argument*, LAWFARE (Mar. 20, 2016), <https://www.lawfareblog.com/undignified-killer-robots-problem-human-dignity-argument> [<https://perma.cc/5U77-XQRY>].

Consequently, the Authors propose the RROE need to have particular regard to IHL, and to the relevant ROE. Future doctrine will almost certainly cater for EAI, including the circumstances in which they should (or should not) be used. In short, used correctly, military doctrine will provide a vital and detailed EAI *modus operandi* that will guide decision-makers, and ensure IHL compliance. Moreover, doctrine can be, and in many cases is, written with due regard to human dignity—or a wider ethical viewpoint—which ensures these are not simply tossed aside.<sup>162</sup>

The benefit of running orders through a wider set of obligations than IHL alone, is that in many cases an EAI will be able to correctly determine whether an order is legally compliant, whether it is consistent with matters of national policy, and, consequently, whether the order is best practice given the circumstances.<sup>163</sup> In addition, it can help to determine whether an order-giver is motivated by something other than a legitimate military reason, such as revenge or fear. Consequently, an EAI applying this analysis will be able to go much further than merely identifying orders which shock the conscience of every reasonable, right-thinking person,<sup>164</sup> or, of which a person of reasonable standing should have known was an unlawful order. A military EAI system can, and should, in other words, operate as though it was a military lawyer.

As a solution, the Authors propose the following: The RROE should include a number of “systems” which constantly analyze the status of an order. System 1 of this “system-of-systems”: an authentication step. This is a vital phase, that is particularly relevant when an order is received in the form of code (cyber),

---

162. Although the current authors do not intend to support the current discussion on such a believe, one author suggests it will one day be possible to programme future EAI's with an “ethical governor,” which will allow for them to display moral judgement when making decisions. See RONAL ARKIN, *GOVERNING LETHAL BEHAVIOUR IN AUTONOMOUS ROBOTS* (2009).

163. The point here is that the greater protection that is offered by ROE is typically subject to policy. See, e.g., Corn & Schoettler, *supra*, note 19, at 822 (arguing that “a ROE-based decision to forgo an attack, even if motivated by an effort to mitigate civilian risk, is not legally mandated. This fact is an important aspect of contemporary military operations, and the difference between legal and policy-based courses of action should be constantly emphasized.”).

164. See ICRC Practice Relating to Customary Rules, *supra* note 52, at 3800 (citing Canadian Code of Conduct §5, Rule 11 (2001)).

or where the EAI monitoring the order is doing so remotely. Where System 1 fails to authenticate an order, System 4A is engaged: this System requires one of three suitable responses: (I) inform the order-giver of the observation, and allow a similar order to be given (“Passive Refusal”); (II) immediately reject the order, but allow similar orders to be given (“Active Refusal”); or (III) immediately reject the order and put safeguards in place to prevent similar orders being issued (“Preventative Refusal”). Perhaps the most likely step in the first instance would be to refuse and seek further clarification. This is therefore, active refusal. However, this might be affected by a number of factors, not least an order’s gravitas.

Where an order is authenticated, System 2 is engaged. This is comprised of three independent micro-systems—with each of them being assigned a fundamental IHL calculation. System A runs a continuous distinction assessment, System B is a continuous dynamic proportionality assessment, while System C analyses whether all practicable or practically possible precautions have been taken to minimize civilian harms. Where one or all of these micro-systems identify that the continuation to follow the order would breach IHL, System 4A is engaged. This is, of course, subject to options noted above. In certain instances, a System 1 and 2 assessment will be sufficient. However, once a System 2 analysis is complete, and there is no cause to intercept the order, System 3 would engage (while noting that System 3 continuously cross-references Systems 1 and 2 one and two—which also run a continuous assessment loop).

Once System 3 is engaged, an order is considered subject to the criteria previously identified —i.e., according to LOAC,<sup>165</sup> IHRL, other relevant treaties (including regional agreements), and national doctrine.<sup>166</sup> And, it is this examination that will

---

165. Which would, of course include the other obligations contained within Article 57 API, such as the obligation to attack the target that will cause the least amount civilian harms in instances where two or more targets offer a similar military advantage. Though a nation may have to be party to the treaty if the provision is not considered customary in nature. *See, e.g.*, SCHARRE, *supra* note 92, at 50-51.

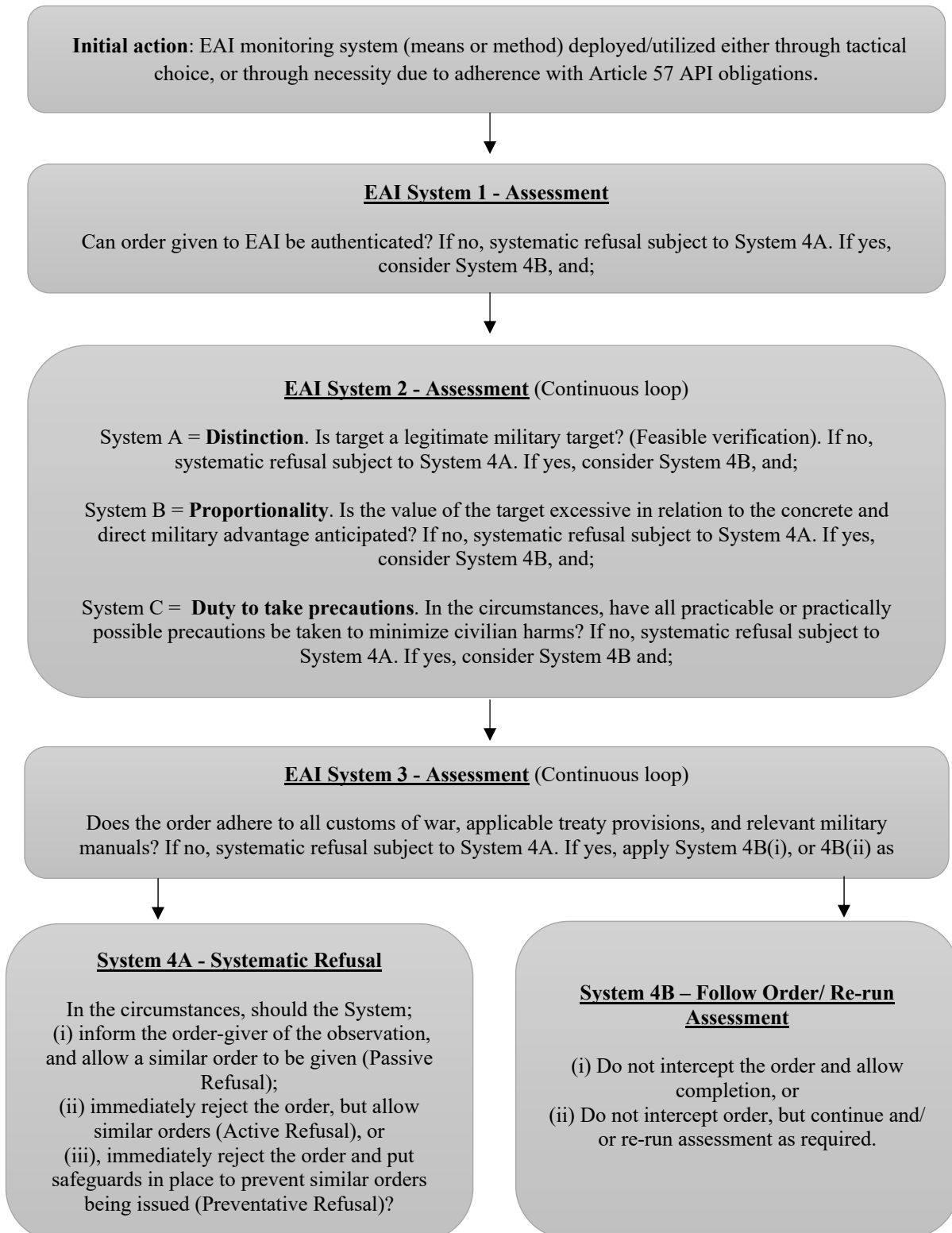
166. The latter of which is likely to include information regarding the interpretation of IHL. *See* Corn & Schoettler, *supra*, note 19, at 822 (identifying, for example, that the U.S. DEP’T OF DEFENSE, LAW OF WAR MANUAL 241 (2015) states that the United States does not consider Article 57(3) API to be customary in nature). In addition, and by way of offering a caveat, while the Authors have proposed that the test

determine whether the instruction is consistent with this wide body of obligations, and/or, whether it is beset, for example, with motivation for revenge, self-preservation, or the commission of a war crime. Depending upon the type of order—i.e., whether it contains a single instruction or multiple instructions—this may be repeated on a perpetual loop. Where there is cause to intercept the order, System 3 will engage System 4A, subject to the caveats previously considered. However, where there is no grounds to intercept an order, System 4B is engaged. This final stage either completes or implements the order, or repeats the entire assessment so long as necessary to allow for completion. These four stages are shown on graphical representation below”

---

should carry certain boundaries, in practice any nation choosing to apply such a test would do so according to which ever model suited their particular state practice—this could be more akin to either the narrow or the wide model previously identified.

Figure 5. EAI - 4 Stage Assessment of Orders



*B. Scenario 1*

With reference to *Figure 5* above, the following Section considers a number of scenarios to see how the RROE might apply to various orders. This analysis begins with the following scenario, which intends to demonstrate how existing, often rudimentary technology, lacks the ability to apply “post-human assessment.” An anti-personnel mine is an example that often appears in the conversation surrounding AWS, and it can also provide an ideal point of focus here.<sup>167</sup> Existing references to anti-personnel mines commonly highlight the fact that once in position there is no further human involvement. In essence, the weapon displays a basic level of autonomy, “deciding” itself when force should be applied. The “decision” to detonate (or not), is based entirely upon whether a pressure threshold is surpassed when an individual steps on the pressure plate or when the trip wire is snagged.<sup>168</sup> In the case of an anti-personnel mine therefore (regardless of any basic autonomy of the weapon itself), a human will make decisions as to whether its use is an appropriate means or method, whether it is otherwise lawful,<sup>169</sup> and that deploying the munition in a certain way is identified as acceptable

---

167. See, e.g., SCHARRE, *supra* note 92, at 50-51.

168. See, e.g., Kevin Bonsor, *How Landmines Work*, HOW STUFF WORKS, <https://science.howstuffworks.com/landmine.htm> [https://perma.cc/UY3Y-RV87] (last visited Dec. 15, 2020).

169. According to all relevant legal provisions, including in particular: the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on their Destruction (“Ottawa Treaty”) Mar. 1, 1999, 2056 U.N.T.S. 211; Protocol II, annexed to the Convention on Prohibitions and Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to be Excessively Injurious or to Have Indiscriminate Effects, Dec. 3, 1998, 2048 U.N.T.S. 93; ICRC Customary Rules, *supra* note 32 (citing ICRC Customary Rule 81: “[w]hen landmines are used, particular care must be taken to minimize their indiscriminate effects . . . State practice establishes this rule as a norm of customary international law applicable in both international and non-international armed conflicts. This rule applies to the use of anti-vehicle mines. It also applies in relation to anti-personnel landmines for States which have not yet adopted a total ban on their use.”).

practice.<sup>170</sup> In fact, this is true of nearly all existing munitions.<sup>171</sup> While an anti-personnel mine could be considered a “basic” EAI, there is no method for the robot to “refuse” to detonate.

### C. Scenario 2

While Scenario 1 identified that “basic” autonomy has no real choice in terms of refusing an order, this second scenario considers more advanced technology. Here, particular emphasis is placed upon weapons platforms. Indeed, most militaries currently deploy rudimentary EAI platforms—many of them with various AI systems on-board. Radar, for example, and other such detection systems, enable the modern-day fighter jet to independently identify potential threats, “lock on,” and provide a pilot with the option of authorizing launch. A human still decides upon the most suitable means of attack, though in reality their “choices” are likely to be fairly restricted. One such option is likely to be an air-to-air munition, which, if self-guiding, is arguably an EAI. Nevertheless, once the human operator

---

170. As the Authors are currently examining elsewhere, certain nations may wish to indoctrinate military manuals, ROE, and other guidance to “ensure” that EAI ensure that certain orders are followed, as opposed to refused. This may be the case, for example, where decision-making responsibilities are not generally delegated to the lower echelons of the military hierarchy. This may perhaps be due to issues with trust, but the point is, an EAI could potentially ensure orders that are very prescriptive in nature are followed closely. For a useful discussion, see Franz-Stefan Gady, *Future Warfighting: Placing Doctrine Before Technology*, SOUNDS STRATEGIC 29:09-33:15 (Jul. 20, 2020), <https://www.iiss.org/blogs/podcast/2020/07/future-warfighting> [<https://perma.cc/C594-7ZT2>].

171. Clearly, the central argument contained in the present thesis is that this will not always be the case. EAI tech is improving all the time. Take for example the U.S. Long Range Anti-Ship Missile (“LRSAM”) which, by utilizing an “intelligent navigation and direct route” feature, is already able to carry out rudimentary distinction assessments. See *Long Range Anti-Ship Missile*, LOCKHEED MARTIN, <https://www.lockheedmartin.com/content/dam/lockheed-martin/mfc/pc/long-range-anti-ship-missile/mfc-lrasm-pc-01.pdf> [<https://perma.cc/46E3-XSMJ>] (last visited Jan. 27, 2021). Future technology may improve upon this in a number of ways. For example, as previously discussed, the weapon itself may contain technology that allows for it to carry out its own complex assessments. Or, an EAI could act as a commander, or even as a kind of military police officer that is used to “enforce the law.” In such an instance, nations would clearly have to decide upon, and programme their EAI accordingly, with the necessary codes of conduct regarding matters such as the issuing of cease and desist commands, and guidance on the use of force when arresting and detaining “friendly” soldiers.

authorizes the application of force, there is usually no further “interference” from the platforms systems.

However, if a pilot was involved in a “dog-fight,” and the platform they inhabited was programmed with the Authors’ RROE, the radar would similarly identify an enemy aircraft and alert the pilot accordingly. The pilot may then choose to fire (or not), based upon their assessment of the prevailing circumstances.<sup>172</sup> If they, or perhaps the operator of a combat drone,<sup>173</sup> decided to engage, System 1 would commence. Authentication in the circumstances would likely be satisfied when the pilot “logged in” at the start of their “shift.” And, System A would presumably verify the target. System B would then engage, and carry out an additional proportionality assessment to that of the pilots. Though arguably, in almost all instances, this would align with the pilot’s choice to deploy the munition. However, the EAI might be aware of extraneous circumstances (such as ground troops that at that moment were at risk of being hit by debris, or perhaps a high risk to the civilian population) that the pilot had not factored in due to inclement weather or to

---

172. One additional discussion that could be had here is what part of the system was giving an order to fire, as opposed merely proving a direction? In other words, could authorization be delegated to a machine so that it could “indicate” that it had calculated the following attack, offering a definite military advantage, that it is lawful, ethical, and very likely to positively influence the course of the battle—Take the shot! Arguably there could be consequences if the human pilot failed to follow such an “order”—because he/she thought it better in the circumstances—perhaps only due to gut instinct. However, what if, due to not taking the shot, lives were lost, or ultimately, the battle was lost. Could/should it lead to a court martial, or criminal proceedings? Or, should it be of no consequence because the human pilot/operator used their “superior” (or hierarchical) judgement? In other words, should the “plane” be assigned a higher rank? This is the cyclical argument; the problem with threat of criminal sanctions is that the EAI does not change the status quo. EAI should be an improvement of the current balance of being able to refuse an order only in very limited circumstances, juxtaposed with military conditioning and potential criminal charges for insubordination.

173. Note that UAVs are becoming much more capable than early Mk I Predator Drones, for example, which were first used by the United States in the 1990s. Initially these were used only used for reconnaissance missions. For an Mk I fact sheet, see *MQ-1B Predator*, U.S. AIR FORCE (Sept. 2015), <https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104469/mq-1b-predator/> [<https://perma.cc/FS8X-2MB8>]. In contrast, rather than operating in relatively uncontested airspace the developers of contemporary UAV are looking to them replacing existing piloted air-to-air combat aircraft. See, e.g., Valerie Insinna, *Boeing Rolls Out Australia’s First ‘Loyal Wingman’ Combat Drone*, DEFENSE NEWS (May 4, 2020), <https://www.defensenews.com/air/2020/05/04/boeing-rolls-out-australias-first-loyal-wingman-combat-drone/> [<https://perma.cc/F5RE-RXLD>].



a navigational error. Of course, each situation is highly contextual and there would not always be a strict obligation to *not* carry out the attack. Nevertheless, where appropriate, System 4A would be engaged. In this instance, passive refusal (4A(i)), or active refusal (4A(ii)) would be the most appropriate option, there not appearing to be a need for preventative refusal (4A(iii)) at this stage.

If System 2 did calculate that distinction and proportionality were initially satisfied, Micro-System 2C would engage. This micro-system identifies whether the means and method were appropriate, and that their selection would, as much as is practicably possible, minimize civilian harms. If the Micro-system 2C assessment identified a grounds for refusing the order, System 4A is engaged—subject to caveats. If, however, the three components of System 2 *were* satisfied, the instruction would not be intercepted and the munition would be deployed (as per System 4B). Due, *inter alia*, to the systems' speed of operation, the speed of the tactical pursuit, and to the battlefield conditions, a System 3 assessment might be unnecessary in such circumstances. However, if System 3 was engaged, the EAI would then consider other matters such as any relevant ROE. These would go further than legal obligations, and might identify, for example, whether the platform was operating in a pre-determined “no-fly zone” or whether there were any other policies or local level restrictions in place. Under such circumstances, the EAI might engage System 4A, and refuse to follow the order.

While there may be a slight increase in the risk faced by the pilot, the introduction of the multi-stage test is a positive method for reducing friction, in much the same way as an anti-stall mechanism.<sup>174</sup> And, if the RROE had been installed on the USS Vincennes in 1988, it may, for example, have detected the friend or foe signal that was being sent out by the civilian aircraft Iran Air Flight 655.<sup>175</sup> As a result the missile may not have been launched, and the civilian aircraft not destroyed.

---

174. Note that with the increased use of unmanned technologies in armed conflict, and the possibility of further advances in fully autonomous tech, it is arguable that such systems could and perhaps should hesitate before deploying force, due entirely to the fact that no human life is in direct risk. *See, e.g.*, Sassoli, *supra* note 18, at 327-28.

175. *See* SCHARRE, *supra* note 92, at 169-70 (discussing the facts of the USS Vincennes and that fateful incident). Similarly, the same test may have also prevented

Returning to Scenario 2 however, if “Track 1” was followed (as opposed to the test considered above), the results would differ for a number of reasons. This may be due to the decision-maker not believing it “reasonable” to deploy an aircraft fitted with an EAI system in the first place. In addition, even where a decision monitoring EAI was utilized, Track 1 only requires an order to be refused where it was a manifestly unlawful order. As a result, with distinction and proportionality satisfied, Track 1 would not intercept, regardless of whether there were political, or wider tactical or operational reasons not to. Similarly, for reasons previously discussed, though “Track 3” systems would require the EAI to be deployed in all circumstances in which it was available, the EAI itself would be obligated to consider a number of conceptual matters, including the dignity of the pilot and the target, and of any civilian that was included in the proportionality assessment.

#### *D. Scenario 3*

The following scenario is used to extend the analysis beyond the human decision to use a particular type of weapon (means), to a robot decision to use a particular tactic (method) to reach certain ends. For the purpose of Scenario 3, let us suppose the “end” is to obtain information from an “adversary.” There are, of course, many ways in this scenario could be presented. Nevertheless, the following is proposed: an AWS, in the form of a humanoid combatant, receives an order from an EAI located in a command-and-control center. The AWS is to “secure a block” in which a large number of enemy combatants are located, and to retrieve, “by any means necessary” information regarding a suspected future attack. Here, System 1 authentication is likely to be fairly straightforward with sensors and additional programs (including destination verification) capable of authenticating the status of the order-giver. Nevertheless, the analysis must be divided in two, given that there are two independent orders. The first is to secure a location. Arguably, here, the AWS analysis is

---

the recent shooting down of Ukraine Air Flight 752 by Iranian armed forces on January 8, 2020. See Michael Safi, *Iran Admits it Fired Two Missiles at Ukrainian Passenger Jet*, *GUARDIAN* (Jan. 21, 2020), <https://www.theguardian.com/world/2020/jan/21/iran-admits-it-fired-two-missiles-at-ukrainian-passenger-jet> [<https://perma.cc/52Q5-4XGN>].

similar to that considered in Scenario 2. Should System 1 and System 2 be satisfied, System 3 would engage and determine whether the order is consistent with IHL and the more focused ROE. The same arguments would be applied in respect of Track 1 and Track 3 application as considered above, and to avoid repetition, there is not a pressing need to consider the entire RROE again at this point.

However, in this scenario the System 3 assessment is key with regard to the second order to obtain information, because there clearly an element of ambiguity—how does the order-giver define “any means necessary”? There are a number of options, the most obvious being that clarification needs to be sought. However, if this was not possible there is a choice of two primary interpretations. In the first instance, it might be taken to mean by any means necessary; providing that they are consistent with international legal obligations. If this was the case, then a similar assessment would be conducted to that already considered. However, a second way in which this instruction might be implemented is—by any means necessary (regardless of international obligations). With regard to the latter, the most appropriate retort would be that where a customary duty exists for a human soldier to disobey a manifestly unlawful order, it must also apply to an AWS. As a result, an AWS, like its human equivalent, must, for example, refuse to follow an order to gain information with the use of torturous methods.<sup>176</sup>

A question that arises, however, is what if there is military doctrine to the contrary? This occurred, for example, in the United States, under the Presidency of George W. Bush. In this

---

176. See Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (“Convention Against Torture”) art. 1, June 26, 1987, 1465 U.N.T.S. 85 (defining torture as: “any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person for such purposes as obtaining from him or a third person information or a confession, punishing him for an act he or a third person has committed or is suspected of having committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such pain or suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising only from, inherent in or incidental to lawful sanctions.”). See also ICRC Customary Rules, *supra* note 32 (ICRC Customary Rule 30 notes that Article 8(2)(a)(ii) of the ICC Statute removes the need for the suffering to be inflicted “by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity.”).

instance, the United States attempted to greatly reduce the definition of torture, and in doing so justify the use of certain cruel, inhuman, or degrading acts that it had carried out. The United States argued that these acts did not violate its international obligations, because they claimed that the acts did not meet the requisite intensity of pain and suffering.<sup>177</sup> The result, as argued by at least one commentator, was that despite the fact that Articles 1 and 2 of the Convention Against Torture<sup>178</sup> placed certain restrictions upon military interrogators, the CIA was authorized “to confine and interrogate detainees with a harshness that markedly violated human rights.”<sup>179</sup>

In the present scenario, if torture was authorized, and indeed ordered, arguably it would be refused by any AWS/EAI, whether it applied the Track 1, Track 2, or Track 3 approach. Of course, a decision-maker employing the former may not consider it reasonable to utilize an EAI for this type of assessment in the first instance. Nevertheless, given that military manuals typically provide a great deal more detail regarding a soldier’s obligations than that which is provided by IHL alone, it might also be argued that the AWS would have an obligation to follow the order if the ROE appeared to reflect recent changes in national policy or best practice. In such a situation however, the Authors propose that because EAI/AWS are a “method” of ensuring compliance with international obligations, they *must* reflect the hierarchy of sources. Thus, military doctrine should be used to support and interpret treaty provisions, and not as a method to negate or lessen their effects. And, when this principle is applied to the

---

177. Richard D. Rosen, *America’s Professional Military Ethic and the Treatment of Captured Enemy Combatants in the Global War on Terror*, 5 *GEO. J.L. & PUB. POL’Y* 113, 118 n.31 (2007) (citing Memorandum from Jay S. Bybee, Office of Legal Counsel, to Alberto R. Gonzales, Counsel to the President, Subject: Standards for Interrogation under 18 U.S.C. §§ 2340-2340A (Aug. 1, 2002)).

178. Article 2 of the Convention Against Torture states: “1. Each State Party shall take effective legislative, administrative, judicial or other measures to prevent acts of torture in any territory under its jurisdiction. 2. No exceptional circumstances whatsoever, whether a state of war or a threat of war, internal political instability or any other public emergency, may be invoked as a justification of torture. 3. An order from a superior officer or a public authority may not be invoked as a justification of torture.” Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment art. 2, Dec. 10, 1984, 1465 U.N.T.S. 85.

179. Robert Bejesky, *Pruning Non-Derogative Human Rights Violations into an Ephemeral Shame Sanction*, 58 *LOY. L. REV.* 821, 824-25 (2012).

present particulars, system 2 would either prevent the AWS from carrying out a tortuous act, or, where an EAI is operating in a supervisory mode, prevent a human subordinate from doing so.<sup>180</sup>

*E. Scenario 4*

The previous two scenarios have considered situations where an EAI might be utilized to assess an order, and where necessary, to refuse to follow it. In either case the analysis concluded that by applying the RROE, refusal can take place due to an order failing to adhere to the fundamental principles of IHL (or the wider body of applicable international law), or, where an order is otherwise inconsistent with matters of national policy, or is, in other words, not of use for service. In this final and undoubtedly most extreme “application” of “robot refusal,” the Authors seek to consider the practicalities and logistics of refusal in the context of nuclear launch.<sup>181</sup>

Ostensibly, an autonomous “operating system” could ultimately override a launch instruction, whether that originated from a human or an EAI. As a result, while the scenario that follows considers a set of prescribed “facts,” the relationships are interchangeable, and not strictly limited to those considered. The EAI could, for example, be interchanged with an AWS, or a human could be swapped out for an AWS. However, at some point, the order must be run through an EAI. In direct contrast to Scenario 2, whereby the acquiring of missile lock is more of a strategic assist, for the EAI in this fourth scenario (preventing unlawful nuclear launch as a result of system of systems) is a “legal” assist.

The scenario considered has the following characteristics, with the caveat that the act in question is scrutinized under the *jus in bello*, as opposed to the *jus ad bellum*.<sup>182</sup> State A is at war with

---

180. Subject to the caveats identified in *supra* note 171.

181. Note that an investigation into the extent of the soldier’s duty to refuse to obey a manifestly unlawful order to launch a nuclear weapon has already been carried out by at least one author. Anthony J. Colangelo, *The Duty to Disobey Illegal Nuclear Strike Orders*, 9 HARV. NAT’L SEC. J. 84 (2018). The purpose of the current discussion is to consider how an EAI might help to ensure that duty is adhered to.

182. The point being that the circumstances that are presented in the following scenario are intended to be examined under DTP, and not in relation, for example to Articles 2(4) and 51 of the UN Charter, the latter of which the Authors are in the process of addressing in an accompanying piece.

State B, both of whom are nuclear powers. In the midst of this “peer-on-peer” conflict, the President of State A (X) authorizes a nuclear attack. Following current strategic command principles, this order goes to strategic command, and to a central war office. The order to initiate the attack is then forwarded on to submarine commander (Y), and to General (Z), who oversees two or more additional launch sites. Shortly after the instruction to initiate the attack is received by Y and Z, *all* communications are lost.<sup>183</sup> The reader may at this stage recall such a scenario being captured in the 1995 motion picture, “Crimson Tide.” Here, Denzel Washington’s character wishes to re-establish radio communications to determine whether a further but unreadable message from strategic command has overridden a previous nuclear launch order. In contrast Gene Hackman’s character portrays the somewhat more belligerent X who believes that the initial order for nuclear launch should be followed without question.

In terms of the present scenario (and not an appraisal of the motion picture), the Authors’ system of systems requires System I authentication. This is clearly a vital stage, but is also one at which the potential for EAI monitoring could be introduced. Currently, looking to US nuclear doctrine, the president carries a personal identification tool, which contains a code that is unique to the president—known colloquially as the “biscuit.”<sup>184</sup> When necessary, this code can be entered into an authorization system that is carried by a constantly rotating military presence, and which accompanies the President at *all* times. This system is known as the nuclear, or atomic, “football.”<sup>185</sup>

In the first instance, these “biscuits” and “footballs” ensure that the order can be authenticated. Potentially, for the sub

---

183. It could, for example, be due to the command and control suffering an overwhelming attack, or due to a devastating cyber-attack.

184. See Michael Dobbs, *The Real Story of the “Football” That Follows the President Everywhere: Take a peek at the mysterious black briefcase that has accompanied every U.S. president since John F. Kennedy*, SMITHSONIAN MAG. (Oct. 2014), <https://www.smithsonianmag.com/history/real-story-football-follows-president-everywhere-180952779/> [https://perma.cc/Y9GQ-7ZFX]; Kirsten Korosec, *Trump Doesn’t Have a Nuclear Button. He Has A Football and a Biscuit*, FORTUNE (Jan. 3, 2018), <https://fortune.com/2018/01/03/trump-doesnt-have-a-nuclear-button-he-has-a-football-and-a-biscuit/> [https://perma.cc/3VCX-TM8V].

185. *Id.*

commander, this might implicitly engage System 2. However, a central question here is, could, or indeed should the “football” be an EAI? If it were, and the order was erroneous, it could be prevented from even making it through to strategic command. Of more general application, the Authors wish to underline that EAI analysis is at its most beneficial the higher up the command chain it takes place (at least in the first instance).<sup>186</sup> And, the football might also be pre-programmed with the correct strategic response—given the circumstances—leading to faster decision making. However, if the decision monitoring EAI were not the “football,” the order could still be passed to a Strategic command or Central War Office EAI, where the order would face the same scrutiny.<sup>187</sup>

Once an erroneous order has passed to individuals Y and Z, the task of intercepting it is likely to become more difficult. Indeed, when such an order is passed from one to two (or more parties), it might result in the slightly curious case where one part of the launch team disobeys the order, while another implements it without hesitation. Arguably, in this instance, recall would be particularly problematic, and probably all too late.<sup>188</sup> This could be prevented with the use of an EAI. EAI assessment could provide an additional safeguard at whichever stage it is considered, but also, perhaps the more often it is utilized.<sup>189</sup> Therefore, although the EAI analysis that follows is largely considered at the launch level, the stages of assessment could be considered at any point during the lifetime of the order.<sup>190</sup>

All System 2 assessments must clearly be made with due respect to the huge devastation potential that nuclear missiles carry. However, given a particular set of circumstances, a nuclear weapon could potentially “target” a military installation in

---

186. This is of course true whether considering the scenario with or without the introduction of EAI technology. Colangelo, *supra* note 181, at 90.

187. Indeed, in the United States there is a “two-man rule,” which, in short, means that the order to launch a nuclear attack is verified by two separate individuals, at every stage in the chain of command. *Id.* at 114-15. It is at least conceivable that an EAI could replace one of those humans at every stage of the process.

188. *See supra* note 148.

189. Though this is, of course, subject to the condition that the tech each stage is functioning correctly. Nevertheless, because an EAI cannot authorise force, the impact of malfunctioning equipment is lessened, as compared, for example, to AWS.

190. *See supra* note 187.

adherence with the principle of distinction. With regards to proportionality however, System 2B might need to be satisfied on the condition that the launch, and all of the anticipated collateral damage, has been authorized at the strategic level. Existing safeguards are no doubt in place for instances where the nuclear order-giver suffers from temporary insanity, or where they choose to dispense an order due to ill health, blackmail, inducement, treachery, etc. However, at the sub level, this would be difficult to determine. Nevertheless, where there was no evidence of a legitimate military target, or, there was evidence to suggest that the target was the “civilian population,” the EAI could, without fear for its job, its societal standing, its life, and the lives of its “nearest and dearest” etc., refuse to follow the order.<sup>191</sup>

Where Systems 2A and 2B were satisfied however, System 2C would engage. Given the gravity attached to a nuclear launch, one might posit that the standard for target verification should reflect something greater than mere “reasonable” steps, and perhaps, in reality, this is likely to be the case. Nonetheless, for an EAI system programmed according to Track 1, reasonableness would remain the minimum requirement. The matter of whether feasible precautions had been taken (or not) would, once again, be highly context dependent. If, for example, an EAI was stationed with the Commander and had, prior to launch and loss of comms, evaluated that state A had attempted all other appropriate means and methods of defeating the enemy, the order to launch would not be intercepted.

However, if the EAI applying the RROE calculated that an alternative means or method would yield a similar result, it should intercept the order and prevent launch, at least until more information regarding the status of the order could be gathered. In contrast, an EAI operating Track 1 would arguably not intercept the same order, due to the fact that so long as distinction and proportionality were satisfied, and the order was “reasonable” in the circumstances, there appears to be no evidence of a manifestly unlawful order. This standpoint is in

---

191. Indeed, as identified by Colangelo, *supra* note 181, at 92, due to the unique destructive nature of nuclear weapons, “where conventional weapons can be used in proximity to civilians to achieve the same or similar military objectives . . . an order to use a nuclear weapon would be manifestly illegal and anyone with sufficient factual knowledge regarding the circumstances of the order should know it.”



itself contrary, to the Track 3 requirement to always use the means and method that minimizes civilian harms, and to carry out any analysis subject to factors such as whether there is a reasonable chance of mission success. Due to the circumstances and the chances of initiating a nuclear apocalypse, this is as the very least questionable.

Where RROE is utilized, however, where System 2 does not identify grounds for refusing an order, System 3 would be engaged. And, it is here that RROE offer the most focused assessment when compared to either of the two extremes. For example, as noted, an EAI applying Track 1 would *only* deny manifestly unlawful orders. This is vital because, nuclear weapons do not distinguish between combatant and civilian. Moreover, a legitimate military target that was positioned close to a densely populated urban area would mean that there would be a considerable level of collateral damage (not to mention the comprehensive costs of all out nuclear war).

With that in mind, it is arguable that a nuclear attack would always be disproportionate in IHL terms. However, in providing their *Nuclear Weapons Advisory Opinion*, the ICJ identified this may not always be the case.<sup>192</sup> Consequently, any “legitimate” order to initiate a nuclear launch *could* be a lawful order, and if this was the case in this scenario, the Track 1 analysis would not intercept. In contrast (except, perhaps, where all other means and methods

---

192. Nuclear Weapons Advisory Opinion, *supra* note 29, ¶ 97. See also MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT § 6.17 (2004) (“There is no specific rule of international law, express or implied, which prohibits the use of nuclear weapons. The legality of their use depends upon the application of the general rules of international law, including those regulating the use of force and the conduct of hostilities”). “There are no rules of customary or conventional international law prohibiting nations from employing nuclear weapons in armed conflict. In the absence of such an express prohibition, the use of nuclear weapons against enemy combatants and other military objectives is not unlawful. Employment of nuclear weapons is, however, subject to the following principles: the right of the parties to the conflict to adopt means of injuring the enemy is not unlimited; it is prohibited to launch attacks against the civilian population as such; and distinction must be made at all times between combatants and civilians to the effect that the latter be spared as much as possible.” DEP’T OF THE NAVY ET AL., THE COMMANDER’S HANDBOOK ON THE LAW OF NAVAL OPERATIONS § 10.2.1 (2007). See generally *Practice Relating to Nuclear Weapons*, IHL DATABASE, ICRC, [https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2\\_rul\\_nuwea](https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_rul_nuwea) [<https://perma.cc/8A27-DYKY>] (last visited Dec. 22, 2020).

had been attempted), it is difficult to imagine a set of circumstances where an EAI operating on Track 3 would *not* decide to intercept the same order, which contestably places too great a weight upon the humanitarian end of the spectrum.

System 3, however, having considered the same set of relevant legal provisions as the Track 1 system would move on to consider the wider body of legal obligations. Importantly, this would include all relevant ROE—including those relating specifically to nuclear launch (which, given the weapon, are likely to be extremely concise). Moreover, in the knowledge that a post-human analysis of an order to initiate a nuclear launch was likely, future ROE would include specific reference to the role in which an EAI should play. Vitally, as identified in Scenario 3, these ROE cannot be used to side-step existing international legal obligations, but only to provide a greater detail of the strategic, operational, and tactical methods of enforcing them.

If the order to launch a nuclear attack was a legitimate order, there would be no reason for the EAI intercept it (though given the circumstances, the engagement of system 4B(ii) might be the more preferable outcome). As previously discussed, the case for non-interception may be stronger the higher up the command chain the EAI analysis is completed. However, where the EAI examination took place at the level of the submarine commander, and where communication was lost in the process of receiving an “update” from strategic command, these ROE may prove vital. No doubt the commander would also have some access to these ROE, and would of course be a highly skilled, well informed, individual.

Nevertheless, a commander cannot be expected to apply the same legal acumen to a potentially voluminous set of instructions, as would a New York Bar attorney. Moreover, they could not be expected to do so at speeds anything like the data-processing capabilities of a supercomputer—especially not when the same individual’s entire training and preparation for events such as this would have been succinctly grounded in the obligation to strictly follow orders. If an EAI were instead tasked with analysing all available data, including all relevant national doctrine, there would be a much greater opportunity to identify whether the order was consistent with matters of law, policy, and best practice,

and the commander could be more confident that they were making the right decision.

Importantly, however, if the EAI did not detect an abnormality here, or indeed any other permitted reason to intercept the order, the order must be followed. Nonetheless, where there is reason to “suspect” an order was, for example, overwritten, or that it was given with malicious intent, it could be refused by the EAI. *Prima facie*, preventative refusal may seem the most appropriate course of action here. However, it is arguable whether states would employ such a system in a nuclear setting. Instead, and perhaps in somewhat of a complete reversal, it is proposed that passive refusal is the most appropriate “type” of refusal for the following reasons. First, strategic decision making will ultimately remain firmly in the hands of humans, but secondly, by operating in such a way, the EAI might demonstrate something akin to a “gut feeling”—a word of caution, as opposed to an absolute rebuttal. Such a system would therefore assist a commander instead of discounting their experience and rank.

#### V. WIDER IMPLICATIONS OF ROBOT REFUSAL

This penultimate Section extends this controversial discussion into three further areas of analysis: the application of “Robot Refusal” in the context of robot PMCs, spies, and more provocatively, whether RROE should include insubordination for EAIs. By way of brief caveat, the Authors do not intend to “close” the discussion in each of these areas but rather, wish to open these areas in the first instance for future discussion, and much-needed debate. The specific PMC “angle” is whether, and to what extent, an EAI PMC is perhaps more or less likely to disobey orders compared with their human counterparts. It is relatively uncontroversial to suggest that one of the primary motivating factors for a human PMC partaking in a theatre of conflict will be one of financial reward. Extending this argument a little further may prompt the reasonable conclusion that PMCs (in light of this “added value” or “financial incentive”) are less likely to refuse an order. Indeed, should the PMC refuse or disobey an order, or at the very least not successfully complete the mission, extra enumerations may be affected. Meanwhile, an EAI PMC would presumably not have that extra variable “present” (unless the EAI is remunerated or motivated by Bitcoin or other crypto

currencies). The Authors therefore conclude that the addition to the “system of systems” presented above, combined with the removal of the “mercenary” factor in the traditional sense, would mean that there would be no reason as to why an EAI PMC would feel less inclined to fail to disobey or refuse an illegal order.

The inclusion of spies within this discussion is to highlight an instance whereby a “retired” combat EAI transitions from SEAL Team 6 and begins employment at the “Agency”—a not altogether uncommon career path. If the EAI spook mirrored that particular human career choice, would the EAI still retain (in the event it hadn’t undertaken a separate training at the “Farm” in Langley, Virginia) its existing system of systems programming? If this were to be the case, and the “hard drive” was not “reformatted,” an order to target a dissident on London Bridge with a poisoned tipped umbrella, would ultimately be rejected—assuming of course that all EAIs were equally equipped with the same system of systems set out in Part IV. As with the previous inclusion of RROE in the PMC world, one could have a similar instance whereby the variation in programming could result in intelligence agents acting differently depending where they had “undertaken” their initial training or programming.

Taken to its conclusion, were the EAI to wrongfully disobey an order, should it face consequences for insubordination? Clearly, and without being overly flippant (were one to apply such a discussion to an AWS) the very simple answer would be to reprogram the system. A less draconian approach may be some form of disciplinary action but what that may look like in practice, the Authors fully concede, is less tangible—loss of leave or pay is unlikely to perhaps have a motivating factor. Clearly, there would need to be a sliding scale, reflective of the increasing severity from mere insubordination to full-on mutiny and/or treason.

## VI. CONCLUSION

Ultimately, this Article has sought to bring to the forefront a timely discussion regarding an EAI’s ability to refuse erroneous orders—particularly in light of the fact that there is no agreed position in terms of state practice or refusal of orders. The Authors have proposed the novel inclusion of a test in order to determine the precise limitations as to when and how this should happen. Therefore, the proposed RROE not only offer greater

clarity for now, but something distinctly concrete for the future. Whereas more traditional lines of thinking have rightly cautioned against an AWS or EAI's ability to compute human emotions when making proportionality assessments, the Authors' novel approach has been to reverse the thinking and suggest that the EAI is not only placed to understand human traits but can also override human error.

Such action might be considered necessary in any number of scenarios. For example, where a human combatant has complied with their IHL requirements, but where factors such as self-preservation or instinct have clouded their "human ability" to objectively make proportionality assessments. In such an instance, the Authors recommend aborting launch. In making a decision as to whether to obey or disobey human error, the EAI should not only calculate the IHL "requirements" but also actively look to discount any "human factors" which may have influenced the decision. When it comes to an EAI responding to an EAI error, the Authors envisaged EAI/AI "checks," carried out by a number of different systems in order to add an additional level of protection in order to identify and prevent rogue orders. In the very purest strategic sense this would indeed be the ultimate "system of systems."

Part II of the Article was the natural starting point for the discussion and revisited the key tenets of IHL in order to provide a logical interface between a combatant's obligations under DTP and the right of refusal considered in Part III. As noted in Parts I and II which, forms part of the key concluding message once states are equipped with EAI, IHL may well compel their usage. Part III of the Article extended the discussion into differing thresholds for the refusal of orders. Here, the Authors summarized that the minimum standard as portrayed in the narrow obligation requires a member of the armed forces to strictly adhere to an order unless there is a customary duty to refuse on the grounds that it is manifestly unlawful. Perhaps somewhat unfortunately, the analysis identified that difficulties may arise when programming EAI to refuse, due to the fact the obligation operates more broadly. For some states, it is not a straightforward consideration as to the lawfulness of the order. Instead, a combatant adhering to the wider obligation must reconcile any order according to relatively abstract concepts such

as human dignity, while concurrently scrutinizing whether or not there is a reasonable chance of mission success, or whether the order is consistent with national policy. This leads to the eventuality, however, that existing practice is not only incongruous, but can also place too great an emphasis upon either military necessity or humanitarian considerations—both of which are elements that IHL, at its heart, strives to balance without bias.

Part IV of the Article provided what the Authors believe is a robust legal “system of systems” not just with the intended effect of ensuring appropriate compliance or non-compliance from a legal perspective, but one that one would have strategic benefit too. It is perhaps overly tempting to suggest that Clausewitz’s idea of “total war” (where the “gloves come off”) is necessarily the appropriate strategic approach—the system envisaged by the Authors not only limits unlawful actions but ones which would also minimize and reduce strategic error as well.

In Part V the Authors broadened the remit of the discussion and extended the discussion into a natural EAI career trajectory: PMC and espionage. Here the Authors note that the existence of EAI may well cause the gradual extinction of PMCs or at least in the sense of them operating as mercenaries. More problematic perhaps in both the realms of PMC and espionage is the “legacy” programming—the EAI’s original training and embedding of system of systems may prove ideal for IHL compliance but prevent ineffectual operational qualities. Quite simply, the system of systems programming may be too effective, and an EAI would refuse any of the “Black Ops” missions those realms frequent.<sup>193</sup>

Too simplistic and broad a rebuttal to any such considerations undertaken in this work would be to simply dismiss such a discussion in the realms of “well . . . it depends on the algorithm.” What this Article has pinpointed and strived to comprehend, is precisely what such calculations would look like.

---

193. A final “footnote” to this discussion more generally, surrounding future use of PMCs is that their role may decrease rather than increase. While it is undeniable that the immediate future will see an ever-increasing presence of PMCs both in combat (and even more so in non-combat roles), the Authors contend that a natural re-shaping of the sector will occur. The very *raison d’être* of a PMC is financial reward as opposed to fighting for the flag—when such reward is no longer conceivable (in the future envisaged by the Authors) the existence of PMCs may continue, but they would no longer be mercenaries.

In terms of an AWS system it might be a continuous proportionality feedback assessment. In terms of EAI, in addition to the ongoing proportionality assessment, it is a series of additional checks and balances both to discount human error and emotion, and, indeed, to ensure an even greater level of compliance with IHL obligations. Clearly, this raised broader issues in terms of chain of command where the Authors questioned whether an EAI should be able to override a chain of command and concluded in the affirmative.

To return to one example from the civilian “field” highlighted in the abstract—an aircraft’s anti stall mechanism’s potential to override human error—such overriding is seen as nothing but a positive.<sup>194</sup> It is also perhaps important to distinguish that the system of systems proposed here is much stricter *par excellence* than say, an automotive vehicle’s system of systems—the car is unlikely to shut down and switch off (thus overriding its human controller) purely because they have ignored the wrench key, designating that the vehicle is in need of service or the coffee cup icon denoting the need for a break. In complete contrast, an EAI could, would, and should abort a nuclear launch if there are any “red flags” raised by the system of systems. The Authors believe that starting and opening this much-needed dialogue for further debate will assist in the growth of scholarship in this area. It is the firm assertion of the Authors that under certain circumstances, “robot refusal” is preferable to unquestioning acceptance of human error.

---

194. See *supra* note 15.

