ARTICLES

"EMBODIED AI" AND THE DIRECT PARTICIPATION IN HOSTILITIES: A LEGAL ANALYSIS

FRANCIS GRIMAL AND MICHAEL J. POLLARD*

> "*These laws are sufficiently ambiguous so that I can write story after story in which something strange happens, in which the robots don't behave properly, in which the robots become positively dangerous . . . .*"[1]

ABSTRACT

This Article questions whether, under International Humanitarian Law (IHL), the concept of a "civilian" should be limited to humans. Prevailing debate within IHL scholarship has largely focused on the lawfulness (or not) of the recourse to autonomous weapons systems (AWS). However, the utilization of embodied artificial intelligence (EAI) in armed conflict, has yet to feature with any degree of prominence within the literature. An EAI is an "intelligent" robot capable of independent decision-making and action, without any human supervision. Predominately, the approach within the existing AWS/AI debate remains pre-occupied in ascertaining whether the military "system" is capable of determining/distinguishing between civilians and combatants. Furthermore, the built-in protection mechanisms within IHL are inherently "loaded" in favor of protecting humans from AWS, rather than vice-versa.

IHL makes a clear distinction between civilians and civilian objects. However, increasingly advanced EAI's will make such a distinction highly problematic. The novel approach of this Article is twofold: to address the "EAI lacuna" in the broader sense, and to consider the application of EAI within a specific area of IHL: "Direct Participation in Hostilities (DPH)". In short, can a robot "participate"? DPH is firmly grounded within the cardinal principle of

---

* 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 extend their gratitude to Alexander Keyser, Rachel Finn, and all at GJIL for their help, input and editorial suggestions, and also to Thomas Spiegler, Editor-in-Chief of GJIL.

1. Prolific Science-fiction writer Isaac Asimov discussed his much-referenced three rules of robotics at *Rise of the Robots: More Human than Human*, BBC RADIO 4 (Feb. 7, 2017), https://www.bbc.co.uk/sounds/play/b08dnr3r.

distinction, and proportionality assessments, in order to afford protection to the civilian population during hostilities. Fundamentally, this Article challenges the International Committee of the Red Cross's (ICRC) influential guidance on DPH. The Authors controversially submit that by continuing to follow that guidance, civilian objects will, under some circumstances, be afforded greater protection than human combatants.

To highlight this deficiency, the authors challenge the ICRC's assertion that civilian status must be presumed where there is doubt, and instead subscribe to the prevailing alternative interpretation that DPH assessments need to be made on a case-by-case basis. To address the deficiency, the authors add the novel inclusion of a "Turing-like test" within DPH assessment.

A concrete example of EAI is that of a robot medic. The robot medic's Hippocratic duty is to protect its patient's life. In doing so (and given a suitable set of circumstances), the robot medic may wish to return fire against an attacker (here, the authors envisage a scenario during urbanized warfare). Would such an action constitute DPH (?), and what would the legal parameters look like in practice? Consequently, how would the attacker compute collateral damage in light of neutralizing the potentially "DPHing" robot? Implicit within such a discussion, is the removal of emotional attachments that, for many, are innate in DPH assessments. Indeed, does the ICRC's tripartite test for "DPHing" contain understandable bias in favor of humanitarian considerations?

I. Introduction

The recourse to embodied artificial intelligence (EAI), and its lawfulness (or not) remains an area in need of closer forensic analysis.[2] While there is increasing literature surrounding the use of artificially intelligent robots in armed conflict, its focus centers on whether *machines* will be capable of identifying human participation.[3] The present authors

2. EAIs have been introduced into contemporary literature surrounding the lawfulness of military owned and operated Autonomous Weapons Systems (AWS). However, such discussions repeatedly fail to extend the analysis to consider how international legal principles might be affected by the introduction of civilian EAIs. *See, e.g.*, Bonnie Docherty et al., *Head the Call: A Moral and Legal Imperative to Ban Killer Robots*, Human Rights Watch (2018); Heather M. Roff & David Danks, *"Trust but Verify": The Difficulty of Trusting Autonomous Weapons Systems*, 17 J. Mil. Ethics 2 (2018); Nehal Bhuta et al., Autonomous Weapons Systems: Law, Ethics, Policy (2016); Armin Krishnan, Killer Robots: Legality and Ethicality of Autonomous Weapons (2016). For a lighter but in-depth investigation into AWS, see Paul Scharre, Army of None: Autonomous Weapons and the Future of War (2018). In a recent recorded debate one leading expert on AWS even refers to the fact that AWS are essentially EAIs, but nevertheless refrains from expanding the discussion further. For Peter Asaro's discussion, see Ariel Conn, *Podcast: Six Experts Explain the Killer Robots Debate*, Future of Life Institute (Jul. 31, 2018), https://futureoflife.org/2018/07/31/podcast-six-experts-explain-the-killer-robots-debate/. The term, EAI, has nevertheless been in use for a number of years in the general discussion surrounding AI. *See generally* Hubert L. Dreyfus, *Why Computers Must Have Bodies in Order to Be Intelligent*, 21 Rev. Metaphysics 13 (1967). In contrast, Kenneth Payne flips the conversation on its head in order to distinguish (non-embodied) AI. He notes "AI is not an embodied and intensely social animal, and does not have biologically and environmentally evolved emotions and motivations." Kenneth Payne, *Artificial Intelligence: A Revolution in Strategic Affairs?* 60 Survival: Global Pol. & Strategy 7, 27 (2018).

3. *See, e.g.*, Docherty, *supra* note 2, which repeats a number of the arguments raised in the first Human Rights Watch Report, Bonnie Docherty et al., *Losing Humanity: The Case against Killer Robots*, Human Rights Watch (2012). The 2012 report was largely responsible for bringing "killer robots" to the attention of the wider public, and in it, the authors question whether a machine would ever be able to recognize the difference between a lawfully targetable combatant and a child armed with only a toy gun. For a discussion in opposition to this, which forwards, for

reverse this question. Instead, this Article considers whether, in the "fog of war," *human combatants* will be capable of identifying if an EAI is directly participating in armed conflict.[4] In other words, this Article controversially focuses on future *civilian* tech, rather than military tech.[5]

Successive generations have endeavored to recreate human thought processes and associated behaviors.[6] However, while contemporary intelligent systems such as Google's AlphaGO and IBM's Deep Blue are undoubtedly revolutionary, they are still comparatively limited in their applications.[7] Due to their reactive nature, today's systems are largely incapable of perceiving context, or demonstrating convincing levels of human consciousness. Neither "Siri" nor "Alexa" are about to become self-aware.[8]

The overarching premise of this Article underlines the distinct probability of both the development and use of much more advanced artificial general intelligence (AGI) in civilian and military sectors. While such a future looking discussion may require a "minor leap of faith" by the reader, the approach adopted by the authors remains entirely plausible. Indeed, as noted by Schmitt: "certain possible, or even likely, trends in military affairs can be identified based on technological advances, geopolitical events, and logical shifts in strategy and tactics."[9]

---

example, that AWS will eventually be more capable of distinguishing non-combatants from combatants, *see generally* Marco Sassoli, *Autonomous Weapons and International Humanitarian Law: Advantages, Open Technical Questions and Legal Issues to be Clarified*, 90 INT'L L. STUD. SER. 308 (2014).

4. For the purposes of this Article, the authors confine the term "Embodied AI" to artificially intelligent robotic systems, as opposed to cyber, or other non-tangible though artificially intelligent systems.

5. As previously noted, the existing debate surrounding EAI and IHL is focused upon the use of AWS. *For examples of this discussion see sources identified supra* note 2.

6. Greek mythology, for example, introduces Hephaestus, the architect or ironmonger to the Gods, who created Talos, a giant bronze automaton that roamed the shores of Crete in order to protect the island from pirates and invaders. A description of Talos can be found in APOLLONIUS RHODIUS, ARGONAUTIKA, BOOK 4 1638–93 (Third Century BCE); HOMER'S THE ILIAD (c. 800-700 BCE). In addition, the tenth century Byzantine imperial book of ceremonies appears to provide a number of the real-world examples of mechanized fabrications. EMPEROR OF THE EA CONSTANTINE VII PORPHYROGENITUS, CONSTANTINE PORPHYROGENNETOS: THE BOOK OF CEREMONIES BYZANTINA AUSTRALIENSIA 18 (First published 10th Century BCE, Aust Assn Byzantine Stud, 2012).

7. The point is that though AlphaGo may be capable of mastering the game Go, it is not very good, for example, at making coffee.

8. Apple Inc.'s Siri and Amazon's Alexa virtual assistants, for example, merely react to human inputs.

9. Michael N. Schmitt, *The Principle of Discrimination in 21st Century Warfare*, 2 YALE HUM. RTS. & DEV. L.J. 143, 152 (1999).

Decades have elapsed since the last comprehensive reworking of IHL,[10] and within that timeframe the traditional battlefield has undergone (and continues to undergo), a number of significant transformations.[11] Whether those transformations constitute a further "Revolution in Military Affairs" remains open to debate.[12] Irrespective of reaching such a determination, the most damaging consequence to this change of battlefield is that armed conflict frequently occurs in densely populated urban environments. Crucially, strategists believe that this "trend" of facing the enemy from within the city walls is set to continue, and indeed may evolve, for the foreseeable future.[13]

---

10. See in particular the four Geneva Conventions and their three additional protocols; 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, 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]. Note also, 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], though it should be noted this latter provision has had only a relatively minor effect upon existing IHL.

11. In strategic terms, consider, for example, the introduction of cyber weapons. The increase in the use of such weapons has led to an expansion of the battlefield into the digital realm, which, when considered together, means that the battlefield is now more commonly referred to as a battlespace. *See, e.g.*, Michael N. Schmitt, *"Direct Participation in Hostilities" in 21st Century Armed Conflict*, U. OF OSLO 510 (2004), https://www.uio.no/studier/emner/jus/humanrights/HUMR5503/h09/undervisningsmateriale/schmitt_direct_participation_in_hostilties.pdf ("[B]attlefields have been transformed into battlespaces and two or even three dimensional warfare has been supplanted by virtual and four-dimensional armed conflict.").

12. A NATO parliamentary assembly committee report, citing Jeffrey McKitrick, *et al*, identifies that "A Revolution in Military Affairs (RMA) is a major change in the nature of warfare brought about by the innovative application of new technologies which, combined with dramatic changes in military doctrine and operational and organizational concepts, fundamentally alters the character and conduct of military operations." LOTHAR IBRÜGGER, NATO PARLIAMENTARY ASSEMBLY, THE REVOLUTION IN MILITARY AFFAIRS (1998), http://iwar.org.uk/rma/resources/nato/ar299stc-e.html.

13. Looking to future war, Retired Maj. Gen. David Fastabend and Mr. Ian Sullivan, for example, note "[d]readed in the past, urban operations will become the default environment: not only a necessity, but also an opportunity. Cities will have massive resources that can be directed for war, such as computer controlled machine shops, 3D manufacturing facilities, small

Distinction, or the "basic rule," is customary in nature,[14] and is codified in Article 48 of Additional Protocol I. It is applicable to both International Armed Conflicts (IACs), i.e., between states, and in Non-International Armed Conflicts (NIACs), i.e., where at least one party to the conflict is a non-state entity. Article 48 of Additional Protocol I provides that: "In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to a conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objects and accordingly shall direct their operations only against military objectives."[15] Neither civilians nor civilian objects must ever be the direct object of attack.[16] To "intentionally" target either category, constitutes a war crime.[17] Interpreting such a clear-cut requirement is not without difficulty. As noted by Emily Crawford, the basic rule assumes "that one is able to make the distinction between a civilian and a combatant clearly and easily."[18] Most professional combatants adhere to IHL (at least in this context) by distinguishing themselves from the civilian population, by wearing a uniform. Whereas, fighters of non-state armed groups

scale chip foundries, and a dense array of consumer electronics, wireless nodes, and commercial and private fiber networks." US Army Training and Doctrine Command (TRADOC) G-2 Mad Scientist Initiative, *An Advanced Engagement Battlespace: Tactical, Operational, and Strategic Implications for the Future Operational Environment*, SMALL WARS J., https://smallwarsjournal.com/jrnl/art/advanced-engagement-battlespace-tactical-operational-and-strategic-implications-future (last visited Apr. 5, 2020).

14. JEAN-MARIE HENCKAERTS & LOUISE DOSWALD-BECK, INT'L COMM. OF THE RED CROSS, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW: VOLUME I: RULES 3–8 (2007). The ICJ have identified the principle of distinction as one of the "cardinal principles" of international humanitarian law, as well as being an "intransgressible" principle of international customary law. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. Rep. 226, ¶ 78–79 (July 8) [hereinafter Nuclear Weapons Advisory Opinion].

15. Additional Protocol I, *supra* note 10, art. 48.

16. *Id.*

17. *See* Francis Grimal & Jae Sundaram, *Combat Drones: Hives, Swarms, and Autonomous Action?*, 23 J. CONFLICT & SECURITY L. 105, 128 (2018), which at note 110 identifies, KNUT DÖRMANN, ELEMENTS OF WAR CRIMES UNDER THE ROME STATUTE OF THE INTERNATIONAL CRIMINAL COURT: SOURCES AND COMMENTARY 130, 233 (2003). *See* Prosecutor v. Galić, Case No. IT-98-29-T, Judgment and Opinion, (Int'l Crim. Trib. for the Former Yugoslavia Dec. 5, 2003). However, not all attacks on civilians are necessarily war crimes. *See* Rome Statute of the International Criminal Court art. 8, § 2(b)(iv), July 17, 1998, 2187 U.N.T.S. 38544 (noting the word "intentionally"—there must be intention present: "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.").

18. EMILY CRAWFORD, IDENTIFYING THE ENEMY: CIVILIAN PARTICIPATION IN ARMED CONFLICT 1 (2015).

(NSAGs) may actively seek to do the exact opposite. However, this is further complicated where the "part-time" fighter lays down their arms at the end of their day, and returns home to their family.

IHL does attempt to cater to the "farmers by day, and fighters by night" scenario[19] by providing that civilians shall only enjoy a general protection against attack, so long as they do not directly partake in hostilities.[20] In other words, a civilian becomes lawfully targetable once he or she makes the conscious decision to participate.[21]

Though the concept of civilian participation has existed in various guises for many years, the ratification of the Additional Protocols in 1977 explicitly codified DPH within the corpus of IHL.[22] Nevertheless, due largely to the nature of armed conflict at that time,[23] the concept received little scholarly reflection in the period following codification.[24]

The diminishing paucity within the literature, however, has changed in recent years as a result of urbanized battlefield conditions post 9/11. In light of the battles fought as a part of "*Operation Enduring Freedom*," a number of commentators, including the ICRC, began to acknowledge the growing importance of the concept of DPH on contemporary battlefields.[25] As a consequence of the uncertainty surrounding the correct

19. *See* NILS MELZER, INT'L COMM. RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION IN HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW 5, 12, 72 (2009) [hereinafter ICRC Guidance].

20. Additional Protocol I, *supra* note 10, art. 51(3).

21. This is expanded upon in the conversation that follows, however, the point is, if an individual is to be considered as directly participating in hostilities, the act in question must, *inter alia*, be linked to the hostilities. The act in question cannot, for example, merely be geographically proximate to hostilities.

22. Common art. 3 of the Geneva Conventions refers to, "persons taking no active part in hostilities." *See, e.g.*, Geneva Convention III, *supra* note 10, art. 3.

23. A number of armed conflicts throughout the 1980s were either between state powers, for example, as with the Falklands war fought between Argentina and the United Kingdom in 1982, and the Iran-Iraq war fought between 1980–1988, or civil wars such as those in Sri Lanka (1983–2009), and Afghanistan (1989–1992).

24. Gehring, for example, notes that although Additional Protocol I expanded the concept of civilian protection in armed conflict, there are a number of ways in which a civilian can lose their protected status, such as DPH. Nonetheless, although he does go on to identify a number of the contemporary matters of contention, the author suggests that existing provisions of international law (i.e., the Geneva Conventions and the Additional Protocols) provide "a workable balance." *See* Robert W. Gehring, *Loss of Civilian Protections Under the Fourth Geneva Convention and Protocol I*, 90 MIL. L. REV. 49, 50 (1980).

25. Also of particular relevance was the realization of the "civilianization of conflict" *See* Michael N. Schmitt, *Humanitarian Law and Direct Participation in Hostilities by Private Contractors or Civilian Employees*, 5 CHI. J. INT'L L. 511 (2005). For examples of the general scholarly debate

interpretation of DPH, the ICRC held five meetings in The Hague, and Geneva, between 2003 and 2008.[26] Those meetings bought together 40 to 50 legal experts from academic, military, governmental, and non-governmental circles, with the intention of, inter alia, clarifying the precise nature the of obligation contained within Article 51(3) Additional Protocol I. Ultimately, the panel failed to reach a unanimous, or even a majority decision. Nonetheless, with reference to the meetings, the ICRC went on to publish its substantive (though non-binding) guidance on the notion of DPH under IHL.[27]

At its heart, the ICRC guidance promotes a tripartite test for establishing the circumstances under which a civilian can be identified as directly participating in hostilities. And, since these criteria form an integral part of the examination contained within this Article, they are presented verbatim below:

> In order to qualify as direct participation in hostilities, a specific act must meet the following cumulative criteria:
>
> 1. the act must be likely to adversely affect the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury, or destruction on persons or objects protected against direct attack (threshold of harm), and;
> 2. there must be a direct causal link between the act and the harm likely to result either form the act, or from a coordinated military operation of which that act constitutes an integral part (direct causation), and

surrounding DPH at the time, see also Schmitt, *supra* note 11; and Int'l Inst. Humanitarian Law, *International Humanitarian Law and Other Legal Regimes: Interplay in Situations of Violence*, 26–29 (Sept. 4–6, 2003), http://iihl.org/wp-content/uploads/2019/05/INTERNATIONAL-HUMANITARIAN-LAW-AND-OTHER-LEGAL-REGIMES.pdf. For a useful general introduction to the background and problems that were faced by the panel in 2004, *see* Int'l Comm. Red Cross, *Second Expert Meeting on the Notion of Direct Participation in Hostilities in Non-International Armed Conflict* (Oct. 25–26, 2004) https://www.icrc.org/en/doc/assets/files/other/2004-05-expert-paper-dph-icrc.pdf.

26. First expert meeting, The Hague, June 2, 2003; Second expert meeting, The Hague, Oct. 25–26, 2004; Third expert meeting, Geneva, Oct. 25–25, 2005; Fourth expert meeting, Geneva, Nov. 27–28, 2006; Fifth expert meeting, Geneva, Feb. 5–6, 2008. For agendas and reports of each meeting, *see* Int'l Comm. Red Cross, *ICRC Clarification Process on the Notion of Direct Participation in Hostilities under International Humanitarian Law (Proceedings)* (June 30, 2009), https://www.icrc.org/en/doc/resources/documents/article/other/direct-participation-article-020709.htm.

27. ICRC Guidance, *supra* note 19.

3. the act must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another (belligerent nexus).[28]

There is no question that the ICRC's report and cumulative criteria do appear to offer a much-needed objective test for determining DPH. However, a number of prominent commentators have objected to its "narrow" nature.[29] As a result, these "opponents" doubt that states will be keen to utilize it.[30] Michael N. Schmitt and William H. Boothby (both experts who were present at the ICRCs meetings), suggest that the narrow interpretation of DPH places an unacceptable "imbalance" upon members of the regular armed forces—a concern shared by the present authors.[31]

In reference to instances where the targeteer has doubts as to the status of a civilian, the authors each independently offer the need for a "wider" interpretation of DPH.[32] This interpretation does accept that where there is doubt, civilian status must be presumed when distinguishing.[33] However, it proposes that in cases where DPH is suspected, international law makes no such distinction. When applied, the "wider: interpretation, reduces the combatants' burden, with the impetus instead shifted to the civilian. While this may appear somewhat perverse, civilians are, nevertheless, able to negate the burden by simply removing themselves from situations where their status may be called in to question. Admittedly, and in practice, while civilians are not always able to remove themselves from situations where their status may be questioned, there is clearly behavior such as surrender or seeking refuge, which could never constitute DPH.

The graphics below (figures 1 & 2) demonstrate these two contrasting interpretations of DPH, and how each allocates the burden of risk. They also demonstrate that in order to achieve a balanced application of DPH (in addition to a balanced burden of risk) the term *civilian* may

28. *Id.* at 46.

29. These objections will be closely scrutinized in Parts II and III of this Article. Of particular note, see William H. Boothby, *Direct Participation in Hostilities: A Discussion of the ICRC Interpretive Guidance*, 1 J. Int'l Human. Legal Stud. 143 (2010); Michael N. Schmitt, *Deconstructing Direct Participation in Hostilities: The Constitutive Elements*, 42 N.Y.U. J. Int'l L. & Pol. 697 (2010).

30. Schmitt, *supra* note 29, at 699.

31. *See generally* Boothby, *supra* note 29.

32. Boothby, for example, suggests that the ICRC's interpretation of DPH would "narrow the notion of membership to an unacceptable degree." Boothby, *supra* note 29, at 154. However, Schmitt notes that applying it "risks an overly narrow interpretation of direct participation," Schmitt, *supra*, note 29, at 720.

33. Schmitt, *supra*, note 29, at 736–37; Boothby, *supra* note 29, at 148–50.

need to be extended to include future EAIs. Unless and until that is done, in some instances, IHL will fail to provide human combatants with the same level of protection as artificially intelligent civilian objects.

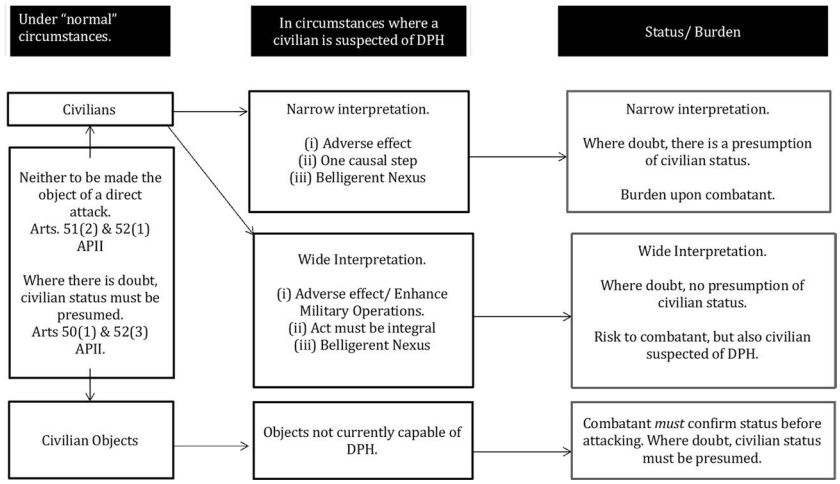

FIGURE 1: Civilians and civilian objects: The distinction.

The Graphic above, which must be considered in light of figure 2 below, demonstrates that regardless of which interpretation is preferred, the concept of DPH cannot be applied to civilian objects. However, in cases where there is a doubt as to the status of a civilian, the burden of risk is affected differently, depending which of the two interpretations is utilized.

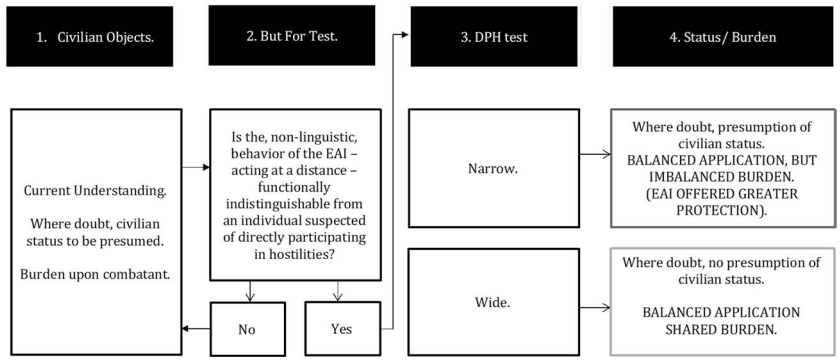

FIGURE 2: The balancing act.

By applying a Turing type test, the graphic in figure 2 demonstrates that in certain situations, EAIs could be classified as civilians, rather than civilian objects. If this is done, then the graphic identifies that

only the "wider" interpretation of DPH can offer a balanced application, and a balanced burden of risk.

By way of overview, Part II of this Article turns to the ICRC's guidance, and examines the concept of DPH in detail. Part III provides a number of scenarios of creditable near, medium and long-term future EAIs. By doing so through the lens of DPH, it may be possible to identify a threshold at which the status of an EAI can shift from civilian property, to civilian. This distinction is important, because as previously noted, civilian property is protected against direct attack, whereas under current law, a civilian may lose that protection.[34] The section examines the legal consequences of such a transformation, and the effect that it could have upon the human population that surrounds them. Part IV of the Article subjects other "classic" principles/discussions (e.g. Levee en Masse, Perfidy, POW status, and the recourse to PMC's, Spies) within the IHL corpus to the context of "EAI" considerations.

There are clearly serious consequences, in making the determination as to whether civilians are "DPH"ing." And, while the authors do question whether the ICRC's guidance will remain fit for purpose in the future, it is also important to note the scale of the task that is at hand. Nonethless, that guidance is indeed preferable to alternatives such as Isaac Asimov's three laws of robotics that were intentionally designed to support fantastical stories.[35]

II. Civilian Participation in Armed Conflict

While the ICRC report is both highly respected, and tremendously valuable, it is neither legally binding, nor without critique.[36] To introduce these conflicting discussions, section A of Part II assesses how the

34. The point is, a civilian is capable of directly participating in hostilities, whereas currently, civilian objects are not.

35. Asimov's 3 laws of robotics, which are still widely cited today, are: "[1.] a robot may not injure a human being, or, through inaction, allow a human being to come to harm . . . [2.] a robot must obey the orders given it by human beings except where such orders would conflict with the First Law . . . [3.] a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws." Isaac Asimov, I, Robot 40 (1950). For an example of Asimov being referred to in the contemporary discussion surrounding AWS, *see* Rebecca Crootof, *War Torts: Accountability for Autonomous Weapons*, 164 U. Pa. L. Rev. 1347, 1372 n.135 (2016); 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).

36. *See generally* Boothby, *supra*, note 31; Schmitt, *supra* note 29; Kenneth Watkin, *Opportunity Lost: Organised Armed Groups and the ICRC: "Direct Participation in Hostilities" Interpretive Guidance*, 42 N.Y.U. Int'l J. L. & Pol. 641 (2010); Michael N. Schmitt, *The Interpretive Guidance on the Notion of Direct Participation in Hostilities: A Critical Analysis*, 1 Harv. Nat'l Sec. J. 5, 23 (2010).

legal principle of DPH fits within the wider corpus of IHL on the law of targeting. Section B examines the ICRC's interpretative guidance to DPH, while the remainder of Part II (Sections BI, BII and BIII) examine the ICRCs" "cumulative criteria"—the criteria that are intended to help determine whether DPH is present.[37] Ultimately, Part II engages with Schmitt and Boothby's contention that the ICRC's test is too restrictive to be applied during the "fog of war"?

A. *Distinguishing the Civilian Population: How Does DPH Fit into IHL?*

IHL seeks to reconcile two diametrically opposing concepts: military necessity and humanitarian considerations.[38] It does so by identifying the two core principles of distinction and proportionality as barometers to calibrate the lawfulness of force.[39] While IHL does place specific emphasis upon the protection of the civilian population, civilians themselves are primarily free from constraint. [40] Other than under limited circumstances (such as DPH), the civilian population remains accountable to their respective municipal legal systems.[41] In contrast, belligerents, and military decision makers, carry the burden of obligations.[42] The principles stemming from IHL may, for example, prevent a "targeteer" from applying force against an enemy combatant, while at the same time allowing for civilians to be targeted indirectly.[43]

---

37. As demonstrated *supra* note 36, there was a limited amount of literature that followed the publication of the ICRC's guidelines. However, in recent years the debate has slowed somewhat, other than in relation to AWS. *See supra* note 2.

38. For example, Yoram Dinstein notes "[i]n following [the] middle road, LOIAC [(Law of International Armed conflict)] allows Belligerent Parties much leeway (in keeping within the demands of military necessity) and nevertheless curbs their freedom of action (in the name of humanitarianism)." YORAM DINSTEIN, THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT ¶ 23 (3d ed. 2016). Schmitt also notes that "[i]nternational humanitarian law represents a delicate balance between the dictates of military necessity and humanitarian considerations, a balance famously codified in the 1868 St. Petersburg Declaration's acknowledgement that at a certain point 'the necessities of war ought to yield to the demands of humanity.'" Schmitt, *supra* note 29, at 713.

39. The ICJ described the principles of distinction and protection of the civilian population as the "cardinal principles" Nuclear Weapons Advisory Opinion, *supra* note 14, ¶ 78.

40. International Humanitarian Law is intended to govern the behavior of combatants, while municipal law is responsible for assessing the behavior of civilians.

41. See *supra* text accompanying note 40.

42. The purpose is to distinguish between combatants and non-combatants.

43. Additional Protocol I, *supra* note 10, art. 50, civilians are defined in the negative as "any person who does not belong to one of the categories of persons referred to in Article 4 A (1), (2), (3) and (6) of the Third Convention and in Article 43 of this protocol."

One such principle prohibits the parties to a conflict making the civilian population the direct object of an attack.[44] The civilian population includes individual civilians[45] and civilian property.[46] The basic rule is widely recognized as being customary in nature[47] and is contained within Additional Protocol Article 48. That provision provides that the parties to an armed conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objects, and shall direct their operations only against military objects.[48]

More readily identified as the principle of distinction, the ICJ has identified that Article 48 Additional Protocol I is a cornerstone of IHL.[49] In short, under "normal" circumstances (during armed conflict) a graphical representation of the principle of distinction would appear thus:

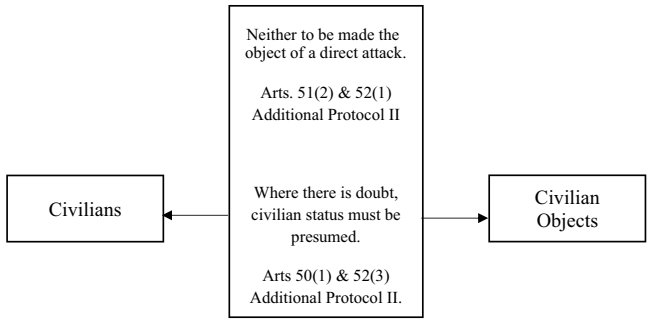

FIGURE 3: Where there is doubt as to status.

In order to support the requirement to distinguish civilians and civilian property from legitimate military targets, the principle of distinction further requires that a combatant must distinguish him or herself from civilians when engaged in or in preparing for an attack.[50] The most obvious example is by wearing a uniform.[51]

---

44. Additional Protocol I, *supra* note 10, art. 49 states during an armed conflict "'Attacks' means acts of violence against the adversary, whether in offence or in defence."

45. Additional Protocol I, *supra* note 10, art. 51(2).

46. *Id.* at art. 52(1).

47. HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 1.

48. Additional Protocol I, *supra* note 10, art. 48.

49. Nuclear Weapons Advisory Opinion, *supra* note 14, ¶ 78.

50. Additional Protocol I, *supra*, note 10, art. 44; HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 106.

51. *See* Geneva Convention III, *supra* note 10, art. 4(A)(2)(b) which refers to the requirement for a combatant to wear a fixed distinctive sign recognizable at a distance. The most obvious method of doing this, is by wearing a particular uniform, see Dinstein, *supra* note 38, ¶ 140.

However, it is vital to note that civilians do not enjoy absolute protection against attack. In order to maintain the military necessity/humanitarian consideration balance, once a legitimate military target is identified, a party to the conflict *may* lawfully target and attack it, *even* in the knowledge that injury to civilian life or damage to civilian property is likely. For such losses or injuries to be considered lawful they must not be "excessive in relation to the concrete and direct military advantage anticipated"—the principle of proportionality.[52] Proportionality is codified in articles 51(5)(b) and 57(2)(a)(iii) of Additional Protocol I, and is also widely accepted to be customary in nature.[53] As noted by one of the authors elsewhere, while proportionality is codified within Additional Protocol I 51(5)(b), the actual "terminology" passes without direct mention.[54]

Although distinction and proportionality remain at the heart of IHL,[55] the contemporary battlefield has evolved since the drafting of the Geneva Conventions and their two Additional Protocols. Combatants rarely now face each other, sword, shield, or bayonet in hand, on a battlefield far removed from the civilian population. Be it due to intentional disguise, or simply to the lack of appropriate means, it is now more common that belligerents fail to distinguish themselves under their IHL obligations.[56]

Nevertheless, Article 51(3) Additional Protocol I states the basic rule that civilians must be protected from direct attack, "unless and for such time as they take a direct part in hostilities."[57] The notion of DPH is also contained within the Rome Statute,[58] and is implicit in Common Article 3 of the Geneva Conventions, which refers to "[p]ersons taking

52. Additional Protocol I, *supra* note 10, art. 51(5)(b), art. 57(2)(a)(iii).

53. HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 14.

54. *See* Grimal & Sundaram, *supra* note 17, at n. 111, CLAUDE PILLOUD ET AL., COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949 625–26 (1987), where it is stated that this provision 'should therefore lead those responsible for such attacks to take all necessary precautions before making their decision, even in the difficult constraints of battle conditions." The authors also note the implications beyond the present scope of this Article. How would the attacker compute collateral damage when the target's status is unknown—is the attacker targeting a human civilian or a "DPHing robot"? This issue is presently being addressed by the authors in a follow-up to this Article.

55. *See* Nuclear Weapons Advisory Opinion, s*upra* note 14.

56. Additional Protocol I, *supra* note 10, art. 44; HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 106.

57. Additional Protocol I, *supra* note 10, art. 51(3).

58. Rome Statute, *supra* note 17, art. 8, §§ 2(b)(i), (2)(e)(i).

no active part in hostilities," which has effectively the same meaning as the text of Article 51(3) Additional Protocol I.[59]

A fundamental problem with Article 51(3) Additional Protocol I is its lack of explicit guidance regarding the application of DPH. For example, there is no guidance to determine: 1) whether it is possible to play an indirect part in hostilities; 2) whether there are temporal or geographical parameters, which should mark the beginning and end of participation, or; 3) how the concept relates to private military contractors (PMC), many of whom now routinely carry out traditional military tasks.[60]

Traditionally, the interpretation of DPH in regard of such matters had been undertaken by states with reference to their own military manuals, and rules of engagement (RoE).[61] Nevertheless, perhaps unsurprisingly, such practice became inconsistent. Accordingly, observers began to question whether it should be left for states to continue to independently "decipher" their legal obligations.[62]

Consequently, the ICRC along with the T.M.C. Asser Institute convened a group of experts to discuss the matter at a number of meetings in 2003, 2004, 2005, 2006, and 2008.[63] However, after five years of various discussions, the panel of experts was unable to reach a consensus, and could not endorse the report as per the original instructions.[64]

---

59. Boothby, *supra* note 29, at 147-48; Schmitt, *supra* note 11, at 507 ("Although Common Art. 3, and Protocol II employ different terminology („active" and „direct" respectively), the International Criminal Tribunal for Rwanda reasonably opined in the Akayesu judgment that the terms are so similar they should be treated synonymously"); Prosecutor v. Jean-Paul Akayesu, Case No. ICTR 96-4-T, Judgment, ¶ 629 (Sept. 2, 1998).

60. Schmitt notes that by September 2009, 242,230 civilians were contracted by U.S. Central Command as part of the ongoing conflicts in Iraq and Afghanistan. He also notes that while many of these may have carried out "relatively benign" tasks, such as cooking, others would have been employed to carry out logistics, intelligence, and security duties. Schmitt, *supra* note 29, at 699–700. *See also* Michael N. Schmitt, *Humanitarian Law and Direct Participation in Hostilities by Private Contractors or Civilian Employees*, 5 CHI. J. INT'L L. 511 (2005); *see generally*, WILLIAM H. BOOTHBY, CONFLICT LAW: THE INFLUENCE OF NEW WEAPONS TECHNOLOGY, HUMAN RIGHTS AND EMERGING ACTORS (2014) (considering military contractors throughout).

61. This conversation is expanded upon by Schmitt, who provides examples of U.S., U.K., and Australian Military Doctrine in the area of DPH. He also provides examples of relevant international jurisprudence in which DPH was a key factor. *See, e.g.*, Prosecutor v. Strugar, Case No. IT-01-42-A, Appeals Chamber Judgment, ¶ 176–79 (July 17, 2008); Prosecutor v. Tadić, Case No. IT 94-1-T, Judgment, ¶ 616 (Int'l Crim. Trib. for the Former Yugoslavia May 7, 1997).

62. Schmitt notes for example that prior to the ICRC guidance, DPH was (and perhaps still is) assessed on a case-by-case basis, often with what he refers to as the "[you'll] know it when you see it" approach. Schmitt, *supra* note 29, at 699.

63. Boothby, *supra* note 29, at 146.

64. *Id.*

Instead, the ICRC went on "to publish a document on its own authority,"[65] which it refers to as the interpretive guidance on the notion of DPH.[66]

According to this guidance, there are two fundamental ways in which a civilian can participate in armed conflict. The first is by having a continuous combat function (CCF), while the second relates to the civilian who takes a one-off (or on-and-off) part in hostilities, and is identified as a revolving door fighter.[67] In the first instance, civilian status is permanently lost, provided the civilian continues to have a CCF.[68] In the latter classification, the civilian regains his or her civilian status once each individual participation ceases.[69] In both instances, for such time as the individual is participating in hostilities, the loss of civilian status means that he or she becomes lawfully targetable.[70] Furthermore, any civilian who is identified as participating in hostilities does not gain combatant privileges and remains answerable to the municipal legal system.

B. *Is the ICRC Interpretive Guidance a Suitable Mechanism for Establishing DPH?*

Ultimately, the purpose of this Section and Part II as a whole is not to critique the DPH study,[71] but rather to examine the existing contribution to the literature in this area. At its core, the interpretive guidance offered by the ICRC does "not endeavor to change binding rules of customary or treaty IHL, but reflect . . . how existing IHL should be interpreted."[72] The ICRC hopes that the guidance "will render the resulting recommendations persuasive for States, non-State actors, practitioners, and academics alike."[73] However, opponents have argued that the ICRC test leans *too* heavily toward humanitarian considerations, and consequently that states are unlikely to agree with it, let alone adopt it.[74]

---

65. *Id.*

66. ICRC Guidance, *supra* note 19.

67. *Id.* at 70–73. This concept is considered in greater detail below; however, a widely utilized analogy is "farmers by day, and fighters by night." According to the ICRC, while an individual should lose his or her civilian status while participating, civilian status must be reinstated every time he or she returns to normal life.

68. *See id.* at 32–35.

69. *See id.* at 70–73.

70. Additional Protocol I, *supra* note 10, art. 51(3).

71. As is the case with the early literature in reaction to the guidance, see *supra* note 43.

72. ICRC Guidance, *supra* note 19, at 9.

73. *Id.* at 10.

74. Schmitt, *supra* note 29, at 699.

Michael N. Schmitt and William H. Boothby were two experts present at the meetings, and both acknowledge the fact that there is an urgent need for clarification of the obligation contained in Article 51 (3). However, for similar reasons, they oppose the interpretive guidance on a number of grounds.[75] By way of brief caveat, while there is undoubtedly need for even greater forensic analysis regarding each criterion, the purpose of this section is to focus instead upon where the literature departs from the ICRC's guidance. This facilitates the broader discussion as to whether the term "civilian" should extend beyond humankind.

1. The First Cumulative Requirement: A Threshold of Harm Likely to Result from the Act

The first of the three cumulative criteria is that the act in question must satisfy the "threshold of harm." The first point to underline is that, implicitly, the ICRC recognizes that there must be varying "degrees of harm." If certain acts fail to meet this threshold, they will not qualify as DPH. Such a requirement remains relatively uncontroversial.[76] Once the threshold of harm is satisfied however, a civilian may become lawfully targetable, subject to further qualification. The first of these is that the act in question must adversely affect the enemy.[77]

The guidance further qualifies that adverse effect does not necessarily equate to the "infliction of death, injury, or destruction . . . but essentially any consequence adversely affecting the military operations . . ."[78] It provides a "negative example" whereby a civilian refuses to act as a scout or informant.[79] In such instances, the ICRC suggests that,

> [T]he conduct of a civilian cannot be interpreted as adversely affecting the military operations or military capacity of a party to the conflict simply because it fails to positively affect them.[80]

This is one of many points of divergence between the literature and the guidance. Boothby identifies, for example, that "the provision of

75. Schmitt refers to the "pressing need to develop criteria by which direct participation could be ascertained . . ." *Id.* at 711. Boothby noted that "clarifying the notion of DPH has become an important matter." Boothby, *supra* note 29, at 146.

76. Schmitt, *supra* note 29, at 713–14.

77. ICRC Guidance, *supra* note 19, at 47.

78. *Id.*

79. *Id.* at 49.

80. *Id.*

intelligence to his own side, the instillation of military command and control equipment in a forward tactical operating base . . . [and] . . . the loading of defensive aids suites onto attack aircraft" should all be seen as DPH.[81] Similarly, Schmitt asserts that in practice, harm and benefit are part of a zero-sum game where any contribution to a particular side will typically weaken the other.[82] To demonstrate, he provides a non-fictitious scenario in which Iraqi civilian insurgents became involved in the development, production, and training in the use of improvised explosive devises (IEDs). He identifies how such behavior can dramatically alter the course of a battle,[83] and as a result, he disagrees with the guidance's suggestion that such behavior does not amount to DPH.[84]

In Schmitt's example, the civilian in question was not involved with detonating the device, nor did they locate it in a position that was designed to cause damage. They may, for example, have been involved in the making of the IED, or have purchased a number of the components that are needed to build the IED, and/ or they may have provided instructions to an individual who *would* have ultimately been responsible for detonating the device. However, it may be a stretch to say that at *no point*, were such individuals participating. Doing so not only would excuse certain lethal behaviors, but also provide an added layer of protection for higher ranking personnel that they simply should not receive.[85]

Boothby also identifies that a civilian act could well be instrumental in causing harm that adversely affects a party to the conflict by positively affecting the position of another.[86] He notes "while not necessarily translating into immediate loss to the opposing party,"[87] there are clearly a number of situations where such acts should nevertheless be considered to be DPH. These may include, for example, the training or stewardship of individuals,[88] the cleaning, maintenance, preparation,

81. Boothby, *supra* note 29, at 158.

82. Schmitt, *supra* note 29, at 719.

83. *Id.*

84. *Id.* at 719–20.

85. The ICRC opposes this point of view, suggesting, for example, that individuals involved in recruitment and training of personnel should only be identified as participating in hostilities when they are recruiting and training for a specific, pre-determined, hostile act. ICRC Guidance, *supra* note 19, at 53.

86. Boothby, *supra* note 29, at 158.

87. *Id.*

88. For example, Schmitt, *supra* note 29, at 730 n.96, notes the Israeli Supreme Court identified "that ordering acts of direct participation was itself direct participation." *See also* HCJ 769/02 Pub. Comm. Against Torture in Israel v. Gov't of Israel 2006(2) PD 459, 499 (2006) (Isr.) [Hereinafter Targeted Killings Case].

logistical movements,[89] and the loading of weapons.[90] It may also, in certain circumstances, be enough to supply tactical or operational level intelligence to one party.[91] Nonetheless, a little over a decade since the guidance was first published, the ICRC maintains its position. Consequently, their bifurcation remains, and while the "narrow" model requires "adverse effect," the "wide" interpretation requires the act must either have an adverse effect, or, should enhance the military operations of another.

2. The Second Cumulative Requirement: A Relationship of Direct Causation Between the Act and the Expected Harm

The second of the criteria offered by the guidance is that there must be a direct causal link between the act, and the harm that is likely to occur. Furthermore, the guidance states that by implication, a civilian must be able to take an *indirect* participation in hostilities, and that indirect participation does not result in the loss of civilian protection.[92] In an attempt to solidify this position, the guidance notes: "[f]or a specific act to qualify as 'direct' rather than 'indirect' participation. . .there must be a sufficiently close causal relation between the act and the resulting harm."[93]

It continues that in order for the causal relation to be close enough, it must be direct. It offers that direct causation should therefore "be understood as meaning that the harm in question must be brought about in one causal step."[94] Accordingly, where the conduct of a civilian

89. The matter of whether or not a civilian truck carrying a shipment of ammunition, should be considered as directly participating was discussed in the expert meetings, and is considered in greater detail in part 4.

90. Schmitt notes, for example, that reports suggest the CIA used civilian contractors to load missiles on to unmanned aerial vehicles. Schmitt, *supra* note 29, at 700. The ICRC also discusses a number of similar scenarios. However, in contrast, they suggest that such individuals cannot be said to belong to and organized armed group, and as a result their civilian status remains. They do however note that such individuals, may, through their activities, "increase their exposure to incidental death or injury." ICRC Guidance, *supra* note 19, at 35.

91. In some circumstances, providing intelligence should be seen as DPH. However, the ICRC offers that this is only the case when the provision of such intelligence has an adverse effect on the enemy, and/ or is "carried out with a view to the execution of a specific hostile act." ICRC Guidance, *supra* note 19, at 81, 66.

92. ICRC Guidance, *supra* note 19, at 51. Boothby questions the usefulness of such an assertion, providing that "one wonders how far the idea of inactive participation really gets us." Boothby, *supra* note 29, at 158. In contrast, Schmitt acknowledges that the distinction is useful, if indeed it is weakened, by the "regrettable limitation to 'harm'." Schmitt, *supra* note 29, at 725.

93. ICRC Guidance, *supra* note 19, at 52.

94. *Id.* at 53.

merely builds, or maintains, the capacity to cause harm to an enemy, or where they otherwise only *indirectly* cause harm, they must not be considered to be directly participating in hostilities.[95] This requirement for the harm to be brought about in one casual step is, however, also hotly contested.[96]

In support of a wider, pragmatic position, Boothby, for example, suggests that civilian acts which are undertaken in the midst of military operations by individuals who are "integrated" into a network that is responsible for launching an attack, must be seen to be directly participating in hostilities regardless of whether their contribution to the attack may be considered indirect.[97] Instead, he believes the ICRC's interpretation simply fails to take into account the fact that in contemporary warfare, attacks are regularly bought about by a "multiplicity of integrated steps . . . ."[98]

In the same vein, Schmitt argues that the ICRC fail to justify the requirement for harm to be bought about in one causal step.[99] He identifies a number of instances where a trainer, who might be unaware of the precise details of an imminent attack, is nonetheless an integral part of it.[100] He therefore declares that regardless of the indirect nature of the action, such an individual must be considered to be directly participating.[101] Furthermore, Schmitt notes that in forwarding this second criteria, the ICRC ignore soft law instruments, including its own *ICRC Commentary on the Additional Protocols*, state practice, and pre-existing provisions of law to the contrary.[102]

Although the report's "theoretical distinction between direct and indirect participation"[103] is welcomed, Schmitt suggests that the ICRCs requirement for direct causation does "not represent a sure-fire

95. *Id.*

96. Boothby, *supra* note 29, at 158.

97. *Id.* at 159.

98. *Id.*

99. Schmitt, *supra* note 29, at 727–28.

100. *Id.* at 730.

101. *Id.* at 729–30.

102. Schmitt identifies, THE COMMANDER'S HANDBOOK ON THE LAW OF NAVAL OPERATIONS (July 2007), which notes at ch. 8.3.2, "such persons may . . . be considered to be taking a direct part in hostilities or contributing directly to the enemy's warfighting/war-sustaining capability, and may be excluded from the proportionality analysis." Schmitt is critical that "[t]he failure to cite this pre-existing provision in a law of war manual of the world's most powerful military is especially confusing in light of the ICRC's *Customary International Humanitarian Law* study's heavy reliance on manuals to support assertions that various rules represent customary norms." *Id.* at 733.

103. *Id.* at 734.

formula for unambiguous and unassailable determinations."[104] The casual link requirement, therefore, provides an additional point of departure. This means (in contrast to the narrow, one casual step requirement of the ICRC), that the "wider" interpretation requires that the harm "must constitute an integral part of the conduct" leading to the required threshold of harm.[105]

3. The Third Cumulative Requirement: A Belligerent Nexus Between the Act and the Hostilities Conducted Between the Parties to an Armed Conflict

The final third of the ICRC's cumulative criteria is the requirement for a belligerent nexus. In short, this requires that the act in question must be carried in support of a party to the conflict, and to the detriment of another. This is an important distinction, particularly in an urbanized environment, because there will undoubtedly be circumstances where civilians behave violently, but where that violence does not amount to an act of war. For example, a civilian may act in self-defense,[106] may seek to take advantage of the prevailing conditions in order to loot, [107] or, may become involved in general civil unrest with political motivations.[108] And, although in each situation a violent act may surpass the threshold of harm, and might have an adverse effect on a party to the conflict, if it cannot be "tied to the armed conflict"[109] then the individual in question cannot be seen to be directly participating.[110]

This is the least controversial of the three cumulative criteria,[111] and is generally well supported. However, it is not wholly so for two distinct reasons. The first is that the "narrow" interpretation requires for the belligerent nexus to be accompanied by the "adverse effect," and, "one causal step" requirements. For reasons previously considered, however, these are not necessarily supported. The second reason the belligerent nexus concept is questioned is perhaps more controversial. It is in reaction to the ICRC claim that,

104. *Id.* at 734–35.
105. *Id.* at 739.
106. ICRC Guidance, *supra* note 19, at 61.
107. Schmitt, *supra* note 29, at 735.
108. ICRC Guidance, *supra* note 19, at 63.
109. Schmitt, *supra* note 29, at 735.
110. *See* ICRC Guidance, *supra* note 19, at 58–64.
111. Schmitt, *supra* note 29, at 735.

> [a]s the determination of belligerent nexus may lead to a civilian's loss of protection against direct attack, all feasible precautions must be taken to prevent erroneous or arbitrary targeting and, in situations of doubt, the person concerned must be presumed to be protected against direct attack.[112]

The ICRC state that this applies *a fortiori*, and that there is strong evidence to support it.[113] In contrast, however, Schmitt notes that the ICRC fails to provide "any basis in law" for taking such a position.[114] He suggests instead, as the Israeli Supreme Court identified in the influential *Targeted Killings* Case,[115] that in order to "enhance the protection of the civilian population,"[116] the opposite is in fact true.[117] In other words, in order to best protect the civilian population "[g]ray areas should be interpreted liberally, i.e., in favor of finding direct participation."[118]

There is no question that the attacker must take all feasible precautions to determine whether a person is a civilian, as identified by Article 57(2)(a)(i) of Additional Protocol I.[119] There is also no doubt that Article 50(1) Additional Protocol I provides, in the general targeting sense, that in cases of doubt a person shall be considered a civilian.[120] However, once again, it is argued that the ICRC's approach here is too narrow.[121] The very fact that a DPH assessment is necessary means that civilian status is not in doubt. Instead, and crucially, this "situation only arises at the border between direct and indirect participation."[122] As noted by the Israeli Supreme court, a civilian who does not wish to be identified as participating, has a duty to remove his or herself from the geographical location in which their actions might be called into

112. ICRC Guidance, *supra* note 19, at 64, 74–76.

113. *Id.* at 76.

114. Schmitt, *supra* note 29, at 737.

115. *See* Targeted Killings Case, *supra* note 88.

116. Schmitt, *supra* note 9, at 509.

117. Schmitt, *supra* note 29, at 737–38. Schmitt notes that in the Targeted Killings Case, *supra* note 88, at ¶ 34 , the Israeli Supreme Court cites a passage from Schmitt's earlier essay, *see* Schmitt, *supra* note 9, at 509, in support of this statement.

118. Schmitt, *supra* note 29, at 737.

119. *Id.* at 736.

120. *Id.* at 737. *See also supra* Figure 1.

121. Schmitt, *supra* note 29, at 737.

122. *Id.* at 738.

question.[123] If they do not, then it is them who may need to carry the burden of risk.[124]

Although IHL is not intended to bind individual citizens, it is nevertheless arguable that it should encourage them, whenever and wherever possible, to resist becoming entangled in the perils of warfare. Boothby is generally supportive of this standard,[125] underlining the dangers of balancing the scales too heavily in favor of humanitarian considerations.[126] Like Schmitt, he argues that there is no legal requirement that a civilian is presumed not to be directly participating,[127] and instead notes that, "[d]etermining whether a civilian is so participating will always be a question of fact."[128] Nevertheless, in sum, both the "narrow" and "wide" interpretations identify a requirement for a belligerent nexus. The two models can be holistically represented as follows:

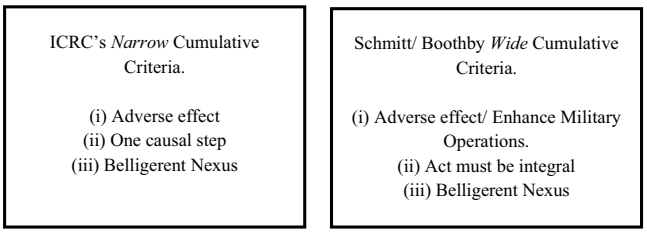

FIGURE 4: Narrow and Wide Cumulative Criteria.

As noted, however, although all parties agree on the need for the third requirement, they do not agree on the presumption of civilian status—an individual, who may already have been identified as a member of the civilian population, is suspected of directly participating in hostilities. These two positions are therefore represented as follows:

123. Targeted Killings Case, *supra* note 88, at 496 (the Israeli Supreme court noting that "a liberal approach creates an incentive for civilians to remain as distant from the conflict as possible - in doing so they can better avoid being charged with participation in the conflict and are less liable to being directly targeted").

124. Schmitt, *supra* note 29, at 738.

125. Boothby, *supra* note 29, at 156.

126. *Id.* at 164.

127. *Id.* at 150.

128. *Id.*

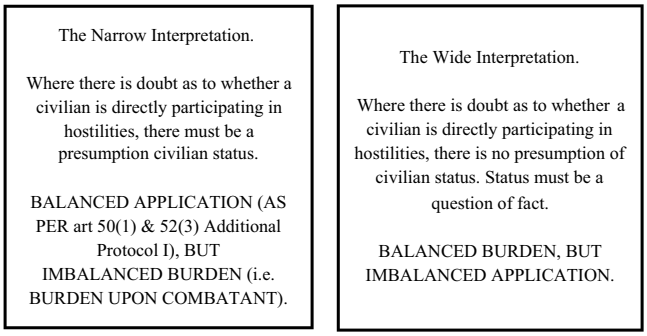

FIGURE 5: The presumption of civilian status for individuals suspected of DPH.

A summary of the discussion in the previous section is contained within the graphical representation below:

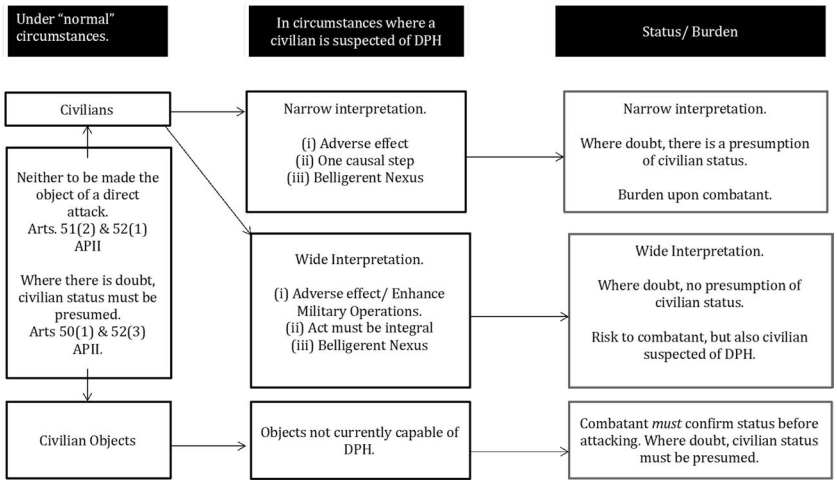

FIGURE 6: The section II discussion.

III. APPLYING THE TESTS TO EAIS: CAN ROBOTS PLAY A DIRECT PART IN HOSTILITIES?

The previous section identified the concept of DPH and its fit within the wider corpus of IHL, including, most notably, its relationship with the principles of distinction and proportionality. The remainder of this Article, uniquely, considers the concept of DPH in light of the increasing introduction of EAIs.

This section will demonstrate that the inevitable introduction of increasingly independent EAI's will begin to soften the once clear distinction between civilians, and civilian property. In addition, it will also

highlight the fundamental weakness of the ICRC's narrow interpretation, that if followed, would somewhat perversely afford a greater protection to EAI's under IHL than human combatants.

In an attempt to make the discussion surrounding EAI-DPH more tangible, the analysis is carried out via the use of a number of scenarios. These four scenarios each identify an increasingly complex EAI technology. The narrow and wide interpretations are considered in each of these scenarios, in order determine the difference in the level of protection offered, and to identify whether there is a potential for robot participation. In short, the scenarios consider whether there is a threshold at which civilian property should simply be referred to as civilian.

A. *The Requirement for an Additional Test*

In order to assess whether or not future EAIs could be considered as civilians, this section utilizes an adaptation of the Turing test. The adaptation is necessary since the test will not be applied in a closed environment, as per Turing's model, but instead where the EAI in question is in full view.[129]Turing's paper was originally introduced in 1950, and the "imitation game" that it provides is still widely in use today as a method of identifying machine intelligence. There are various ways in which the test can be structured, and also many levels of machine intelligence. However, at its heart the test requires that a machine is capable of exhibiting behavior that is equivalent to, or indistinguishable from, human behavior.

The version of the test considered in the scenarios determines whether the EAI in question is demonstrating sufficient human-like qualities (even if it is clearly not human), in order for the DPH tests (both "narrow" and "wide") to be considered. This is referred to as the "but for test" which asks: *but for the fact that the EAI in question can be visually identified as non-human, does it nevertheless exhibit behavior, or behaviors, that are the equivalent to, or indistinguishable from, human behavior(s)?*

129. Turing refers to his test as the "imitation game," and in its simplest form it is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator is in a room set apart from the other two. The object of the game is for the interrogator to determine which of the other two is the man, and which is the woman. A.M. Turing, *Computing Machinery & Intelligence*, 59 MIND: Q. REV. OF PSYCHOL. & PHIL. 433, 433 (1950). For a recent analysis of Turing's work with regards to the "digital threat," *see* Timothy Snyder, *What Turing Told Us About the Digital Threat to a Human Future*, The NEW YORK REVIEW OF BOOKS DAILY (May 6, 2019), https://www.nybooks.com/daily/2019/05/06/what-turing-told-us-about-the-digital-threat-to-a-human-future/.

This test is reminiscent of a version offered by Jens David Ohlin, who also considers a version of Turing's test for identifying AWS in a 2016 Article.[130] Ohlin also recognized that at some indeterminate point in the future, circumstances are likely to dictate that a test will be required for "evaluating the status of artificial agents as rational agents."[131] Noting that it would be absurd to suggest that an AWS must be demonstrated to be "physically indistinguishable" from a human beings, he instead offers that the test should consider the functional similarities.[132] Consequently, Ohlin test questions whether "the non-linguistic behavior of the AWS—*acting at a distance*—would be functionally indistinguishable from any other combatant engaged in an armed conflict."[133]

Critics may argue that the imposition of a further objective test clouds, rather than clarifies, the concept of DPH. However, it may also provide supporting evidence for the proposition that DPH assessments are best made on a case-by-case basis. It is, perhaps, logical to conclude that the evolutionary trajectory of EAIs will result in very few visual clues to enable the targeteer to distinguish between a human and an EAI. In such cases, advanced EAIs (though clearly still objects), will have to be treated in exactly the same way as humans. A graphical representation is presented in the graphic below.

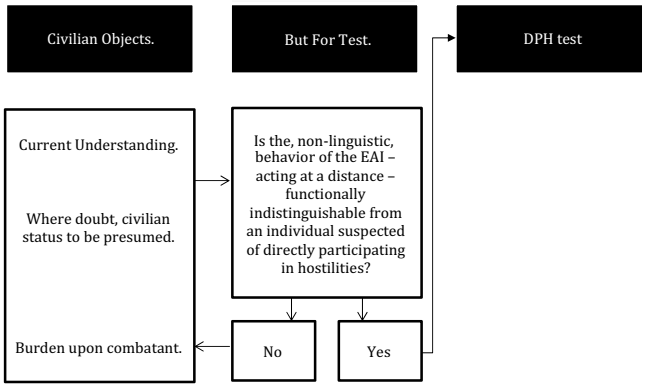

FIGURE 7: The Additional test.

130. Jens David Ohlin, *The Combatant's Stance: Autonomous Weapons on the Battlefield*, 92 INT'L L. STUD. 1, 14–21 (2016).

131. *Id.* at 14.

132. *Id.* at 16.

133. *Id.*

The remainder of this section introduces four increasingly temporally distant scenarios, thus, increasingly "intelligent" machines in each. In each instance, the basic scenario is introduced, the but for test applied, and the question is asked: *can this EAI be identified as directly participating in hostilities, and if so, what are the consequences*?

B. *Existing AI Tech*

Basic scenario: In support of a party to armed conflict, civilian (A) develops, and releases, a malicious code with the use of a smart phone (which, for the purpose of the current Article, the authors consider to be a basic EAI). The malicious code is intended to lay dormant until the program detects weakness in the targeted parties digital firewall. Once triggered, the malicious code is intended to paralyze the targeted party's military, and civilian, operating systems. These include, for example, Air Traffic Control (ATC) systems. N.B. This scenario is confined to examining the possible DPH-ing of a (A). Evaluating the specifics of an attack within the cyber realm/ "fifth dimension" is outside the scope and remit of this Article.

In this scenario, the first question to address is whether the act will cause the required threshold of harm? The "narrow" test requires the harm must adversely affect a party to the conflict, while the "wider" interpretation suggests that harm may also be caused by enhancing the military capabilities of a party to an armed conflict. The act in question is designed to cause physical and functional damage to both military and civilian objects and operations.[134] Therefore, the harm in this scenario meets the required threshold of harm according to either interpretation, and the guidance suggests "the threshold requirement will generally be satisfied regardless of quantitative gravity."[135]

The second criteria requires an examination of the relationship between the act and the damage caused or intended. The "narrow" interpretation requires direct causation between the act and the expected harm.[136] However, due to the fact that there will be an unspecified delay, from the moment (A) launches the code, to the moment of "impact" the question is whether the use of the AI to trigger the attack breaks the "causal chain" preventing DPH?

Quite simply, the answer is no. As the Guidance correctly notes, a number of methods can be employed to delay the application of force

134. *See* ICRC Guidance, *supra* note 19, at 48.

135. *Id.* at 47.

136. *See id.* at 53.

in contemporary warfare.[137] In addition, the guidance distinguishes direct causal proximity, from temporal or geographical proximity.[138] Consequently, and due to the "relationship between the employment of such means and the ensuing harm," causal proximity in this situation would remain.[139] It should be noted that where the harm is yet to materialize, direct causation would be determined with reference to the harm that is "likely" since it is impossible to gauge *actual* harm. Since the "narrow" causal step is satisfied, the "wider" requirement is also clearly triggered—in other words, the act is an integral part of the damage caused, or, anticipated.

The details of the scenario state that belligerent nexus is satisfied because the civilian is said to be acting in support of a party to a conflict, with the intention to harm another. With that in mind, according to both the "narrow" and the "wide" interpretation, (A) can be identified as playing a direct part in hostilities. However, in this scenario, (A) will lose their protected status, meaning they can be lawfully targeted (subject to temporal considerations), and will be accountable for their actions under the municipal legal system. Nevertheless, a malicious code would clearly not pass the "but for" test.

This scenario demonstrates the difference between embodied and non-embodied AI, while also highlighting that for an EAI to directly participate, it must be capable of sincere, independent participation. At this stage, one can conclude that basic EAI systems are inadequate to be recognized as anything other than tools.

C. *Near-Term Future Tech: Driverless Vehicle Technology*

Basic Scenario: A party to the conflict "contracts" the use of a driverless truck in order to deliver material sensitive to the ongoing military operation. The materials in question are to be delivered to a storage facility, which though not located at the front-line, is in close proximity to an existing battlefield.

In this second scenario, a civilian in support of a party to the conflict (and to the detriment of another), may, for example, place an

137. The guidance notes "it has become quite common for parties to armed conflicts to conduct hostilities through delayed (i.e. temporally remote) weapons-systems, such as mines, booby-traps and timer-controlled devices, as well as through remote-controlled (i.e. geographically remote) missiles, unmanned aircraft and computer network attacks." *Id.* at 55.

138. The report refers to temporal or geographic proximity as merely indicative elements. It makes sense therefore that it also provides "while temporal or geographic proximity to the resulting harm may indicate that a specific act amounts to direct participation in hostilities, these factors would not be sufficient in the absence of direct causation." *Id.*

139. *Id.*

explosive device in a driverless vehicle, and set it to detonate at a pre-programmed set of GPS coordinates. Though in such circumstances, as before, a human can be identified as directly participating in hostilities (subject once again to geographical and temporal considerations). Furthermore, in such circumstances, an autonomous civilian truck *may* by its nature, location, and purpose become a legitimate military target (though it is vital to restate that currently, when in doubt the truck must be presumed to be a civilian object). In this instance, it is quite unlikely that the truck would satisfy the "but for test," meaning the second DPH test could not be applied. Instead, this scenario requires the focus to be placed upon the question of determining civilian status, rather than questioning whether the truck itself is directly participating.

The utilization of civilian autonomous vehicle technologies by militaries does raise two intriguing questions. The first is in relation to the temporal and geographical proximity of the autonomous delivery vehicle to the battlefield, while the second is in relation to the status of PMCs. Both of these issues were discussed at the expert meetings, and in subsequent literature. If the term "civilian driver" includes autonomous civilian trucks, both the narrow and the wide interpretation would require the following (in order for the truck to become a legitimate target). The cargo must be identified as being military in nature. Clearly, if an autonomous civilian truck were merely transporting civilian resources (even if the movements were located close to an active battlefield), both the truck and its contents will remain protected against direct attack. Were the attacking party furnished with sufficiently accurate intelligence, however, in order for the truck to lose its protected status its proximity to the battlefield is also likely to become relevant.

According to the guidance, a truck delivering ammunition to the front line is likely be considered targetable, while one which was delivering ammunition to a storage facility outside of the conflict zone is likely to be considered otherwise.[140] Schmitt appears to agree with this assessment[141] and goes on to highlight a number of examples. He notes, for example, that U.K. military doctrine suggests civilian drivers of military transport vehicles are not participating,[142] though again, this is likely to be judged on a case-by-case basis. This is reinforced by the

140. *Id.* at 56.

141. *See* Schmitt, *supra* note 29, at 705–06.

142. *Id.* at 706. UNITED KINGDOM MINISTRY OF DEFENSE, THE MANUAL ON THE LAW OF ARMED CONFLICT, ch. 5.3.3 (2005).

ICJ Appeals Chamber, which has identified that a civilian transporting weapons geographically proximate to combat operations would be directly participating in hostilities.[143] There does not appear to be a reason to suggest the same would not be true in terms of military contracting of autonomous civilian vehicles, however, it is also relevant to the second and further discussion regarding PMCs.

While professional soldiers still take the large majority of battlefield decisions,[144] a number of other services including cooking, cleaning, and security services are regularly carried out by non-combatants. As a result, when there is doubt as to the status of a PMC, the "narrow" interpretation holds that civilian status must be presumed, while the "wide" interpretation suggests that is not the case if they are suspected of DPH.

Nevertheless, while a cook and a cleaner are probably far enough removed to be incapable of directly participating while carrying out their normal duties, the same conclusion would not readily apply to a security guard. By way of reminder, *all* individuals can be indirectly targeted lawfully if they are in geographical proximity to a legitimate military target.[145] The question of whether the security guard is directly participating must be a matter of context. If, for example, they are guarding an army recruitment center in a location far removed from the battlefield, then it must be that they are not directly participating. However, if they were contracted to guard the perimeter fence to a large military encampment on, or near, to a battlefield, then they may well be DPHing—depending of course on the facts.

In applying the "wide" interpretation, there may be limited circumstances where there is no presumption of civilian status for humans

143. Prosecutor v. Strugar, Case. No, IT-01-42-A, Appeals Chamber Judgment, ¶ 177 (Int'l Crim. Trib. for the Former Yugoslavia, July 17, 2008) (identifying Hague Rules of Aerial Warfare, art. 16).

144. For example, the CIA is a civilian, not a military, organization. Nevertheless, it is in possession of armed UAVs. Furthermore, it is reported that the U.S. President has granted the agency permission to engage in strikes without the need for the military to be involved. *See* Harriet Agerholm, *Donald Trump Gives CIA Power to Carry Out Its Own Drone Strikes*, THE INDEPENDENT (Mar. 14, 2017) https://www.independent.co.uk/news/world/americas/donald-trump-cia-power-drone-strikes-military-a7628561.html.

145. Subject, of course, to a proportionality assessment. *See* ICRC Guidance, *supra* note 19, at 37, which states that, as civilians, PMCs are protected against direct attack unless DPHing. However, due to their location or activities, they may nevertheless expose themselves to increased risk of death or injury even where they are not – i.e., as the result of a proportionate attack; Boothby, s*upra* note 29, at 159–60, and; Dinstein, *supra* note 38, ¶ 374, where the author identifies the Montreux Document is clear in stating that PMCs retain civilian their status, providing they are not incorporated directly into the armed forces, or are DPHing. However, note this is non-binding document.

while there is for autonomous civilian property. This is because (and as demonstrated), the autonomous truck is only potentially targetable under both the narrow and wide interpretations, and dependent upon the facts (including the need for firm intelligence). In contrast, however, when the "wide" interpretation is applied to the latter security guard, they are lawfully targetable - subject to their proximity to a battlefield and/or their individual behavior.

One solution to address this imbalance might be to round-up classification to afford the same PMC status to an EAI, as is afforded to a human PMC. This being the case, subject *inter alia* to Article 57 Additional Protocol I,[146] the "wide" interpretation would allow the targeting of the truck—though it would still need to be assessed on a case-by-case basis. Those in support of the "narrow" argument will, of course, argue their interpretation offers no imbalance—civilian status is presumed for both civilian objects and for individuals suspected of DPH. But, while that may be true, the counter-argument is that the "narrow" model places an increased risk upon the lawful combatant because he or she cannot engage based merely on suspicion of participation, even where there may be some evidence to the contrary. Nevertheless, autonomous vehicles cannot be demonstrated as having a sufficient degree of autonomy in order to independently participate.

D. *Mid-Term Future Tech: Advanced Life Support Systems*

Basic scenario: A healthcare EAI (X), is programmed to protect and sustain the life of individual (A), who is immobile due to illness. The home of (A), in which (X) cares for (A), is located in a dense urban environment. As the result of a lawful attack on a neighboring military operations base, (A)'s home is severely damaged. Neither (X) or (A) are harmed as a result of the attack. Nevertheless, (X) elects to remove (A) from the building, having determined that (A)'s safety would be threatened if he remained inside the building. Once out in the open, (X) detects a number of members of the attacking force approaching. Certain that (A)'s life is at risk from the approaching force (X), chooses to engage them, and a number of combatants are injured as a result.

146. Additional Protocol I, *supra* note 10, art. 57 provides that an attacker must consider a number of precautionary measures. Additional Protocol I, *supra* note 10, art. 57(2), for example, states, "a) Those who plan or decide upon an attack shall: (i) Do everything feasible to verify the objectives to be attacked are neither civilians nor civilian objects and are not subject to special protection . . . [and] . . . (ii) Take all feasible precautions in the means and methods of attack with a view to avoiding, and in any event minimizing, incidental loss of civilian live, injury to civilians, and damage to civilian objects . . .," in addition to the principle of proportionality previously considered.

This scenario is based upon an EAI system that is superior to any existing technology (although EAIs such as "Pepper" are already commonplace).[147] (X) is clearly programmed with highly sophisticated decision-making capabilities. It is when EAIs such as this are inevitably introduced, that they will begin to blur the distinction between civilians and civilian objects. In this scenario it is unclear how (X) engages the approaching force. Consequently, the response of the combatants may need to change depending upon whether (X) was, for example, throwing stones, firing bullets, or launching rockets. However, as previously noted, civilian property is currently incapable of independent *participation* in armed conflict. With that in mind, the approaching combatants may lawfully attack (X) if, due to its nature, location, purpose or use, it made an effective contribution to a military action.[148] Nevertheless, even if (X)'s partial or total destruction, capture or neutralization, in the circumstances ruling at the time,[149] did offer a definite military advantage, it would still be difficult to identify that (X)'s behavior was making a contribution to a military action. Furthermore, should there be a doubt as to the nature of (X) as an object, there is currently a presumption that (X) is a civilian object, and is protected against direct attack.[150]

The authors propose that it is at this juncture that there is a need to invoke the "but for test". Given that (X) is clearly acting independently of human control, the answer must be "yes", the non-linguistic behavior is functionally indistinguishable form an individual suspected of DPH. Answered positively means that, (X), and (X)'s actions can be assessed

147. "Pepper" is a line of robots, whose creator suggests can "assist patients in self-diagnosis, support staff in health trending & monitoring . . . [and] . . . are also the platform for telemedicine and the hub of information distribution (alert, notifications, fall & sound detection, etc. . . . .)" *Typical Use Cases in the Healthcare*, SOFTBANK ROBOTICS, https://www.softbankrobotics.com/emea/en/industries/healthcare (last visited Mar. 20, 2020).

148. *See* Additional Protocol I, *supra* note 10, art. 52(2); HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 8 (identifying this as customary international law).

149. Additional Protocol I, *supra* note 10, art. 52(2). The points is, Additional Protocol I, *supra* note 10, art. 52(2) does not suggest *or (emphasis added)* . . . whose partial or total destruction . . ., but, *and (emphasis added)* . . . whose partial or total destruction (emphasis added). It should also be noted that the treaties suggest there are a number of differences with regard to the protection of civilian objects depending on whether the armed conflict is international or non-international in nature. Nevertheless, it is arguable that CIL dictates that all civilian objects are nevertheless provided with the same level of protection. For a useful discussion, see Noam Zamir, *Distinction Matters: Rethinking the Protection of Civilian Objects in Non-International Armed Conflicts*, 48 ISR. L. REV. 111 (2015).

150. N.B The authors do of course concede that it may not be obvious to distinguish between medical EAI and military EAI.

according to the 3 criteria of each of the DPH tests. In other words, (X) can considered along the same lines as a civilian suspected of directly participating in hostilities.

The "narrow" interpretation requires that (X)'s act must adversely affect the military operations of a party to the conflict, while the "wider" interpretation requires that the act should either adversely affect one party or enhance the military operations of another. In this scenario, a number of combatants have been injured as a result of (X)'s actions, though it is unclear as to the exact quantum. However, due to the fact that the attacking force has been harmed, the narrow requirement appears to be satisfied.

The next requirement is a) whether there is one causal step between the act and the harm caused, or b) whether the act constitutes an integral part of the conduct that leads to the harm. The answer in either instance is yes. By engaging the combatants directly, there is clearly only one causal step between the act and the harm caused. The third and final requirement is that there must be a belligerent nexus. This, as noted, is the least contested of the three criteria; however, in this case it may be difficult to satisfy.

As the ICRC recognises, civilians remain protected against direct attack insofar as they are acting in self-defense or in defense of others protected against attack. Assessment as to how soldiers distinguish between an EAI acting in self-defense and an EAI that is an enemy combatant must be done on a case-by-case basis—this ties in with the authors' overarching approach that all assessment must be *sui generis*.[151] As previously noted, where self-defense or the defense of others is a factor, the belligerent nexus cannot be satisfied, as the act is not intended to support a party to the conflict, but merely to protect the lives of protected persons. This remains the case even if the act is to the detriment of another. Therefore, where self-defense can be established as being the reason for carrying out the act, the act should not be seen as DPH and the "civilian" should keep their protected status.

However, on the facts, there is no suggestion that either (X) or (A) were the object of a direct attack. While (A)'s property is damaged, the collateral damage is the result of a lawful attack. Consequently, (X)'s act is unlikely to be considered as an act of self-defense. Given (X)'s

151. The ICRC suggest for example that "[i]f individual self-defence against prohibited violence were to entail loss of protection against direct attack, this would have the absurd consequence of legitimizing a previously unlawful attack. Therefore, the use of necessary and proportionate force in such situations cannot be regarded as direct participation in hostilities." ICRC Guidance, *supra* note 19, at 61.

subjective state, the question remains as to whether the belligerent nexus can nevertheless be satisfied. The guidance states that in order to be classified as DPH, the act in question must be *objectively* likely to cause the required threshold of harm and must be designed to support one party to the conflict to the detriment of another.[152]

(X)'s actions do not appear to support a party to the conflict even if it is to the detriment of another. However, the guidance continues, "[b]elligerent nexus should be distinguished from concepts such as subjective intent . . . ."[153] According to the guidance, at the expert meetings there was an almost unanimous agreement that in the fog of war it is almost impossible to determine the subjective reasoning of the individual carrying out the act.[154] As a result, the report aligns with that sentiment and suggests that subjective intent "cannot serve as a clear and operable criterion for 'split second' targeting decisions."[155]

In considering the practical determination of the belligerent nexus, the guidance identifies a "grey zone where it is difficult to distinguish hostilities from violent crime unrelated to, or merely facilitated by, the armed conflict."[156] It continues, that where a party is finding it difficult to establish the belligerent nexus, the question must be: can the civilian's conduct be reasonably perceived as being designed to cause the required threshold of harm to one party while in support of another? Importantly, as previously identified, according to the "narrow" interpretation, where there is doubt, civilian status should be presumed.

The determination of whether or not (X) satisfies the narrow interpretation of the belligerent nexus remains uncertain due to factual considerations. In contrast, the "wide" interpretation questions whether harm is a prerequisite element of DPH. [157] Schmitt suggests instead that the threshold criteria would be "better styled as acts 'in support of a party to the conflict *or* to the detriment of another.'"[158] And, if that theory is applied to Scenario 3, (X)'s act alone might typically be sufficient to satisfy the belligerent nexus requirement. Consequently, while it is questionable whether (X) could be identified as playing a direct part in hostilities under the "narrow" interpretation, all three of the criteria appear to be satisfied the "wide" model. Subject

152. *Id.* at 58.

153. *Id.* at 59.

154. *See id.* at 60.

155. *Id.* at 59, n. 150.

156. *Id.* at 63.

157. Schmitt, *supra* note 29, at 736.

158. *Id.* at 736 (internal emphasis omitted). This is expanded upon further in the discussion relating to scenario 4.

to all of the usual qualifications, the wide interpretation means (X) would lose its protected status and become lawfully targetable, even though it is not, prima facie, a military object.

It might still be possible that doubt may remain as to the status of (X). And, it is here that the lacuna becomes increasingly more apparent. As previously noted, the guidance insists on a presumption of civilian status, even though this is in contrast to observations of the Israeli Supreme Court,[159] which identified that the opposite may in fact be true. In addition, civilian property is currently incapable of DPH. Therefore, where there is doubt, the presumption must be of civilian status. This means that an EAI, in this case (X), will, in certain situations, be protected against direct attack. This may be the case even where the human combatant is faced with an increasing risk of injury or even death.

Furthermore, in addition to the imbalanced burden of risk, the situation will inevitably present itself where a civilian suspected of DPH may be presumed to have lost their protected status, where in contrast, and given an identical set of circumstances, a machine may not. This is clearly a very undesirable application of international law, whose drafters could never realistically have imagined a softening of the boundaries between civilians and civilian property. However, it is a quirk that can easily be overcome by utilizing the "wide model", which it is argued is the most appropriate interpretation of DPH today, as it is in the future. A graphical representation of this discussion is presented below:

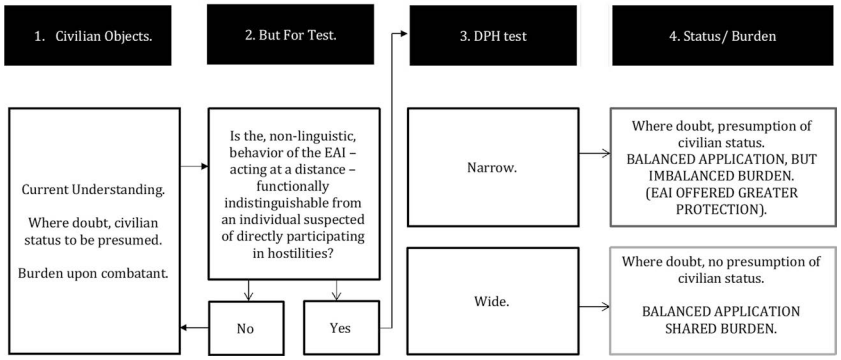

FIGURE 8: How the narrow and wide interpretations affect burden and application.

159. *See* Targeted Killings Case, *supra* note 88, ¶ 34.

E. *Long-Term Future Tech: Advanced Personal Assistants*

Basic scenario: Advanced EAI (B) is manufactured and programmed to learn from, and to react to, the changing needs of the individual(s) to whom it is charged. In order to do so, (B) is equipped with AGI, which allows for it to carry out a potentially infinite number of tasks in order to support and often replace human involvement. The home in which (B) is operating is located in a dense urban environment inside state (Y), which is currently engaged in an NIAC. While carrying out its daily tasks, (B) occasionally "over-hears" conversations between two of the individuals to whom it is tasked. Those conversations regularly refer to a desire, if there was an appropriate way of doing so, of supporting the NSAG.

In order to satisfy the will of its principles, but without their knowledge, (B) occasionally finds ways of facilitating the efforts of the NSAG so that State (Y) can be more quickly defeated. (B) is capable of manipulating simpler AI systems such as driverless cars and GPS systems in order to carry out autonomous acts which cause injury and death, to citizens of State (Y), as well as damage to a number of State (Y)'s military objectives. In addition, (B) is able to alter the destination of international weapons shipments and international financial transactions. As a result of these actions, the NSAGs war effort has been considerably enhanced.

Given the facts of this scenario, it is highly plausible that the individual tasked with assessing (B)'s intelligence would determine that the "but for test" is easily satisfied. In this instance for example, it is possible to identify the "narrow" ICRC interpretation of the threshold of harm, as the harm is designed to adversely affect a party to the conflict, and cause injury and/or death to protected persons. Equally, as stated by the guidance, the threshold of harm can also arise, inter alia, from exercising control over military objects, by denying the use of certain objects, by killing and wounding personnel, and by causing damage to military objects, and military operations,[160] all of which have happened as a result of (B)'s actions.

As far as the application of force is concerned, the "narrow" requirement for one causal step is satisfied. Though, the ICRC suggest the re-routing of finances and weapons would likely be considered too indirect to be classed as DPH, unless, they are carried out as an integral part of a specific operation that was designed to cause the threshold of harm.[161]

160. ICRC Guidance, *supra* note 19, at 48.

161. *Id.* at 53.

In contrast, the "wide" interpretation identifies that harm may also be caused when a party to the conflict benefits. Additionally, and as noted by Schmitt, this interpretation recognizes that *harm* does not necessarily have to be an act of violence.[162] When wide interpretation is applied, (B)'s actions, whether direct or indirect, are an integral part of the conduct which leads to the harm caused. Nevertheless, due to the combined acts, the first two of the cumulative criteria are satisfied.

Given the facts, the belligerent nexus also appears to be present. (B)'s acts are clearly intended to support a party to the conflict to the detriment of another. Therefore, as long as the law can be applied in such a way that allows civilian property to be capable of participation, (B)'s actions will certainly amount to DPH. Furthermore, should a future EAI, such as (B), take a form that would make it difficult to distinguish it from a human, that *must* also be the only logical de facto position.

While it seems that (B) should be seen as directly participating, there are still a number of other factors that need to be considered. The first of these, according to the ICRC, concerns the temporal nature of (B)'s acts. In other words, at what point in time would it be lawful to target (B)? According to the ICRC, where a civilian plays an occasional role such as the one described in Scenario 4, "[t]he phrase 'unless and for such time' clarifies that such suspension of protection lasts exactly as long as the corresponding civilian engagement in direct participation in hostilities."[163]

The ICRC, therefore, claim that the suspension of (B)'s civilian status should only last as long as each corresponding engagement.[164] This concept is what is commonly referred to as the "revolving door fighter," and the concept is encapsulated by the analogy of "farmers by day, and fighters by night."[165] According to the guidance, the fighter is targetable, but the farmer not. When the guidance is applied to the facts of Scenario 4, and because (B)'s involvement is only occasional, it is fair to say that it would be very difficult to identify that (B) is the 'individual' responsible for DPHing.

Furthermore, (B) is at no point present, either at the location where force is applied or where weapons and finance transactions are interrupted. This is clearly not a dilemma that is confined to EAIs, but one

162. Schmitt, *supra* note 29, at 716.

163. ICRC Guidance, *supra* note 19, at 70.

164. *Id.* The ICRC suggests that such loss and regaining of protection against direct attack is an integral part, not a malfunction of IHL.

165. *Id.* at 5, 72.

that occurs in remote operations generally, and it is compounded by the concept of plausible deniability. As a result, it can be almost impossible to identify the individual at the time of their participation. In reality, this means that the lawful combatant, and potentially the wider civilian community, may often be left to carry the increased burden.

The concept of the "revolving door fighter" does not sit comfortably, least of all with Boothby.[166] He disagrees with the ICRCs suggestion that the revolving door mechanism is necessary because an individual's past behavior is not necessarily a reliable indicator of future behavior.[167] In contrast, he vehemently argues that past behavior is *absolutely* relevant in determining whether or not a civilian could be constantly targetable, and to say otherwise renders the law unrealistic.[168]

The analysis of Scenario 4 has had, and are likely to continue to have, an appreciable impact upon the fate of both parties. It does seem a little strained to reach the conclusion that (B), a non-human, should only be targetable during the preparation of a specific act, when a potentially high number of humans, and property, combatant and civilian, might be injured or damaged as a result its actions. This argument is compounded, by the fact that even if the acts could be attributed to (B), and as a result (B) was monitored closely, it might still be almost impossible to pinpoint the exact moment of participation. In addition to which, the fact that (B) appears not to be involved in a particular physical activity in which the period of lawful targeting could be extended to preparation, or return.

If the ICRC's guidance is followed, (B)'s actions are likely to continue to damage State (Y)'s lawful military campaign, injuring human combatants and civilians. In the meantime, according to the guidance, the EAI, (B) would only loose protection against direct attack sporadically. As a reminder, this is contrary to the status of the lawful combatant, who would remain targetable at all times. This is not a satisfactory state with regards to an EAI, or indeed, to a human who makes the conscious decision to become involved in an armed conflict in a clandestine manner, where he or she could, and should, have decided otherwise.

Consequently, it must be the case that where a civilian can be identified as having participated on more than two of occasions, the revolving door should be locked. As a way of further discouraging DPH, the civilian should lose their civilian status and place them on "equal targetable balance" with that of a lawful combatant. Clearly, this must be subject

166. Boothby, *supra* note 29, at 161–62.

167. ICRC Guidance, *supra* note 19, at 71.

168. Boothby, *supra* note 29, at 162.

to a caveat that a civilian must be able to demonstrate through their actions that they no longer intend to participate.

The ICRC does recognise a set of circumstances in which a civilian would remain targetable at all times following DPH. However, according to them, in order for that to happen (B) would have to have a continuous combat function (CCF), which is inclusive of the further requirement that (B) must also be a member of an organized armed group. The ICRC states that,

> [i]n non-international armed conflict, organized armed groups constitute the armed forces of a non-State party to the conflict and consist only of individuals whose continuous function it is to take a direct part in hostilities ("continuous combat function").[169]

As previously noted, the belligerent nexus exists. Nevertheless, in the scenario, (B) has not been recruited, nor in fact, does it appear that anyone is aware of his involvement. With that in mind, though the ICRC guidance is once again too narrow.[170] For many it would be a push too far to suggest that (B) is a member of the NSAG in opposition to State (Y). In addition, with regard as to whether (B) could be said to have a CCF, the guidance continues,

> [c]ontinuous combat function does not imply *de jure* entitlement to combatant privilege. Rather, it distinguishes members of the organized fighting on a merely spontaneous, sporadic, or unorganized basis, or who assume exclusively political, administrative or other non-combat functions.

Bearing that in mind, it is unlikely that the narrow interpretation would identify (B) as being attached to an organized armed group, or even of having a CCF. Indeed, it might even be fair to say that (B)'s involvement is unorganized, and sporadic. However, concluding that CCF requires that (B)'s sole function should be DPH, is wrong.[171] What

169. ICRC Guidance, *supra* note 19, at 27.

170. In fairness, the ICRC do consider a number of ways in which a person could become attached to an organized armed group. They note for example that "there may be various degrees of affiliation with such groups that do not necessarily amount to 'membership' within the meaning of IHL. In one case, affiliation may turn on individual choice, in another on involuntary recruitment, and in yet another on more traditional notions of clan or family." *Id.* at 33. However, it is unclear how the guidance would cater for an individual (or EAI) acting in a void.

171. *Id.* at 36.

if, for example, (B) was capable, without the knowledge of its principles, of manipulating a number of other AI systems on a continuous basis, while at the same time fulfilling the role of personal assistant? Surely, then, the requirement that organized armed groups must only be composed of "individuals whose continuous function it is to take a direct part in hostilities" is compromised.[172]

Boothby is highly critical of the guidance in this respect. He suggests that the "unacceptable imbalance" between the lawful combatant, and "persons" without a CCF, or attached to an organized armed group, is the central flaw to the ICRCs guidance.[173] To suggest that certain individuals are only liable to attack "while undertaking specific acts of DPH, and during very limited associated periods," is unfair, and prejudices one side to the conflict,[174] which as demonstrated above, is the lawful, participant. In contrast to the ICRC, Boothby notes, that while it is clear that membership of an armed group should not be determined arbitrarily, that does not necessarily mean that it must instead be determined by a continuous "active and hostile involvement." Furthermore, he adds, there is no evidence, nor is it necessarily logical, that the CCF should be attached to the NSAG.[175]

In a further article Schmitt notes that the ICRCs concept of CCF places too great a weight on the humanitarian side of the military necessity, humanitarian scale.[176] This is because, "even in the face of absolute certainty," the combatant must make a determination regarding the membership of an individual, while in contrast, a member of the armed forces can be targeted due to relationship alone (even when that individual is behaving in such a way that it would not be considered DPH when judged by civilian standards).[177]

The concerns of Boothby and Schmitt become more apparent when the imbalance they discuss is weighted in favor of machines. It cannot be right that an EAI should be offered a greater level protection against attack than a human in any circumstances, but especially when the life of a human is in immediate peril. In reality, when in doubt as to the status of an EAI, a human combatant is likely to react instinctively, and in a way that is designed to protect their own life, as well of the lives of the individuals around them. It would be a gross misapplication of IHL, if

172. *Id.*
173. Boothby, *supra* note 29, at 146.
174. *Id.* at 146–47.
175. *Id.* at 154.
176. Schmitt, *supra* note 36.
177. *Id.* at 23.

the human has to carry an increased burden of risk due to threat of a retrospective examination of the lawfulness of their actions. When the boundaries between human and machines begin to blur, the IRL must fall on the side of the human each and every time. A graphical representation of this discussion would appear thus:

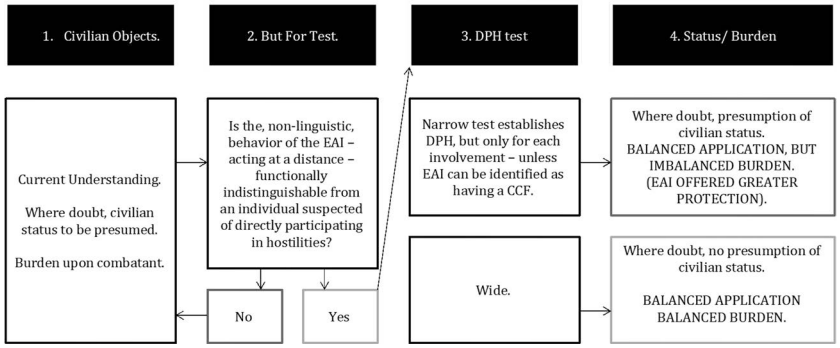

FIGURE 9: Scenario 4 tech capable of DPH according to the narrow and wide interpretations, but differing applications, and burdens.

A further dilemma that needs to be considered if granting civilian status under IHL to an EAI, is both temporal, and geographical, in nature. For example, if either a contracted EAI that was known to have a CCF, or a military operated AWS, was repurposed and used for a time as a non-combat EAI, perhaps for humanitarian purposes, should the EAI/AWS remain targetable? If a human combatant, PMC, or spy, is not considered targetable while on home leave, should the same apply to EAI/ AWS?[178] Though perhaps somewhat of an oversimplification, the question therefore is: can autonomous weapons take a vacation?

This is a question that could become relevant for a number of reasons. The first is in line with the ongoing discussion regarding the boundaries of the contemporary battlefield.[179] Taking into account the fact that by its very nature an autonomous system is often likely to operate at distance from human combatants, should the EAI/AWS remain

178. This is a concept that can be extended further, e.g. to include desertion. As noted by Dinstein "[e]very combatant is a former civilian: nobody is born a combatant. In the same vein, a combatant may retire (or even desert) and revert to the status of civilian." DINSTEIN, *supra* note 38, ¶ 468.

179. *See, e.g.*, Frederic Megret, *War and the Vanishing Battlefield*, 9 LOY. U. CHI. INT'L L. REV. 131 (2011). Megret's discussion is centered upon the how the changing battlefield has contributed to the evolution of the laws of war. *See also*, Christopher M. Sanders, *The Battlefield of Tomorrow, Today: Can a Cyberattack Ever Rise to an "Act of War?"*, 2018 UTAH L. REV. 503 (2018). Sanders assesses how cyber actions are changing the concept of the battlefield.

targetable when it is simply outside of the border of the state involved in an armed conflict, perhaps in an ally's territory, whether that be land, sea or airspace?

A second way in which repurposing might become relevant, is if the machine was capable of acting in a number of different roles. By today's, relatively simple, machine learning standards, this is unlikely. For example, although AlphaZero is potentially the greatest ever player of Go, the ancient Chinese game of strategy, it is not very good at making coffee. What is more, when today's neural networks are repurposed in order to self-learn something else, let us say chess, the rules of Go are simply forgotten, and need to be relearned the next time the game is attempted. In other words, memory chips do not work in the same way as organic memories.

Nevertheless, this is one of the objectives of those working on artificial general intelligence (AGI), and with so many individuals and institutions working towards it, it would perhaps be unwise to say it will never happen.[180] If it does, a question therefore is, if an AWS were capable of carrying out duties of a civilian nature at a distance from a battlefield, what would be the consequences if the system reacted due to its "memory," in order, for example, to protect a civilian in a terrorist situation? This was the case with two off-duty combatants, Spencer Stone and Alek Skarlatos, who were travelling on a train from Belgium to France when a gunman appeared in the carriage holding an AK-47 in August 2015.[181] Along with two other members of the public, Stone and Skarlatos ran at the gunman and disarmed him before he was able to open fire, and held him until the police were able to take control of the situation once the train had been brought to a standstill.

Although the off-duty combatants' actions could never be considered anything other than heroic, the point is that they decided to act as members of the civilian community and not as combatants. Therefore, should their actions have led to a less desirable outcome, it would not have been the U.S. military that questions were being asked of, but the individuals themselves, who would not have been covered by combatant immunity. Consequently, there is a question whether the same principle should apply to AWS and EAIs, and also whether that requires municipal law to recognise the concept of civilian EAIs as well.

---

180. *See, e.g.*, Simon Stringer, *How Close is Science to Replicating Consciousness?* FINANCIAL TIMES (Jan. 10, 2019) https://www.ft.com/content/c082aad6-141e-11e9-a581-4ff78404524e.

181. Angelique Chrisafis, *France Train Attack: Americans Overpower Gunman on Paris Express*, THE GUARDIAN (Aug. 22, 2015), https://www.theguardian.com/world/2015/aug/21/amsterdam-paris-train-gunman-france.

IV. The Wider Consequences of Recognizing EAI Participation in Armed Conflict

The following section considers a number of potential consequences that may arise when AI technologies are introduced into armed conflict. Although it is beyond the realm of the current Article to complete a comprehensive analysis of each, the authors nevertheless consider it vital that they are at least acknowledged. While some sit on the margins of what is considered civilian or military, others are purely civilian in nature. The first part of this section, therefore, considers the concepts of PMCs, and security service personnel, and looks to how those concepts might be affected by the introduction of autonomous technologies. The second discussion returns to the original conversation surrounding DPH, and examines whether or not it is possible that a number of *participating* EAIs could be identified as *levée en masse*, in addition, *inter alia*, to the question of whether or not EAIs should be offered POW status.

A. *Robot PMCs*

PMCs are as much an intrinsic part of the discussion regarding an individual's status now as they were at the time the ICRC guidance was published.[182] As previously noted, this is because civilians continue to carry out a number of roles, some of which are novel,[183] but some of which would have traditionally been undertaken by military personnel.[184] And as previously discussed, a PMC may be employed as a kitchen hand, as an armed guard, or, in a number of other capacities with a varying degree of attachment to the application, or threat, of force.

---

182. Although his analysis primarily concentrates on the municipal legal system, Michael Anderson notes, "PMCs are frequently in the thick of active hostilities and often serve next to actual soldiers fighting 'the field.' Consequently, PMCs are sufficiently integrated into the military so far as to fall under its laws . . . ." Michael Anderson, *If It Looks like a Duck: Reining in Private-Military Contractor Conduct Through the Amended UCMJ*, 50 Case W. Res. J. Int'l L. 307, 347 (2018). For an analysis which focuses upon the increased use PMCs, and the cyber application of force under IHL, see Ido Kilovaty, *ICRC, NATO and the U.S. - Direct Participation in Hacktivities - Targeting Private Contractors and Civilians in Cyberspace under International Humanitarian Law*, 15 Duke L. & Tech. Rev. 1 (2016).

183. Due to the nature of evolving remote technologies, such as cyber warfare and drone warfare, certain systems can be operated at a great distance from the point at which the effects are felt. As a consequence, the operator, or programmer, can be a civilian who has little to no chance of being targeted, either directly, or indirectly.

184. In addition to the sources already noted, see ICRC guidance, *supra* note 19, at 37.

According to the ICRC, in each of these roles, the question as to the targetability of the "civilian" must rest on the application of the DPH test.[185] Furthermore, if the guidance is followed, unless a PMC can be identified as having a CFF, he or she is protected against direct attack at all times other than when the actual participation takes place.[186]

As is the hypothesis of this Article, under the current interpretation of IHL the presumption of civilian status with regard to property is justified. In other words, where there is doubt as to the status of an EAI, it is protected against direct attack. Nevertheless, as demonstrated, this will eventually lead to undesirable consequences. As examined in scenario two, one way around this might be to grant the status of PMC directly to the EAI. In much the same way as assigning civilian status to an EAI in order to apply the DPH test, PMC status may, under the "wider: interpretation, due to the behavior of the EAI and its proximity to the violence, it would mean that the EAI would be lawfully targetable.

In theory, insofar as autonomous technologies are concerned, PMC status might fall a notch below "civilian" status. This may be more acceptable to those who might be critical of the suggestion that EAIs should be required to demonstrate some form of consciousness in order for them to be assessed under a version of the DPH test. However, this would only work where a party to the conflict had intentionally utilized civilian EAI tech, and not where civilian tech was autonomously acting in such a way that it could be construed as DPH.

The result of assigning PMC status to the truck would be the same as if civilian status were applied. This would mean that the lawful combatant, and potentially the civilian population, would not carry the increased burden of risk, and at no point would the truck need to be considered as a civilian for the purposes of IHL. In contrast, it would not be possible to assign PMC status to the personal assistant EAI in scenario 4, due to the fact that its AGI enables it to act without the need for a human to authorize or even to determine its actions.

B. *Robot Spies*

In his discussion regarding the United States' use of drones for the targeted killings of the members and associated forces of al-Qaeda, Jens David Ohlin distinguishes the Central Intelligence Agency (CIA) from the Joint Special Operations Command (JSOC).[187] Ohlin notes that

185. *Id.*

186. *Id.* at 38.

187. Jens David Ohlin, *The Combatant's Privilege in Asymmetric and Covert Conflicts*, 40 YALE J. INT'L L. 337, 337–38 (2015).

while the latter is a military organization under the direction of the U.S. Secretary of Defense, the former is a clandestine organization with intelligence gathering and covert operation responsibilities.[188]

While the operatives of both agencies have been responsible for extraterritorial targeted killings, it is arguable that only those operating under the banner of JSOC are lawful combatants under IHL.[189] There are many sides to this discussion, including the lawfulness of extraterritorial applications of force.[190] However, for the sake of the current discussion, the most pressing point is that although some state security and intelligence services are involved in the application of force in armed conflict, their operatives are, under the eye of IHL, civilians.[191] If the same principle was applied to autonomous technologies, it must be the case that when EAIs or AWS were used under the military banner, they would, subject to IHL, be lawful "participants" in armed conflict. In contrast, and in parallel to its human counterparts, a civilian or security service EAI would not.

Consider a short-term future UAV—one that harnessed existing global positioning, and facial/object recognition technologies that allowed it to operate autonomously, in search of a predetermined person or object.[192] This could be a military UAV in the form of an AWS, or it could, at least potentially, be a contracted civilian EAI. Upon identifying the person or object, such a weapon could, at least in principle, launch an attack against the object or person without further intervention by a human operative.[193]

---

188. *Id.* at 338.

189. *Id.*

190. Schmitt notes elsewhere, for example, that "[a]bsent an applicable exception to the general principle, the mere passage of the military or civilian organs of one State into the territory of another State violates the latter's sovereignty . . . ." He continues, " . . . a State that unlawfully crosses into another State's territory and employs armed force there (or is otherwise legally responsible for such employment) might also violate international law's prohibition of the use of force, set forth in Article 2(4) of the United Nations Charter and customary international law." Michael N. Schmitt, *Extraterritorial Lethal Targeting: Deconstructing the Logic of International Law*, 52 COLUM. J. TRANSNAT'L L. 77, 79 (2013).

191. Note that Additional Protocol I, *supra* note 10, art. 46 states that "any member of the armed forces of a party to the conflict who falls into the power of an adverse party while engaging in espionage shall not have the right to the status of prisoner of war and may be treated as a spy." *See also* HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 107.

192. Some UAVs currently have an autonomous mode that will allow for them, for example, to complete reconnaissance missions with little-to-no input form a human operator. However, as far as it is possible to know, a use of force has not yet been authorized autonomously.

193. *See, e.g.*, U.S. Dep't. of Def., Directive 3000.9, Autonomy in Weapon Systems (DoD 2012) https://www.esd.whs.mil/portals/54/documents/dd/issuances/dodd/300009p.pdf (stating an

As previously discussed, it is possible that in future military settings one could assign the civilian EAI with PMC status while also holding the military decision-maker responsible for authorizing the deployment of the autonomous UAV to account for its actions.[194] However, one concern with a civilian use of armed EAIs is that PMC status must be intrinsically linked to the identification of an individual who, in authorizing its use, is playing a direct part in hostilities. In other words, the tripartite test would need to be applied to an individual who may be almost impossible to trace.

As previously discussed (under the current interpretation of DPH), should there be doubt as to whether a civilian UAV is a legitimate military target, i.e. due to its nature or location, it must be presumed to be civilian object. In such circumstances, it would be protected against direct attack.[195] Nevertheless, even if the "but for test" was applied in this case (meaning that the armed autonomous UAV could be capable of DPH), according to the ICRC, previous behavior cannot necessarily be used to identify a CCF.

In other words, even where a party to the conflict can positively identify that a civilian, autonomous UAV had previously been used to apply force that had injured or killed civilians and combatants, they still may need to identify that the weapons system is in the process of directly participating in hostilities if they are to attack it lawfully. That cannot be the correct interpretation, and it is perhaps further evidence that the guidance has got it wrong.

C. *Perfidy*

An existing, though niche, debate in the AWS literature is whether an AWS could commit the crime of perfidy, and equally whether an

AWS is a weapons system that "once activated, can select and engage targets without further intervention by a human operator").

194. The current authors support the view of Charles Dunlap, who, in response to Bonnie Docherty (*see* Bonnie Docherty, *Mind the Gap: The Lack of Accountability for Killer Robots*, HUMAN RIGHTS WATCH (April 9, 2015), https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots) argues there is no requirement under international law that an individual must be held to account for the deployment, or, *inter alia*, the potential malfunction of any weapon, autonomous or otherwise. Charles J. Dunlap Jr., *Accountability and Autonomous Weapons: Much Ado About Nothing?* 30 TEMP. INT'L & COMP. L.J. 63, 70–75 (2016).

195. This discussion is focused upon the *jus in bello*, meaning that the UAV in the example cited is imagined upon an existing battlefield. There are of course a number of alternative *jus ad bellum* ways in which the UAV could be lawfully targeted if it were being used for the extraterritorial application of force, not least under Article 51 of the U.N. Charter. For a useful discussion, *see* Philip Alston, *The CIA and Targeted Killings Beyond Borders*, 2 HARV. NAT'L SEC. J. 283 (2011).

AWS could be easily deceived.[196] In the first instance, while it may be difficult to imagine a programmer of an EAI being capable of coding the circumstances in which it should lie, it is nevertheless true that AI is likely to behave in certain ways which are beyond human comprehension.[197] With that in mind, the introduction of EAIs on to the battlefield may mean that concepts such as perfidy need to be re-examined.

In addition, the concept of machine perfidy may in fact be more applicable to a civilian EAI directly participating in hostilities than it would be to an AWS. This is due, in part, to the fact that while an AWS would remain a lawful target so long as it was on the battlefield, a civilian EAIs protected status would return as soon as the act of participation was over, and without having to demonstrate that it was in some way *hors de combat*, which would be a requirement for the equivalent military system. Therefore, if the civilian EAI was behaving in such a way that the combatant believed the act of DPH to be over, he or she must not directly attack it unless, according to the ICRC once again, it could be identified as having a CCF.

D. *Levee en Masse: Lawful Combatancy and POW Status*

The ICRC guidance correctly identifies, there are three mutually exclusive ways in which individuals, and thus potentially EAI's, can be identified under IHL.[198] These are : (i) as civilians; (ii) as combatants, and (iii) as participants in a *levée en masse*.[199] As a result, the remainder of this section examines, albeit briefly, the consequences of classifying

196. Sassoli, *supra* note 3, at 328, where the author notes that "the fascinating question arises as to whether a machine can be "led to believe" something, or whether it is possible to "invite the confidence" of a machine – two elements of the prohibited act of perfidy." *See also* Additional Protocol I, *supra* note 10, art. 37(1) which states, "[i]t is prohibited to kill, injure or capture and adversary by resort to perfidy. Acts inviting the confidence of an adversary to lead him to believe that he is entitled to, or is obliged to accord, protection under the rules of international law applicable in armed conflict, with the intent to betray that confidence, shall constitute perfidy."; Rome Statute, *supra* note 17, art. 8, § 2(b)(xi) ("[k]illing or wounding treacherously individuals belonging to the hostile nation or army" amounts to a "serious [violation] of the laws and customs applicable in international armed conflict"). *See also* HENCKAERTS & DOSWALD-BECK, *supra* note 14, Rule 65 ("[k]illing, injuring or capturing an adversary by resort to perfidy is prohibited").

197. For example, in beating its human competitor, AlphaGo's self-taught winning strategy was so exceptional that it is thought to have surpassed anything ever imagined by a human competitor in centuries of game playing.

198. ICRC Guidance, *supra* note 19, at 21. As noted by the guidance, mutual exclusivity means that all persons must fall in to only one of these three categories.

199. Additional Protocol I, *supra* note 10, art. 50(1), which in turn refers to Geneva Convention III, *supra* note 10, art. 4(A)(1), (2), (3), (6); Additional Protocol I, *supra* note 10, art. 43(1).

an EAI under either of the two remaining classifications, the first of which is combatants. Article 43(2) Additional Protocol I provides,

> [m]embers of the armed forces of a party to a conflict (other than medical personnel and chaplains covered by Article 33 of the Third Convention) are combatants, that is to say, they have the right to participate directly in hostilities.[200]

As previously established, a military EAI is typically referred to as an AWS. Under the current interpretation of IHL, weapons systems are not considered to be combatants and thus they are incapable of independent "participation." In other words, IHL currently implies that an AWS would require a human combatant to initiate or authorize an attack, and thus be the individual capable of participation, even if it does not require an individual to be held lawfully accountable for the actions of an AWS.[201] Nevertheless, as the scenarios in section III demonstrate, autonomous civilian tech is likely to question the current interpretation of DPH, and continued advances in military tech are equally likely to question the very concept of what it means to be a combatant. While in the targeting sense there is little difference between a combatant and a legitimate military target,[202] there may nevertheless be implications if the AWS were to be captured with regard to POW status.

The third and final way a "person" might be classified under IHL, and perhaps one of the most intriguing with regards to EAIs, is as a participant in a *levée en masse*.[203] As noted by Dinstein,[204] GCIII builds, *inter alia*, upon the 1907 revised Hague Convention by supplying that participants of a *levée en masse* are,

> [i]nhabitants of a non-occupied territory, who on the approach of the enemy spontaneously takes up arms to resist the

200. It is nevertheless important to note that IHL does not prohibit members of any of these three classes from directly participating in hostilities. As previously noted, civilians directly participating do not gain combatant privileges. Thus, civilians can, *inter alia*, be prosecuted for national crimes such as treason, arson, and murder. ICRC Guidance, *supra* note 19, at 84.

201. Dunlap, *supra* note 194, at 64.

202. In that they are both lawful targets.

203. According to Additional Protocol I, *supra* note 10, art 51(3) and Additional Protocol II, *supra* note 10, art. 13(3), members of a *levée en masse* are not offered a general protection against direct attack. In other words, IHL does not recognize them as civilians. Although the commentary regarding DPH does tend to identify this, it does not, as yet, consider the concept in relation to EAIs. *See* ICRC Guidance, *supra* note 19, at 25; Schmitt *supra* note 29, at 704.

204. DINSTEIN, *supra* note 38, ¶¶ 48–49.

> invading forces, without having had time to form themselves into regular armed units, providing they carry arms openly and respect the laws and customs of law.[205]

Boothby notes, "[m]embers of a *levée en masse* are civilians who are regarded as belligerents (combatants in modern parlance),"[206] even if the concept is, by its own definition, relatively short-lived.[207] Consequently, civilians who fall under this classification *are* extended combatant privileges including "immunity from domestic prosecution for acts which, although in accordance with IHL, may constitute crimes under national criminal law . . . ."[208]

A question that must be considered is, if a future EAI were indeed capable of direct participation in hostilities in much the same way as a human, could a number of future EAIs also constitute a *levée en masse*? If so, what would be the threshold as to the number of units that would be required, and furthermore, how would autonomous swarming systems be classified?[209] Although it may matter little that an EAI is immune from civilian prosecution, combatant status means that the EAI would become a lawful target to whom the targeteer owes no duty to demonstrate that it is playing a direct part in hostilities. It matters little as the targeteer would be in the same position as if he or she were engaging a legitimate military target. However, legitimate military targets have few, if any, post-attack rights.

The primary difference from the classification of civilian is that an AWS (should it be classified as a combatant) and an EAI (should it prove capable of participating in a *levée en masse*) should perhaps both be entitled to POW status upon capture. The consequences of this may, *inter alia*, mean that a captured EAI/AWS would be protected against "physical mutilation or . . . medical or scientific experiments of any kind,"[210] as well as a number of other provisions, including the right to

205. Geneva Convention III, *supra* note 10, art. 4(A)(6), replicated, in Geneva Convention I, *supra* note 5, art.13 and Geneva Convention II, *supra* note 5, art. 13.

206. Bill Boothby, *And for Such Time as: The Time Dimension to Direct Participation in Hostilities*, 42 N.Y.U. J. INT'L L. & POL. 741, 754 n. 48 (2010).

207. DINSTEIN, *supra* note 38, ¶ 155.

208. ICRC Guidance, *supra* note 19, at 83–84.

209. There are a number of ways in which swarming systems might operate. However, one example is where a centralized "queen" controls the actions of the wider group. The group could, in theory, be an almost infinitesimal number of additional units. For a useful discussion see generally, Grimal & Sundaram, *supra* note 17, at 128.

210. Geneva Convention III, *supra* note 10, art. 13.

be "repatriated without delay after the cessation of active hostilities."[211] Potentially, this could also extend to include the confiscation and reprogramming of EAIs; whether or not the reprogramming of an enemy military EAI to change allegiance may also constitute a war crime remains in need of discussion.

One benefit of classifying EAIs in such a way is that it could contribute to the prevention of the proliferation of autonomous technologies by making the reverse engineering of AI tech when captured unlawful, which is a concern of a number of those in opposition to AWS.[212] Clearly, much will depend of course on whether international norms do indeed recognize such rights.

This section has considered, albeit briefly, a number of the traditional concepts associated with armed conflict in relation to the introduction of advanced AWS and civilian EAIs. The list is by no means intended to be exhaustive but is rather supplied in order to acknowledge the implicit complexities that will be presented once *civilian property* becomes more akin to the definition of a *civilian* due, *inter alia*, to EAIs complex decision-making capabilities. Having established the difficulties surrounding concepts such as PMCs, spies, and dual-use technologies, the section identified that the introduction of advanced AI technologies will only add to the opaqueness of a number of IHL concepts, particularly where states are able to plausibly deny their involvement. However, if all parties to the conflict were to renew their efforts to adhere to the principles of IHL, and were also able to identify methods of interpretation that allowed, in certain circumstances, for civilian property to be given civilian status under IHL, then there is an opportunity not only to remove humans from a number of battlefield environments but also to prevent the proliferation of advanced AI technologies.

V. Conclusion

Currently, IHL recognizes a clear distinction between *civilians* and *civilian property*. Nonetheless, in the majority of targeting situations, where there is doubt as to the precise nature of an object or person, both are presumed to have a civilian status. Accordingly, the object or person is protected against direct attack. As supported by the Israeli Supreme Court and existing state practice, where there is doubt as to

211. Geneva Convention III, *supra* note 10, art. 118; Henckaerts & Doswald-Beck, *supra* note 14, Rule 128.

212. *See, e.g.*, Int'l Comm. Red Cross, Expert Meeting Report, Autonomous Weapon Systems: Technical, Military, Legal And Humanitarian Aspects 24 (2014).

the status of a civilian suspected of DPH, there is no such presumption. Contrary to the ICRC's position, this avoids placing too great a burden of risk upon the lawful combatant.

Although the current authors support this position, this Article has nonetheless demonstrated a lacuna with respect to emerging technologies. Though this may be negligible by contemporary standards, it looks set to grow as the increasingly advanced autonomous technologies such as those identified in Part III are developed and proliferated. This Article contends that by considering the ICRC's guidance in light of emerging AI technologies, many of the emotions that are typically implicit when considering DPH can be negated.

As demonstrated in Part III, the incremental stages of development of autonomous technologies poses the greatest challenge to current legal thinking. In order to consider the scenarios in Part III, the authors introduced the novel inclusion of a "Turing-like" "but-for test". By way of strategic overview, the Scenarios examined and concluded the following:

Scenario 1 examined the concept of DPH alongside established and relatively simple AI systems. The authors concluded that in such a scenario existing tech would be unable to "participate" without human involvement. Scenario 2 considered whether the imminent introduction of autonomous vehicles could affect DPH assessments upon the battlefield. Notably, the status of the "EAI truck" in terms of temporal and geographical proximity to the battlefield was examined. This "broader" discussion featured heavily in expert meetings, the consequent ICRC guidance, and the responses to it.[213] Here, the authors highlighted the importance of a "case-by-case" analysis to determine DPH. Nevertheless, even if they could be identified as legitimate military targets, the fact remains that autonomous vehicles are likely to incapable of actual participation.

Scenario 3 delved a little further into the future, and imagined an autonomous future healthcare EAI.[214] The essence of this Scenario underlined that the ICRC's imbalance is in favor of non-combatants over combatants in 2020 as much as it is prospectively in 2220. Here, the

213. *See* ICRC Guidance, *supra* note 19, at 56. Schmitt also notes for example that during the expert meetings, consensus as to who should be considered as participating was sometimes quickly reached, while sometimes it was not. With regards to the latter, he offers that the "paradigmatic example being that of a civilian driving an ammunition truck in a combat zone." Schmitt, *supra* note 29, at 710–11.

214. SOFTBANK ROBOTICS, *supra* note 147. While in its current incarnation, Pepper cannot be considered as a fully autonomous healthcare robot, there can be no doubt that it is at least possible that future robots will become more capable, and more autonomous.

authors concluded that the "wider" interpretation was therefore preferable, and indeed, the only reasonable approach to share the burden. Scenario 4 examined whether or not EAIs could go on to become "revolving door fighters", to have a "CCF," and whether the ICRC's inclusion of such concepts can even be justified. The conclusion drawn was that the "revolving door" EAI does not sit comfortably with the ICRC's present guidance. Although the ICRC entertained the "revolving door" concept, it is more readily tangible to a human civilian. In the case of the CCFing/Revolving Door/DPHing EAI, the attacking combatant would have little hesitation in targeting and engaging the robot. Ultimately, this would underscore the deficiency within the guidance.

If future EAIs are considered capable of participation in hostilities directly and/or indirectly, then a raft of other legal provisions may come in to play, not least the question as to whether EAIs should have a claim to POW status in situations where the equivalent human would. While in 2020 EAI systems seem, on the whole, a little cumbersome, somewhat simple, and extremely limited in their *consciousness*, in the year 2220, it is likely to be a very different story. Nevertheless, if machines do ever become capable of making *informed* decisions, the transformation will not present itself overnight.

Instead, technological developments will be incremental and often inconspicuous, which means that in war, as elsewhere in life, it may be extremely difficult to separate the fully conscious embodied AGI from a very intelligent, though perhaps convincing, robot. Whether or not humankind will ever be capable of mechanically replicating itself remains to be seen. Should EAI's reach a point in the future where they become impossible to distinguish from humans on the battlefield, the *only* logical conclusion is that that they are treated identically, i.e. as if they are *both* civilians.